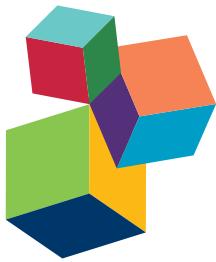


# APPLICATIONS OF STEM (SCIENCE, TECHNOLOGY, ENGINEERING AND MATHEMATICS) TOOLS IN MICROBIOLOGY OF INFECTIOUS DISEASES

EDITED BY: Julio Alvarez and Andres M. Perez

PUBLISHED IN: Frontiers in Microbiology



# frontiers

## **Frontiers Copyright Statement**

© Copyright 2007-2017 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

**ISSN 1664-8714**

**ISBN 978-2-88945-183-8**

**DOI 10.3389/978-2-88945-183-8**

## **About Frontiers**

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## **Frontiers Journal Series**

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## **Dedication to Quality**

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## **What are Frontiers Research Topics?**

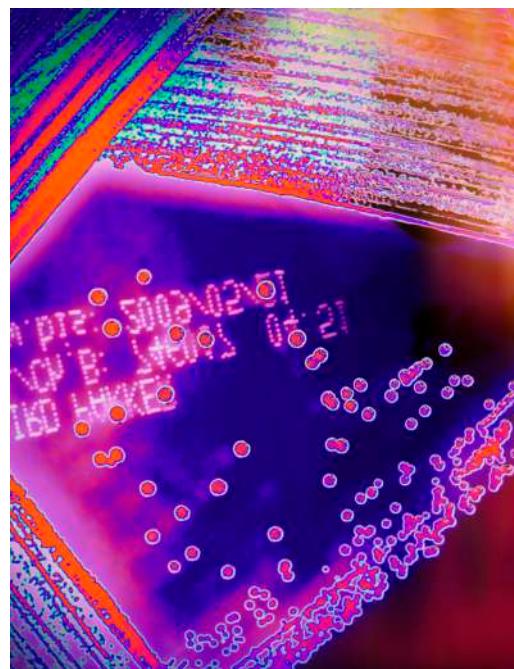
Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: [researchtopics@frontiersin.org](mailto:researchtopics@frontiersin.org)

# APPLICATIONS OF STEM (SCIENCE, TECHNOLOGY, ENGINEERING AND MATHEMATICS) TOOLS IN MICROBIOLOGY OF INFECTIOUS DISEASES

Topic Editors:

**Julio Alvarez**, University of Minnesota, USA

**Andres M. Perez**, University of Minnesota, USA



Bacterial growth in solid agar. VISAVENT Health Surveillance Centre. Universidad Complutense de Madrid (UCM).

Epidemiology is a discipline intended to systematically investigate, and ideally quantify, disease dynamics in populations (Perez, 2015). Epidemiological assessments may be divided into four large areas, namely, (a) identification and characterization of a pathogen, (b) development of systems for detection of cases, (c) descriptive epidemiology and quantification of disease patterns, and (d) advanced analytical methods to design intervention strategies. Briefly, there is an initial need for understanding the pathogeny of a disease and condition, which may also include experimental studies and development of new models of infection and proliferation under different conditions. Subsequently, such knowledge may be applied to support the identification of cases, which typically includes the design, evaluation, and validation of diagnostic

tests. Disease may then be quantified in a population, leading to the identification of patterns and application of molecular characterization techniques to understand disease spread, and ultimately to identify factors preventing or promoting disease. Finally, those factors may be incorporated into advanced quantitative methods and epidemiological models, which are used to design and evaluate strategies aimed at preventing, controlling, or eliminating disease in the population.

Recent years have seen a dramatic increase in the application of science, technology, engineering, and mathematical (STEM) tools and approaches intended to enhance such analytical epidemiology process, with the ultimate goal of supporting disease prevention, control, and eradication. This eBook comprises a series of research articles that, through current state-of-the-art scientific knowledge on the application of STEM tools to the microbiology of infectious diseases, demonstrate their usefulness at the various components of an integral epidemiological approach, divided into the four large components of (a) experimental studies, (b) novel diagnostic techniques, (c) epidemiological characterization, and (d) population modeling and intervention.

**Citation:** Alvarez, J., Perez, A. M., eds. (2017). Applications of STEM (Science, Technology, Engineering and Mathematics) Tools in Microbiology of Infectious Diseases. Lausanne: Frontiers Media. doi: 10.3389/978-2-88945-183-8

# Table of Contents

- 07 Editorial: Applications of STEM (Science, Technology, Engineering and Mathematics) Tools in Microbiology of Infectious Diseases**  
Julio Alvarez and Andres M. Perez
- 1. Experimental studies**
- 10 Prophylactic Use of Ganoderma lucidum Extract May Inhibit Mycobacterium tuberculosis Replication in a New Mouse Model of Spontaneous Latent Tuberculosis Infection**  
Lingjun Zhan, Jun Tang, Shuzhu Lin, Yanfeng Xu, Yuhuan Xu and Chuan Qin
- 19 Proteomics Analysis of Three Different Strains of Mycobacterium tuberculosis under In vitro Hypoxia and Evaluation of Hypoxia Associated Antigen's Specific Memory T Cells in Healthy Household Contacts**  
Santhi Devasundaram, Akilandeswari Gopalan, Sulochana D. Das and Alamelu Raja
- 33 Comparative proteomic analysis of extracellular proteins expressed by various clonal types of Staphylococcus aureus and during planktonic growth and biofilm development**  
Salman S. Atshan, Mariana N. Shamsudin, Zamberi Sekawi, Leslie T. Thian Lung, Fatemeh Barantalab, Yun K. Liew, Mateg Ali Alreshidi, Salwa A. Abduljaleel and Rukman A. Hamat
- 42 In vitro and in vivo analysis of antimicrobial agents alone and in combination against multi-drug resistant Acinetobacter baumannii**  
Songzhe He, Hui He, Yi Chen, Yueming Chen, Wei Wang and Daojun Yu
- 53 Pharmacokinetic/Pharmacodynamic Correlation of Cefquinome Against Experimental Catheter-Associated Biofilm Infection Due to Staphylococcus aureus**  
Yu-Feng Zhou, Wei Shi, Yang Yu, Meng-Ting Tao, Yan Q. Xiong, Jian Sun and Ya-Hong Liu
- 63 RNA-seq de novo Assembly Reveals Differential Gene Expression in Glossina palpalis gambiensis Infected with Trypanosoma brucei gambiense vs. Non-Infected and Self-Cured Flies**  
Illiassou Hamidou Soumana, Christophe Klopp, Sophie Ravel, Ibouriyamine Nabihoudine, Bernadette Tchicaya, Hugues Parrinello, Luc Abate, Stéphanie Rialle and Anne Geiger
- 2. Novel diagnostic techniques**
- 82 Giant Magnetoresistance-based Biosensor for Detection of Influenza A Virus**  
Venkatramana D. Krishna, Kai Wu, Andres M. Perez and Jian-Ping Wang

- 90 Rapid Identification and Multiple Susceptibility Testing of Pathogens from Positive-Culture Sterile Body Fluids by a Combined MALDI-TOF Mass Spectrometry and Vitek Susceptibility System**  
Yueru Tian, Bing Zheng, Bei Wang, Yong Lin and Min Li
- 106 Rapid Detection of K1 Hypervirulent Klebsiella pneumoniae by MALDI-TOF MS**  
Yonglu Huang, Jiaping Li, Danxia Gu, Ying Fang, Edward W. Chan, Sheng Chen and Rong Zhang
- 113 Assessment of MALDI-TOF MS as alternative tool for Streptococcus suis identification**  
Marta Pérez-Sancho, Ana Isabel Vela, Teresa García-Seco, Marcelo Gottschalk, Lucas Domínguez and José Francisco Fernández-Garayzábal
- 119 Evaluation of a High-Intensity Green Fluorescent Protein Fluorophage Method for Drug- Resistance Diagnosis in Tuberculosis for Isoniazid, Rifampin, and Streptomycin**  
Xia Yu, Yunting Gu, Guanglu Jiang, Yifeng Ma, Liping Zhao, Zhaogang Sun, Paras Jain, Max O'Donnell, Michelle Larsen, William R. Jacobs Jr and Hairong Huang
- 125 Detection of Mycobacterium tuberculosis in extrapulmonary biopsy samples using PCR targeting IS6110, rpoB, and nested-rpoB PCR Cloning**  
Hossein Meghdadi, Azar D. Khosravi, Ata A. Ghadiri, Amir H. Sina and Ameneh Alami
- 133 Survey and Visual Detection of Zaire ebolavirus in Clinical Samples Targeting the Nucleoprotein Gene in Sierra Leone**  
Huan Li, Xuesong Wang, Wei Liu, Xiao Wei, Weishi Lin, Erna Li, Puyuan Li, Derong Dong, Lifei Cui, Xuan Hu, Boxing Li, Yanyan Ma, Xiangna Zhao, Chao Liu and Jing Yuan
- 140 Corrigendum: Survey and Visual Detection of Zaire Ebolavirus in Clinical Samples Targeting the Nucleoprotein Gene in Sierra Leone**  
Huan Li, Xuesong Wang, Wei Liu, Xiao Wei, Weishi Lin, Erna Li, Puyuan Li, Derong Dong, Lifei Cui, Xuan Hu, Boxing Li, Yanyan Ma, Xiangna Zhao, Chao Liu and Jing Yuan
- 141 A Novel Typing Method for Streptococcus pneumoniae Using Selected Surface Proteins**  
Arnaud Domenech, Javier Moreno, Carmen Ardanuy, Josefina Liñares, Adela G. de la Campa and Antonio J. Martín-Galiano

### 3. Epidemiological characterization

- 150 Prevalence of Escherichia coli Virulence Genes in Patients with Diarrhea and a Subpopulation of Healthy Volunteers in Madrid, Spain**  
Adriana Cabal, María García-Castillo, Rafael Cantón, Christian Gortázar, Lucas Domínguez and Julio Álvarez
- 156 Phenotypic and Genetic Heterogeneity in Vibrio cholerae O139 Isolated from Cholera Cases in Delhi, India during 2001–2006**  
Raikamal Ghosh, Naresh C. Sharma, Kalpataru Halder, Rupak K. Bhadra, Goutam Chowdhury, Gururaja P. Pazhani, Sumio Shinoda, Asish K. Mukhopadhyay, G. Balakrish Nair and Thadavarayan Ramamurthy
- 165 Prevalence and genetic diversity of clinical Vibrio parahaemolyticus isolates from China, revealed by multilocus sequence typing scheme**  
Dongsheng Han, Hui Tang, Chuanli Ren, Guangzhou Wang, Lin Zhou and Chongxu Han

**173 Sero-Prevalence and Genetic Diversity of Pandemic V. parahaemolyticus Strains Occurring at a Global Scale**

Chongxu Han, Hui Tang, Chuanli Ren, Xiaoping Zhu and Dongsheng Han

**4. Population modeling and intervention**

**183 Evaluation of the risk factors contributing to the African swine fever occurrence in Sardinia, Italy**

Beatriz Martínez-López, Andres M. Perez, Francesco Feliziani, Sandro Rolesu, Lina Mur and José M. Sánchez-Vizcaíno

**193 Applications of Bayesian Phylodynamic Methods in a Recent U.S. Porcine Reproductive and Respiratory Syndrome Virus Outbreak**

Mohammad A. Alkhamis, Andres M. Perez, Michael P. Murtaugh, Xiong Wang and Robert B. Morrison

**203 Advances and Limitations of Disease Biogeography Using Ecological Niche Modeling**

Luis E. Escobar and Meggan E. Craft



# Editorial: Applications of STEM (Science, Technology, Engineering and Mathematics) Tools in Microbiology of Infectious Diseases

*Julio Alvarez\** and *Andres M. Perez*

Department of Veterinary Population Medicine, College of Veterinary Medicine, University of Minnesota, Saint Paul, MN, USA

**Keywords:** STEM, quantitative methods, epidemiology, modeling, pathogen detection

## Editorial on the Research Topic

Applications of STEM (Science, Technology, Engineering and Mathematics) Tools in Microbiology of Infectious Diseases

## INTRODUCTION

Epidemiology is a discipline intended to systematically investigate, and ideally quantify, disease dynamics in populations (Perez, 2015). Epidemiological assessments may be divided into four large areas, namely, (a) identification and characterization of a pathogen, (b) development of systems for detection of cases, (c) descriptive epidemiology and quantification of disease patterns, and (d) advanced analytical methods to design intervention strategies. Briefly, there is an initial need for understanding the pathogenicity of a disease and condition, which may also include experimental studies and development of new models of infection and proliferation under different conditions. Subsequently, such knowledge may be applied to support the identification of cases, which typically includes the design, evaluation, and validation of diagnostic tests. Disease may then be quantified in a population, leading to the identification of patterns and application of molecular characterization techniques to understand disease spread, and ultimately to identify factors preventing or promoting disease. Finally, those factors may be incorporated into advanced quantitative methods and epidemiological models, which are used to design and evaluate strategies aimed at preventing, controlling, or eliminating disease in the population.

Recent years have seen a dramatic increase in the application of science, technology, engineering, and mathematical (STEM) tools and approaches intended to enhance such analytical epidemiology process, with the ultimate goal of supporting disease prevention, control, and eradication. The research topic here provides an update of the current state-of-the-art scientific knowledge of the application of STEM tools to the microbiology of infectious diseases, at the various components of an integral epidemiological approach, divided into the four large components of (a) experimental studies, (b) novel diagnostic techniques, (c) epidemiological characterization, and (d) population modeling and intervention.

## EXPERIMENTAL STUDIES

The increased knowledge gained in the last decades on the importance of host-pathogen and environment-pathogen interactions has provided a better understanding on the complexities associated with the design of effective therapeutic and preventive strategies against infectious

## OPEN ACCESS

### Edited by:

Vitaly V. Ganusov,  
University of Tennessee, USA

### Reviewed by:

Kirsten McCabe,  
Los Alamos National Laboratory, USA

### \*Correspondence:

Julio Alvarez  
jalvarez@umn.edu

### Specialty section:

This article was submitted to  
Infectious Diseases,  
a section of the journal  
*Frontiers in Microbiology*

**Received:** 19 November 2016

**Accepted:** 31 January 2017

**Published:** 14 February 2017

### Citation:

Alvarez J and Perez AM (2017)  
Editorial: Applications of STEM (Science, Technology, Engineering and Mathematics) Tools in Microbiology of Infectious Diseases.  
*Front. Microbiol.* 8:215.  
doi: 10.3389/fmicb.2017.00215

agents, but also raised new questions on how to evaluate those potential new approaches in a capturing their complexity. Six papers looking into the use of STEM tools for the design of experimental models for replicating this environment and host-pathogen interactions and ultimately the assessment of alternative strategies to prevent pathogen replication were selected here. Those papers were focused on emerging and/or neglected pathogens whose control is impaired due, at least in part, to its fastidious nature (*Mycobacterium tuberculosis*) (Zhan et al.; Devasundaram et al.), its ability to form environmentally resistant structures such as biofilms and/or resist common therapeutic approaches (*Staphylococcus aureus* and *Acinetobacter baumannii*) (Atshan et al.; He et al.; Zhou et al.) or the lack of effective treatments to prevent or treat infection (*Trypanosoma brucei gambiense*) (Hamidou Soumana et al.). An *in-vitro* dormancy model was used to study genes that were overexpressed under hypoxia by *M. tuberculosis* strains recovered from disease outbreaks (as opposed to *M. tuberculosis* laboratory strains) (Devasundaram et al.). Nevertheless, complementation of evidences found *in-vitro* with results from *in-vivo* models is important because different results may be obtained in those alternative settings. As an example, murine models were used to assess antibacterial activities of a panel of 14 antimicrobial agents against multi-drug resistant *A. baumannii* (He et al.). Using an alternative approach the efficacy of a fourth generation cephalosporin (cefquinome) against planktonic and biofilm *S. aureus* cells was evaluated in another study (Zhou et al.). A mouse model was also of critical importance to mimic human latent tuberculosis infection and assess the efficacy of a fungus (*Ganoderma lucidum*) against *M. tuberculosis* H37Rv (Zhan et al.). Finally, protein and gene expression found in a pathogen (*S. aureus*) and a vector (the tsetse fly *Glossina palpalis gambiensis*, vector of *T. brucei gambiense*) were analyzed in two studies that both considered different conditions [planktonic growth vs. biofilm development for *S. aureus* (Atshan et al.) and infected vs. non-infected specimens for the *G. palpalis* (Hamidou Soumana et al.)].

## NOVEL DIAGNOSTIC TECHNIQUES

Eight papers were selected demonstrating the use of novel diagnostic techniques or the evaluation of established approached for the detection of global emergencies (Huang et al.; Li et al.; Meghdadi et al.; Pérez-Sancho et al.; Domenech et al.; Krishna et al.; Tian et al.; Yu et al.). The combined use of magnetic nanoparticles (MNPs) and giant magnetoresistance (GMR) biosensors has recently gained attention because of their potential for simultaneous, rapid and affordable detection of multiple pathogens. Monoclonal antibodies in combination with MNPs and GMR biosensors were used to detect influenza virus in swine, with a potential application to nasal samples, which are routinely collected by the swine industry worldwide (Krishna et al.). Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry is becoming increasingly popular. In the research topic here, applications of MALDI-TOF has been demonstrated in three case-studies,

including antimicrobial susceptibility testing in positive body fluid cultures collected from a hospital in Shanghai, China (Tian et al.), characterization of hypervirulent strains of *Klebsiella pneumoniae* (Huang et al.), and identification of *Streptococcus suis* isolates obtained from multiple host species (Pérez-Sancho et al.). Finally, rapid detection of emergent diseases and conditions of global importance require of novel, affordable tools that could potentially be applied at large scale in developing settings to early detect disease cases. Here, it was presented (1) a novel method using mycobacteriophage φ2GFP10, which was evaluated for detecting drug resistance in clinical isolates of *Mycobacterium tuberculosis* (Yu et al.), (2) nested PCR for the detection of *Mycobacterium tuberculosis* DNA in patients with extra pulmonary tuberculosis (Meghdadi et al.), (3) a reverse transcription loop-mediated isothermal amplification (RT-LAMP) method to detect Zaire ebolavirus in Sierra Leone using the nucleoprotein gene as a target sequence (Li et al.), and a novel typing method for *Streptococcus pneumoniae* using selected surface proteins (Domenech et al.).

## EPIDEMIOLOGICAL CHARACTERIZATION

Once the tools for detection and characterization of the pathogen of interest have been developed the challenge shifts to the interpretation of the results obtained in different populations. Four papers focusing on three bacterial pathogens (*Escherichia coli*, *Vibrio parahaemolyticus*, and *V. cholerae*) that exemplify these challenges and include the application of molecular biology methods were selected (Han et al.; Cabal et al.; Ghosh et al.; Han et al.). Use of molecular characterization techniques allow identification of the pathogen beyond the species level, thus offering a much higher degree of refinement that can help to understand better the epidemiology of a disease (Muellner et al., 2011) and clarify the biological relationship between different strains. In addition, when used directly on clinical samples, it may help to improve the sensitivity of the diagnostic approach used eliminating the need for isolation of the pathogen, what can be particularly useful when its concentration is low and there are other microorganisms that could outgrow it. For example, a set of specific real-time (RT) PCRs aiming at a panel of virulence genes characteristic of different *E. coli* pathotypes was used to assess their distribution and, when detected, quantify the number of copies present, in samples from clinical human cases and healthy volunteers in Spain and thus provide qualitative and quantitative information on the significance of their detection (Cabal et al.). Detection of virulence genes, coupled with other molecular (DNA sequencing, ribotyping, restriction fragment length polymorphism—RFLP—and pulsed field gel electrophoresis—PFGE) and conventional (serotyping, antimicrobial susceptibility determination) was also used for a detailed characterization of a collection of *V. cholerae* strains recovered in a 5-year period in India, revealing a considerable diversity and thus a dynamic evolution of the pathogen that must be taken into account when evaluating individual isolates (Ghosh et al.). Finally, a well-established molecular characterization technique, multi-locus sequence typing (MLST), was used to

assess the genetic diversity of *V. parahaemolyticum* at the national (China) (Han et al.) and, for pandemic strains (those harboring a given set of virulence markers), global levels (Han et al.) by analyzing two collections of clinical and environmental isolates. While the first study demonstrated a high genetic diversity among strains recovered from the same country, comparable to what had been described at a global scale (Han et al.), the second evidenced the persistence of certain pandemic strains in different countries and their occasional detection in environmental samples (Han et al.).

## POPULATION MODELING AND INTERVENTION

Population modeling for the characterization of spatial risk is prerequisite for the design and implementation of prevention and control interventions in a region. Three techniques (Bayesian analysis, ecological niche modeling, and phylogeography tools) have been demonstrated in this research topic (Martínez-López et al.; Alkhamis et al.; Escobar and Craft). Bayesian techniques have the potential to support the characterization of disease risk in a region because of their flexibility to adapt the coding to different data structure and variables distribution. A Bayesian multivariable logistic regression mixed model was used to assess the relation between hypothesized risk factors and African swine fever virus (ASFV) distribution in Sardinia (Italy) after the beginning of the eradication program in 1993 (Martínez-López et al.). Additionally, one of the challenges epidemiologists often face is the absence of control or population data. Ecological niche modeling is a tool that allows quantification of spatial risk using presence-only (case) data. An overview of the background, history, and conceptual framework of ecological niche modeling applied to epidemiology

and public health was presented (Escobar and Craft). Finally, the use of Bayesian phylodinamic models was demonstrated using a large volume of porcine reproductive and respiratory syndrome virus (PRRSV) collected in the United States (Alkhamis et al.).

## FINAL REMARKS

The acknowledgment that health of humans, animals, and the environment is inter-connected has resulted in the establishment of the One Health approach (Perez, 2015). Cross-cutting a myriad of pathogens, host species, and settings, and under the One Health umbrella, novel STEM tools and methods have a tremendous potential to increase our ability to understand pathogens dynamics, improve effectiveness of detection, characterize the epidemiological setting in which disease spread, and, ultimately, develop effective strategies to mitigate or eliminate the impact of disease.

## AUTHOR CONTRIBUTIONS

JA and AP co-edited the Research Topic and wrote this editorial.

## ACKNOWLEDGMENTS

The authors would like to thank the authors that submitted their work to this Research Topic, the reviewers that critically evaluated these and the Frontiers Editorial Office for their help producing it. The authors would also like to thank the crew of the University of Minnesota Science Technology Engineering and Mathematics for Minnesota Advancement (STEMMA) laboratory for the support in the organization of this editorial and research topic.

## REFERENCES

- Muellner, P., Zadoks, R. N., Perez, A. M., Spencer, S. E., Schukken, Y. H., and French, N. P. (2011). The integration of molecular tools into veterinary and spatial epidemiology. *Spat. Spatiotemporal Epidemiol.* 2, 159–171. doi: 10.1016/j.sste.2011.07.005
- Perez, A. M. (2015). Past, Present, and Future of Veterinary Epidemiology and Economics: One Health, Many Challenges, No Silver Bullets. *Front. Vet. Sci.* 2:60. doi: 10.3389/fvets.2015.00060

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial

relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Alvarez and Perez. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Prophylactic Use of *Ganoderma lucidum* Extract May Inhibit *Mycobacterium tuberculosis* Replication in a New Mouse Model of Spontaneous Latent Tuberculosis Infection

Lingjun Zhan <sup>†</sup>, Jun Tang <sup>†</sup>, Shuzhu Lin, Yanfeng Xu, Yuhuan Xu and Chuan Qin <sup>\*</sup>

Key Laboratory of Human Diseases and Comparative Medicine, Ministry of Health, Institute of Laboratory Animal Science, Peking Union Medical College, Chinese Academy of Medical Sciences and Comparative Medicine Center, Beijing, China

## OPEN ACCESS

**Edited by:**

Andres M. Perez,  
University of Minnesota, USA

**Reviewed by:**

Michelle Martinez-Montemayor,  
Universidad Central del Caribe,

Puerto Rico

Paras Jain,

Albert Einstein College of Medicine,  
USA

**\*Correspondence:**

Chuan Qin  
qinchuan@pumc.edu.cn

<sup>†</sup>Co-first authors.

**Specialty section:**

This article was submitted to  
Infectious Diseases,  
a section of the journal  
*Frontiers in Microbiology*

**Received:** 14 May 2015

**Accepted:** 10 December 2015

**Published:** 08 January 2016

**Citation:**

Zhan L, Tang J, Lin S, Xu Y, Xu Y and Qin C (2016) Prophylactic Use of *Ganoderma lucidum* Extract May Inhibit *Mycobacterium tuberculosis* Replication in a New Mouse Model of Spontaneous Latent Tuberculosis Infection. *Front. Microbiol.* 6:1490.  
doi: 10.3389/fmicb.2015.01490

A mouse model of spontaneous latent tuberculosis infection (LTBI) that mimics LTBI in humans is valuable for drug/vaccine development and the study of tuberculosis. However, most LTBI mouse models require interventions, and a spontaneous LTBI mouse model with a low bacterial load is difficult to establish. In this study, mice were IV-inoculated with 100 CFU *Mycobacterium tuberculosis* H37Rv, and a persistent LTBI was established with low bacterial loads (0.5~1.5 $\log_{10}$  CFU in the lung; <4 $\log_{10}$  CFU in the spleen). Histopathological changes in the lung and spleen were mild during the first 20 weeks post-inoculation. The model was used to demonstrate the comparative effects of prophylactic and therapeutic administration of *Ganoderma lucidum* extract (spores and spores lipid) in preventing H37Rv replication in both lung and spleen. H37Rv was inhibited with prophylactic use of *G. lucidum* extract relative to that of the untreated control and therapy groups, and observed in the spleen and lung as early as post-inoculation week 3 and week 5 respectively. H37Rv infection in the therapy group was comparable to that of the untreated control mice. No significant mitigation of pathological changes was observed in either the prophylactic or therapeutic group. Our results suggest that this new LTBI mouse model is an efficient tool of testing anti-tuberculosis drug, the use of *G. lucidum* extract prior to *M. tuberculosis* infection may protect the host against bacterial replication to some extent.

**Keywords:** mouse model, spontaneous latent tuberculosis infection, *Ganoderma lucidum* extract, prophylactic use, inhibited *Mycobacterium tuberculosis* replication

## INTRODUCTION

Approximately one third of the world's population is infected with *Mycobacterium tuberculosis*, and about 90% of these infections are latent. A better understanding of mechanisms leading to latency is required to help prevent, control, treat, and eliminate tuberculosis infection and disease.

Similar to other latent infections, *M. tuberculosis* in latent tuberculosis infection (LTBI) replicates at low levels in the infected organs, and therefore histopathology results

are absent or mild. There are several established LTBI animal models, including mice (Scanga et al., 1999), guinea pigs (Kashino et al., 2008), rats (Singhal et al., 2011), rabbits (Manabe et al., 2008), and nonhuman primates (Lin et al., 2009a). These models have been used to identify host factors that contribute to the establishment and maintenance of *M. tuberculosis* latency, and reactivation of replication. One of the most recognized LTBI mouse models is the Cornell model, which is created by inhibiting *M. tuberculosis* replication with intervening factors (Lenaerts et al., 2004; Woolhiser et al., 2007). However, there is no currently available mouse model of spontaneously paucibacillary tuberculosis.

The three primary parameters used for evaluating vaccine or drug candidate against tuberculosis are the bacterial load, pathological change in the lung, and the tuberculosis relapse rate in the latter phase of LTBI. With the current LTBI mouse model, the period of latency is relatively long, and to complete the evaluation can require 3–7 months (Ziv et al., 2001; Ha et al., 2003; Zhang et al., 2011). Thus, the current models are inefficient and costly, and slow the development of new tuberculosis vaccines and drug treatments.

A previous study showed that a proteoglycan extracted from the fruiting bodies of the bracket (polypore) fungus *G. lucidum* could be a preventative of diabetic complications (Pan et al., 2013). Furthermore, triterpenes of *G. lucidum* were shown to exert anti-lung cancer activity *in vitro* and *in vivo*, and such anti-cancer activity was mediated by enhancing immunomodulation and induction of cellular apoptosis (Feng et al., 2013). Also of note is that polysaccharides purified from the submerged culture of *G. formosanum* can activate macrophages and protect mice against *Listeria monocytogenes* infection (Wang et al., 2011). *G. lucidum* can also regulate natural killer cells (Chien et al., 2004), macrophages (Yeh et al., 2010), T cells (Lai et al., 2010; Yoshida et al., 2012), and dendritic cells (Meng et al., 2011), all of which are actively involved in the innate and adaptive immune responses to *M. tuberculosis* infection. Thus, we wondered whether *G. lucidum* could protect the host from *M. tuberculosis* infection.

To test our hypothesis, in the present study we established a novel mouse model of spontaneous LTBI, with a shorter latency period and lower bacterial load than previous models. We then evaluated the protective effects against *M. tuberculosis* infection in this new model exerted by the Takayama *G. Lucidum* (MeiShanTang, Hong Kong), which is rich in triterpenes, and is proven effective in treating simian acquired immune deficiency syndrome via the immune system (Lu et al., 2011).

## MATERIALS AND METHODS

### Ethics Statement

The Institute of Animal Use and Care Committee of the Institute of Laboratory Animal Science, Peking Union Medical College approved all protocols and procedures that involved animals (ILAS-PC-2013-015). All mice were housed in plastic cages (6/cage) with free access to drinking water and a pellet diet, under controlled humidity ( $50 \pm 10\%$ ), light (12/12 h light/dark cycle),

and temperature ( $23 \pm 2^\circ\text{C}$ ) conditions in an Animal Biosafety Level 3 (ABSL-3) facility. Prior to procedures performed at each time-point, mice were fasted overnight and then given anesthesia.

### Establishment of the LTBI Mouse Model

#### Bacterial Strain

The *M. tuberculosis* strain H37Rv was first cultured in Löwenstein-Jensen plates (L-J plates) for 3 weeks, to mid-log phase. Then the bacteria were harvested using a sterilized L-shaped glass rod and resuspended in 0.9% NaCl in a glass grinder. The suspension was filtered through a  $5\ \mu\text{m}$  membrane and the bacterial density adjusted to  $\sim 1 \times 10^7$  colony-forming units (CFU)/mL. An aliquot of bacilli were then plated on an L-J plate for CFU enumeration.

#### Mice

Specific-pathogen free female C57BL/6 mice (6–8 weeks old,  $n = 450$ ) were obtained from Vital River Laboratory Animal Technology (China). Mice were maintained in an ABSL-3 specific pathogen-free facility and allowed food and water *ad libitum*.

#### Inoculation of Mice with *M. tuberculosis*

The mice were randomly and equally apportioned to an infection or a blank control group, 150 mice for each experiment, 75 mice for infection and blank control group each, the experiment would be repeated 3 times with 1 week interval. Mice in the infection group were inoculated through the tail vein with 0.1 mL of H37Rv ( $1 \times 10^3$  CFU/ mL), that was 100CFU for each mouse, and control mice were similarly injected with 0.1 mL 0.9% NaCl. All procedures were performed in a biosafety cabinet located in the ABSL-3 facility.

### Analysis of Pathologic Changes

Six mice were sacrificed at each time-point (weeks 1, 3, 5, 8, 12, 16, 20, and 24), for the first experiment, the rest mice were sacrificed at week 34, 38, 44 and 52 (see Supplementary Image 2), six for each time-point, and for the latter two experiments, the rest mice were sacrificed at 52th week. The lung, spleen, and liver were removed from each animal; gross lesions were examined and described. The necropsy tissues were fixed and paraffin-embedded. The sections were cut and stained with hematoxylin and eosin, and reviewed by a veterinary pathologist. Semi-quantitative scores were assigned to reflect the lesion size and extent of inflammation of the entire lung field: +, 25%; ++, 50%; + + +, 75%.

### Quantitative Culture of Lung and Spleen Homogenates

The lung and spleen tissues taken at necropsy were washed in 4%  $\text{H}_2\text{SO}_4$  and homogenized in 0.9% NaCl. Serial dilutions of the homogenous lung fluid were used to inoculate the L-J medium culture in tubes, which were incubated at  $37^\circ\text{C}$  and 5%  $\text{CO}_2$ . The bacterial counts in the tubes were read after 21 days. At each time-point, the mean of the bacterial loads in each organ type of 6 mice were calculated.

## Evaluation of the Protective Effect of *G. lucidum* in the LTBI Mouse Model

### Study Design

The mice were divided into an LTBI mouse model control group, a *G. lucidum* prophylaxis group, and a therapy group. Mice in the control group were inoculated with *M. tuberculosis* as described above, but were fed normally and not given *G. lucidum* extract. The prophylaxis group received daily doses of *G. lucidum* mushroom extract (described below) beginning 1 month before *M. tuberculosis* inoculation and until 16 weeks after inoculation. In the therapy group, mice were given daily doses of *G. lucidum* mushroom extract beginning at the time of inoculation and lasting 16 weeks.

### Preparation of *G. lucidum* Extract-Containing Food

Takaya Shell-broken Ganoderma Lucid Spores and Spores Lipid, which were rich in terpene carbon dioxide extraction from included species of *G. lucidum* (MeiShanTang, Hong Kong) were mixed with the mouse feed powder and water, and then the mixture was manually formulated into ~15-cm-long feed strips with diameter of 1.5 cm. The strips were dried in the 37°C oven for 48 h. Each mouse received daily 15 mg of *G. lucidum* spores and 15 mg spore lipids in 4 g of feed (Lu et al., 2011).

### Mouse Trails

In the prophylaxis group, the feeding duration of *G. lucidum* was 20 weeks, starting 1 month before *M. tuberculosis* inoculation and continuing to the 16th week post-inoculation. In the therapy group, dietary *G. lucidum* (including spores and spore lipids) was administrated for 16 weeks, from the inoculation of *M. tuberculosis* to the 16th week post-inoculation. The effect of *G. lucidum* on *M. tuberculosis* infection in mice was evaluated according to the bacterial load in the lung and spleen, and the pathological changes in the lung, spleen, and liver at 3, 5, 8, and 16 weeks post-inoculation.

### Fluorescence-Activated Cell Sorting (FACS) Analysis of Immune Cells

Dendritic, natural killer, CD4<sup>+</sup>/CD8<sup>+</sup> T cells, and regulatory T (Treg) cells in the peripheral blood and lung were stained with specific antibody (all from eBioscience) for 30 min at 4°C. The following mouse antibodies were used: CD11b-fluorescein isothiocyanate (FITC) and CD4-FITC; I-AB-phycerythrin (PE), CD86-PE, CD8-PE, CD49b-PE, and FOXP3-PE; CD80-peridinin chlorophyll protein complex (PERCP) and CD3-PERCP; and CD11c-allophycocyanin (APC), perforin-APC, and CD25-APC. The stained cells were analyzed by FACS.

### Statistical Analyses

Quantitative data are expressed as the mean ± standard error of the mean (SEM), and analyzed using a two-tailed Student's *t*-test and ANOVA. *P* < 0.05 was considered statistically significant.

## RESULTS

### Bacterial Loads in the Lung and Spleen of the LTBI Mouse

The bacterial loads of lung and spleen in the infection and blank control groups at 1, 3, 5, 8, 12, 16, 20, and 24 weeks after inoculation with *M. tuberculosis* was determined (Figures 1A,B). No *M. tuberculosis* was detected in the blank control mice at any time-point. In the lungs of the infection group, the mean bacterial load was ~1.5log<sub>10</sub> CFU at 3rd week, at the lowest level found (~0.5 log<sub>10</sub> CFU) at 8th week, fluctuated from indeterminably low to ~2log<sub>10</sub> CFU at 8–20th week, and was ~2.5 log<sub>10</sub> CFU at week 24 (Figure 1A). In the spleens of the mice in the infection group, the mean bacterial load at 3rd week was 4.5log<sub>10</sub> CFU, then 2.1log<sub>10</sub> CFU at 8th week, and was subsequently progressively higher at each time-point, to >4 log<sub>10</sub> CFU at week 24 (Figure 1B). Compared with blank control, both the bacteria loads of spleen and lung in infected group had significant statistic difference (\*\*P < 0.001).

There were no observed gross pathological changes throughout the infection course in any group, but histopathological changes were observed under the light microscope. In the lung at week 3, mild infiltration with inflammatory cells in the peripheral vascular was observed in the lung. At week 5, a granuloma-like structure in the lung was observed, and scored as a severe lesion (+++). However, at week 8, the histopathologic lesion was not seen, and only a few inflammatory cells infiltrated vessels. Infiltration of inflammatory cells in the lung was greater at week 16 (Figure 2).

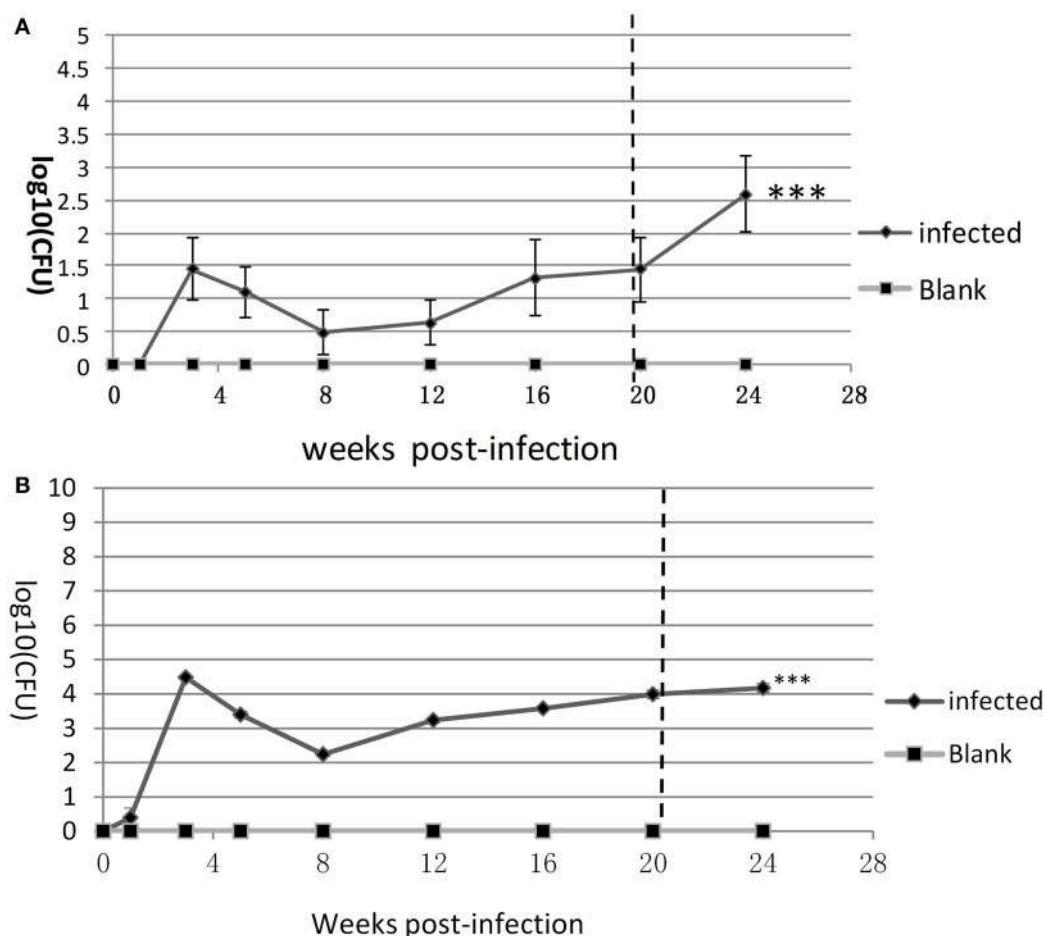
In the spleen at 3rd week, several small granulomas (diameter, 5–50 μm) were detected in the white pulp. At week 5, the granulomas in the white pulp of the spleen were of diameters 20–150 μm, but at week 8 granuloma sizes were 5–20 μm. At week 20, several granuloma-like lesions of size 10–50 μm were observed in the spleen white pulp (Figure 2).

At week 3–16, granuloma-like lesions were occasionally observed in the hepatic lobule. The size of irregularly shaped granuloma-like lesions at week 3 was ~50 × 350 μm, and smaller granuloma-like lesions (20–50 μm) were present from weeks 5 to 20 (Figure 2).

### Bacterial Loads in the Lung and Spleen of *G. lucidum* Extract-Treated Mice

The condition of mice inoculated with 100 CFU *M. tuberculosis* resembled LTBI during the first 20 weeks post-inoculation. Therefore, to evaluate the effect of the *G. lucidum* extract, we determined the bacterial loads of the control and model mice at time-points up to 16 weeks post-inoculation.

In the prophylaxis group, in the lung the bacterial load was essentially none up through week 8, but was (0.42 ± 0.34) log<sub>10</sub> CFU at week 16 (Figure 3A). In the therapy group, the *M. tuberculosis* level in the lung remained essentially undetectable through the first 3 weeks, but was (1.39 ± 0.77) log<sub>10</sub> CFU at week 5, (0.42 ± 0.42) log<sub>10</sub> CFU at week 8, and again undetectable at week 16. In the spleens of the prophylaxis group, the mean



**FIGURE 1 |** The replication kinetics of *M. tuberculosis* in the lung and spleen. Bacterial loads ( $\log_{10}$ CFU) in (A) the lung and (B) spleen were counted on the L-J medium from week 1 to week 24. Six mice for each time-point in each group. One of three similar experiments is shown. The SEMs are plotted as error bars. *T*-test and ANOVA analysis, \*\*\* $P$  < 0.001.

*M. tuberculosis* levels were  $(1.69 \pm 0.71) \log_{10}$  CFU at week 3,  $(1.12 \pm 0.77) \log_{10}$  CFU at week 5,  $(3.48 \pm 0.11) \log_{10}$  CFU at week 8, and  $(2.41 \pm 0.41) \log_{10}$  CFU at week 16. In the spleens of the therapy group, the bacterial loads were similar to that of the control group: a peak value at  $(4.33 \pm 0.75) \log_{10}$  CFU at week 3,  $(3.56 \pm 0.05) \log_{10}$  CFU at week 5,  $(3.22 \pm 0.17) \log_{10}$  CFU at week 8, and  $(3.04 \pm 0.17) \log_{10}$  CFU at week 16 (Figure 3B). As for spleen bacteria loads, the prophylaxis group showed statistical differences at both week 3 and 5, compared with both blank control and therapy group ( $W_3, P < 0.05$ ;  $W_5, P < 0.01$ ), and only with control group at week 8 ( $P < 0.05$ ); but for lung bacteria loads, the prophylaxis group revealed significant statistic difference only at week 5, compared with therapy group (Figures 3A,B).

### Histopathologic Changes in *G. lucidum*-Treated Mice

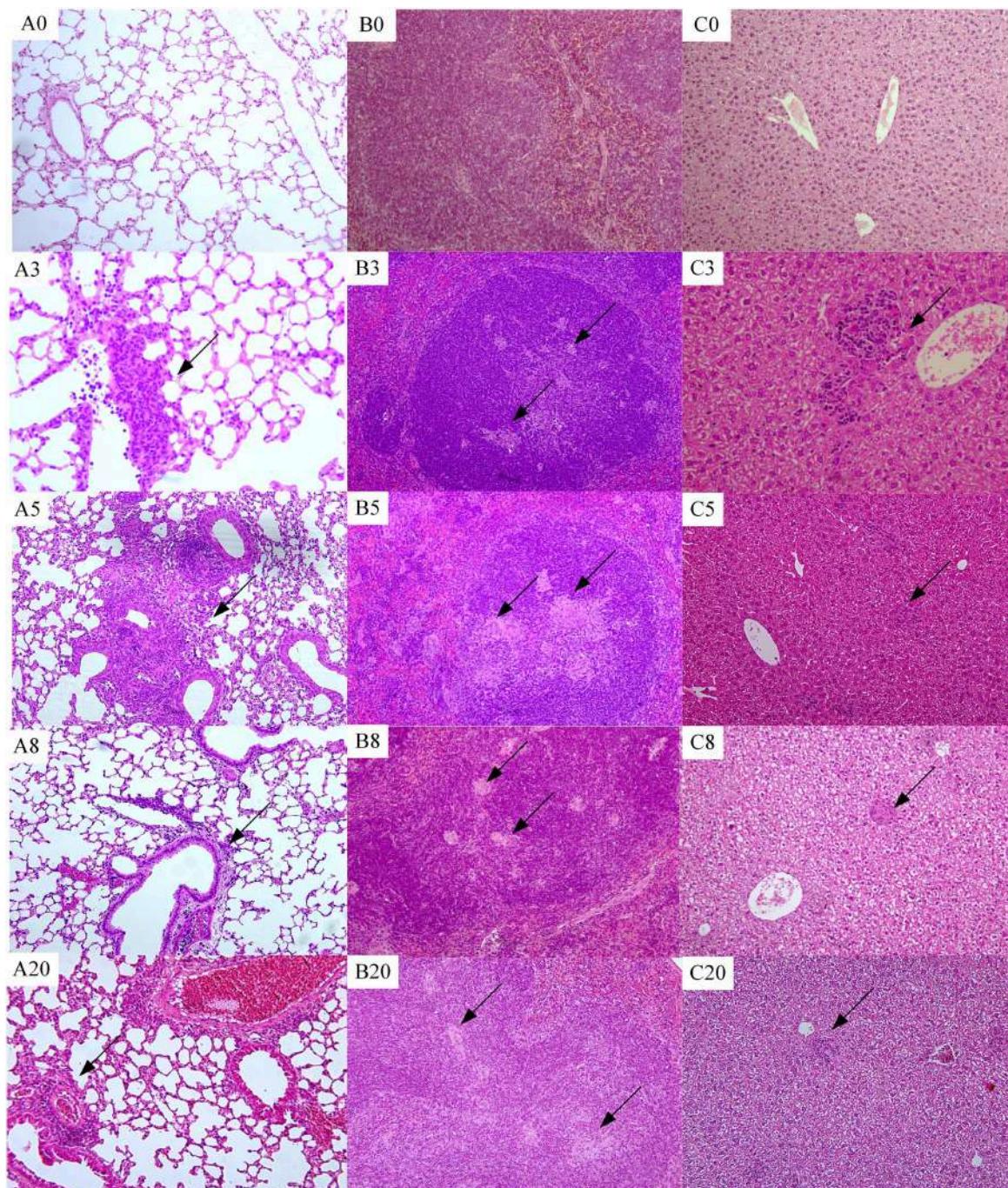
There were no macroscopic changes in the lungs, spleens, or livers in any of the groups, and no visible differences in

the microscopic histology of the spleens or livers. However, histological differences in the lung were detected.

In the LTBI model control (untreated) group, inflammatory lesions were aggravated at week 5 compared with week 3, but were not seen from week 8 to 16. Similar to the control group, in the prophylaxis and therapy group small lesions peaked at week 5, and then were absent at week 8–16. At week 5, inflammation in the prophylaxis group was milder than inflammation in the control group (Figure 4A). Compared with the control group, mice in the *G. lucidum* prophylaxis group had lower pathology scores during the infection course, but the difference was not significant ( $P > 0.05$ ; Figure 4B).

### Percentages of Dendritic Cell, Natural Killer Cell, and Treg Cell, and the Ratios of CD4<sup>+</sup>/CD8<sup>+</sup> T cell in Blood

The number of dendritic cell, natural killer cell, Treg cell, and the ratio of CD4<sup>+</sup>/CD8<sup>+</sup> T cell in the peripheral blood

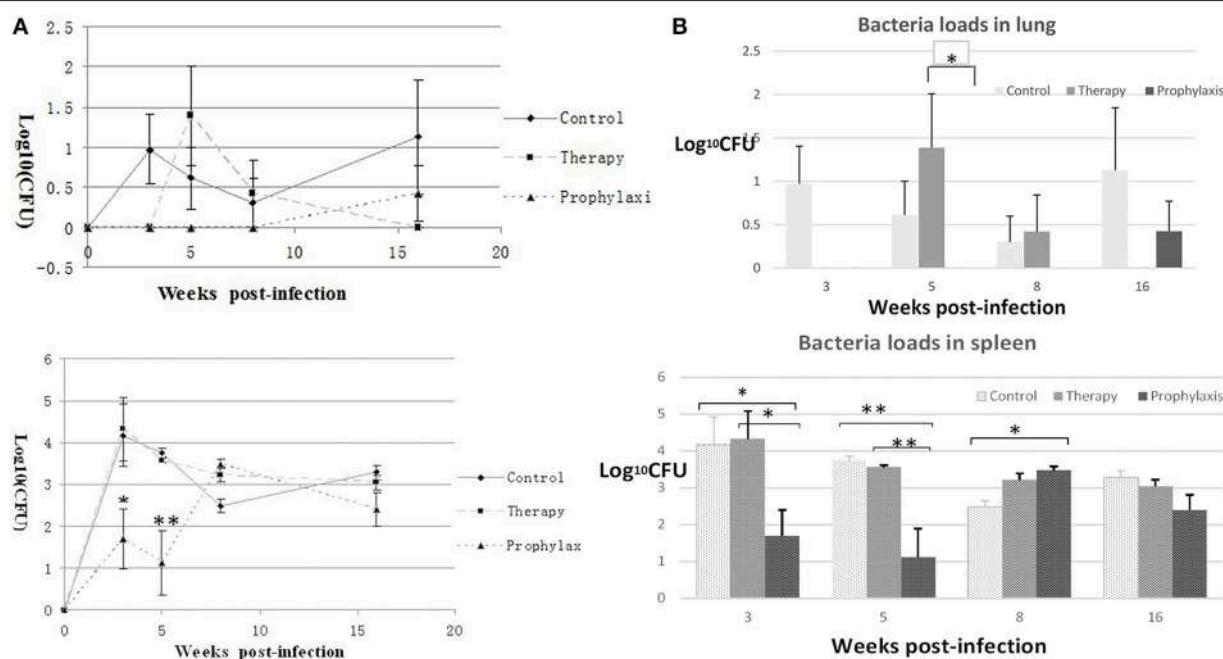


**FIGURE 2 | Histopathological changes over time (100 $\times$ ).** (A0–A20) Lung histology at weeks 0, 3, 5, 8, and 20, respectively. (B0–B20) Spleen histopathology at weeks 0, 3, 5, 8, and 20. (C0–C20) Liver histopathology at weeks 0, 3, 5, 8, and 20. Arrow indicates granulomas-like lesion.

were analyzed by FACS (Figure 5). In the prophylaxis group, the percentages of dendritic cell were  $14.8 \pm 5.04\%$  at week 3,  $6.56 \pm 1.93\%$  at week 5,  $1.22 \pm 0.10\%$  at week 8, and  $7.64 \pm 2.69\%$  at week 16. In the therapy and control group, the percentages of dendritic cells were similar to that of the prophylaxis group, progressively increasing over

time from  $9.24 \pm 1.55\%$  at week 3 to  $18.48 \pm 2.00\%$  at week 16.

The percentages of natural killer cell in the prophylaxis group was  $2.70 \pm 0.68\%$  at week 3,  $9.82 \pm 1.48\%$  at week 5,  $15.76 \pm 2.63\%$  at week 8, and  $7.15 \pm 1.20\%$  at week 16. The percentages of natural killer cell in the control and therapy group were



comparable: ~2, ~10, ~5, and ~7% at weeks 3, 5, 8, and 16, respectively.

The ratios of CD4<sup>+</sup>/CD8<sup>+</sup>T cells in all three groups were similar and stable at ~1.8 throughout the infection period, except for the prophylaxis group, in which a mean ratio of ~4.2 was observed at week 8.

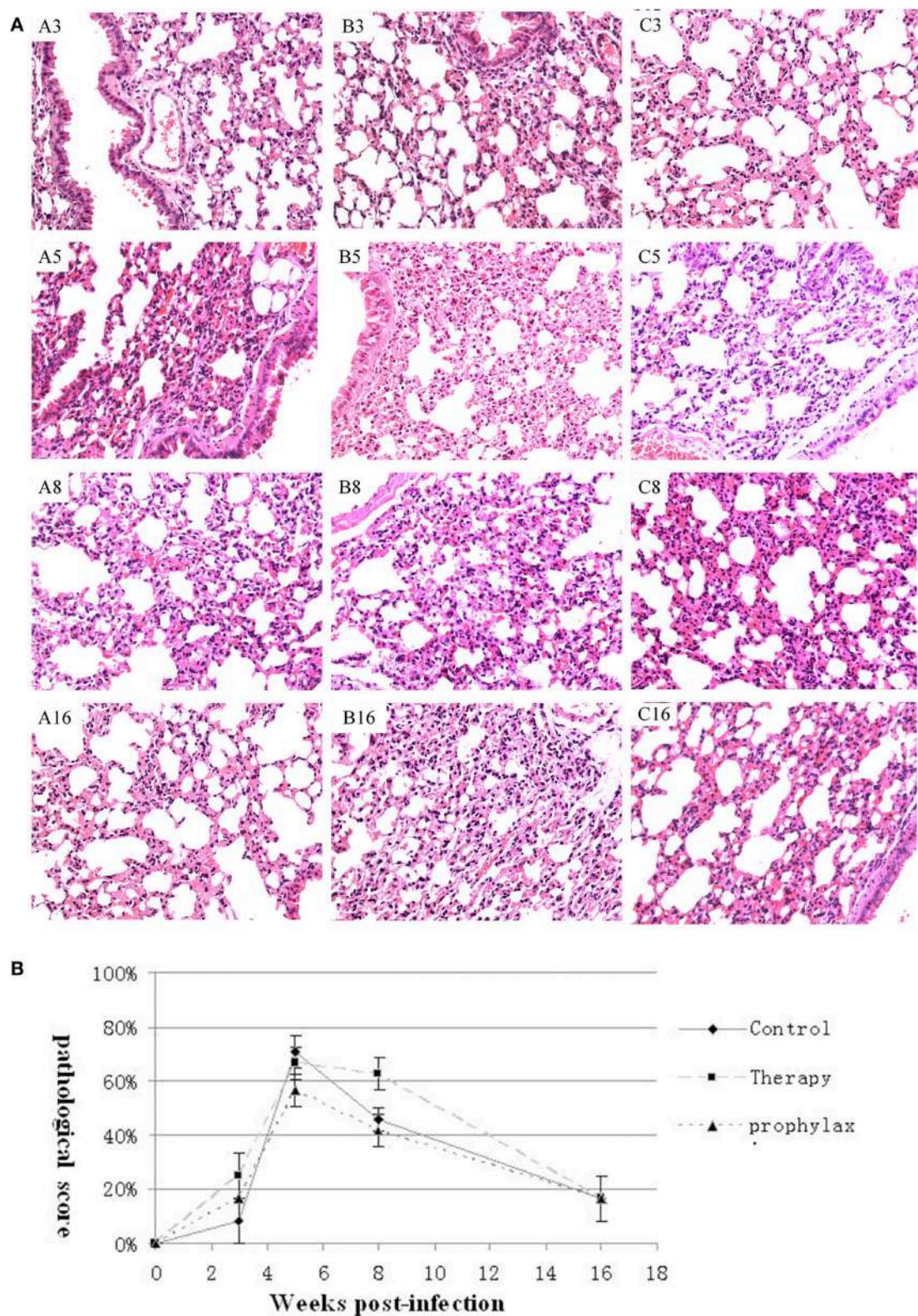
The percentages of Treg cells in the prophylaxis group were  $0.86 \pm 0.05\%$  at week 3,  $0.49 \pm 0.08\%$  at week 5,  $1.63 \pm 0.08\%$  at week 8, and  $0.76 \pm 0.12\%$  at week 16. In the control and therapy group, the percentages of Treg cells were similar (~1%) at week 5, and progressively and similarly decreased from week 5 to 16 (Figure 5).

## DISCUSSION

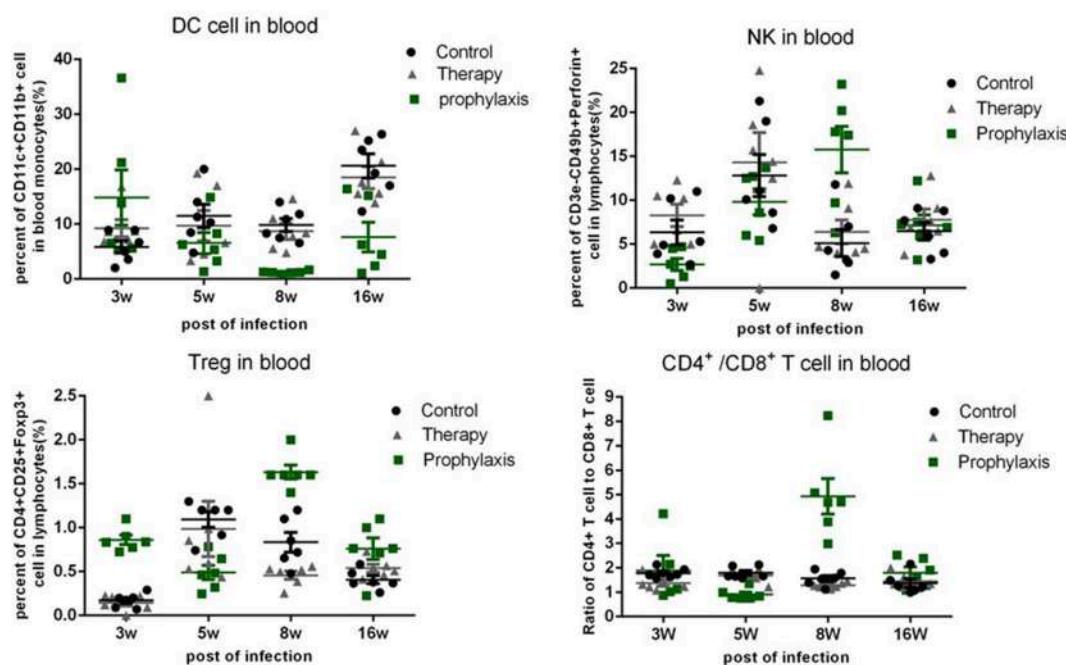
We present here a novel mouse model of spontaneous LTBI, which closely resembles human LTBI. The model was generated through experimental inoculation of *M. tuberculosis* without resorting to other means such as anti-tuberculosis drugs treatment or *Mycobacterium bovis* Bacillus Calmette-Guerin vaccination. A useful application of this model is to evaluate anti-tuberculosis drugs or vaccines during the early latency phase, which would be more efficient than the previous models (Ziv et al., 2001; Ha et al., 2003; Zhang et al., 2011). In the present study, we demonstrated the efficiency of the new mouse LTBI model by testing the effect of *G. lucidum* extract on *M. tuberculosis* infection, by inhibiting *M. tuberculosis* replication.

To initiate the LTBI model, C57BL/6 mice were IV-injected with 100 CFU *M. tuberculosis* inoculum. *M. Tuberculosis* initially replicated in the lung and spleen, with bacterial loads rising until week 3, after which infection entered into a latent phase with *M. tuberculosis* levels declining from week 3 and stabilizing at a lower level till week 20. In accord with the bacterial load findings, pathological changes were detected at week 3, which became prominent at week 5, and then subsided to only mild perivascular inflammation from week 8 to 20. *M. tuberculosis* replication shifted from relatively active to dormant, resulting in lower bacterial loads and mild pathological changes in both the lung and spleen from weeks 8 to 20 post-inoculation. Thus, the period within 8 weeks would be regarded as “pre-latency phase,” and the period from week 8 to week 20 comprised the real latency phase of *M. tuberculosis* infection in this model (Figures 2, 3). The bacteria load curves of spleen and lung were similar to others study, but with lower value and shorter latency period, which would be more efficient in drug evaluation.

Treatment of mice with *G. lucidum* extract beginning 1 month before inoculation with *M. tuberculosis* (the prophylaxis group) inhibited replication of the bacterium in the lung and spleen during the first 8 weeks post-inoculation, relative to untreated control mice and mice treated from the time of inoculation (the therapy group). This inhibition was especially evident in the spleen 5 weeks after inoculation. No significant inhibition of *M. tuberculosis* replication was observed in the therapy group, suggesting



**FIGURE 4 | Pathology of the lung. (A)** Histopathology of the lungs at weeks 3, 5, 8, and 16 weeks of representative mice in the (A3–A16) control, (B3–B16) prophylaxis, and (C3–C16) therapy groups (200 $\times$ ). **(B)** Pathological lesion scores in the lung of mice in the control, therapy and prophylaxis groups at weeks 3, 5, 8, and 16. SEMs are plotted ( $P > 0.05$ ).



**FIGURE 5 |** Percentages of dendritic cells (DC), natural killer (NK) cells, and Treg cells, and the ratios of CD4<sup>+</sup>/CD8<sup>+</sup>T cells in the blood. The results are expressed as mean  $\pm$  SEM, plotted using GraphPad Prism version 4.01 (GraphPad, San Diego, CA).

that early use of *G. lucidum* extract is probably required for exerting anti- *tuberculosis* activity in this mouse model (**Figure 4**).

The changes of immune cell percentages in the peripheral blood did not correlate with the bacterial loads of *M. tuberculosis* or pathological changes observed in the lung and spleen. However, responses of the peripheral blood and lung to inoculation were opposed, with regard to percentages of dendritic cells. At post-inoculation week 3, the number of peripheral dendritic cells in prophylaxis group was higher than that in the control group, but at week 5 was lower than the control group, while the number of dendritic cells in the prophylaxis group lung was higher at week 5 than week 3 (Supplementary Data), and higher than the control group. These changes suggest that dendritic cells may have transferred from the peripheral blood to the lung from week 3 to week 5, and the accumulation of dendritic cells in the lung might participate in and augment the repression of *M. tuberculosis* replication, inferred from the immune effect of *G. lucidum* on dendritic cells in previous research (**Figure 5**) (Lin et al., 2009b; Meng et al., 2011).

However, the immune mechanism of *G. lucidum* extract could not be discerned, and nor if the effective constituent was terpenes, which should be investigated in a further study.

In conclusion, we established a mouse model resembling human *M. tuberculosis* LTBI. The early acute phase of infection (from inoculation to week 8) and the relatively short latency period makes this model useful for evaluating anti-tuberculosis

drugs. We utilized the model to investigate the effect of *G. lucidum* extract on *M. tuberculosis* infection. We found that administration of Takaya *G. lucidum* extract prior to inoculation of *M. tuberculosis* was associated with lower levels of bacterial replication in this model.

## AUTHOR CONTRIBUTIONS

LZ and CQ conceived the concept and designed the experiment. LZ, JT, and SL conducted all the experiments except for the pathological analysis. YX and YX performed the pathology examinations. YL provided the Takaya *Ganoderma lucidum* extract. LZ performed the data analysis and wrote the manuscript. All the authors read and approved this manuscript for publication.

## ACKNOWLEDGMENTS

This study was supported by the National Science and Technology Major Projects of Infectious Disease (2012ZX10004501-001-005).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2015.01490>

## REFERENCES

- Chien, C. M., Cheng, J. L., Chang, W. T., Tien, M. H., Tsao, C. M., Chang, Y. H., et al. (2004). Polysaccharides of *Ganoderma lucidum* alter cell immunophenotypic expression and enhance CD56+ NK-cell cytotoxicity in cord blood. *Bioorg. Med. Chem.* 12, 5603–5609. doi: 10.1016/j.bmc.2004.08.004
- Feng, L., Yuan, L., Du, M., Chen, Y., Zhang, M. H., Gu, J. F., et al. (2013). Anti-lung cancer activity through enhancement of immunomodulation and induction of cell apoptosis of total triterpenes extracted from *Ganoderma lucidum* (Leyss. ex Fr.) Karst. *Molecules* 18, 9966–9981. doi: 10.3390/molecules18089966
- Ha, S. J., Jeon, B. Y., Kim, S. C., Kim, D. J., Song, M. K., Sung, Y. C., et al. (2003). Therapeutic effect of DNA vaccines combined with chemotherapy in a latent infection model after aerosol infection of mice with *Mycobacterium tuberculosis*. *Gene Ther.* 10, 1592–1599. doi: 10.1038/sj.gt.3302057
- Kashino, S. S., Napolitano, D. R., Skobe, Z., and Campos-Neto, A. (2008). Guinea pig model of *Mycobacterium tuberculosis* latent/dormant infection. *Microbes Infect.* 10, 1469–1476. doi: 10.1016/j.micinf.2008.08.010
- Lai, C. Y., Hung, J. T., Lin, H. H., Yu, A. L., Chen, S. H., Tsai, Y. C., et al. (2010). Immunomodulatory and adjuvant activities of a polysaccharide extract of *Ganoderma lucidum* in vivo and in vitro. *Vaccine* 28, 4945–4954. doi: 10.1016/j.vaccine.2010.05.037
- Lenaerts, A. J., Chapman, P. L., and Orme, I. M. (2004). Statistical limitations to the Cornell model of latent tuberculosis infection for the study of relapse rates. *Tuberculosis (Edinb.)* 84, 361–364. doi: 10.1016/j.tube.2004.03.002
- Lin, P. L., Rodgers, M., Smith, L., Bigbee, M., Myers, A., Bigbee, C., et al. (2009a). Quantitative comparison of active and latent tuberculosis in the cynomolgus macaque model. *Infect. Immun.* 77, 4631–4642. doi: 10.1128/IAI.00592-09
- Lin, Y. L., Liang, Y. C., Tseng, Y. S., Huang, H. Y., Chou, S. Y., Hseu, R. S., et al. (2009b). An immunomodulatory protein, Ling Zhi-8, induced activation and maturation of human monocyte-derived dendritic cells by the NF-kappaB and MAPK pathways. *J. Leukoc. Biol.* 86, 877–889. doi: 10.1189/jlb.0708441
- Lu, Y. Z., Wu, X. X., Chen, S., Yuan, J., Lai, C. H., Bao, L. L., et al. (2011). Effective of *Ganoderma lucidum* preparation in treating simian aquire immune deficiency syndrome. *Acta Acad. Med. Sin.* 33, 318–324. doi: 10.3881/j.issn.1000-503X.2011.03.021
- Manabe, Y. C., Kesavan, A. K., Lopez-Molina, J., Hatem, C. L., Brooks, M., Fujiwara, R., et al. (2008). The aerosol rabbit model of TB latency, reactivation and immune reconstitution inflammatory syndrome. *Tuberculosis (Edinb.)* 88, 187–196. doi: 10.1016/j.tube.2007.10.006
- Meng, J., Hu, X., Shan, F., Hua, H., Lu, C., Wang, E., et al. (2011). Analysis of maturation of murine dendritic cells (DCs) induced by purified *Ganoderma lucidum* polysaccharides (GLPs). *Int. J. Biol. Macromol.* 49, 693–699. doi: 10.1016/j.ijbiomac.2011.06.029
- Pan, D., Zhang, D., Wu, J., Chen, C., Xu, Z., Yang, H., et al. (2013). Antidiabetic, antihyperlipidemic and antioxidant activities of a novel proteoglycan from *Ganoderma lucidum* fruiting bodies on db/db mice and the possible mechanism. *PLoS ONE* 8:e68332. doi: 10.1371/journal.pone.0068332
- Scanga, C. A., Mohan, V. P., Joseph, H., Yu, K., Chan, J., and Flynn, J. L. (1999). Reactivation of latent tuberculosis: variations on the cornell murine model. *Infect. Immun.* 67, 4531–4538.
- Singhal, A., Mathys, V., Kiass, M., Creusy, C., Delaire, B., Aliouat el, M., et al. (2011). BCG Induces Protection against *Mycobacterium tuberculosis* Infection in the Wistar rat model. *PLoS ONE* 6:e28082. doi: 10.1371/journal.pone.0028082
- Wang, C. L., Pi, C. C., Kuo, C. W., Zhuang, Y. J., Khoo, K. H., Liu, W. H., et al. (2011). Polysaccharides purified from the submerged culture of *Ganoderma formosanum* stimulate macrophage activation and protect mice against *Listeria monocytogenes* infection. *Biotechnol. Lett.* 33, 2271–2278. doi: 10.1007/s10529-011-0697-2
- Woolhiser, L., Tamayo, M. H., Wang, B., Gruppo, V., Belisle, J. T., Lenaerts, A. J., et al. (2007). In vivo adaptation of the Wayne model of latent tuberculosis. *Infect. Immun.* 75, 2621–2625. doi: 10.1128/IAI.00918-06
- Yeh, C. H., Chen, H. C., Yang, J. J., Chuang, W. I., and Sheu, F. (2010). Polysaccharides PS-G and protein LZ-8 from Reishi (*Ganoderma lucidum*) exhibit diverse functions in regulating murine macrophages and T lymphocytes. *J. Agric. Food Chem.* 58, 8535–8544. doi: 10.1021/jf100914m
- Yoshida, H., Suzuki, M., Sakaguchi, R., Tani, I., Kotani, H., Shudo, N., et al. (2012). Preferential induction of Th17 cells in vitro and in vivo by Fucogalactan from *Ganoderma lucidum* (Reishi). *Biochem. Biophys. Res. Commun.* 422, 174–180. doi: 10.1016/j.bbrc.2012.04.135
- Zhang, T., Li, S.-Y., Williams, K. N., Andries, K., and Nuermberger, E. L. (2011). Short-course chemotherapy with TMC207 and Rifapentine in a murine model of latent tuberculosis infection. *Am. J. Respir. Crit. Care Med.* 184, 732–737. doi: 10.1164/rccm.201103-0397OC
- Ziv, E., Daley, C. L., and Blower, S. M. (2001). Early therapy for latent tuberculosis infection. *Am. J. Epidemiol.* 153, 381–385. doi: 10.1093/aje/153.4.381
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Copyright © 2016 Zhan, Tang, Lin, Xu, Xu and Qin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Proteomics Analysis of Three Different Strains of *Mycobacterium tuberculosis* under *In vitro* Hypoxia and Evaluation of Hypoxia Associated Antigen's Specific Memory T Cells in Healthy Household Contacts

OPEN ACCESS

**Edited by:**

Julio Alvarez,  
University of Minnesota Twin Cities,  
USA

**Reviewed by:**

Yunlong Li,  
Wadsworth Center, USA  
Yusuf Akhter,  
Central University of Himachal  
Pradesh, India

**\*Correspondence:**

Alamelu Raja  
alameluraja@gmail.com

**Specialty section:**

This article was submitted to  
Infectious Diseases,  
a section of the journal  
*Frontiers in Microbiology*

**Received:** 20 January 2016

**Accepted:** 02 August 2016

**Published:** 09 September 2016

**Citation:**

Devasundaram S, Gopalan A,  
Das SD and Raja A (2016)

Proteomics Analysis of Three Different  
Strains of *Mycobacterium*  
*tuberculosis* under *In vitro* Hypoxia  
and Evaluation of Hypoxia Associated  
Antigen's Specific Memory T Cells  
in Healthy Household Contacts.

*Front. Microbiol.* 7:1275.  
doi: 10.3389/fmicb.2016.01275

*In vitro* mimicking conditions are thought to reflect the environment experienced by *Mycobacterium tuberculosis* inside the host granuloma. The majority of *in vitro* dormancy experimental models use laboratory-adapted strains H37Rv or Erdman instead of prevalent clinical strains involved during disease outbreaks. Thus, we included the most prevalent clinical strains (S7 and S10) of *M. tuberculosis* from south India in addition to H37Rv for our *in vitro* oxygen depletion (hypoxia) experimental model. Cytosolic proteins were prepared from hypoxic cultures, resolved by two-dimensional electrophoresis and protein spots were characterized by mass spectrometry. In total, 49 spots were characterized as over-expressed or newly emergent between the three strains. Two antigens (ESAT-6, Lpd) out of the 49 characterized spots were readily available in recombinant form in our lab. Hence, these two genes were overexpressed, purified and used for *in vitro* stimulation of whole blood collected from healthy household contacts (HHC) and active pulmonary tuberculosis patients (PTB). Multicolor flow cytometry analysis showed high levels of antigen specific CD4<sup>+</sup> central memory T cells in the circulation of HHC compared to PTB ( $p < 0.005$  for ESAT-6 and  $p < 0.0005$  for Lpd). This shows proteins that are predicted to be up regulated during *in vitro* hypoxia in most prevalent clinical strains would indicate possible potential immunogens. *In vitro* hypoxia experiments with most prevalent clinical strains would also elucidate the probable true representative antigens involved in adaptive mechanisms.

**Keywords:** *M. tuberculosis*, hypoxia, prevalent clinical strains, two-dimensional electrophoresis, mass spectrometry, multicolor flow cytometry

## INTRODUCTION

The success of *Mycobacterium tuberculosis* lies in its ability to persist within humans for long periods without causing any disease symptoms, known as latent tuberculosis infection (LTBI). About two billion people are estimated to have latent infections that could reactivate into TB disease (Kasprowicz et al., 2011). Because of the huge reservoir of latently infected individuals, diagnosis, and treatment of latent TB infections have obtained increasing importance as public health measures to control TB. In-depth knowledge about the biology of dormant *M. tuberculosis* is important to develop new therapeutic tools for latent TB (Barry et al., 2009). Several lines of evidence link latent tuberculosis and inhibition of *M. tuberculosis* growth with hypoxic conditions. Depletion of oxygen prevents aerobic respiration by the obligate aerobe within the host (Fang et al., 2012). In order to persist within the host, *M. tuberculosis* possess the ability to adapt to the hypoxic environment which is considered a crucial part of the adaptation mechanism (Muttucumaru et al., 2004).

To better understand this state of dormancy, numerous *in vitro* experiments have been developed to mimic, at least in part, the host's intracellular environment experienced by *M. tuberculosis*. *In vitro* mimicking conditions include oxygen deprivation (hypoxia), low pH, and nutrient starvation which either inhibits or slows down bacterial growth (non-replicating stage). Among these, hypoxia is extensively studied and considered a potential factor for transformation into the non-replicating dormant form of *M. tuberculosis* (Rustad et al., 2009; Fang et al., 2012). The most frequently used experimental model for hypoxia-induced *M. tuberculosis* dormancy is the defined headspace model of non-replicating persistence (NRP; Wayne and Hayes, 1996) and is adapted for the present study.

To date, many *in vitro* hypoxia experimental models use common laboratory mycobacterial strains like H37Rv and Erdmann (Sherman et al., 2001; Voskuil et al., 2004). However, laboratory strains might not completely represent the virulence of naturally occurring clinical strains involved in disease outbreaks. A few hypoxia reports (Boon et al., 2001; Starck et al., 2004) used prevalent clinical strains but these strains were not from TB endemic areas. The unique feature of the present report is that we have used two prevalent clinical strains (S7 and S10) from a TB endemic region like India and evaluated their adaptation mechanisms, in terms of protein expression, under *in vitro* hypoxia.

The strains S7 and S10 were first reported by Das et al. (1995) from a restriction fragment length polymorphism (RFLP) study which showed that most (38–40%) of the clinical isolates of *M. tuberculosis*, taken from the *Bacillus Calmette–Guerin* (BCG) trial area of Tiruvallur district, south India, harbored a single copy of *IS6110* in their genome. Among the other strains studied by Rajavelu and Das (2005), S7, and S10 drew our attention due to their distinct immune responses (S7 induced Th-2 response while Th-1 response was induced by strain S10) despite having a single copy of *IS6110* at the same locus in the genome.

Genes/proteins that are over-expressed during *in vitro* stress are likely to be crucial for intracellular survival of *M. tuberculosis* and are potential targets for anti-TB drug and

vaccine development (Betts, 2002; Andersen, 2007; Hingley-Wilson et al., 2010).

We hypothesized that proteins identified as up regulated especially in clinical isolates, during *in vitro* hypoxia, would be better potential vaccine candidates than proteins predicted from laboratory adapted strains.

To analyze hypoxia associated proteins, we compared protein expression profiles of each strain (H37Rv, S7, and S10) during well aerated growth conditions (aerobic) and oxygen depleted growth conditions (anaerobic/hypoxic).

Proteins spots, expressed under hypoxia were characterized by mass spectrometry. Two antigens were selected for *in vitro* recombinant antigen preparation to test our hypothesis of using clinical strains to obtain the most promising potential antigens. These two antigens were used to stimulate samples of whole blood collected from latently infected healthy household contacts (HHCs) and active TB individuals [pulmonary TB (PTB)]. The LTBI population is presumed to be protected against active TB disease and antigens that are preferentially detected by LTBI can be considered as novel vaccine targets (Andersen, 2007; Govender et al., 2010).

## MATERIALS AND METHODS

### **Mycobacterial Strains**

The laboratory strain H37Rv (ATCC 27294) of *M. tuberculosis* was obtained from Colorado State University (CSU), Fort Collins, CO, USA. H37Rv was originally derived from H37, a clinical isolate isolated from a pulmonary tuberculosis patient in 1905 (Steenken et al., 1934). The clinical isolates S7 and S10 of *M. tuberculosis* were first isolated from the BGC trial area of the Tiruvallur District, Tamil Nadu, India during the Model Dots trial (Das et al., 1995). These isolates are maintained as glycerol stocks and can be obtained through proper request to National Institute for Research in Tuberculosis, Chennai, India.

### ***In vitro* Culture Method**

Three mycobacterial strains (H37Rv, S7, and S10) were grown in Middlebrook 7H9 media (MB7H9) containing 2% glycerol (v/v), 10% albumin-dextrose-catalase (ADC), and 0.05% Tween 80 (v/v) at 37°C 200 rpm to obtain aerobic cultures.

Wayne's *in vitro* oxygen depletion method was followed to generate hypoxic(anaerobic) cultures (Wayne and Hayes, 1996). All three strains (H37Rv, S7, and S10) were inoculated into screw capped test tubes (20 mm × 125 mm, with a total fluid capacity of 25.5 mL) pre-filled with supplemented MB7H9. Test tubes were initially filled with 17 ml of MB7H9 broth leaving 8.5 ml head space to give a head to air space ratio of 0.5. After inoculation, these tubes were incubated at 37°C. Sterile 8-mm Teflon-coated magnetic stirring bars were used in hypoxic cultures to stir gently at 120 rpm. This stirring maintains the uniform dispersion and the rate of O<sub>2</sub> depletion was under control.

The O<sub>2</sub> depletion was monitored by reduction and decolorization of the methylene blue indicator. A final concentration of 1.5 µg mL<sup>-1</sup> of sterile solution of methylene blue (Sigma-Aldrich, St. Louis, MO, USA) was added into the

hypoxia cultures during inoculation. In *M. tuberculosis* *in vitro* cultures methylene blue decolorization starts when the dissolved oxygen concentration is declined below 3% (Leistikow et al., 2010). Hence, complete decolorization of methylene blue was taken to indicate oxygen depletion.

The culture tube containing supplemented MB7H9, methylene blue and no bacterial inoculum was set-up as a “blank.” Growth was measured at OD<sub>600 nm</sub> (optical density) in both aerobic and anaerobic cultures. Triplicate cultures of both aerobic and anaerobic were set up for each strain.

## Cytosolic Proteins Preparation

Triplicate aerobic and anaerobic cultures of H37Rv, S7, and S10 were harvested by centrifugation at 4000 × g for 15 min at 4°C. The pellets, from triplicate cultures, were washed twice with 40 mM Tris-buffer, centrifuged (4000 × g, 15 min, 4°C), and the supernatant was discarded. The pellets were resuspended in lysis buffer containing 20 mM Tris-HCl, 100 mM dithiothreitol (DTT), 1 mM PMSF (phenylmethylsulfonyl fluoride), complete protease inhibitor cocktail (Sigma-Aldrich, St. Louis, MO, USA), and 10 mg/mL lysozyme in ice. The cell membrane was then disturbed by ultra sonication (amplitude 40%) and homogenate was collected after high speed centrifugation (18000 × g for 25 min). Likewise, three cytosolic protein fractions for H37Rv, S7, and S10 were available and were separated by 2DE (two-dimensional electrophoresis) experiments.

## Two-Dimensional Electrophoresis (2DE)

Four hundred (400 µg) micrograms of cytosolic fraction proteins from H37Rv, S7, and S10, from both aerobic and anaerobic cultures, were taken and impurities were removed by a 2D cleanup kit (Bio-Rad Laboratories, Hercules, CA, USA). Protein concentration was then estimated by BCA Protein assay – Reducing agent compatibility kit (Thermo Fisher Scientific Inc, Waltham, MA, USA).

For first dimension isoelectric focusing (IEF), protein samples were solubilized in rehydration buffer (8 M urea, CHAPS 2%, ampholytes 3–7 and 4–6 and 15–100 mM DTT and separated by 17 cm immobilized pH gradient (IPG) strips (Bio-Rad Laboratories, USA) of pH range 4–7. Each IPG strip was rehydrated with 300 µl rehydration buffer containing 200 µg protein samples. IEF was performed at 20°C in an IEF cell as per manufacturer's instructions (Bio-Rad Laboratories, Hercules, CA, USA) by following electrical conditions; 150 V for 15 min, end voltage was 10,000 V and volt hours were 40–60,000. The electrical limit was set at 50 µA per strips. After IEF strips were first equilibrated with equilibration buffer 1 (2% DTT, 2% SDS, and 6 M urea) for 10 min and followed by equilibration buffer 2 (2.5% iodoacetamide, 2% SDS, and 6 M urea) for 10 min. Each of the IPG strips was loaded on a vertical SDS-PAGE gel (12%; Second dimension) and sealed with 1% low melting agarose dissolved in SDS running buffer. SDS-PAGE was performed under reducing conditions at constant voltage (150 V). Then, gels were developed by Coomassie brilliant blue (CBB) R 250 (Bio-Rad Laboratories, Hercules, CA, USA) to visualize proteins and images of gels were acquired by Chemidoc XRS gel documentation system (Bio-Rad Laboratories, Hercules, CA,

USA). Image analysis was performed using the PDQuest software (version 7.0.0; Bio-Rad Laboratories, Hercules, CA, USA) by stepwise spot detection and spot matching. Protein samples from all three strains were loaded in equal concentrations and 2-DE (two-dimensional electrophoresis) experiments were done for all three lysates prepared per strain. Both aerobic and anaerobic lysates were loaded simultaneously for IEF and SDS-PAGE to maintain the same running conditions. Uniform staining and destaining was followed for all gels (aerobic and anaerobic).

To identify protein expression differences between the aerobic and anaerobic cultures of all three *M. tuberculosis* strains, cytosolic proteins were separated by 2-DE and protein profiles were compared by PDQuest software. A threshold value of 1.5-fold difference was assigned to identify differentially expressed protein spots between the aerobic and anaerobic cultures of each strain. Student's *t*-test with confidential hit 0.05 reliability score was performed to evaluate the significance of each differentially expressed protein spot. Differentially expressed protein spots that were identified consistently in two out of three cytosolic fraction's 2-DE experiments were only included for further characterization.

Two types of criteria were used to identify the possible hypoxia associated proteins from these strains. First, protein spots whose intensity was higher in anaerobic gel (hypoxic) compared to its aerobic counterpart gel (aerobic) were selected and designated as “over-expressed” proteins during hypoxia. These “over-expressed” proteins are expressed at basal levels during aerobic growth and increase expression in a hypoxic environment. Secondly, protein spots which were completely absent (no basal level expression) during aerobic growth, but expressed only during anaerobic growth were selected and designated as “newly expressed” protein spots. These newly expressed spots were unique to anaerobic gel (hypoxia) and no corresponding spots were seen in the aerobic counterpart gel (aerobic).

## In-gel Digestion with Trypsin

Protein spots of interest were excised from gels and digested with modified trypsin (Roche Molecular Biochemicals, Indianapolis, IN, USA). The gel plugs were washed several times with 100 mM ammonium bicarbonate (NH<sub>4</sub>HCO<sub>3</sub>) in 50% acetonitrile, the gel pieces were subjected to a reduction step using 10 mM DTT in 100 mM NH<sub>4</sub>HCO<sub>3</sub> buffer (45 min at 56°C). Alkylation was performed with a solution of 55 mM iodoacetamide in 100 mM NH<sub>4</sub>HCO<sub>3</sub> (30 min at room temperature in the dark) followed by in-gel digestion with 20 µl of trypsin (10 ng/µl) in 50 mM NH<sub>4</sub>HCO<sub>3</sub> (overnight at 37°C). Subsequently, the peptides were extracted in NH<sub>4</sub>HCO<sub>3</sub> buffer with 5% formic acid. Samples were vacuum-dried and reconstituted in 25 µl of sample preparation solution (98% water, 2% acetonitrile, and 0.5% formic acid).

## Mass Spectrometric Analysis

The protein digest spectrum was acquired on a Q-STAR Elite (QTOF) mass spectrometer equipped with Applied Biosystems (Waltham, MA, USA) Nano Spray II ion source. The solution containing peptides was injected into Nano-LC through an autosampler system and then eluted with a gradient of water

and acetonitrile. The flow rate of nano-reverse phase column (Michrom C18 5  $\mu$  300 Å) was 400 nL/min for 1 h. This nano-reverse phase column is connected to the Nano Spray ESI- QTOF system (Qstar Elite, Applied Biosystems, Waltham, MA, USA). Eluted peptides from the column were ionized using ESI source with ion spray voltage 2250 V and temperature 120°C. Ionized peptides were analyzed by one full MS scan and four consecutive product ion scans of the four most intense peaks, using rolling collision energy. Ion fragmentation included a selection of ions in m/z range: >400 and <1600, of charge state of +2 to +5, exclusion of former target ions for 30 s, accumulation time of 1 s for a full scan and 2 s for MS/MS.

The resultant MS/MS data was searched against the NCBI non-redundant database<sup>1</sup> option in the Protein Pilot 5.0 software (AB Sciex, Haryana, India) for the identification of proteins. During the analysis, in the search, parameter scope was allowed to include modification of cysteine by iodoacetamide and biological modifications programmed in algorithm were allowed. Mass tolerance for precursor ion and fragment ions were set to 100 ppm and 0.2 Da, respectively. The number of missed cleavages permitted was two. Differentially expressed proteins during hypoxia, in H37Rv, S7, and S10 were categorized according to function based on "TubercuList" database<sup>2</sup>.

## Study Subjects and Antigens Used for *In vitro* Whole Blood Culture

This study was approved by the Institutional Ethics Committee of National Institute for Research in Tuberculosis (NIRT) and informed consent was obtained from all study participants. We included 20 individuals (10 HHC and 10 PTB) patients who visited in Government Thiruvateeswarar Hospital of Thoracic Medicine, Otteri, Chennai. HHC participants, generally parents, spouses, and children, were selected if only they shared the living quarters for a minimum of 3 months with 10 h per day of close contact with sputum positive, active TB patients (index TB case) who were naive for anti-tubercular therapy. Their infection state was confirmed by positive QuantiFERON-TB Gold in Tube (QFT-IT; Cellestis, a company of Qiagen GmbH) results. All HHC were negative for acid fast bacilli sputum smear microscopy and had a negative chest x-ray indicating no active TB disease symptoms.

Pulmonary TB patients were recruited by positive sputum smear microscopy. Three sputum samples were collected at various days from all the active TB patients for smear and cultures. All PTB patients were positive for culture and were excluded if they had symptoms of immunosuppressive disease like diabetes, HIV, or other co-infections. All PTB patients were naive for anti-tuberculosis treatment.

Five ml of blood was collected and diluted 1:1 with RPMI1640 (Sigma-Aldrich, St. Louis, MO, USA) medium with penicillin/streptomycin (100 U/100 mg/mL), L-glutamine (2 mM), and HEPES (10 mM) and distributed into tissue culture plates. The cultures were then stimulated, at the final concentration of 5  $\mu$ g/ml as determined earlier (Kumar et al.,

2010), with *M. tuberculosis* ESAT-6 (E6), which was received from CSU, USA and Lpd (Rv0462) obtained by *in vitro* cloning and overexpression (Devasundaram et al., 2014) along with no antigen control (unstimulated). Phytohemeagglutinin (PHA) was used as mitogen control to show the proliferative capacity of lymphocytes from donors at a concentration of 1  $\mu$ g/ml. Plates were incubated for 16 h with the prior addition of Brefeldin A (10 mg/mL) at fourth h. After incubation, cells were harvested with PBS and RBC were lysed with BD FACS lysing solution (Becton Dickinson, San Jose, CA, USA) as prescribed by the manufacturer. The cells were then stained for memory T cell markers.

## Surface Staining of Memory T Cells and Data Analysis

A single cell suspension of antigen stimulated cells was prepared with PBS and stained with antibodies against CD3 and CD4 conjugated to PerCP 5.5, APC-Cy7 (BD Biosciences, USA), respectively. Cells were then stained with PE-CD62L, a reliable surface marker to distinguish central and effector memory T cell subtypes. In addition, APC-CD45RA was used to differentiate naive cells from central and memory T cells. All antibodies were used at a final concentration of 5  $\mu$ l/1 million cells and were incubated for 20–30 min in the dark followed by washing with PBS. Stained cells were immediately analyzed on a FACSCanto II flow cytometer with FACSDiva software, version 6 (Becton Dickinson and Company, Cockeysville, MD, USA). A total of 100,000 lymphocyte events were recorded via forward and side scatter and data were analyzed in Flow Jo software (TreeStar). All data were depicted as the percentage of CD4<sup>+</sup> T cells expressing memory surface markers.

CD3<sup>+</sup> CD4<sup>+</sup> T-cells were gated for expression of CD45RA and CD62L and defined as central memory (CD45RA<sup>-</sup> CD62L<sup>+</sup>), effector memory (CD45RA<sup>-</sup> CD62L<sup>-</sup>), naive (CD45RA<sup>+</sup> CD62L<sup>+</sup>), and CD45RA<sup>+</sup> effectors (T<sub>EMRA</sub>; CD45RA<sup>+</sup> CD62L<sup>-</sup>).

For all antibodies utilized, fluorescence-minus-one (FMO) controls were used to define positive and negative boundaries. Compensation was calculated with signals from fluorochrome monoclonal antibodies linked to CompBeads (purchased from BD Biosciences, USA). The Mann-Whitney *U* test was performed using GraphPad Prism software (version 5.0; GraphPad Prism) with *p* < 0.05 considered to be statistically significant.

## RESULTS

### Growth Patterns in Aerobic and Anaerobic Cultures of H37Rv, S7, and S10

At intervals, tubes (both aerobic and anaerobic) were removed for growth measurements at OD<sub>600nm</sub> and methylene blue decolorization was also monitored. Similar patterns of growth curves were observed in both aerobic and anaerobic *in vitro* cultures. Complete decolorization of methylene blue indicated

<sup>1</sup><http://www.ncbi.nlm.nih.gov/refseq/>

<sup>2</sup><http://tuberculist.epfl.ch/>

oxygen depletion in the anaerobic cultures (exemplary images of **Figure 1A**). Initial growth patterns were similar up to 5 days for both aerobic and anaerobic cultures in all three strains (**Figures 1A–C**) but growth declined from day 9 in hypoxic culture tubes. Decolorization of methylene blue was also similar in three strains under hypoxia. The “blank tube” remained the same color till the end of incubation as no bacterium was inoculated, as observed in our earlier experiment (Devasundaram et al., 2015).

## Protein Expression during Hypoxia in H37Rv, S7, and S10

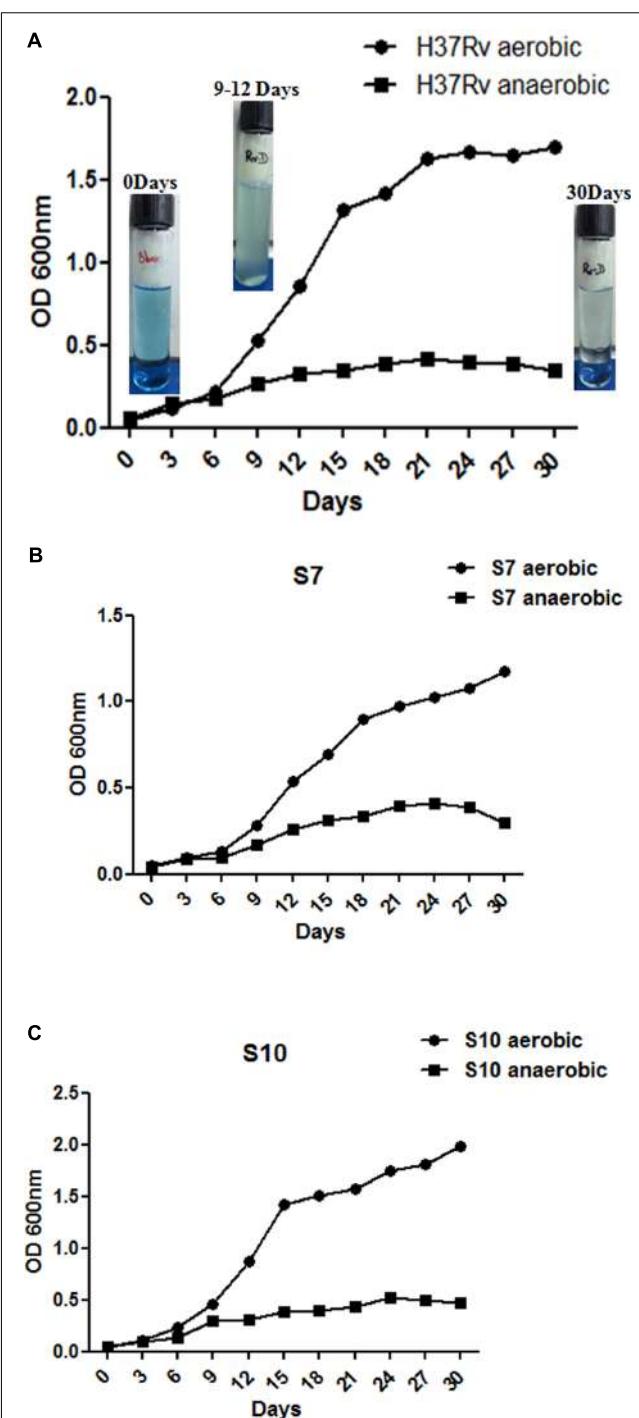
Aerobic and anaerobic cultures were harvested at 30 days and complete methylene blue decolorization in anaerobic cultures was observed between 25 and 30 days.

Two-dimensional electrophoresis gels were stained with CBB R 250 as prescribed (Sharma et al., 2010; Kumar B. et al., 2013). Spots with a consistent increase in intensity (over-expressed spots) and spots that emerged only during hypoxia (newly appeared) were selected and identified by mass spectrometry. The majority of cytosolic proteins from *M. tuberculosis* H37Rv, the clinical strains S7 and S10 were focused in the acidic pH range of 4–6.5. 2-DE gels of cytosolic proteins under aerobic and anaerobic growth conditions are shown in **Figure 2**.

By comparing the protein spots between aerobic and anaerobic cultures of H37Rv, a total of 15 spots either over-expressed (encircled) or newly appeared (boxed) during hypoxia were identified, designated as RvD (H37Rv dormant) and numbered sequentially (**Figures 2A,B**). The isoelectric point (pI) and molecular weight (Mr) of the proteins were identified on the gel against each spot position. Details of the open reading frame (ORF) number, predicted gene product, number of matched peptides, and the percentage sequence coverage obtained for each protein spots are given in **Table 1**. A few differentially expressed protein spots, identified by PDQuest from hypoxic conditions, were not characterized by mass spectrometry. This was either due to low concentration or because a confident peptide match was not obtained during the peptide search in Protein pilot software. The majority of spots in the gel were of single protein identity while a few proteins were found to exist as multiple spots (spot RvD1 and RvD2 in **Figure 2B**).

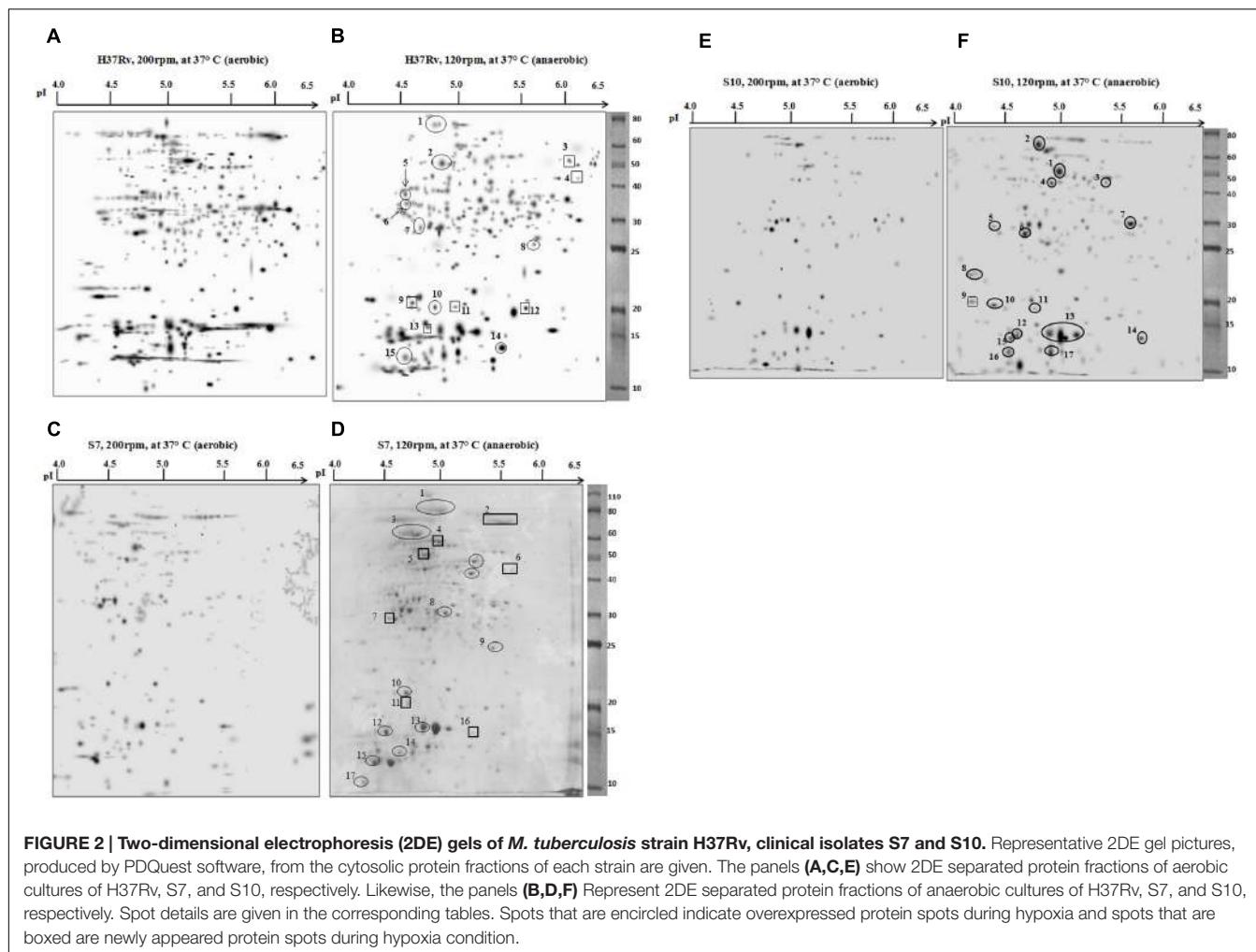
Among the 15 protein spots, 9 (RvD 1, 2, 5, 6, 7, 8, 10, 14, and 15) were over-expressed as a result of hypoxia in H37Rv and characterized as Rv0440 (GroEL2), Rv1240 [probable malate dehydrogenase (MDH)], Rv2145c (Wag31), Rv3028c (Flavoprotein), Rv1886c (Ag85B), Rv2780 (alanine dehydrogenase), Rv0854 (conserved protein), Rv2445c (diphosphate kinase), and Rv3418c (GroES), respectively. Newly appeared spots during depleted oxygen conditions in H37Rv were RvD 3, 4, 9, 11, 12, and 13 and identified as Rv0462 (dihydrolipoamide dehydrogenase), Rv2953 (enoyl reductase), Rv0054c (single strand stabilizing protein), Rv2185c (TB16.3), Rv1284 ( $\beta$ -carbonic anhydrase), and Rv2031c [HspX (Heat Shock Protein)], respectively.

Eleven protein spots increased in intensity (over-expressed) and six newly appeared spots were identified in S7 during



**FIGURE 1 |** Growth curves for *Mycobacterium tuberculosis* H37Rv aerobic and anaerobic cultures over the time course studied.

(A–C) Show growth patterns of H37Rv, S7, and S10, in duplicates, under aerated and anaerobic culture conditions, respectively. Aerated cultures of all three strains were obtained by growing at 37°C at 200 rpm with loose cap tubes. Dormant cultures (denoted as “D” to the suffix of each strain name) were obtained by growing all three strains for 30 days at 120 rpm in 20- by 125 mm screw-cap tubes containing MB7H9 broth. Cultures were stirred with 8 mm magnetic bars. Exemplary photos of H37Rv anaerobic cultures are given inside the graph. Mean values with standard error from duplicate cultures, at optical density (OD), 600 nm are shown.



**FIGURE 2 |** Two-dimensional electrophoresis (2DE) gels of *M. tuberculosis* strain H37Rv, clinical isolates S7 and S10. Representative 2DE gel pictures, produced by PDQuest software, from the cytosolic protein fractions of each strain are given. The panels (**A,C,E**) show 2DE separated protein fractions of aerobic cultures of H37Rv, S7, and S10, respectively. Likewise, the panels (**B,D,F**) Represent 2DE separated protein fractions of anaerobic cultures of H37Rv, S7, and S10, respectively. Spot details are given in the corresponding tables. Spots that are encircled indicate overexpressed protein spots during hypoxia and spots that are boxed are newly appeared protein spots during hypoxia condition.

hypoxia, designated as S7D (dormant S7) and sequentially numbered in the gel (**Figures 2C,D**). Their identifications are given in **Table 2**. Interestingly, Rv0440-GroEL2 (spot no. S7D1) was also identified in H37Rv (spot RvD1) under hypoxia, hence considered to be a common protein spot between S7 and H37Rv. In addition, Rv2445c was also identified as over-expressed and common to both S7 (S7D16) and H37Rv (RvD14).

Rv2953, trans-acting enoyl reductase, appeared newly in both S7 (S7D6) and H37Rv (RvD4) during hypoxia. Rv2031c which was identified as a newly appeared spot in H37Rv (RvD13) was identified as overexpressed and present as multiple spots in S7 during hypoxia (S7D13 and S7D14).

Protein expression in the clinical strain S10, under aerobic and anaerobic conditions is given in **Figures 2E,F**. Hypoxia induced protein spots in S10 were designated as S10D (dormant S10), numbered sequentially and spot characterization by mass spectrometry is given in **Table 3**. As observed in H37Rv and S7, Rv0440-GroEL2 was also found to be overexpressed under hypoxia in S10 (S10D1). Hence, it is grouped with “common spots.”

A well known hypoxia protein  $\alpha$ -crystalline, encoded by Rv2031c, was also found to be overexpressed in S10 (S10D13)

and found to be a “common spots” in all three strains of the present study. Apart from this, the protein spot Rv0462 (dihydrolipoamide dehydrogenase) and Rv1240 (MDH) were identified as common spots between H37Rv and S10. Magnified regions of gel portions are given in **Figure 3** for better visualization and representative MS/MS data for a randomly selected spot from each strain is given in Supplementary Figure S1.

## Functional Classification of Differentially Expressed Proteins

The majority of differentially expressed proteins from H37Rv were predicted to be involved in “Intermediary metabolism and respiration” (Rv0462, Rv1240, Rv1284, Rv2445c, and Rv3028c). In case of S7, the majority of proteins were classified under “lipid metabolism” (Rv0405, Rv0632c, Rv1679, Rv2953, and Rv3667) and in S10, expressed proteins were predicted to be involved in virulence, detoxification, and adaptation (Rv0350, Rv0351, Rv0440, Rv1636, Rv2031c, and Rv3418c).

Proteins involved in cell wall synthesis (Rv2145c, Rv3875) and conserved hypothetical proteins were also found to be

**TABLE 1 | Overexpressed and newly appearing proteins identified by mass spectrometry from *Mycobacterium tuberculosis* laboratory strain H37Rv under hypoxia compared to aerated cultures.**

Spot no	Rv no	Protein Name	Predicted M.Wt (kDa)	Predicted pI	Coverage (%)	Peptide matched
RvD1	Rv0440	GroEl2	65	4.8	7.60%	TDDVAGDGTTATVLAQALVR, GLNALADAVK
RvD2	Rv1240	Probable malate dehydrogenase (MDH)	48	4.9	16.7%	NAAEVVNDQAWIEDEFIPTVAK, SDLLEANGAIFTAQGK
RvD3	Rv0462	Dihydrolipoamide dehydrogenase (LpdC)	50	5.8	35%	DAKAFGISGEVTFDYGIAYDR, VAGVHFLMK, LPVGTLSANVTYEEQILSR, ATFCQPNVASFGLTEQQAR, HGELLGGHLVGHDAEELLPELTQQRWDLTASELAR
<b>RvD4</b>	Rv2953	<i>Trans</i> -acting enoyl reductase	42	6.2	20%	MMLGPNAADWPLILADASQPLTEAMAAR, STAVLLAQSGLALALDRDR
RvD5	Rv2145c	Wag31	35	4.5	45%	INELDQELAAGGGAGVTPQATQAIPAYEPEPGK, LTNTAKAESDK, TRLKTYPESQLEELGQ, GSAAPVDSNADAGGFQDFNRGK
RvD6	Rv3028c	Probable electron transfer flavo protein	32	4.5	8%	TVSPQLYIALGISGAIQHR
RvD7	Rv1886c	Ag85B mycolyl transferase	30	5.3	24%	VQFQSGGNNSPAVYLLDGLR, NDPTQQIPK
RvD8	Rv2780	Secreted L-alanine dehydrogenase ALD	32	6.1	21%	GHEVLIQAGAGEGSAITDADFK, ADLVIGAVLVPGA, TSTYALTNTATMPVYLELADHGWR
<b>RvD9</b>	Rv0054c	Single-strand binding protein Ssb	23	4.6	12%	FTPSGAAVANFTVASTPR, ASRSGGFGSGSR
RvD10	Rv0854	Conserved protein	20	4.9	6%	TAGITDEQWVAYSWTDR
<b>RvD11</b>	Rv2185c	Conserved protein TB16.3	19	5.1	18%	TTQTYYIDADPGEVMK, EVEILEADDEGYPKR, GSGTEVTYELAVDLAVPMIGMLK
<b>RvD12</b>	Rv1284	Beta-carbonic anhydrase	18	5.4	21.50%	TVTDDYLANNVDYASGFK, GFVFDVATGK
<b>RvD13</b>	Rv2031c	Heat shock protein (HSPX)	16	4.9	29.0%	ATTLPVQRHPR, AELPGVDPDKDVDIMVR, SEFAYGSFVR, HIQIR
RvD14	Rv2445c	Probable nucleoside diphosphate kinase K (NDK)	15	5.3	22.80%	GLTIAALQLR
RvD15	Rv3418c	10 kDa chaperonin GROES	12	4.4	58%	IPLDVAEGDTVIYSK, RIPLDVAEGDTVIYSK

Details of proteins characterization by mass spectrometry from the anaerobic cultures (designated as RvD) of H37Rv. Bold underlined spot numbers indicate the newly appearing proteins during hypoxia.

differentially expressed during hypoxia and their distributions among the strains are given in **Table 4**.

## Higher Frequency of Memory T Cell Markers in Circulation of Healthy Infected Individuals

Antigen specific memory T cells were analyzed in the stimulated blood culture of HHC and PTB. The gating strategy followed (**Figure 4A**) along with representative flow cytometry is given in **Figure 4B**. Significantly higher antigen specific memory cells were present in HHC, ESAT-6 ( $p < 0.005$ ), and for Lpd ( $p < 0.0005$ ) **Figure 4C**, when compared to PTB with respect to central memory cell phenotype. Mitogen response, shown only in the representative flow diagram, was equal in both HHC and PTB showing proliferative capacity was not defective. These antigens specific Th1, Th2, and poly functional T cell response was also found to be significantly higher in HHC ( $N = 30$ ) when compared to PTB ( $N = 30$ ; Communicated manuscript).

## DISCUSSION

Laboratory strains (such as H37Rv) might not completely mimic the virulence of naturally occurring clinical strains. Vaccines based on proteins that are predicted to be over-expressed only in the laboratory strain might not prevent infections caused by virulent strains. Hence, we have included the most prevalent clinical *M. tuberculosis* strains (S7 and S10) to study protein expression under hypoxia. In our experiments, growth related differences were minimized between the strains by terminating all cultures during late exponential growth (25–30 days) and expression of already reported hypoxia genes and DosS-DosR regulon genes shows faithful achievement of hypoxia (Devasundaram et al., 2015). For a few protein spots, the percentage of peptide coverage was less than 10, which is still acceptable (Mehaffy et al., 2010; Ang et al., 2014) and could be due to the lowest concentration of peptides obtained from the protein spots.

The expression of molecular chaperones (assist proper protein folding) and proteases (degrades unfolded proteins) was observed to be an adaptive bacterial response during various

**TABLE 2 | Details of overexpressed and newly appearing proteins identified by mass spectrometry from *M. tuberculosis* laboratory strain S7 under hypoxia compared to aerated cultures.**

Spot no	Rv no	Protein Name	Predicted M.Wt (kDa)	Predicted pI	Coverage (%)	Peptide matched
S7D1	Rv0440	GroEL2	65	4.8	11%	GLNALADAVK, EIELEDPYEK
<b>S7D2</b>	Rv3667	Acetyl coA synthetase	80	5.9	10%	LLITSDGQFRR, RGHPAPLKDAADEAVSQPDSPVEHVLVRRTGIDVSWNDER, AEVAEAIISPIARPR
S7D3	Rv0405	Polyketide synthase-6 (pkS6)	110	5.4	1%	TASALAAQAGRLGR
<b>S7D4</b>	Rv2220	Glutamine synthetase	62	5.3	16.70%	SVFDDGLAFDGSSIR, GGYFPVAPNDQYVDLR, MLTNINSGFILEK, LPVGYEAPINLVYSQR
<b>S7D5</b>	Rv1679	Possible acyl-CoA dehydrogenase FadE16	55	4.9	23%	VDTDCAFPAEAVDALRK, ASVNDAAITTESAMR
<b>S7D6</b>	Rv2953	Trans-acting enoyl reductase	42	6.1	20%	MMLGPNAADWPLILADASQPLTEAMAAR, STAVILLAQSGLALALDRDR
<b>S7D7</b>	Rv3029c	Probable electron transfer flavo protein beta subunit	31	4.8	4.90%	EAADAVLDEINER
S7D8	Rv3060c	GntR family transcriptional regulator	33	5.2	2.30%	TYGASGMPSR
S7D9	Rv0632c	Probable enoyl-CoA hydratase EchA3	25	5.8	13.80%	VFSGGFDLK, GGFELAYR, ILTGEVQPAIDMLR
S7D10	gil 183985444	Integral membrane protein	21	4.7	0.90%	GPIPFDAPRER
<b>S7D11</b>	gil 518086696	Cyclase [M. tuberculosis]	20	4.8	11.10%	AIADIEAYPQWISEYK
S7D12	Rv3716c	Conserved protein	15	4.4	19%	VVDPDDIELQDLIVGAMR
S7D13 & 14	Rv2031c	Heat shock protein (HspX)	16	4.6	29%	ATTLPVQRHPR, AELPGVDPDKDVIMVR, SEFAYGSFVR
S7D15	Rv3418c	10 kDa chaperonin (GroES)	12	4.2	58%	IPLDVAEGDTVIYSK, RIPLDVAEGDTVIYSK
S7D16	Rv2445c	Probable nucleoside diphosphate kinase K (NDK)	15	5.3	22.80%	GLTIAALQLR
S7D17	Rv3875	ESAT-6	12	4.3	17.90%	WDATATELNNAQNQLAR

Details of proteins characterization by mass spectrometry from the S7 anaerobic cultures (designated as S7D). Bold underlined spot numbers indicate the newly appearing proteins during hypoxia. For few spots gene number was not described in protein search tool, in such case accession number was given.

environmental stresses (Gumber and Whittington, 2009). DosR antigens and members of the small heat-shock protein family, also known as chaperones, were identified as over-expressed during hypoxia in all three strains H37Rv, S7, and S10. The small heat shock protein family includes chaperones like Rv2031c [ $\alpha$ -crystallin (ACR) protein–HspX], Rv3418c (10 kDa chaperonin GroES), and Rv0440 (Hsp65), encoded in the second copy of a gene for a 60 kDa chaperonin in the *M. tuberculosis* genome and all are identified in the present study.

GroEL2 (Rv0440), also a chaperone, prevents protein misfolding, and promotes the refolding and proper assembly of unfolded/misfolded polypeptides generated during stress conditions (Sharma et al., 2010). This protein was predicted to be over-expressed in all three strains under hypoxia. GroES (Rv3418c), also a chaperone protein, was already reported to be expressed by Rustad et al. (2009) and our results also confirmed overexpression during hypoxia. Collectively, this shows hypoxia associated proteins' expression in our experimental model.

A number of reports observed the expression of HspX (Rv2031c) under hypoxia (Starck et al., 2004; Siddiqui et al.,

2011), and its expression was also observed during aerobic growth based on H37Rv as a model strain (Desjardin et al., 2001). Adding to these reports, we also observed no change in expression of HspX in H37Rv under hypoxia. Since HspX expression was not unique to hypoxia, as observed based on H37Rv protein analysis, many researchers might exclude them for vaccine development. But, in contrast to H37Rv, HspX was identified as over-expressed in both the clinical isolates (S7 and S10) during hypoxia indicating its possible role during hypoxia in clinical isolates. This further highlights the need to include the most relevant clinical strains for *in vitro* experiments to minimize bias.

Heat shock protein accumulates as the dominant marker during LTBI, which is estimated to affect almost one-third of the world's population (Dubaniewicz et al., 2013). Hence, HspX can still be considered a hypoxia associated protein. In our experiments, an additional new spot of HspX was found adjacent to the actual HspX spot position which might have been a result of proteolytic degradation or post-translational modification. Different spots representing HspX (isoforms or degraded spots)

**TABLE 3 | Details of overexpressed and newly appearing proteins identified by mass spectrometry from *M. tuberculosis* laboratory strain S10 under hypoxia compared to aerated cultures.**

Spot no	Rv no	Protein Name	Predicted M.Wt (kDa)	Predicted pI	Coverage (%)	Peptide matched
S10D1	Rv0440	GroEl2	65	4.8	11%	TDDVAGDGTTATVLAQALVR
S10D2	Rv0350	Probable chaperone protein DNAK	75	4.9	23%	HMGSDWSIEIDGKK, GIPQIEVTFDIDANGIVHVTAK
S10D3	Rv0462	Dihydrolipoamide dehydrogenase (LpdC)	50	5.8	25%	VAGVHFLMK, ATFCQPNVASFGLTEQQAR
S10D4	Rv1240	Probable malate dehydrogenase (MDH)	48	4.9	16.7%	NAAEVVNDQAWIEDEFIPTVAK, SDLLEANGAIFTAQGK
S10D5	Rv0351	GrpE	27	4.2	31%	PDGNSGEQVTVDKRR, PVIGTVMRQGYQLGEQVLR
S10D6	Rv3029c	Probable electron transfer flavo protein ( $\beta$ -subunit)	31	4.8	4.90%	EAADAVLDEINER
S10D7	Rv2032	Conserved protein acg	30	5.7	7%	IDVIADDMRPELAAASK
S10D8	gil 33742142	Sequence 142 from patent US 6583266 (corresponding atpA)	23	4.2	3.70%	VVNPLGQPIDGR, GTTIASVRR
<b>S10D9</b>	Rv3841	Possible bacterioferritin (BfrB)	20	4.3	13.30%	VEIPGVDTVR, AGANLFELENFVAR
S10D10	Rv3841	Possible bacterioferritin (BfrB)	20	4.6	42%	NHAMMLVQHLLDR, VEIPGVDTVR, EALALALDQER, VALMATLVR, AGANLFELENFVAR, EVDVAPAASGAPHAAGGRL
S10D11	gil 518086696	Cyclase [ <i>M. tuberculosis</i> ]	20	4.8	20.80%	AIADIEAPQWSEYK, QSLSWTLESSLLK
S10D12	Rv3716c	Conserved protein	15	4.4	19%	VVDPPDDIELQDLIVGAMR
S10D13	Rv2031c	HspX	16	4.7	29%	ATTLPVQRHPR, AELPGVDPDKVDIMVR, SEFAYGSFVR
S10D14	Rv1636	Iron-regulated universal stress protein family protein TB15.3	14	5.8	37%	LIIASAYLPQHEDAR, DESYKVTGTAPIYEILHDAK, LLGSVPANVSRRAK
S10D15	Rv3418c	10 kDa chaperonin (GROES)	12	4.5	15%	IPLDVAEGDTVIYSK
S10D16	Rv3418c	10 kDa chaperonin GROES	12	4.5	27%	IPLDVAEGDTVIYSK, YNGEEYLILSAR
S10D17	Rv2626c	Hypoxic response protein 1 (Hrp1)	12	4.7	9%	DIMNAGVTCVGEHETLTAAQYMR

Mass spectrometry characterization of anaerobic proteins of S10 (designated as S10D). Underlined spot numbers indicate the newly appearing proteins during hypoxia. For few protein spots accession number was given since the gene number was not described in protein search tool used here.

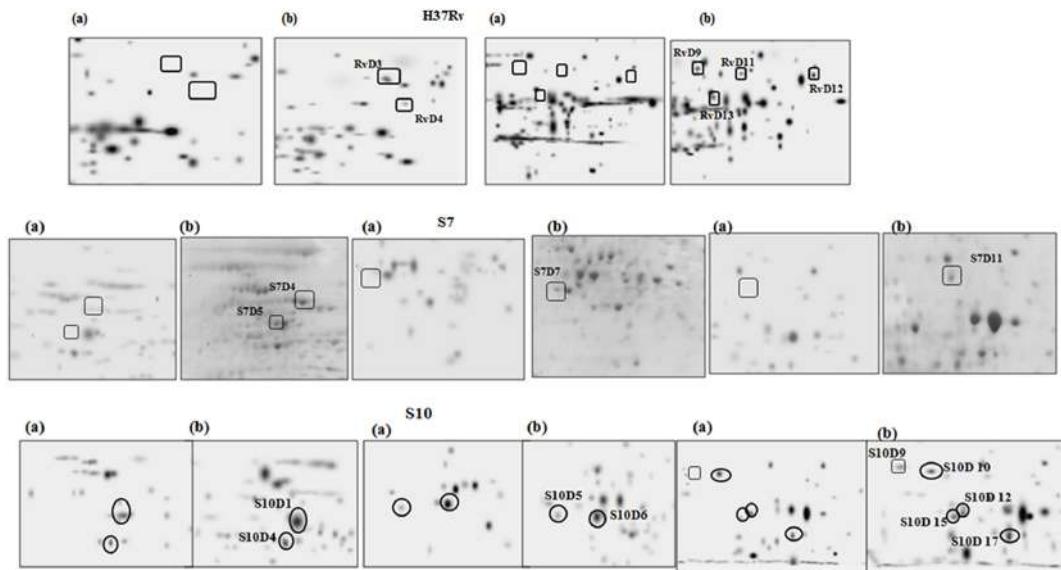
at lower molecular weights with differing pI have already been observed during 2-DE analysis of mycobacterial proteins (Betts et al., 2000).

Proteins of the Acr family were reported to be involved in the oxidative stress response in *M. tuberculosis* and persistence mechanisms in host macrophages (Stewart et al., 2005). Its expression during heat shock stress response was also observed (Gumber and Whittington, 2009); but its role under hypoxia in clinical isolates (S7 and S10) is evidently reported by our results. This observation extends support for the ACR family proteins to still serve as a better target for understanding the latency mechanisms of *M. tuberculosis*.

The lipid-rich outer cell wall layer, a unique feature of mycobacteria, contributes to their resilience and contains many compounds known to be involved in virulence (Camacho et al., 2001). Phthiocerol dimycocerosate (PDIM) constitutes a major virulence factor and functionally important surface-exposed lipid of *M. tuberculosis*. Biosynthesis of the PDIM core domain requires gene *ppsD* that encodes one of the five modular type-I

polyketide synthases (*ppsA-E*) of *M. tuberculosis* (Perez et al., 2004). But *ppsD* lacks functional enoyl reductase activity which is required for the synthesis of these lipids. Rv2953 encodes a trans-acting enoyl reductase that acts along with *ppsD* in phthiocerol and phenolphthiocerol biosynthesis and completes the final steps in PDIM biosynthesis (Simeone et al., 2007). We observed over-expression of Rv2953 in the most prevalent clinical isolates and H37Rv in our hypoxia experiment. This supports the idea that Rv2953 expression is needed during dormancy, where thickening of mycobacterial cell walls is generally observed, to complete cell wall lipid biosynthesis (PDIM; Murry et al., 2009). This strengthens the idea of targeting Rv2953 for antibody development to neutralize virulence factor biosynthesis or for use as a biomarker for latency.

The second common overexpressed protein between H37Rv and S7 was Rv2445c (nucleoside phosphate kinase-Ndk). Ndk is known for its interactions with host signaling molecules Rab5 and Rab7 (Hop et al., 2015), two small GTPases that control phagosome lysosome fusion, the consequence of which is



**FIGURE 3 |** Magnified regions of gel pictures showing proteins (a) aerobic cultures (b) anaerobic cultures from H37Rv, S7, and S10.

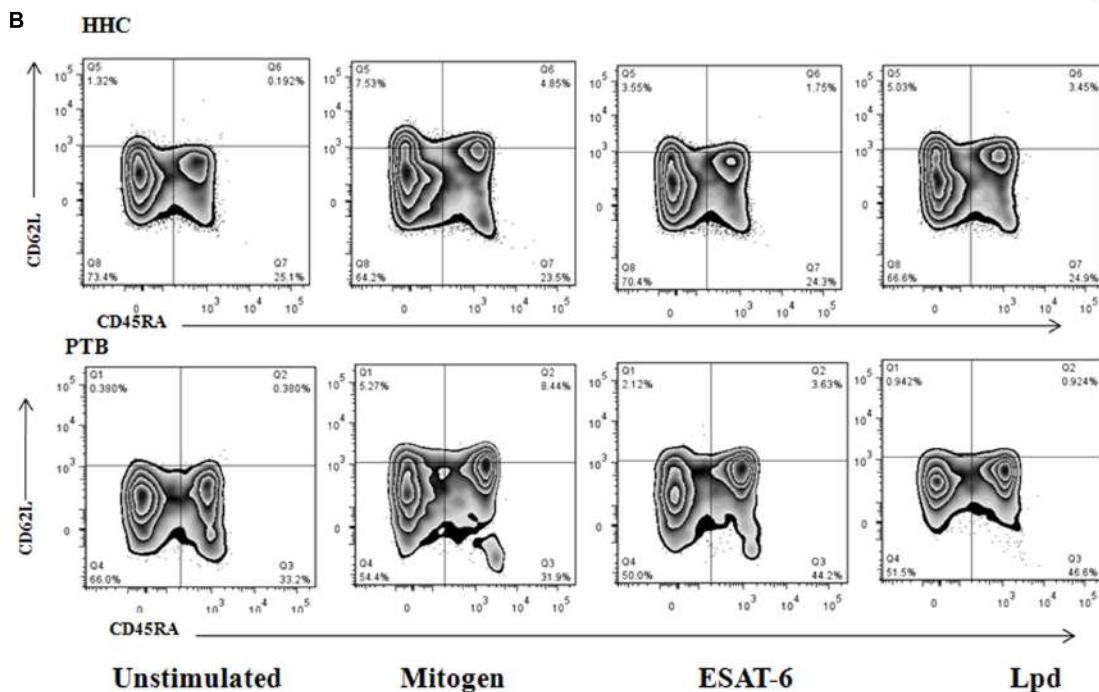
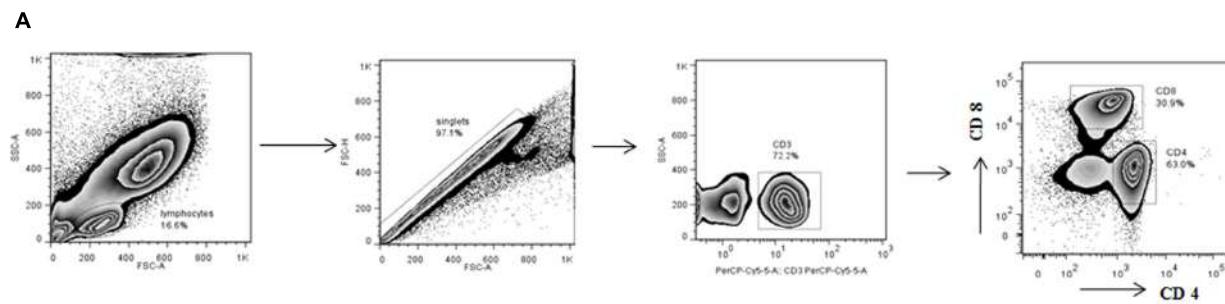
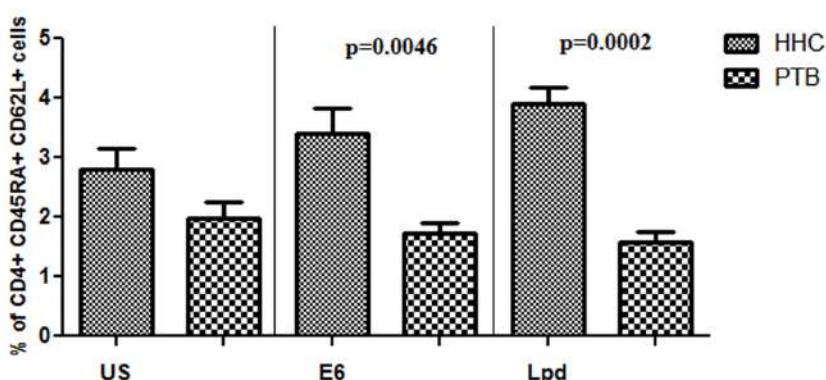
inhibition of phagolysosome fusion (Dar et al., 2011). Ndk can also interact with the host Rac1 signaling molecule that leads to an NADPH oxidase assembly; defect in this assembly would cause impaired reactive oxygen species production (Sun J. et al., 2010). In doing so, Ndk contributes to intracellular survival and subsequent establishment of mycobacterial infection. Though the GTPase activity of Ndk has been reported earlier (Chopra et al., 2003) the contribution of Ndk to *M. tuberculosis* pathogenesis has only recently been addressed (Sun J. et al., 2010; Sun J.N. et al., 2010). Ndk expression was not observed in any of the earlier *in vitro* based *M. tuberculosis* stress model studies. A recent nutrient starvation model study revealed Ndk expression, but with no significant difference in protein expression (Albrethsen et al., 2013). Though the role of Ndk in *M. tuberculosis* was well elucidated, its role during

hypoxia has not been clarified so far. We report for the first time, to our knowledge, on the possible role of Ndk during hypoxia. We identified Rv1240 (MDH) and Rv0462 (dihydrolipoamide dehydrogenase) as over expressed proteins common between H37Rv and S10. It is a well-known fact that during the persistent phase of infection, *M. tuberculosis* switches to tricarboxylic acid (TCA) metabolism to utilize fatty acids as a carbon source (Bishai, 2000). Thus, TCA metabolic enzymes are likely to have a role during mycobacterial dormancy. MDH (Rv1240), a TCA metabolic enzyme, was found to be unique to intraphagosomal mycobacteria (Mattow et al., 2006), which supports its overexpression under hypoxia, a condition observed within phagosomes. Along with these observations, our results also highlighted the possible role of Rv1240 during hypoxia.

**TABLE 4 |** Categorization of over expressed and newly appeared proteins under hypoxia in H37Rv, S7, and S10.

S. No	Classifications	H37Rv	S7	S10
1	Amino acid biosynthesis	None	Rv2220	None
2	Cell envelope	Rv0440, Rv2145c	Rv0440	Rv0440
3	Cellular processes	None	Rv0405, Rv3875	Rv1636
4	Central intermediary metabolism	Rv1284	None	None
5	Energy metabolism	Rv0462, Rv2780, Rv1240	Rv1679	Rv0462, Rv1240
6	Hypothetical proteins—Conserved	Rv0854, Rv2185c	Rv3716c	Rv2032
7	Protein fate	None	Rv3418c, Rv3875	Rv3418c
8	Regulatory functions	None	Rv3060c	None
9	Transport and binding proteins	None	None	Rv3841
10	Unclassified	Rv1886c, Rv2031c, Rv2445c, Rv2953, Rv3028c	Rv0632c, Rv2031c, Rv2445c, Rv2953, Rv3029c, Rv3667	Rv0350, Rv0351, Rv2031c, Rv2626c, Rv3029c

Over-expressed and newly appeared proteins from H37Rv, S7 and S10 during hypoxia was categorized according to their predicted biological functions based on the information given in *TuberuList* database.

**C**

**FIGURE 4 | (A)** General gating strategy for T cell is given. After selecting the singlets from lymphocyte population, CD3<sup>+</sup> cells were gated followed by CD4<sup>+</sup> cells were gated and selected for memory phenotyping. **(B)** CD4<sup>+</sup> cells were analyzed for the expression of CD62L and CD45RA surface markers to define the subtypes of memory cells. Antigen specific expansion of the central memory phenotype is evident in healthy household contacts (HHC) and representative flow diagram from HHC and pulmonary TB (PTB) is given. **(C)** Percentage of central memory cells specific to ESAT-6 (E6) and Lpd were shown as bar graphs representing mean with SEM. p-values were calculated using the Mann–Whitney U test and value <0.05 considered as significant.

Lpd (Rv0462), *M. tuberculosis*'s sole dihydrolipoamide dehydrogenase (Argyrou and Blanchard, 2001) is a flavin-adenine dinucleotide-containing NADH-dependent oxidoreductase that plays an essential role in intermediary metabolism as the E3 component of the pyruvate dehydrogenase complex. Earlier reports indicated that *M. tuberculosis* becomes vulnerable when Lpd (Rv0462) is inhibited (Bryk et al., 2002) and Lpd helps *M. tuberculosis* to resist host reactive nitrogen intermediates (Venugopal et al., 2011). These dynamics clearly convey a role for the metabolic enzymes Rv1240 and Rv0462 during hypoxia. Lpd gene expression under hypoxia was not reported earlier and a unique feature of our results are that they show the first evidence for Lpd expression during hypoxia as well as its expression in clinical strains of *M. tuberculosis*.

The electron transfer flavoprotein β subunit (Rv3029c) and a conserved hypothetical protein (Rv3716c) were found exclusively in both the clinical isolates S7 and S10 as over-expressed proteins during hypoxia, but not in H37Rv. Rv3029c is a well-known gene that participates in β-oxidation of fatty acids (Covert et al., 2001) and produces energy when fatty acids are used as the sole carbon source during persistence. Not much data is available for Rv3716c, but it was reported under "contact specific antigens" in a study conducted on healthy contacts of TB and PTB patients using H37Rv culture filtrate antigen fractions (Deenadayalan et al., 2010).

Rv3667 (acetyl CoA synthetase) and Rv2220 (glutamine synthetase), well known genes of hypoxia from H37Rv reported by many studies (Sherman et al., 2001; Rustad et al., 2009) were found only in the S7 strain, in addition to Rv3875 (ESAT-6) and Rv1679 (FadE6). In contrast, gene Rv0405 (polyketide synthase-6) reported as repressed under hypoxia with H37Rv (Rustad et al., 2009) was found to be expressed during hypoxia in the clinically prevalent strain S7. Genes Rv1679 (possible FadE16), Rv3060c (GntR family transcription regulator), Rv0632c (probable enoyl-coA hydratase) were not found to be reported from H37Rv-based *in vitro* hypoxic model, but appeared in the S7 clinical strain. We were tempted to speculate on differences between the laboratory and clinically prevalent strains, given similar treatment conditions and to create a list of possible true representative antigens. But, these variations were found only with S7 and not with S10, which responded similarly to H37Rv.

Our earlier microarray study with total RNA from these three strains under aerobic and anaerobic conditions also found similar hypoxic gene expression patterns during hypoxia between S10 and H37Rv. Whereas gene expression in S7 differed from the laboratory strain H37Rv (Devasundaram et al., 2015). Thus, studying the most prevalent virulent strains would help to better understand their pathogenesis and minimize variations in targeting virulence factors for effective control of infectious diseases.

The complete list of genes that are predicted to be overexpressed and common in all three strains during hypoxia was discussed in our earlier microarray report. In contrast, the present mass spectrometry report is preliminary and total spot characterization has yet to be completed. Hence, the 134 common hits identified during microarray reporting might not correlate completely with our present mass spectrometry data. But, a few

characterized protein spots like Rv2953 (RvD4), Rv3060c (S7D8), Rv1240 (S10D4) was also found in our gene expression data (GEO accession no: GSE55863).

Cellular immune response in LTBI would reflect the type of immunity responsible for efficient disease control and serve as a good experimental model (Kumar et al., 2010). The availability of two antigens in our lab, ESAT-6 and Lpd allowed us to evaluate the predicted antigens from the established *in vitro* oxygen depletion model in human experimental set-up. Similar outcomes were observed with both peripheral blood mononuclear cells (PBMC) and whole blood (WB) during T lymphocyte assays. WB assays are advantageous compared to PBMC assays since they require less blood volume. Hence, we preferred to stimulate WB to assess T-lymphocyte response. Generally many researchers dilute the blood with RPMI to screen more antigens simultaneously at dilutions of 1:1 (Kumar N.P. et al., 2013), 1:2 (Kumar et al., 2010), 1:5 and 1:10 (Deenadayalan et al., 2013) and we followed 1:2 as was followed by (Kumar et al., 2010). Antigen specific central memory cells were identified in HHC showing their association during latency. It has already been shown that CD4<sup>+</sup> memory T cells are predominant functional subsets in LTBI (Pollock et al., 2013). The decreased frequency of central memory cells in PTB shows the absence or decline of antigen (ESAT-6, Lpd) specific memory cells in the circulation of active TB patients. This clearly suggests that these antigens are expressed in the dormancy state and have a possible role during latency. Since latent TB populations are considered to be "protective against active TB disease," antigens that are predominantly recognized by their sensitized T cells could be a potential target for vaccine development. Global searches have allowed the identification of more than a 100 potential virulence genes in pathogenic mycobacteria. But, controlling TB infection is still a major challenge due to the distinct behavior of the laboratory model strains and the clinical strains. Thus, it is crucial to study clinically relevant infectious strains to identify common, abundant proteins between strains for drug development and vaccines. Putative drug targets, vaccine candidates, and diagnostic markers for TB were identified by comparative proteome analyses of *M. tuberculosis* strains/clinical isolates of varying virulence. Thus genes/proteins observed from our results can be explored further for their use in diagnosis or vaccine development against TB.

## AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: AR and SD. All the experiments and analysis was performed: SD. Mycobacterial culture work and 2-DE work assisted: AG. All authors contributed equally for manuscript writing.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2016.01275>

## REFERENCES

- Albrethsen, J., Agner, J., Piersma, S. R., Hojrup, P., Pham, T. V., Weldingh, K., et al. (2013). Proteomic profiling of *Mycobacterium tuberculosis* identifies nutrient-starvation-responsive toxin-antitoxin systems. *Mol. Cell. Proteomics* 12, 1180–1191. doi: 10.1074/mcp.M112.018846
- Andersen, P. (2007). Vaccine strategies against latent tuberculosis infection. *Trends Microbiol.* 15, 7–13. doi: 10.1016/j.tim.2006.11.008
- Ang, K. C., Ibrahim, P., and Gam, L. H. (2014). Analysis of differentially expressed proteins in late-stationary growth phase of *Mycobacterium tuberculosis* H37Rv. *Biotechnol. Appl. Biochem.* 61, 153–164. doi: 10.1002/bab.1137
- Argyrou, A., and Blanchard, J. S. (2001). *Mycobacterium tuberculosis* lipoamide dehydrogenase is encoded by Rv0462 and not by the lpdA or lpdB genes. *Biochemistry* 40, 11353–11363. doi: 10.1021/bi010575o
- Barry, C. E. III, Boshoff, H. I., Dartois, V., Dick, T., Ehrhart, S., Flynn, J., et al. (2009). The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. *Nat. Rev. Microbiol.* 7, 845–855. doi: 10.1038/nrmicro2236
- Betts, J. C. (2002). Transcriptomics and proteomics: tools for the identification of novel drug targets and vaccine candidates for tuberculosis. *IUBMB Life* 53, 239–242. doi: 10.1080/15216540212651
- Betts, J. C., Dodson, P., Quan, S., Lewis, A. P., Thomas, P. J., Duncan, K., et al. (2000). Comparison of the proteome of *Mycobacterium tuberculosis* strain H37Rv with clinical isolate CDC 1551. *Microbiology* 146(Pt. 12), 3205–3216. doi: 10.1099/00221287-146-12-3205
- Bishai, W. (2000). Lipid lunch for persistent pathogen. *Nature* 406, 683–685. doi: 10.1038/35021159
- Boon, C., Li, R., Qi, R., and Dick, T. (2001). Proteins of *Mycobacterium bovis* BCG induced in the Wayne dormancy model. *J. Bacteriol.* 183, 2672–2676. doi: 10.1128/JB.183.8.2672-2676.2001
- Bryk, R., Lima, C. D., Erdjument-Bromage, H., Tempst, P., and Nathan, C. (2002). Metabolic enzymes of mycobacteria linked to antioxidant defense by a thioredoxin-like protein. *Science* 295, 1073–1077. doi: 10.1126/science.1067798
- Camacho, L. R., Constant, P., Raynaud, C., Laneelle, M. A., Triccas, J. A., Gicquel, B., et al. (2001). Analysis of the phthiocerol dimycocerosate locus of *Mycobacterium tuberculosis*. Evidence that this lipid is involved in the cell wall permeability barrier. *J. Biol. Chem.* 276, 19845–19854. doi: 10.1074/jbc.m100662200
- Chopra, P., Singh, A., Koul, A., Ramachandran, S., Drlica, K., Tyagi, A. K., et al. (2003). Cytotoxic activity of nucleoside diphosphate kinase secreted from *Mycobacterium tuberculosis*. *Eur. J. Biochem.* 270, 625–634. doi: 10.1046/j.1432-1033.2003.03402.x
- Covert, B. A., Spencer, J. S., Orme, I. M., and Belisle, J. T. (2001). The application of proteomics in defining the T cell antigens of *Mycobacterium tuberculosis*. *Proteomics* 1, 574–586. doi: 10.1002/1615-9861(200104)1:4<574::AID-PROT574>3.3.CO;2-8
- Dar, H. H., Prasad, D., Varshney, G. C., and Chakraborti, P. K. (2011). Secretory nucleoside diphosphate kinases from both intra- and extracellular pathogenic bacteria are functionally indistinguishable. *Microbiology* 157(Pt. 11), 3024–3035. doi: 10.1099/mic.0.049221-0
- Das, S., Paramasivan, C. N., Lowrie, D. B., Prabhakar, R., and Narayanan, P. R. (1995). IS6110 restriction fragment length polymorphism typing of clinical isolates of *Mycobacterium tuberculosis* from patients with pulmonary tuberculosis in Madras, south India. *Tuber. Lung Dis.* 76, 550–554. doi: 10.1016/0962-8479(95)90533-2
- Deenadayalan, A., Heaslip, D., Rajendiran, A. A., Velayudham, B. V., Frederick, S., Yang, H. L., et al. (2010). Immunoproteomic identification of human T cell antigens of *Mycobacterium tuberculosis* that differentiate healthy contacts from tuberculosis patients. *Mol. Cell. Proteomics* 9, 538–549. doi: 10.1074/mcp.M900299-MCP200
- Deenadayalan, A., Maddineni, P., and Raja, A. (2013). Comparison of whole blood and PBMC assays for T-cell functional analysis. *BMC Res. Notes* 6:120. doi: 10.1186/1756-0500-6-120
- Desjardin, L. E., Hayes, L. G., Sohaskey, C. D., Wayne, L. G., and Eisenach, K. D. (2001). Microaerophilic induction of the alpha-crystallin chaperone protein homologue (hspX) mRNA of *Mycobacterium tuberculosis*. *J. Bacteriol.* 183, 5311–5316. doi: 10.1128/JB.183.18.5311-5316.2001
- Devasundaram, S., Deenadayalan, A., and Raja, A. (2014). In silico analysis of potential human T cell antigens from *Mycobacterium tuberculosis* for the development of subunit vaccines against tuberculosis. *Immunol. Invest.* 43, 137–159. doi: 10.3109/08820139.2013.857353
- Devasundaram, S., Khan, I., Kumar, N., Das, S., and Raja, A. (2015). The influence of reduced oxygen availability on gene expression in laboratory (H37Rv) and clinical strains (S7 and S10) of *Mycobacterium tuberculosis*. *J. Biotechnol.* 210, 70–80. doi: 10.1016/j.biote.2015.04.017
- Dubaniewicz, A., Holownia, A., Kalinowski, L., Wybierska, M., Dobrucki, I. T., and Singh, M. (2013). Is mycobacterial heat shock protein 16 kDa, a marker of the dormant stage of *Mycobacterium tuberculosis*, a sarcoid antigen? *Hum. Immunol.* 74, 45–51. doi: 10.1016/j.humimm.2012.10.007
- Fang, X., Wallqvist, A., and Reifman, J. (2012). Modelling phenotypic metabolic adaptations of *Mycobacterium tuberculosis* H37Rv under hypoxia. *PLoS Comp. Biol.* 8:e1002688. doi: 10.1371/journal.pcbi.1002688
- Govender, L., Abel, B., Hughes, E. J., Scriba, T. J., Kagina, B. M., de Kock, M., et al. (2010). Higher human CD4 T cell response to novel *Mycobacterium tuberculosis* latency associated antigens Rv2660 and Rv2659 in latent infection compared with tuberculosis disease. *Vaccine* 29, 51–57. doi: 10.1016/j.vaccine.2010.10.022
- Gumber, S., and Whittington, R. J. (2009). Analysis of the growth pattern, survival and proteome of *Mycobacterium avium* subsp. *paratuberculosis* following exposure to heat. *Vet. Microbiol.* 136, 82–90. doi: 10.1016/j.vetmic.2008.10.003
- Hingley-Wilson, S. M., Lougheed, K. E., Ferguson, K., Leiva, S., and Williams, H. D. (2010). Individual *Mycobacterium tuberculosis* universal stress protein homologues are dispensable in vitro. *Tuberculosis (Edinb).* 90, 236–244. doi: 10.1016/j.tube.2010.03.013
- Hop, H. T., Simborio, H. L., Reyes, A. W., Arayan, L. T., Min, W., Lee, H. J., et al. (2015). Immunogenicity and protective effect of recombinant *Brucella abortus* Ndk (rNdk) against a virulent strain *B. abortus* 544 infection in BALB/c mice. *FEMS Microbiol. Lett.* 362, 1–6. doi: 10.1093/femsle/fnv003
- Kasprowicz, V. O., Churchyard, G., Lawn, S. D., Squire, S. B., and Lalvani, A. (2011). Diagnosing latent tuberculosis in high-risk individuals: rising to the challenge in high-burden areas. *J. Infect. Dis.* 204(Suppl. 4), S1168–S1178. doi: 10.1093/infdis/jir449
- Kumar, B., Sharma, D., Sharma, P., Katoch, V. M., Venkatesan, K., and Bisht, D. (2013). Proteomic analysis of *Mycobacterium tuberculosis* isolates resistant to kanamycin and amikacin. *J. proteomics* 94, 68–77. doi: 10.1016/j.jprot.2013.08.025
- Kumar, M., Meenakshi, N., Sundaramurthi, J. C., Kaur, G., Mehra, N. K., and Raja, A. (2010). Immune response to *Mycobacterium tuberculosis* specific antigen ESAT-6 among south Indians. *Tuberculosis* 90, 60–69. doi: 10.1016/j.tube.2009.10.003
- Kumar, N. P., Sridhar, R., Banurekha, V. V., Jawahar, M. S., Nutman, T. B., and Babu, S. (2013). Expansion of pathogen-specific T-helper 1 and T-helper 17 cells in pulmonary tuberculosis with coincident type 2 diabetes mellitus. *J. Infect. Dis.* 208, 739–748. doi: 10.1093/infdis/jit241
- Leistikow, R. L., Morton, R. A., Bartek, I. L., Frimpong, I., Wagner, K., and Voskuil, M. I. (2010). The *Mycobacterium tuberculosis* DosR regulon assists in metabolic homeostasis and enables rapid recovery from nonrespiring dormancy. *J. Bacteriol.* 192, 1662–1670. doi: 10.1128/JB.00926-09
- Mattow, J., Siejak, F., Hagens, K., Becher, D., Albrecht, D., Krah, A., et al. (2006). Proteins unique to intraphagosomally grown *Mycobacterium tuberculosis*. *Proteomics* 6, 2485–2494. doi: 10.1002/pmic.200500547
- Mehaffy, C., Hess, A., Prenni, J. E., Mathema, B., Kreiswirth, B., and Dobos, K. M. (2010). Descriptive proteomic analysis shows protein variability between closely related clinical isolates of *Mycobacterium tuberculosis*. *Proteomics* 10, 1966–1984. doi: 10.1002/pmic.200900836
- Murry, J. P., Pandey, A. K., Sasse, C. M., and Rubin, E. J. (2009). Phthiocerol dimycocerosate transport is required for resisting interferon-gamma-independent immunity. *J. Infect. Dis.* 200, 774–782. doi: 10.1086/605128
- Muttucumaru, D. G., Roberts, G., Hinds, J., Stabler, R. A., and Parish, T. (2004). Gene expression profile of *Mycobacterium tuberculosis* in a non-replicating state. *Tuberculosis* 84, 239–246. doi: 10.1016/j.tube.2003.12.006
- Perez, E., Constant, P., Laval, F., Lemassu, A., Laneelle, M. A., Daffe, M., et al. (2004). Molecular dissection of the role of two methyltransferases in the biosynthesis of phenolglycolipids and phthiocerol dimycocerosate in the

- Mycobacterium tuberculosis* complex. *J. Biol. Chem.* 279, 42584–42592. doi: 10.1074/jbc.M406134200
- Pollock, K. M., Whitworth, H. S., Montamat-Sicotte, D. J., Grass, L., Cooke, G. S., Kapembwa, M. S., et al. (2013). T-cell immunophenotyping distinguishes active from latent tuberculosis. *J. Infect. Dis.* 208, 952–968. doi: 10.1093/infdis/jit265
- Rajavelu, P., and Das, S. D. (2005). Th2-type immune response observed in healthy individuals to sonicate antigen prepared from the most prevalent *Mycobacterium tuberculosis* strain with single copy of IS6110. *FEMS Immunol. Med. Microbiol.* 45, 95–102. doi: 10.1016/j.femsim.2005.02.011
- Rustad, T. R., Sherrid, A. M., Minch, K. J., and Sherman, D. R. (2009). Hypoxia: a window into *Mycobacterium tuberculosis* latency. *Cell. Microbiol.* 11, 1151–1159. doi: 10.1111/j.1462-5822.2009.01325.x
- Sharma, P., Kumar, B., Gupta, Y., Singhal, N., Katoh, V. M., Venkatesan, K., et al. (2010). Proteomic analysis of streptomycin resistant and sensitive clinical isolates of *Mycobacterium tuberculosis*. *Proteome sci.* 8, 59. doi: 10.1186/1477-5956-8-59
- Sherman, D. R., Voskuil, M., Schnappinger, D., Liao, R., Harrell, M. I., and Schoolnik, G. K. (2001). Regulation of the *Mycobacterium tuberculosis* hypoxic response gene encoding alpha-crystallin. *Proc. Natl. Acad. Sci. U.S.A.* 98, 7534–7539. doi: 10.1073/pnas.121172498
- Siddiqui, K. F., Amir, M., and Agrewala, J. N. (2011). Understanding the biology of 16 kDa antigen of *Mycobacterium tuberculosis*: scope in diagnosis, vaccine design and therapy. *Crit. Rev. Microbiol.* 37, 349–357. doi: 10.3109/1040841X.2011.606425
- Simeone, R., Constant, P., Malaga, W., Guilhot, C., Daffe, M., and Chalut, C. (2007). Molecular dissection of the biosynthetic relationship between phthiocerol and phthiodiolone dimycocerosates and their critical role in the virulence and permeability of *Mycobacterium tuberculosis*. *FEBS J.* 274, 1957–1969. doi: 10.1111/j.1742-4658.2007.05740.x
- Starck, J., Kallenius, G., Marklund, B. I., Andersson, D. I., and Akerlund, T. (2004). Comparative proteome analysis of *Mycobacterium tuberculosis* grown under aerobic and anaerobic conditions. *Microbiology* 150(Pt 11), 3821–3829. doi: 10.1099/mic.0.27284-0
- Steenken, W., Oatway, W. H., and Petroff, S. A. (1934). Biological studies of the tubercle *Bacillus*: iii. Dissociation and pathogenicity of the R and S variants of the human tubercle *Bacillus* (H(37)). *J. Exp. Med.* 60, 515–540. doi: 10.1084/jem.60.4.515
- Stewart, G. R., Patel, J., Robertson, B. D., Rae, A., and Young, D. B. (2005). Mycobacterial mutants with defective control of phagosomal acidification. *PLoS Pathog.* 1:269–278. doi: 10.1371/journal.ppat.0010033
- Sun, J., Wang, X., Lau, A., Liao, T. Y., Bucci, C., and Hmama, Z. (2010). Mycobacterial nucleoside diphosphate kinase blocks phagosome maturation in murine RAW 264.7 macrophages. *PLoS ONE* 5:e8769. doi: 10.1371/journal.pone.0008769
- Sun, J. N., Solis, N. V., Phan, Q. T., Bajwa, J. S., Kashleva, H., Thompson, A., et al. (2010). Host cell invasion and virulence mediated by *Candida albicans* Ssa1. *PLoS Pathog.* 6:e1001181. doi: 10.1371/journal.ppat.1001181
- Venugopal, A., Bryk, R., Shi, S., Rhee, K., Rath, P., Schnappinger, D., et al. (2011). Virulence of *Mycobacterium tuberculosis* depends on lipoamide dehydrogenase, a member of three multienzyme complexes. *Cell Host Microbe* 9, 21–31. doi: 10.1016/j.chom.2010.12.004
- Voskuil, M. I., Visconti, K. C., and Schoolnik, G. K. (2004). *Mycobacterium tuberculosis* gene expression during adaptation to stationary phase and low-oxygen dormancy. *Tuberculosis* 84, 218–227. doi: 10.1016/j.tube.2004.02.003
- Wayne, L. G., and Hayes, L. G. (1996). An in vitro model for sequential study of shiftdown of *Mycobacterium tuberculosis* through two stages of nonreplicating persistence. *Infect. Immun.* 64, 2062–2069.
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Copyright © 2016 Devasundaram, Gopalan, Das and Raja. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Comparative proteomic analysis of extracellular proteins expressed by various clonal types of *Staphylococcus aureus* and during planktonic growth and biofilm development

## OPEN ACCESS

**Edited by:**

Julio Alvarez,

University of Minnesota, USA

**Reviewed by:**

Tamas Szakmany,

Cardiff University, UK

Magdalena Chirila,

Iuliu Hatieganu University of Medicine

and Pharmacy, Romania

**\*Correspondence:**Salman S. Atshan and  
Rukman A. Hamat,Department of Medical Microbiology  
and Parasitology, Faculty of Medicine  
and Health Sciences, Universiti Putra  
Malaysia, Serdang, Selangor 43400,  
Malaysiasalmanatshan@yahoo.com;  
rukman@upm.edu.my**Specialty section:**This article was submitted to  
Infectious Diseases,  
a section of the journal  
*Frontiers in Microbiology***Received:** 14 March 2015**Accepted:** 12 May 2015**Published:** 03 June 2015**Citation:**Atshan SS, Shamsudin MN, Sekawi Z,  
Thian Lung LT, Barantalab F, Liew YK,  
Alreshidi MA, Abduljaleel SA and  
Hamat RA (2015) Comparative  
proteomic analysis of extracellular  
proteins expressed by various clonal  
types of *Staphylococcus aureus* and  
during planktonic growth and biofilm  
development. *Front. Microbiol.* 6:524.  
doi: 10.3389/fmicb.2015.00524

Salman S. Atshan<sup>1,2,3\*</sup>, Mariana N. Shamsudin<sup>1</sup>, Zamberi Sekawi<sup>1</sup>, Leslie T. Thian Lung<sup>1</sup>, Fatemeh Barantalab<sup>4</sup>, Yun K. Liew<sup>1</sup>, Mateg Ali Alreshidi<sup>5</sup>, Salwa A. Abduljaleel<sup>6</sup> and Rukman A. Hamat<sup>1\*</sup>

<sup>1</sup> Department of Medical Microbiology and Parasitology, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Serdang, Malaysia, <sup>2</sup> Department of Medical Laboratory Sciences, University College of Humanity Studies, Najaf, Iraq,

<sup>3</sup> Department of Clinical Laboratory Sciences, Faculty of Pharmacy, Basrah University, Basrah, Iraq, <sup>4</sup> Department of Immunology, Faculty of Medicine and Health Science, Universiti Putra Malaysia, Serdang, Malaysia, <sup>5</sup> Department of Basic Medical Sciences, Faculty of Medicine, Sulaiman Alraji Colleges, Albukairiyah, Saudi Arabia, <sup>6</sup> Department of Biology, Faculty of Science, Basrah University, Basrah, Iraq

*Staphylococcus aureus* is well known for its biofilm formation with rapid emergence of new clones circulating worldwide. The main objectives of the study were (1) to identify possible differences in protein expression among various and closely related clonal types of *S. aureus*, (2) to establish the differences in protein expression in terms of size of protein spots and its intensities between bacteria which are grown statically (biofilm formation) with that of under aeration and agitation, and (3) to compare the differences in protein expression as a function of time (in hours). In this study, we selected six clinical isolates comprising two similar (MRSA-527 and MRSA-524) and four different (MRSA-139, MSSA-12E, MSSA-22d, and MSSA-10E) types identified by spa typing, MLST and SCCmec typing. We performed 2D gel migration comparison. Also, two MRSA isolates (527 and 139) were selected to determine quantitative changes in the level of extracellular proteins at different biofilm growth time points of 12, 24, and 48 h. The study was done using a strategy that combines 2-DGE and LC-MS/MS analysis for absolute quantification and identification of the extracellular proteins. The 2DGE revealed that the proteomic profiles for the isolates belonging to the similar spa, MLST, and SCCmec types were still quite different. Among the extracellular proteins secreted at different time points of biofilm formation, significant changes in protein expression were observed at 48 h incubation as compared to the exponential growth at 12 h incubation. The main conclusion of the work is that the authors do observe differences among isolates, and growth conditions do influence the protein content at different time points of biofilm formation.

**Keywords:** biofilm, clone, *S. aureus*, 2DGE, LC-MS/MS

## Introduction

*Staphylococcus aureus* is one of the top nosocomial and public health pathogens and causes a wide range of infections, ranging from mild to fatal diseases (Gordon and Lowy, 2008). *S. aureus* clones are highly adaptable and have the ability to form structures known as biofilms, leading to surface colonization and the creation of a niche where the bacteria appear more resistant to host defenses and antimicrobials (Costerton et al., 1999). The process of colonization is initiated by the attachment of *S. aureus* to the surfaces, which is mediated by adhesion factors including “Microbial Surface Components Recognizing Adhesive Matrix Molecules” (MSCRAMMs) (Van Belkum et al., 2009), and secreted expanded-repertoire adhesive molecules (SERAMs) such as the extracellular adhesive protein (Eap), extracellular fibrinogen-binding protein (Efb), and extracellular matrix protein (Emp) (Clarke and Foster, 2006). The cell wall-associated virulence factors and extracellular proteins are controlled by the accessory gene regulator *agr* and the staphylococcal accessory regulator *sar* (Novick, 2000; Gotz, 2002). The *agr* locus regulates the expression of cell wall-associated proteins and secreted exoproteins in response to the density of the bacterial population (Otto, 2001). In addition to two regulatory systems, the alternative sigma factor has been identified as one of the most important virulence regulatory genes, which can regulate the expression of several exoproteins and surface proteins in response to changing environmental conditions. The *sigB* operon of *S. aureus* represents a global regulatory system and has been shown to be intimately involved in biofilm formation and enables the organism to deal with environmental stress (Lauderdale et al., 2009).

Two-dimensional gel electrophoresis (2DGE) is a famous traditional tool that is widely applied for protein separation and quantitation (Tonella et al., 2001; Gupta et al., 2002; Rosen and Ron, 2002; Scherl et al., 2005; Resch et al., 2006). This technique was initially applied on biological fluids to identify potential biomarkers and study model organisms such as *Escherichia coli* and *Bacillus subtilis*. However, to date, studies on the use of proteomic techniques to quantify the protein content of genotypically different clones at multiple time points are still lacking. This manuscript aims at comparing extracellular protein production for six different *S. aureus* isolates. To this extend, we decided to first analyse the differences by using 2D gel migration and followed by liquid chromatography tandem mass spectrometry (LC-MS/MS) to identify proteins that are differentially produced at different time points of biofilm formation. The goal of this study could be of great help in the characterization of genotypically different clones in terms of types of protein involved and could provide potential biomarkers in identifying specific MRSA or MSSA biofilm producers among identical *spa* types in a clinical environment.

**Abbreviations:** *spa*, sequence region of the protein A gene; MLST, multi locus sequence typing; SCCmec, staphylococcal cassette chromosome methicillin type gene; 2-DGE, Two-dimensional Gel Electrophoresis; LC-MS/MS, Liquid chromatography tandem mass spectrometry.

## Materials and Methods

### Bacterial Strains and Culture Conditions

In this study, six clinical *S. aureus* isolates were subjected to 2D gel sodium dodecyl sulfate-polyacrylamide gel electrophoresis (2DG-SDS-PAGE) for comparative secretomic analysis. These isolates (Table 1) were characterized into different clones through SCCmec, *spa*, and MLST typing. In addition, two representative MRSA isolates were selected (527 and 139) for proteomic analyses at different biofilm growth time points of 12, 24, and 48 h in the second experiment. The liquid chromatography tandem mass spectrometry (LC-MS/MS) was then used to identify proteins which were differentially produced at different times of biofilm formation for the only one representative isolate (MRSA-527).

The isolates are well known for their ability to form stable biofilms (Salman et al., 2012). The isolates were received in the form of stock culture from the Medical Microbiology Laboratory/Faculty of Medicine and Health Sciences /UPM, which was previously garnered from Kuala Lumpur General Hospital (HKL), Malaysia. The sources of the isolates were from different infection sites. For the first experiment, approximately  $5 \times 10^5$  CFU/ml of log-phase' cells from the six *S. aureus* isolates were inoculated in 250 ml sterile glass bottles containing 100 ml tryptic soy broth supplemented with 1% glucose (TSBG; Merck, Darmstadt, Germany Baker, UK) and incubated at 37°C for 24 h. In the second experiment, the two MRSA isolates were cultured in a tryptic soy broth supplemented with 1% glucose and were grown aerobically in 6-well tissue culture polystyrene plates (Roskilde, Denmark) and incubated statically at 37°C for 12, 24 and 48 h using the method previously described by Stepanović et al. (2007).

### Bacterial Secreted Protein Preparation

The cultured cells were centrifuged at 6000 × g for 15 min at 4°C in a refrigerated centrifuge after incubation. The supernatant was removed carefully and mixed with 10% tri-chloro acetic acid (TCA), and left overnight at 4°C to precipitate. After the overnight incubation, the precipitate was centrifuged at 10,000 × g at 4°C for 30 min. The supernatant was carefully decanted, and the protein pellets were washed several times by an ice-cold absolute acetone (Engelmann and Hecker, 2009). The resultant pellets were then air-dried for 3 min and solubilized with a rehydration buffer (Bio-Rad Laboratories,

TABLE 1 | Characteristics of isolates used in this study.

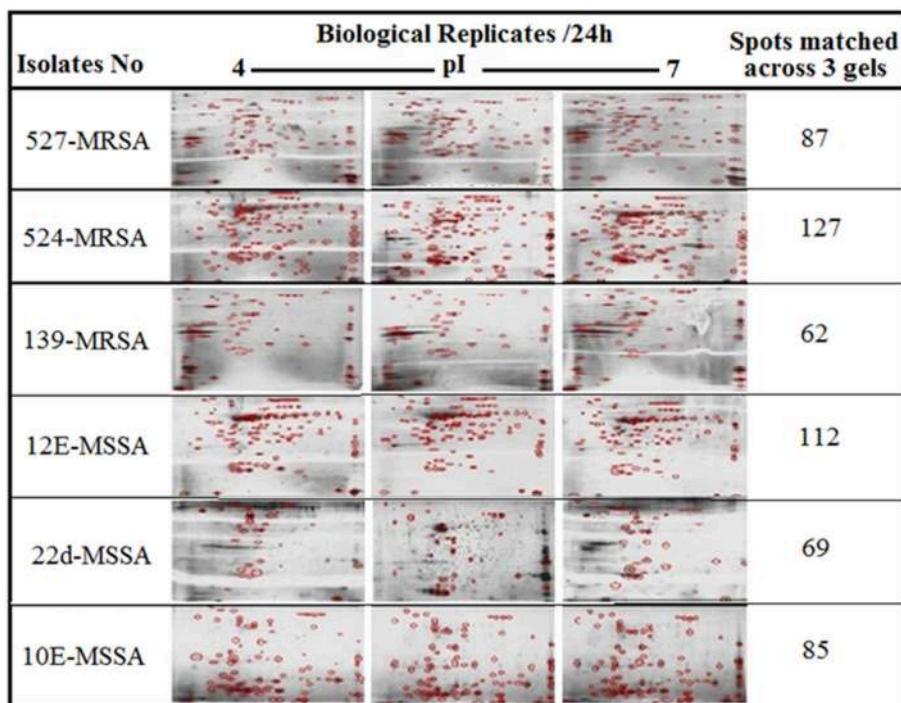
Strains no.	Spa types	MLST		SCCmec	Isolation site
		ST	CC		
1 MRSA-527	t037	ST 239	CC8	IIIA	Pus
2 MRSA-524	t037	ST 239	CC8	IIIA	Pus
3 MRSA-139	t138	ST1283	CC8	IIIA	Blood
4 MSSA-12E	t701	ST152	CC8	–	Haematoma
5 MSSA-22d	t548	ST5	CC5	–	Urine
6 MSSA-10E	t084	ST15	CC15	–	CSF

Ltd.). Since the isoelectric focusing was not successful, the protein preparations were cleaned with 2D Clean-up Kit (Bio-Rad Laboratories, Ltd.) to eliminate detergents, salts, lipids, phenolics, and nucleic acids. The proteins were solubilized again with Bio-Rad rehydration solution and RC DC Protein Assay Kit (Bio-Rad Laboratories, Ltd.) was used to quantify the concentration of proteins according to the manufacturer's instruction. Bovine serum albumin (BSA) was utilized as the protein standard. The

solubilized proteins were utilized directly or stored at  $-80^{\circ}\text{C}$  until further use.

### Two-dimension Gel Electrophoresis (2-DGE)

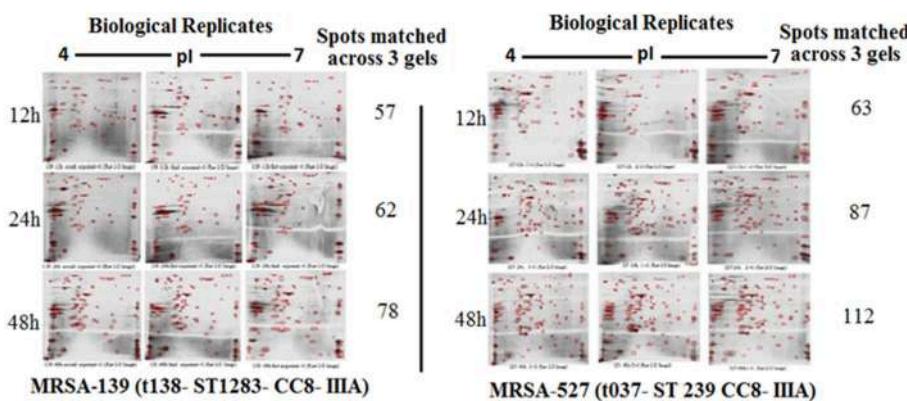
For the separation of extracellular proteins in the first dimension, 25  $\mu\text{g}$  of solubilized exoproteins were passively rehydrated in 125  $\mu\text{l}$  rehydration buffer containing 1% DTT (Bio-Red) for 14 h on a 7 cm IPG strip 4-7 (GE Healthcare Biosciences). Each strip



**FIGURE 1 | 2DE gel protein patterns of six clinical isolates of *S. aureus*.**

A total of 25  $\mu\text{g}$  protein extract of each isolate was separated on 2D gels, using IPG strips (pl 4-7) for the first dimension. Protein spots were stained with silver stain and scanned

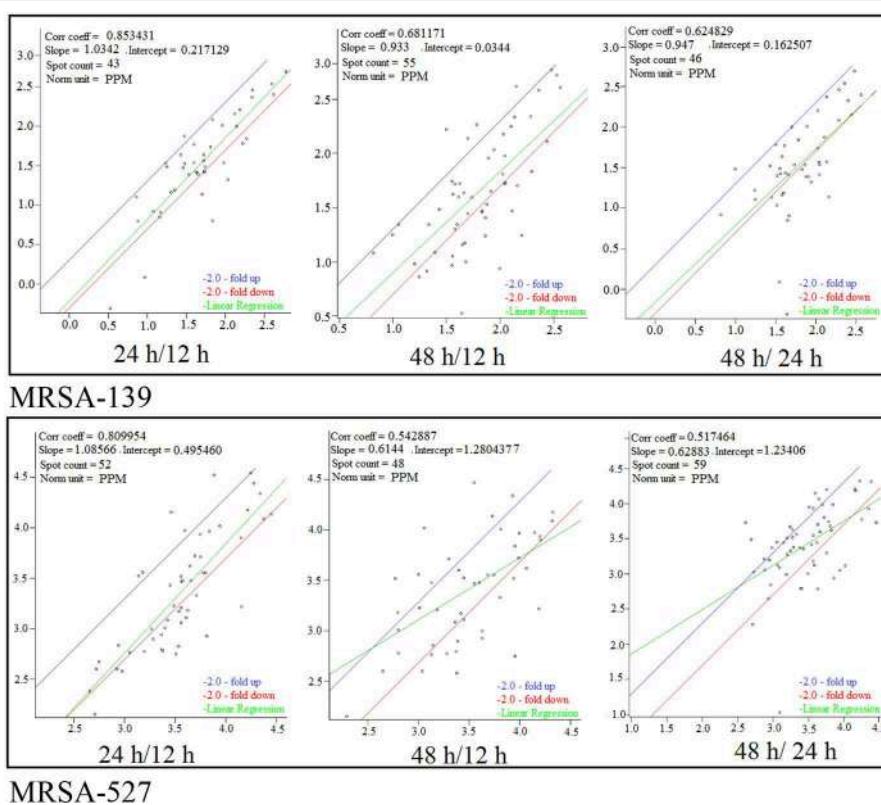
using Densitometer GS-800 Mode Imager. PDQuest software was utilized to analyze the data in which 2DE images from the internal pooled standard from all six different isolates were employed as a reference for comparative analyses.



**FIGURE 2 | Comparing 2DE gel protein patterns of *S. aureus* isolate number 527 and 139 under different growth time points.**

A total of 25  $\mu\text{g}$  protein extract of each isolate was separated on 2D

gels, using IPG strips (pl 4-7) for the first dimension. Protein spots were stained with silver stain and scanned using Densitometer GS-800 Mode Imager.



**FIGURE 3 |** Scatter plot of logarithmic spot quantities on three gels from different growth time points of *S. aureus* clinical isolate 527 and 139 under 12, 24, and 48 h. Spots were normalized based on total spot quantity.

was overlaid with 2 mL of mineral oil to prevent evaporation and urea crystallization. The IPG strips were then placed in an isoelectric focusing instrument (PROTEAN IEF cell). The IEF program was performed with the appropriate three-step protocol as shown in the Table S1. Each strip was then equilibrated in 2 ml SDS-equilibration buffer I (50 mM Tris-HCl, pH 8.8, 6 M urea, 30% glycerol, 2% SDS, and 0.002% bromophenol blue) containing 50 mg dithiothreitol (DTT) and buffer II containing 250 mg iodoacetamide (IAA) for 15 min at room temperature. These incubation procedures were performed to reintroduce SDS and provide permanent reduction. Second-dimension SDS-PAGE was performed with 1 mm thick, 12% polyacrylamide gel according to the procedures described by Ziebandt et al. (2010) (Table S2). The IPG strips were then overlaid with agarose sealing solution (1% (w/v), 0.002% (w/v) bromophenol blue in tris-glycine SDS electrophoresis buffer) (Bio-Rad) and subjected to electrophoresis with 1x tris/glycine/SDS running buffer (Bio-Rad) at 200 V until the dye front reached the bottom of the resolving gel. The gels were stained with a silver stain plus kit (Bio-Rad Laboratories, Ltd.) according to the manufacturer's instructions and scanned with a Bio-Rad GS-800 scanner.

### Protein Analysis by PDQuest Software

The scans from the three independent experiments (25 µg) were compared to determine the differences in protein

**TABLE 2 |** Comparing high expression levels of spots from *S. aureus* clinical isolate 139 and 527, under different point time growth using PDQuest analyses software.

Isolates	Analysis set name	Significantly high expression spots
MRSA-527(t037- ST 239 CC8- IIIA)	24/12 h	19
	48/12 h	21
	48/24 h	5
MRSA-139(t138- ST1283- CC8- IIIA)	24/12 h	22
	48/12 h	28
	48/24 h	8

production between genotypically different isolates and biofilm cells at different time points. The PDQuest advanced 8.0.1 2D gel analysis software with the total density in gel image normalization method associated with parts per million (PPM) was utilized as a scaling factor (Bio-Rad). Data were normalized through the local regression model as recommended by PDQuest. Student's *t*-test (95% confidence interval) was employed in the statistical analysis to determine any significant differences in spot intensity ( $p < 0.05$ ). For the experiment on biofilm development at 12, 24, and 48 h incubation, the protein changes in the 2D gels were considered when

protein intensity was highly expressed. Overexpressed proteins occurring after 48 h were excised and kept at  $-80^{\circ}\text{C}$  for further analysis.

### In-gel Digestion and Protein Identification through Liquid Chromatography-Mass Spectrometry (LC-MS)

The procedures in this section were conducted by Proteomics International Pty. Ltd. (Broadway, Nedlands, Western Australia 6009). The gel pieces were subjected to in-gel digestion with trypsin after the protein spots were excised from a single representative of MRSA-527 that was grown at 48 h of biofilm formation. The peptides were extracted according to standard techniques (Bringans et al., 2008). The peptides were analyzed through LC MS/MS with Ultimate 3000 Nano HPLC system

(Dionex) coupled with a 4000 Q TRAP mass spectrometer (Applied Biosystems). The tryptic peptides were loaded onto a 3  $\mu\text{m}$  C18 PepMap100 column (LC Packings) and separated with a linear gradient of water/acetonitrile/0.1% formic acid (v/v). The Spectra were analyzed to identify proteins of interest using Mascot sequence matching software [Matrix Science] using the Ludwig NR database. For database searching the following parameters were used: Database, Ludwig NR; taxonomy, bacteria; enzyme: trypsin; mass tolerance:  $\pm 1.2$  Da; MS/MS tolerance,  $\pm 0.05$  Da; mass value, monoisotopic; protein mass, unrestricted; and fragment mass tolerance,  $\pm 0.6$  Da. One missed cleavage and variable modifications of methionine oxidation were allowed in the analysis. Protein identification was performed based on a statistically significant MOWSE score ( $p < 0.05$ ).

**TABLE 3 | A total of 32 strongly expression spots were identified by LC-MS/MS from *S. aureus* isolate number 527 under biofilms developed growth of 48 h.**

Spot no.	Identified proteins	Accession no.	Mascot score	Mass Mw	No. of matched unique peptides
1	50S ribosomal protein L17	H0DFF7	84	13,765	2
2	30S ribosomal protein S9	D2GJI4	125	13,485	2
3	30S ribosomal protein S10	F0P6L8	84	11,553	2
4	Phosphoglycerate kinase	D0K349	219	42,603	6
5	Succinyl-CoA ligase	I0TS27	76	41,668	2
6	Ornithine aminotransferase	F0D5L4	68	37,755	2
7	Glyceraldehyde-3-phosphate dehydrogenase	B0FZE8	77	36,268	2
8	Putative uncharacterized protein	H4GCR7	105	35,764	2
9	Exotoxin 15	H4CJ65	105	25,362	2
10	50S ribosomal protein L25	I0C1W8	108	24,320	3
11	Superoxide dismutase	I0JDL1	44	22,709	1
12	Superoxide dismutase	D2GRV5	159	22,755	3
13	Putative uncharacterized protein	H1SYF7	62	21,902	2
14	Peroxiredoxin	I0TYP8	68	21,024	1
15	Alkyl hydroperoxide reductase subunit C	Q6GJR7	161	20,963	3
16	Ribosome-recycling factor	E5RBF6	130	20,328	1
17	Alkaline shock protein 23	P0AOP6	141	19,180	4
18	Putative septation protein	F0D890	79	10,854	2
19	Transmembrane sulfatase	F0DG99	92	74,385	2
20	Chaperonin	D0K6R6	104	57,567	1
21	Enolase	D2N5I4	284	47,087	3
22	N-acetyl muramoyl-L-alanine amidase	H1TS15	283	69,208	6
23	IgG-binding protein SBI	E0P4N0	56	50,013	1
24	Epidermal cell differentiation inhibitor	P24121	53	27,663	1
25	Exotoxin 15	H4A246	153	26,305	3
26	Staphylococcal exotoxin 1	Q2YVR4	101	25,998	1
27	Ornithine aminotransferase 1	F0D9Y3	142	44,240	2
28	Phosphoglycerate kinase	H3YI93	69	42,576	3
29	Ornithine carbamoyltransferase	F0DEE9	79	37,527	2
30	Alcohol dehydrogenase	H1SYV0	61	35,926	2
31	Superoxide dismutase	I0JDL1	93	22,709	2
32	protein SA211940967	H0C9Z5	66	19,301	2

Sequences similarity of identified proteins in the present study are available as an Ludwig NR and NCBI BLAST database. Individual ions scores  $>57$  indicate identity or extensive homology ( $p < 0.05$ ).

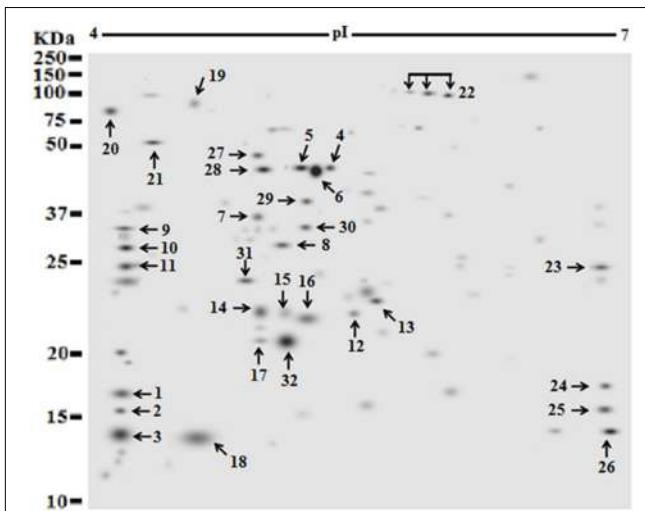
## Results

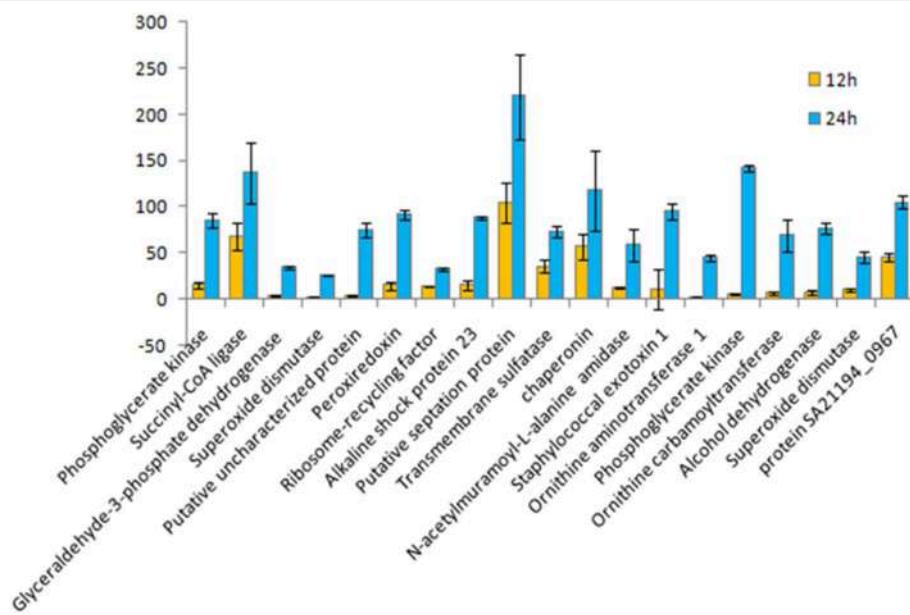
### Extracellular Protein Analysis

**Figure 1** provides a visual of the 2DE protein profiles with a high degree of exoproteome heterogeneity among the isolates belonging to different *spa* types in the first experiment. Surprisingly, isolates with similar MLST and SCCmec types secreted proteins displayed remarkable differences either on the positional shifts (location) or number or intensity of the protein spots within the gel map. For example in descending order: isolate number 524, 12-E, 527, 10-E, 22-d, and 139 secreted 127, 112, 87, 85, 69, and 62 proteins, respectively. The overall mean coefficient of variation (C.V) was 84.13. The most obvious similarity was observed among isolates 12-E, 22-d, 524, and 527 as they shared 76, 62, 100, and 39 proteins, respectively. When the 2DE images obtained from two MRSA strains (139 and 527) at 12, 24, and 48 h incubation were imported into the PDQuest software, the number of expressed secreted proteins and spot volume were again found to be different between the isolates and even between all time-points (**Figure 2**). Each spot on a particular gel has to be mapped to the corresponding spot on the other gels in a process called spot matching in order to compare the spot intensities (**Figure 3**). Since the size and intensity of the spots differed from gel to gel; thus, a spot remodeling step was applied. This is to allow the spot boundaries to fit the gray level distributions of the original gel images and also to determine the spots that may differ significantly in terms of high levels of expression (**Table 2**).

### Proteins Identified Through Mass Spectrometry

Of 112 differentially expressed proteins from MRSA- 527 at 48 h, 32 strongly expressed spots were selected for the identification through Mass Spectrometry using LC/MS as summarized in the **Table 3** and **Figure 4**. In planktonic phase at 12 h incubation, there was an initial decreased in the expression of extracellular proteins expression in terms of the number spots and its intensities, and was followed by an increase of these features at 24 and 48 h of biofilm development. The highly expressed protein spots are shown in Table S3 and Figure S1. Of these, the simultaneous increase in the expression levels of protein such as alkaline shock protein 23, phosphoglycerate kinase, ligase, aminotransferase, phosphate dehydrogenase, exotoxin 15, dismutase, sulfatase, enolase, alanine amidase, aminotransferase1, phosphoglycerate kinase, and others we found in this biofilm producer, and the maximum peak of expression occurred at 48 h compared to 24 and 12 h growth phases ( $p < 0.05$ ) as summarized in **Table 2**, **Figures 5–7** and Figure S2. The results obtained from the LC-MS/MS were generated from mass spectrometry with several protein scores made for each protein. The peptide summary option was utilized to view the results (results can be viewed at <https://sysbio-mascot.wehi.edu.au/mascot>). Protein scores are the sum of a series of peptide scores which could determine the rank of each protein hit. For scores higher than 57 are considered significant at  $p < 0.05$ , where  $p$  is the probability that the observed





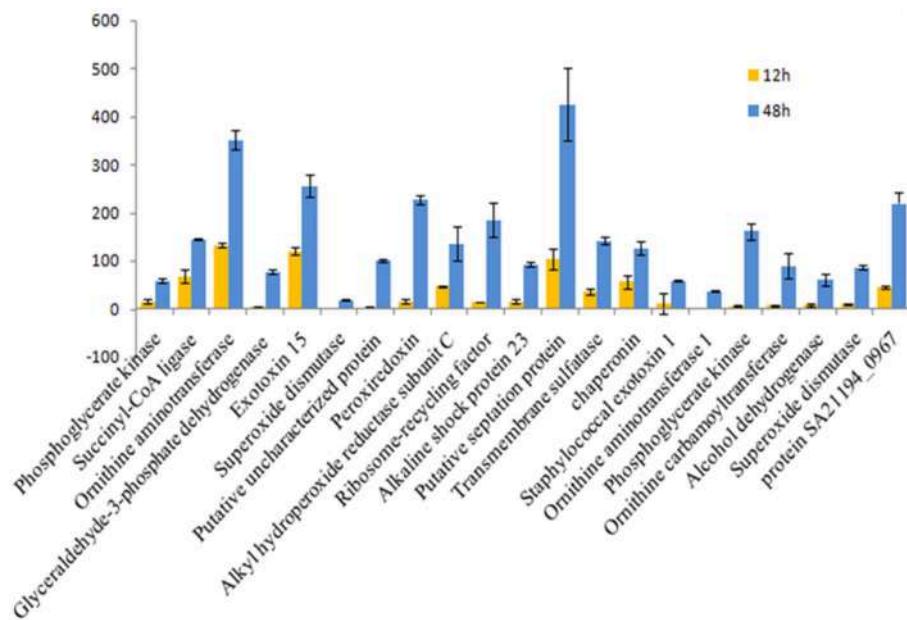
**FIGURE 5 | Fold change for 12 vs. 24 h growth time points of *S. aureus* clinical isolate (527).** A chart displaying each extracellular protein (spot) type measures the differences in protein quantity production under different growth time points of *S. aureus* biofilm developed, using a PDQuest advanced Software analysis for all spots in 2DE gel. Data represent the mean of triplicate determinations of the increase in total

spot quantity. An OD value of totalspot quantity in gel image normalization method associated with parts per million (PPM) was utilized as a scaling factor (Bio-Rad). Data were normalized through the local regression model as recommended by PDQuest Software. The significant difference in spot intensity was more than two-fold changes ( $p < 0.05$ ) between 24 and 12 h.

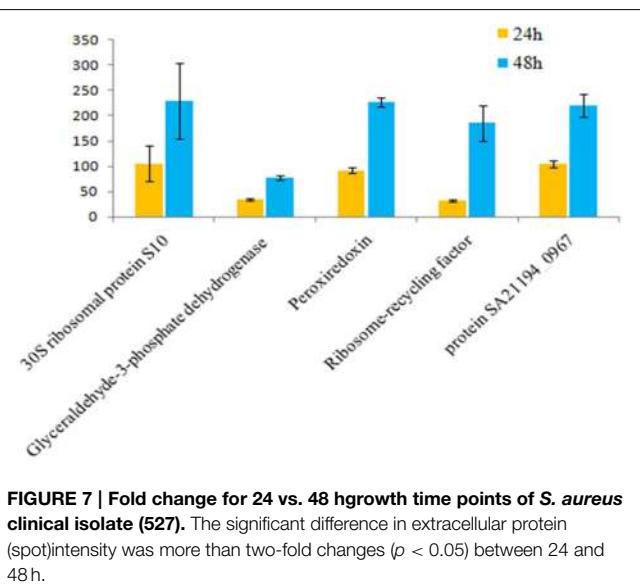
(Hecker et al., 2010). In the biofilm experiments, the rate of extracellular protein production was found to increase in term of spot intensity and the number of spots at 48 and 24 h compared to the exponential phase growth at 12 h incubation (Figure 2). Although this change was greater at 48 and 24 h, the maximum value of the other spots remained similar to that of spots achieved at 12 h. This may explain the increase in the rate of extracellular proteins production during biofilm growth and suggests that it can be utilized for regulation and as potential markers for the identification of biofilms. Therefore, we concur with the previous finding which suggested that a change in the growth status and culture medium causes a change in the relative rate of extracellular protein formation (Yang et al., 2014). This observation is clearly relevant in biofilm development; thus, further research should be conducted to identify the genes which are responsible for encoding these extracellular proteins. The main difficulty encountered through Mass Spectrometry in our study is that it was performed to analyse the type of secreted proteins. However, some of these proteins were found to be highly expressed at different time points of the biofilm growth which are not described as exoproteins, and have cytosolic localization (Table S3, Figure 4 and Figure S1). Hence, the presence of these proteins in culture supernatants may be due to death and/or lysis of biofilm cells during maturation. However, the alkaline shock protein 23 'asp23' was found to be increased with a higher expression level in older biofilms ( $p < 0.05$ ) (Figures 5, 6, Figure S2). To our knowledge, this finding is very interesting and corroborate with Goerke et al.

(2005) in which alkaline shock protein 23 activity would peak upon entry into the stationary growth phase when *S. aureus* is grown under stressful condition and it has been confirmed that the upregulation of  $\sigma^B$  took place in the older biofilms (Goerke et al., 2005). The alternative sigma factor,  $\sigma^B$ , and the accessory global regulator locus, *agr*, are the two important virulence regulatory genes, which regulate the expression of several exoproteins and surface proteins in response to changing environmental conditions. In that way, bacteria can adapt and survive in a biofilm (Knobloch et al., 2004). Our study identified the exoproteins that were strongly expressed at 48 h incubation in only one isolate, which may not, however, be the similar exoproteins produced in other isolates types. This would make sense because six different isolates in the first experiment did not show homogeneous expression of their exoproteins. Thus, we can partially conclude that there are differences in exoproteins profiles, and their protein expression at various time points of biofilm growth among the different isolates. However, this could be used as different surrogate markers for biofilm identification within the various isolates type. Hence, this type of extracellular protein production can only be applied to the MRSA isolates in this present study. It may not be suitable for other *S. aureus* isolates, even for those that exhibit similar characteristics.

Our findings could provide new insight into the genetic background of closely related *S. aureus* isolates. A slight difference in their genetic make-ups would lead to different protein profiles. It is known that bacteria could not have



**FIGURE 6 | Fold change for 12 vs. 48 h growth time points of *S. aureus* clinical isolate (527).** The significant difference in extracellular protein (spot)intensity was more than two-fold changes ( $p < 0.05$ ) between 12 and 48 h.



**FIGURE 7 | Fold change for 24 vs. 48 h growth time points of *S. aureus* clinical isolate (527).** The significant difference in extracellular protein (spot)intensity was more than two-fold changes ( $p < 0.05$ ) between 24 and 48 h.

a complete identical genetic background as interspecies or intraspecies transmission would continuously occur even within the same clonal lineage (Liew and Neela, 2012). In addition to the variability of the 2DE protein patterns of the different clinical isolates, the results also present differences in the quantities of individual proteins. Variations were observed in terms of spot intensity among the *S. aureus* isolates despite these isolates having the same MLST and CC type. Thus, variation in protein spot intensities may be related to the differential activities of the staphylococcal gene regulators, for instance in the case

of virulence factors, the activity of staphylococcal *agr* could influence those factors (Novick, 2003). The observed variations can also be attributed either to post-translational modification (PTM) or to modification occurring during the preparation of the protein samples. All these results reflect the extraordinary capacity of *S. aureus* clones to adapt and produce biofilms in various environments.

In conclusion, the present study shows that a considerable proteomic difference exists among similar and various types of *S. aureus*. A significant variation in spot size intensities was observed in the production of extracellular proteins. This variation could possibly have contributed to the degree of virulence even within the same clonal genotype and could enhance individual heterogeneity in the infection potential. This diversity suggests that the overexpression of extracellular protein production can only be applied as a diagnostic marker for a single clone of MRSA. It cannot be applied to other clones with or without similar characteristics. Thus, the development of a rapid and precise identification profile for each clone type in human infections is important on the outcome of patients with invasive infections, in order to prescribe the correct therapeutic or reduce empirical treatment. Further immunological studies, such as studies that employ two-dimensional immunoblotting, should be conducted to screen the sera of patients infected with various *S. aureus* clones in future.

## Acknowledgments

We are indebted to the Department of Medical Microbiology and Parasitology, Faculty of Medicine

and Health Sciences, Universiti Putra Malaysia for providing research facilities. This work was supported by the Ministry of Science, Technology and Innovation Malaysia (MOSTI, Putrajaya, Malaysia) Grant 5524112 to UPM.

## References

- Bringans, S., Eriksen, S., Kendrick, T., Gopalakrishnakone, P., Livk, A., and Lock, R. (2008). Proteomic analyses of the venom of *Heterometruslongimanus* (Asian black scorpion). *Proteomics* 8, 1081–1096. doi: 10.1002/pmic.200700948
- Clarke, S. R., and Foster, S. J. (2006). Surface adhesins of *Staphylococcus aureus*. *Adv. Microb. Physiol* 51, 187–224. doi: 10.1016/S0065-2911(06)51004-5
- Costerton, J. W., Stewart, P. S., and Greenberg, E. P. (1999). Bacterial biofilms: a common cause of persistent infections. *Science* 284, 1318–1322.
- Engelmann, S., and Hecker, M. A. (2009). Proteomics view of virulence factors of *Staphylococcus aureus*. *Genome Dyn.* 6, 187–197. doi: 10.1159/000235771
- Goerke, C., Fluckiger, U., Steinhuber, A., Bisanzio, V., Ulrich, M., Bischoff, M., et al. (2005). Role of *Staphylococcus aureus* global regulators sae and sigma B in virulence gene expression during device-related infection. *Infect. Immun.* 73, 3415–3421. doi: 10.1128/IAI.73.6.3415-3421.2005
- Gordon, R. J., and Lowy, F. D. (2008). Pathogenesis of methicillin-resistant *Staphylococcus aureus* infection. *Clin. Infect. Dis.* 46, S350–S359. doi: 10.1086/533591
- Gotz, F. (2002). *Staphylococcus* and biofilms. *Mol. Microbiol.* 43, 1367–1378. doi: 10.1046/j.1365-2958.2002.02827.x
- Gupta, S., Pandit, S. B., Srinivasan, N., and Chatterji, D. (2002). Proteomics analysis of carbon-starved *Mycobacterium smegmatis*: induction of Dps-like protein. *Protein Eng.* 15, 503–512. doi: 10.1093/protein/15.6.503
- Hecker, M., Becher, D., Fuchs, S.d, and Engelmann, S. A. (2010). Proteomic view of cell physiology and virulence of *Staphylococcus aureus*. *Int. J. Med. Microbiol.* 300, 76–87. doi: 10.1016/j.ijmm.2009.10.006
- Knobloch, J. K. M., Jager, S., Horstkotte, M. A., Rohde, H., and Mack, D. (2004). RsbU-dependent regulation of *Staphylococcus epidermidis* biofilm formation is mediated via the alternative sigma factor sB by repression of the negative regulator gene icaR. *Infect. Immun.* 72, 3838–3848. doi: 10.1128/IAI.72.7.3838-3848.2004
- Lauderdale, K. J., Boles, B. R., Cheung, A. L., and Horswill, A. R. (2009). Interconnections between Sigma B, agr, and proteolytic activity in *Staphylococcus aureus* biofilm maturation. *Infect. Immun.* 77, 1623–1635. doi: 10.1128/IAI.01036-08
- Liew, Y. K., and Neela, V. (2012). Comparative exoproteome profile of MRSA-ST9 isolated from pigs and pig handler. *Indian J. Microbiol.* 5, 507–509. doi: 10.1007/s12088-012-0262-2
- Novick, R. P. (2000). “Pathogenicity factors and their regulation,” in *Gram-positive Pathogens*, eds V. A. Fischetti, R. P. Novick, J. J. Ferretti, D. A. Portnoy, and J. I. Rood (Washington, DC: ASM Press), 392–407.
- Novick, R. P. (2003). Autoinduction and signal transduction in the regulation of staphylococcal virulence. *Mol. Microbiol.* 48, 1429–1449. doi: 10.1046/j.1365-2958.2003.03526.x
- Otto, M. (2001). *Staphylococcus aureus* and *Staphylococcus epidermidis* peptide pheromones produced by the accessory gene regulator agr system. *Peptides* 22, 1603–1608. doi: 10.1016/S0196-9781(01)00495-8
- Resch, A., Leicht, S., Saric, M., Pasztor, L., and Jakob, A. (2006). Comparative proteome analysis of *Staphylococcus aureus* biofilm and planktonic cells and correlation with transcriptome profiling. *Proteomics* 6, 1867–1877. doi: 10.1002/pmic.200500531
- Rosen, R., and Ron, E. Z. (2002). Proteome analysis in the study of the bacterial heat-shock response. *Mass Spectrom. Rev.* 21, 244–265. doi: 10.1002/mas.10031
- Salman, S. A., Mariana, N. S., Leslie, T. T. L., Zamperi, S., and Ehsanollah, G. (2012). Comparative characterisation of genotypically different clones of MRSA in the production of biofilms. *J. Biomed. Biotech.* 2012:417247. doi: 10.1155/2012/417247
- Scherl, A., François, P., Bento, M., Deshusses, J. M., and Charbonnier, Y. (2005). Correlation of proteomic and transcriptomic profiles of *Staphylococcus aureus* during the post-exponential phase of growth. *J. Microbiol. Methods* 60, 247–257. doi: 10.1016/j.mimet.2004.09.017
- Stepanović, S., Vukoviæ, D., and Hola, D. (2007). Quantification of biofilm in microtiter plates: overview of testing conditions and practical recommendations for assessment of biofilm production by *Staphylococci*. *APMIS* 115, 891–899. doi: 10.1111/j.1600-0463.2007.apm\_630.x
- Tonella, L., Hoogland, C., Binz, P. A., Appel, R. D., and Hochstrasser, D. F. (2001). New perspectives in the *Escherichia coli* proteome investigation. *Proteomics* 1, 409–402. doi: 10.1002/1615-9861(200103)1:3<409::AID-PROT409>3.0.CO;2-M
- Van Belkum, A. V., Damian, C., Melles, J. N., Willem, B., van Leeuwen, W. W., and Margreet, C. (2009). Coevolutionary aspects of human colonization and infection by *Staphylococcus aureus* infection. *Infect. Genet. Evol.* 9, 32–47. doi: 10.1016/j.meegid.2008.09.012
- Yang, Q., Xuezhi, D., Xuemei, L., Shuang, L., Yunjun, S., and Ziquan, Y. (2014). Differential proteomic profiling reveals regulatory proteins and novel links between primary metabolism and spinosad production in *Saccharopolysporaspinosa*. *Microb. Cell Fact.* 13:27. doi: 10.1186/1475-2859-13-27
- Ziebandt, A. K., Kusch, H., Degner, M., Jaglitz, S., and Sibbald, M. J. (2010). Proteomics uncovers extreme heterogeneity in the *Staphylococcus aureus* exoproteome due to genomic plasticity and variant gene regulation. *Proteomics* 10, 1634–1164. doi: 10.1002/pmic.200900313

## Supplementary Material

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2015.00524/abstract>

# In vitro and in vivo analysis of antimicrobial agents alone and in combination against multi-drug resistant *Acinetobacter baumannii*

Songzhe He<sup>1,2</sup>, Hui He<sup>1,2</sup>, Yi Chen<sup>1,2</sup>, Yueming Chen<sup>2</sup>, Wei Wang<sup>2</sup> and Daojun Yu<sup>1,2\*</sup>

<sup>1</sup> The Affiliated First Hospital of Hangzhou, Zhejiang Chinese Medical University, Hangzhou, China, <sup>2</sup> Department of Clinical Laboratories, Hangzhou First People's Hospital, Hangzhou, China

## OPEN ACCESS

### Edited by:

Julio Alvarez,  
University of Minnesota, USA

### Reviewed by:

Li Xu,  
Cornell University, USA  
Fiona Walsh,  
National University of Ireland  
Maynooth, Ireland

### \*Correspondence:

Daojun Yu,  
The Affiliated First Hospital of  
Hangzhou, Zhejiang Chinese Medical  
University;  
Department of Clinical Laboratory,  
Hangzhou First People's Hospital,  
Hangzhou, Zhejiang, 310006, China  
yudaojun98@163.com

### Specialty section:

This article was submitted to  
Infectious Diseases,  
a section of the journal  
Frontiers in Microbiology

Received: 03 February 2015

Accepted: 08 May 2015

Published: 27 May 2015

### Citation:

He S, He H, Chen Y, Chen Y, Wang W and Yu D (2015) In vitro and in vivo analysis of antimicrobial agents alone and in combination against multi-drug resistant *Acinetobacter baumannii*. *Front. Microbiol.* 6:507.  
doi: 10.3389/fmicb.2015.00507

**Objective:** To investigate the *in vitro* and *in vivo* antibacterial activities of tigecycline and other 13 common antimicrobial agents, alone or in combination, against multi-drug resistant *Acinetobacter baumannii*.

**Methods:** An *in vitro* susceptibility test of 101 *A. baumannii* was used to detect minimal inhibitory concentrations (MICs). A mouse lung infection model of multi-drug resistant *A. baumannii*, established by the ultrasonic atomization method, was used to define *in vivo* antimicrobial activities.

**Results:** Multi-drug resistant *A. baumannii* showed high sensitivity to tigecycline (98% inhibition), polymyxin B (78.2% inhibition), and minocycline (74.2% inhibition). However, the use of these antimicrobial agents in combination with other antimicrobial agents produced synergistic or additive effects. *In vivo* data showed that white blood cell (WBC) counts in drug combination groups C (minocycline + amikacin) and D (minocycline + rifampicin) were significantly higher than in groups A (tigecycline) and B (polymyxin B) ( $P < 0.05$ ), after administration of the drugs 24 h post-infection. Lung tissue inflammation gradually increased in the model group during the first 24 h after ultrasonic atomization infection; vasodilation, congestion with hemorrhage were observed 48 h post infection. After 3 days of anti-infective therapy in groups A, B, C, and D, lung tissue inflammation in each group gradually recovered with clear structures. The mortality rates in drug combination groups(groups C and D) were much lower than in groups A and B.

**Conclusion:** The combination of minocycline with either rifampicin or amikacin is more effective against multi-drug resistant *A. baumannii* than single-agent tigecycline or polymyxin B. In addition, the mouse lung infection by ultrasonic atomization is a suitable model for drug screening and analysis of infection mechanism.

**Keywords:** *Acinetobacter baumannii*, multi-drug resistant, ultrasonic atomization, pneumonia infection model, combination treatment

## Introduction

*Acinetobacter baumannii* is a nonfermentative, gram-negative bacillus, whose natural reservoir still remains to be determined. It can represent an opportunistic pathogen in humans, and

often causes nosocomial infections in immunocompromised patients, such as pneumonia, urinary tract infection, and sepsis (Dettori et al., 2014). Until recently, most studies on *A. baumannii* have focused on antibiotic resistance, treatment and epidemiological analysis (Ercan et al., 2014). With the large amount of clinical applications of antibiotics, the isolation rate of drug-resistant *A. baumannii* has been gradually rising, and the emergence of multi-drug resistant strains poses a big challenge for antibiotic treatment (Lee et al., 2011; Sievert et al., 2013). In recent years, the drug resistance issue has attracted worldwide attention. New therapeutic strategies against *A. baumannii* are urgently needed.

The treatment choices available for this infection are limited. Tigecycline and polymyxin B have shown some efficacy, as evidenced by both *in vitro* and *in vivo* experiments (Durante-Mangoni et al., 2013; Stein and Babinchak, 2013; Chuang et al., 2014). However, due to the lack of large-scale clinical studies, as well as the high cost of tigecycline and the potential nephrotoxic effects of polymyxin, the clinical use of these drugs has been limited. Combinations of two or more antimicrobial drugs are often used for the treatment of multi-drug resistant *A. baumannii* infections. It has been reported that meropenem, polymyxin B and minocycline have synergistic effects *in vitro* against *A. baumannii* (Zusman et al., 2013; Ning et al., 2014). In addition, the combination of sulbactam with imipenem displays synergistic bactericidal activity in the lung tissue (Dinc et al., 2013). The combination of rifampicin with imipenem, sulbactam, and colistin has the ability to potentiate the anti-infection activity of these drugs (Pachón-Ibáñez et al., 2010).

Although both *in vitro* and *in vivo* data support the efficacy of certain antibiotics against *A. baumannii*, discrepancies have been found in the results, due to the unstable or inefficient animal model (Mutlu Yilmaz et al., 2012). Hence, it is of particular importance to establish a stable model for drug screening or for investigating infection mechanisms. At present, the mouse model of *A. baumannii* infection has been shown not to be successful, as only a self-limiting bacterial pneumonia is induced, even if a high dose of bacteria is administered. To improve this model, some research groups used immunocompromised mice or mucin-treated mice to increase their sensitivity to *A. baumannii* (van Faassen et al., 2007; Pichardo et al., 2010). Lung infections in murine models have been produced by direct tracheotomy infection (Eveillard et al., 2010), micro-tracheal injection (Eveillard et al., 2010), or intranasal administration (Russello et al., 2008). However, those methods have clear disadvantages, leading to low infection rate (Qiu et al., 2009).

Given the low infection rate and instability of the current pneumonia models, we intended to establish a new *A. baumannii* infected mouse model using ultrasonic atomization. The drugs that were found to have antibacterial effect *in vitro*, tigecycline and polymyxin B were validated in this *in vivo* model. Additionally, some relatively cost-effective antibiotics were compared, including amikacin, minocycline and rifampicin, and the efficacy of the combinations including those drugs was analyzed. In conclusion, this study presented a new *in vivo*

model for future studies, and provided experimental evidence of an effective combination therapy for multi-drug resistant *A. baumannii* infection.

## Materials and Methods

### Strains

One hundred and one multi-drug resistant *A. baumannii* strains were obtained from Hangzhou First People's Hospital and were identified by the Vitek 2 Compact analyzer (BioMérieux SA, France). Multi-drug resistant strains were identified by drug susceptibility test and stored at  $-80^{\circ}\text{C}$ . *Pseudomonas aeruginosa* ATCC27853 were used as control strain.

### Experimental Animals

Five hundred specific pathogen-free BALB/c mice (half male and half female, weight 12–14 g, age 4 weeks) were bred at a temperature of  $18\text{--}25^{\circ}\text{C}$  and humidity of 40–70%. The license number was SCXK (Shanghai) 2013–0016. According to "Animal Quality Management Approach" (1997), the experimental procedures were under the approval by the Experimental Animal Center of Zhejiang Chinese Medical University.

### Experimental Drugs and Main Instruments

The drugs purchased and used in this experiments were the following: imipenem/cilastatin sodium (IMP/CS) (Merek Sharp & Dohme Corp., New Jersey, United States); piperacillin/tazobactam sodium (TZP) (Wyeth Lederle SPA, New Jersey, United States); cefoperazone/sulbactam sodium (SCF) (Pfizer, New York, United States); ceftazidime (CAZ) (Hailing Chemical Pharmaceutical Co., Ltd., Haikou, Hainan); rifampicin (RIF) (Shuangding Pharmaceutical Co., Ltd., Shenyang, Liaoning); amikacin (AMK) (Qiliu Pharmaceutical Co., Ltd., Jinan, Shandong); levofloxacin (LEV) (Yangtze River Pharmaceutical Group Ltd., Taizhou, Jiangsu); polymyxin B (PB) (Japan Pharmaceutical Industry Co., Ltd., Taipei, Taiwan); tigecycline (TIG) (Hisun Pharmaceutical Co., Ltd., Taizhou, Zhejiang); minocycline (MNO) (Wyeth Pharmaceutical Co., Ltd., Suzhou, Jiangsu); chloramphenicol (C) (Modern Pharmaceutical Co., Ltd., Shanghai); erythromycin (E) (Kelun Pharmaceutical Co., Ltd., Chengdu, Sichuan); fosfomycin sodium (FOS) (Northeast Pharmaceutical Group Shenyang No.1 Pharmaceutical Co., Ltd., Shenyang, Liaoning); methotrexate (MTX) (Hengrui Pharmaceutical Co., Ltd., Lianyungang, Jiangsu). Ultrasonic Nebulizer (402A1 type, Jiangsu Diving Medical Equipment Co., Ltd., Suzhou, Jiangsu).

### Minimum Inhibitory Concentration (MIC)

According to standard Regulations of Clinical Laboratory (Piewngam and Kiratisin, 2014), the broth dilution method was used to detect MIC. Briefly, solutions with different concentrations of antimicrobial agents were added to a sterile 96-well polystyrene plate. A concentration of 0.5 McFarland units ( $5 \times 10^8$  CFU/ml) of bacterial suspension was diluted with Lysogeny Broth (LB) (final concentration  $5 \times 10^5$  CFU/ml) and was added to each well of the plate. The plate was sealed

and incubated at 35°C for 18–24 h. *Pseudomonas aeruginosa* ATCC27853 was used as control. The concentration of the drugs that completely inhibited bacterial growth was defined as MIC. The evaluation of tigecycline was based on FDA standards and the other antibiotics were in accordance with Clinical and Laboratory Standards Institute (2014).

### Chequerboard Assay

The drug combination regimens are listed in **Table 1**. A micro-dilution method associated with checkerboard was applied in the drug combination screening on three randomly selected strains. Drug interactions were determined by the fractional inhibitory concentration index (FICI). FICI was defined as  $FICI = \text{MIC}_{A2}/\text{MIC}_{A1} + \text{MIC}_{B2}/\text{MIC}_{B1}$  and FICI index  $\leq 0.5$ ,  $0.5–1$ ,  $1–4$ ,  $>4$  were used to define synergism, addition, non-relation or antagonism, respectively (Sopirala et al., 2010).

### Time-Kill Curve Experiments

Tigecycline, polymyxin B, minocycline, rifampicin, chloramphenicol and fosfomycin sodium were used in the time-kill curve experiments. Drug concentrations of  $0.5 \times \text{MIC}$ ,  $1 \times \text{MIC}$ ,  $2 \times \text{MIC}$ , and  $4 \times \text{MIC}$  were chosen for these experiments. Briefly, tubes containing LB with antibiotics were inoculated with *A. baumannii* in a log-phase inoculum of roughly  $5 \times 10^5 \text{ CFU/ml}$ . Tubes were incubated in an ambient atmosphere at 35°C. At time 0, 2, 4, 8, and 24 h after inoculation, serial 10-fold dilutions were performed and aliquots were plated onto nutrient agar. The time-kill curve experiments were performed twice and results were analyzed by mean colony count values from the duplicate plates for each isolate (Rodriguez et al., 2010). The bactericidal activity of single antibiotics or combinations was defined as  $\geq 3 \log_{10} \text{ CFU/ml}$  decrease in the viable count compared with the initial inoculum. Synergism and antagonism were respectively defined as  $\geq 2 \log_{10} \text{ CFU/ml}$  decrease or increase in the viable count with the combination compared with the most active agent alone at different time points (Tångdén et al., 2014).

### Establishment of Pneumonia Model and Drug Treatment

After 1 week adaptation, the median lethal dose of methotrexate was detected (data not shown). The experiment included control (10 mice), model (90 mice) and treatment groups (divided into A,

**TABLE 1 |** The scheme for antibiotics combination.

Antibiotics	RIF	TZP	C	FOS	E	MNO	AMK	IMP/CS	PB
PB	+	+	+	+	+	-	-	-	-
TIG	+	-	+	+	+	+	+	-	-
MNO	+	-	+	+	+	-	+	-	+
FOS	-	-	+	-	+	-	+	+	-
C	-	-	-	-	-	-	-	+	-

PB, Polymyxin B; TIG, Tigecycline; MNO, Minocycline; FOS, Fosfomycin sodium; C, Chloramphenicol; RIF, Rifampicin; TZP, Piperacillin/tazobactam sodium; E, Erythromycin; IMP/CS, Imipenem/cilastatin sodium; AMK, Amikacin; "+" represents the combination of two drugs; "-" represents the non-combination of two drugs.

B, C, and D group, 80 mice per group). The mice in the control group were fed normally, while those in the model and treatment groups received an intraperitoneal injection of methotrexate (0.3 mg/day) for 3 consecutive days. The dose of methotrexate was calculated based on body surface area (Men and Mice medication ratio is 1:0.0026). Three days later, the mice in the model and treatment groups were given with 10% chloral hydrate (250 mg/kg) by intraperitoneal injection. The anesthetized mice were placed in a plastic container with two ports. A concentration of  $5 \times 10^8 \text{ CFU/ml}$  of multi-drug resistant *A. baumannii* was placed in a small ultrasonic nebulizer. *A. baumannii* flowed into the plastic container through an entry, and was discharged from the other end. In total, the ultrasonic frequency was 1.7 MHz  $\pm$  10%, atomization speed was 2 ml/min and the time of continuous atomization was 30 min. All operations were performed in a biological safety cabinet.

After infection with multi-drug resistant *A. baumannii*, the mice were randomly assigned to one of the following treatment groups: tigecycline (group A), polymyxin B (group B), minocycline + amikacin (group C), minocycline + rifampicin (group D). Mice within each group were also randomly assigned to four sub-groups of different treatment time (post infection 0, 4, 24, and 48 h), each sub-group was 20 mice (10 mice for recording the body symptoms and mortality, 10 mice for detection on counts of white blood cells and pathological examination of lung). Antimicrobial agents were given by intraperitoneal injection, with dosages as follows: group A (tigecycline, 10 mg/kg, q12h), group B (polymyxin B, 5 mg/kg, q6h), group C (minocycline, 7.5 mg/kg, q12h and amikacin, 7.5 mg/kg, q12h), and group D (minocycline, 7.5 mg/kg, q12h and rifampicin, 25 mg/kg, per day). These doses were chosen according to previous pharmacokinetic and pharmacodynamic data from experimental models (Song et al., 2009). The drugs in each group were administered for three consecutive days. The body symptoms and mortality of each sub-group (10 mice) were recorded at 0, 4, 24, and 48 h post infection, correspondently.

### Counts of White Blood Cells (WBC) and Pathological Examination of Lung

Cardiac blood samples were drawn from the mice infected with multi-drug resistant *A. baumannii* at 0, 4, 24, and 48 h post infection. The white blood cell counts were determined. The lungs were fixed in 40% formaldehyde, and paraffin-embedded sections were stained by hematoxylin and eosin. The lungs were subjected to pathological examination to evaluate morphology and inflammation (Hardy et al., 2009; Giladi et al., 2010). The mice in drug-treated groups (for three consecutive days) were sacrificed after therapy for 24, 48, and 72 h. WBC counts and lung pathological examination were conducted.

### Statistical Analysis

The data were presented as mean  $\pm$  standard deviation (SD). Differences between comparison groups were analyzed by Analysis of Variance (ANOVA) using SPSS19.0 software.  $P < 0.05$  was considered significant difference.

## Results

### In vitro Antibacterial Activity of Antimicrobial Agents Alone

The 101 multi-resistant strains tested in this study were completely resistant to piperacillin/tazobactam sodium, ceftazidime, levofloxacin, amikacin, fosfomycin sodium, chloramphenicol (Resistance rate: 100%), and had high resistance to imipenem/cilastatin sodium, cefoperazone/sulbactam, erythromycin (Resistance rate: >79%). By contrast, these strains had high sensitivity to rifampicin (Sensitivity: 79.2%), polymyxin B (Sensitivity: 78.2%) and minocycline (Sensitivity: 74.2%). The *in vitro* antimicrobial resistance values are listed in **Table 2**.

### In vitro Antibacterial Activity of Antimicrobial Agents in Combination

Chloramphenicol had no additive effects in combination with other antimicrobial agents, with the exception of polymyxin B. Similarly, fosfomycin had no additive effects with other antimicrobial agents, with the exception of erythromycin. However, all the remaining drug combinations showed either synergistic or additive effects (**Table 3**).

### Results of Time-Kill Curve Experiments

Chloramphenicol, polymyxin B and fosfomycin sodium at the concentrations of  $4 \times$  MIC or  $2 \times$  MIC significantly reduced the number of colonies within 4 h and completely eliminated the colonies within 8 h. However, at low concentrations ( $1 \times$  MIC,  $0.5 \times$  MIC), they had no effects on bacterial growth. The strains proliferated during the first two hours with tigecycline and rifampicin. With minocycline treatment, the number of colonies increased within 4 h; if higher drug concentrations were used, they initially decreased, but displayed regrowth 8 h after treatment (**Figure 1**).

**TABLE 2 | Thirteen normal antibiotics activity profile of multi-drug resistant *Acinetobacter baumannii*.**

Antibiotics	S/%	I/%	R/%	MIC range/(μg/ml)	MIC <sub>50</sub>	MIC <sub>90</sub>
IMP/CS	14.3	2.2	83.5	1–256	32	64
TZP	-	-	100	64–4096	512	1024
SCF	-	8.8	91.2	4–256	32	128
CAZ	-	-	100	4–512	128	512
LEV	-	-	100	4–256	8	32
AMK	-	-	100	128–8192	2048	8192
E	15.8	5.0	79.2	2–256	64	256
RIF	79.2	20.8	-	0.5–32	4	16
TIG	98.0	2.0	-	0.125–2	0.25	0.5
PB	78.2	21.8	-	0.125–4	0.5	2
FOS	-	-	100	64–4096	512	4096
MNO	74.2	14.9	10.9	1–64	4	16
C	-	-	100	32–512	256	512

PB, Polymyxin B; TIG, Tigecycline; MNO, Minocycline; FOS, Fosfomycin sodium; C, Chloramphenicol; RIF, Rifampicin; TZP, Piperacillin/tazobactam sodium; E, Erythromycin; IMP/CS, Imipenem/cilastatin sodium; AMK, Amikacin; SCF, Cefoperazone/sulbactam sodium; CAZ, Ceftazidime; LEV, Levofloxacin; S, I, R represents Susceptible, Intermediate and Resistant, respectively.

### In vivo Antibacterial Activity of Antimicrobial Agents

After ultrasonic atomization infection with *A. baumannii* for 0, 4, 24, and 48 h, WBCs in cardiac blood of model group were  $1.31 \pm 0.31$ ,  $1.84 \pm 0.20$ ,  $2.73 \pm 0.47$ , and  $4.13 \pm 1.10 (\times 10^9/L)$ , respectively. As shown by the data, the WBC counts increased significantly in a time-dependent manner. After drug treatment (consecutive 3 days), initiating from 0 h after infection by multi-drug resistant *A. baumannii*, WBCs were measured at 24, 48, and 72 h after drug treatments. WBCs in group D (MNO + RIF) were significantly higher than those in model group, group A (TIG), group B (PB) and group C (MNO + AMK) ( $P < 0.05$ ). Compared with group C, WBCs in group A and group B were not significantly affected. After drug treatment (consecutive 3 days), initiating from 4 h post infection, WBCs in group C and group D were comparable ( $P > 0.05$ ) at 24 h after drug treatment. However, compared with group A and group B, WBCs in group C and group D were significantly increased at 48 h after drug treatments. After drug treatment (consecutive 3 days) at 24 h after infection, WBCs in group C were significantly different from those in groups A, B, and D at 24 h after treatments. By contrast, WBCs in group D were comparable with the values in group A and group B at 24, 48, and 72 h after drug treatments. After drug treatment (consecutive 3 days), initiating from 48 post infection, WBCs in groups A, B, C, and D at 24, 48, and 72 h after drug treatments were significantly different from those in model group. However, WBCs in groups A, B, C, and D were comparable at 48 and 72 h after drug treatments. In addition, after treatments with single drug initiating from 24, 48 and 72 h post infection, there was no difference regarding WBCs at groups A and B (**Table 4**).

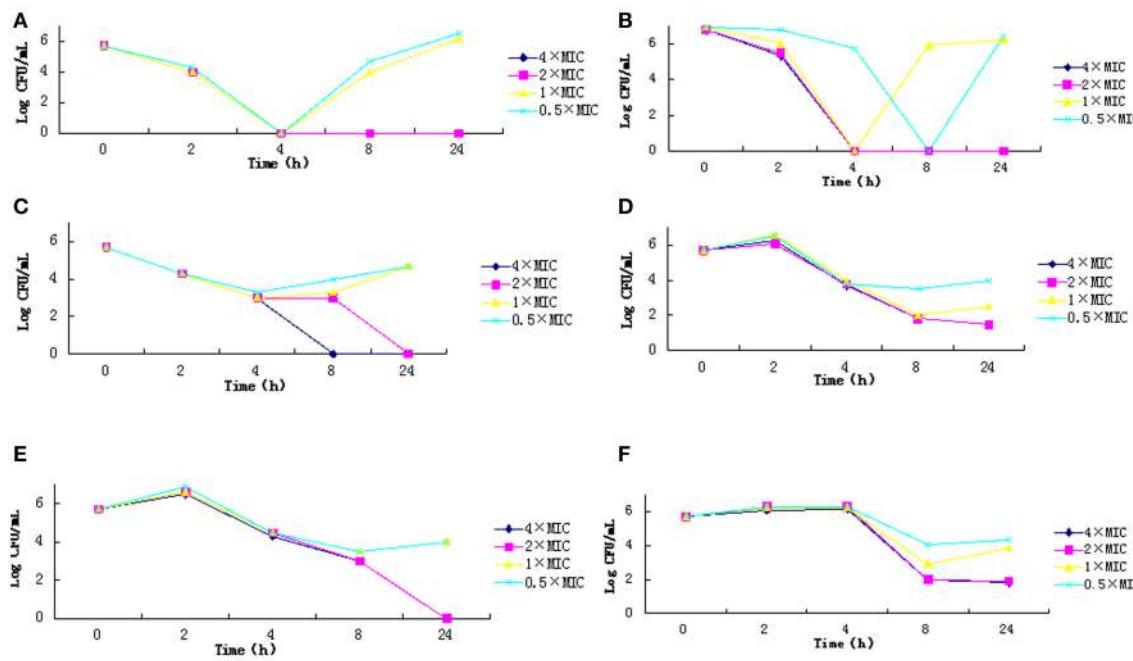
### Changes of Vital Signs and Mortality Rate

We established a new lung infection model by ultrasonic atomization. Immediately after infection, there was no mortality in any group. In the early stage after infection (0–4 h), a few mice in the model group presented with symptoms such as shortness of breath and loss of activity. However, after more prolonged infection, mice mortality rates gradually increased to 80%. Except for group D, some mice of other three groups died in the first 4 h after infection. However, at longer infection

**TABLE 3 | The fractional inhibitory concentration index (FICI) values of antibiotics combination.**

Antibiotics	RIF	TZP	C	FOS	E	MNO	AMK	IMP/CS	PB
PB	0.500	2.250	0.750	1.125	0.875	N.D.	N.D.	N.D.	N.D.
TIG	0.750	N.D.	1.125	1.125	0.750	0.625	0.875	N.D.	N.D.
MNO	0.750	N.D.	1.250	1.125	0.750	N.D.	0.750	N.D.	0.750
FOS	N.D.	N.D.	1.125	N.D.	0.875	N.D.	1.125	2.000	N.D.
C	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	2.000	N.D.	N.D.

PB, Polymyxin B; TIG, Tigecycline; MNO, Minocycline; FOS, Fosfomycin sodium; C, Chloramphenicol; RIF, Rifampicin; TZP, Piperacillin/tazobactam sodium; E, Erythromycin; IMP/CS, Imipenem/cilastatin sodium; AMK, Amikacin; FICI was defined as  $FICI = MIC_{A2}/MIC_{A1} + MIC_{B2}/MIC_{B1}$ . FICI index  $\leq 0.5$ ,  $0.5–1$ ,  $1–4$ ,  $>4$  were defined as synergistic, addition, non-relation or antagonism, respectively. N.D., not done.



**FIGURE 1 |** Time-kill curve for the 6 types of antibiotics. **(A)** C, Chloramphenicol; **(B)** PB, Polymyxin B; **(C)** FOS, Fosfomycin sodium; **(D)** TIG, Tigecycline; **(E)** RIF, Rifampicin; **(F)** MNO, Minocycline.

times (24 or 48 h post-infection), the medication groups showed increased mortality. The mortality rates in drug combination groups (groups C and D) were much lower than in groups A and B. The mortality rates in medication groups at all time points were lower than the corresponding values in the model group (Table 5). These results suggested that the drug combinations were much more effective than single drug treatment.

### Pathological Changes in the Pneumonia Model and After Drug Treatment

In the model group (Figure 2) at baseline the morphology of lung tissue was normal, with regular structure, no interstitial inflammation, slightly dilated blood vessels and no infiltration of inflammatory cells. Four hours after infection, there was a lung tissue inflammation reaction composed mostly of lymphocytes (Grade 1: the amount of inflammatory cell infiltration was less than 20%). Twenty-four hours after infection, the pulmonary infiltration of lymphocytes was greatly increased (Grade 2: the amount of inflammatory cell infiltration was between 20 and 40%). Meanwhile, a small number of neutrophils and macrophages were observed. Between 24 and 48 h after infection, a large number of inflammatory cells were found (mainly neutrophils, Grade 3: the amount of inflammatory cell infiltration was 41–80%); 48 h after infection, there were severe inflammation findings, including vascular dilation, congestion with hemorrhage, neutrophils, lymphocyte and macrophage infiltration in bronchial and alveolar (Grade 4: the amount of inflammatory cell infiltration was >80%), a significant dilation of blood vessels, congestion with hemorrhage, collapse of part of

the alveolar structure and increased visible bacterial colonies in alveolar abscesses.

Within 4 h after infection, the medication in groups A, B, C, and D caused decreased inflammatory infiltration (Grade 2). Inflammatory cell infiltration in group D was much lower than in other groups (Grade 1) (Figure 3). Within 24–48 h after infection, obvious inflammation infiltration and local hemorrhage were observed in groups A and B (Grade 3+), and obvious inflammatory reactions were observed in groups C and D (Grade 3-) without lung tissue disintegration (Figure 4). After consecutive 3 days of anti-infective therapy in groups A, B, C, and D, lung tissue inflammation in each group gradually recovered with clear structures (Grade 2) (Figure 5).

### Discussion

*A. baumannii* is becoming an important pathogen with the ability of causing nosocomial infections (Ozturk et al., 2014). With the increasingly widespread use of antimicrobial drugs, *A. baumannii* has developed resistance to several drugs, with the emergence of multi-drug resistant strains (Bassetti et al., 2011). In this study, we found that *A. baumannii* displayed low sensitivities to carbon penicillins, cephalosporins, and aminoglycosides. The MIC<sub>90</sub> of these drugs were more than 64 µg/ml. The loss of sensitivity to those drugs might be due to the wide use of antibiotics in the clinical practice. However, *A. baumannii* was still sensitive to tigecycline (MIC<sub>90</sub>: 0.5 µg/ml), polymyxin B (MIC<sub>90</sub>: 2 µg/ml), minocycline (MIC<sub>90</sub>: 16 µg/ml) and rifampicin (MIC<sub>90</sub>: 16 µg/ml). Meanwhile, the combinational experiments showed that tigecycline, polymyxin

**TABLE 4 |** Mean values of WBC counts and standard deviations for different antibiotics treatment in several infection time (0, 4, 24, 48 h).

Groups	24 h post-treatment				48 h post-treatment				72 h post-treatment			
	0 h <sup>1a</sup>	4 h <sup>2a</sup>	24 h <sup>3a</sup>	48 h <sup>4a</sup>	0 h <sup>1b</sup>	4 h <sup>2b</sup>	24 h <sup>3b</sup>	48 h <sup>4b</sup>	0 h <sup>1c</sup>	4 h <sup>2c</sup>	24 h <sup>3c</sup>	48 h <sup>4c</sup>
A	1.90 ± 0.93	1.80 ± 0.62	3.10 ± 0.54	4.90 ± 0.56	2.30 ± 0.37	2.45 ± 0.67	2.97 ± 0.40	5.43 ± 0.13	2.05 ± 0.34	2.10 ± 0.36	2.72 ± 0.45	4.89 ± 0.24
B	1.90 ± 0.35	1.60 ± 0.41	3.20 ± 0.37	5.10 ± 0.12	1.65 ± 0.22	1.98 ± 0.15	2.95 ± 0.52	4.94 ± 0.62	2.34 ± 0.51	2.05 ± 0.24	2.89 ± 0.33	4.85 ± 0.32
C	1.70 ± 0.47	3.60 ± 0.69	4.26 ± 0.89	5.30 ± 0.53	2.00 ± 0.23	3.80 ± 0.33	4.60 ± 0.21	4.70 ± 0.18	2.02 ± 0.14	3.73 ± 0.28	4.50 ± 0.30	4.64 ± 0.41
D	4.60 ± 0.52	4.30 ± 0.25	4.03 ± 0.31	5.01 ± 0.40	5.20 ± 0.34	5.00 ± 0.13	4.30 ± 0.45	4.62 ± 0.69	4.90 ± 0.31	4.30 ± 0.40	4.10 ± 0.23	4.57 ± 0.37

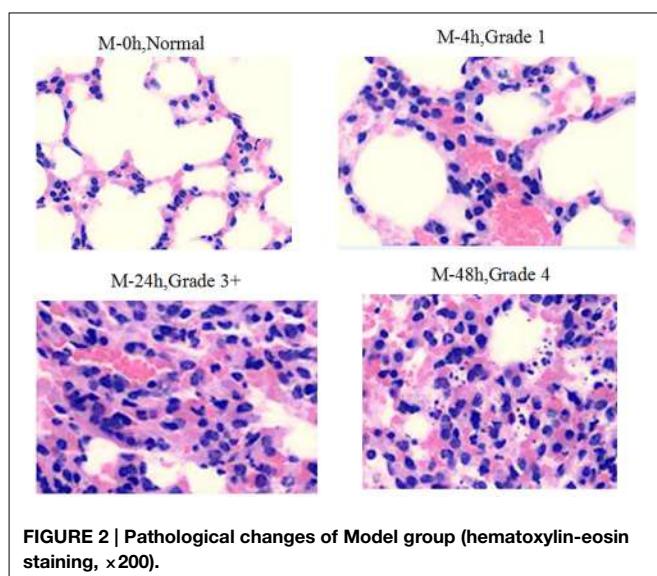
Group A, Tigecycline (TIG); Group B, Polymyxin B (PB); Group C, Minocycline +Ampicillin (MNO+AMK); Group D, Minocycline + Rifampicin (MNO+RIF).

*P*-values come from multiple comparisons for different antibiotics treatment group in the same infection time.*P*<sub>AB</sub> represents the *P*-value of multiple comparisons between Group A and Group B; *P*<sub>AC</sub> represents the *P*-value of multiple comparisons between Group A and Group C; *P*<sub>AD</sub> represents the *P*-value of multiple comparisons between Group A and Group D; *P*<sub>BC</sub> represents the *P*-value of multiple comparisons between Group B and Group C; *P*<sub>BD</sub> represents the *P*-value of multiple comparisons between Group B and Group D; *P*<sub>CD</sub> represents the *P*-value of multiple comparisons between Group C and Group D.1a: *P*<sub>AB</sub> = 0.283; *P*<sub>AC</sub> = 0.381; *P*<sub>AD</sub> = 0.018; *P*<sub>BC</sub> = 0.834; *P*<sub>BD</sub> = 0.028; *P*<sub>CD</sub> = 0.040.2a: *P*<sub>AB</sub> = 0.927; *P*<sub>AC</sub> = 0.154; *P*<sub>AD</sub> = 0.178; *P*<sub>BC</sub> = 0.009; *P*<sub>BD</sub> = 0.010; *P*<sub>CD</sub> = 0.125.3a: *P*<sub>AB</sub> = 0.727; *P*<sub>AC</sub> = 0.022; *P*<sub>AD</sub> = 0.665; *P*<sub>BC</sub> = 0.001; *P*<sub>BD</sub> = 0.392; *P*<sub>CD</sub> = 0.005.4a: *P*<sub>AB</sub> = 0.261; *P*<sub>AC</sub> = 0.004; *P*<sub>AD</sub> = 0.013; *P*<sub>BC</sub> = 0.031; *P*<sub>BD</sub> = 0.011; *P*<sub>CD</sub> = 0.024.1b: *P*<sub>AB</sub> = 0.055; *P*<sub>AC</sub> = 0.191; *P*<sub>AD</sub> = 0.000; *P*<sub>BC</sub> = 0.061; *P*<sub>BD</sub> = 0.000; *P*<sub>CD</sub> = 0.000.2b: *P*<sub>AB</sub> = 0.083; *P*<sub>AC</sub> = 0.000; *P*<sub>AD</sub> = 0.000; *P*<sub>BC</sub> = 0.000; *P*<sub>BD</sub> = 0.000; *P*<sub>CD</sub> = 0.001.3b: *P*<sub>AB</sub> = 0.946; *P*<sub>AC</sub> = 0.028; *P*<sub>AD</sub> = 0.058; *P*<sub>BC</sub> = 0.032; *P*<sub>BD</sub> = 0.065; *P*<sub>CD</sub> = 0.069.4b: *P*<sub>AB</sub> = 0.408; *P*<sub>AC</sub> = 0.245; *P*<sub>AD</sub> = 0.191; *P*<sub>BC</sub> = 0.719; *P*<sub>BD</sub> = 0.605; *P*<sub>CD</sub> = 0.873.1c: *P*<sub>AB</sub> = 0.296; *P*<sub>AC</sub> = 0.971; *P*<sub>AD</sub> = 0.000; *P*<sub>BC</sub> = 0.312; *P*<sub>BD</sub> = 0.000; *P*<sub>CD</sub> = 0.000.2c: *P*<sub>AB</sub> = 0.841; *P*<sub>AC</sub> = 0.000; *P*<sub>AD</sub> = 0.000; *P*<sub>BC</sub> = 0.000; *P*<sub>BD</sub> = 0.000; *P*<sub>CD</sub> = 0.039.3c: *P*<sub>AB</sub> = 0.800; *P*<sub>AC</sub> = 0.022; *P*<sub>AD</sub> = 0.062; *P*<sub>BC</sub> = 0.034; *P*<sub>BD</sub> = 0.097; *P*<sub>CD</sub> = 0.561.4c: *P*<sub>AB</sub> = 0.931; *P*<sub>AC</sub> = 0.635; *P*<sub>AD</sub> = 0.547; *P*<sub>BC</sub> = 0.697; *P*<sub>BD</sub> = 0.605; *P*<sub>CD</sub> = 0.897.

**TABLE 5 | Effect on mortality rates [% (n/N)] at different infected time for each group.**

Groups	0 h	4 h	24 h	48 h
Model	0.0 (0/10)	30.0 (3/10)	60.0 (6/10)	80.0 (8/10)
A	0.0 (0/10)	10.0 (1/10)	10.0 (1/10)	50.0 (5/10)
B	0.0 (0/10)	10.0 (1/10)	30.0 (3/10)	60.0 (6/10)
C	0.0 (0/10)	10.0 (1/10)	20.0 (2/10)	40.0 (4/10)
D	0.0 (0/10)	0.0 (0/10)	10.0 (1/10)	50.0 (5/10)

Group A, Tigecycline (TIG); Group B, Polymyxin B (PB); Group C, Minocycline + Amikacin (MNO+AMK); Group D, Minocycline + Rifampicin (MNO+RIF).



**FIGURE 2 | Pathological changes of Model group (hematoxylin-eosin staining,  $\times 200$ ).**

B and rifampicin displayed synergistic or additive effect when combined with other drugs. The FICI in all the combination groups in this study were between 0.5 and 2.25.

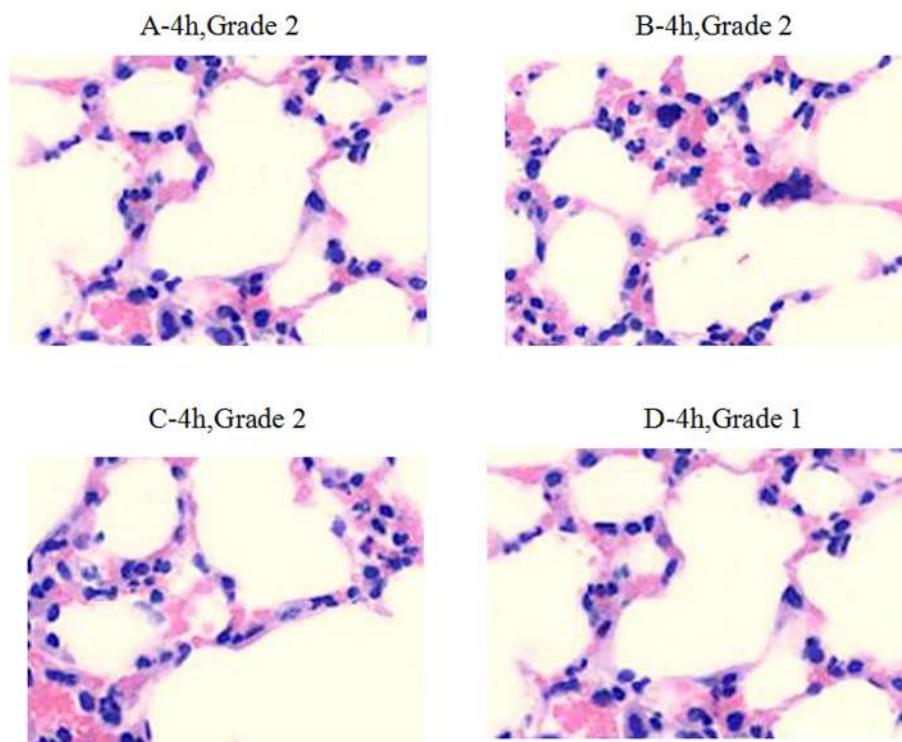
Tigecycline is a new type of glycyprostacyclin antimicrobial agent, which inhibits bacterial protein translation and produces antibacterial effects which are modulated by different factors, including resistance-nodulation-cell division (RND)-type transporters and other efflux pumps (Sun et al., 2013). Polymyxin B inhibits bacterial growth by increasing membrane permeability (Liu et al., 2014). However, because of its nephrotoxicity and neurotoxicity, this drug is preferably used in combination in the clinical practice, allowing a dose reduction to alleviate the toxic effects. Minocycline achieves its bactericidal effects through protein synthesis inhibition (Rumbo et al., 2013). Because of dose-dependent effects on the gastrointestinal tract and the vestibular system, minocycline alone might not be suitable for anti-infection of multi-drug resistant *A. baumannii*. In combination with other antimicrobial agents, minocycline has been able to achieve significant anti-infective effect (Zhang et al., 2013).

The antibacterial activity *in vitro* does not necessarily reflect the activity *in vivo*. Hence, a stable and effective animal model is required to evaluate novel therapeutic approaches and clearly identify bacterial virulence factors (McConnell et al., 2013). Mice

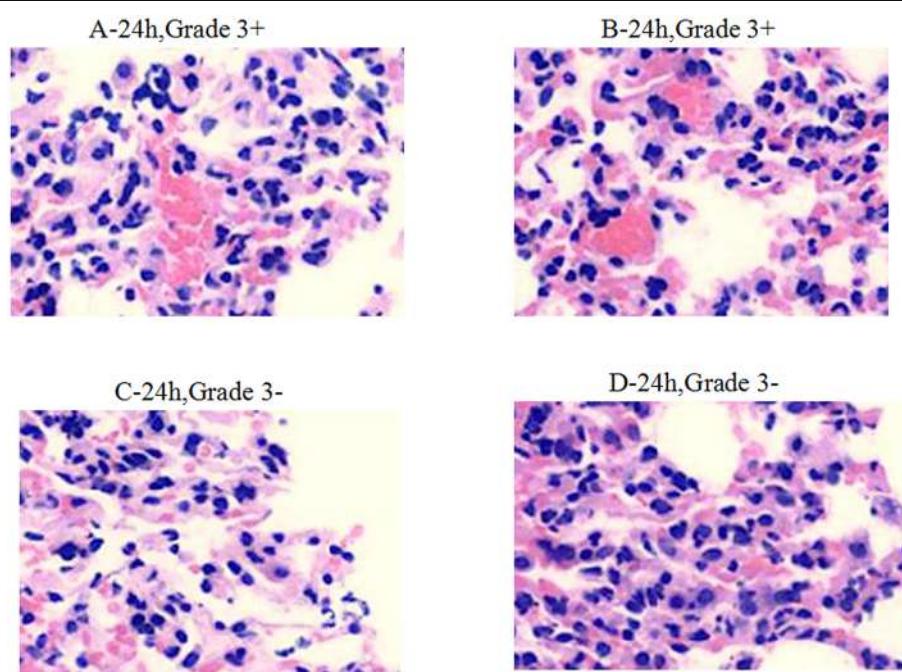
(such as C57BL/6, BALB/c, and A/J, etc.) have a predisposition to a variety of pathogens, and mice models are frequently used to study antimicrobial infection and related pathogenesis. BALB/c line (inbred) mice are susceptible to pneumonia. In this study BALB/c mice were used for an animal model of bacterial pneumonia (Chiang et al., 2013). The ultrasonic atomization method was used to establish the pneumonia model. Before the establishment of the model, methotrexate was used to reduce mouse white blood cells, resulting in immunodeficient mice. After atomization, the mortality and WBCs gradually increased with infection time. Several inflammatory cells (mainly neutrophils) were observed in the lungs 24–48 h after infection. In addition, clinical manifestations of lung inflammation were detected in model mice, including shortness of breath, weight loss, appetite loss, reduced activity. These symptoms mimicked the clinical presentation of pneumonia, which confirmed the validity of our model. This allowed us to screen novel therapeutic strategies for multi-drug resistant *A. baumannii* pulmonary infection, as well as to investigate the infection mechanisms.

After successful establishment of the mouse pneumonia model, the efficacy of drug treatment and pathological changes in lung tissue were analyzed. According to *in vitro* susceptibility test, polymyxin B and tigecycline were effective in inhibiting multi-drug resistant *A. baumannii*, and the combinations (minocycline + amikacin and minocycline + rifampicin) had synergistic antibiotic effects. We tried to confirm the *in vitro* data using our pneumonia model. In this study, within 4 h after infection, WBC counts in treatment group D was higher than in groups A, B, and C. These data indicated that minocycline and rifampicin had synergistic effect in the early stage after infection. Regarding the time-kill curve experiments, proliferation of bacteria within 4 h was found in mice treated with minocycline or rifampicin, although the number of colonies gradually decreased with the duration of treatment time. In contrast, the minocycline and rifampicin combination was largely effective in mice within 4 h after infection. The possible mechanisms might be due to the inhibition of protein synthesis by minocycline and RNA translation at early phase by rifampicin (Jamal et al., 2014). However, within 24–48 h after infection, the inhibition effect in group C became clear and WBC counts were significantly higher compared with that in group D. Amikacin inhibits bacterial growth by increasing membrane permeability at logarithmic growth, leading to release of the functional substance. On the contrary, minocycline functions through inhibiting protein synthesis. The combination of these two drugs has synergistic or additive effect (Cunha, 2006).

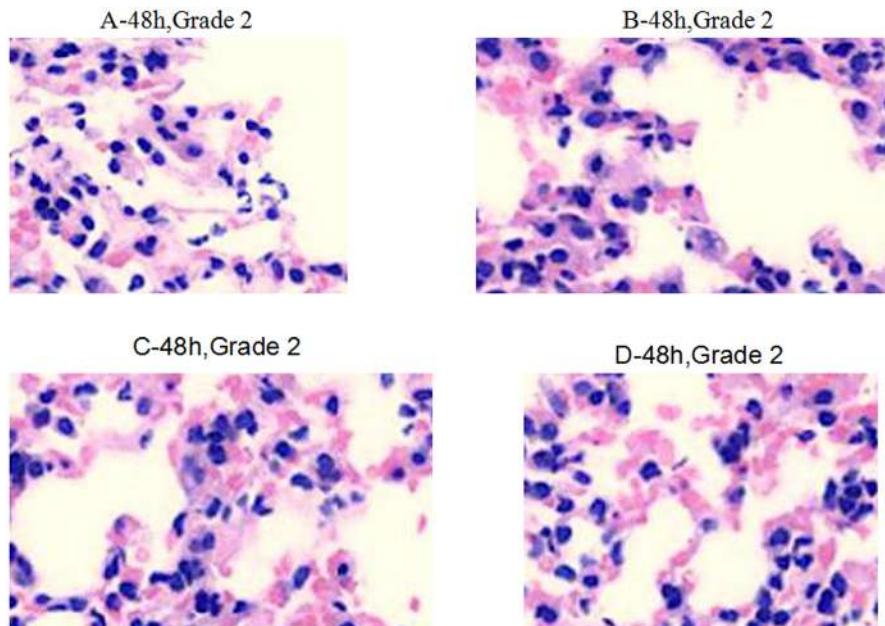
Song et al also reported that the combination of two drugs among polymyxin B, imipenem and rifampicin was effective against multi-drug resistant *A. baumannii* (Song et al., 2009). In a late phase after infection (24–48 h), the medication group had obvious effects, when compared with the model group. Meanwhile, according to the time-kill curve experimental data, six high concentrations ( $4 \times \text{MIC}$ ,  $2 \times \text{MIC}$ ) of antimicrobial agents reduced the amount of the colony count more than  $3 \log_{10}$  within 24 h, suggesting that an increase in the concentration of antibacterial drug to some extent, can achieve a good bactericidal effect *in vitro*. However, in time-kill curve



**FIGURE 3 |** Pathological changes of treatment groups at 4 h post infection (hematoxylin-eosin staining,  $\times 200$ ).



**FIGURE 4 |** Pathological changes of treatment groups at 24 h post infection (hematoxylin-eosin staining,  $\times 200$ ).



**FIGURE 5 |** Pathological changes of treatment groups at 48 h post infection (hematoxylin-eosin staining,  $\times 200$ ).

experiments, bacteria treated with the studied agents at low concentrations displayed an activity of regrowth after 24 h. The degradation of tigecycline might contribute to the loss of drug effect. For other antimicrobial agents, the regrowth after 24 h may be due to the pharmaceutical gradual failure.

The pathological changes after drug treatments also suggested that drugs in combination had synergistic or additive effects compared with the same drugs administered as single agents. Importantly, minocycline and rifampicin combination had a prominent effect in mice within 4 h after infection, while amikacin and minocycline combination had synergistic or additive effect 24 h after infection.

Minocycline is a well-characterized and safe second-line antimicrobial drug. Both *in vivo* and *in vitro* studies indicated that minocycline has antibacterial effect on multi-drug resistant *A. baumannii*. Also, thanks to its low cost, minocycline can play an important role in the treatment of multi-drug resistant *A. baumannii*. However, this drug should be avoided as single agent. The combination with other antibacterial drugs not only reduces the doses of single drugs used, but also lowers the risk of bacterial resistance (Zavascki et al., 2010). On the other side, rifampicin and amikacin are safe, effective, economical drugs, which are widely used in the clinic. Therefore, either minocycline + amikacin or minocycline + rifampicin not only can reduce the treatment cost, but can also increase the inhibition activity against multi-drug resistant *A. baumannii*.

Although multi-drug resistant *A. baumannii* was sensitive to tigecycline and polymyxin B in *in vitro* experiments, as showed in the time-kill curve experiments, the effects of these drugs *in vivo* are not certain. These data further confirmed the disagreement of *in vivo* and *in vitro* effects in antibacterial activity. The

following reasons might explain this discrepancy. First, based on pharmacokinetic analysis, the drug half-life *in vivo* affects the absorption and distribution of drugs; second, different routes of administration directly affect drug absorption, distribution, metabolism and excretion, thus modifying the concentration of drug in the body and the behavior over time, ultimately affecting drug efficacy; third, relevant uncontrollable factors affect animal experiments; fourth, *A. baumannii* can be divided into mucinous and non-mucinous types. The mucinous type can easily form colonies in the lung and recovery is difficult following treatment (Neonakis et al., 2014). Therefore, it is necessary to establish an animal model for *in vivo* screening of anti-bacterial drugs.

In conclusion, this study demonstrated that tigecycline and polymyxin B were highly sensitive to multi-drug resistant *A. baumannii* in an *in vitro* susceptibility test and, when combined with other drugs, they can produce synergistic or additive effects. *In vivo* experimental data indicated that minocycline in combination with either rifampicin or amikacin was more effective against multi-drug resistant *A. baumannii* than tigecycline or polymyxin B alone. In addition, the ultrasonic atomization lung infection model can simulate the entire processes of clinical infectious pneumonia. This model can be used to explore the mechanisms and to screen new drugs against multi-drug resistant *A. baumannii* infection.

## Acknowledgments

This work was supported, in whole or in part, by the Science Technology Department of Zhejiang Province Grant 2012C37094, Science Technology Bureau of Hangzhou Grant 20130733Q04 and Hangzhou Health Bureau Grant 2012ZD001.

## References

- Bassetti, M., Ginocchio, F., and Mikulska, M. (2011). New treatment options against gram-negative organisms. *Crit Care* 15, 215. doi: 10.1186/cc9997
- Chiang, S. R., Tang, H. J., Chen, C. H., Chen, C. C., Lee, W. Y., Chang, P. C., et al. (2013). Acid aspiration provokes the pneumonia caused by multidrug-resistant *Acinetobacter baumannii* in BALB/c mice. *Int. J. Infect. Dis.* 17, e454–e460. doi: 10.1016/j.ijid.2013.01.022
- Chuang, Y. C., Cheng, C. Y., Sheng, W. H., Sun, H. Y., Wang, J. T., Chen, Y. C., et al. (2014). Effectiveness of tigecycline-based versus colistin-based therapy for treatment of pneumonia caused by multidrug-resistant *Acinetobacter baumannii* in a critical setting: a matched cohort analysis. *BMC Infect. Dis.* 14:102. doi: 10.1186/1471-2334-14-102
- Cunha, B. A. (2006). New uses for older antibiotics: nitrofurantoin, amikacin, colistin, polymyxin B, doxycycline, and minocycline revisited. *Med. Clin. North Am.* 90, 1089–1107. doi: 10.1016/j.mcna.2006.07.006
- Dettori, M., Piana, A., Deriu, M. G., Lo Curto, P., Cossu, A., Musumeci, R., et al. (2014). Outbreak of multidrug-resistant *Acinetobacter baumannii* in an intensive care unit. *New Microbiol.* 37, 185–191.
- Dinc, G., Demiraslan, H., Elmali, F., Ahmed, S. S., Metan, G., Alp, E., et al. (2013). Efficacy of sulbactam and its combination with imipenem, colistin and tigecycline in an experimental model of carbapenem-resistant *Acinetobacter baumannii* sepsis. *Cancer Chemotherapy* 59, 325–329. doi: 10.1159/000356755
- Durante-Mangoni, E., Signoriello, G., Andini, R., Mattei, A., De Cristoforo, M., Murino, P., et al. (2013). Colistin and rifampicin compared with colistin alone for the treatment of serious infections due to extensively drug-resistant *Acinetobacter baumannii*: a multicenter, randomized clinical trial. *Clin. Infect. Dis.* 57, 349–358. doi: 10.1093/cid/cit253
- Erac, B., Yilmaz, F. F., Hosgor Limoncu, M., Ozturk, I., and Aydemir, S. (2014). Investigation of the virulence factors of multidrug-resistant *Acinetobacter baumannii* isolates. *Mikrobiyol. Bul.* 48, 70–81. doi: 10.5578/mb.6981
- Eveillard, M., Soltner, C., Kempf, M., Saint-André, J. P., Lemarié, C., Randrianarivelo, C., et al. (2010). The virulence variability of different *Acinetobacter baumannii* strains in experimental pneumonia. *J. Infect.* 60, 154–161. doi: 10.1016/j.jinf.2009.09.004
- Giladi, M., Porat, Y., Blatt, A., Shmueli, E., Wasserman, Y., Kiron, E. D., et al. (2010). Microbial growth inhibition by alternating electric fields in mice with *Pseudomonas aeruginosa* lung infection. *Antimicrob. Agents Chemother.* 54, 3212–3218. doi: 10.1128/AAC.01841-09
- Hardy, R. D., Coalson, J. J., Peters, J., Chaparro, A., Techasaensiri, C., Cantwell, A. M., et al. (2009). Analysis of pulmonary inflammation and function in the mouse and baboon after exposure to *Mycoplasma pneumoniae* CARDs toxin. *PLoS ONE* 4:e7562. doi: 10.1371/journal.pone.0007562
- Tångdén, T., Hickman, R. A., Forsberg, P., Lagerbäck, P., Giske, C. G., and Cars, O. (2014). Evaluation of double- and triple-antibiotic combinations for VIM- and NDM-producing *Klebsiella pneumoniae* by *in vitro* time-kill experiments. *Antimicrob. Agents Chemother.* 58, 1757–1762. doi: 10.1128/AAC.00741-13
- Jamal, M. A., Rosenblatt, J. S., Hachem, R. Y., Ying, J., Pravinkumar, E., Nates, J. L., et al. (2014). Prevention of biofilm colonization by Gram-negative bacteria on minocycline-rifampin-impregnated catheters sequentially coated with chlorhexidine. *Antimicrob. Agents Chemother.* 58, 1179–1182. doi: 10.1128/AAC.01959-13
- Lee, K., Yong, D., Jeong, S. H., and Chong, Y. (2011). Multidrug-resistant *Acinetobacter* spp.: increasingly problematic nosocomial pathogens. *Yonsei Med. J.* 52, 879–891. doi: 10.3349/ymj.2011.52.6.879
- Liu, Q., Li, W., Feng, Y., and Tao, C. (2014). Efficacy and safety of polymyxins for the treatment of *Acinetobacter baumannii* infection: a systematic review and meta-analysis. *PLoS ONE* 9:e98091. doi: 10.1371/journal.pone.0098091
- McConnell, M. J., Actis, L., and Pachon, J. (2013). *Acinetobacter baumannii*: human infections, factors contributing to pathogenesis and animal models. *FEMS Microbiol. Rev.* 37, 130–155. doi: 10.1111/j.1574-6976.2012.00344.x
- Mutlu Yilmaz, E., Sunbul, M., Aksoy, A., Yilmaz, H., Guney, A. K., and Guvenc, T. (2012). Efficacy of tigecycline/colistin combination in a pneumonia model caused by extensively drug-resistant *Acinetobacter baumannii*. *Int. J. Antimicrob. Agents* 40, 332–336. doi: 10.1016/j.ijantimicag.2012.06.003
- Neonakis, I. K., Spandidos, D. A., and Petinaki, E. (2014). Is minocycline a solution for multidrug-resistant *Acinetobacter baumannii*? *Future Microbiol.* 9, 299–305. doi: 10.2217/fmb.13.167
- Ning, F., Shen, Y., Chen, X., Zhao, X., Wang, C., Rong, Y., et al. (2014). A combination regimen of meropenem, cefoperazone-sulbactam and minocycline for extensive burns with pan-drug resistant *Acinetobacter baumannii* infection. *Chin. Med. J.* 127, 1177–1179. doi: 10.3760/cma.j.issn.0366-6999.20132047
- Ozturk, S., Ustun, C., Pehlivan, S., and Ucak, H. (2014). Acute generalized exanthematous pustulosis associated with tigecycline. *Ann. Dermatol.* 26, 246–249. doi: 10.5021/ad.2014.26.2.246
- Pachón-Ibáñez, M. E., Docobo-Pérez, F., López-Rojas, R., Domínguez-Herrera, J., Jiménez-Mejías, M. E., García-Curiel, A., et al. (2010). Efficacy of rifampin and its combinations with imipenem, sulbactam, and colistin in experimental models of infection caused by imipenem-resistant *Acinetobacter baumannii*. *Antimicrob. Agents Chemother.* 54, 1165–1172. doi: 10.1128/AAC.00367-09
- Pichardo, C., Pachón-Ibáñez, M. E., Docobo-Pérez, F., López-Rojas, R., Jiménez-Mejías, M. E., García-Curiel, A., et al. (2010). Efficacy of tigecycline vs. imipenem in the treatment of experimental *Acinetobacter baumannii* murine pneumonia. *Eur. J. Clin. Microbiol. Infect. Dis.* 29, 527–531. doi: 10.1007/s10096-010-0890-6
- Piewngam, P., and Kiratisin, P. (2014). Comparative assessment of antimicrobial susceptibility testing for tigecycline and colistin against *Acinetobacter baumannii* clinical isolates, including multidrug-resistant isolates. *Int. J. Antimicrob. Agents* 33, 396–401. doi: 10.1016/j.ijantimicag.2014.06.014
- Qiu, H., KuoLee, R., Harris, G., and Chen, W. (2009). High susceptibility to respiratory *Acinetobacter baumannii* infection in A/J mice is associated with a delay in early pulmonary recruitment of neutrophils. *Microbes Infect.* 11, 946–955. doi: 10.1016/j.micinf.2009.06.003
- Rodriguez, C. H., De Ambrosio, A., Bajuk, M., Spinazzi, M., Nastro, M., Bombicino, K., et al. (2010). *In vitro* antimicrobials activity against endemic *Acinetobacter baumannii* multiresistant clones. *J. Infect. Dev. Ctries.* 4, 164–167. doi: 10.3855/jidc.604
- Rumbo, C., Gato, E., López, M., Ruiz de Alegría, C., Fernández-Cuenca, F., Martínez-Martínez, L., et al. (2013). Contribution of efflux pumps, porins, and beta-lactamases to multidrug resistance in clinical isolates of *Acinetobacter baumannii*. *Antimicrob. Agents Chemother.* 57, 5247–5257. doi: 10.1128/AAC.00730-13
- Russo, T. A., Beanan, J. M., Olson, R., MacDonald, U., Luke, N. R., Gill, S. R., et al. (2008). Rat pneumonia and soft-tissue infection models for the study of *Acinetobacter baumannii* biology. *Infect. Immun.* 76, 3577–3586. doi: 10.1128/IAI.00269-08
- Sievert, D. M., Ricks, P., Edwards, J. R., Schneider, A., Patel, J., Srinivasan, A., et al. (2013). Antimicrobial-resistant pathogens associated with healthcare-associated infections: summary of data reported to the national healthcare safety network at the centers for disease control and prevention, 2009–2010. *Infect. Control Hosp. Epidemiol.* 34, 1–14. doi: 10.1086/668770
- Song, J. Y., Cheong, H. J., Lee, J., Sung, A. K., and Kim, W. J. (2009). Efficacy of monotherapy and combined antibiotic therapy for carbapenem-resistant *Acinetobacter baumannii* pneumonia in an immunosuppressed mouse model. *Int. J. Antimicrob. Agents* 33, 33–39. doi: 10.1016/j.ijantimicag.2008.07.008
- Spirala, M. M., Mangino, J. E., Gebreyes, W. A., Biller, B., Bannerman, T., Balada-Llasat, J. M., et al. (2010). Synergy testing by etest, microdilution checkerboard, and time-kill methods for pan-drug-resistant *Acinetobacter baumannii*. *Antimicrob. Agents Chemother.* 54, 4678–4683. doi: 10.1128/AAC.00497-10
- Stein, G. E., and Babinchak, T. (2013). Tigecycline: an update. *Diagn. Microbiol. Infect. Dis.* 75, 331–336. doi: 10.1016/j.diagmicrobio.2012.12.004
- Sun, Y., Cai, Y., Liu, X., Bai, N., Liang, B., and Wang, R. (2013). The emergence of clinical resistance to tigecycline. *Int. J. Antimicrob. Agents* 41, 110–116. doi: 10.1016/j.ijantimicag.2012.09.005
- van Faassen, H., KuoLee, R., Harris, G., Zhao, X., Conlan, J. W., and Chen, W. (2007). Neutrophils play an important role in host resistance to respiratory infection with *Acinetobacter baumannii* in mice. *Infect. Immun.* 75, 5597–5608. doi: 10.1128/IAI.00762-07
- Zavascki, A. P., Carvalhaes, C. G., Picao, R. C., and Gales, A. C. (2010). Multidrug-resistant *Pseudomonas aeruginosa* and *Acinetobacter baumannii*: resistance

- mechanisms and implications for therapy. *Expert Rev. Anti Infect. Ther.* 8, 71–93. doi: 10.1586/eri.09.108
- Zhang, Y., Chen, F., Sun, E., Ma, R., Qu, C., and Ma, L. (2013). Antibacterial activity of combinations of fosfomycin, minocycline and polymyxin B on pan-drug-resistant. *Exp. Ther. Med.* 5, 1737–1739. doi: 10.3892/etm.2013.1039
- Zusman, O., Avni, T., Leibovici, L., Adler, A., Friberg, L., Stergiopoulou, T., et al. (2013). Systematic review and meta-analysis of *in vitro* synergy of polymyxins and carbapenems. *Antimicrob. Agents Chemother.* 57, 5104–5111. doi: 10.1128/AAC.01230-13

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 He, He, Chen, Chen, Wang and Yu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Pharmacokinetic/Pharmacodynamic Correlation of Cefquinome Against Experimental Catheter-Associated Biofilm Infection Due to *Staphylococcus aureus*

Yu-Feng Zhou<sup>1,2</sup>, Wei Shi<sup>1,2</sup>, Yang Yu<sup>1,2</sup>, Meng-Ting Tao<sup>1,2</sup>, Yan Q. Xiong<sup>3,4</sup>, Jian Sun<sup>1,2</sup> and Ya-Hong Liu<sup>1,2\*</sup>

<sup>1</sup> National Risk Assessment Laboratory for Antimicrobial Resistance of Animal Original Bacteria, South China Agricultural University, Guangzhou, China, <sup>2</sup> Laboratory of Veterinary Pharmacology, College of Veterinary Medicine, South China Agricultural University, Guangzhou, China, <sup>3</sup> Los Angeles Biomedical Research Institute, Harbor-UCLA Medical Center, Torrance, CA, USA, <sup>4</sup> Division of Infectious Diseases, Department of Medicine, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA, USA

## OPEN ACCESS

### Edited by:

Andres M. Perez,  
University of Minnesota, USA

### Reviewed by:

A. Gnanamani,  
Central Leather Research Institute,  
India  
Margaret Ip,  
Chinese University of Hong Kong,  
Hong Kong

### \*Correspondence:

Ya-Hong Liu  
lyh@scau.edu.cn

### Specialty section:

This article was submitted to  
Infectious Diseases,  
a section of the journal  
*Frontiers in Microbiology*

Received: 13 September 2015

Accepted: 15 December 2015

Published: 07 January 2016

### Citation:

Zhou Y-F, Shi W, Yu Y, Tao M-T, Xiong YQ, Sun J and Liu Y-H (2016) Pharmacokinetic/Pharmacodynamic Correlation of Cefquinome Against Experimental Catheter-Associated Biofilm Infection Due to *Staphylococcus aureus*. *Front. Microbiol.* 6:1513.  
doi: 10.3389/fmicb.2015.01513

Biofilm formations play an important role in *Staphylococcus aureus* pathogenesis and contribute to antibiotic treatment failures in biofilm-associated infections. The aim of this study was to evaluate the pharmacokinetic/pharmacodynamic (PK/PD) profiles of cefquinome against an experimental catheter-related biofilm model due to *S. aureus*, including three clinical isolates and one non-clinical isolate. The minimal inhibitory concentration (MIC), minimal biofilm inhibitory concentration (MBIC), biofilm bactericidal concentration (BBC), minimal biofilm eradication concentration (MBEC) and biofilm prevention concentration (BPC) and *in vitro* time-kill curves of cefquinome were studied in both planktonic and biofilm cells of study *S. aureus* strains. The *in vivo* post-antibiotic effects (PAEs), PK profiles and efficacy of cefquinome were performed in the catheter-related biofilm infection model in murine. A sigmoid  $E_{max}$  model was utilized to determine the PK/PD index that best described the dose-response profiles in the model. The MICs and MBICs of cefquinome for the four *S. aureus* strains were 0.5 and 16  $\mu$ g/mL, respectively. The BBCs (32–64  $\mu$ g/mL) and MBECs (64–256  $\mu$ g/mL) of these study strains were much higher than their corresponding BPC values (1–2  $\mu$ g/mL). Cefquinome showed time-dependent killing both on planktonic and biofilm cells, but produced much shorter PAEs in biofilm infections. The best-correlated PK/PD parameters of cefquinome for planktonic and biofilm cells were the duration of time that the free drug level exceeded the MIC ( $fT > MIC$ ,  $R^2 = 96.2\%$ ) and the MBIC ( $fT > MBIC$ ,  $R^2 = 94.7\%$ ), respectively. In addition, the  $AUC_{24h}/MBIC$  of cefquinome also significantly correlated with the anti-biofilm outcome in this model ( $R^2 = 93.1\%$ ). The values of  $AUC_{24h}/MBIC$  for biofilm-static and 1-log<sub>10</sub>-unit biofilm-bactericidal activity were 22.8 and 35.6 h, respectively. These results indicate that the PK/PD profiles of cefquinome could be used as valuable guidance for effective dosing regimens treating *S. aureus* biofilm-related infections.

**Keywords:** biofilms, *Staphylococcus aureus*, PK/PD, cefquinome, catheter-associated infection

## INTRODUCTION

Biofilm-related infections are major medical problems and are usually refractory to antibiotic therapy (Costerton et al., 1999). The treatment failures in clinical cases are consistently reported by both clinician and veterinarians (Fabres-Klein et al., 2015). *Staphylococcus aureus* is a pathogen commonly associated with biofilm-related infections such as endocarditis, osteomyelitis, prosthetic joint infections, and catheter-related infections (Parra-Ruiz et al., 2012). Antibiotics that are effective against planktonic bacteria often do not prove satisfactory in eradicating biofilms, as biofilm cells are physiologically distinct from non-adherent and planktonic cells (Widmer et al., 1990). *S. aureus* cells within biofilm are significantly resistant to host defense systems as well as the antimicrobial therapy (Begun et al., 2007). The poor therapeutic outcome may be due to slow bacterial growth rate, limited penetration of the antibiotic and the presence of persister cells (e.g., small-colony variants) within the biofilm matrix (Davies et al., 1998). Therefore, there is a growing need for new approaches to optimize antibiotic regimens *in vivo* for the treatment of biofilm-related infections.

Cefquinome is a fourth generation cephalosporin which used widely in the veterinary industry with antimicrobial activity against a broad spectrum of Gram-positive and -negative bacterial species, and is regarded as highly stable to  $\beta$ -lactamases (CVMP, 1995). The PK/PD profiles of antibiotics could provide an important approach to establish more-effective treatment strategies and to predict the antimicrobial efficacies (Craig, 1998). However, most previous PK/PD studies were focused on planktonic cells and very limited results regarding biofilm infections were reported (Shan et al., 2014). The extrapolation of these results to biofilm cells was problematic for predicting an efficient doing regimen in biofilm-related infections (Blaser et al., 1995). Therefore, in the present study, we evaluated the *in vivo* PK/PD profiles of cefquinome against an experimental catheter-related biofilm infection model in murine due to three clinical *S. aureus* isolates and one non-clinical *S. aureus* isolate.

## MATERIALS AND METHODS

### Antibiotics and Bacterial Strains

Cefquinome was pharmaceutical grade and purchased from Qilu Animal Health Products Co., Ltd. (Jinan, China). Three veterinary clinical isolates from the endocarditis cases [One methicillin-susceptible *S. aureus* (MSSA; S45) and two methicillin-resistant *S. aureus* (MRSA; M4 and M21)], and one non-clinical isolate (MSSA; F27) were included in the present study. All *S. aureus* strains were identified by MALDI-TOF MS system (Axima-Assurance-Shimadzu). For MRSA isolates, the 16S rRNA and *mecA* genes were detected using a multiplex PCR assay. MRSA strain ATCC 43300 served as a control.

### In Vitro Susceptibility Testing and Biofilm Susceptibility Assay

The minimal inhibitory concentrations (MICs) of cefquinome against planktonic *S. aureus* cells were determined using standard Clinical and Laboratory Standards Institute (CLSI) microdilution method (CLSI, 2008). The minimal biofilm inhibitory concentrations (MBICs), biofilm bactericidal concentrations (BBCs), minimal biofilm eradication concentrations (MBECs), and biofilm prevention concentrations (BPCs) of cefquinome for biofilms were determined using the Calgary Biofilm Device as previously reported (Ceri et al., 1999; Moskowitz et al., 2004; Fernandez-Olmos et al., 2012; Macia et al., 2014; Details in Supplementary Material).

### In Vitro Time-Kill Curves

The *in vitro* time-kill curves were determined as previously described (Zhao et al., 2014). In brief, cefquinome was added into MH broth containing approximately  $5 \times 10^5$  CFU/mL exponentially growing *S. aureus* cells to obtain drug concentrations of 0, 0.5, 1, 2, 4, 8, and  $16 \times$  MIC, and incubated at 37°C for 24 h. Samples were removed at 0, 3, 6, 9, 12, and 24 h after incubation and then subjected to 10-fold serial dilutions. Twenty-five microliter of each dilution was then plated onto quadrants of MH agar and incubated at 37°C for 24 h for viable counts enumeration. Results were expressed as  $\log_{10}$  CFU/mL and the limit of detection was 40 CFU/mL.

### Biofilm Formation Assay

Stationary phase of *S. aureus* cells were resuspended in physiological NaCl solution to an OD<sub>650nm</sub> of 0.5 (~10<sup>8</sup> CFU/mL) and diluted 1:100 into Brain Heart Infusion (BHI) broth supplemented with 0.5% glucose (Seidl et al., 2008). The 96-well plate was inoculated with 200  $\mu$ L of this suspension and incubated for 18 h at 37°C. Subsequently, the wells were rinsed twice to remove planktonic cells. Biofilm was stained with crystal violet (0.1% in distilled water) for 1 min and washed with PBS three times. After visual observation, the adhering dye was dissolved with 75% alcohol to quantify the biomass measuring optical density at 650 nm.

### Experimental Catheter-Associated Biofilm Infection Model in Murine

The animals used for the *in vivo* experiments were 6-week-old (24–27 g) and pathogen-free female ICR mice (Guangdong Medical Lab Animal Center, Guangzhou, China). Animals were maintained in accordance with the American Association for Accreditation of Laboratory Animal Care criteria. All animal experiments were approved by Animal Research Committees of South China Agricultural University.

The catheter-associated biofilm infection in a murine model was established as previously described (Kadurugamuwa et al., 2003). Briefly, 1-cm segments of 14-gage Teflon intravenous catheter (Abbocath-T; Burns Vet Supply, Vancouver, WA, USA) were infected by study *S. aureus* strains in 3 mL of BHI broth supplemented with 0.5% glucose. After 6–8 h incubation at 37°C, the infected catheters were washed twice with PBS to remove

**TABLE 1 |** *In vitro* susceptibility testing and biofilm susceptibility assays of cefquinome vs. *Staphylococcus aureus* isolates used in this study.

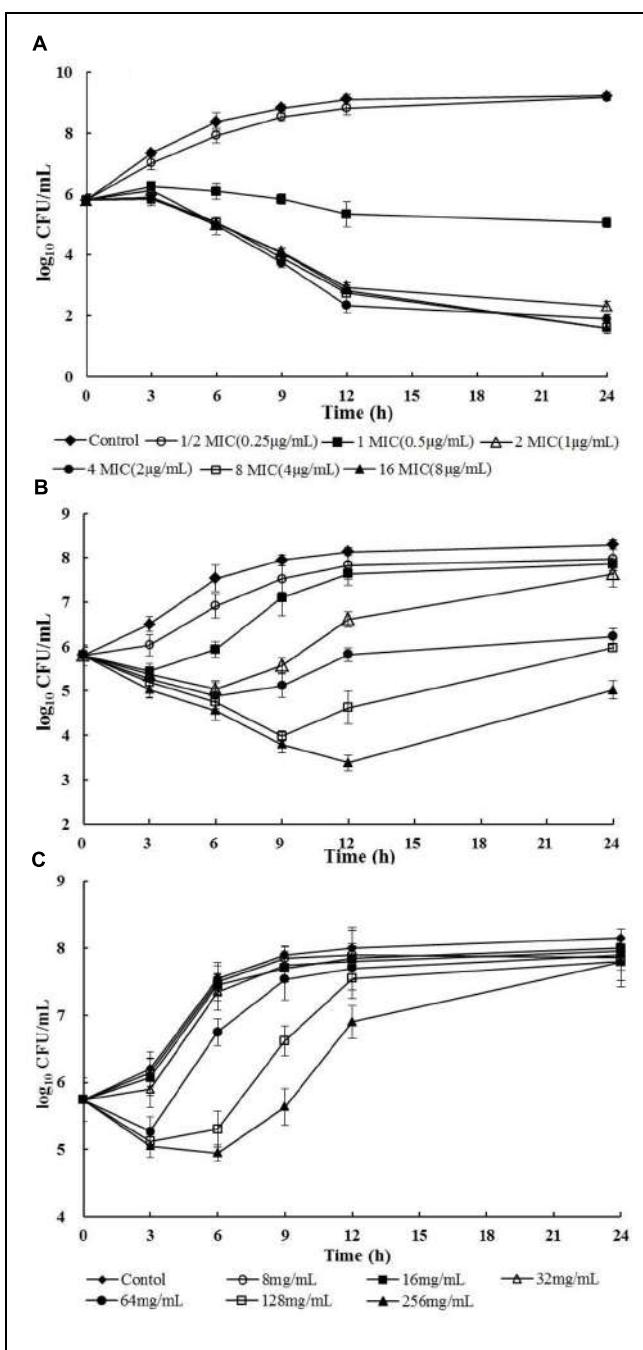
S. aureus Strains	Cefquinome ( $\mu\text{g/mL}$ )					Oxacillin	mecA
	MIC <sup>§</sup>	MBIC <sup>†</sup>	BBC	MBEC	BPC	MIC	
S45 (MSSA)	0.5	16	64	256	2	0.125	ND
F27 (MSSA)	0.5	16	32	64	1	0.125	ND
M4 (MRSA)	0.5	16	64	128	2	16	+
M21 (MRSA)	0.5	16	32	64	1	16	+

ND, not determined. <sup>§</sup>MIC, minimal inhibitory concentration of cefquinome for planktonic cells; <sup>†</sup>MBIC, minimal biofilm inhibitory concentration, defined as the lowest concentration of drug that inhibited visible biofilm growth in the recovery medium (Moskowitz et al., 2004); BBC, biofilm bactericidal concentration, defined as the lowest concentration that killed 99.9% of the biofilm cells (Fernandez-Olmos et al., 2012); MBEC, minimal biofilm eradication concentration, defined as the minimal concentration of antibiotic required to eradicate the biofilm (0 CFU/peg on plate counts; Ceri et al., 1999); BPC, biofilm prevention concentration, determined using a modification of MBIC assay in which bacterial inoculation and drug exposure occur simultaneously (Fernandez-Olmos et al., 2012).

unbound bacteria, and then implanted subcutaneously on each side of each mouse. Our preliminary data demonstrated that the infected catheter contained  $\sim 5 \times 10^5$  CFU/catheter (data not shown). At specific time points, tissue fluid (planktonic cells) was aspirated around each catheter segment and plated on MH agar plates for CFU determination. At sacrifice, catheters were removed from the subcutaneous tunnels and rinsed twice with PBS. Subsequently, the catheters were transferred to a separate tube containing 1 ml of sterile PBS. The tubes were placed in an ultrasonic bath (100 W, 40 kHz) and sonicated for 10 min, followed by vortexing for 1 min to remove biofilm cells from the catheter surface. The disaggregated biofilm was then processed to quantify the number of viable cells in the suspension.

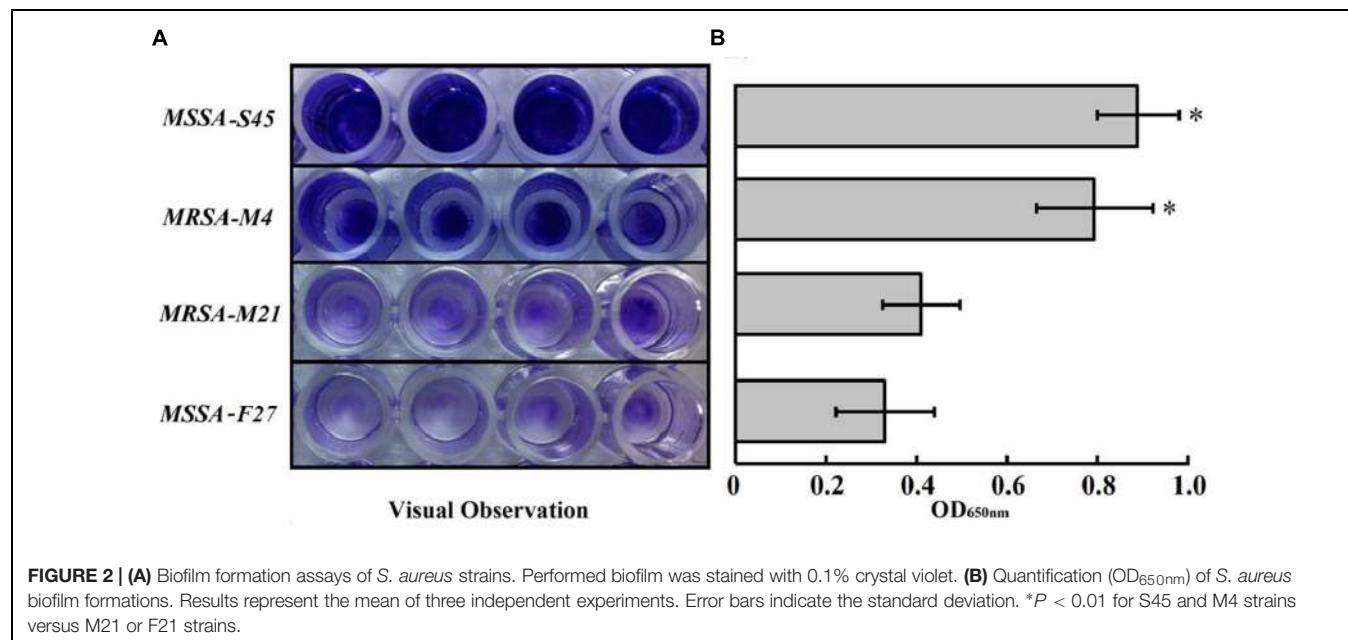
## Pharmacokinetics

The animals were administrated intramuscularly using varying doses of cefquinome (2, 8, 16, 32, 64, 128, or 256 mg/kg; six animals/group) as a single administration at 24 h after catheter implantation. Blood samples were collected by retro-orbital puncture following time points: 0, 0.08, 0.17, 0.25, 0.5, 0.75, 1, 2, 3, 4, and 6 h after antibiotic administration. Plasma was immediately isolated by centrifugation at  $3000 \times g$  for 10 min at  $4^\circ\text{C}$ , and drug concentrations in plasma were determined using a HPLC-ESI-MS/MS method as described previously (Zhou et al., 2015). The limit of quantification (LOQ) and detection were 0.01 and 0.005  $\mu\text{g/mL}$ , respectively. The time-concentration curves of cefquinome were best fitted to a one-compartment model with first-order absorption. PK parameters including half-lives of first-order absorption ( $T_{1/2\text{Ka}}$ ) and elimination ( $T_{1/2\text{Ke}}$ ), volume of distribution during the terminal phase as a function of bioavailability ( $V_d/F$ ), total area under time-concentration curve (AUC), body clearance as a function of bioavailability ( $\text{Cl}/F$ ), the peak plasma concentration ( $C_{\max}$ ) and the time of maximum concentration ( $T_{\max}$ ) were conducted using WinNonlin software (version 6.1, Pharsight, St. Louis, MO, USA). The bioavailability ( $F$ ) was calculated as  $F = (\text{AUC}_{\text{i.m.}}/\text{AUC}_{\text{i.v.}}) \times 100\%$  ( $F = 98.3\%$ , intravenous PK



**FIGURE 1 |** *In vitro* time-kill curves and *in vivo* treatment efficacies of cefquinome against planktonic and biofilm MRSA-M4 cells in a catheter-associated infection model after a single intramuscular dose of cefquinome (MIC = 0.5  $\mu\text{g/mL}$ ). **(A)** Cefquinome versus planktonic bacteria *in vitro*; **(B)** Cefquinome versus planktonic bacteria *in vivo*; **(C)** Cefquinome versus biofilm bacteria *in vivo*.

data not shown). The time courses of multiple administrations in PK/PD analysis were extrapolated from the corresponding single dose PK data obtained in the present study. The protein blinding of cefquinome in mouse was previously reported (7.4%; Wang et al., 2014). In addition, cefquinome PKs in healthy animals were also determined as reference.



**FIGURE 2 | (A)** Biofilm formation assays of *S. aureus* strains. Performed biofilm was stained with 0.1% crystal violet. **(B)** Quantification (OD<sub>650nm</sub>) of *S. aureus* biofilm formations. Results represent the mean of three independent experiments. Error bars indicate the standard deviation. \*P < 0.01 for S45 and M4 strains versus M21 or F21 strains.

**TABLE 2 | PK parameters of cefquinome after a single intramuscular administration in catheter-associated biofilm infection model of mice.**

Cefquinome (mg/kg)	T <sub>1/2KeI</sub> <sup>†</sup> (h)	T <sub>1/2Ka</sub> (h)	T <sub>max</sub> (h)	C <sub>max</sub> (μg/mL)	AUC (μg/mL × h)	V <sub>d/F</sub> (L/kg)	Cl/F (L/kg × h)
2	0.33	0.02	0.10	3.02	1.79	0.54	1.12
8	0.31	0.04	0.15	10.1	6.25	0.57	1.28
16	0.22	0.06	0.16	22.2	11.8	0.43	1.36
32	0.35	0.03	0.12	35.6	22.7	0.71	1.41
64	0.30	0.03	0.11	67.2	37.9	0.74	1.69
128	0.47	0.04	0.17	178.5	165.1	0.57	0.78
256	0.48	0.05	0.19	287.6	331.1	0.73	0.75
Mean ± SD	0.38 ± 0.14	0.04 ± 0.01	0.14 ± 0.03	—	—	0.61 ± 0.11	1.20 ± 0.32

<sup>†</sup>T<sub>1/2KeI</sub>, elimination half-life; T<sub>1/2Ka</sub>, absorption half-life; C<sub>max</sub>, peak plasma concentration; T<sub>max</sub>, time of maximum concentration; AUC, total area under the time-concentration curve; V<sub>d/F</sub>, volume of distribution during terminal phase as a function of bioavailability; Cl/F, body clearance as a function of bioavailability.

## In Vivo Efficacy of Cefquinome in Planktonic and Biofilm Bacteria and PAEs

To evaluate the *in vivo* efficacy of cefquinome in the experimental catheter-related biofilm model caused by a representative MRSA strain, M4, animals were treated with a single intramuscular dose of cefquinome (8–256 mg/kg) at 24 h after infection. The control groups received physiological NaCl. At 0, 3, 6, 9, 12, and 24 h post-dosing, animals were sacrificed and tissue fluid around the catheter was collected for planktonic bacterial culture. In addition, the catheter segments were removed aseptically, and sonicated for 10 min as described above, then quantitatively cultured. MRSA densities were expressed as mean log<sub>10</sub> CFU/mL and log<sub>10</sub> CFU/catheter ± SD for in tissue fluid and biofilm; respectively.

The post-antibiotic effect (PAE) was calculated with the equation: PAE = T - C, where T is the time for the mean growth of 1 log<sub>10</sub> CFU in planktonic or biofilm bacteria of treated mice after free drug levels in plasma fell below the MIC or MBIC, and C is the corresponding time for the untreated control mice (Spivey, 1992).

## PD Parameter Determination in Catheter-Associated Biofilm Infection Model

To evaluate the regimens of cefquinome, including dose levels as well as dosing intervals, in planktonic and biofilm

**TABLE 3 | PAE durations of cefquinome against MRSA-M4 after a single dose of administration in catheter-associated biofilm infection model of mice.**

Cefquinome (mg/kg)	T > MIC (h)	T > MBIC(h)	PAE duration (h)	
			Planktonic <sup>†</sup>	Biofilm <sup>*</sup>
8	1.9	—	0.6	—
16	3.1	0.28	0.9	-0.5
32	4.8	0.47	1.8	-0.2
64	6.0	0.75	2.3	0.6
128	8.7	3.16	3.8	0.3
256	11.5	4.38	5.3	0.7

<sup>†</sup>A positive relationship between cefquinome dosing and PAEs in planktonic cells ( $R^2 = 94.9\%$ ); \*P < 0.01 for PAEs of MRSA biofilm versus planktonic cells.

**TABLE 4 | Integration of PK/PD indices of cefquinome against planktonic and biofilm bacteria in *S. aureus* catheter-associated biofilm infection model of mice (MRSA-M4).**

Cefquinome (mg/kg)	Planktonic			Biofilm*		
	T > MIC	C <sub>max</sub> /MIC	AUC/MIC	T > MBIC	C <sub>max</sub> /MBIC	AUC/MBIC
2	1.2	6.1	3.6	–	0.19	0.11
8	1.9	20.2	12.5	–	0.63	0.39
16	3.1	44.5	23.6	0.28	1.39	0.74
32	4.8	71.1	45.4	0.47	2.22	1.42
64	6.0	134.3	75.8	0.75	4.20	2.37
128	8.7	357.1	330.3	3.16	11.2	10.3
256	11.5	575.1	662.2	4.38	17.9	20.7

\*P < 0.05 for PK/PD indices of biofilm versus planktonic cells.

bacteria, the anesthetized mice infected with catheter segments carrying  $\sim 5 \times 10^5$  CFU *S. aureus* were treated at 24 h after implantation with single or multiple intramuscular administration of cefquinome. Treatment regimens included total doses ranging from 2 to 512 mg/kg/day administered using twice-daily and once-daily. An untreated group received physiological NaCl intramuscularly. All groups of mice were sacrificed after 24 h of therapy. The catheter segments were removed and tissue fluid was collected for CFU determination as described above.

## PK/PD Modeling and Data Analysis

For PK/PD integration of cefquinome in planktonic and biofilm bacteria, the surrogate indices the duration of time that free drug level exceed the MIC or MBIC (T > MIC or MBIC), C<sub>max</sub>/MIC or MBIC and AUC from 0 to 24 h (AUC<sub>24h</sub>)/MIC or MBIC, were calculated for each animal. The numbers of bacteria in planktonic cells and biofilms were correlated to these PK/PD indices for each of the dosing regimens studied. The *in vivo* PK/PD analysis was performed using the inhibitory sigmoid dose-effect model derived from the following formula:  $E = E_0 + E_{\max} \times C_e^N / (EC_{50i}^N + C_e^N)$ , where E<sub>0</sub> is the change in log<sub>10</sub> CFU/mL of untreated controls (absence of drug), E<sub>max</sub> is the maximal antibacterial effect determined as the difference in log<sub>10</sub> CFU/mL, EC<sub>50</sub> is the value of the target PK/PD index required to achieve 50% of E<sub>max</sub>, C<sub>e</sub> is the target PK/PD indices (T > MIC or MBIC, C<sub>max</sub>/MIC or MBIC and AUC<sub>24h</sub>/MIC or MBIC), and N is the Hill coefficient that described the slope of dose effect curve (Aliabadi et al., 2003). The correlation between the efficacy and each of these PK/PD indices was calculated using the non-linear WinNonlin regression program (version 6.1, Pharsight; Gebru et al., 2009). R<sup>2</sup> was used to estimate the variance of regression with each of the PK/PD parameters.

To further compare the antibacterial efficacy of cefquinome between planktonic and biofilm cells, the sigmoid dose-response model derived from the Hill equation also was used to calculate the target values of cefquinome that produced the bacteriostatic action, 0.5-log<sub>10</sub>-unit and 1-log<sub>10</sub>-unit of the net bactericidal effect over 24 h (biofilm-static, 0.5-log<sub>10</sub>-unit and 1-log<sub>10</sub>-unit biofilm-cidal values, respectively; Zhang et al., 2014).

## RESULTS

### In Vitro Susceptibility Testing and Biofilm Susceptibility Assays

The biofilm susceptibility of the four *S. aureus* isolates is shown in Table 1. The MICs and BPCs of cefquinome for the four strains used in this study were nearly identical (0.5 and 1–2 µg/mL, respectively), indicating that cefquinome has high antibacterial activity *in vitro* against planktonic *S. aureus* cells and potential ability on preventing of early biofilm formations. The MBICs (16 µg/mL), BBCs (32–64 µg/mL) and MBECs (64–256 µg/mL) of cefquinome against the study *S. aureus* strain cells within biofilms were significantly higher than their corresponding MIC values. None of these parameters correlated with the methicillin-resistance status of these strains.

### In Vitro Time-Kill Curves

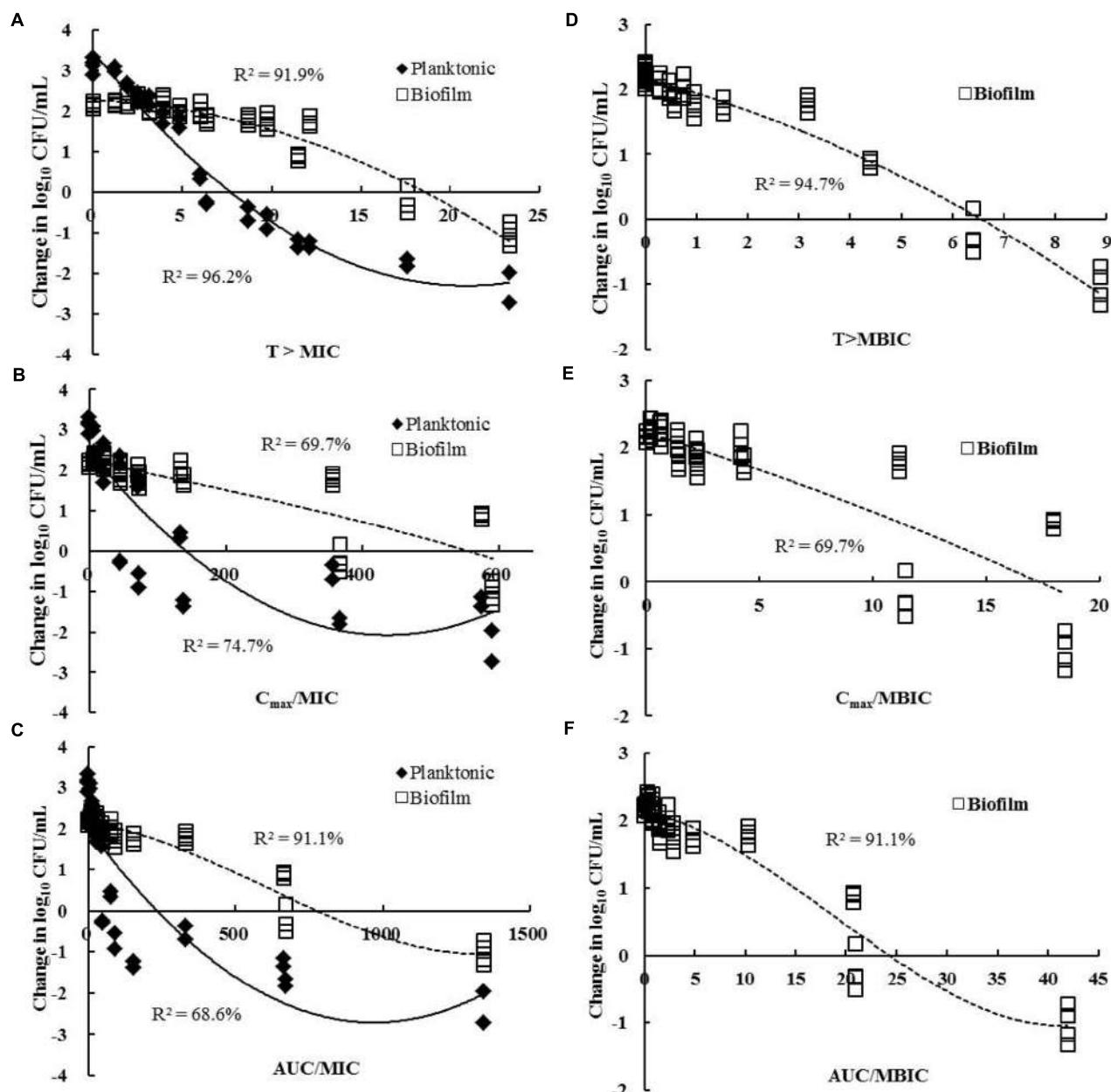
The *in vitro* time-kill curves of cefquinome against planktonic MRSA-M4 were illustrated in Figure 1A. *In vitro* killing profiles demonstrated a time- and concentration-dependent feature. Persisting inhibition of planktonic bacterial growth was observed when *S. aureus* was exposed to cefquinome at a concentration of 0.5 µg/mL. At 2 × MIC and all higher concentrations of cefquinome, either bactericidal effect or elimination of *S. aureus* (3- or 4-log<sub>10</sub>-units reduction) was observed during 12–24 h of incubation, while less than 12 h of incubation was insufficient to eliminate all bacteria.

### In Vitro Biofilm Formation Assays

All study strains possessed the ability to form biofilms. Interestingly, MRSA-M4 and MSSA-S45 strains formed significantly greater biofilms as compared to MRSA-M21 or MSSA-F21 strain ( $P < 0.01$ ; Figure 2). However, based on the OD<sub>650nm</sub> values (OD > 2 × OD of media control; 0.2), the MRSA-M21 and MSSA-F21 strains were considered as the strong biofilm producers (Jin et al., 2006). No significant difference was observed between MRSA-M4 and MSSA-S45 biofilms.

### Pharmacokinetics of Cefquinome

The PK parameters of cefquinome in the representative *S. aureus* strain M4 infected mice after a single intramuscular dose of



**FIGURE 3 | Relationships between the change in  $\log_{10}$  number of CFU for MRSA-M4 and PK/PD parameters of cefquinome.** Each symbol represents the data from each catheter segment or tissue fluid. (A)  $T > \text{MIC}$ ; (B)  $C_{\max}/\text{MIC}$ ; (C)  $AUC_{24h}/\text{MIC}$ ; (D)  $T > \text{MBIC}$ ; (E)  $C_{\max}/\text{MBIC}$ ; (F)  $AUC_{24h}/\text{MBIC}$ . The  $R^2$  value is the determination coefficient. Data points below the horizontal line represent killing and points above the horizontal line represent growth. Planktonic cells, filled diamonds; biofilms, and open squares.

2–256 mg/kg are shown in Table 2. The drug was absorbed and eliminated according to a one-compartmental model with first-order absorption (Supplementary Figure S1). A dose dependency was observed for  $C_{\max}$  and AUC values of cefquinome with ranges of from 3.02 to 287.6  $\mu\text{g}/\text{mL}$  and 1.79 to 331.1  $\mu\text{g} \times \text{h}/\text{mL}$ , respectively. The  $T_{\max}$  varied from 0.10 to 0.19 h with a mean of 0.14 h. The elimination half-life ( $T_{1/2\text{Ket}}$ ) of cefquinome ranged from 0.22 to 0.48 h. We also demonstrated that there were no significant cefquinome PK differences between healthy and

infected animals with different *S. aureus* strains (PK data in healthy animals not shown).

### Antimicrobial Efficacy of Cefquinome and PAEs In Vivo

As expected, the *in vivo* activity of cefquinome exhibited time-dependent features both on planktonic cells and on biofilm formations. However, a higher cefquinome administration was acquired to suppress the regrowth of biofilms. The dosage

**TABLE 5 | PK/PD model parameter estimates and target values of cefquinome for T > MIC or MBIC required to achieve the various antibacterial efficacies against a representative MRSA isolates M4 in planktonic or in catheter-associated biofilm infection model.**

Parameters and cell type	$E_{max}^{\dagger}$ ( $\log_{10}$ CFU/mL)	$E_0$ ( $\log_{10}$ CFU/mL)	$EC_{50}$ (h)	Static (h)	0.5- $\log_{10}$ drop (h)	1.0- $\log_{10}$ drop (h)
Planktonic (T > MIC)	-5.64	3.11	6.61	7.25	8.61	10.4
Biofilm (T > MIC)**	-4.40	2.16	17.4	17.6	19.7	22.7
Biofilm (T > MBIC)*	-3.57	2.13	5.42	6.17	7.35	8.76

\* $E_{max}$  is the maximal drug effect of cefquinome against biofilm bacteria;  $E_0$ , difference in number of biofilm bacteria (CFU/mL) in untreated group between time 0 and 24 h;  $EC_{50}$  is the T > MIC or MBIC value required to achieve 50% of the  $E_{max}$ ; \*P < 0.05 for T > MBIC in biofilm versus in planktonic cells; \*\*P < 0.01 for T > MIC in biofilm versus in planktonic cells.

regimen of cefquinome at 256 mg/kg inhibited the planktonic cells for 24 h, but regrowth was observed in the biofilm infections at the same time point (Figures 1B,C). Importantly, an approximately 3- $\log_{10}$  CFU reduction occurred for planktonic cells, while about 1.5  $\log_{10}$  CFU increased in the biofilm infection following 12 h treatment with cefquinome at 256 mg/kg (Figures 1B,C).

A positive relationship between cefquinome dosing and PAEs was seen in planktonic *S. aureus* cells ( $R^2 = 94.9\%$ ). However, less than 1 h PAE with the dose range of cefquinome (8–256 mg/kg) was demonstrated for the biofilm cells (Table 3). Therefore, significant longer PAE of planktonic *S. aureus* cells was observed as compared to bacteria cells within biofilms ( $P < 0.01$ ; Table 3).

## PK/PD Parameters Integration and Determination

Integration of PK/PD indices of cefquinome with planktonic and biofilm bacteria are showed in Table 4. The PK/PD parameters of planktonic cells were significantly greater as compared with the PK/PD profiles of biofilm when single doses of cefquinome at 2–256 mg/kg were administrated ( $P < 0.05$ ; Table 4). For instance, T > MICs were ranged from 1.2 to 11.5 h, but the T > MBICs were from 0.28 to 4.38 h. In addition,  $C_{max}$ /MICs and  $C_{max}$ /MBICs were from 6.1 to 575.1 and 0.19 to 17.9; respectively. More importantly, the AUC/MICs were significantly larger than AUC/MBICs (3.6–662 h vs. 0.11–20.7 h).

The relationship of the change of CFU and the PK/PD parameters of cefquinome between planktonic and biofilm cells was shown in Figure 3. T > MIC was the PK/PD index that best correlated with antimicrobial efficacy for planktonic cells ( $R^2 = 96.2\%$ ; Figure 3A). However, for biofilm infections, the T > MBIC index ( $R^2 = 94.7\%$ ) showed a better correlation than the T > MIC ( $R^2 = 91.9\%$ ; Figures 3A,D). Interestingly, the AUC/MIC parameter exhibited a strong correlation with the *in vivo* efficacy of cefquinome for biofilm cells ( $R^2 = 91.1\%$ ), but a poor correlation for planktonic bacteria ( $R^2 = 68.6\%$ ; Figure 3C).

## PK/PD Model Parameter Estimates for the Target Efficacy Against Study Isolates

The PK/PD indices and the corresponding target values of cefquinome required to achieve various efficacies against *S. aureus* M4 in planktonic cells and biofilm infections are listed in Table 5. The  $EC_{50}$  value of cefquinome was 6.61 h for planktonic cells versus 17.4 h for a biofilm infection ( $P < 0.01$ ).

The 1- $\log_{10}$ -unit reduction effect of cefquinome required value of T > MIC at least 10.4 h for planktonic cells, but 22.7 h for biofilm formations (T > MBIC; 8.76 h; Table 5). The *in vivo* dose-effect relationship of cefquinome for three additional strains of *S. aureus* in both planktonic and biofilm cells were also calculated using the inhibitory sigmoid  $E_{max}$  model and a similar result was observed (Supplementary Table S1).

In order to predict an effective dose, we determined the dose-response relationships between PK/PD index ( $AUC_{24h}/MBIC$ ) and the *in vivo* efficacy of cefquinome. The dose response curve strongly correlated with the  $AUC_{24h}/MBIC$  index for all study *S. aureus* biofilm infections ( $R^2 = 93.1\%$ ; Supplementary Figure S2). The target values of cefquinome necessary to produce a biofilm-static action and a 1- $\log_{10}$ -unit biofilm-cidal action as well as the corresponding  $AUC_{24h}/MBIC$  index values were listed in Table 6. The mean values of  $AUC_{24h}/MBIC$  associated with stasis and 1- $\log_{10}$ -unit reduction were 22.8 and 35.6 h, respectively.

## DISCUSSION

In this study, an experimental catheter-associated biofilm murine infection model was developed for evaluation of the PK/PD profiles of cefquinome against *S. aureus*, including MSSA and MRSA strains, growing as planktonic cells and within biofilms. To the best of our knowledge, the present study represents the

**TABLE 6 | In vivo pharmacokinetics and pharmacodynamics model of cefquinome against a *S. aureus* catheter-associated biofilm infection using  $AUC_{24h}/MBIC$  as the predictive PK-PD index (all study *S. aureus* strains included).**

Biofilm PK-PD Parameters (units)	Mean values	SD
$E_{max}$ ( $\log_{10}$ CFU/mL) <sup>†</sup>	-3.31	0.17
$E_0$ ( $\log_{10}$ CFU/mL)	1.95	0.09
$EC_{50}$ (h)	20.3	0.15
Slope (N)	3.92	0.79
$AUC_{24h}/MBIC$ for biofilm-static action (h)	22.8	1.46
$AUC_{24h}/MBIC$ for 1- $\log_{10}$ -unit biofilm-cidal action (h)	35.6	5.74

\* $E_{max}$  is the maximal drug effect of cefquinome against biofilm bacteria;  $E_0$ , difference in number of biofilm bacteria (CFU/mL) in untreated group between time 0 and 24 h;  $EC_{50}$  is the  $AUC_{24h}/MBIC$  value required to achieve 50% of the  $E_{max}$ ; N, slope of  $AUC_{24h}/MBIC$ -response curve;  $AUC_{24h}/MBIC$  for biofilm-static action, values which produced  $E = 0$  (no change in bacterial count after 24 h treatment).

first time PK/PD evaluation of cefquinome *in vivo* in *S. aureus* biofilm-related infections.

Several key insights emerged from this study. First, we investigated *in vivo* PAEs of cefquinome in a catheter-related biofilm infection model due to *S. aureus*, including MSSA and MRSA strains, and observed that the PAEs in biofilms showed significantly shorter than those in planktonic cells (Table 3). It is well accepted that the importance of long PAEs for optimizing treatment regimens in clinical practice (Spivey, 1992). Thus, longer PAE is positively correlated with greater *S. aureus* counts reduction in the catheter-related biofilm model which is consistent with previous reports (den Hollander et al., 1998; Ahmad et al., 2015). For instance, a recent study developed in a neutropenic mouse thigh model demonstrated that exposure of planktonic *S. aureus* ATCC 29213 cells to cefquinome led to a visibly longer PAE of 2.9 h as compared to below 1 h PAE observed *in vitro* (Wang et al., 2014). Similarly, another study also proved that the longer *in vitro* PAE could achieve a better *in vivo* treatment efficacy in a rabbit meningitis model (Tauber et al., 1984).

In addition, we demonstrated that cefquinome was fairly effective, causing  $\sim$ 2- to 3-log<sub>10</sub> CFU reductions for planktonic cells but only  $\sim$ 1.0- to 1.5-log<sub>10</sub> CFU reductions for biofilm cells within the catheter. This observation proved that higher concentrations and longer treatment times of cefquinome are requested to kill *S. aureus* cells within biofilm as compared with planktonic cells in the catheter-associated biofilm infections in murine duo to *S. aureus* strains. Unfortunately, the treatment strategies aimed at eradicating biofilm cells was not achievable in the present model, even at the highest dosing regimen (256 mg/kg; q12h). The lack of the antimicrobial efficacy for biofilm-related infections may be related to intrinsic tolerance of biofilm-grown bacterial populations to antibiotics, as well as the poor penetration of cefquinome through extracellular polymeric matrix of biofilm formations and inadequate antibiotic-exposure for bacterial cells embedded in biofilms (Macia et al., 2014). Other mechanisms such as the contributions of matrix components and chromosomal  $\beta$ -lactamase increases may also lead to this resistance (Hengzhuang et al., 2013).

It is notable that we were able to take advantage of the PK/PD parameters to optimize the *in vivo* efficacy against planktonic cells, as well as biofilm infections in the catheter-related biofilm model. The dose-response profiles of cefquinome for planktonic and biofilm cells showed the best correlation with T > MIC and T > MBIC, respectively ( $R^2$ , 96.2% for planktonic cells and 94.7% for biofilms; Figures 3A,D). More recently, Wang et al. (2014) investigated the pharmacodynamics of cefquinome for planktonic MRSA infections and also noted that the same index (T > MIC) correlated the best with treatment efficacy in a murine thigh infection model. In our study, although the T > MBIC index exhibited the best correlation, the AUC/MIC of cefquinome was significantly correlated with biofilm killing ( $R^2$  = 91.1%) as compared to planktonic cell infections ( $R^2$  = 68.6%; Figure 3C). This was almost equivalent with the correlation of the T > MIC for biofilms ( $R^2$  = 91.9%). Interestingly, the AUC/MBIC index which is based on the antibiotic susceptibility assay of biofilms showed a significant correlation with the *in vivo* efficacy

for all study *S. aureus* biofilm cells infections ( $R^2$  = 93.1%). Additionally, considering the higher dosage and longer the treatment periods required to treat biofilm infections *in vivo*, our results suggest that the AUC/MBIC would be the recommended predictive PK/PD index in computation of dosing regimens for *S. aureus* biofilm-related infections.

The development of device-related biofilm infections is correlated with the density of adherent cells on the implant surfaces (Joo and Otto, 2012). Once the bacterial density increases and the colony changes to biofilm formation, the eradication is virtually impossible *in vivo* (Fernandez-Olmos et al., 2012). Therefore, the important treatment strategy for biofilm-producing bacteria is to prevent the progression of early biofilm formation instead of finding a final dosing regimen for eradication of the mature biofilms. In this study, the BPCs of cefquinome for all study *S. aureus* isolates were only slightly higher than their MICs, which is an interesting parameter that could be used with the aim of reducing the cell density to prevent biofilm formation (Macia et al., 2014). More importantly, the present PK/PD model on *S. aureus* biofilms provided an AUC<sub>24h</sub>/MBIC ratio of 35.6 h that could theoretically predict 1-log<sub>10</sub> CFU biofilm-cidal effect. These results in conjunction with BPC<sub>90</sub> data of *S. aureus* biofilms may provide an additional approach to the design of dosing regimens that prevent the early stages of biofilms caused by planktonic *S. aureus* cells during antibiotic treatment. Additionally, a recent study also showed that daptomycin and rifampin alone and in combination were successful in preventing *S. aureus* biofilm infections at the early stages in a subcutaneous rat pouch model (Cirioni et al., 2010).

For the treatment of biofilm infections, more frequent administrations are needed to obtain a longer treatment period in form of T > MBIC. However, it is usually impractical to administer drug more frequently in the veterinary clinical trials. Routinely, once-daily or twice-daily schedule is considered a good compliance target. Furthermore, cefquinome (and the majority of  $\beta$ -lactams in general) have short elimination half-lives and the limitations are obvious with this routine dosing strategy (Turnidge, 1998). Thus, new formulations of cefquinome with prolonged half-life profiles should be developed to achieve a better therapeutic outcome in biofilm-related infections.

Our investigations have several limitations. For example, only four representative *S. aureus* strains were evaluated in this study. Thus, the results need to be verified in a larger population of strains. In addition, the combination regimens were not tested against *S. aureus* biofilm infections model. Studies to define the efficacy of antibiotic combinations in the same catheter-related biofilm model are ongoing in our laboratory. Nonetheless, this is the first study to our best knowledge to demonstrate PK/PD relationships of cefquinome against MSSA and MRSA growing as planktonic and biofilm cells in an *in vivo* experimental catheter-associated biofilm infection model. In the present study, cefquinome showed time-dependent activities against *S. aureus* biofilms *in vitro* as well as *in vivo*. In addition, significantly shorter PAEs were observed on biofilms vs. planktonic cells. The PK/PD index of cefquinome that best correlated with anti-biofilm efficacy was the T > MBIC in this study. More importantly, the AUC/MBIC of cefquinome also significantly

correlated with therapeutic outcomes for biofilm-related infections. These results could potentially provide a new perspective for establishing appropriate strategies of antibiotic treatment in *S. aureus* biofilm related formation.

## AUTHOR CONTRIBUTIONS

Y-HL conceived of this study and participated in its design and coordination. Y-FZ designed the experiment and drafted the manuscript. Y-FZ, WS, and YY carried out the *in vivo* animal experiments and *in vitro* time-kill curve studies. M-TT carried out the *in vivo* experiment about the additional non-clinical *S. aureus* strain in the revision of manuscript. YX and JS participated in the data analysis and revision of manuscript. All authors read and approved the final manuscript.

## REFERENCES

- Ahmad, I., Hao, H., Huang, L., Sanders, P., Wang, X., Chen, D., et al. (2015). Integration of PK/PD for dose optimization of Cefquinome against *Staphylococcus aureus* causing septicemia in cattle. *Front. Microbiol.* 6:588. doi: 10.3389/fmicb.2015.00588
- Aliabadi, F. S., Ali, B. H., Landoni, M. F., and Lees, P. (2003). Pharmacokinetics and PK-PD modelling of danofloxacin in camel serum and tissue cage fluids. *Vet. J.* 165, 104–118. doi: 10.1016/S1090-0233(02)0258-7
- Begun, J., Gaiani, J. M., Rohde, H., Mack, D., Calderwood, S. B., Ausubel, F. M., et al. (2007). Staphylococcal biofilm exopolysaccharide protects against *Caenorhabditis elegans* immune defenses. *PLoS Pathog.* 3:e57. doi: 10.1371/journal.ppat.0030057
- Blaser, J., Vergeres, P., Widmer, A. F., and Zimmerli, W. (1995). In vivo verification of *in vitro* model of antibiotic treatment of device-related infection. *Antimicrob. Agents Chemother.* 39, 1134–1139. doi: 10.1128/AAC.39.5.1134
- Ceri, H., Olson, M. E., Stremick, C., Read, R. R., Morck, D., and Buret, A. (1999). The calgary biofilm device: new technology for rapid determination of antibiotic susceptibilities of bacterial biofilms. *J. Clin. Microbiol.* 37, 1771–1776.
- Cirioni, O., Moccagno, F., Ghiselli, R., Silvestri, C., Gabrielli, E., Marchionni, E., et al. (2010). Daptomycin and rifampin alone and in combination prevent vascular graft biofilm formation and emergence of antibiotic resistance in a subcutaneous rat pouch model of staphylococcal infection. *Eur. J. Vasc. Endovasc. Surg.* 40, 817–822. doi: 10.1016/j.ejvs.2010.08.009
- CLSI (2008). *Performance Standards for Antimicrobial Disk and Dilution Susceptibility Tests for Bacteria Isolated From Animals: Approved Standard*, 3rd Edition. Wayne, PA: CLSI.
- Costerton, J. W., Stewart, P. S., and Greenberg, E. P. (1999). Bacterial biofilms: a common cause of persistent infections. *Science* 284, 1318–1322. doi: 10.1126/science.284.5418.1318
- Craig, W. A. (1998). Pharmacokinetic/pharmacodynamic parameters: rationale for antibacterial dosing of mice and men. *Clin. Infect. Dis.* 26, 1–10. doi: 10.1086/516284
- CVMP (1995). *Cefquinome Summary Report EMEA/MRL/005/95*. London: The European Agency for the Evaluation of Medicinal Products.
- Davies, D. G., Parsek, M. R., Pearson, J. P., Iglesias, B. H., Costerton, J. W., and Greenberg, E. P. (1998). The involvement of cell-to-cell signals in the development of a bacterial biofilm. *Science* 280, 295–298. doi: 10.1126/science.280.5361.295
- den Hollander, J. G., Fuursted, K., Verbrugh, H. A., and Mouton, J. W. (1998). Duration and clinical relevance of postantibiotic effect in relation to the dosing interval. *Antimicrob. Agents Chemother.* 42, 749–754.
- Fabres-Klein, M. H., Caizer Santos, M. J., Contelli Klein, R., Nunes De Souza, G., De Oliveira Barros, and Ribon, A. (2015). An association between milk and slime increases biofilm production by bovine *Staphylococcus aureus*. *BMC Vet. Res.* 11:3. doi: 10.1186/s12917-015-0319-7
- Fernandez-Olmos, A., Garcia-Castillo, M., Maiz, L., Lamas, A., Baquero, F., and Canton, R. (2012). In vitro prevention of *Pseudomonas aeruginosa* early biofilm formation with antibiotics used in cystic fibrosis patients. *Int. J. Antimicrob. Agents* 40, 173–176. doi: 10.1016/j.ijantimicag.2012.04.006
- Gebru, E., Lee, J. S., Chang, Z. Q., Hwang, M. H., Cheng, H., and Park, S. C. (2009). Integration of pharmacokinetic and pharmacodynamic indices of orbifloxacin in beagle dogs after a single intravenous and intramuscular administration. *Antimicrob. Agents Chemother.* 53, 3024–3029. doi: 10.1128/AAC.01346-08
- Hengzhuang, W., Ciofu, O., Yang, L., Wu, H., Song, Z., Oliver, A., et al. (2013). High beta-lactamase levels change the pharmacodynamics of beta-lactam antibiotics in *Pseudomonas aeruginosa* biofilms. *Antimicrob. Agents Chemother.* 57, 196–204. doi: 10.1128/AAC.01393-12
- Jin, H., Zhou, R., Kang, M., Luo, R., Cai, X., and Chen, H. (2006). Biofilm formation by field isolates and reference strains of *Haemophilus parasuis*. *Vet. Microbiol.* 118, 117–123. doi: 10.1016/j.vetmic.2006.07.009
- Joo, H. S., and Otto, M. (2012). Molecular basis of *in vivo* biofilm formation by bacterial pathogens. *Chem. Biol.* 19, 1503–1513. doi: 10.1016/j.chembio.2012.10.022
- Kadurugamuwa, J. L., Sin, L., Albert, E., Yu, J., Francis, K., Deboer, M., et al. (2003). Direct continuous method for monitoring biofilm infection in a mouse model. *Infect. Immun.* 71, 882–890. doi: 10.1128/IAI.71.2.882-890.2003
- Macia, M. D., Rojo-Molinero, E., and Oliver, A. (2014). Antimicrobial susceptibility testing in biofilm-growing bacteria. *Clin. Microbiol. Infect.* 20, 981–990. doi: 10.1111/1469-0691.12651
- Moskowitz, S. M., Foster, J. M., Emerson, J., and Burns, J. L. (2004). Clinically feasible biofilm susceptibility assay for isolates of *Pseudomonas aeruginosa* from patients with cystic fibrosis. *J. Clin. Microbiol.* 42, 1915–1922. doi: 10.1128/JCM.42.5.1915-1922.2004
- Parra-Ruiz, J., Bravo-Molina, A., Pena-Monje, A., and Hernandez-Quero, J. (2012). Activity of linezolid and high-dose daptomycin, alone or in combination, in an *in vitro* model of *Staphylococcus aureus* biofilm. *J. Antimicrob. Agents Chemother.* 57, 2682–2685. doi: 10.1093/jac/dks272
- Seidl, K., Goerke, C., Wolz, C., Mack, D., Berger-Bachi, B., and Bischoff, M. (2008). *Staphylococcus aureus* CcpA affects biofilm formation. *Infect. Immun.* 76, 2044–2050. doi: 10.1128/IAI.00035-08
- Shan, Q., Liang, C., Wang, J., Li, J., and Zeng, Z. (2014). In vivo activity of cefquinome against *Escherichia coli* in the thighs of neutropenic mice. *Antimicrob. Agents Chemother.* 58, 5943–5946. doi: 10.1128/AAC.03446-14
- Spivey, J. M. (1992). The postantibiotic effect. *Clin. Pharm.* 11, 865–875.
- Tauber, M. G., Zak, O., Scheld, W. M., Hengstler, B., and Sande, M. A. (1984). The postantibiotic effect in the treatment of experimental meningitis caused by *Streptococcus pneumoniae* in rabbits. *J. Infect. Dis.* 149, 575–583. doi: 10.1093/infdis/149.4.575
- Turnidge, J. D. (1998). The pharmacodynamics of beta-lactams. *Clin. Infect. Dis.* 27, 10–22. doi: 10.1086/514622

## ACKNOWLEDGMENTS

This work was supported by the National Science Fund for Distinguished Young Scholars (grant no. 31125026); Program for Changjiang Scholars and Innovative Research Team in University of Ministry of Education of China (grant no. IRT13063); the Natural Science Foundation of Guangdong Province (grant no. S2012030006590); Science and Technology Planning Project of Guangdong Province China (grant no. 2012A020800004).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2015.01513>

- Wang, J., Shan, Q., Ding, H., Liang, C., and Zeng, Z. (2014). Pharmacodynamics of cefquinome in a neutropenic mouse thigh model of *Staphylococcus aureus* infection. *Antimicrob. Agents Chemother.* 58, 3008–3012. doi: 10.1128/AAC.01666-13
- Widmer, A. F., Frei, R., Rajacic, Z., and Zimmerli, W. (1990). Correlation between in vivo and in vitro efficacy of antimicrobial agents against foreign body infections. *J. Infect. Dis.* 162, 96–102. doi: 10.1093/infdis/162.1.96
- Zhang, B. X., Lu, X. X., Gu, X. Y., Li, X. H., Gu, M. X., Zhang, N., et al. (2014). Pharmacokinetics and ex vivo pharmacodynamics of cefquinome in porcine serum and tissue cage fluids. *Vet. J.* 199, 399–405. doi: 10.1016/j.tvjl.2013.12.015
- Zhao, D. H., Zhou, Y. F., Yu, Y., Shi, W., Yang, X., Xiao, X., et al. (2014). Integration of pharmacokinetic and pharmacodynamic indices of valnemulin in broiler chickens after a single intravenous and intramuscular administration. *Vet. J.* 201, 109–115. doi: 10.1016/j.tvjl.2014.05.010
- Zhou, Y. F., Zhao, D. H., Yu, Y., Yang, X., Shi, W., Peng, Y. B., et al. (2015). Pharmacokinetics, bioavailability and PK/PD relationship of cefquinome for *Escherichia coli* in Beagle dogs. *J. Vet. Pharmacol. Ther.* 38, 543–548. doi: 10.1111/jvp.12225

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Zhou, Shi, Yu, Tao, Xiong, Sun and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# RNA-seq *de novo* Assembly Reveals Differential Gene Expression in *Glossina palpalis gambiensis* Infected with *Trypanosoma brucei gambiense* vs. Non-Infected and Self-Cured Flies

Illiassou Hamidou Soumana<sup>1</sup>, Christophe Klopp<sup>2</sup>, Sophie Ravel<sup>1</sup>, Ibouniyamine Nabihoudine<sup>2</sup>, Bernadette Tchicaya<sup>1</sup>, Hugues Parrinello<sup>3,4,5,6</sup>, Luc Abate<sup>7</sup>, Stéphanie Rialle<sup>3,4,5,6</sup> and Anne Geiger<sup>1\*</sup>

## OPEN ACCESS

### Edited by:

Andres M. Perez,  
University of Minnesota, USA

### Reviewed by:

Paras Jain,

Albert Einstein College of Medicine,  
USA

Li Xu,

Cornell University, USA

### \*Correspondence:

Anne Geiger  
anne.geiger@ird.fr

### Specialty section:

This article was submitted to  
Infectious Diseases,  
a section of the journal  
*Frontiers in Microbiology*

Received: 26 June 2015

Accepted: 29 October 2015

Published: 13 November 2015

### Citation:

Hamidou Soumana I, Klopp C, Ravel S, Nabihoudine I, Tchicaya B, Parrinello H, Abate L, Rialle S and Geiger A (2015) RNA-seq *de novo* Assembly Reveals Differential Gene Expression in *Glossina palpalis gambiensis* Infected with *Trypanosoma brucei gambiense* vs. Non-Infected and Self-Cured Flies.

Front. Microbiol. 6:1259.  
doi: 10.3389/fmicb.2015.01259

<sup>1</sup> UMR 177, Institut de Recherche Pour le Développement-CIRAD, CIRAD TA A-17/G, Montpellier, France, <sup>2</sup> Institut National de la Recherche Agronomique, GenoToul, UR875, Castanet-Tolosan, France, <sup>3</sup> Centre National de la Recherche Scientifique, Unité Mixte de Recherche 5203, Institut de Génomique Fonctionnelle, Montpellier, France, <sup>4</sup> Institut National de la Santé et de la Recherche Médicale U661, Montpellier, France, <sup>5</sup> Universités de Montpellier 1 and 2, UMR 5203, Montpellier, France, <sup>6</sup> Montpellier GenomiX, Institut de Génomique Fonctionnelle, Montpellier, France, <sup>7</sup> UMR MIVEGEC (Institut de Recherche pour le Développement 224-Centre National de la Recherche Scientifique 5290-UM1-UM2), Institut de Recherche pour le Développement, Montpellier, France

*Trypanosoma brucei gambiense* (Tbg), causing the sleeping sickness chronic form, completes its developmental cycle within the tsetse fly vector *Glossina palpalis gambiensis* (Gpg) before its transmission to humans. Within the framework of an anti-vector disease control strategy, a global gene expression profiling of trypanosome infected (susceptible), non-infected, and self-cured (refractory) tsetse flies was performed, on their midguts, to determine differential genes expression resulting from *in vivo* trypanosomes, tsetse flies (and their microbiome) interactions. An RNAseq *de novo* assembly was achieved. The assembled transcripts were mapped to reference sequences for functional annotation. Twenty-four percent of the 16,936 contigs could not be annotated, possibly representing untranslated mRNA regions, or Gpg- or Tbg-specific ORFs. The remaining contigs were classified into 65 functional groups. Only a few transposable elements were present in the Gpg midgut transcriptome, which may represent active transpositions and play regulatory roles. One thousand three hundred and seventy three genes differentially expressed (DEGs) between stimulated and non-stimulated flies were identified at day-3 post-feeding; 52 and 1025 between infected and self-cured flies at 10 and 20 days post-feeding, respectively. The possible roles of several DEGs regarding fly susceptibility and refractoriness are discussed. The results provide new means to decipher fly infection mechanisms, crucial to develop anti-vector control strategies.

**Keywords:** *de novo* assembly, *Glossina palpalis gambiensis*, human African trypanosomiasis, *in vivo* metatranscriptomics, RNA-seq, *Trypanosoma brucei gambiense*

## INTRODUCTION

Human African Trypanosomiasis (HAT), one of the most neglected tropical diseases in the world (Brun et al., 2010), is endemic to 36 countries of sub-Saharan Africa, where it results in a loss of 1.5 million disability-adjusted life years every year (Hotez et al., 2009). This devastating disease has been targeted for elimination by the WHO and PATTEC (Pan-African Tsetse and Trypanosomiasis Eradication Campaign), and subsequently by the London declaration on neglected tropical diseases. In terms of mortality, the disease is ranked ninth out of 25 human infectious and parasitic diseases in Africa (Welburn et al., 2009). Sleeping sickness remains responsible to this day for major hindrances to social, agricultural, and economic development in Africa.

HAT is caused by two subspecies of African trypanosomes transmitted by tsetse flies: *Trypanosoma brucei gambiense* (Tbg) is responsible for the chronic form of HAT in Western and Central Africa, while *Trypanosoma brucei rhodesiense* is responsible for the acute form of the disease in East Africa (Kennedy, 2008; Franco et al., 2014). In recent years, the number of new cases has begun to decrease, mirroring a situation previously observed in the 1960s and which preceded the last heavy outbreak in the 1990s.

To date, no vaccine is available to prevent sleeping sickness. Moreover, several currently used drugs cause harmful side effects, in addition to inducing trypanosome-resistant strains (Baker et al., 2013). Furthermore, some diagnostic tools are inefficient for proper HAT detection (Simarro et al., 2008; Geiger et al., 2011). The search for novel strategies, including alternative vector-based strategies (Rio et al., 2004; Aksoy et al., 2008; Medlock et al., 2013), must therefore be pursued further. One such approach, the release of sterile *Glossina* males to drastically decrease the targeted population size, was successfully tested in Zanzibar (Vreysen et al., 2000; Abd-Alla et al., 2013). However, even though these sterile males are not trypanosome-infected, they are still able to acquire trypanosomes from an infected host and transmit them to non-infected humans. Therefore, releasing flies that are both sterile and resistant to trypanosome infection (refractory flies) could be more effective and a lesser risk for humans. This type of approach requires deciphering the physiological mechanisms that govern fly refractoriness to trypanosome infection, in order to develop methodologies for enhancing tsetse fly resistance.

Refractoriness is the status of most tsetse flies, as shown by the typical low prevalence of trypanosome infections in natural fly populations in HAT foci, as well as in flies submitted to an experimental infection. In the latter case, flies are fed on trypanosome-infected mice displaying high levels of parasitemia. Even though all the flies ingested trypanosomes, typically only around 15% become infected; the others were either self-cured from the ingested parasites, or they did not produce mature parasites and therefore never became infective (Moloo et al., 1986; Dukes et al., 1989; Frézil and Cuisance, 1994; Maudlin and Welburn, 1994; Jamonneau et al., 2004). Understanding the biological processes leading to the elimination of ingested trypanosomes or parasite maturation failure, identifying the key steps and the key factors involved, and investigating different

means to stimulate refractoriness will all help to effectively combat sleeping sickness.

The existence of two distinct pathways, one in which ingested trypanosomes are eliminated by refractory flies and the other in which trypanosomes are established in the gut and achieve their developmental cycle in susceptible flies, clearly demonstrates the occurrence of complex molecular interactions. These interactions are not restricted to cross-talk between the invading trypanosomes and the tsetse fly. For example, tsetse can harbor the primary obligate symbiont *Wigglesworthia glossinidia* and the secondary symbiont *Sodalis glossinidius*, which are known to favor fly infection by trypanosomes (Geiger et al., 2007; Farikou et al., 2010). Significant modulation of *Wigglesworthia* and *Sodalis* gene expression was previously recorded following fly trypanosome invasion (Hamidou Soumana et al., 2014a,b). Finally, field flies have been shown to harbor a large diversity of bacterial species (Geiger et al., 2013; Hamidou Soumana et al., 2013), suggesting that the whole microbiome may be involved in modulating the fly global response to trypanosome invasion, and consequently the fly vector competence.

The physiological mechanisms involved in this vector competence (i.e., its ability to acquire the parasite, to favor its maturation, and to transmit it to a mammalian host) are not well understood, and the genes that control it remain largely unknown. Nevertheless, some responses have been identified (reviewed in Geiger et al., 2015). For example, several studies have reported that an RNAi approach used to silence genes controlling the Imd pathway (Hao et al., 2001) and the tsetse fly immune-responsive glutamine/proline-rich (EP) protein (Haines et al., 2010) increase midgut colonization efficiency. The importance of reactive oxygen species as determinants of resistance have been similarly demonstrated (MacLeod et al., 2007a; Macleod et al., 2007b; Nayduch and Aksoy, 2007). Recently, Weiss et al. (2013, 2014) demonstrated the importance of microbiome-regulated host immune barriers in establishing the trypanosome infection. In addition, as trypanosomes migrate from the gut toward the salivary glands they reach the proventriculus, an immune-active tissue expressing the nitric oxide synthase gene and containing increased levels of nitric oxide, reactive oxygen intermediates, and hydrogen peroxide (Hao et al., 2003). Only a few trypanosomes will survive and complete their migration to the salivary glands, where they multiply and evolve into their infectious metacyclic form.

We previously investigated 12 immune genes selected from those formerly reported by Lehane et al. (2003) to be highly over-expressed in *Glossina morsitans morsitans* challenged with *T. b. brucei* (Hamidou Soumana et al., 2014c). Nevertheless, deciphering the mechanisms that allow trypanosomes to adapt to the different tsetse fly microenvironments and thereby escape insect immune responses requires a more global approach. We have therefore performed a large comparative transcriptome analysis of trypanosome-infected, non-infected and self-cured (refractory) *Glossina palpalis gambiensis* (Gpg) flies, the vector of *T. b. gambiense*, the trypanosome causing the chronic form of HAT in West Africa. The present work follows our previous investigations of differential gene expression in *Sodalis* and *Wigglesworthia* (Hamidou Soumana et al., 2014a,b) by focusing

on the differentially expressed genes (DEGs) from both flies and trypanosomes, since some genes could be used as targets to enhance tsetse fly refractoriness to trypanosome infection. Since the establishment step is fundamental to the trypanosome life cycle within its vector, our investigation has once again focused on the tsetse fly midgut, where the ingested trypanosome is established (or not). To investigate global infection dynamics at key early time points, samples were collected at 3, 10, and 20 days post-feeding on either trypanosome-infected or non-infected bloodmeals.

The analyses were performed using an RNA-seq *de novo* assembly approach—“*a revolutionary tool for transcriptomics*” (Wang et al., 2009). Our report presents results of transcriptome read assembly. *G. p. gambiensis* and *G. m. morsitans* being two different species, functional annotation was performed with reference to a broad panel of insect data bases including *G. m. morsitans*. Here, we identified DEGs in susceptible and refractory tsetse. In addition, we have identified single nucleotide polymorphisms (SNPs) and their variants (insertions and deletions) and have evaluated their relationships within the levels of gene expression in the different samples. Finally, this study highlights molecular interactions on the basis of biosynthesis pathways controlled by genes shown to be differentially expressed.

## MATERIALS AND METHODS

### Ethical Statement

All reported experiments on animals were conducted according to internationally recognized guidelines. The experimental protocols (numbers 12TRYP03, 12TRYP04, and 12TRYP06) were approved by the Ethics Committee on Animal Experiments and the Veterinary Department of the Centre International de Recherche Agronomique pour le Développement (CIRAD), Montpellier, France.

### Glossina Species and Trypanosome Strains used for Experimental Infections

*G. p. gambiensis* flies and the *T. b. gambiense* isolate T.b.g. S7/2/2 used in this study have been previously described (Hamidou Soumana et al., 2014a,b).

### Experimental Design and Sampling Procedures

Preliminary note: the samples analyzed in the present study were previously used to identify DEGs in *Sodalis* and *Wigglesworthia* (Hamidou Soumana et al., 2014a,b). The experimental steps described in this Section (“Experimental Design and Sampling Procedures”) are similar to those described in the latter studies. Additional experimental steps (described in “RNA-seq: Sample Preparation and Sequencing” and the following Sections) are specific to the present study.

Briefly, a set of 100 randomly chosen *G. p. gambiensis* teneral (<32 h old) female flies were fed on non-infected mice. Three days after feeding, two biological replicates of seven flies each were dissected and the seven midguts from each replicate were

pooled (=sample NS3 in two replicates). A second set of 900 *G. p. gambiensis* teneral (<32 h old) female flies were fed on *T. b. gambiense*-infected mice (averaging 20 flies per mouse), which displayed parasitemia levels ranging between 16 and 64 × 10<sup>6</sup> parasites/ml of blood. Flies were then randomly separated into three groups.

The first group of flies was recovered 3 days after the infective feeding (=“stimulated flies” or S3 samples) and randomly separated into two biological replicates of seven flies each. The flies from each replicate were dissected separately and the corresponding midguts were pooled in RNAlater (Ambion) and stored at –80°C until RNA extraction.

The second group of flies was recovered 10 days after the infective feeding. DNA was extracted from anal drops by the chelex method (Ravel et al., 2003), and the presence of *T. b. gambiense* in their anal drops was confirmed by PCR using specific primers (Moser et al., 1989). Based on the PCR results, flies were separated into one of two subgroups: (a) those with trypanosomes in their anal drops (=infected flies or I10 samples); and (b) those not displaying trypanosomes in their anal drops (=self-cured flies or NI10 samples). Each subgroup was further divided randomly into two biological replicates of three flies each (at this sampling time the prevalence of infected flies was <5%). The flies from each replicate were then processed as above.

The third group of flies was recovered 20 days after feeding on trypanosome-infected mice and was processed similarly to the second group. Infection prevalence was high enough at this sampling time to establish two replicates of seven flies each (infected flies = I20; self-cured flies = NI20 samples).

Finally, transcriptome analyses were performed on a total of 12 samples, representing six “categories” of differently treated flies (S3 and NS3; I10 and NI10; I20 and NI20). Each category was further subdivided into two biological replicates.

### RNA-seq: Sample Preparation and Sequencing

RNA was extracted from the pooled midguts of each biological replicate using TRIzol reagent (Gibco-BRL, Life Technologies), according to the manufacturer’s protocols. RNA pellets were resuspended in nuclease-free water and the concentration was quantified using a NanoDrop spectrophotometer. RNA quality and the absence of DNA contamination were confirmed on a 2100 Bioanalyzer chip (Agilent Technologies, Santa Clara, CA, USA) prior to cDNA library synthesis.

cDNA libraries were prepared (using 4 µg of total RNA from each sample) for subsequent Illumina sequencing with the mRNA-seq Sample Prep kit (Illumina, San Diego, CA, USA). Specifically, RNA was fragmented and used as a template for a randomly primed PCR. After amplification, ends were repaired and ligated to Illumina adapters. The cDNA library was then verified for appropriate fragment size (200–300 bp) on a BioAnalyzer chip. Libraries were amplified onto flow cells using an Illumina cBot and the fragments were sequenced, using a paired-ends strategy, on an Illumina HiSeq2000 (Illumina, San Diego, CA, USA) for 2 × 101 cycles, according to the manufacturer’s protocols. The barcoded libraries were multiplexed by 4 on a single lane. Paired-end raw reads were

automatically trimmed and validated by screening for low quality (e.g., short sequences or ambiguous nucleotides), low complexity, and contaminants. These false reads were removed from the study and the remaining reads were assembled *de novo*.

The  $2.91 \times 10^8$  raw sequencing reads were filtered to remove bad quality bases and reads, resulting in  $2.76 \times 10^8$  remaining reads (95.14%). All reads were then used for *de novo* assembly of the transcriptome. Datasets for the reads are available from the NCBI Short Read Archive (SRA), accession number SRP046074.

## De novo Assembly and Transcriptome Analysis

To construct the *G. p. gambiensis* assembled whole transcriptome, all short reads obtained from infected, stimulated and non-infected, non-stimulated tsetse flies at 3, 10, and 20 days post-feeding were first assembled into contigs with no gap, using the *de novo* transcriptome assembly software programs Velvet and Oases (Zerbino and Birney, 2008). Each read was then mapped back to the contigs using the Bwa short-read aligner (Li and Durbin, 2009), to generate the gross count per contig for each biological replicate representing the different conditions. The assembled contigs were annotated by BLASTX alignment ( $E < 0.00001$ ) to protein databases such as the NCBI NR (<http://www.ncbi.nlm.nih.gov>), Swiss-Prot (<http://www.expasy.ch/sprot>), ensembl-pep, refseq-rna, refseq-protein, and FlyBase databases. Contig sequences were deposited at the NCBI Transcriptome Shotgun Assembly (TSA) Database under BioProject PRJNA260242. Gene Ontology (GO) annotation assignment (Ashburner et al., 2000) was used to perform functional gene annotation by mapping GO terms using the NCBI NR, GO (<http://www.geneontology.org/>), and UniProt (<http://www.ebi.ac.uk/UniProt/>) databases; *E*-value cutoff of  $10^{-5}$  (Conesa et al., 2005).

## Technical Description of the Assembly Process

### Per Condition Assembly

Read pairs were first cleaned from remaining sequencing adapter sequences using the trim\_galore script ([http://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/); Smallwood et al., 2014). Over-represented reads were then filtered out using the normalize\_by\_kmer\_coverage.pl script from the Trinity software package (Haas et al., 2013). In the next step, invalid base calls were discarded by extracting the longest sub-sequence without Ns from each read. Specifically, if the length of the longest sub-sequence did not exceed half of the sequence length, the read, and its pair were removed. The final step was performed using an in-house script.

Nine assemblies using nine different k-mers (25, 31, 37, 43, 49, 55, 61, 65, and 69) were performed on pre-processed input data. Each assembly produces a transcripts.fa file and each raw transcripts.fa file is organized into loci. Rather than referring to genetic loci, each locus is actually a collection of similar sequences including (but not limited to) splice variations and partial assemblies of the longer transcripts in the locus. We chose to keep only the best transcript for each locus, using

the script OasesV0.2.04OutputToCsvDataBase.py (<http://code.google.com/p/oases-to-csv/>; Schulz et al., 2012). Subsequently, all files were merged and anti-sense chimeras (accidentally produced by the assembly step) were cut with a homemade script.

Identical contigs produced by different k-mers were removed using the cd-hit-est program (Li and Godzik, 2006). Because different k-mers sometimes construct different transcript parts, we used TGICL (Pertea et al., 2003), an OLC (overlap layout consensus) assembler, to assemble contigs displaying significant overlaps. The contigs were also filtered to a minimum length of 200 bp.

Input reads were then mapped back to the contigs using the bwa aln function (Li and Durbin, 2009). The resulting alignment files were used to correct contig sequences from spurious insertions and deletions resulting from an in-house script, and to filter out contigs with very low coverage. The filter excludes contigs with less than two mapped reads per million.

### Meta-assembly

All single condition contig fasta files were concatenated, and each contig was renamed by adding the condition name to the beginning of its name. The longest open reading frame (ORF) of each contig was then searched using the getorf program from EMBOSS (Rice et al., 2000). A cd-hit clustering was performed on ORFs with a sequence identity  $\geq 0.9$ , in order to extract from each cluster the contig with the longest ORF, or the longest contig (if the ORF sizes were identical). A clustering using cd-hit-est was then done on selected ORFs with a sequence identity  $\geq 0.95$ . Input reads from all conditions were then mapped back to the contigs using bwa (Li and Durbin, 2009). Contigs with very low coverage (<1 mapped read per million) were filtered out.

## Analysis of DEGs

DEG's were identified using the DESeq software, version 1.16.0 (Anders and Huber, 2010). This method represents widely accepted and complementary analytical approaches for RNASeq data. The raw read counts were produced by realigning the read on the contigs. These counts were used as inputs for DESeq to calculate the normalized expression for each contig in the different samples (e.g., trypanosome-stimulated and non-stimulated tsetse flies at 3 days and trypanosome-infected and non-infected tsetse flies at 10 and 20 days). Differential expression was then reported as fold change, with associated *p*-values. DESeq calculates *p*-values using a negative binomial distribution that accounts for technical as well as biological variability. The resulting raw *p*-values were corrected for multiple tests using the False Discovery Rate (Benjamini and Hochberg, 1995). Contig pairs whose read numbers displayed a greater than two-fold difference between the selected conditions (with  $p < 0.05$ ) were identified as DEGs. The DESeq approach is well suited for count data (i.e., read counts), as is the case for RNA-Seq experiments, and the method estimates variance in a local fashion for varying signal strength (Trapnell et al., 2010).

For functional analysis, all DEGs were mapped to terms in the GO database, which requires *E*-values adjusted for multiple testing to be  $<0.001$ . The annotation of all significant genes was further supplemented by BLASTX, conserved domains, and

literature searches. Using these combined approaches resulted in a functionally driven classification.

## SNP Identification

Many SNP panels have been build using an RNA-Seq assembly reference for species without reference genome (Salem et al., 2012; Swaminathan et al., 2012; see also GATK Program which is the classical tool to evidence SNPs).

In order to call putative variants (SNPs, insertions and deletions), the alignment files were cleaned from reads with low mapping quality and PCR duplicates using the samtools software, v 0.1.19-44428cd. The remaining reads were recalibrated and realigned with the GATK Program 2.4-9-g532efad (DePristo et al., 2011), which was also used for variant calling (Unifiedgenotyperalgorithm). Variants were filtered using a Phred quality score of 20.

Variants were deposited at the NCBI single nucleotide polymorphism database (dbSNP) with the accession number SUB833398.

## RESULTS

### Infection Time Course

At day 10 after fly feeding, the anal drops from 13 out of 262 tsetse flies (4.96%) were PCR-positive for trypanosomes. At day 20 after fly feeding, the anal drops of 43 out of 349 flies (12.32%) were PCR-positive for trypanosomes.

### Transcriptome Sequencing by RNA-seq and de novo Assembly

Twelve RNA-seq libraries were prepared from total RNA extracted from pooled midguts of non-stimulated and non-infected (refractory) tsetse flies [representing 3 (NS3), 10 (NI10), and 20 days (NI20) post-bloodmeal uptake] and from pooled midguts of stimulated or infected tsetse flies [representing 3 (S3), 10 (I10), and 20 days (I20) post-bloodmeal uptake].

A total of approximately 520 million raw reads (50 million paired-end reads, 2 × 100 bp) representing approximately 165 GB of sequence data were generated from 12 independent 200 bp insert libraries. Prior to *de novo* assembly, the quality of the reads was assessed using the base-calling quality scores (Cock et al., 2010) from Illumina's base-caller RTA software. Most reads displayed Phred-like quality scores at the Q20 level, indicating a sequencing error probability of 0.01%. After trimming and cleaning, between 58,370,072 and 87,387,674 read pairs were kept, depending on the sample library.

*De novo* assembly was then performed using the Velvet software. Oases is a *de novo* transcriptome assembler designed to produce extended contigs from short read sequencing technologies in the absence of any genomic reference. It clusters

the contigs from a preliminary assembly by Velvet into small groups and uses a de Bruijn graph-based algorithm to construct transcript isoforms (Schulz and Zerbino, 2010). The contigs, produced by Velvet, were post-processed using Oases.

This yielded a total of 16,936 contigs ranging from 137 to 4836 bp (average length: 2302 bp), with an N50 at 3036 and a N90 at 1215 (Table 1, Figure 1). The assembled transcriptome size is 38,986,687 bp.

### Functional Annotation and Classification of Assembled Contigs

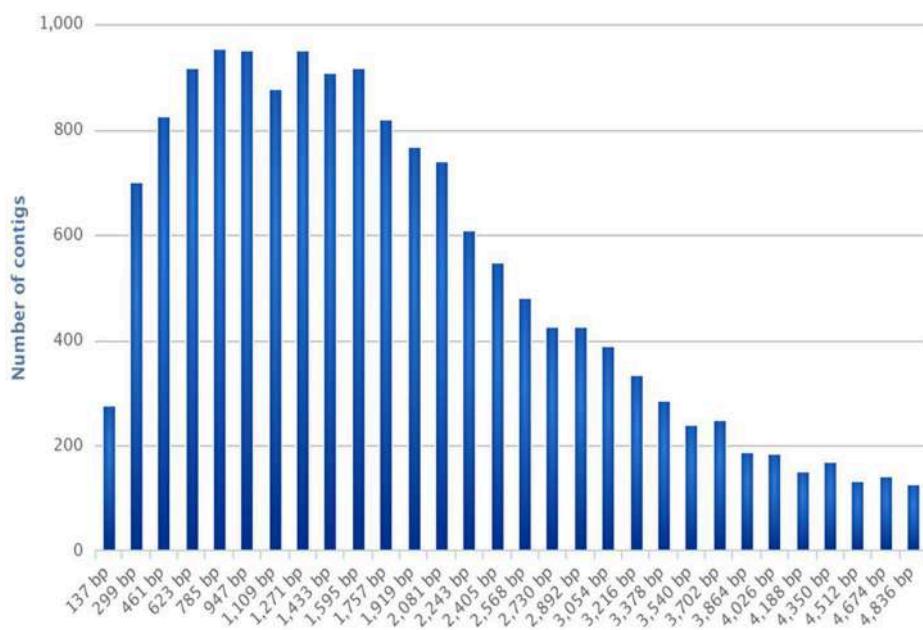
A total of 12,806 contigs could be annotated (out of 16,936) from which 9698 were annotated with reference to the sequences recorded in the Refseq-proteins database. Subsequently, 1303 and 1805 contigs were annotated with reference to the sequences recorded in the Refseq-rna and Swiss-Prot databases, respectively. In the end, 24.39% of the contigs could not be annotated. Nevertheless, these orphan sequences may be of great interest, as they could refer to putative *G. p. gambiensis* and *T. b. gambiense* specific biological functions (Figure 2), and therefore specific genes.

Our annotated dataset including 16,936 contigs (Supplementary Table S1) is most likely representative of the *G. p. gambiensis* gene catalog. In terms of the total number of hits, BLASTX hits and top hits were mostly identified with *Ceratitis capitata* (5443 hits), *Drosophila melanogaster* (1656 hits), *Trypanosoma brucei* (838 hits), *Drosophila willistoni* (626 hits), *Drosophila virilis* (608 hits), and *Drosophila mojavensis* (561 hits). Less than 24% of the *G. p. gambiensis* annotated consensus transcriptome had orthologous hits in 14 other species, including several *Drosophila* species, *Acyrtosiphon pisum*, *Hydra magnipapillata*, *Anopheles* sp., *Bombyx* sp., *Aedes* sp., and *Glossina morsitans* (Figure 3).

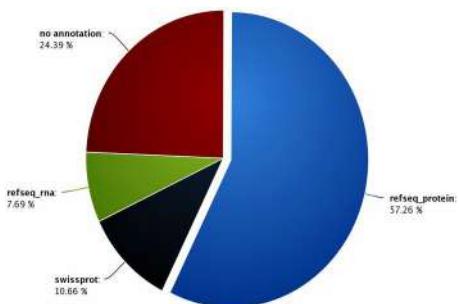
Among the 16,936 contigs, 7207 could be assigned to three main GO: “biological process” (3702 contigs) was the predominant domain, followed by the “molecular function” (2191 contigs) and “cellular component” (1314 contigs) domains (Figure 4). GO annotation assignments classified contigs into 30 subcategories within the biological process domain, 10 within the molecular function domain, and 25 within the cellular component domain (Figure 4). The “biological process” domain subcategories that displayed the most highly abundant transcripts include: gene expression (348 transcripts—9.4% of the “biological process” domain transcripts), system development (345 transcripts—9.3%), neurological system process (300 transcripts—8.1%), responses to stimuli (293 transcripts—7.9%), transport (292 transcripts—7.9%), signal transduction (234 transcripts—6.3%), coagulation (230 transcripts—6.2%), cellular process (216 transcripts—5.8%), and differentiation (207 transcripts—5.6%). The “molecular

**TABLE 1 | Measurement of contigs from de novo assembly.**

Sample ID	N75	N50	N25	Mini	Maxi	Ave	Count	Total (bp)
Assembly from 12 combined libraries	2910	1796.5	1042	137	4836	2302	16,936	38,986,687



**FIGURE 1 | Frequencies of contig length distribution.** The histogram represents the number of contigs per contig length (expressed in bp).



**FIGURE 2 | Percentage of contigs annotated with reference to the Refseq-proteins, Refseq-rna, and Swiss-Prot databases.**

function” domain subcategories that displayed the most highly abundant transcripts include: binding (1559 transcripts—71.2% of the “molecular function” domain transcripts), catalytic activity (260 transcripts—11.9%), and channel activity (191 transcripts—8.7%). Finally, the “cellular component” domain subcategories that displayed the most common groups of proteins include: membrane (281 transcripts—21.4% of the “cellular component” domain transcripts), nucleus (246 transcripts—18.7%), and macromolecular complex (161 transcripts—12.3%) (Figure 4).

## Detection and Identification of DEGs in Response to Tsetse Infection by Trypanosomes

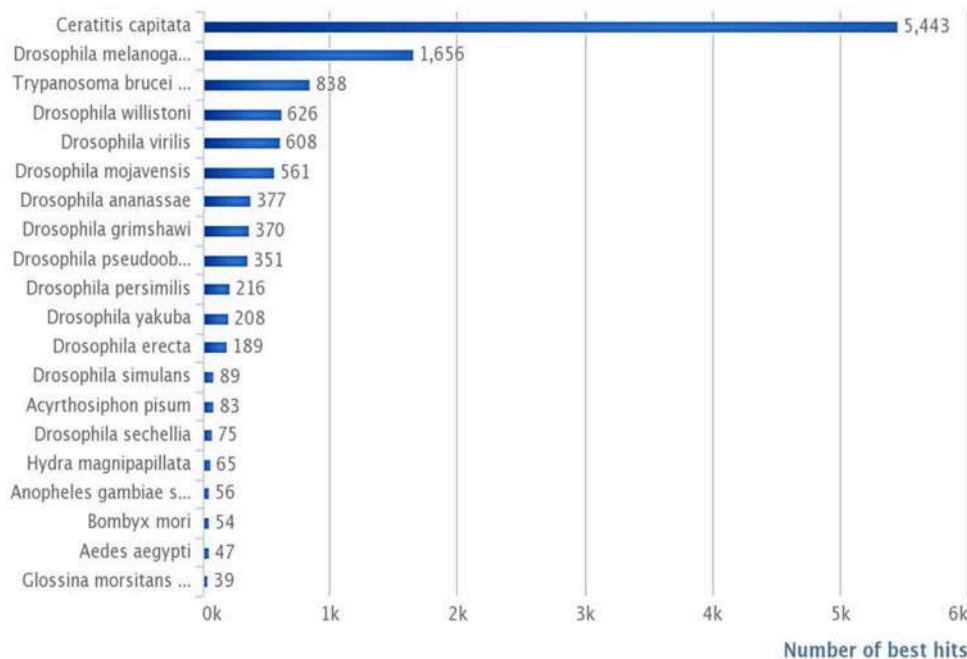
We compared 12 tsetse fly (*G. p. gambiensis*) transcriptome profiles to better understand their pathosystem at the

transcriptome level (S3 vs. NS3; I10 vs. NI10; I20 vs. NI 20 tsetse flies).

We observed significantly differentially expressed genes (Figures 5, 6;  $p < 0.05$ ) between S3 and NS3 flies (1373 genes), I10 and NI10 flies (52 genes), and I20 and NI20 flies (1025 genes) (Supplementary Tables S2–S4). Among the DEGs identified for 3-day samples, names could be assigned to 797 contigs with reference to the *T. brucei* database, and to 435 contigs with reference to the insect database; 141 contigs remained “hypothetical.” Among the DEGs identified for 10-day samples, names could be assigned to 39 contigs with reference to the insect database, and 13 remained “hypothetical.” Finally, among the DEGs identified for the 20-day samples, names could be assigned to 866 contigs with reference to the *T. brucei* database, and 112 contigs with reference to the insect database; 47 remained “hypothetical.”

When comparing day three sampled flies that ingested a non-infected bloodmeal (NS3 flies) with flies that ingested an infected bloodmeal (S3 flies), 208 transcripts showed an up-regulated expression (Fold Change  $> 1$ ) in non-stimulated flies, whereas 1165 transcripts were down-regulated (Fold change  $< 1$ ) (Supplementary Table S2). In self-cured flies sampled 10 days after ingesting an infected bloodmeal (NI10 flies), 19 transcripts were up-regulated and 33 were down-regulated when compared to the corresponding genes of infected flies (I10 flies) (Supplementary Table S3). Finally, in self-cured flies sampled 20 days after ingesting an infected bloodmeal (NI20 flies), 49 contig-derived transcripts were up-regulated and 976 were down-regulated when compared to the corresponding genes of infected (I20) flies (Supplementary Table S4).

GO-based classification was performed on the characterized DEGs and categories, in order to identify which ones were



**FIGURE 3 | Number of best hits per species, in reference to the total number of contigs.**

significantly altered during invasion and infection of tsetse flies by trypanosomes (Figures 7–9). At day 3 sampling, GO analysis classified 151 of the annotated DEGs into 26, 8, and 16 subgroups within the biological process, molecular function, and cellular component categories, respectively (Figure 7). In the biological process category, the subcategories that were most affected by trypanosome stimulation were: metabolic process (8.9%), system development (6.9%), response to stimulus (10%), signal transduction (7.3%), transport (6.6%), and gene expression (6.6%); in addition, 18 identified metabolic pathways were significantly affected by trypanosome stimulation.

At day 10 sampling, GO analysis classified 17 of the annotated DEGs into 13, 5, and 9 subgroups within the biological process, molecular function, and cellular component categories, respectively (Figure 8). In the biological process category, the subcategories most affected by trypanosome infection were: metabolic process (e.g., carbohydrate metabolic process; 18.5%), biological process, gene expression, neurological system process, and response to stimuli (11.1% each). The binding subcategory was the most affected by trypanosome infection (53.8%) within the molecular function category.

Finally, at day 20 post infected blood meal uptake, GO analysis classified 57 of the annotated DEGs into 24, 9, and 20 subgroups within the biological process, molecular function, and cellular component categories, respectively (Figure 9). In the biological process category, the subcategories that were most affected by trypanosome infection were: gene expression (12.1%), metabolic process (11.5%), transduction (9.5%), response to stimulus (8.8%), and morphogenesis (8.1%). In the molecular function category, the most affected subcategories were: binding (47.4%) and catalytic activity (24.4%). Finally, in the cellular

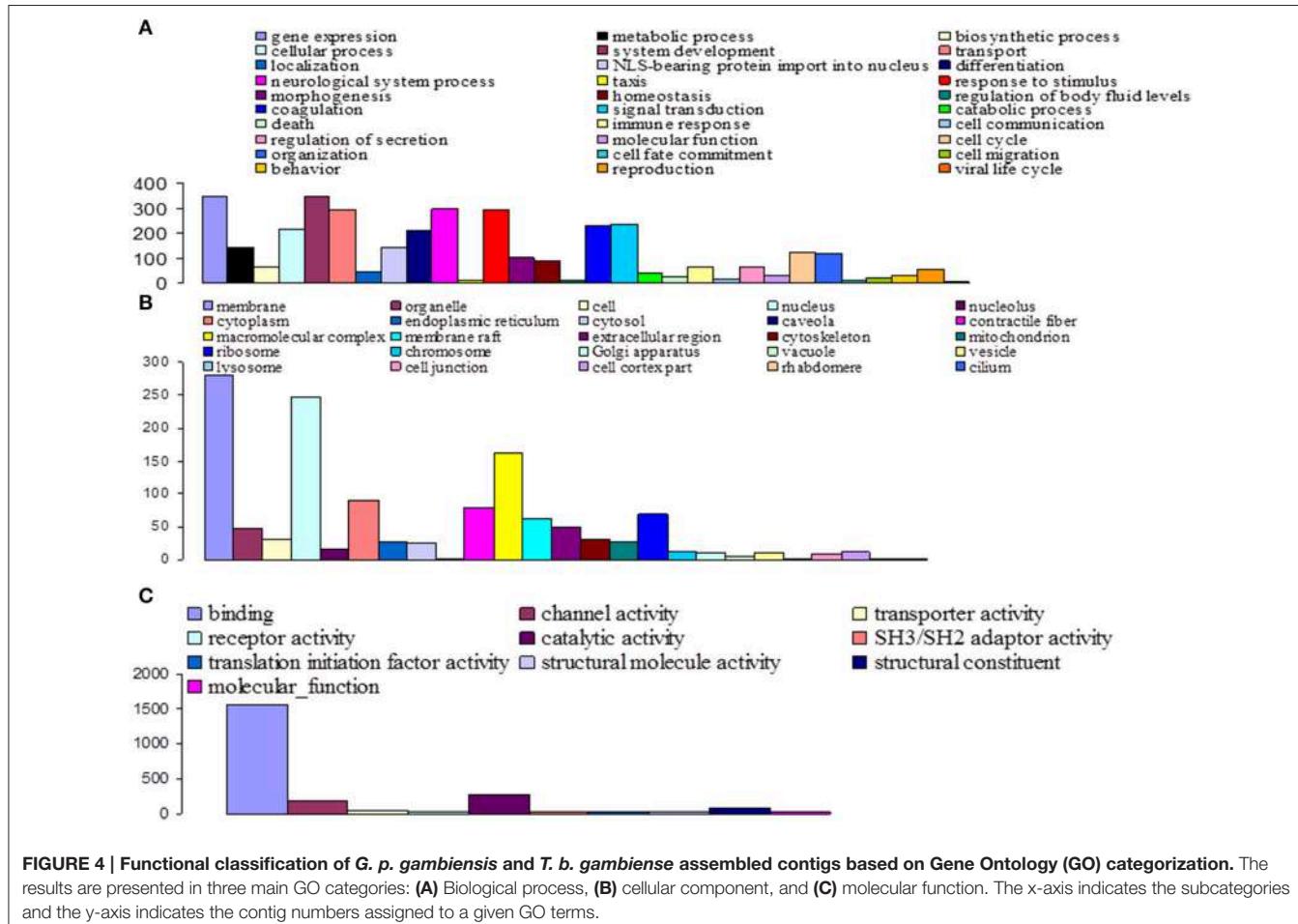
component category, the most affected subcategories were: membrane (17.4%), cytoplasm (16.3%), nucleus (13%), and ribosome (10.9%).

### Refined List of DEGs of Interest

Some DEGs appear *a priori* to be of greater interest than others, owing either to their level of over-expression or down-expression in S3, I10 and I20 samples vs. NS3, NI10, and NI20 samples, or the protein function that they encode. These particular DEGs selected for days 3, 10, and 20 are presented together in Table 2. As expected, trypanosome genes (noted in the identification column as “GLOS\_TB ...”) were expressed only in samples from flies that had ingested a trypanosome-infected bloodmeal; similar results are presented for the overall DEGs in Supplementary Table S2 (day 3 samples), Supplementary Table S3 (day 10 samples), and Supplementary Table S4 (day 20 samples). However, we could not detect any evidence for trypanosome gene expression in the I10 samples (see the Discussion section). Some DEGs were observed to be expressed in both S3 and I20 flies; their mean expression levels are compared in Table 3. The set of genes that were previously identified as being mostly trypanosome genes were expressed much higher in I20 vs. S3 samples (with the exception of two genes). Finally, the set of genes reported in Table 2 was sorted on the basis of the function (mostly catalytic activity) of the proteins they encode. Interestingly, genes encoding proteases were predominant, whether they belong to the tsetse or trypanosome genome.

### SNP Detection

In our study, SNPs were identified after realigning the reads on the 16,936 contigs. After applying filters, the analysis performed



**FIGURE 4 | Functional classification of *G. p. gambiensis* and *T. b. gambiense* assembled contigs based on Gene Ontology (GO) categorization.** The results are presented in three main GO categories: (A) Biological process, (B) cellular component, and (C) molecular function. The x-axis indicates the subcategories and the y-axis indicates the contig numbers assigned to a given GO terms.

on all 16,936 contigs resulted in the identification of 195,464 high confidence SNPs from 14,929 contigs (average = 11 SNPs per contig). Detected polymorphisms were more due to transition (178,328) than to insertion (11,269) and deletion (5867) processes.

SNPs were also revealed in the DEGs from 3-, 10-, and 20-day tsetse fly samples (Supplementary Tables S5–S7, respectively). SNPs could be assigned to 625, 17, and 480 annotated contigs from the 3-, 10-, and 20-day tsetse samples, respectively. Three hundred and ninety nine annotated contigs showing SNPs were identified both in 3 and 20-days tsetse samples.

DEGs from 3-day samples in which SNPs were identified encode such proteins as proteases, antimicrobial peptides, glucose metabolism enzymes, nucleotide metabolism enzymes, proteins involved in transcription process, chitinases, and aquaporins (Supplementary Table S5). DEGs from 10-day samples displaying SNPs were found to encode glucose metabolism enzymes, lectizyme, glutathione S-transferase, and thrombin inhibitor (Supplementary Table S6). Finally, DEGs from 20-day samples displaying SNPs were found to encode such proteins as proteases, glucose metabolism enzymes, nucleotide metabolism enzymes, and proteins involved in transcription process, as well as the

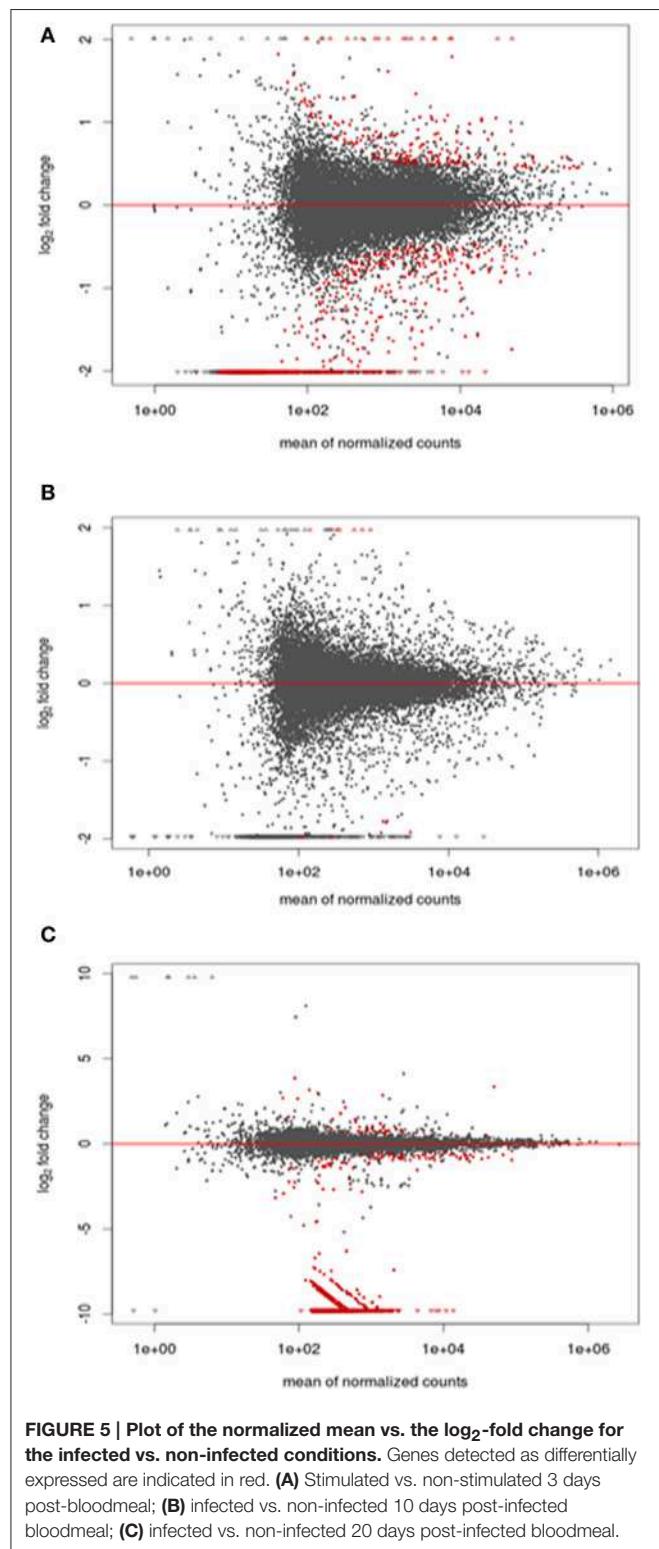
trypanosome reactive oxygen species detoxification system (Supplementary Table S7). Sequences encoding stress-related proteins, such as heat-shock proteins, were also identified among the DEGs that carried SNPs; these were referenced as belonging to the *G. p. gambiensis* and *T. b. gambiense* transcriptomes.

## DISCUSSION

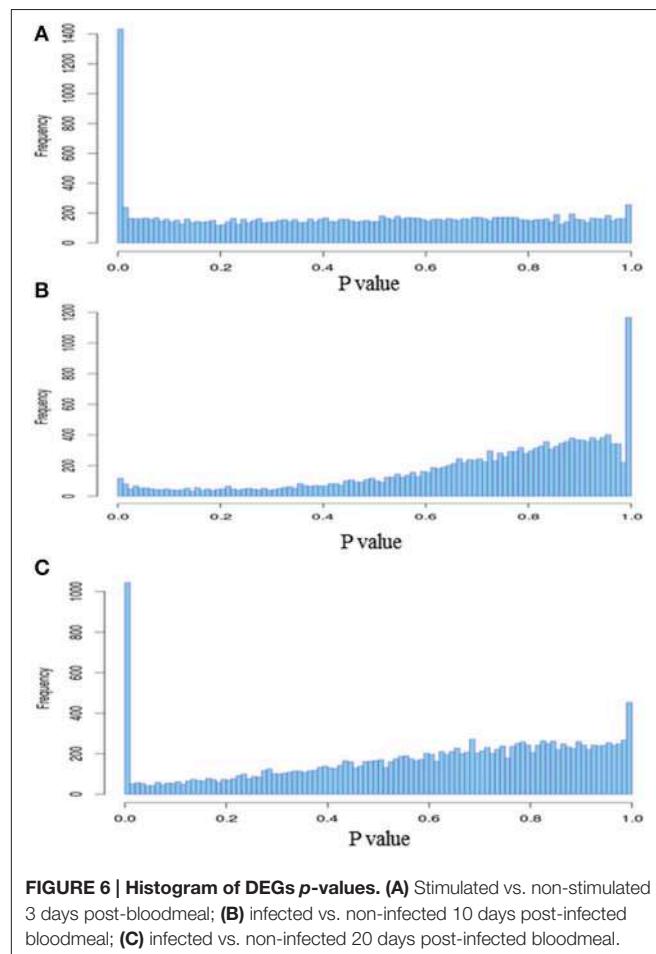
### General Aspects

Deciphering the mechanisms involved in the facilitation of (or refractoriness to) tsetse fly infection by trypanosomes is crucial for developing anti-vector strategies to fight sleeping sickness. In this frame, *G. p. gambiensis* (Gpg) and *T. b. gambiense* (Tbg) transcripts were identified using the RNAseq *de novo* assembly approach. The transcripts were mapped not only on the *G. morsitans morsitans* (Gmm) genome but on a panel of other reference sequences allowing the identification of Gpg genes that may not be represented into the Gmm genome.

The sampling times were chosen according to a previously determined time course of susceptible fly infection by trypanosomes (Van den Abbeele et al., 1999; Ravel et al.,



2003): 3 days post-feeding to target DEGs involved in early events associated with trypanosome entry into the midgut; 10 days post-feeding to target DEGs involved in the establishment of infection; and 20 days post-infected bloodmeal feeding, in

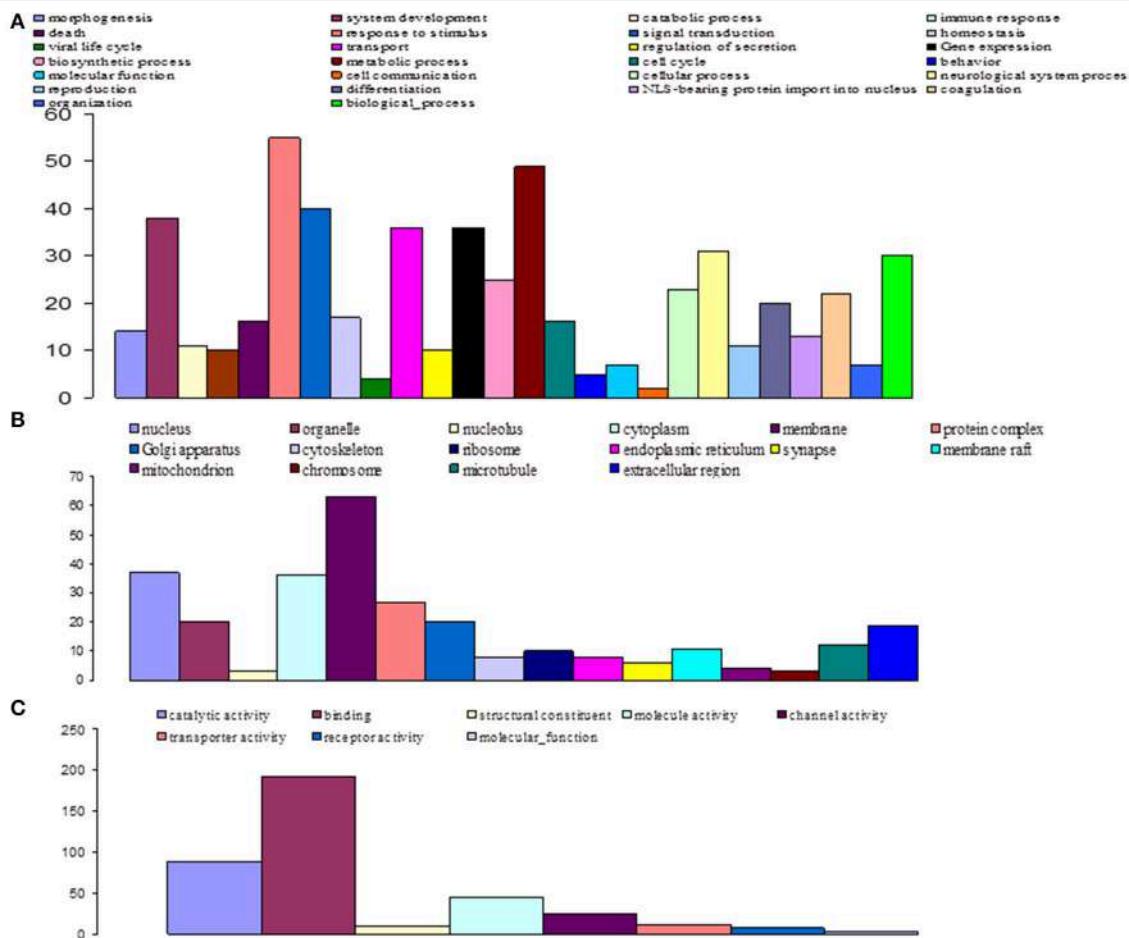


order to target genes involved in events occurring relatively late during trypanosome infection.

A limited number of transposable element sequences (such as protein LTV1 homologs and the viral A-type inclusion protein) were present in the *G. p. gambiensis* midgut transcriptome. This is in contrast to data reported for the *G. m. morsitans* sialotranscriptome (Alves-Silva et al., 2010), and which were obtained following the sequencing of the *G. morsitans* genome (International Glossina Genome Initiative, 2014). As suggested by Alves-Silva et al. (2010), these sequences may represent active transposition, as well as expression of regulatory sequences (Silva et al., 2004).

Numerous DEGs were identified that may be specific to the post-infected bloodmeal time. Nevertheless, some DEGs appear to be of greater interest than others regarding the objectives of the study, and are presented in Table 2.

Although S3, I10, and I20 flies were all fed on trypanosome-infected bloodmeal, we preliminarily observed that trypanosome gene expression was only recorded in S3 and I20 samples, even though the anal drops from I10 flies were positive. This apparent discrepancy is likely due to the attrition phenomenon (Gibson and Bailey, 2003) that occurs several days after trypanosome ingestion, which leads to the elimination of most of the ingested trypanosomes (even in susceptible flies). In contrast, the number



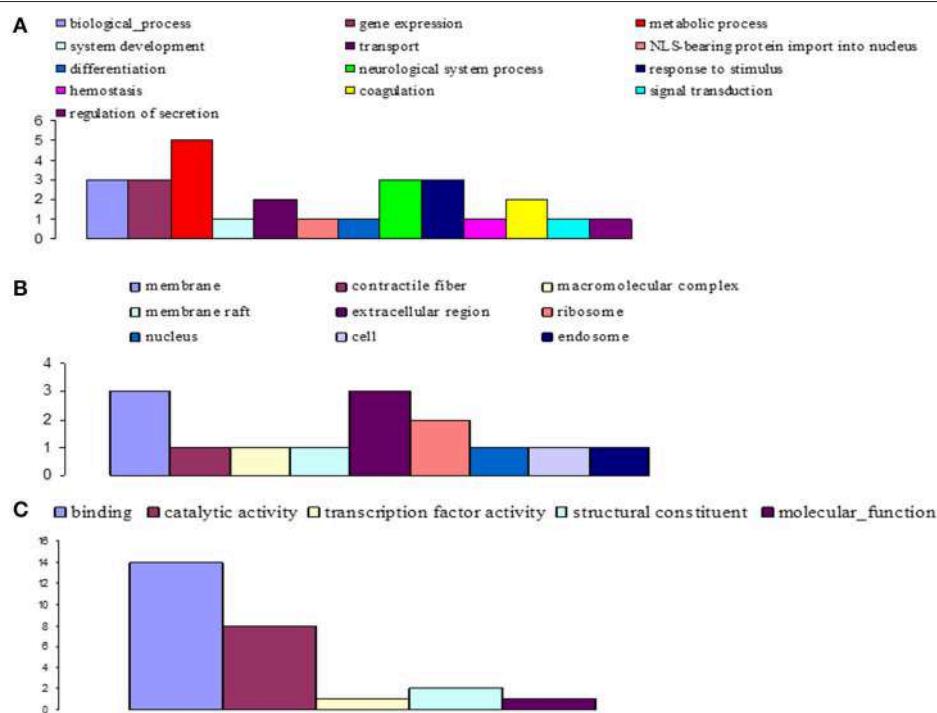
**FIGURE 7 | Distribution of statistically significant contigs for *Glossina palpalis gambiensis*, 3 days post-bloodmeal, based on Gene Ontology (GO) categorization.** The results are presented in three main GO categories: (A) biological process, (B) cellular component, and (C) molecular function. The x-axis indicates the subcategories and the y-axis indicates the number of contigs assigned to a given GO term.

of trypanosomes surviving in I10 flies is probably too low to be detected by an indirect and less sensitive detection method, i.e., recording the transcripts resulting from the expression of some of their genes.

In our experiment, the evolution of trypanosome populations between the day 3 and day 20 sampling points could not be measured (i.e., by trypanosome counting or use of specific DNA probes), since the total midgut extracts were dedicated to total mRNA extraction. Nevertheless, data shown in Table 3 for genes expressed in both S3 and I20 flies support the hypothesis of an increased trypanosome population in I20 vs. S3 (and I10) flies. In fact, the trypanosome expression levels of these genes are 10- to 40-fold higher (depending on the gene) in I20 flies than in S3 flies. However, these differences in gene expression between the different genes also support the idea that gene expression could be differentially stimulated, thus leading to the noted differences in expression levels. Thus, both the increase in the midgut trypanosome population and the modulation in trypanosome gene expression could contribute to the differences in trypanosome gene expression recorded

in I20 flies, as compared to S3 flies. Finally, as expected, no trypanosome expression could be recorded in NI20 flies. This is in agreement with the absence of trypanosome detection in the anal drops of these flies, and confirms their refractoriness to trypanosome infection. Non-infected control flies (NS3) also did not display any trypanosome gene expression (Table 2).

In Table 2, the set of DEGs were classified according to a major function of the proteins they encode. We observed a very high percentage of both tsetse and trypanosome genes encoding a wide array of proteases. Trypanosome genes were only expressed in S3 and I20, whereas tsetse genes were also expressed in the day 10 samples. The high number of trypanosome protease genes identified is in agreement with previous studies of the trypanosome secretome, which characterized a number of proteases suspected to be involved in the trypanosome infective process (Atyame Nten et al., 2010; Geiger et al., 2010). The number of tsetse (but not trypanosome) laccase encoding genes was also surprising. One trypanosome lectin gene was identified and is expressed in S3 and I20 samples, whereas several tsetse lectin genes are only expressed in I10.



**FIGURE 8 | Distribution of statistically significant contigs for *Glossina palpalis gambiensis*, 10 days post-bloodmeal, based on Gene Ontology (GO) categorization.** The results are presented in three main GO categories: (A) biological process, (B) cellular component, and (C) molecular function. The x-axis indicates the subcategories and the y-axis indicates the number of contigs assigned to a given GO term.

As illustrated by our results, several representatives can be listed for a given protein (e.g., laccase, lectin, or serine protease). Each of these representatives (i.e., an isoprotein/isoenzyme) appears to be encoded by a specific gene, suggesting that the isoproteins do not result from posttranscriptional events, and that their expression could be differentially regulated.

## Specific Aspects

Thrombin inhibitor was under-expressed in stimulated flies as compared to flies fed on a non-infected bloodmeal. By contrast, thrombin was over-expressed in infected flies as compared to self-cured flies at 10 and 20 days post-infected bloodmeal. Adult tsetse flies require several molecules that are essential for efficient blood feeding, which counteract the coagulation and blood platelet aggregation responses of the host (Alves-Silva et al., 2010). Thrombin inhibitor may be associated with such anti-clotting activities (Parker and Mant, 1979; Cappello et al., 1998; Alves-Silva et al., 2010). Furthermore, its under-expression early in the midgut invasion process could represent a defense mechanism to immobilize parasites and avoid their dissemination into other tissues.

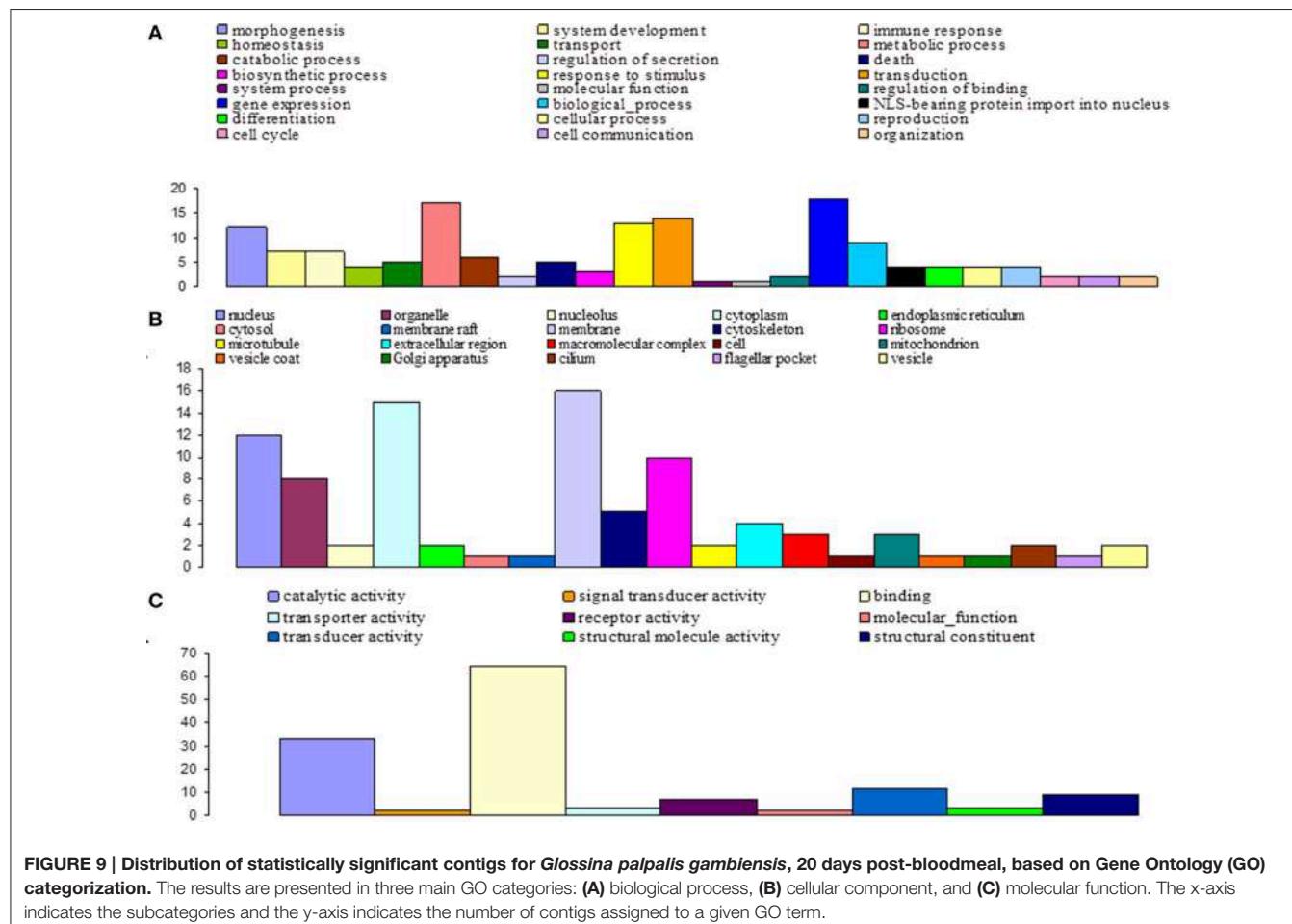
The peritrophic matrix protein three precursor and mucin genes were over-expressed in stimulated flies. *Glossina* possess a peritrophic membrane, continuously built by the proventriculus, which separates the lumen of the midgut from the epithelial cells (Lehane, 1997). It is generally composed of chitin, peritrophin proteins, glycosaminoglycans, and mucin-like molecules (International Glossina Genome Initiative, 2014).

Importantly, the peritrophic membrane is involved in regulating the host immune induction timing, following the parasite challenge (Weiss et al., 2013). Thus, over-expression of these genes in stimulated *G. p. gambiensis* flies could delay the activation of immune gene expression, which would further favor *T. b. gambiense* establishment.

Serine proteases could be involved in such diverse functions as digestion, clotting activity, control of proteolytic cascades in the immunity process, and control of pro-phenoloxidase activation, which causes pathogen melanization (Stark and James, 1998; Kanost, 1999; Alves-Silva et al., 2010). Serine proteases and serpin were previously reported in the *G. m. morsitans* transcriptome (Lehane et al., 2003) and sialotranscriptome (Alves-Silva et al., 2010). Several serine proteases as well as serpin were over-expressed in *G. p. gambiensis*, although they were under-expressed in stimulated or infected flies (depending on the sampling time post-infected bloodmeal ingestion).

Innate immune response products have previously been considered as contributors to fly refractoriness (Hao et al., 2001), and we observed that the antimicrobial peptide cecropin was over-expressed in stimulated flies. There is evidence that innate immune responses, particularly the antimicrobial peptides regulated via the Imd pathway, are among the factors contributing to the tsetse refractoriness to trypanosomes (Hu and Aksoy, 2006).

Lectins display carbohydrate recognition domains associated with innate immunity (Kanost et al., 2004). *S. glossinidius* was previously shown to favor parasite establishment in the insect



**FIGURE 9 | Distribution of statistically significant contigs for *Glossina palpalis gambiensis*, 20 days post-bloodmeal, based on Gene Ontology (GO) categorization.** The results are presented in three main GO categories: (A) biological process, (B) cellular component, and (C) molecular function. The x-axis indicates the subcategories and the y-axis indicates the number of contigs assigned to a given GO term.

midgut through a complex biochemical mechanism involving N-acetyl glucosamine, which would result from pupal chitin hydrolysis by a *S. glossinidius*-produced endochitinase; this product could then inhibit the tsetse midgut lectin otherwise lethal to procyclic forms of the trypanosome (Welburn and Maudlin, 1999). In our experiments, the insect chitinase gene was over-expressed in day 3 stimulated tsetse flies, which is in line with previous hypotheses. However, the lectin gene was over-expressed in infected flies 10 days post-infected bloodmeal ingestion contradicting previous hypotheses. Lectins have also been suggested to display anti-clotting activities (Alves-Silva et al., 2010); thus, the possibility of differential catalytic specificities between the insect and the *Sodalis* chitinase and lectins.

In contrast, the lectizyme gene was over-expressed in flies that cleared the infection 10 days post-infected bloodmeal. This enzyme was previously reported to display lectin (trypanosome agglutination capability) and protease activity involved in the establishment of trypanosomes in tsetse flies (Abubakar et al., 2006).

In our study we identified the presence of both LysM and the putative peptidoglycan-binding domain-containing protein 1-like isoform X1. This protein, homologous to the *C. capitata*

protein, was under-expressed in day 3 stimulated *G. p. gambiensis* flies. Such proteins, closely related to *Drosophila*, have been found in the sialotranscriptome of *G. m. morsitans* (Alves-Silva et al., 2010). A pathogen recognition protein implicated in the initiation of innate defense mechanisms was also previously identified in *G. m. morsitans* fat body (Attardo et al., 2006).

Laccase genes were over-expressed in both stimulated and day 20 infected flies. In *Anopheles gambiae*, laccases have been suggested to oxidize toxic molecules in the bloodmeal, resulting in detoxification or cross-linking of the molecules to the peritrophic matrix, and thus targeting them for excretion (Lang et al., 2012).

For a successful *Glossina* infection, trypanosomes must be transferred from a mammal to an insect host, and therefore they must express specialized proteins to escape multispecies host immune responses (Atyame Nten et al., 2010). Tsetse flies use a proline-alanine shuttle system for energy distribution instead of carbohydrate metabolism (International Glossina Genome Initiative, 2014). Proline is used as a major carbon source during tsetse flight, as well as by trypanosomes (Bursell, 1963). Delta-1-pyrroline-5-carboxylate dehydrogenase is involved in proline catabolism. In all flies, whether stimulated (3-day samples) or infected (10 and 20 days post-infected

**TABLE 2 | DEGs distribution in stimulated vs. non-stimulated (S3 vs. NS3) flies, and in infected vs. non-infected (refractory) flies either 10 days (I10 vs. NI10) or 20 days (I20 vs. NI20) post-bloodmeal.**

Identification	Base mean A (S) or (I)	Base mean B (NS) or (NI)	Fold change NS/S or NI/I	Best hit description
<b>PROTEASES AND PROTEASE INHIBITORS</b>				
GLOS_TB10.61.1870.1.1	552.2672	0.0000	0.000	Aminopeptidase—partial mRNA—Tbb strain 927/4
GLOS_TB10.61.1870.1.1	28.0979	0.0000	0.000	Aminopeptidase—partial mRNA—Tbb strain 927/4
GLOS_TB11.02.0070.1.1	341.6449	0.0000	0.000	Aminopeptidase—partial mRNA—Tbb strain 927/4
GLOS_TB11.02.1070.1.1	64.4924	0.0000	0.000	Aminopeptidase—partial Tbb strain 927/4
GLOS_TB927.3.4750.1.1	789.7420	0.0000	0.000	Aminopeptidase—Tbb strain 927/4
GLOS_TB927.6.1520.1.1	1555.5389	2.2036	0.001	Aquaporin 3 partial mRNA—Tbb strain 927/4
GLOS_TB927.6.1520.1.1	67.2871	0.0000	0.000	Aquaporin 3 partial mRNA—Tbb strain 927/4
GLOS_TB927.3.3410.1.1	337.3614	0.0000	0.000	Aspartyl aminopeptidase—Tbb strain 927/4
GLOS_TB927.3.3410.1.1	29.1621	0.0000	0.000	Aspartyl aminopeptidase—Tbb strain 927/4
GLOS_TB11.47.0036.1.1	440.7008	0.0000	0.000	Calpain, partial mRNA—Tbb strain 927/4
GLOS_TB11.47.0036.1.1	14.0309	0.0000	0.000	Calpain, partial mRNA—Tbb strain 927/4
GLOS_TB927.7.4070.1.1	951.2004	0.0000	0.000	Calpain-like cysteine peptidase—Tbb strain 927/4
GLOS_TB927.7.4070.1.1	50.4615	0.0000	0.000	Calpain-like cysteine peptidase—Tbb strain 927/4
GLOS_TB927.1.2100.1.1	976.7119	1.1018	0.001	Calpain-like cysteine peptidase, cysteine peptidase – Tbb
GLOS_TB927.1.2230.1.1	582.3075	0.0000	0.000	Calpain-like protein fragment—partial mRNA—Tbb
GLOS_LOC101462601.1.5	8394.7979	4561.9074	0.543	Chymotrypsin-1-like [ <i>Ceratitis capitata</i> ]
GLOS_TB927.6.1020.1.1	261.7561	0.0000	0.000	Cysteine peptidase precursor partial mRNA—Tbb strain 927/4
GLOS_TB927.8.3060.1.1	467.3909	0.0000	0.000	Cytosolic leucyl aminopeptidase—partial mRNA—Tbb
GLOS_TB10.389.1480.1.1	801.1016	0.0000	0.000	Cytosolic nonspecific dipeptidase—partial mRNA—Tbb
GLOS_GPL.11.22	28498.3972	40606.4756	1.425	Lectizyme <i>Glossina austeni</i> (lectin and protease activity)
GLOS_GPL.16.22	88429.6303	127653.4331	1.444	Lectizyme <i>Glossina fuscipes fuscipes</i> (lectin and protease activity)
GLOS_LOC101461571.2.2	2911.7794	1267.1640	0.435	Lysosomal aspartic protease-like [ <i>Ceratitis capitata</i> ]
GLOS_TB927.8.1620.1.1	1486.2183	0.0000	0.000	Major surface protease gp63—Tbb strain 927/4
GLOS_TB09.211.4760.1.1	42.3819	0.0000	0.000	Metacaspase 5 partial mRNA (protease arg-lys specific)—Tbb
GLOS_TB11.01.6360.1.1	327.1803	0.0000	0.000	Metalloprotease—partial mRNA—Tbb strain 927/4
GLOS_TB11.01.6360.1.1	34.1576	0.0000	0.000	Metalloprotease—partial mRNA—Tbb strain 927/4
GLOS_TB11.52.0003.1.1	587.8647	0.0000	0.000	Oligopeptidase b—Tbb strain 927/4
GLOS_TB11.52.0003.1.1	14.1393	0.0000	0.000	Oligopeptidase b—Tbb strain 927/4
GLOS_MTRNR21.10.1.3	875.3765	451.0249	0.515	Pan <i>troglodytes</i> MTRNR2-like 10, mRNA
GLOS_LOC101456159.4.4	10357.1870	14556.9484	1.405	Retinoid-inducible serine carboxypeptidase-like [ <i>C. capitata</i> ]
GLOS_TB10.70.7100.1.1	1243.2538	0.9682	0.001	Serine carboxypeptidase III precursor—partial mRNA—Tbb
GLOS_TB10.70.7100.1.1	29.2706	0.0000	0.000	Serine carboxypeptidase III precursor—partial mRNA—Tbb
GLOS_LOC101457953.1.5	467.5594	28.6144	0.061	Serine protease easter-like [ <i>Ceratitis capitata</i> ]
GLOS_LOC101457953.2.5	1048.7914	149.3232	0.142	Serine protease easter-like [ <i>Ceratitis capitata</i> ]
GLOS_LOC101457953.3.5	257.1584	104.7158	0.407	Serine protease easter-like [ <i>Ceratitis capitata</i> ]
GLOS_LOC101457953.5.5	535.7352	31.6210	0.059	Serine protease easter-like [ <i>Ceratitis capitata</i> ]
GLOS_LOC101455604.4.10	4147.7876	8446.3133	2.036	Serine protease SP24D-like [ <i>Ceratitis capitata</i> ]
GLOS_LOC101455604.7.10	7256.5753	4107.4021	0.566	Serine protease SP24D-like [ <i>Ceratitis capitata</i> ]
GLOS_TB10.6K15.3800.1.1	675.8737	0.0000	0.000	Dipeptidyl-peptidase 8-like serine peptidase—Tbb strain 927/4
GLOS_TB927.7.190.1.1	471.9197	0.0000	0.000	Thimet oligopeptidase A—Tbb strain 927/4
GLOS_TB927.7.190.1.1	20.0182	0.0000	0.000	Thimet oligopeptidase A—Tbb strain 927/4
GLOS_LOC101450759.15.17	101392.5605	76275.2604	0.752	Transmembrane protease serine 9-like [ <i>Ceratitis capitata</i> ]
GLOS_TRYDG.4.5	205.3854	100.1940	0.488	Trypsin delta/gamma [ <i>Drosophila melanogaster</i> ]
GLOS_TRYDG.5.5	25628.5380	15491.6912	0.604	Trypsin delta/gamma [ <i>Drosophila erecta</i> ]
GLOS_LOC101463325.1.1	70961.0232	106614.8609	1.502	Trypsin-like [ <i>Ceratitis capitata</i> ]
GLOS_LOC101460475.2.2	622.8445	282.1185	0.453	Venom carboxylesterase-6-like [ <i>Ceratitis capitata</i> ]
GLOS_CBPA1.4.11	81.9328	203.9497	2.489	Zinc carboxypeptidase A 1 [ <i>Drosophila p. pseudoobscura</i> ]
GLOS_CBPA1.11.11	12571.5017	25447.6241	2.024	Zinc carboxypeptidase A [ <i>Drosophila p. pseudoobscura</i> ]

(Continued)

**TABLE 2 | Continued**

Identification	Base mean A (S) or (I)	Base mean B (NS) or (NI)	Fold change NS/S or NI/I	Best hit description
GLOS_LOC101459622.1.22	1301.8232	2242.4272	1.723	zinc metalloproteinase nas-4-like [ <i>Ceratitis capitata</i> ]
GLOS_TB927.8.6450.1.1	439.0594	0.0000	0.000	Inhibitor of cysteine peptidase—Tbb strain 927/4
GLOS_LOC101459846.1.2	2433.4913	1664.4399	0.684	Alaserpin-like (Serin protease inhibitor) [ <i>Ceratitis capitata</i> ]
GLOS_LOC101459846.2.2	1420.6193	838.5583	0.590	Alaserpin-like (Serin protease inhibitor) [ <i>Ceratitis capitata</i> ]
<b>OXYDASES AND DEHYDROGENASES</b>				
GLOS_LOC101457181.1.5	73109.9421	21931.6983	0.300	Laccase-2-like [ <i>Ceratitis capitata</i> ]
GLOS_LOC101457181.2.5	17482.0138	5829.7342	0.333	Laccase-2-like [ <i>Ceratitis capitata</i> ]
GLOS_LOC101457181.3.5	1802.2092	1078.3896	0.598	Laccase-2-like [ <i>Ceratitis capitata</i> ]
GLOS_LOC101457181.3.5	4844.5886	1509.7455	0.312	Laccase-2-like [ <i>Ceratitis capitata</i> ]
GLOS_LOC101457181.4.5	25283.1929	8470.9491	0.335	Laccase-2-like [ <i>Ceratitis capitata</i> ]
GLOS_LOC101457181.5.5	11790.1166	7254.7118	0.615	Laccase-2-like [ <i>Ceratitis capitata</i> ]
GLOS_LOC101457181.5.5	34357.2827	8198.3799	0.239	Laccase-2-like [ <i>Ceratitis capitata</i> ]
GLOS_TB927.7.210.1.1	411.4912	0.0000	0.000	Proline oxidase—Tbb strain 927/4
GLOS_TB10.70.4280.1.1	841.4256	0.0000	0.000	Delta-1-pyrroline-5-carboxylate dehydrogenase [ <i>T. brucei</i> ]
GLOS_TB10.70.4280.1.1	25.2307	0.0000	0.000	Delta-1-pyrroline-5-carboxylate dehydrogenase [ <i>T. brucei</i> ]
GLOS_TB11.02.1990.1.1	319.6615	0.0000	0.000	Ferric reductase—Tbb strain 927/4
GLOS_ND1.2.4	9129.9740	5081.8960	0.557	NADH dehydrogenase subunit 1 (mit.) [ <i>C. megacephala</i> ]
GLOS_ND3.1.1	1793.2711	869.6530	0.485	NADH dehydrogenase subunit 3 (mitochondrion) [ <i>P. utilis</i> ]
GLOS_ND4.5.6	12286.9040	7978.0201	0.649	NADH dehydrogenase subunit 4 (mitochondrion) [ <i>C. bezziana</i> ]
<b>LECTINS</b>				
GLOS_TB11.02.1680.1.1	456.9371	0.0000	0.000	Lectin—partial mRNA—Tbb strain 927/4
GLOS_TB11.02.1680.1.1	29.1621	0.0000	0.000	Lectin—partial mRNA—Tbb strain 927/4
GLOS_LECA.10.13	3842.8084	1551.5134	0.404	Lectin subunit alpha [ <i>Sarcophaga peregrina</i> ]
GLOS_LECA.3.13	8003.5218	3988.4579	0.498	Lectin subunit alpha [ <i>Sarcophaga peregrina</i> ]
GLOS_LECA.4.13	1427.7877	656.6802	0.460	Lectin subunit alpha [ <i>Sarcophaga peregrina</i> ]
GLOS_LECA.5.13	2385.9247	1113.9812	0.467	Lectin subunit alpha [ <i>Sarcophaga peregrina</i> ]
GLOS_LECA.6.13	1794.0384	740.5790	0.413	Lectin subunit alpha [ <i>Sarcophaga peregrina</i> ]
GLOS_LECA.9.13	3360.6422	1983.1933	0.590	Lectin subunit alpha [ <i>Sarcophaga peregrina</i> ]
GLOS_GPL.11.22	28498.3972	40606.4756	1.425	Lectizyme [ <i>Glossina austeni</i> ] (lectin and protease activity)
GLOS_GPL.16.22	88429.6303	127653.4331	1.444	Lectizyme [ <i>G. fuscipes fuscipes</i> ] (lectin and protease activity)
<b>HYDROLASES</b>				
GLOS_LOC101452734.1.1	299.0185	126.9519	0.425	Alpha-N-acetylgalactosaminidase-like [ <i>Ceratitis capitata</i> ]
GLOS_LOC101462140.1.1	5318.9872	2975.0751	0.559	Chitinase 3-like [ <i>Ceratitis capitata</i> ]
GLOS_FDL.1.2	34480.1169	48644.5776	1.411	Probable beta-hexosaminidase [ <i>D. melanogaster</i> ]
GLOS_TB927.7.6850.1.1	1626.9116	0.0000	0.000	Trans-sialidase partial mRNA—Tbb strain 927/4
GLOS_TB927.7.6850.1.1	102.4366	0.0000	0.000	Trans-sialidase partial mRNA—Tbb strain 927/4
<b>INHIBITORS</b>				
GLOS_LOC101459846.1.2	2433.4913	1664.4399	0.684	Alaserpin-like (Serine protease inhibitor) [ <i>C. capitata</i> ]
GLOS_LOC101459846.2.2	1420.6193	838.5583	0.590	Alaserpin-like (Serine protease inhibitor) [ <i>C. capitata</i> ]
GLOS_TTI.10.16	988.9930	513.9447	0.520	Thrombin inhibitor (tsetse) [ <i>G. morsitans morsitans</i> ]
GLOS_TTI.3.16	55.7102	131.7672	2.365	Thrombin inhibitor (tsetse) [ <i>G. morsitans morsitans</i> ]
GLOS_TTI.4.16	9152.8285	3778.1122	0.413	Thrombin inhibitor (tsetse) [ <i>G. morsitans morsitans</i> ]
GLOS_TTI.1.16	1304.6401	185.2590	0.142	Thrombin inhibitor (tsetse) [ <i>G. morsitans morsitans</i> ]
<b>OTHER FUNCTIONS</b>				
GLOS_LOC101454505.3.12	8857.1384	5131.7479	0.579	Acyl-CoA-binding protein homolog isoform X1 [ <i>Ceratitis capitata</i> ]
GLOS_LOC101461142.7.9	6823.2923	10238.5963	1.501	Acyl-CoA Delta(11) desaturase-like isoform X3 [ <i>Ceratitis capitata</i> ]
GLOS_TB927.8.7410.1.1	49.3250	0.0000	0.000	Calreticulin, putative (signaling)—partial mRNA—Tbb strain 927/4
GLOS_TB927.8.7410.1.1	1024.9149	1.1018	0.001	Calreticulin, putative (signaling)—partial mRNA—Tbb strain 927/4
GLOS_CEC.2.2	761.3134	160.3734	0.211	Cecropin G.m.m. (antimicrobial peptide)

(Continued)

**TABLE 2 | Continued**

Identification	Base mean A	Base mean B	Fold change	Best hit description
	(S) or (I)	(NS) or (NI)	NS/S or NI/I	
GLOS_CECC.1.1	679.9491	104.6283	0.154	Cecropin-C <i>D. yakuba</i> (antimicrobial peptide)
GLOS_CC2H.1.1	321.8132	0.0000	0.000	Cell division control protein 2 homolog—Tbb
GLOS_TB10.26.0510.1.1	860.7592	0.0000	0.000	CYC2-like cyclin—partial mRNA—Tbb strain 927/4
GLOS_HYPB.1.2	4304.3151	8966.5053	2.083	Hypodermin—[ <i>Hypoderma lineatum</i> ]
GLOS_GST.1.1	1624.8670	583.4978	0.359	Glutathione S-transferase [ <i>Musca domestica</i> ]
GLOS_TB927.7.3980.1.1	380.9405	1.1018	0.003	Immunodominant antigen—partial mRNA—Tbb
GLOS_TB09.211.4513.1.1	49.1803	0.0000	0.000	Kinetoplastid membrane protein KMP-11—Tbb strain 927/4
GLOS_TB927.3.3580.1.1	1112.2710	1.1018	0.001	Lipophosphoglycan biosynthetic protein—partial mRNA—Tbb
GLOS_TB927.3.3580.1.1	43.0844	0.0000	0.000	Lipophosphoglycan biosynthetic protein—partial mRNA—Tbb
GLOS_LOC101449088.1.1	3335.7392	6030.1811	1.808	LysM and peptidoglycan-binding domain-containing protein 1-like isoform X1
GLOS_LOC101460827.1.2	2004.2841	643.2004	0.321	Mucin-5AC-like (glycosylated protein)[ <i>Ceratitis capitata</i> ]
GLOS_TB927.8.2160.1.1	24.3474	0.0000	0.000	Multidrug resistance protein A—partial mRNA—Tbb strain 927/4
GLOS_TB11.03.0140.1.1	448.9753	0.0000	0.000	Nucleoporin—partial mRNA—Tbb strain 927/4
GLOS_PMP3.2.2	20992.2730	13567.5959	0.646	Peritrophic matrix protein 3 precursor [ <i>Tribolium castaneum</i> ]
GLOS_LOC101453268.1.1	733.2798	1317.6387	1.797	Period circadian protein-like [ <i>Ceratitis capitata</i> ]
GLOS_LOC101456033.1.1	1049.0807	701.7895	0.669	Platelet binding protein GspB-like [ <i>Ceratitis capitata</i> ]
GLOS_LOC101456285.1.1	32755.6635	60800.4454	1.856	Protein FAM188A homolog [ <i>Ceratitis capitata</i> ]
GLOS_LTV1.1.1	1777.9362	2547.2329	1.433	Protein LTV1 homolog—[ <i>Drosophila melanogaster</i> ]
GLOS_PMAR_PMAR029216.1.2	1356.3842	2415.7386	1.781	Protein conserved hypothetical[ <i>Perkinsus marinus</i> ATCC 50983]
GLOS_LOC101461327.1.1	631.4029	1439.7920	2.280	Protein Uncharacterized LOC101461327 [ <i>C. capitata</i> ]
GLOS_LOC101450586.2.9	7835.2507	4885.1378	0.623	Protein uncharacterized—LOC101450586 [ <i>Ceratitis capitata</i> ]
GLOS_LOC101462034.1.2	1260.9956	595.6506	0.472	Protein uncharacterized—LOC101462034 [ <i>Ceratitis capitata</i> ]
GLOS_TB11.01.5310.1.1	638.3697	0.9682	0.002	Receptor-type adenylate cyclase—partial mRNA—Tbb
GLOS_RS25.5.10	44.5773	300.9004	6.750	Ribosomal protein (40S) [ <i>Drosophila melanogaster</i> ]
GLOS_RL31.2.3	5894.0883	3872.9890	0.657	Ribosomal protein (60S) [ <i>Drosophila melanogaster</i> ]
GLOS_LOC101450592.2.2	381.2922	201.2523	0.528	Sialin-like (transporter) [ <i>Ceratitis capitata</i> ]
GLOS_LOC101461145.1.1	623.4847	291.2207	0.467	Stress response protein NST1-like [ <i>Ceratitis capitata</i> ]
GLOS_TB927.5.2940.1.1	455.0980	0.0000	0.000	Stress induced protein sti1 partial mRNA—Tbb strain 927/4
GLOS_TRF.1.1	5009.4602	116.1075	0.023	Transferrin—[ <i>Sarcophaga peregrina</i> ]
GLOS_TB927.8.6750.1.1	483.9673	0.0000	0.000	Translationally controlled tumor protein (TCTP)—Tbb strain 927/4
GLOS_TB927.8.6750.1.1	35.2218	0.0000	0.000	Translationally controlled tumor protein (TCTP)—Tbb strain 927/4
GLOS_MLC_9020.1.1	318.8201	1124.6933	3.528	Transmembrane protein [ <i>Mycoplasma mycooides</i> subsp. Capri]
GLOS_TVAG_198570.1.2	1670.6333	838.7765	0.502	Viral A-type inclusion protein [ <i>Trichomonas vaginalis</i> G3]

DEGs are sorted within each section with reference to the proteins that they encode (listed alphabetically). The gene sets are a part of those presented in Supplementary Tables S2 (day 3 samples), S3 (day 10 samples), and S4 (day 20 samples).

Fonts colors indicate genes differentially expressed between tsetse flies sampled at day 3, day 10, and day 20 post-fly feeding.

The text "GLOS\_TB ..." refers to Trypanosoma genes; all others are Glossina genes.

For day three samples, S, refers to flies that ingested a trypanosome-infected bloodmeal; NS, refers to flies that ingested a non-infected blood meal.

"I10" and "NI10" are 10 day samples; "I20" and "NI20" are 20 day samples. I, susceptible flies, which received an infected bloodmeal and became infected. NS, refractory flies that were self-cured following an infected bloodmeal.

bloodmeal), the gene encoding this dehydrogenase was over-expressed. This gene was found to be homologous with that of *T. brucei*.

Several different peptidase families were shown in the present study to be expressed by trypanosomes, including members of the serine and cysteine proteinases, and metallopeptidases. These peptidases could: (a) act as virulence factors thus favoring parasite invasion and growth in the host environment; (b) allow trypanosomes to evade the host immune defenses; (c) produce nutrients by hydrolyzing host proteins (Atyame Nten et al., 2010); and (d) be involved in the blood clotting, thus in immobilizing

invading parasites (Lehane et al., 2003; Alves-Silva et al., 2010; International Glossina Genome Initiative, 2014).

Proteins involved in signaling were also identified in the secretome of procyclic trypanosomes (Atyame Nten et al., 2010). Several of these proteins, such as calreticulin, could play physiopathological roles. In the present study, transcripts corresponding to these proteins were found expressed by trypanosomes.

Transferrin has also been demonstrated as an important part of the immune system in insects and vertebrates (Nichol et al., 2002; Guz et al., 2007). Up-regulation of its transcription

**TABLE 3 | Comparison of gene expression levels differentially expressed in both stimulated (S3) and infected (I20) flies.**

Identification	Base mean S3	Base mean I20	I20/S3	Best hit description
GLOS_TB10.61.1870.1.1	28.0979	552.2672	19.655	Aminopeptidase—partial mRNA—Tbb strain 927/4
GLOS_TB927.6.1520.1.1	67.2871	1555.5389	23.118	Aquaporin three partial mRNA—Tbb strain 927/4
GLOS_TB927.3.3410.1.1	29.1621	337.3614	11.568	Aspartyl aminopeptidase—Tbb strain 927/4
GLOS_TB11.47.0036.1.1	14.0309	440.7008	31.409	Calpain, partial mRNA—Tbb strain 927/4
GLOS_TB927.7.4070.1.1	50.4615	951.2004	18.850	Calpain-like cysteine peptidase—Tbb strain 927/4
GLOS_TB11.01.6360.1.1	34.1576	327.1803	9.579	Metalloprotease—partial mRNA—Tbb strain 927/4
GLOS_TB11.52.0003.1.1	14.1393	587.8647	41.577	Oligopeptidase b—Tbb strain 927/4
GLOS_TB10.70.7100.1.1	29.2706	1243.2538	42.475	Serine carboxypeptidase III precursor—partial mRNA—Tbb
GLOS_TB927.7.190.1.1	20.0182	471.9197	23.574	Thimet oligopeptidase A—Tbb strain 927/4
GLOS_LOC101457181.3.5	4844.5886	1802.2092	0.372	Laccase-2-like [ <i>Ceratitis capitata</i> ]
GLOS_LOC101457181.5.5	34357.2827	11790.1166	0.343	Laccase-2-like [ <i>Ceratitis capitata</i> ]
GLOS_TB10.70.4280.1.1	25.2307	841.4256	33.349	Delta-1-pyrroline-5-carboxylate dehydrogenase [ <i>T. brucei</i> ]
GLOS_TB11.02.1680.1.1	29.1621	456.9371	15.669	Lectin—partial mRNA—Tbb strain 927/4
GLOS_TB927.7.6850.1.1	102.4366	1626.9119	15.882	Trans-sialidase partial mRNA—Tbb strain 927/4
GLOS_TB927.3.3580.1.1	43.0844	1112.2710	25.816	Lipophosphoglycan biosynthetic protein, putative—mRNA—Tbb
GLOS_TB927.8.6750.1.1	35.2218	483.9673	13.741	Translationally controlled tumor protein—Tbb strain 927/4
GLOS_TB927.8.7410.1.1	49.3250	1024.9149	20.779	Tbb strain 927/4 calreticulin, putative—partial mRNA

This list of genes is extracted from the genes presented in **Table 2**.

I20/S3 indicates the fold change in gene expression levels in I20 vs. S3 samples.

following an immune challenge is reported for a number of insects including *Aedes aegypti*, *Bombyx mori*, and *Drosophila* (Yoshiga et al., 1997, 1999; Yun et al., 1999). In the present study, tsetse flies stimulated by trypanosomes (S-3days) were observed to over-express the transferrin gene. These results are in agreement with Guz et al. (2007), who reported an increase in transferrin expression levels upon microbial challenge in tsetse flies.

As previously reported (Hamidou Soumana et al., 2014b), genes encoding ribosomal proteins can also be differentially expressed (either up-regulated or down-regulated). For instance, in NI10 samples (i.e., self-cured flies) a tsetse fly 40S ribosomal protein gene was 6.75-fold over-expressed. In contrast, a 60S ribosomal protein gene was under-expressed as compared to the expression levels of the same genes recorded in I10 samples ( $NI10/I10 = 0.657$ ). Wang et al. (2013) reported similar results for phosphate- or iron-deficient *Arabidopsis* roots vs. control roots, in response to a changing environment. These findings raise the question of whether modulations of ribosomal protein gene expression could also be involved in tsetse fly adaptation to the stress of trypanosome invasion.

We identified over 195,464 high confidence SNPs from 14,929 contigs across the whole transcriptome assemblies of *G. p. gambiensis* and *T. b. gambiense*. These SNPs overlap genes that exhibit both up- and down-regulation of homologous transcripts from different insect and parasite species. The SNP genetic sites identified in our dataset will provide useful marker resources for fine mapping experiments and marker-assisted *G. p. gambiensis* control programs.

These results will have immediate applications for exploring *G. p. gambiensis* genome diversity and co-expression networks involved in tsetse infection by trypanosomes, as well as the development of stochastic and metabolic networks. In addition,

these resources can be used to identify novel genes, transcript models and eQTLs, and to study trypanosome adaptation to diverse fly tissue environments. These findings will also be useful when undertaking comparative studies with the *G. morsitans* transcriptome (Attardo et al., 2006; Alves-Silva et al., 2010) and genome (International Glossina Genome Initiative, 2014).

To conclude: our study is the first to investigate key steps of tsetse fly infection by trypanosomes through characterization of the *G. p. gambiensis* transcriptome and the complete set of tsetse fly and trypanosome DEGs. This approach revealed genes that interact in well-defined patterns and the various characterized DEGs provide insights into the complexity of the host-parasite interactions. Future investigations should aim to characterize the involvement of identified genes in tsetse refractoriness to trypanosome infection.

## ACKNOWLEDGMENTS

The authors thank the “Région Languedoc-Roussillon—Appel d’Offre Chercheur d’Avenir 2011,” the “Service de Coopération et d’Action Culturelle de l’Ambassade de France au Niger” and the “Institut de Recherche pour le Développement” for their financial support. IH is a PhD student supported by the French Embassy to Niger, Service de Coopération et d’Action Culturelle (SCAC). We acknowledge the support of Laboratoire d’Excellence (Labex) Parafrap N°ANR-11-LABX-0024.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2015.01259>

## REFERENCES

- Abd-Alla, A. M., Bergoin, M., Parker, A. G., Maniania, N. K., Vlak, J. M., Bourtzis, K., et al. (2013). Improving sterile insect technique (SIT) for tsetse flies through research on their symbionts and pathogens. *J. Invertebr. Pathol.* 112(Suppl.), S2–S10. doi: 10.1016/j.jip.2012.07.009
- Abubakar, L. U., Bulimo, W. D., Mulaa, F. J., and Osir, E. O. (2006). Molecular characterization of a tsetse fly midgut proteolytic lectin that mediates differentiation of African trypanosomes. *Insect Biochem. Mol. Biol.* 36, 344–352. doi: 10.1016/j.ibmb.2006.01.010
- Aksoy, S., Weiss, B., and Attardo, G. (2008). Paratransgenesis applied for control of tsetse transmitted sleeping sickness. *Adv. Exp. Med. Biol.* 627, 35–48. doi: 10.1007/978-0-387-78225-6\_3
- Alves-Silva, J., Ribeiro, J. M., Van Den Abbeele, J., Attardo, G., Hao, Z., Haines, L. R., et al. (2010). An insight into the sialome of *Glossina morsitans morsitans*. *BMC Genomics* 11:213. doi: 10.1186/1471-2164-11-213
- Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11, R10. doi: 10.1186/gb-2010-11-10-r106
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556
- Attardo, G. M., Strickler-Dinglasan, P., Perkin, S. A., Caler, E., Bonaldo, M. F., Soares, M. B., et al. (2006). Analysis of fat body transcriptome from the adult tsetse fly, *Glossina morsitans morsitans*. *Insect Mol. Biol.* 15, 411–424. doi: 10.1111/j.1365-2583.2006.00649.x
- Atyame Nten, C. M., Sommerer, N., Rofidal, V., Hirtz, C., Rossignol, M., Cuny, G., et al. (2010). Excreted/secreted proteins from trypanosome procyclic strains. *J. Biomed. Biotechnol.* 2010:212817. doi: 10.1155/2010/212817
- Baker, N., de Koning, H. P., Mäser, P., and Horn, D. (2013). Drug resistance in African trypanosomiasis: the melarsoprol and pentamidine story. *Trends Parasitol.* 29, 110–118. doi: 10.1016/j.pt.2012.12.005
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* 57, 289–300.
- Brun, R., Blum, J., Chappuis, F., and Burri, C. (2010). Human African trypanosomiasis. *Lancet* 375, 148–159. doi: 10.1016/S0140-6736(09)60829-1
- Bursell, E. (1963). Aspects of the metabolism of amino acids in the tsetse fly, *Glossina* (Diptera). *J. Insect Physiol.* 9, 439–452. doi: 10.1016/0022-1910(63)90054-4
- Cappello, M., Li, S., Chen, X., Li, C. B., Harrison, L., Narashimhan, S., et al. (1998). Tsetse thrombin inhibitor: bloodmeal-induced expression of an anticoagulant in salivary glands and gut tissue of *Glossina morsitans morsitans*. *Proc. Natl. Acad. Sci. U.S.A.* 95, 14290–14295. doi: 10.1073/pnas.95.24.14290
- Cock, P. J. A., Fields, C. J., Goto, N., Heuer, M. L., and Rice, P. M. (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res.* 38, 1767–1771. doi: 10.1093/nar/gkp1137
- Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676. doi: 10.1093/bioinformatics/bti610
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 15, 491–498. doi: 10.1038/ng.806
- Dukes, P., Kaukas, A., Hudson, K. M., Asonganyi, T., and Gashumba, J. K. (1989). A new method for isolating *Trypanosoma brucei gambiense* from sleeping sickness patients. *Trans. R. Soc. Trop. Med. Hyg.* 83, 636–639. doi: 10.1016/0035-9203(89)90379-9
- Farikou, O., Njiokou, F., Mbida Mbida, J. A., Njitchouang, G. R., Djemba, H. N., Asonganyi, T., et al. (2010). Tripartite interactions between tsetse flies, *Sodalis glossinidius* and trypanosomes—an epidemiological approach in two historical human African trypanosomiasis foci in Cameroon. *Infect. Genet. Evol.* 10, 115–121. doi: 10.1016/j.meegid.2009.10.008
- Franco, J. R., Simarro, P. P., Diarra, A., and Jannin, J. G. (2014). Epidemiology of human African trypanosomiasis. *Clin. Epidemiol.* 6, 257–275. doi: 10.2147/CLEP.S39728
- Frézil, J. L., and Cuisance, D. (1994). Trypanosomiasis, diseases with future: prospects and uncertainty. *Bull. Soc. Pathol. Exot.* 87, 391–393.
- Geiger, A., Fardeau, M. L., Njiokou, F., and Ollivier, B. (2013). *Glossina* spp. gut bacterial flora and their putative role in fly-hosted trypanosome development. *Front. Cell Infect. Microbiol.* 3:34. doi: 10.3389/fcimb.2013.00034
- Geiger, A., Hirtz, C., Bécue, T., Bellard, E., Centeno, D., Gargani, D., et al. (2010). Exocytosis and protein secretion in *Trypanosoma*. *BMC Microbiol.* 10:20. doi: 10.1186/1471-2180-10-20
- Geiger, A., Ponton, F., and Simo, G. (2015). Adult blood-feeding tsetse flies, trypanosomes, microbiota and the fluctuating environment in sub-Saharan Africa. *ISME J.* 9, 1496–1507. doi: 10.1038/ismej.2014.236
- Geiger, A., Ravel, S., Mateille, T., Janelle, J., Patrel, D., Cuny, G., et al. (2007). Vector competence of *Glossina palpalis gambiensis* for *Trypanosoma brucei* s.l. and genetic diversity of the symbiont *Sodalis glossinidius*. *Mol. Biol. Evol.* 24, 102–109. doi: 10.1093/molbev/msl135
- Geiger, A., Simo, G., Grébaut, P., Peltier, J. B., Cuny, G., and Holzmüller, P. (2011). Transcriptomics and proteomics in human African trypanosomiasis: current status and perspectives. *J. Proteomics* 74, 1625–1643. doi: 10.1016/j.jprot.2011.01.016
- Gibson, W., and Bailey, M. (2003). The development of *Trypanosoma brucei* within the tsetse fly midgut observed using green fluorescent trypanosomes. *Kinetoplastid Biol. Dis.* 2:1. doi: 10.1186/1475-9292-2-1
- Guz, N., Attardo, G. M., Wu, Y., and Aksoy, S. (2007). Molecular aspects of transferrin expression in the tsetse fly (*Glossina morsitans morsitans*). *J. Insect Physiol.* 53, 715–723. doi: 10.1016/j.jinsphys.2007.03.013
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi: 10.1038/nprot.2013.084
- Haines, L. R., Lehane, S. M., Pearson, T. W., and Lehane, M. J. (2010). Tsetse EP protein protects the fly midgut from trypanosome establishment. *PLoS Pathog.* 6:e1000793. doi: 10.1371/journal.ppat.100793
- Hamidou Soumana, I., Loriod, B., Ravel, S., Tchicaya, B., Simo, G., Rihet, P., et al. (2014a). The transcriptional signatures of *Sodalis glossinidius* in the *Glossina palpalis gambiensis* flies negative for *Trypanosoma brucei gambiense* contrast with those of this symbiont in tsetse flies positive for the parasite: possible involvement of a *Sodalis*-hosted prophage in fly *Trypanosoma* refractoriness? *Infect. Genet. Evol.* 24, 41–56. doi: 10.1016/j.meegid.2014.03.005
- Hamidou Soumana, I., Simo, G., Njiokou, F., Tchicaya, B., Abd-Alla, A. M., Cuny, G., et al. (2013). The bacterial flora of tsetse fly midgut and its effect on trypanosome transmission. *J. Invertebr. Pathol.* 112(Suppl.), S89–S93. doi: 10.1016/j.jip.2012.03.029
- Hamidou Soumana, I., Tchicaya, B., Chuchana, P., and Geiger, A. (2014c). Midgut expression of immune-related genes in *Glossina palpalis gambiensis* challenged with *Trypanosoma brucei gambiense*. *Front. Microbiol.* 5:609. doi: 10.3389/fmicb.2014.00609
- Hamidou Soumana, I., Tchicaya, B., Simo, G., and Geiger, A. (2014b). Comparative gene expression of *Wigglesworthia* inhabiting non-infected and *Trypanosoma brucei gambiense*-infected *Glossina palpalis gambiensis* flies. *Front. Microbiol.* 5:620. doi: 10.3389/fmicb.2014.00620
- Hao, Z., Kasumba, I., and Aksoy, S. (2003). Proventriculus (cardia) plays a crucial role in immunity in tsetse fly (Diptera: *Glossinidae*). *Insect Biochem. Mol. Biol.* 33, 1155–1164. doi: 10.1016/j.ibmb.2003.07.001
- Hao, Z., Kasumba, I., Lehane, M. J., Gibson, W. C., Kwon, J., and Aksoy, S. (2001). Tsetse immune responses and trypanosome transmission: implications for the development of tsetse-based strategies to reduce trypanosomiasis. *Proc. Natl. Acad. Sci. U.S.A.* 98, 12648–12653. doi: 10.1073/pnas.221363798
- Hotez, P. J., Fenwick, A., Savioli, L., and Molyneux, D. H. (2009). Rescuing the bottom billion through control of neglected tropical diseases. *Lancet* 373, 1570–1575. doi: 10.1016/S0140-6736(09)60233-6
- Hu, C., and Aksoy, S. (2006). Innate immune responses regulate trypanosome parasite infection of the tsetse fly *Glossina morsitans morsitans*. *Mol. Microbiol.* 60, 1194–1204. doi: 10.1111/j.1365-2958.2006.05180.x

- International Glossina Genome Initiative (2014). Genome sequence of the tsetse fly (*Glossina morsitans*): vector of African trypanosomiasis. *Science* 344, 380–386. doi: 10.1126/science.1249656
- Jamonneau, V., Ravel, S., Koffi, M., Kaba, D., Zeze, D. G., Ndri, L., et al. (2004). Mixed infections of trypanosomes in tsetse and pigs and their epidemiological significance in a sleeping sickness focus of Côte d'Ivoire. *Parasitology* 129, 693–702. doi: 10.1017/S0031182004005876
- Kanost, M. R. (1999). Serine proteinase inhibitors in arthropod immunity. *Dev. Comp. Immunol.* 23, 291–301. doi: 10.1016/S0145-305X(99)00012-9
- Kanost, M. R., Jiang, H., and Yu, X. Q. (2004). Innate immune responses of a lepidopteran insect, *Manduca sexta*. *Immunol. Rev.* 198, 97–105. doi: 10.1111/j.0105-2896.2004.0121.x
- Kennedy, P. G. (2008). The continuing problem of human African trypanosomiasis (sleeping sickness). *Ann. Neurol.* 64, 116–126. doi: 10.1002/ana.21429
- Lang, M., Kanost, M. R., and Gorman, M. J. (2012). Multicopper oxidase-3 is a laccase associated with the peritrophic matrix of *Anopheles gambiae*. *PLoS ONE* 7:e33985. doi: 10.1371/journal.pone.0033985
- Lehane, M. J. (1997). Peritrophic matrix structure and function. *Annu. Rev. Entomol.* 42, 525–550. doi: 10.1146/annurev.ento.42.1.525
- Lehane, M. J., Aksoy, S., Gibson, W., Kerhennou, A., Berriman, M., Hamilton, J., et al. (2003). Adult midgut expressed sequence tags from the tsetse fly *Glossina morsitans morsitans* and expression analysis of putative immune response genes. *Genome Biol.* 4:R63. doi: 10.1186/gb-2003-4-10-r63
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. doi: 10.1093/bioinformatics/btl158
- Macleod, E. T., Darby, A. C., Maudlin, I., and Welburn, S. C. (2007b). Factors affecting trypanosome maturation in tsetse flies. *PLoS ONE* 2:e239. doi: 10.1371/journal.pone.0000239
- Macleod, E. T., Maudlin, I., Darby, A. C., and Welburn, S. C. (2007a). Antioxidants promote establishment of trypanosome infections in tsetse. *Parasitology* 134, 827–831. doi: 10.1017/S0031182007002247
- Maudlin, I., and Welburn, S. C. (1994). Maturation of trypanosome infections in tsetse. *Exp. Parasitol.* 79, 202–205. doi: 10.1006/expa.1994.1081
- Medlock, J., Atkins, K. E., Thomas, D. N., Aksoy, S., and Galvani, A. P. (2013). Evaluating paratransgenesis as a potential control strategy for African trypanosomiasis. *PLoS Negl. Trop. Dis.* 7:e2374. doi: 10.1371/journal.pntd.0002374
- Moloo, S. K., Asonganyi, T., and Jenni, L. (1986). Cyclical development of *Trypanosoma brucei gambiense* from cattle and goats in *Glossina*. *Acta Trop.* 43, 407–408.
- Moser, D. R., Cook, G. A., Ochs, D. E., Bailey, C. P., McKane, M. R., and Donelson, J. E. (1989). Detection of *Trypanosoma congolense* and *Trypanosoma brucei* subspecies by DNA amplification using the polymerase chain reaction. *Parasitology* 99, 57–66. doi: 10.1017/S0031182000061023
- Nayduch, D., and Aksoy, S. (2007). Refractoriness in tsetse flies (*Diptera: Glossinidae*) may be a matter of timing. *J. Med. Entomol.* 44, 660–665. doi: 10.1093/jmedent/44.4.660
- Nichol, H., Law, J. H., and Winzerling, J. J. (2002). Iron metabolism in insects. *Annu. Rev. Entomol.* 47, 535–559. doi: 10.1146/annurev.ento.47.091201.145237
- Parker, K. R., and Mant, M. J. (1979). Effects of tsetse (*Glossina morsitans morsitans* Westw.) (*Diptera: Glossinidae*) salivary gland homogenate on coagulation and fibrinolysis. *Thromb. Haemost.* 42, 743–751.
- Pertea, G., Huang, X., Liang, F., Antonescu, V., Sultana, R., Karamycheva, S., et al. (2003). TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics* 19, 651–652. doi: 10.1093/bioinformatics/btg034
- Ravel, S., Grébaut, P., Cuisance, D., and Cuny, G. (2003). Monitoring the developmental status of *Trypanosoma brucei gambiense* in the tsetse fly by means of PCR analysis of anal and saliva drops. *Acta Trop.* 88, 161–165. doi: 10.1016/S0001-706X(03)00191-8
- Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite. *Trends Genet.* 16, 276–277. doi: 10.1016/S0168-9525(00)02024-2
- Rio, R. V., Hu, Y., and Aksoy, S. (2004). Strategies of the home-team: symbioses exploited for vector-borne disease control. *Trends Microbiol.* 12, 325–336. doi: 10.1016/j.tim.2004.05.001
- Salem, M., Vallejo, R. L., Leeds, T. D., Palti, Y., Liu, S., Sabbagh, A., et al. (2012). RNA-Seq identifies SNP markers for growth traits in rainbow trout. *PLoS ONE* 7:e36264. doi: 10.1371/journal.pone.0036264
- Schulz, M., and Zerbino, D. (2010). *Oases - De novo Transcriptome Assembler for Very Short Reads*. Technical Report, MPI for Molecular Genomics, European Bioinformatics Institute (EMBL-EBI), UC Santa Cruz.
- Schulz, M. H., Zerbino, D. R., Vingron, M., and Birney, E. (2012). Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* 28, 1086–1092. doi: 10.1093/bioinformatics/bts094
- Silva, J. C., Loreto, E. L., and Clark, J. B. (2004). Factors that affect the horizontal transfer of transposable elements. *Curr. Issues Mol. Biol.* 6, 57–71.
- Simarro, P. P., Jannin, J., and Cattand, P. (2008). Eliminating human African trypanosomiasis: where do we stand and what comes next? *PLoS Med.* 5:e55. doi: 10.1371/journal.pmed.0050055
- Smallwood, S. A., Lee, H. J., Angermueller, C., Krueger, F., Saadeh, H., Peat, J., et al. (2014). Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat. Methods* 11, 817–820. doi: 10.1038/nmeth.3035
- Stark, K. R., and James, A. A. (1998). Isolation and characterization of the gene encoding a novel factor Xa-directed anticoagulant from the yellow fever mosquito, *Aedes aegypti*. *J. Biol. Chem.* 273, 20802–20809. doi: 10.1074/jbc.273.33.20802
- Swaminathan, K., Chae, W. B., Mitros, T., Varala, K., Xie, L., Barling, A., et al. (2012). A framework genetic map for *Misanthus sinensis* from RNAseq-based markers shows recent tetraploidy. *BMC Genomics* 13:142. doi: 10.1186/1471-2164-13-142
- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. doi: 10.1038/nbt.1621
- Van den Abbeele, J., Claes, Y., Bockstaele, D., Ray, D., and Coosemans, M. (1999). *Trypanosoma brucei* spp. development in the tsetse fly: characterization of the post-mesocyclic stages in the foregut and proboscis. *Parasitology* 118, 469–478. doi: 10.1017/S0031182099004217
- Vreyen, M. J., Saleh, K. M., Ali, M. Y., Abdulla, A. M., Zhu, Z. R., Juma, K. G., et al. (2000). *Glossina austeni* (Diptera: Glossinidae) eradicated on the island of Unguja, Zanzibar, using the sterile insect technique. *J. Econ. Entomol.* 93, 123–135. doi: 10.1603/0022-0493.93.1.123
- Wang, J., Lan, P., Gao, H., Zheng, L., Li, W., and Schmidt, W. (2013). Expression changes of ribosomal proteins in phosphate- and iron-deficient Arabidopsis roots predict stress-specific alterations in ribosome composition. *BMC Genomics* 14:783. doi: 10.1186/1471-2164-14-783
- Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63. doi: 10.1038/nrg2484
- Weiss, B. L., Savage, A. F., Griffith, B. C., Wu, Y., and Aksoy, S. (2014). The peritrophic matrix mediates differential infection outcomes in the tsetse fly gut following challenge with commensal, pathogenic, and parasitic microbes. *J. Immunol.* 193, 773–782. doi: 10.4049/jimmunol.1400163
- Weiss, B. L., Wang, J., Maltz, M. A., Wu, Y., and Aksoy, S. (2013). Trypanosome infection establishment in the tsetse fly gut is influenced by microbiome-regulated host immune barriers. *PLoS Pathog.* 9:e100331. doi: 10.1371/journal.ppat.100331
- Welburn, S. C., and Maudlin, I. (1999). Tsetse-trypanosome interactions: rites of passage. *Parasitol. Today* 15, 399–403. doi: 10.1016/S0169-4758(99)01512-4
- Welburn, S. C., Maudlin, I., and Simarro, P. P. (2009). Controlling sleeping sickness-a review. *Parasitology* 136, 1943–1949. doi: 10.1017/S0031182009006416
- Yoshiga, T., Georgieva, T., Dunkov, B. C., Harizanova, N., Ralchev, K., and Law, J. H. (1999). *Drosophila melanogaster* transferrin. Cloning, deduced protein sequence, expression during the life cycle, gene localization and up-regulation on bacterial infection. *Eur. J. Biochem.* 260, 414–420. doi: 10.1046/j.1432-1327.1999.00173.x

- Yoshiga, T., Hernandez, V. P., Fallon, A. M., and Law, J. H. (1997). Mosquito transferrin, an acute-phase protein that is up-regulated upon infection. *Proc. Natl. Acad. Sci. U.S.A.* 94, 12337–12342. doi: 10.1073/pnas.94.23.12337
- Yun, E. Y., Kang, S. W., Hwang, J. S., Goo, T. W., Kim, S. H., Jin, B. R., et al. (1999). Molecular cloning and characterization of a cDNA encoding a transferrin homolog from *Bombyx mori*. *Biol. Chem.* 380, 1455–1459. doi: 10.1515/BC.1999.188
- Zerbino, D. R., and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18, 821–889. doi: 10.1101/gr.074492.107

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Hamidou Soumana, Klopp, Ravel, Nabihouidine, Tchicaya, Parrinello, Abate, Rialle and Geiger. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Giant Magnetoresistance-based Biosensor for Detection of Influenza A Virus

Venkatramana D. Krishna<sup>1†</sup>, Kai Wu<sup>2†</sup>, Andres M. Perez<sup>1\*</sup> and Jian-Ping Wang<sup>2\*\*</sup>

<sup>1</sup> Department of Veterinary Population Medicine, College of Veterinary Medicine, University of Minnesota, St. Paul, MN, USA

<sup>2</sup> Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN, USA

## OPEN ACCESS

### Edited by:

Fabrice Merien,  
Auckland University of Technology,  
New Zealand

### Reviewed by:

Bernard A. P. Lafont,  
National Institutes of Health, USA  
Benoit Chassaing,  
Georgia State University, USA

### \*Correspondence:

Jian-Ping Wang  
jpwang@umn.edu;  
Andres M. Perez  
aperez@umn.edu

<sup>†</sup>These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Infectious Diseases,  
a section of the journal  
*Frontiers in Microbiology*

Received: 18 January 2016

Accepted: 14 March 2016

Published: 29 March 2016

### Citation:

Krishna VD, Wu K, Perez AM  
and Wang J-P (2016) Giant  
Magnetoresistance-based Biosensor  
for Detection of Influenza A Virus.  
*Front. Microbiol.* 7:400.  
doi: 10.3389/fmicb.2016.00400

We have developed a simple and sensitive method for the detection of influenza A virus based on giant magnetoresistance (GMR) biosensor. This assay employs monoclonal antibodies to viral nucleoprotein (NP) in combination with magnetic nanoparticles (MNPs). Presence of influenza virus allows the binding of MNPs to the GMR sensor and the binding is proportional to the concentration of virus. Binding of MNPs onto the GMR sensor causes change in the resistance of sensor, which is measured in a real time electrical readout. GMR biosensor detected as low as  $1.5 \times 10^2$  TCID<sub>50</sub>/mL virus and the signal intensity increased with increasing concentration of virus up to  $1.0 \times 10^5$  TCID<sub>50</sub>/mL. This study showed that the GMR biosensor assay is relevant for diagnostic application since the virus concentration in nasal samples of influenza virus infected swine was reported to be in the range of  $10^3$  to  $10^5$  TCID<sub>50</sub>/mL.

**Keywords:** giant magnetoresistance, biosensor, magnetic nanoparticle, GMR chip, influenza A virus

## INTRODUCTION

Influenza viruses belong to the family *Orthomyxoviridae*, which are enveloped single strand negative sense RNA viruses with segmented RNA genome. Based on their matrix (M) and nucleoprotein (NP), influenza viruses are classified into type A, B, or C. Influenza A viruses (IAVs) are further classified into subtypes based on their surface glycoproteins, hemagglutinin (HA), and neuraminidase (NA). IAV is a common respiratory pathogen infecting many hosts including humans, pigs (swine influenza virus or SIV) and birds (avian influenza virus or AIV). In addition to SIV, pigs are susceptible to infection with influenza viruses of human and avian origin and this is believed to contribute to novel reassortant influenza viruses with pandemic potential (Kida et al., 1994). Surveillance of swine and avian influenza viruses in the wild, in farms, and in live bird markets is critical for detection of newly emerging influenza viruses with significant impact on human and veterinary public health. Rapid, sensitive, and reliable method for detection of IAV in the environment, tissues and body fluids is important for controlling the infection and reducing the impact of possible influenza pandemic by early detection and rapid intervention. Currently, laboratory diagnosis of IAV relies on isolation of virus in embryonated chicken eggs or cell culture, detection of viral antigens, serological tests to detect virus specific antibodies, and detection of viral RNA by reverse transcription-quantitative polymerase chain reaction (RT-qPCR; Lee et al., 1993; Townsend et al., 2006; Leuwerke et al., 2008; Chen et al., 2011). Virus isolation is sensitive method and considered gold standard for virus diagnosis (Amano and Cheng, 2005), however, this labor intensive technique requires average of 3–7 days to obtain the results (Ellis and Zambon, 2002). Detection of viral antigens and serological test for antibody detection are either poor in specificity

or low in sensitivity. Although, RT-qPCR is highly sensitive and specific method, its requirement for expensive laboratory instruments and technical expertise (Ellis and Zambon, 2002; Payungporn et al., 2006) in addition to longer time for the completion of the test, as it involves RNA extraction step, limits its application in the field. The objective of this study is to develop sensitive and specific method for detection of swine influenza viruses with minimum sample handling and laboratory skill requirements.

Various technologies have been developed for rapid, sensitive, and specific detection of virus using nanotechnology-based approaches (Lee et al., 2011; Nidzworski et al., 2014). These technologies use nanoparticles in combination with electrical or electrochemical detection (Patolsky et al., 2004; Tam et al., 2009; Shirale et al., 2010; Driskell et al., 2011; Li et al., 2011; Singh et al., 2014). To date, chip-based giant magnetoresistance (GMR) spin valves along with magnetic nanoparticles (MNPs) have become a powerful tool for high sensitivity, real-time electrical readout, and rapid biomolecule detection (Baselt et al., 1998; Rife et al., 2003; Graham et al., 2004; Schotter et al., 2004; Millen et al., 2005; Loureiro et al., 2009, 2011; Gaster et al., 2011; Wang et al., 2014). The fabrication and integration of GMR biosensors are compatible with the large multiplex technology and the current Very Large Scale Integration (VLSI) technology (Wang et al., 2015) so it is possible to lower down the cost if the mass production is carried out. Moreover, GMR chips can be integrated with not only electronics but also microfluidics for immunoassay applications (Xu et al., 2008; Zhi et al., 2012). In addition, GMR biosensors are matrix-insensitive (Zhang and Zhou, 2012) and therefore their performance are very robust and not affected by environmental factors such as temperature and pH.

Giant magnetoresistance-based immunoassay detection is based on the principle that stray field from MNPs that bound on sensor surface will alter the magnetization in free layer (Supplementary Figure S1), thus changing the resistance of GMR sensors (Baibich et al., 1988; Binash et al., 1989). A higher number of MNPs bound to GMR sensors per unit area leads to a higher detection signal. GMR sensors have been utilized previously in biomolecule and chemical detection (Srinivasan et al., 2009, 2011; Zhi et al., 2012). Unlike fluorescent labels used in immunofluorescence methods, MNPs do not bleach (Eickenberg et al., 2013). In addition, there is no ferromagnetism property in biological samples, allowing the detection of magnetic signals with less background noise (Zhang et al., 2013). Nowadays, the size of MNPs can be controlled to the identical size as the biomolecules to which they will interact with (Hsing et al., 2007). Furthermore, labeling large molecules as well as nano- or micro-particles with small biomolecules can be successfully realized (Hsing et al., 2007; Ladj et al., 2013; Zhou et al., 2015).

In the present study, we demonstrated sensitive detection of influenza virus using GMR biosensors. The virus type specific broadly reactive monoclonal antibodies to NP employed in this study were able to detect IAV of swine and human origin in direct antigen capture enzyme linked immunosorbent assay (ELISA). Using swine influenza virus H3N2v as a representative virus

we found the limit of detection of GMR biosensor assay was  $1.5 \times 10^2$  TCID<sub>50</sub>/mL virus. Comparison of GMR biosensor-based detection with antigen capture ELISA showed that GMR biosensor was more sensitive. In addition, GMR biosensor-based assay allows for a real time measurement of signals. The signals are captured and processed immediately as it is generated and can be monitored continuously without operator intervention.

## MATERIALS AND METHODS

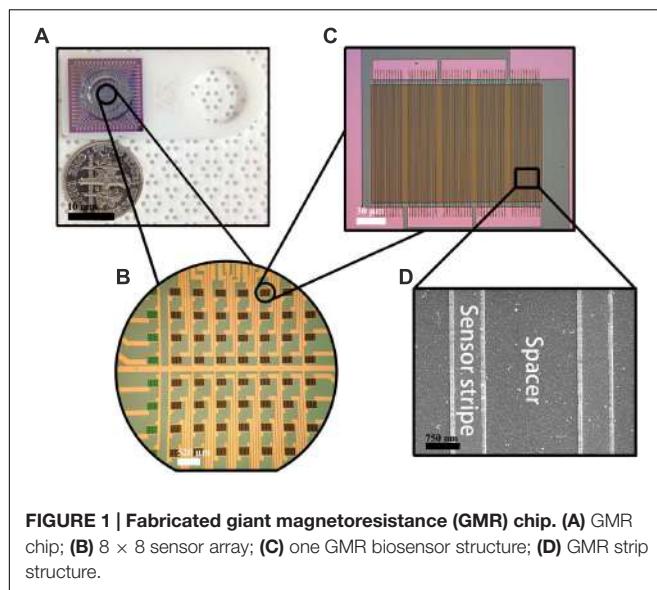
### Viruses

The human pandemic influenza A/California/04/2009 (H1N1 CA/09), the swine influenza viruses A/Sw/Iowa/73 (H1N1 IA/73), A/Sw/Illinois/2008 (H1N1 IL/08), and A (H3N2) variant virus (H3N2v) were obtained from the University of Minnesota Veterinary Diagnostic Laboratory (St Paul, MN). Viruses were propagated in Madin-Darby canine kidney (MDCK) cells (ATCC CCL-34) in Dulbecco's modified Eagle medium (DMEM) containing 0.5 µg/mL TPCK-trypsin (Worthington Biochemical Corporation, Lakewood, NJ, USA) and purified from the clarified cell culture supernatants by ultracentrifugation through a 30% (w/v) sucrose cushion and stored in aliquots at -80°C. Culture supernatant from un-infected MDCK cells were processed similarly to use for mock virus preparation. The concentration of purified virus was determined by TCID<sub>50</sub> assay. For immunoassays the virus was inactivated at 60°C for 1 h. To disrupt the virus particles, the mock and virus preparation were treated with 1% IGEPAL CA-630 (Sigma-Aldrich, Product No. I8896) for 10 min at 37°C.

### GMR Chip Fabrication and Sensor Array Structure

The multilayer GMR spin valve films with top-down structure of Ta (50 Å)/NiFe (20 Å)/CoFe(10 Å)/Cu(33 Å)/CoFe(25 Å)/IrMn(80 Å)/Ta (25 Å) were deposited by a Shamrock Magnetron Sputter System onto Si/SiO<sub>2</sub> (1000 Å) substrate at the University of Minnesota. A 4-inch GMR wafer containing 21 usable chips is manufactured by photolithography, ion beam milling, and electron beam evaporation techniques. An 18 nm thick Al<sub>2</sub>O<sub>3</sub> layer was coated onto chip surface by atomic layer deposition (ALD) followed by a 20 nm SiO<sub>2</sub> layer by plasma-enhanced chemical vapor deposition (PECVD) in order to prevent current leakage and in the meanwhile SiO<sub>2</sub> layer paves the way for future surface functionalization.

Each GMR chip is in the size of 16 mm × 16 mm with 8 × 8 sensor array in its center (Figures 1A,B). Each sensor is in the size of 120 µm × 120 µm containing five GMR strip groups connected in series and each group contains 10 GMR strips connected in parallel (Figure 1C). Each strip with the size of 120 µm × 750 nm is separated by 2 µm (Figure 1D). All the GMR chips were annealed at 200°C under an applied magnetic field of 0.5 Tesla along the minor axis (Figure 1C) for 1 h then naturally cooled down to room temperature in order to fully align the magnetization in the pinned layer.



**FIGURE 1 |** Fabricated giant magnetoresistance (GMR) chip. **(A)** GMR chip; **(B)** 8 × 8 sensor array; **(C)** one GMR biosensor structure; **(D)** GMR strip structure.

## GMR Biosensor Surface Functionalization

Giant magnetoresistance chips are first exposed to ultraviolet light and ozone (UVO) for 15 min to remove organic material from the sensor surface as well as to expose the hydroxyl group bonding sites. Each chip is then soaked in 5 mL anhydrous toluene mixed with 1% of 3-aminopropyltriethoxy silane (APTES) for 1 h at room temperature to allow APTES to covalently bind to the hydroxyl group (Figure 2A) from silica layer that is on-top-of GMR biosensors. Chips are thoroughly rinsed with acetone followed by ethanol and dried with nitrogen gas. The surface of APTES modified chips contain amino groups. To attach aldehyde groups onto sensor surface, the 64-sensor-array area of each chip is covered with 5% glutaraldehyde (Glu) solution (100 µL) and incubated at room temperature for 5 h under a relative humidity of ~97%. The terminal aldehyde groups generated on the sensor surface allow subsequent covalent bonding of biomolecules containing amino groups onto GMR sensor (Wang et al., 2013, 2014, 2015).

## Influenza A Virus Immunoassay

3-Aminopropyltriethoxysilane–Glu modified GMR sensors were robotically printed with 500 µg/mL influenza A capture antibody (MAB8800; EMD Millipore Corporation, Temecula, CA, USA, specific to IAV NP) in a volume of 1.2 nL per sensor using the sci-FLEXARRAYER S5 (Scienion, Germany; Supplementary Figure S2). For the control reactions bovine serum albumin (BSA; 1 mg/mL) and biotinylated bovine serum albumin (biotin-BSA; 1 mg/mL) were similarly printed onto GMR sensors. The 8 × 8 sensor array were divided into three regions (Supplementary Figure S2). Four columns (32 sensors) were spotted with influenza A capture antibody, two columns (16 sensors) with biotin-BSA, and the rest two columns (16 sensors) with BSA. Printed chips were incubated at 4°C for 12 h under a relative humidity of ~97%. A bottomless reaction well made

of polymethyl methacrylate (PMMA) was attached onto chip centered at the sensor area. This reaction well can hold as much as 100 µL liquid. Next, the sensor area was rinsed with PBST [0.05% tween 20 in phosphate buffered saline (PBS)] for three cycles to remove unbound biomolecules. To block any potential binding sites on the sensor, 100 µL of 10 mg/mL BSA was added to the reaction well and incubated at room temperature for 30 min. After removing BSA and washing the sensor area with 100 µL of PBST for three cycles, 100 µL of antigen (heat inactivated virus) of different concentrations were added to the reaction well and incubated at room temperature for 1 h. After washing the sensor area with 100 µL of PBST for three cycles, 100 µL of 5 µg/mL biotinylated detection antibody (MAB8257B, EMD Millipore Corporation, Temecula, CA, USA, a mouse anti influenza A monoclonal antibody specific for NP) was added and incubated at room temperature for another 1 h. Subsequently, detection antibody was aspirated and sensor area was rinsed with PBST for three cycles. Chips were kept at 4°C, 97% humidity condition before real-time testing. In order to detect all IAV subtypes, the capture and detection antibodies specific to influenza A NP were used. These antibodies were certified by the manufacturer as influenza A specific and the detection antibody was not shown to cross react with influenza B or other respiratory viruses.

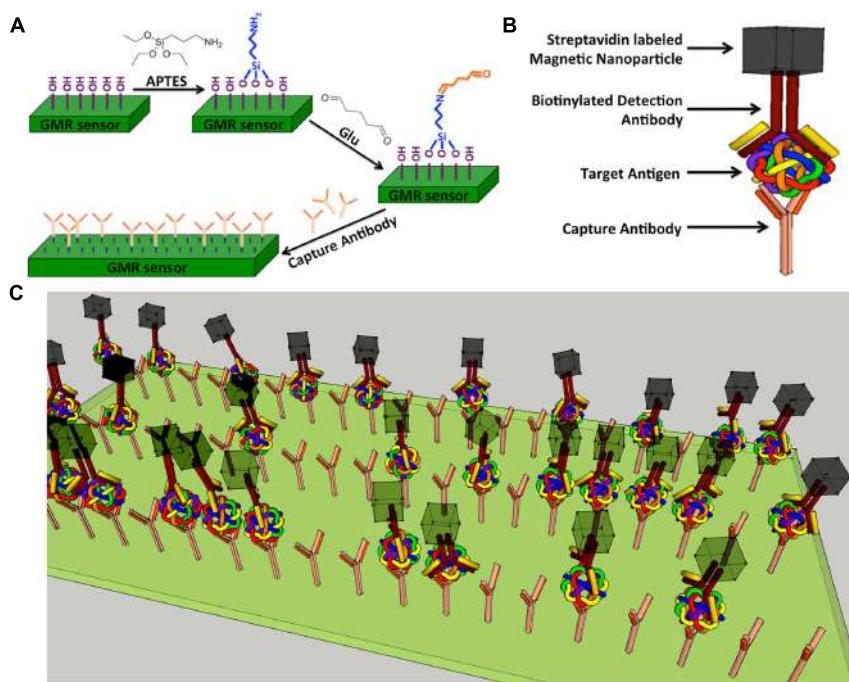
## Detection Principle and Signal Flow

A sandwich assay structure (Srinivasan et al., 2009, 2011) used in this study is illustrated in Figure 2B. The detailed detection architecture is shown in Figure 2C. First, capture antibody was immobilized on the GMR sensor, then the antigen and biotinylated detection antibody were added successively and allowed to bind. Finally, streptavidin labeled MNPs (Miltenyi Biotec, Inc., Auburn, CA, USA; Catalog No. 130-048-101) were added and specifically bound to the near surface of sensor through the biotin-streptavidin interaction. Number of bound MNPs is proportional to number of target antigen. It is worthwhile to mention that, since there are 64 sensors in one GMR chip, it is possible to detect 64 types of biomolecules in one test.

In a bench top system, a probe station with 17 × 4 pin array (Supplementary Figure S2) is connected to the pads of GMR chip (Wang et al., 2015). An alternating current with frequency of 1000 Hz flows through the main bus. An in-plane magnetic field with amplitude of 30 Oe, frequency of 50 Hz is applied along the minor axis direction. A Digital Acquisition card (DAQ, NI USB-6289, 18-Bit, 625 kS/s) collects analog signals from side tones at 950 and 1050 Hz and carries out fast Fourier transform (FFT) before sending the data points back. It takes 1 s to collect one data points on one sensor, since there are 64 sensors in one GMR chip; it takes 1 min to go through all the sensors. Signals are extracted from the background noise using a Wheatstone bridge and then amplified by low-noise, low-distortion instrumentation amplifier (INA163, Texas Instruments).

## Enzyme Linked Immunosorbent Assay

Microtiter plates (Corning, Inc., Corning, NY, USA) were coated with 100 µL of 3 µg/mL anti-influenza A monoclonal antibody

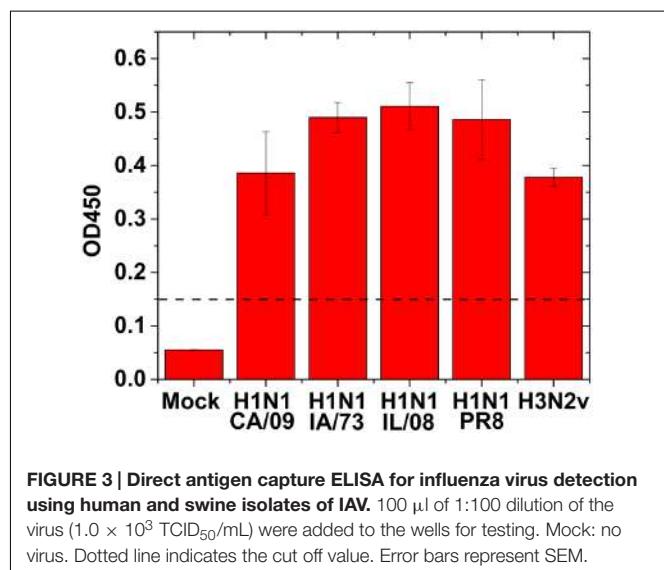


**FIGURE 2 | Schematic representation of GMR biosensor. (A)** Schematic diagram of GMR biosensor surface functionalization. **(B)** Schematic drawing of a typical sandwich structure; **(C)** Schematic illustration of influenza A virus detection.

(MAB8800; EMD Millipore Corporation, Temecula, CA, USA) specific for influenza A NP. After overnight incubation at 4°C, the wells were blocked with 5% skim milk in PBS for 2 h at room temperature. 100 µL of heat inactivated virus diluted in sample diluent (3% BSA in PBS) was then added and incubated for 1 h at 37°C. After washing the wells three times with wash buffer (0.05% tween 20 in PBS), 100 µL of 1 µg/mL biotinylated anti-influenza A monoclonal antibody (MAB8257B; EMD Millipore Corporation, Temecula, CA, USA) was added and incubated for 1 h at room temperature. Wells were washed three times with wash buffer and incubated for 30 min at room temperature with 100 µL of 1:1000 diluted Pierce high sensitivity streptavidin-horseradish peroxidase (HRP; Thermo scientific, Rockford, IL, USA). After washing the wells three times with wash buffer, 100 µL of one step ultra TMB (Thermo scientific, Rockford, IL, USA) was added and the reaction was stopped after 30 min incubation at room temperature by adding 100 µL of 1 N H<sub>2</sub>SO<sub>4</sub>. The absorbance at 450 nm was measured by microtiter plate reader (Thermo Scientific). The cut off value was calculated as mean of negative control multiplied by 2.

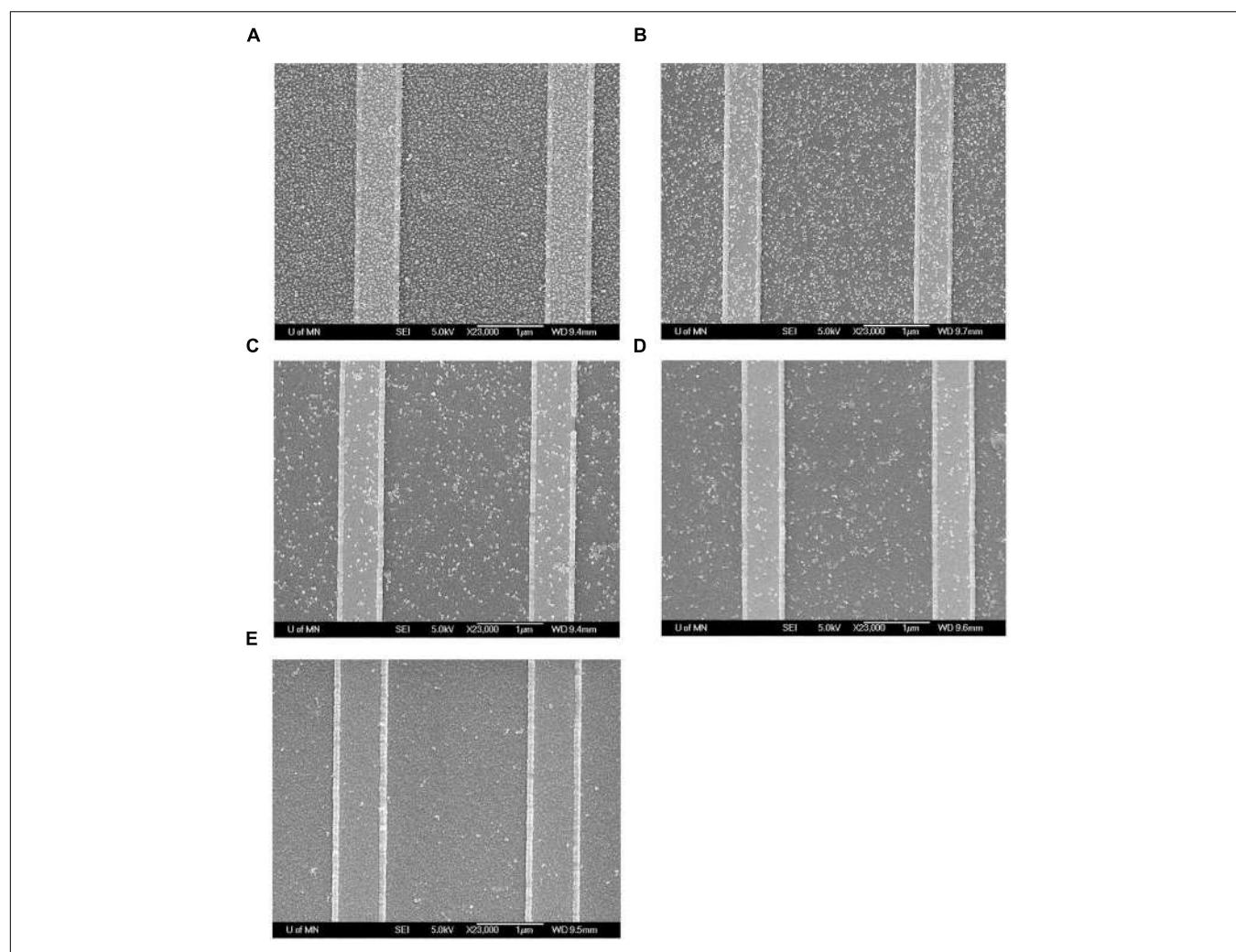
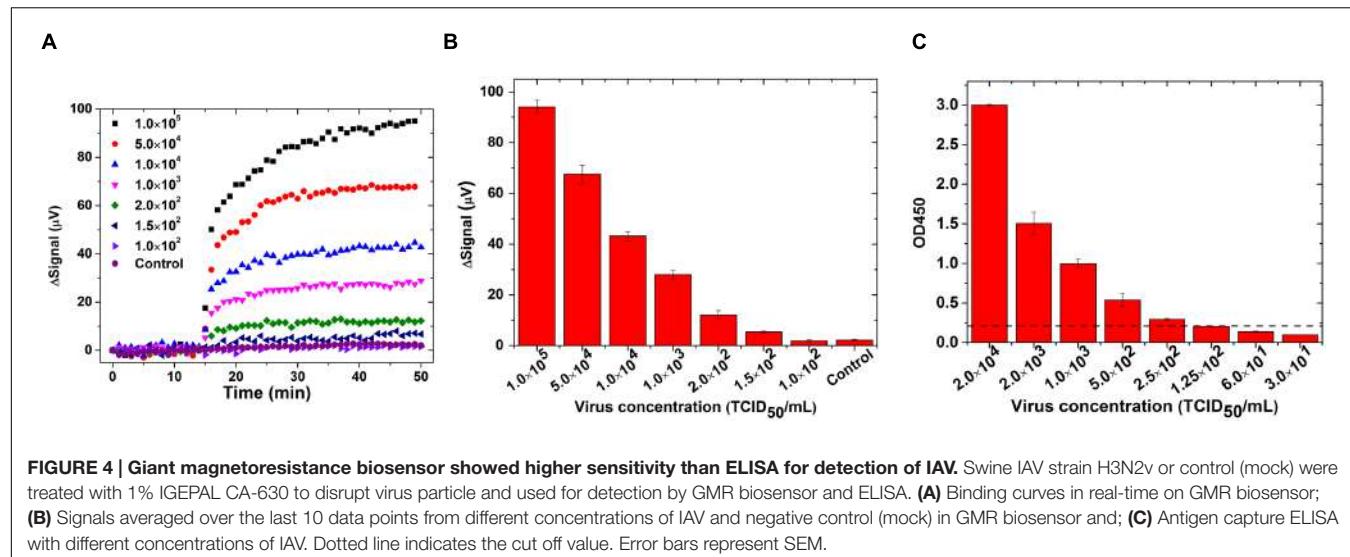
## RESULTS

To validate the IAV specific antibodies to use in GMR biosensor assay, we performed antigen capture ELISA using human or swine isolates of IAV. All the IAV strains tested showed positive result by ELISA (**Figure 3**) suggesting that these antibodies are capable of detecting multiple IAV strains.



**FIGURE 3 | Direct antigen capture ELISA for influenza virus detection using human and swine isolates of IAV.** 100 µL of 1:100 dilution of the virus ( $1.0 \times 10^3$  TCID<sub>50</sub>/mL) were added to the wells for testing. Mock: no virus. Dotted line indicates the cut off value. Error bars represent SEM.

The real-time binding curves for influenza biosensor are shown in **Figure 4A**. Initially 30 µL of PBS was preloaded into the reaction well. After 15 min of stabilization, 30 µL MNP solution was added into reaction and signals were collected for another 35 min. Signals from reactions with different concentrations of virus increased immediately after addition of MNP solution, which indicates a real-time binding of MNPs onto GMR sensors. Furthermore, the signal increased with increasing concentration of virus. The averaged signal from  $1.5 \times 10^2$  TCID<sub>50</sub>/mL viruses



is  $5.45 \mu\text{V}$ , and it goes up with increased virus concentration and reaches to  $94.1 \mu\text{V}$  for  $1.0 \times 10^5 \text{ TCID}_{50}/\text{mL}$  virus (**Figure 4B**). The signal from negative control group did not show any obvious rise and the averaged signal was  $2.2 \mu\text{V}$ , indicating that the signals are specific to IAV. The cut-off value for distinguishing positive from negative was set as  $3.0 \mu\text{V}$ , since the mean value of negative control plus three times standard deviation of the mean is  $3.0$ . As the signal from  $1.0 \times 10^2 \text{ TCID}_{50}/\text{mL}$  viruses was below  $3.0 \mu\text{V}$  ( $1.52 \pm 0.37 \mu\text{V}$ ), the detection limit of GMR sensor was estimated as  $1.5 \times 10^2 \text{ TCID}_{50}/\text{mL}$ . We compared GMR biosensor assay for influenza virus with antigen capture ELISA and found that the GMR biosensor was more sensitive than ELISA. The limit of detection of antigen capture ELISA using these monoclonal antibodies was  $2.5 \times 10^2 \text{ TCID}_{50}/\text{mL}$  (**Figure 4C**).

To further confirm the binding of MNPs to the GMR sensor, these chips were investigated by field-emission gun scanning electron microscopy (FEG-SEM) at the Characterization Facility, University of Minnesota. The GMR chips were rinsed with DI water to wash away unbound MNPs and dried by nitrogen gas. The chips were then coated with  $50 \text{ \AA}$  of Platinum (Pt) and observed by FEG-SEM. As shown in **Figure 5**, number of bound MNPs per unit area increases as the concentration of influenza virus increases.

## DISCUSSION

This work extends the application of GMR-based assay for virus detection. In this study, a sensitive GMR biosensor was developed for the detection of influenza virus. We selected viral NP as target antigen to detect all strains of IAV, since large scale sequence analysis of this viral protein showed high degree of conservation among different subtypes of IAV from multiple hosts and lineages (Kukol and Hughes, 2014). As NP is localized within the virus particle, a non-ionic, non-denaturing detergent IGEPAL CA-630 was used to disrupt the virus particles in the sample. The results demonstrated that the GMR biosensor is able to detect viral concentrations ranging from  $1.5 \times 10^2$  to  $1.0 \times 10^5 \text{ TCID}_{50}/\text{mL}$ . This is relevant to nasal samples of infected swine, which has been reported to contain  $10^3$  to  $10^5 \text{ TCID}_{50}/\text{mL}$  viral particles (Lekcharoensuk et al., 2006). In addition, our results showed that the GMR biosensor is more sensitive than ELISA. Therefore, GMR biosensor assay has a potential for diagnosis and monitoring of influenza virus. In this study, we used purified virus diluted in buffer for the assay. Further work is required to address the effect of sample matrix on the sensitivity and specificity of GMR sensor. Although, detection antibody was not reported to cross react with influenza B or other respiratory

## REFERENCES

- Amano, Y., and Cheng, Q. (2005). Detection of influenza virus: traditional approaches and development of biosensors. *Anal. Bioanal. Chem.* 381, 156–164. doi: 10.1007/s00216-004-2927-0
- Baibich, M. N., Broto, J. M., Fert, A., Van Dau, F. N., Petroff, F., Etienne, P., et al. (1988). Giant magnetoresistance of (001) Fe/(001) Cr magnetic superlattices. *Phys. Rev. Lett.* 61:2472. doi: 10.1103/PhysRevLett.61.2472

viruses, further validation of GMR biosensor using these related viruses as well as multiple strains of swine influenza viruses are needed to assess the practical application of this technique.

Although, IAV H3N2v was used as a representative influenza virus in our GMR biosensor system, this assay can be further extended to other viruses and infectious agents. Since there are 64 sensor arrays in one GMR chip, this assay can be further optimized to detect different subtypes of influenza viruses in a single test or simultaneously detect different pathogens from single sample. The assay is simple with minimum sample preparation, although GMR biosensor surface functionalization and antibody immobilization on the sensor requires time and labor. Protein incubation, sample handling and washing could be improved by integrating with the well-developed microfluidic channel platform. In addition, it is possible to integrate this assay into portable, hand held device for on-site application.

## AUTHOR CONTRIBUTIONS

J-PW and AP conceptualized this research. VK and KW planned the experiments. KW carried out the GMR biosensing experiments. VK carried out the ELISA reference experiments. The manuscript was written through contributions of all authors. All authors have given approval to the final version.

## FUNDING

This work is supported by MNDrive: OVPR STEMMA program, Institute of Engineering in Medicine of the University of Minnesota, National Science Foundation MRSEC facility program, the Distinguished McKnight University Professorship, Centennial Chair Professorship and UROP program from the University of Minnesota.

## ACKNOWLEDGMENTS

We thank Maxim Cheeran for his advice and constant help. We also thank Yinglong Feng, Yi Wang and Wei Wang for the fruitful discussions.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2016.00400>

Baselt, D. R., Lee, G. U., Natesan, M., Metzger, S. W., Sheehan, P. E., and Colton, R. J. (1998). A biosensor based on magnetoresistance technology. *Biosens. Bioelectron.* 13, 731–739. doi: 10.1016/S0956-5663(98)00037-2

Binasch, G., Grünberg, P., Saurenbach, F., and Zinn, W. (1989). Enhanced magnetoresistance in layered magnetic structures with antiferromagnetic interlayer exchange. *Phys. Rev. B* 39, 4828. doi: 10.1103/PhysRevB.39.4828

- Chen, Y., Cui, D., Zheng, S., Yang, S., Tong, J., Yang, D., et al. (2011). Simultaneous detection of influenza A, Influenza B, and respiratory syncytial viruses and subtyping of influenza A H3N2 Virus and H1N1 (2009) virus by multiplex real-time PCR. *J. Clin. Microbiol.* 49, 1653–1656. doi: 10.1128/JCM.02184-10
- Driskell, J. D., Jones, C. A., Tompkins, S. M., and Tripp, R. A. (2011). One-step assay for detecting influenza virus using dynamic light scattering and gold nanoparticles. *Analyst* 136, 3083–3090. doi: 10.1039/c1an15303j
- Eickenberg, B., Meyer, J., Helmich, L., Kappe, D., Auge, A., Wedemann, A., et al. (2013). Lab-on-a-chip magneto-immunoassays: how to ensure contact between superparamagnetic beads and the sensor surface. *Biosensors* 3:327. doi: 10.3390/bios3030327
- Ellis, J. S., and Zambon, M. C. (2002). Molecular diagnosis of influenza. *Rev. Med. Virol.* 12, 375–389. doi: 10.1002/rmv.370
- Gaster, R. S., Xu, L., Han, S.-J., Wilson, R. J., Hall, D. A., Osterfeld, S. J., et al. (2011). Quantification of protein interactions and solution transport using high-density GMR sensor arrays. *Nat. Nanotechnol.* 6, 314–320. doi: 10.1038/nnano.2011.45
- Graham, D. L., Ferreira, H. A., and Freitas, P. P. (2004). Magnetoresistive-based biosensors and biochips. *Trends Biotechnol.* 22, 455–462. doi: 10.1016/j.tibtech.2004.06.006
- Hsing, I., Xu, Y., and Zhao, W. (2007). Micro-and nano-magnetic particles for applications in biosensing. *Electroanalysis* 19, 755–768. doi: 10.1002/elan.200603785
- Kida, H., Ito, T., Yasuda, J., Shimizu, Y., Itakura, C., Shortridge, K. F., et al. (1994). Potential for transmission of avian influenza viruses to pigs. *J. Gen. Virol.* 75, 2183–2188. doi: 10.1099/0022-1317-75-9-2183
- Kukol, A., and Hughes, D. J. (2014). Large-scale analysis of influenza A virus nucleoprotein sequence conservation reveals potential drug-target sites. *Virology* 454–455, 40–47. doi: 10.1016/j.virol.2014.01.023
- Ladj, R., Bitar, A., Eissa, M., Mugnier, Y., Le Dantec, R., Fessi, H., et al. (2013). Individual inorganic nanoparticles: preparation, functionalization and in vitro biomedical diagnostic applications. *J. Materials Chem. B* 1, 1381–1396. doi: 10.1039/c2tb00301e
- Lee, B. W., Bey, R. F., Baarsch, M. J., and Simonson, R. R. (1993). ELISA method for detection of influenza a infection in swine. *J. Vet. Diagn. Invest.* 5, 510–515. doi: 10.1177/104063879300500402
- Lee, D., Chander, Y., Goyal, S. M., and Cui, T. (2011). Carbon nanotube electric immunoassay for the detection of swine influenza virus H1N1. *Biosens. Bioelectron.* 26, 3482–3487. doi: 10.1016/j.bios.2011.01.029
- Lekcharoensuk, P., Lager, K. M., Vemulapalli, R., Woodruff, M., Vincent, A. L., and Richt, J. A. (2006). Novel swine influenza virus subtype H3N1, United States. *Emerg. Infect. Dis.* 12, 787–794. doi: 10.3201/eid1205.051060
- Leuwerke, B., Kitikoon, P., Evans, R., and Thacker, E. (2008). Comparison of three serological assays to determine the cross-reactivity of antibodies from eight genetically diverse U.S. swine influenza viruses. *J. Vet. Diagn. Invest.* 20, 426–432. doi: 10.1177/104063870802000403
- Li, D., Wang, J., Wang, R., Li, Y., Abi-Ghanem, D., Berghman, L., et al. (2011). A nanobeads amplified QCM immunosensor for the detection of avian influenza virus H5N1. *Biosens. Bioelectron.* 26, 4146–4154. doi: 10.1016/j.bios.2011.04.010
- Loureiro, J., Andrade, P., Cardoso, S., Da Silva, C., Cabral, J., and Freitas, P. (2011). Magnetoresistive chip cytometer. *Lab Chip* 11, 2255–2261. doi: 10.1039/c0lc00324g
- Loureiro, J., Ferreira, R., Cardoso, S., Freitas, P., Germano, J., Fermon, C., et al. (2009). Toward a magnetoresistive chip cytometer: integrated detection of magnetic beads flowing at cm/s velocities in microfluidic channels. *Appl. Phys. Lett.* 95:034104. doi: 10.1063/1.3182791
- Millen, R. L., Kawaguchi, T., Granger, M. C., Porter, M. D., and Tondra, M. (2005). Giant magnetoresistive sensors and superparamagnetic nanoparticles: a chip-scale detection strategy for immunosorbent assays. *Anal. Chem.* 77, 6581–6587. doi: 10.1021/ac0509049
- Nidzworski, D., Pranszke, P., Grudniewska, M., Król, E., and Gromadzka, B. (2014). Universal biosensor for detection of influenza virus. *Biosens. Bioelectron.* 59, 239–242. doi: 10.1016/j.bios.2014.03.050
- Patolsky, F., Zheng, G., Hayden, O., Lakadamyali, M., Zhuang, X., and Lieber, C. M. (2004). Electrical detection of single viruses. *Proc. Natl. Acad. Sci. U.S.A.* 101, 14017–14022. doi: 10.1073/pnas.0406159101
- Payungporn, S., Chutinimitkul, S., Chaiseng, A., Damrongwantanapokin, S., Buranathai, C., Amonsin, A., et al. (2006). Single step multiplex real-time RT-PCR for H5N1 influenza A virus detection. *J. Virol. Methods* 131, 143–147. doi: 10.1016/j.jviromet.2005.08.004
- Rife, J., Miller, M., Sheehan, P., Tamanaha, C., Tondra, M., and Whitman, L. (2003). Design and performance of GMR sensors for the detection of magnetic microbeads in biosensors. *Sensors Actuat.* 107, 209–218. doi: 10.1016/S0924-4247(03)00380-7
- Schotter, J., Kamp, P.-B., Becker, A., Pühler, A., Reiss, G., and Brückl, H. (2004). Comparison of a prototype magnetoresistive biosensor to standard fluorescent DNA detection. *Biosens. Bioelectron.* 19, 1149–1156. doi: 10.1016/j.bios.2003.11.007
- Shirale, D. J., Bangar, M. A., Park, M., Yates, M. V., Chen, W., Myung, N. V., et al. (2010). Label-Free chemiresistive immunosensors for viruses. *Environ. Sci. Technol.* 44, 9030–9035. doi: 10.1021/es102129d
- Singh, R., Sharma, A., Hong, S., and Jang, J. (2014). Electrical immunosensor based on dielectrophoretically-deposited carbon nanotubes for detection of influenza virus H1N1. *Analyst* 139, 5415–5421. doi: 10.1039/c4an01335b
- Srinivasan, B., Li, Y., Jing, Y., Xing, C., Slaton, J., and Wang, J.-P. (2011). A three-layer competition-based giant magnetoresistive assay for direct quantification of endoglin from human urine. *Anal. Chem.* 83, 2996–3002. doi: 10.1021/ac2005229
- Srinivasan, B., Li, Y., Jing, Y., Xu, Y., Yao, X., Xing, C., et al. (2009). A detection system based on giant magnetoresistive sensors and high-moment magnetic nanoparticles demonstrates zeptomole sensitivity: potential for personalized medicine. *Angew. Chem. Int. Edit.* 48, 2764–2767. doi: 10.1002/anie.200806266
- Tam, P. D., Van Hieu, N., Chien, N. D., Le, A.-T., and Anh Tuan, M. (2009). DNA sensor development based on multi-wall carbon nanotubes for label-free influenza virus (type A) detection. *J. Immunol. Methods* 350, 118–124. doi: 10.1016/j.jim.2009.08.002
- Townsend, M. B., Dawson, E. D., Mehlmann, M., Smagala, J. A., Dankbar, D. M., Moore, C. L., et al. (2006). Experimental evaluation of the FluChip diagnostic microarray for influenza virus surveillance. *J. Clin. Microbiol.* 44, 2863–2871. doi: 10.1128/JCM.00134-06
- Wang, W., Wang, Y., Tu, L., Feng, Y., Klein, T., and Wang, J.-P. (2014). Magnetoresistive performance and comparison of supermagnetic nanoparticles on giant magnetoresistive sensor-based detection system. *Sci. Rep.* 4:5716. doi: 10.1038/srep05716
- Wang, W., Wang, Y., Tu, L., Klein, T., Feng, Y., and Wang, J.-P. (2013). Surface modification for Protein and DNA immobilization onto GMR biosensor. *Magnet. IEEE Trans.* 49, 296–299. doi: 10.1109/TMAG.2012.2224327
- Wang, Y., Wang, W., Yu, L., Tu, L., Feng, Y., Klein, T., et al. (2015). Giant magnetoresistive-based biosensing probe station system for multiplex protein assays. *Biosens. Bioelectron.* 70, 61–68. doi: 10.1016/j.bios.2015.03.011
- Xu, L., Yu, H., Akhras, M. S., Han, S.-J., Osterfeld, S., White, R. L., et al. (2008). Giant magnetoresistive biochip for DNA detection and HPV genotyping. *Biosens. Bioelectron.* 24, 99–103. doi: 10.1016/j.bios.2008.03.030
- Zhang, X., Reeves, D. B., Perreard, I. M., Kett, W. C., Griswold, K. E., Gim, B., et al. (2013). Molecular sensing with magnetic nanoparticles using magnetic spectroscopy of nanoparticle Brownian motion. *Biosens. Bioelectron.* 50, 441–446. doi: 10.1016/j.bios.2013.06.049
- Zhang, Y., and Zhou, D. (2012). Magnetic particle-based ultrasensitive biosensors for diagnostics. *Expert Rev. Mol. Diagn.* 12, 565–571. doi: 10.1586/erm.12.54
- Zhi, X., Liu, Q., Zhang, X., Zhang, Y., Feng, J., and Cui, D. (2012). Quick genotyping detection of HBV by giant magnetoresistive biochip combined with PCR and line probe assay. *Lab Chip* 12, 741–745. doi: 10.1039/c2lc20949g

Zhou, J., Wang, C., Wang, P., Messersmith, P. B., and Duan, H. (2015). Multifunctional magnetic nanochains: exploiting self-polymerization and versatile reactivity of mussel-inspired polydopamine. *Chem. Materials* 27, 3071–3076. doi: 10.1021/acs.chemmater.5b00524

**Conflict of Interest Statement:** Dr. J-PW has equity and royalty interests in, and serves on the Board of Directors and the Scientific Advisory Board, for Zepto Life Technology LLC, a company involved in the commercialization of GMR Biosensing technology. The University of Minnesota also has equity and royalty

interests in Zepto Life Tech LLC. These interests have been reviewed and managed by the University of Minnesota in accordance with its Conflict of Interest policies.

Copyright © 2016 Krishna, Wu, Perez and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Rapid Identification and Multiple Susceptibility Testing of Pathogens from Positive-Culture Sterile Body Fluids by a Combined MALDI-TOF Mass Spectrometry and Vitek Susceptibility System

Yueru Tian<sup>1†</sup>, Bing Zheng<sup>2†</sup>, Bei Wang<sup>1</sup>, Yong Lin<sup>1\*</sup> and Min Li<sup>2\*</sup>

## OPEN ACCESS

### Edited by:

Andres M. Perez,  
University of Minnesota, USA

### Reviewed by:

Biswapriya Biswas Misra,  
University of Florida, USA  
Santi M. Mandal,  
Vidyasagar University, India

### \*Correspondence:

Yong Lin  
linyong7007@126.com;  
Min Li  
ruth\_limin01@126.com

<sup>†</sup>These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Infectious Diseases,  
a section of the journal  
*Frontiers in Microbiology*

**Received:** 02 January 2016

**Accepted:** 30 March 2016

**Published:** 20 April 2016

### Citation:

Tian Y, Zheng B, Wang B, Lin Y and  
Li M (2016) Rapid Identification and  
Multiple Susceptibility Testing of  
Pathogens from Positive-Culture  
Sterile Body Fluids by a Combined  
MALDI-TOF Mass Spectrometry and  
Vitek Susceptibility System.  
*Front. Microbiol.* 7:523.

doi: 10.3389/fmicb.2016.00523

Infections of the bloodstream, central nervous system, peritoneum, joints, and other sterile areas are associated with high morbidity and sequelae risk. Timely initiation of effective antimicrobial therapy is crucial to improving patient prognosis. However, standard final identification and antimicrobial susceptibility tests (ASTs) are reported 16–48 h after a positive alert. For a rapid, effective and low-cost diagnosis, we combined matrix-assisted laser desorption/ionization time of flight mass spectrometry with a Vitek AST system, and performed rapid microbial identification (RMI) and rapid multiple AST (RMAST) on non-duplicated positive body fluid cultures collected from a hospital in Shanghai, China. Sterile body fluid positive culture and blood positive culture caused by Gram negative (GN) or polymicrobial were applied to the MALDI-TOF measurement directly. When positive blood culture caused by Gram positive (GP) bacteria or yeasts, they were resuspended in 1 ml brain heart infusion for 2 or 4 h enrichment, respectively. Regardless of enrichment, the RMI (completed in 40 min per sample) accurately identified GN and GP bacteria (98.9 and 87.2%, respectively), fungi (75.7%), and anaerobes (94.7%). Dominant species in multiple cultures and bacteria that failed to grow on the routine plates were correctly identified in 81.2 and 100% of cases, respectively. The category agreements of RMAST results, determined in the presence of various antibiotics, were similarly to previous studies. The RMI and RMAST results not only reduce the turnaround time of the patient report by 18–36 h, but also indicate whether a patient's antibiotic treatment should be accelerated, ceased or de-escalated, and adjusted the essential drugs modification for an optimized therapy.

**Keywords:** MALDI-TOF mass spectrometry, Vitek AST system, sterile body fluids positive culture, rapid diagnosis, clinical impact

**Abbreviations:** A list of abbreviations is shown in Table S2 in Supplementary Material.

## INTRODUCTION

Infections of the bloodstream, central nervous system, peritoneum, joints, and other sterile areas are associated with high morbidity and risk of sequelae (Thigpen et al., 2011; Goto and Al-Hasan, 2013; Chon et al., 2014; Ascione et al., 2015; Bagheri-Nesami et al., 2015). Among these, bloodstream infection (BSI) is most serious, because it can rapidly deteriorate into sepsis, severe sepsis, or septic shock. BSIs have become a major cause of death in European intensive care units, incurring a mortality rate of 30–50% (Vincent et al., 2006). To improve this prognosis, timely initiation of effective antimicrobial therapy is essential (Kumar et al., 2006; Vincent et al., 2006; Dellinger et al., 2013; Chon et al., 2014; Ascione et al., 2015; Bagheri-Nesami et al., 2015). Currently, culture remains the gold standard of infection diagnosis. However, standard final identification (ID) and ASTs are reported 16–48 h after a positive alert. During this delay, the clinician must administer an empirical antimicrobial therapy, typically a broad-spectrum antibiotic or an antibiotic cocktail to cover all likely pathogens. However, inappropriate antimicrobial therapy will worsen the outcome (Ibrahim et al., 2000; Tumbarello et al., 2010; Cain et al., 2015). Moreover, the long-term use of dispensable broad-spectrum antibiotics promotes antibiotic resistance and spread, increases cost and lengthens hospital stays (Blot et al., 2002; Tumbarello et al., 2010).

Matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF MS) promises a revolutionary breakthrough in clinical microbiology. The technology identifies bacteria within 6 min (Seng et al., 2009), with a species-level accuracy of 84.1–93.6% (Bizzini and Greub, 2010) and high sensitivity ( $\sim 10^5$  CFU). Currently, MALDI-TOF MS has shown the ability to distinguish vanB-positive *Enterococcus faecium* from isolates that do not possess this resistance gene (Griffin et al., 2012) and demonstrate the carbapenemase activity by detection of meropenem and the relevant degradation products (Hrabak et al., 2011). Another study focused on a proteome analysis of ampicillin-resistant *Fusobacterium nucleatum* was reported (Al-Haroni et al., 2008). Moreover, MALDI-TOF MS fingerprinting has been used to distinguish the different expression levels of cell wall components between resistant isolates and sensitive isolates (Xu et al., 2006). Recently, RMI by MALDI-TOF MS has been adopted in various protocols, including the commercial Sepsityper kit (Bruker Daltonics, Bremen, Germany; Martiny et al., 2012; Hazelton et al., 2014; Idelevich et al., 2014; Martinez et al., 2014; Schieffer et al., 2014; Morgenthaler and Kostrzewska, 2015), serum separator tubes (Stevenson et al., 2010; Schubert et al., 2011) and in-house methods (Martiny et al., 2012). These techniques return the RMI from positive bottles within 30 min to a few hours. They also reliably identify species-level GN and GP bacteria (with accuracies of 90 and 76%, respectively) and yeasts (66%) (Morgenthaler and Kostrzewska, 2015). However, the Sepsityper kit is expensive, whereas the in-house method involves multi-step washing/centrifugation and yields relatively low RMI accuracy for GP bacteria and yeasts.

Rapid ASTs have been attempted in several studies. For instance, methicillin-resistant staphylococci and vancomycin-resistant enterococci have been detected in real-time PCR-melt curve analysis (Chan et al., 2015). MALDI-TOF MS can detect subtle differences in isogenic *Staphylococcus aureus*, which determine the organism's resistance to methicillin or teicoplanin (Majcherczyk et al., 2006), ampicillin susceptibility (Grundt et al., 2012), and carbapenemase (Hrabák et al., 2012). All of the above methods are efficient; however, molecular methods cannot identify the expression of resistance genes, and other methods are limited to one or a few specific antibiotic determination profiles.

To overcome these difficulties, we determined the RMI and RMAST by a combined MALDI-TOF MS/Vitek AST system. Bacterial cells were directly extracted from positive sterile body fluid cultures in serum separator tubes, and proliferated in brain heart infusion broth (BHI) for 2–4 h.

This study aims to evaluate the reliability and accuracy of this protocol in fast pathogen diagnosis without additional costs and efforts.

## MATERIALS AND METHODS

### Location

This study was conducted in Huashan Hospital (affiliated with Fudan University), located in the center of Shanghai, China. Huashan Hospital is one of the largest (1300 beds) comprehensive teaching hospitals in China, handling ~8000 admissions per day.

### Clinical Samples

Sterile body fluids, including blood, cerebrospinal fluid (CSF), pleural fluid, ascitic fluid, pericardial effusion, joint cavity fluid, and vitreous fluid, were injected into blood culture bottles to improve the positive rate of clinical samples. From September 2014 to August 2015, we enrolled the first positive culture from each patient. Thus, we collected 485 non-duplicated positive cultures.

### Blood Culture (BC) Bottles and BC System

The BC bottles were BacT/Alert aerobic/SA and anaerobic/SN (bioMérieux, Marcy l'Etoile, France), and BACTEC Plus Aerobic/F, BACTEC Plus Anaerobic/F and BACTEC Mycosis-IC/F (Becton Dickinson, USA; Figure S1). All bottles were incubated in automated BC systems (BacT/Alert 3D, bioMérieux and BACTEC FX, Becton Dickinson; Figure S1) until they tested positive. Positive cultures were analyzed immediately after a positive alert during laboratory hours (8 a.m. to 5 p.m.). Cultures that became positive later than this period were stored in the BC systems and analyzed the next morning.

### Positive Cultures Processing for RMI Analysis and RMAST

Based on the Gram staining results, 5 ml (sterile body fluid positive culture and blood positive culture caused by GN or polymicrobial) or 10 ml (blood positive culture caused by other bacteria) were drawn into one or two serum separator tubes

(BD Vacutainer SSTII Advance, USA). Bacteria were pelleted by centrifugation at 4000 g for 10 min. The bacteria became aggregated at the surface of the polymeric gel, while the red-blood-cell component was sedimented beneath the gel layer. The supernatant was discarded and the bacterial pellet was gently resuspended in 1 ml sterile distilled water, without disrupting the gel layer. The suspension was transferred to a sterile Eppendorf tube, mixed thoroughly, and centrifuged at 16,000 g for 1 min. The supernatant was discarded and the washing/centrifugation steps were repeated once. The pellets were resuspended in 100  $\mu$ l sterile distilled water. A portion of the resuspension was drawn and subjected to an ethanol/formic acid extraction procedure. The rest of the resuspension was prepared for 0.5 McFarland (McF) and subjected to RMAST as described for standard AST.

Prior to the ethanol/formic acid extraction, pellets of blood cultures caused by other bacteria (but not those caused by GN or polymicrobial) were resuspended in 1 ml BHI for enrichment. GP bacteria and yeasts were inoculated at 37°C with shaking at 200 rpm for 2 and 4 h, respectively. The enrichments were pelleted at 16,000 g for 1 min. After discarding the supernatant, the pellets were washed with 1 ml sterile distilled water, and re-centrifuged at 16,000 g for 1 min. Again, the supernatant was discarded.

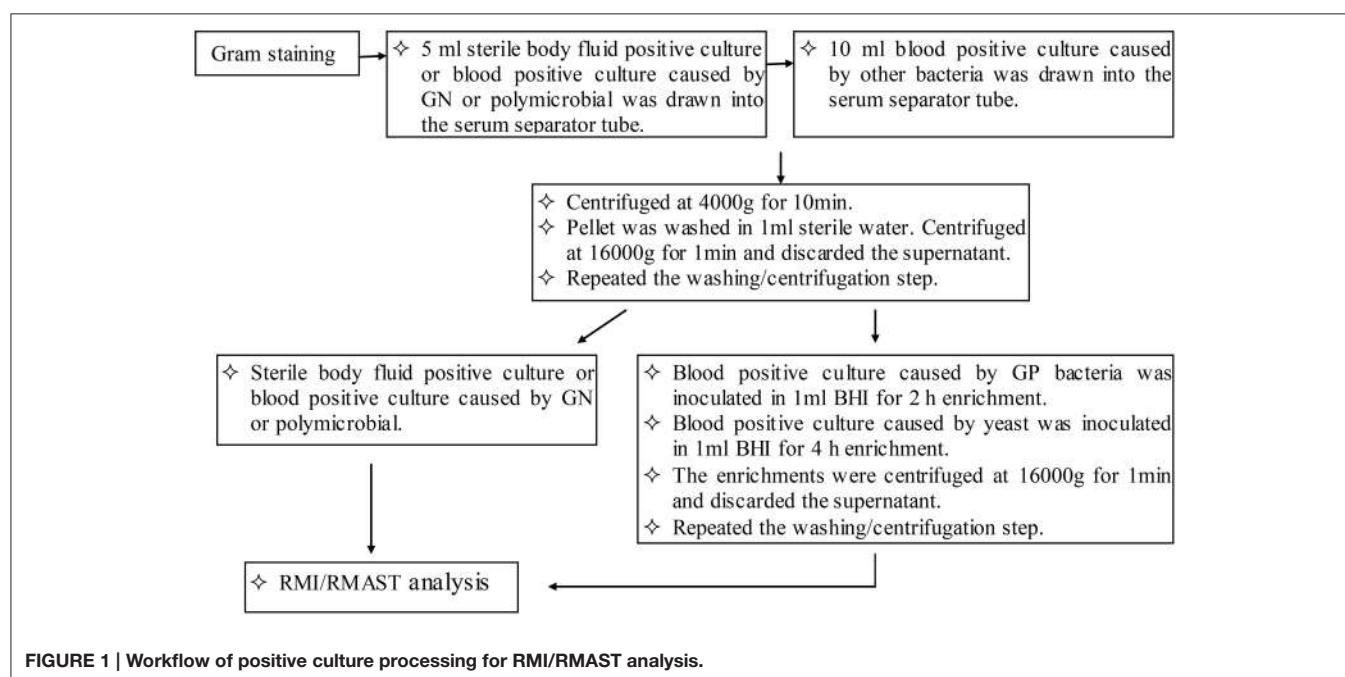
In the ethanol/formic acid extraction procedure, a portion of the bacterial pellet was resuspended in 300  $\mu$ l water by vortexing. The suspension was thoroughly mixed with 900  $\mu$ l absolute ethanol and then centrifuged at 16,000 g for 1 min. The supernatant was discarded and the residual ethanol was removed after a repeat centrifugation. The cell pellet was air dried and dissolved in 30  $\mu$ l of 70% formic acid by thorough vortexing. After adding 30  $\mu$ l acetonitrile, the dissolved pellet was centrifuged at 16,000 g for 2 min and 1  $\mu$ l supernatant was spotted onto a steel target plate for

MALDI-TOF MS (Bruker Daltonics, Bremen, Germany) analysis (Figure 1).

## Standard Identification and Antimicrobial Susceptibility Testing

The positive broths were sub-cultured onto routine plates of 5% sheep blood agar, chocolate, or anaerobic blood agar. Bacteria that failed to grow on the routine plates were inoculated on self-made plates composed of 20 ml sterile blood culture broth (bioMérieux, Marcy l'Etoile, France; Becton Dickinson, USA) and 3.9% agar powder (Oxoid, Thermo Fisher Scientific, England). Plates were grown in the incubator (Thermo Scientific Forma, USA) at 35°C in 5% CO<sub>2</sub> or an anaerobic atmosphere until visible colonies appeared. A pure bacterial colony was smeared onto a steel target plate for identification by MALDI-TOF MS. For yeasts, 1  $\mu$ l of formic acid was added to the plate and air dried for 5 min.

In the antimicrobial susceptibility testing, Vitek cards AST-GN13, AST-Gp67, and AST-Gp68 were used for GN bacteria, staphylococci/enterococci and *Streptococcus pneumoniae*, respectively. Other streptococci and yeasts were tested with ATB-STREP5 and ATB-FUNGUS3 (bioMérieux, Marcy l'Etoile, France), respectively. The GN bacteria, staphylococci, enterococci, streptococci, or yeast pellets were dissolved in 0.45% saline solution to prepare 0.5–0.63 McF or 2 McF suspension, respectively. The Vitek card AST-GN13 or AST-Gp67/AST-Gp68 were filled with suspension composed of 3 ml 0.45% saline solution and 145 or 280  $\mu$ l 0.5–0.63 McF suspension, respectively. The ATB-STREP5 strip was filled with suspension composed of ATB S medium and 200  $\mu$ l 0.5 McF suspension. The ATB-FUNGUS3 strip was filled with suspension composed of ATB F2 medium and 20  $\mu$ l 2McF suspension.



## Matrix-Assisted Laser Desorption/Ionization-Time of Flight Mass Spectrometry

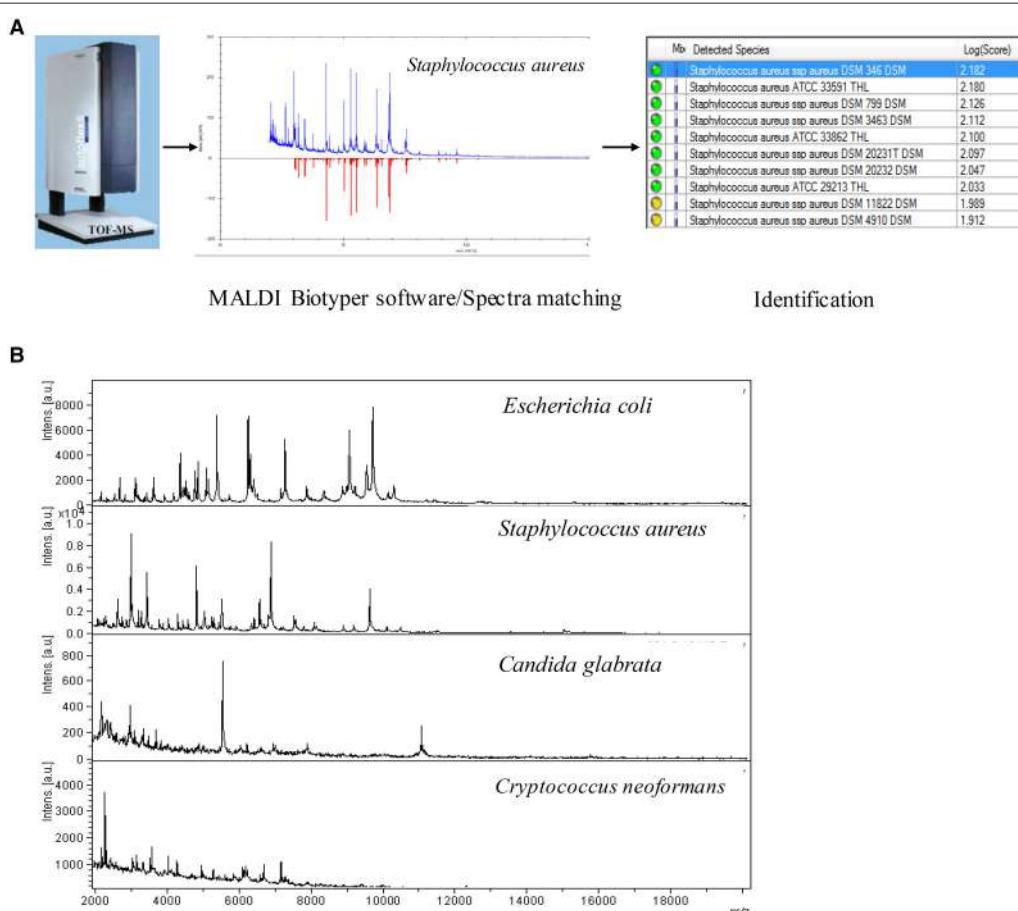
The spot was overlaid with 1  $\mu$ l MALDI matrix (a saturated solution of  $\alpha$ -cyano-4-hydroxycinnamic acid in 50% acetonitrile–2.5% trifluoroacetic acid). After drying at room temperature for 5 min, sample was subjected to analysis of the bacteria protein using MALDI-TOF MS system. The spectrum was obtained in linear positive-ion mode range from 2000 to 20 000 Da. Each spot was measured manually on five different positions by using 1000 laser shots at 25 Hz in groups of 40 shots. The MALDI Bruker Biotype 3.0 software and library (Bruker Daltonics) were used for spectra analysis. According to the spectra matching and score criterion, MALDI Bruker Biotype 3.0 software obtained appropriate scores and identification results. The workflow was shown in Figure 2A.

## Interpretation of RMI and RMAST

RMI results were scored by the manufacturer's standard criterion (cut-off values of 1.7 and 2.0 for acceptable identification to the genus and species levels, respectively) and by the modified criterion (cut-off values of 1.5 and 1.8 for at least

three identical results in the list at the genus and species levels, respectively). These criteria have been described in previous studies (Schmidt et al., 2012; Machen et al., 2014; Martinez et al., 2014; Morgenthaler and Kostrzewska, 2015). The RMI results were compared with the final reports. Non-identifiable samples included correctly identified samples with scores below the genus level cut-off value and samples that presented no peak or a very weak signal. Discordant identification included samples that the RMI results were inconsistent with the final reports which proved correct. Discrepancies were resolved by 16S rRNA (for bacteria; primer F, 5'-AGAGTTGATGATGGCTCAG-3'; primer R, 5'-ACCGCAACTGCTGGCAC-3'; expected PCR product size: ~800 bp) or 18S rRNA (for fungi; primer F, 5'-GATACCGTCGTAGTCTTA-3'; primer R, 5'-ATTCCCTCGTTGAAGAGC-3'; expected PCR product size: ~800 bp) amplification. DNA was extracted from colonies sub-cultured for 24 h using Genomic DNA isolation kit (Sangon Biotech, Shanghai, China). Thermal cycler conditions were 94°C for 5 min, followed by 35 cycles of 94°C (30 s), 55°C (30 s), and 72°C (30 s), with a final extension at 72°C (7 min).

RMAST results were compared with those obtained from the standard method. The minimum inhibitory concentrations (MIC) obtained by both methods were translated into clinical



**FIGURE 2 | (A)** The RMI workflow of MALDI-TOF MS system. **(B)** RMI spectra of representative bacterial species.

categories (susceptible, intermediate, resistant), following the CLSI recommendations. The comparison between the direct and standard inoculation methods was categorized as agreement, very major error (VME, false susceptibility), major error (ME, false resistance), or minor error (mE, susceptible/resistant versus intermediate susceptibility). Discrepancies in MICs were resolved by broth dilution methods according to the CLSI (2015) guidelines.

## Evaluation of Clinical Relevance

To prevent the development of resistance, to reduce toxicity, and to reduce costs, the antimicrobial regimen should be ceased when detected typical contaminants (Monoculture grew Coagulase-negative *staphylococcus* among multiple cultures, *Micrococcus* spp, *Corynebacterium* spp, *Bacillus* spp, *Propionibacterium acnes*) and assessed for potential de-escalation (De-escalation refers to narrow the spectrum of antimicrobial coverage and choose the most appropriate single-agent therapy). Antibiotic therapy should be installed when detected pathogenic species that cannot be distinguished by Gram staining and modified when detected intrinsic resistant bacteria (Bacteria were intrinsic antibiotic-resistance. *Stenotrophomonas maltophilia*, *Enterococcus casseliflavus*, and *Candida glabrata* are intrinsic resistant to carbapenem, vancomycin, and fluconazole, respectively) or

acquired resistant bacteria (Antibiotic-resistance was mediated by plasmid, resistant enzymes, or other resistance mechanisms). RMI/RMAST results were reported to the clinician. To assess the impact of our approach on the management of sterile body infections, we determined whether the results can instructively recommend an optimal therapeutic scheme. Unreliable results were excluded. Indications leading to a treatment change are listed in Table 1.

## Statistical Analysis

Statistical time comparisons were conducted by a Wilcoxon signed-rank test (GraphPad Prism 5.0, CA, USA).  $P < 0.05$  was considered statistically significant. All statistical tests were two-tailed.

## RESULTS

### Rapid Microorganism Identification

Regardless of enrichment, the RMI test of each sample was completed in 40 min (15 min for the pellet collection/washing/centrifugation steps, 15 min for the extraction procedure, 5 min for the sample spotting/drying steps, 5 min for MALDI-TOF MS measurement). The RMI was evaluated in 485 non-duplicated positive cultures

**TABLE 1 |** Indications for optimized antibiotic therapy and recommendations based on RMI /RMAST results.

	RMI result (0.58–4.58 h)	RMAST result (8.4–34 h)
Cessation of antibiotic therapy	Detection of typical contaminants: * Monoculture grew Coagulase-negative <i>staphylococcus</i> among multiple cultures * <i>Micrococcus</i> spp. * <i>Corynebacterium</i> spp. * <i>Bacillus</i> spp. * <i>Propionibacterium acnes</i>	
Installation of antibiotic therapy	Detection of pathogenic species that cannot be distinguished by Gram staining: * <i>Staphylococcus aureus</i> * <i>Staphylococcus lugdunensis</i> * <i>Listeria monocytogenes</i>	Assessment and verification the initiated measure.
Modification of antibiotic therapy	Detection of intrinsic resistant bacteria: * <i>Stenotrophomonas maltophilia</i> with carbapenem resistance * <i>Enterococcus casseliflavus</i> with vancomycin resistance * <i>Candida glabrata</i> with fluconazole resistance	Assessment and verification the initiated measure.
De-escalation antibiotic therapy		De-escalation of broad spectrum antibiotics or last-resort antibiotics (carbapenems, vancomycin, linezolid) or combination therapy in case of results indicates low-risk intrinsic resistant bacteria (e.g., <i>Escherichia coli</i> , <i>Staphylococci</i> ) associated with negative acquired antibiotic-resistant history.
Modification of antibiotic therapy		Detection of acquired resistant bacteria (e.g., carbapenem resistant GN bacteria, vancomycin, or linezolid non-susceptible <i>Enterococci</i> and <i>Staphylococci</i> ).

(**Table 2**). The compositions of the positive cultures are listed in Table S1.

The RMI spectra of representative bacterial species were shown in **Figure 2B**. Scored by the standard criterion, the correct identification rates of RMI were 96.3% for GN bacteria (182/189), 82.8% for GP bacteria (168/203), 11.8% for fastidious bacteria (2/17), 62.2% for fungi (23/37), and 89.5% for anaerobic bacteria (17/19). In cultures of polymicrobial, the dominant bacterium was correctly identified in 81.2% (13/16) of cases and bacteria that failed to grow on the routing plates were identified with a success rate of 50.0% (2/4). Scored by the modified criterion, the correct identification rates of RMI were 98.9% (187/189) for GN bacteria, 87.2% (177/203) for GP bacteria, 11.8% (2/17) for fastidious bacteria, 75.7% (28/37) for fungi, and 94.7% (18/19) for anaerobic bacteria (18/19). In cultures containing polymicrobial and bacteria that failed to grow on the routing plates, the correct identification rates were 81.2% (13/16; dominant species) and 100.0% (4/4), respectively. The modified criterion provided more accurate RMI scores than the standard criterion, with no relevant misidentification at the genus level (**Tables 2, 3**). The RMI accuracy was lower in blood samples than in other sterile body fluids (Table S1). The mean RMI scores and the RMI results of polymicrobial are listed in **Tables 2, 3**, respectively.

## Rapid Antimicrobial Susceptibility Test

Among the 320 non-duplicated positive cultures, 140 GN bacteria, 105 *Staphylococcus* spp., 28 *Enterococcus* spp., 14 *Streptococcus* spp., 33 fungi, and 3349 bacterial /antimicrobial combinations, were analyzed in the RMAST investigation (**Table 4**).

The RMAST was actually performed on 142 GN bacteria (109 Enterobacteriaceae, 14 *Acinetobacter* spp., and 19 non-fermentative bacteria), but two cultures (1 *Acinetobacter junii*, 1 *Brevundimonas diminuta*) grew poorly so were excluded from the analysis. In total, 1828 bacterial/antimicrobial combinations were analyzed with a category agreement of 96.77%, a mE rate of 2.52%, an ME rate of 0.22%, and a VME rate of 0.49%. Errors were found in *Escherichia coli* (1.09% mE, 0.05% ME, 0.27% VME), *Klebsiella pneumoniae* (0.66% mE, 0.05% ME, 0.71% VME), *Enterobacter cloacae* (0.22% mE), *Enterobacter aerogenes* (0.05% mE), *Morganella morganii* (0.16% mE), *Raoultella ornithinolytica* (0.11% mE), *Providencia rettgeri* (0.05% mE), *Pseudomonas aeruginosa* (0.22% mE, 0.05% ME, 0.11% VME), and other non-fermenters (0.22% mE, 0.11% ME, 0.05% VME).

Although, RMAST results were available for all 28 *Enterococcus* spp., they were available for only 105 of the 114 *Staphylococcus* spp.; 10 cases failed because of poor growth (4 *Staphylococcus epidermidis*, 2 *S. hominis*, 1 *S. capitis*, 1 *S. auricularis*, 1 *S. simulans*, and 1 *S. caprae*). In total, 1246 bacterial/antimicrobial combinations were analyzed with a category agreement of 93.50%, an mE rate of 3.69%, an ME rate of 0.72%, and a VME rate of 2.09%. Errors were found in *S. aureus* (0.32% mE, 0.16% VME), Coagulase-negative *staphylococcus* (1.28% mE, 0.56% ME, 1.76% VME), and *Enterococci* spp. (0.32% mE, 0.16% ME, 0.16% VME).

RMAST was originally performed on 21 *Streptococcus* spp., including 5 *S. pneumoniae*, 3 *S. agalactiae*, 1 *S. pyogenes*, and 12

*S. viridans*, but 7 of the *S. viridans* failed because of poor growth, yielding 14 RMAST results. In total, 140 bacterial/antimicrobial combinations were analyzed with a category agreement of 98.57%, an ME rate of 0.71%, and a VME rate of 0.71%.

RMAST results were available for 33 of the 37 fungal isolates (7 *Candida albicans*, 6 *C. tropicalis*, 3 *C. parapsilosis*, 3 *C. glabrata*, 1 *C. lusitaniae*, 1 *Trichosporon asahii*, 15 *Cryptococcus neoformans*, and 1 *C. albidus*). Four cases (1 *C. albicans*, 1 *C. parapsilosis*, 1 *C. glabrata*, and 1 *Cryptococcus albidus*) grew poorly and were thus excluded. In total, 90 bacterial/antimicrobial combinations were analyzed with a category agreement of 95.56%, an mE rate of 3.33% and a VME rate of 1.11% in the yeast group. Forty-five bacterial/antimicrobial combinations were also analyzed with a category agreement of 100% in the *C. neoformans* group.

Forty-three multi-drug resistant bacteria were detected in this study, including 1 vancomycin-intermediate *Staphylococcus epidermidis*, 1 vancomycin-resistant *Enterococcus faecium*, 1 linezolid-resistant *S. aureus*, 8 linezolid-resistant *S. capitis*, 16 carbapenem non-susceptible Enterobacteriaceae, 8 carbapenem non-susceptible *Acinetobacter* spp., and 6 carbapenem non-susceptible non-fermenting spp. The RMAST category agreement of multi-drug resistant bacteria was 97.68%, with 1 mE in carbapenem non-susceptible Enterobacteriaceae.

## Time to Identification and Antimicrobial Susceptibility Testing

The average times of the RMI (vs. final identification report) were 0.58 h for GN, polymicrobial and sterile body fluid bacteria (vs. 18.1 h), 2.58 h for GP bacteria (vs. 18.1 h), 3.53 h for fungi (0.58 h in sterile body fluid and 4.58 h in blood, vs. 32.2 h), 1.36 h for anaerobic bacteria (0.58 h for GN and 2.58 h for GP bacteria, vs. 27.8 h), and 2.5 h for bacteria that failed to grow on the routing plates (vs. 64 h). The results are presented in **Figure 3**.

The average times of RMAST (vs. the conventional method) were  $8.4 \pm 2.6$  h for GN bacteria (vs.  $26.4 \pm 2.6$  h;  $p < 0.0001$ ),  $13.1 \pm 3.0$  h for GP bacteria (vs.  $31.1 \pm 3.0$  h;  $p < 0.0001$ ), 34 ± 12 h for fungi (vs.  $58 \pm 12$  h;  $p < 0.0001$ ; **Figure 3**).

## Analysis of Clinical Relevance

Among 485 RMI cases (**Figure 4**), 14.85% of the results indicated that clinicians should adjust their treatment. Specifically, antibiotic therapy should be ceased in 3.71% of patients, initiated in 9.28% of patients, and modified in 1.86% of patients. Among the 320 RMAST results (**Figure 4**), 65% indicated that clinicians could optimize the therapeutic regimen by de-escalating (51.6%) or modifying (13.44%) the antibiotic therapy.

## DISCUSSION

Sterile body fluid infections tend to trigger systemic or local tissue damage. Because of their severity and morbidity, BSIs and acute bacterial meningitis have placed a great burden on health care settings (Adhikari et al., 2010; Thigpen et al., 2011; Goto and Al-Hasan, 2013; Portnoy et al., 2015). Early diagnosis and rapid intervention are critical to improving patient prognosis.

This study incorporated a combined MALDI-TOF MS/Vitek AST system into the clinical microbiology laboratory workflow.

**TABLE 2 |** PMI results scored by standard criterion and modified criterion.

Species	Standard cut-off values						Modified cut-off values						Not identified	Discordant ID	Mean scores (min-max)			
	No ID		Correct ID		Not identified		No ID		Correct ID		Not identified							
	n (%)	species	genus	Score ≥ 2.0	Score ≥ 1.7	n (%)	species	genus	Score ≥ 1.8	Score ≥ 1.5	n (%)	species						
<b>AEROBIC GN BACTERIA</b>																		
<b>Enterobacteriaceae</b>																		
<i>Escherichia coli</i>	63 (100)	55 (87.3)	8 (12.7)	0	0	63 (100)	57 (90.5)	6 (9.5)	0	0	2,097 (1,712–2,417)							
<i>Klebsiella pneumoniae</i>	38 (100)	33 (86.8)	5 (13.2)	0	0	38 (100)	37 (97.4)	1 (2.6)	0	0	2,076 (1,799–2,380)							
<i>Klebsiella oxytoca</i>	1 (100)	1 (100)	0	0	0	1 (100)	1 (100)	0	0	0	2,261							
<i>Enterobacter cloacae</i>	7 (100)	4 (57.1)	3 (42.9)	0	0	7 (100)	5 (71.4)	2 (28.6)	0	0	1,952 (1,762–2,087)							
<i>Enterobacter aerogenes</i>	1 (100)	1 (100)	0	0	0	1 (100)	1 (100)	0	0	0	2,183							
<i>Serratia marcescens</i>	4 (100)	2 (50.0)	2 (50.0)	0	0	4 (100)	2 (50.0)	2 (50.0)	0	0	1,928 (1,733–2,154)							
<i>Morganella morganii</i>	3 (100)	3 (100)	0	0	0	3 (100)	3 (100)	0	0	0	2,217 (2,059–2,415)							
<i>Proteus mirabilis</i>	3 (100)	2 (66.7)	1 (33.3)	0	0	3 (100)	3 (100)	0	0	0	2,079 (1,828–2,233)							
<i>Proteus vulgaris</i>	2 (100)	1 (50.0)	1 (50.0)	0	0	2 (100)	1 (50.0)	1 (50.0)	0	0	1,901 (1,787–2,015)							
<i>Salmonella</i>	2 (100)	0	2 (100)	0	0	2 (100)	1 (50.0)	1 (50.0)	0	0	1,752 (1,703–1,800)							
<i>Pantoea</i>	2 (100)	1 (50.0)	1 (50.0)	0	0	2 (100)	2 (100)	0	0	0	2,079 (1,828–2,233)							
<i>Raoultella ornithinolytica</i>	2 (100)	1 (50.0)	1 (50.0)	0	0	2 (100)	1 (50.0)	1 (50.0)	0	0	2,083 (1,794–2,372)							
<i>Providencia rettgeri</i>	2 (100)	0	2 (100)	0	0	2 (100)	0	2 (100)	0	0	1,720 (1,712–1,727)							
<i>Providencia stuartii</i>	1 (100)	0	1 (100)	0	0	1 (100)	1 (100)	0	0	0	1,890							
<i>Citrobacter freundii</i>	2 (100)	2 (100)	0	0	0	2 (100)	2 (100)	0	0	0	2,044 (2,001–2,087)							
<i>Citrobacter braakii</i>	1 (100)	0	1 (100)	0	0	1 (100)	1 (100)	0	0	0	1,924							
Subtotal	134 (100)	106 (79.1)	28 (20.9)	0	0	134 (100)	118 (88.1)	16 (11.9)	0	0	0							
<b>ACINETOBACTER spp</b>																		
<i>Acinetobacter baumannii</i>	21 (100)	17 (81.0)	4 (19.0)	0	0	21 (100)	18 (85.7)	3 (14.3)	0	0	2,054 (1,738–2,237)							
<i>Acinetobacter junii</i>	3 (100)	2 (66.7)	1 (33.3)	0	0	3 (100)	2 (66.7)	1 (33.3)	0	0	1,947 (1,720–2,087)							
<i>Acinetobacter haemolyticus</i>	3 (100)	0	3 (100)	0	0	3 (100)	2 (66.7)	1 (33.3)	0	0	1,755 (1,702–1,817)							
<i>Acinetobacter pittii</i>	1 (100)	0	1 (100)	0	0	1 (100)	1 (100)	0	0	0	1,826							
Subtotal	28 (100)	19 (67.9)	9 (32.1)	0	0	28 (100)	23 (82.1)	5 (17.9)	0	0	0							
<b>NON-FERMENTERS</b>																		
<i>Pseudomonas aeruginosa</i>	12 (100)	8 (66.7)	4 (33.3)	0	0	12 (100)	10 (83.3)	2 (16.7)	0	0	2,054 (1,738–2,237)							
<i>Stenotrophomonas maltophilia</i>	5 (20.0)	0	1 (20.0)	4 (80.0)	0	5 (20.0)	3 (60.0)	1 (20.0)	1 (20.0)	0	N/A							
<i>Pseudomonas fulva</i>	1 (100)	0	1 (100)	0	0	1 (100)	1 (100)	0	0	0	1,994							
<i>Myrodes odoratim</i>	1 (100)	1 (100)	0	0	0	1 (100)	1 (100)	0	0	0	2,206							
<i>Pseudomonas ozzhhabitans</i>	1 (100)	0	1 (100)	0	0	1 (100)	1 (100)	0	0	0	1,878							

(Continued)

TABLE 2 | Continued

Species	Standard cut-off values						Modified cut-off values						Not identified	Discordant ID	Mean scores (min-max)			
	No ID		Correct ID		Not identified		No ID		Correct ID		Not identified							
	n (%)	species	genus	Score ≥ 2.0	Score ≥ 1.7	n (%)	species	genus	Score ≥ 1.8	species	genus	Score ≥ 1.5						
<i>Brevundimonas diminuta</i>	1 (0)	0	0	1 (100)	0	1 (0)	0	0	0	1 (100)	0	0	0	0	N/A			
<i>Aeromonas hydrophila</i>	2 (100)	1 (50.0)	1 (50.0)	0	0	2 (100)	2 (100)	0	0	0	0	0	0	0	1.986 (1.927–2.045)			
<i>Aeromonas caviae</i>	1 (0)	0	0	1 (100)	0	0	1 (100)	0	0	1 (100)	0	0	0	0	1.682			
<i>Aeromonas veronii</i>	1 (100)	0	0	1 (100)	0	0	1 (100)	0	0	1 (100)	0	0	0	0	1.765			
<i>Methylbacterium</i>	1 (0)	0	0	1 (100)	0	0	1 (100)	0	0	1 (100)	0	0	0	0	1.623			
<i>Vibrio cholerae</i>	1 (100)	0	0	1 (100)	0	0	1 (100)	0	0	1 (100)	0	0	0	0	1.876			
Subtotal	27 (74.0)	10 (37.0)	10 (37.0)	7 (26.0)	0	27 (92.6)	19 (70.4)	6 (22.2)	2 (7.4)	0	0	0	0	0				
Total	189 (96.3)	135 (71.4)	47 (24.9)	7 (3.7)	0	189 (98.9)	160 (84.7)	27 (16.9)	2 (1.1)	0	0	0	0	0				
<b>AEROBIC GP BACTERIA</b>																		
<b>Staphylococci</b>																		
<i>Staphylococcus aureus</i>	41 (100)	19 (46.3)	22 (53.7)	0	0	41 (100)	31 (75.6)	10 (24.4)	0	0	0	0	0	0	1.989 (1.700–2.410)			
<i>Staphylococcus epidermidis</i>	29 (82.8)	6 (20.7)	18 (62.1)	3 (10.3)	2 (6.9)	29 (86.2)	14 (48.3)	11 (37.9)	2 (6.9)	2 (6.9)	2 (6.9)	2 (6.9)	2 (6.9)	2 (6.9)	1.850 (1.658–2.098)			
<i>Staphylococcus capitis</i>	28 (85.7)	6 (21.4)	18 (64.3)	2 (6.9)	2 (6.9)	28 (89.3)	12 (42.6)	13 (46.4)	1 (3.8)	1 (3.8)	2 (6.9)	2 (6.9)	2 (6.9)	2 (6.9)	1.865 (1.616–2.159)			
<i>Staphylococcus hominis</i>	23 (82.6)	4 (17.4)	15 (65.2)	2 (8.7)	2 (8.7)	23 (82.6)	9 (39.1)	10 (43.5)	2 (8.7)	2 (8.7)	2 (8.7)	2 (8.7)	2 (8.7)	2 (8.7)	1.857 (1.703–2.195)			
<i>Staphylococcus haemolyticus</i>	9 (88.9)	0	8 (88.9)	1 (11.1)	0	9 (88.9)	0	8 (88.9)	1 (11.1)	0	0	0	0	0	1.748 (1.701–1.882)			
<i>Staphylococcus cohnii</i>	3 (66.7)	0	2 (66.7)	1 (33.3)	0	3 (66.7)	0	2 (66.7)	1 (33.3)	0	0	0	0	0	1.769 (1.741–1.797)			
<i>Staphylococcus warneri</i>	1 (100)	0	1 (100)	0	0	1 (100)	1 (100)	1 (100)	0	0	0	0	0	0	1.934			
<i>Staphylococcus xylosus</i>	1 (100)	1 (100)	0	0	0	1 (100)	1 (100)	0	0	0	0	0	0	0	2.068			
<i>Staphylococcus saprophyticus</i>	1 (100)	0	1 (100)	0	0	1 (100)	1 (100)	0	0	0	0	0	0	0	1.920			
<i>Staphylococcus aunicularis</i>	1 (100)	0	1 (100)	0	0	1 (100)	1 (100)	0	0	0	0	0	0	0	1.912			
<i>Staphylococcus simulans</i>	1 (100)	0	1 (100)	0	0	1 (100)	1 (100)	0	0	0	0	0	0	0	1.819			
<i>Staphylococcus caprae</i>	1 (0)	0	0	1 (100)	0	1 (100)	0	0	0	1 (100)	0	0	0	0	N/A			
Subtotal	139 (88.5)	36 (25.9)	87 (62.6)	10 (7.2)	6 (4.3)	139 (89.9)	71 (51.1)	54 (38.8)	8 (5.8)	8 (5.8)	8 (5.8)	8 (5.8)	8 (5.8)	8 (5.8)	6 (4.3)			
<b>ENTEROCOCCI</b>																		
<i>Enterococcus faecium</i>	13 (84.6)	2 (15.4)	9 (69.2)	1 (7.7)	1 (7.7)	13 (92.3)	8 (61.5)	4 (30.8)	0	0	1 (7.7)	1 (7.7)	1 (7.7)	1 (7.7)	1.856 (1.606–2.113)			
<i>Enterococcus faecalis</i>	10 (100)	4 (40.0)	6 (60.0)	0	0	10 (100)	8 (80.0)	2 (20.0)	0	0	0	0	0	0	1.926 (1.742–2.186)			
<i>Enterococcus cassel/savvus</i>	4 (100)	3 (75.0)	1 (25.0)	0	4 (100)	3 (75.0)	1 (25.0)	0	0	0	0	0	0	0	1.989 (1.710–2.097)			
<i>Enterococcus avium</i>	3 (66.7)	0	2 (66.7)	1 (33.3)	3 (66.7)	0	2 (66.7)	1 (33.3)	0	0	0	0	0	0	1.778 (1.760–1.795)			
Subtotal	30 (90.0)	9 (30.0)	18 (60.0)	2 (6.7)	1 (3.3)	30 (93.3)	19 (63.3)	9 (30.0)	1 (3.3)	1 (3.3)	1 (3.3)	1 (3.3)	1 (3.3)	1 (3.3)				

(Continued)

TABLE 2 | Continued

Species	Standard cut-off values						Modified cut-off values				Not identified	Discordant ID	Mean scores (min-max)			
	No ID		Correct ID		Not identified n (%)	Discordant ID n (%)	Correct ID									
	species	genus	Score ≥ 2.0	Score ≥ 1.7			species	genus	Score ≥ 1.8	Score ≥ 1.5						
<b>STREPTOCOCCUS</b>																
<i>Streptococcus pneumoniae</i>	5 (40.0)	0	2 (40.0)	1 (20.0)	5 (60.0)	2 (40.0)	1 (20.0)	1 (20.0)	1 (20.0)	1 (20.0)	0	0	N/A			
<i>Streptococcus agalactiae</i>	3 (100)	0	3 (100)	0	3 (100)	0	0	0	0	0	0	0	1.898 (1.870–1.934)			
<i>Streptococcus pyogenes</i>	1 (0)	0	0	1 (100)	0	1 (0)	0	0	1 (100)	0	0	0	N/A			
<i>Streptococcus salivarius</i>	5 (60.0)	1 (20.0)	2 (40.0)	2 (40)	5 (80.0)	1 (20)	3 (60.0)	1 (20)	3 (60.0)	1 (20)	0	0	1.862 (1.726–2.088)			
<i>Streptococcus sanguis</i>	2 (0)	0	0	2 (100)	0	2 (50.0)	0	1 (50.0)	1 (50)	1 (50)	0	0	N/A			
<i>Streptococcus mitis</i>	2 (100)	0	2 (100)	0	2 (100)	0	2 (100)	0	2 (100)	0	0	0	1.729 (1.702–1.755)			
<i>Streptococcus gordoni</i>	1 (100)	0	1 (100)	0	0	1 (100)	0	1 (100)	0	0	0	0	1.721			
<i>Streptococcus oralis</i>	1 (0)	0	0	1 (100)	0	1 (100)	0	1 (100)	0	0	0	0	N/A			
<i>Streptococcus massiliensis</i>	1 (0)	0	0	1 (100)	0	1 (0)	0	0	1 (100)	0	0	0	N/A			
Subtotal	21 (52.4)	1 (4.8)	10 (47.6)	9 (42.9)	1 (4.8)	21 (71.4)	6 (28.6)	9 (42.9)	5 (23.8)	5 (23.8)	1 (4.8)					
<b>OTHER GP BACTERIA</b>																
<i>Listeria monocytogenes</i>	4 (100)	4 (100)	0	0	0	4 (100)	0	0	0	0	0	0	2.125 (2.078–2.168)			
<i>Micrococcus</i> spp	2 (50.0)	0	1 (50.0)	0	1 (50.0)	2 (50.0)	0	1 (50.0)	0	0	1 (50.0)	0	1.773			
<i>Corynebacterium</i> spp	6 (16.7)	0	1 (16.7)	5 (83.3)	0	6 (50.0)	0	3 (50.0)	3 (50.0)	0	0	0	N/A			
<i>Bacillus amyloliquefaciens</i>	1 (100)	1 (100)	0	0	0	1 (100)	1 (100)	0	0	0	0	0	2.083			
Subtotal	13 (53.8)	5 (38.5)	2 (15.4)	5 (38.5)	1 (7.7)	13 (69.2)	5 (38.5)	4 (30.8)	3 (23.1)	3 (23.1)	1 (7.7)					
Total	203 (82.8)	51 (25.1)	117 (57.6)	26 (12.8)	9 (4.4)	203 (87.2)	101 (49.8)	76 (37.4)	17 (8.4)	17 (8.4)	9 (4.4)					
<b>FASTIDIOUS BACTERIA</b>																
<i>Haemophilus influenzae</i>	2 (50.0)	0	1 (50.0)	1 (50.0)	0	2 (50.0)	1 (50.0)	0	1 (50.0)	0	0	0	1.888			
<i>Erysipelothrix rhusiopathiae</i>	1 (100)	0	1 (100)	0	0	1 (100)	0	1 (100)	0	0	0	0	1.795			
<i>Actinobacillus</i>	1 (0)	0	0	1 (100)	0	1 (100)	0	0	0	1 (100)	0	0	N/A			
<i>actinomycetemcomitans</i>																
<i>Actinomyces neui</i>	1 (0)	0	0	1 (100)	0	1 (0)	0	0	0	0	1 (100)	0	N/A			
<i>Actinomyces meyeri</i>	1 (0)	0	0	1 (100)	0	1 (0)	0	0	0	0	1 (100)	0	N/A			
<i>Nocardia</i> spp	4 (0)	0	0	4 (100)	0	4 (0)	0	0	0	0	4 (100)	0	N/A			
<i>Streptomyces</i>	1 (0)	0	0	1 (100)	0	1 (0)	0	0	0	0	1 (100)	0	N/A			
<i>Granulicatella adiacens</i>	2 (0)	0	0	2 (100)	0	2 (0)	0	0	0	0	2 (100)	0	N/A			
<i>Moraxella</i> spp	1 (0)	0	0	1 (100)	0	1 (0)	0	0	0	0	1 (100)	0	N/A			
<i>Cardiobacterium hominis</i>	1 (0)	0	0	1 (100)	0	1 (0)	0	0	0	0	1 (100)	0	N/A			
<i>Lactobacillus</i>	1 (0)	0	0	1 (100)	0	1 (0)	0	0	0	0	1 (100)	0	N/A			
<i>Microbacterium liquefaciens</i>	1 (0)	0	0	1 (100)	0	1 (0)	0	0	0	0	1 (100)	0	N/A			
Total	17 (11.8)	0 (0.0)	2 (11.8)	15 (88.2)	0	17 (11.8)	1 (5.9)	1 (5.9)	15 (88.2)	15 (88.2)	0					

(Continued)

TABLE 2 | Continued

Species	Standard cut-off values				Modified cut-off values				Not identified	Discordant ID	Mean scores (min-max)	
	No ID		Correct ID		No ID		Correct ID					
	n (%)	species	genus	Score ≥ 2.0	Score ≥ 1.7	n (%)	species	genus	Score ≥ 1.8	Score ≥ 1.5		
<b>FUNGI</b>												
<i>Candida albicans</i>	7 (42.9)	2 (28.6)	1 (14.3)	3 (42.8)	1 (14.3)	7 (71.4)	3 (42.9)	2 (28.6)	2 (14.3)	1 (14.3)	N/A	
<i>Candida tropicalis</i>	6 (50.0)	0	3 (50.0)	2 (33.3)	1 (16.7)	6 (83.3)	3 (50.0)	2 (33.3)	0	1 (16.7)	1.808 (1.696-1.884)	
<i>Candida glabrata</i>	3 (33.3)	0	1 (33.3)	1 (33.3)	1 (33.3)	3 (33.3)	1 (33.3)	0	1 (33.3)	1 (33.3)	N/A	
<i>Candida parapsilosis</i>	3 (33.3)	0	1 (33.3)	2 (66.7)	0	3 (66.7)	1 (33.3)	1 (33.3)	0	0	N/A	
<i>Candida lusitaniae</i>	1 (100)	0	1 (100)	0	0	1 (100)	0	1 (100)	0	0	1.782	
<i>Cryptococcus neoformans</i>	15 (86.7)	9 (60.0)	4 (26.7)	1 (6.7)	1 (6.7)	15 (86.7)	12 (80.0)	1 (6.7)	1 (6.7)	1 (6.7)	1.928 (1.626-2.237)	
<i>Cryptococcus albidus</i>	1 (0)	0	0	1 (100)	0	1 (0)	0	0	1 (100)	0	N/A	
<i>Trichosporon asahii</i>	1 (100)	1 (100)	0	0	0	1 (100)	1 (100)	0	0	0	2.029	
Total	37 (62.2)	12 (32.5)	11 (29.7)	10 (27.0)	4 (10.8)	37 (75.7)	21 (56.8)	7 (18.9)	5 (13.5)	4 (10.8)		
<b>ANAEROBIC BACTERIA</b>												
<i>Bacteroides fragilis</i>	9 (10)	5 (55.6)	4 (44.4)	0	0	9 (100)	9 (100)	0	0	0	2.016 (1.820-2.283)	
<i>Propionibacterium acnes</i>	3 (100)	2 (66.7)	1 (33.3)	0	0	3 (100)	2 (66.7)	1 (33.3)	0	0	2.020 (1.736-2.217)	
<i>Fusobacterium necrophorum</i>	2 (50.0)	0	1 (50.0)	1 (50.0)	0	2 (50.0)	0	1 (50.0)	1 (50.0)	0	1.798	
<i>Bacteroides uniformis</i>	1 (100)	1 (100)	0	0	0	1 (100)	1 (100)	0	0	0	2.008	
<i>Veillonella</i>	1 (100)	1 (100)	0	0	0	1 (100)	1 (100)	0	0	0	2.038	
<i>Staphylococcus</i>	1 (100)	0	1 (100)	0	0	1 (100)	1 (100)	0	0	0	1.828	
<i>saccharolyticus</i>	1 (100)	0	1 (100)	0	0	1 (100)	1 (100)	0	0	0		
<i>Prevotella bivia</i>	1 (0)	0	0	1 (100)	0	1 (100)	1 (100)	0	0	0	1.894	
<i>Campylobacter fetus</i>	1 (0)	0	0	1 (100)	0	1 (100)	0	1 (100)	0	0	N/A	
Total	19 (89.5)	9 (47.4)	8 (42.1)	2 (10.5)	0	19 (94.7)	15 (78.9)	3 (15.8)	1 (5.3)	0		
Polymicrobial	16 (81.2)	5 (31.3)	8 (50.0)	3 (18.8)	0	16 (81.2)	7 (43.8)	6 (37.5)	3 (18.8)	0	1.924 (1.703-2.241)	
<b>BACTERIA GROW IN THE SPECIAL MEDIUM</b>												
<i>Methylbacterium</i>	1 (100)	0	0	1 (100)	0	1 (100)	0	1 (100)	0	0	1.588	
<i>Streptococcus salivarius</i>	1 (100)	0	1 (100)	0	0	1 (100)	1 (100)	0	0	0	1.573	
<i>thermophilus</i>	1 (100)	0	1 (100)	0	0	1 (100)	1 (100)	0	0	0	1.828	
<i>Staphylococcus</i>	1 (0)	0	0	1 (100)	0	1 (100)	0	0	0	0		
<i>saccharolyticus</i>	1 (0)	0	0	1 (100)	0	1 (100)	0	1 (100)	0	0		
<i>Campylobacter fetus</i>	1 (0)	0	0	1 (100)	0	1 (100)	0	1 (100)	0	0	N/A	
Total	4 (50.0)	0 (0.0)	2 (50.0)	2 (50.0)	0	4 (100)	2 (50.0)	2 (50.0)	0 (0.0)	0		

FID, final identification; FAST, final AST; N/A, not available.

**TABLE 3 | RMI results of polymicrobial.**

Species	Correct ID	Scores
<i>Stenotrophomonas maltophilia/Staphylococcus haemolyticus</i>	<i>Stenotrophomonas maltophilia</i>	1.732
<i>Klebsiella pneumoniae/Enterococcus faecalis</i>	<i>Klebsiella pneumoniae</i> <i>Enterococcus faecalis</i>	2.316/2.482
<i>Enterobacter cloacae/Streptococcus oralis</i>	<i>Not identified (very weak signal)</i>	N/A
<i>Candida tropicalis/Acinetobacter baumannii</i>	<i>Acinetobacter baumannii</i>	1.703
<i>Streptococcus oralis/Staphylococcus epidermidis</i>	<i>Not identified (Misidentified for Staphylococcus warneri)</i>	1.752
<i>Acinetobacter baumannii/Klebsiella pneumoniae</i>	<i>Klebsiella pneumoniae</i>	2.176
<i>Staphylococcus capitis/Klebsiella pneumoniae</i>	<i>Staphylococcus capitis</i>	1.739
<i>Pseudomonas aeruginosa/Klebsiella pneumoniae</i>	<i>Klebsiella pneumoniae</i>	1.824
<i>Pseudomonas aeruginosa/Klebsiella pneumoniae</i>	<i>Klebsiella pneumoniae</i>	2.219
<i>Pseudomonas aeruginosa/Klebsiella pneumoniae</i>	<i>Klebsiella pneumoniae</i>	2.22
<i>Staphylococcus epidermidis/Streptococcus salivarius</i>	<i>Staphylococcus epidermidis</i>	1.902
<i>Enterococcus faecium/Staphylococcus haemolyticus</i>	<i>Not identified (very weak signal)</i>	N/A
<i>Staphylococcus haemolyticus/Staphylococcus capitis</i>	<i>Staphylococcus capitis</i>	1.749
<i>Acinetobacter baumannii/Klebsiella pneumoniae</i>	<i>Klebsiella pneumoniae</i>	2.241
<i>Staphylococcus aureus/Pantoea spp</i>	<i>Staphylococcus aureus</i>	1.715
<i>Acinetobacter baumannii/Klebsiella pneumoniae</i>	<i>Klebsiella pneumoniae</i>	2.241

The aim was to provide clinicians with fast and accurate identification results and multiple precise MIC values for optimizing the therapeutic regimen.

Numerous strategies can identify RMI pathogens from positive blood samples by MALDI-TOF MS (Hazelton et al., 2014; Idelevich et al., 2014; Martinez et al., 2014; Schieffer et al., 2014; Chan et al., 2015; Morgenthaler and Kostrzewska, 2015). The RMI accuracy of GN and GP bacteria or yeasts identified by these strategies has been widely reported, but the investigated species were mostly single. Comprehensive analyses are relatively rare. In this study, we collected 485 non-duplicated positive cultures from blood, csf, ascitic fluid, vitreous fluid, and eye tissues samples, and initiated by GN and GP bacteria, streptococci, fungi, fastidious bacteria, anaerobic bacteria, polymicrobial, and bacteria that failed to grow on the routine plates, and systematically evaluated the accuracy and feasibility of RMI. Data showed the RMI accuracies of CSF, ascitic fluid, vitreous fluid, and eye tissues (almost 100%) were obviously higher than blood sample (57.1–100%), especially for CSF, ascitic fluid, vitreous fluid, and eye tissues samples which might attribute to the blood protein interference. Moreover, the correct RMI rate of GN and GP bacteria reached 98.9 and 87.2%, respectively, similar to the results of previous studies (Hazelton et al., 2014; Schieffer et al., 2014), whereas that of fungi (75.7%) was slightly higher than in previous studies (Idelevich et al., 2014). We attribute this improvement to the 4 h enrichment and constitution of the sample source. The RMI accuracy of csf infections caused by *C. neoformans* was especially high (100%,  $n = 9$ ). Fastidious bacteria yielded the least accurate results because they grow poorly; this problem deserves further exploration. The dominant species in polymicrobial was correctly identified in 81.2% of cases. In addition, to the best of our knowledge, we were the first to characterize the RMIs of rare bacteria that failed to grow on routine plates (including *Methylobacterium*, *Streptococcus salivarius*, *Staphylococcus thermophilus*, *Staphylococcus saccharolyticus*, and

*Campylobacter fetus*). The RMI accuracy reached 100% for these species ( $n = 4$ ), although the sample size was small, and should be increased for an accurate evaluation. The accurate and early detection of salmonella is important for the control and prevention of hospital infections.

A variety of studies have focused on the resistance determinants and activities of resistant enzymes (Majcherczyk et al., 2006; Grundt et al., 2012; Hrabák et al., 2012; Chan et al., 2015), but the obtained data were relatively scarce and unable to provide accurate MIC values to clinicians (Idelevich et al., 2014; Machen et al., 2014). Unlike the former methods, the Vitek AST system is highly sensitive and well-standardized. In this study, the category agreements of GN bacteria, *Staphylococcus* and *Enterococcus* spp., *Streptococcus* spp., and fungi reached 96.77, 93.5, 98.57, and 95.56%, respectively, similarly to previous studies (Romero-Gómez et al., 2012; Wimmer et al., 2012). The percentage of major and very major errors was low among the 3349 bacteria/antimicrobial combinations. In a RMAST of GN bacteria, most of the VMEs (0.49%,  $n = 9$ ) occurred in amikacin, gentamicin, cefazolin, cefotetan, cefepime, and piperacillin-tazobactam, as observed for *E. coli* and *P. aeruginosa* in a previous study (Wimmer et al., 2012). Moreover, the present study obtained VMEs from *K. pneumoniae* and other non-fermenters. In the RMAST of *Staphylococcus* spp., most of the VMEs ( $n = 26$ ) occurred in the presence of gentamicin, levofloxacin, clindamycin, and sulfamethoxazole/trimethoprim. It is worth mentioning that the VITEK ATB AST system could read the data automatically or semi-automatically. The automatic interpretation of low-growth streptococcal and fungal species could be supplemented by manual assistance. Thus, the RMAST category agreement of fungi is obviously higher in our study than in previous study (Idelevich et al., 2014). Importantly, the RMAST category agreement of *C. neoformans* reached 100%.

Accurate RMI/RMAST results are useful for optimizing a therapeutic regimen. First, the RMI and RMAST results are

**TABLE 4 | Bacterial/antimicrobial combinations and errors in bacterial isolates from positive-culture sterile body fluids.**

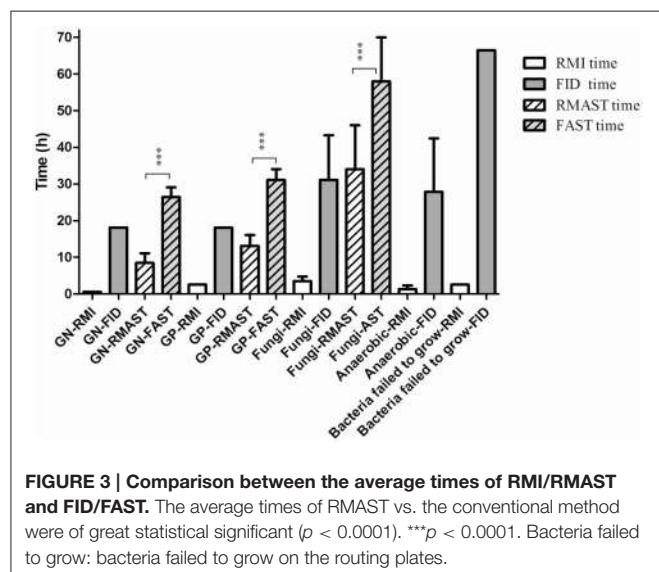
Antimicrobial agent	No	Category agreement		mE	ME		VME	
<b>GN BACTERIA (n = 140)</b>								
Amikacin	122	120	98.36%	1	0.82%		1	0.82%
Gentamicin	135	131	97.04%	3	2.22%		1	0.74%
Ciprofloxacin	135	131	97.04%	3	2.22%	1	0.74%	
Ampicillin	117	117	100.0%					
Cefazolin	135	127	94.07%	7	5.19%		1	0.74%
Cefotetan	135	129	95.56%	4	2.96%		2	1.48%
Ceftazidime	135	127	94.07%	6	4.44%	2	1.48%	
Cefotaxime	135	134	99.26%	1	0.74%			
Cefepime	135	123	91.11%	9	6.67%		3	2.22%
Imipenem	135	133	98.52%	2	1.48%			
Ertapenem	112	112	100.0%					
Piperacillin-tazobactam	122	116	95.08%	5	4.10%		1	0.82%
Sulfamethoxazole/trimethoprim	140	137	97.86%	2	1.43%	1	0.71%	
Tobramycin	135	132	97.78%	3	2.22%			
Total	1828	1769	96.77%	46	2.52%	4	0.22%	9
<b>STAPHYLOCOCCUS SPP (n = 105), ENTEROCOCCUS SPP (n = 28)</b>								
Penicillin	105	104	99.05%				1	0.95%
Oxacillin	105	104	99.05%				1	0.95%
Gentamicin	105	83	79.05%	10	9.52%	5	4.76%	7
Levofloxacin	133	124	93.23%	6	4.51%		3	2.26%
Erythromycin	133	130	97.74%	2	1.50%		1	0.75%
Clindamycin	133	127	95.49%	4	3.01%		2	1.50%
Linezolid	133	133	100.0%					
Vancomycin	133	133	100.0%					
Rifampicin	105	105	100.0%					
Sulfamethoxazole/trimethoprim	105	93	88.57%	1	0.95%	2	1.90%	9
Ampicillin	28	4	14.29%	23	82.14%		1	3.57%
Gentamicin-High level	28	25	89.29%			2	7.14%	1
Total	1246	1165	93.50%	46	3.69%	9	0.72%	26
<b>STREPTOCOCCUS SPP (n = 14)</b>								
Penicillin	14	13	92.86%			1	7.14%	
Cefotaxime	14	14	100.0%					
Ceftriaxone	5	5	100.0%					
Imipenem	5	5	100.0%					
Meropenem	5	5	100.0%					
Levofloxacin	14	14	100.0%					
Erythromycin	14	14	100.0%					
Clindamycin	14	14	100.0%					
Linezolid	5	5	100.0%					
Vancomycin	14	14	100.0%					
Quinupristin-Dalflo	9	8	88.89%			1	11.11%	
Tetracycline	9	9	100.0%					
Chloramphenicol	9	9	100.0%					
Sulfamethoxazole/trimethoprim	9	9	100.0%					
Total	140	138	98.57%			1	0.71%	1

(Continued)

**TABLE 4 | Continued**

Antimicrobial agent	No	Category agreement	mE	ME	VME	
<b>YEAST (n = 18)</b>						
5-Fluorocytosine	18	17	94.44%	1	5.56%	
Fluconazole	18	17	94.44%		1	5.56%
Voriconazole	18	18	100.0%			
Amphotericin B	18	18	100.0%			
Itraconazole	18	16	88.89%	2	11.11%	
Total	90	86	95.56%	3	3.33%	
<b>CRYPTOCOCCUS NEOFORMANS (n = 15)</b>						
5-Fluorocytosine	15	15	100.0%			
Fluconazole	15	15	100.0%			
Amphotericin B	15	15	100.0%			
Total	45	45	100.0%			

The total numbers of isolates differed for each specific antimicrobial agent tested according to the CLSI (2015) guidelines.



**FIGURE 3 | Comparison between the average times of RMI/RMAST and FID/FAST.** The average times of RMAST vs. the conventional method were of great statistical significant ( $p < 0.0001$ ). \*\*\* $p < 0.0001$ . Bacteria failed to grow: bacteria failed to grow on the routing plates.

returned at least 18–36 h earlier than the final reports. Notably, the RMI of rare bacteria that failed to grow on routing plates might be advanced by 66.5 h. Second, the RMI results may accelerate the installation antibiotic therapy of BSIs caused by *S. aureus* (8.45%), *S. lugdunensis* (not detected in the present study) and *Listeria monocytogenes* (0.82%), which due to morphology consistent with the contaminated bacteria and gram stain are indistinguishable. The RMI also hastens the modification of empirical antibiotic treatment in cases showing intrinsic resistance (1.86% of cultures in the present study). From another viewpoint, 3.71% of the cultures indicated cessation of antibiotic therapy. Early recognitions of these contaminants would avoid wastage of medical resources. Third, the RMAST results provide a variety of accurate MIC values, by which clinicians can choose drug multi-directionally and estimate therapeutic doses precisely. In the present RMAST results,

13.44% of the cases showed resistance to last-resort antibiotics (such as vancomycin, linezolid, and carbapenem), further alerting clinicians to adjust the treatment or initiate essential combination therapy. In contrast, 51.56% of the bacteria were susceptible to third-generation cephalosporin and carbapenem (GN bacteria) or vancomycin and linezolid (GP bacteria), suggesting that antibiotic therapy might be de-escalated in patients infected by these organisms.

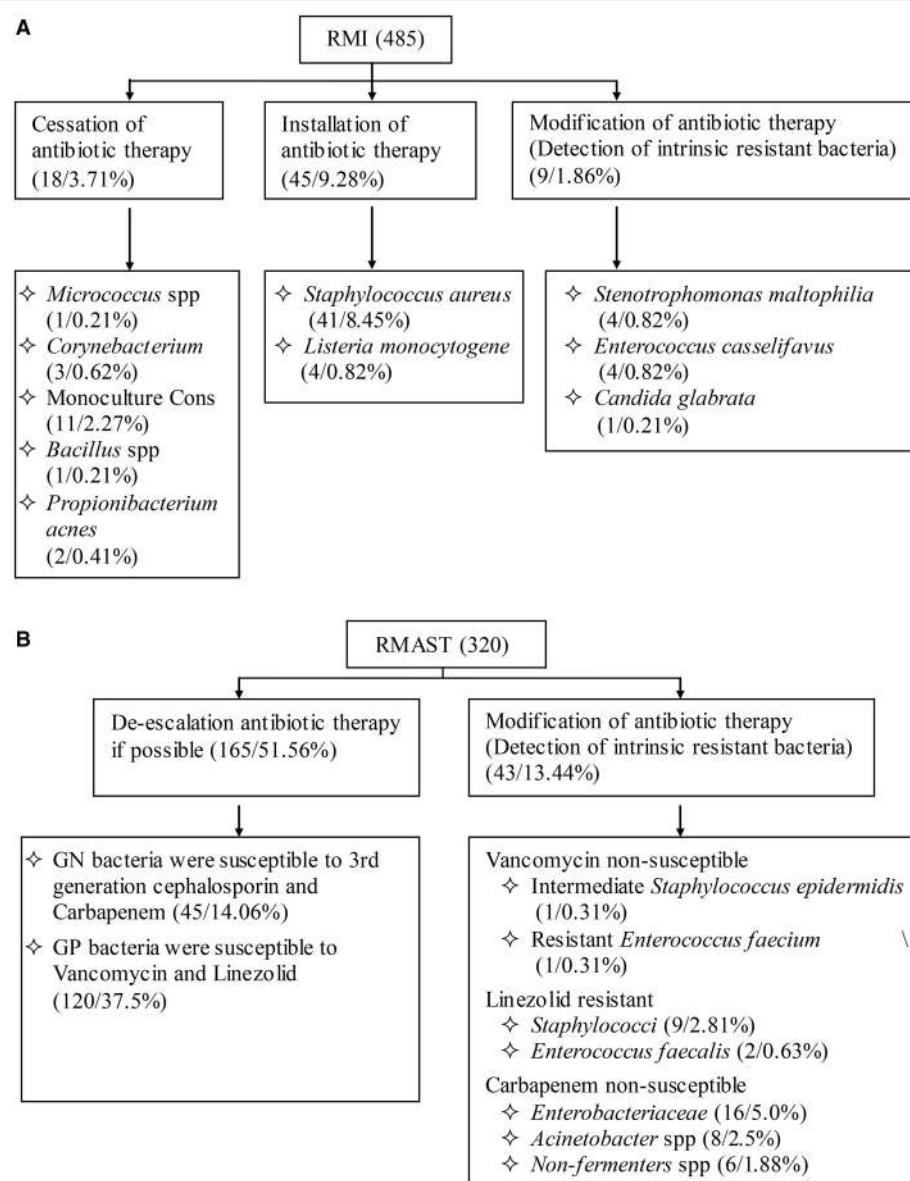
In this study, we theoretically examined the utility of RMI/RMAST in optimizing therapeutic regimens. We did not conduct a practical investigation. Several studies (Martiny et al., 2013; Perez et al., 2013) prospectively demonstrated that rapid pathogen identification and antimicrobial stewardship reduces the hospital length of stay and total costs. However, these studies were limited to BSI caused by GN bacteria. Therefore, we tentatively propose a protocol that establishes the control and intervention groups and detects bacterial, yeast and fungal infections in diverse sterile body fluids. The impact of RMI/RMAST in decreasing the hospital length of stay, cost, and antibiotic selective pressure deserves further investigation.

## CONCLUSIONS

The combined MALDI-TOF MS and Vitek AST system could obtain a rapid, accurate, reliable identification, and ASTs reports. The RMI and RMAST results not only reduce the turnaround time of the patient, but also guide clinicians whether a patient's antibiotic treatment should be accelerated, ceased or de-escalated, and adjusted the essential drugs modification for an optimized therapy.

## ETHICS STATEMENT

This study was approved by the ethics committee of Huashan Hospital, Shanghai Medical College, Fudan University, Shanghai,



**FIGURE 4 | (A):** Influence of RMI results on empirical antimicrobial therapy. **(B):** Influence of RMAST results on empirical antimicrobial therapy.

People's Republic of China (protocol HS-H-2014-0213). All subjects provided written informed consent before their inclusion in the study.

## AUTHORS CONTRIBUTIONS

YT and BZ performed all experiments. BW assisted in antimicrobial susceptibility testing. ML, YL, YT, and BZ conceived the study and analyzed the results. ML and YL supervised the study and wrote the manuscript. All authors read and approved the final manuscript.

## FUNDING

This study was supported by the National Natural Science Foundation of China (grants 81322025, 81171623 and 81371875), Outstanding Young Talent Plan of Shanghai (XYQ2011039), and Shanghai Shuguang Talent Project (12SG03).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2016.00523>

## REFERENCES

- Adhikari, N. K., Fowler, R. A., Bhagwanjee, S., and Rubenfeld, G. D. (2010). Critical care and the global burden of critical illness in adults. *Lancet* 376, 1339–1346. doi: 10.1016/S0140-6736(10)60446-1
- Al-Haroni, M., Skaug, N., Bakken, V., and Cash, P. (2008). Proteomic analysis of ampicillin-resistant oral *Fusobacterium nucleatum*. *Oral Microbiol. Immunol.* 23, 36–42. doi: 10.1111/j.1399-302XX.2007.00387.x
- Ascione, T., Pagliano, P., Mariconda, M., Rotondo, R., Balato, G., Toro, A., et al. (2015). Factors related to outcome of early and delayed prosthetic joint infections. *J. Infect.* 70, 30–36. doi: 10.1016/j.jinf.2014.07.008
- Bagheri-Nesami, M., Babamahmoodi, F., and Nikkhah, A. (2015). Types, risk factors, clinical symptoms and diagnostic tests of acute adult meningitis in northern iran during 2006–2012. *J. Clin. Diagn. Res.* 9, IC01–IC05. doi: 10.7860/JCDR/2015/11991.5936
- Bizzini, A., and Greub, G. (2010). Matrix-assisted laser desorption ionization time-of-flight mass spectrometry, a revolution in clinical microbial identification. *Clin. Microbiol. Infect.* 16, 1614–1619. doi: 10.1111/j.1469-0691.2010.03311.x
- Blot, S., Vandewoude, K., De Bacquer, D., and Colardyn, F. (2002). Nosocomial bacteremia caused by antibiotic-resistant gram-negative bacteria in critically ill patients: clinical outcome and length of hospitalization. *Clin. Infect. Dis.* 34, 1600–1606. doi: 10.1086/340616
- Cain, S. E., Kohn, J., Bookstaver, P. B., Albrecht, H., and Al-Hasan, M. N. (2015). Stratification of the impact of inappropriate empirical antimicrobial therapy for gram-negative bloodstream infections by predicted prognosis. *Antimicrob. Agents Chemother.* 59, 245–250. doi: 10.1128/AAC.03935-14
- Chan, W. S., Chan, T. M., Lai, T. W., Chan, J. F., Lai, R. W., Lai, C. K., et al. (2015). Complementary use of MALDI-TOF MS and real-time PCR-melt curve analysis for rapid identification of methicillin-resistant staphylococci and VRE. *J. Antimicrob. Chemother.* 70, 441–447. doi: 10.1093/jac/dku411
- Chon, Y. E., Kim, S. U., Lee, C. K., Park, J. Y., Kim, Y., Han, K. H., et al. (2014). Community-acquired vs. nosocomial spontaneous bacterial peritonitis in patients with liver cirrhosis. *Hepatogastroenterology* 61, 2283–2290.
- Clinical and Laboratory Standards Institute (CLSI) (2015). *Performance Standards for Antimicrobial Susceptibility Testing; 25th Informational Supplement M100-S25*. Wayne, PA: CLSI.
- Dellinger, R. P., Levy, M. M., Rhodes, A., Annane, D., Gerlach, H., Opal, S. M., et al. (2013). Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock, 2012. *Intensive Care Med.* 39, 165–228. doi: 10.1007/s00134-012-2769-8
- Goto, M., and Al-Hasan, M. N. (2013). Overall burden of bloodstream infection and nosocomial bloodstream infection in North America and Europe. *Clin. Microbiol. Infect.* 19, 501–509. doi: 10.1111/1469-0691.12195
- Griffin, P. M., Price, G. R., Schooneveldt, J. M., Schlebusch, S., Tilse, M. H., Urbanski, T., et al. (2012). Use of matrix-assisted laser desorption ionization-time of flight mass spectrometry to identify vancomycin-resistant enterococci and investigate the epidemiology of an outbreak. *J. Clin. Microbiol.* 50, 2918–2931. doi: 10.1128/JCM.01000-12
- Grundt, A., Findeisen, P., Miethke, T., Jäger, E., Ahmad-Nejad, P., and Neumaier, M. (2012). Rapid detection of ampicillin resistance in Escherichia coli by quantitative mass spectrometry. *J. Clin. Microbiol.* 50, 1727–1729. doi: 10.1128/JCM.00047-12
- Hazelton, B., Thomas, L. C., Olma, T., Kok, J., O'Sullivan, M., Chen, S. C., et al. (2014). Rapid and accurate direct antibiotic susceptibility testing of blood culture broths using MALDI Sepsityper combined with the BD Phoenix automated system. *J. Med. Microbiol.* 63, 1590–1594. doi: 10.1099/jmm.0.075580-0
- Hrabák, J., Studentová, V., Walková, R., Zemlicková, H., Jakubová, V., Chudácková, E., et al. (2012). Detection of NDM-1, VIM-1, KPC, OXA-48, and OXA-162 carbapenemases by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J. Clin. Microbiol.* 50, 2441–2443. doi: 10.1128/JCM.01002-12
- Hrabak, J., Walkova, R., Studentova, V., Chudackova, E., and Bergerova, T. (2011). Carbapenemase activity detection by matrix-assisted laser desorption-ionization time of flight mass spectrometry. *J. Clin. Microbiol.* 49, 3222–3227. doi: 10.1128/JCM.00984-11
- Ibrahim, E. H., Sherman, G., Ward, S., Fraser, V. J., and Kollef, M. H. (2000). The influence of inadequate antimicrobial treatment of bloodstream infections on patient outcomes in the intensive care unit setting. *CHEST* 118, 146–155. doi: 10.1378/chest.118.1.146
- Idelevich, E. A., Grunewald, C. M., Wüllenweber, J., and Becker, K. (2014). Rapid identification and susceptibility testing of *Candida* spp. from positive blood cultures by combination of direct MALDI-TOF mass spectrometry and direct inoculation of Vitek 2. *PLoS ONE* 9:e114834. doi: 10.1371/journal.pone.0114834
- Kumar, A., Roberts, D., Wood, K. E., Light, B., Parrillo, J. E., Sharma, S., et al. (2006). Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Crit. Care Med.* 34, 1589–1596. doi: 10.1097/01.CCM.0000217961.75225.E9
- Machen, A., Drake, T., Wang, Y. F. (2014). Same day identification and full panel antimicrobial susceptibility testing of bacteria from positive blood culture bottles made possible by a combined lysis-filtration method with MALDI-TOF VITEK mass spectrometry and the VITEK2 system. *PLoS ONE* 9:e87870. doi: 10.1371/journal.pone.0087870
- Majcherczyk, P. A., McKenna, T., Moreillon, P., and Vaudaux, P. (2006). The discriminatory power of MALDI-TOF mass spectrometry to differentiate between isogenic teicoplanin-susceptible and teicoplanin-resistant strains of methicillin-resistant *Staphylococcus aureus*. *FEMS Microbiol Lett.* 255, 233–239. doi: 10.1111/j.1574-6968.2005.00060.x
- Martinez, R. M., Bauerle, E. R., Fang, F. C., and Butler-Wu, S. M. (2014). Evaluation of three rapid diagnostic methods for direct identification of microorganisms in positive blood cultures. *J. Clin. Microbiol.* 52, 2521–2529. doi: 10.1128/JCM.00529-14
- Martiny, D., Debaugnies, F., Gateff, D., Gérard, M., Aoun, M., Martin, C., Konopnicki, D., et al. (2013). Impact of rapid microbial identification directly from positive blood cultures using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry on patient management. *Clin. Microbiol. Infect.* 19, E568–E581. doi: 10.1111/1469-0691.12282
- Martiny, D., Dediste, A., and Vandenberg, O. (2012). Comparison of an in-house method and the commercial Sepsityper<sup>TM</sup> kit for bacterial identification directly from positive blood culture broths by matrix-assisted laser desorption-ionisation time-of-flight mass spectrometry. *Eur. J. Clin. Microbiol. Infect. Dis.* 31, 2269–2281. doi: 10.1007/s10096-012-1566-1
- Morgenthaler, N. G., and Kostrzewska, M. (2015). Rapid identification of pathogens in positive blood culture of patients with sepsis: review and meta-analysis of the performance of the sepsityper kit. *Int. J. Microbiol.* 2015:827416. doi: 10.1155/2015/827416
- Perez, K. K., Olsen, R. J., Musick, W. L., Cernoch, P. L., Davis, J. R., Land, G. A., et al. (2013). Integrating rapid pathogen identification and antimicrobial stewardship significantly decreases hospital costs. *Arch. Pathol. Lab. Med.* 137, 1247–1254. doi: 10.5858/arpa.2012-0651-OA
- Portnoy, A., Jit, M., Lauer, J., Blommaert, A., Ozawa, S., Stack, M., et al. (2015). Estimating costs of care for meningitis infections in low- and middle-income countries. *Vaccine* 33, A240–A247. doi: 10.1016/j.vaccine.2014.11.061
- Romero-Gómez, M. P., Gómez-Gil, R., Paño-Pardo, J. R., and Mingorance, J. (2012). Identification and susceptibility testing of microorganism by direct inoculation from positive blood culture bottles by combining MALDI-TOF and Vitek-2 Compact is rapid and effective. *J. Infect.* 65, 513–520. doi: 10.1016/j.jinf.2012.08.013
- Schieffer, K. M., Tan, K. E., Stamper, P. D., Somogyi, A., Andrea, S. B., Wakefield, T., et al. (2014). Multicenter evaluation of the Sepsityper extraction kit and MALDI-TOF MS for direct identification of positive blood culture isolates using the BD BACTEC FX and VersaTREK diagnostic blood culture systems. *J. Appl. Microbiol.* 116, 934–941. doi: 10.1111/jam.12434
- Schmidt, V., Jarosch, A., März, P., Sander, C., Vacata, V., and Kalka-Moll, W. (2012). Rapid identification of bacteria in positive blood culture by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Eur. J. Clin. Microbiol. Infect. Dis.* 31, 311ee7. doi: 10.1007/s10096-011-1312-0
- Schubert, S., Weinert, K., Wagner, C., Gunzl, B., Wieser, A., Maier, T., et al. (2011). Novel, improved sample preparation for rapid, direct identification from positive blood cultures using Matrix-Assisted Laser Desorption/Ionization Time-of-Flight (MALDI-TOF) Mass Spectrometry. *J. Mol. Diagn.* 13, 701–706. doi: 10.1016/j.jmoldx.2011.07.004

- Seng, P., Drancourt, M., Gouriet, F., La Scola, B., Fournier, P. E., Rolain, J. M., et al. (2009). Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Clin. Infect. Dis.* 49, 543–551. doi: 10.1086/600885
- Stevenson, L. G., Drake, S. K., and Murray, P. R. (2010). Rapid identification of bacteria in positive blood culture broths by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J. Clin. Microbiol.* 48, 444–447. doi: 10.1128/JCM.01541-09
- Thigpen, M. C., Whitney, C. G., Messonnier, N. E., Zell, E. R., Lynfield, R., Hadler, J. L., et al. (2011). Bacterial meningitis in the United States, 1998–2007. *N. Engl. J. Med.* 364, 2016–2025. doi: 10.1056/NEJMoa1005384
- Tumbarello, M., Spanu, T. D., Bidino, R., Marchetti, M., Ruggeri, M., Trecarichi, E. M., et al. (2010). Costs of bloodstream infections caused by *Escherichia coli* and influence of extended-spectrum-β-lactamase production and inadequate initial antibiotic therapy. *Antimicrob. Agents Chemother.* 54, 4085–4091. doi: 10.1128/AAC.00143-10
- Vincent, J. L., Sakr, Y., Sprung, C. L., Ranieri, V. M., Reinhart, K., Gerlach, H., et al. (2006). Sepsis in European intensive care units: results of the SOAP study. *Crit. Care Med.* 34, 344–353. doi: 10.1097/01.CCM.0000194725.48928.3A
- Wimmer, J. L., Long, S. W., Cernoch, P., Land, G. A., Davis, J. R., Musser, J. M., et al. (2012). Strategy for rapid identification and antibiotic susceptibility testing of Gram-negative bacteria directly recovered from positive blood cultures using the Bruker MALDI Biotyper and the BD Phoenix system. *J. Clin. Microbiol.* 50, 2452–2454. doi: 10.1128/JCM.00409-12
- Xu, C., Lin, X., Ren, H., Zhang, Y., Wang, S., and Peng, X. (2006). Analysis of outer membrane proteome of *Escherichia coli* related to resistance to ampicillin and tetracycline. *Proteomics* 6, 462–473. doi: 10.1002/pmic.200500219

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Tian, Zheng, Wang, Lin and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Rapid Detection of K1 Hypervirulent *Klebsiella pneumoniae* by MALDI-TOF MS

**Yonglu Huang<sup>1†</sup>, Jiaping Li<sup>1†</sup>, Danxia Gu<sup>1</sup>, Ying Fang<sup>1</sup>, Edward W. Chan<sup>2,3</sup>, Sheng Chen<sup>2,3\*</sup> and Rong Zhang<sup>1\*</sup>**

<sup>1</sup> Department of Clinical Microbiology, Second Affiliated Hospital of Zhejiang University, Hangzhou, China, <sup>2</sup> Shenzhen Key Lab for Food Biological Safety Control, Food Safety and Technology Research Center, Hong Kong PolyU Shenzhen Research Institute, Shenzhen, China, <sup>3</sup> State Key Lab of Chirosciences, Department of Applied Biology and Chemical Technology, The Hong Kong Polytechnic University, Hong Kong, China

## OPEN ACCESS

### Edited by:

Julio Alvarez,  
University of Minnesota, USA

### Reviewed by:

Tamas Szakmany,  
Cardiff University, UK  
Frederic Lamothe,  
Lausanne University Hospital,  
Switzerland

### \*Correspondence:

Sheng Chen  
sheng.chen@polyu.edu.hk;  
Rong Zhang  
brigitte\_zx@163.com

<sup>†</sup>These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Infectious Diseases,  
a section of the journal  
*Frontiers in Microbiology*

**Received:** 26 August 2015

**Accepted:** 01 December 2015

**Published:** 21 December 2015

### Citation:

Huang Y, Li J, Gu D, Fang Y, Chan EW, Chen S and Zhang R (2015)  
Rapid Detection of K1 Hypervirulent *Klebsiella pneumoniae* by MALDI-TOF  
MS. *Front. Microbiol.* 6:1435.  
doi: 10.3389/fmicb.2015.01435

Hypervirulent strains of *Klebsiella pneumoniae* (hvKP) are genetic variants of *K. pneumoniae* which can cause life-threatening community-acquired infection in healthy individuals. Currently, methods for efficient differentiation between classic *K. pneumoniae* (cKP) and hvKP strains are not available, often causing delay in diagnosis and treatment of hvKP infections. To address this issue, we devised a Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry (MS) approach for rapid identification of K1 hvKP strains. Four standard algorithms, genetic algorithm (GA), support vector machine (SVM), supervised neural network (SNN), and quick classifier (QC), were tested for their power to differentiate between K1 and non-K1 strains, among which SVM was the most reliable algorithm. Analysis of the receiver operating characteristic curves of the interest peaks generated by the SVM model was found to confer highly accurate detection sensitivity and specificity, consistently producing distinguishable profiles for K1 hvKP and non-K1 strains. Of the 43 *K. pneumoniae* modeling strains tested by this approach, all were correctly identified as K1 hvKP and non-K1 capsule type. Of the 20 non-K1 and 17 K1 hvKP validation isolates, the accuracy of K1 hvKP and non-K1 identification was 94.1 and 90.0%, respectively, according to the SVM model. In summary, the MALDI-TOF MS approach can be applied alongside the conventional genotyping techniques to provide rapid and accurate diagnosis, and hence prompt treatment of infections caused by hvKP.

**Keywords:** K1 hvKP, MALDI-TOF MS, rapid detection, SVM model, typical spectra

## INTRODUCTION

*Klebsiella pneumoniae*, a facultative anaerobic gram-negative bacillus (Fang et al., 2004), is an important opportunistic pathogen associated with both community-acquired and nosocomial infection such as pneumonia, urinary tract infections, septicemia, and wound infections, especially among patients in ICU (Vardakas et al., 2015). The first case of hypervirulent *Klebsiella pneumoniae* (HvKP) infection was reported to have originated from a patient with liver abscess in China in 1980s (Siu et al., 2012). Hvkp is a variant which is morphologically different from the classic strain in terms of appearance of colonies grown on agar plate. HvKP is not only able to cause nosocomial infection in immunocompromised patients, but more importantly, it often causes life-threatening

community-acquired (CA) infection in healthy individuals, eliciting a great concern worldwide (Liu et al., 2014). In recent years, the incidence of hvKP infection has increased markedly in various countries including Asia (Zhang et al., 2014), Europe (Decré et al., 2011), and South America (Vila et al., 2011).

Different from classical *K. pneumoniae* (cKP), hvKP has high iron acquisition ability, an increase in capsule production mediated by rmpA/rmpA2, which confers the hypermucoviscous, and association with Mucoviscosity-associated gene A (magA) and are commonly seen in K1, K2, K5, K20, K54, and K57 with K1 and K2 being the most dominant serotypes (Yeh et al., 2010; Shon et al., 2013). Emergence of hvKP strains represents a huge threat to human health (Shen et al., 2013) but methods for efficient differentiation between classic and hvKP strains are not available.

Matrix assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) is considered as an efficient tool which can accurately identify both commonly encountered pathogenic bacterial species and microbial pathogens that are difficult to identify, such as yeasts, anaerobes, and fastidious microorganisms (Martiny et al., 2012). Recently, some studies have shown that this technology exhibits the capacity for rapid discrimination of antibiotic resistant strains such as methicillin-resistant *Staphylococcus aureus* (Madhava Charyulu et al., 2012; Hu Yy et al., 2015) and carbapenem-resistant Enterobacteriaceae (Lau et al., 2014) from the sensitive organisms, detection of virulence factors such as *S. aureus* delta-toxin (Gagnaire et al., 2012; Josten et al., 2014), and epidemiological typing (Josten et al., 2013). In this work, we developed a MALDI-TOF MS method for rapid identification of K1 *K. pneumoniae* isolates, and evaluated its

reliability in rapid detection of major *K. pneumoniae* virulence factors.

## MATERIALS AND METHODS

### Conventional K1 hvKP Identification

A string test was performed to identify hvKP from clinical *K. pneumoniae* strains. A positive string test is defined as the formation of a mucoviscous string of >5 mm in length when using a bacteriology inoculation loop to touch and stretch a colony grown overnight on an blood agar plate at 35°C (Fang et al., 2007). The capsular polysaccharide synthesis virulence genes (K1, K2, K5, K20, K54, and K57) and other relevant genes (*wcaG*, *rmpA*, *magA*, and Aerobactin) were amplified through a TPersonal cycler (Biometra, Germany) to further identify K1 hvKP strains. Primer sequence and annealing temperature were shown in Table 1. PCR products were sequenced using an ABI 3730 sequencer (Applied Biosystems, Foster City, CA); and the data obtained were compared with the reported sequences retrieved from GenBank. A strain with positive string test, positive K1 positive K1 capsular polysaccharide gene and one or more of the virulence was defined as hvKP.

### Molecular Typing

Epidemiological relatedness of the K1 strains was studied by pulsed-field gel electrophoresis (PFGE) and multilocus sequence typing (MLST). Twenty-three K1-positive *K. pneumoniae* strains were genotyped by PFGE following the PulseNet protocol provided by the website of the U.S. Centers for Disease Control and Prevention (<http://www.cdc.gov/pulseNet/pathogens/index.html>). The bacterial cells were digested with the XbaI restriction

**TABLE 1 |** List of primers for detection of *Klebsiella pneumoniae* virulence genes.

Gene	Primer	Primer sequence	Amplicon (bp)	Annealing temperature (°C)	References
K1	K1-F	GTTAGGTATTGCAAGCCATGC	1047	55	Fang et al., 2007
	K1-R	GCCCCAGGTTAAATGAATCCGT			
K2	K2-F	GACCCGATATTCTACATCTGACAGAG	641	57	Turton et al., 2008
	K2-R	CCTGAAGTAAAATCGTAAATAGATGGC			
K5	K5-F	TGGTAGTGATGCTCGCGA	280	55	Turton et al., 2008
	K5-R	CCTGAACCCACCCAATC			
K20	K20-F	CGGTGCTACAGTCATCATT	741	55	Fang et al., 2007
	K20-R	GTTATACGATGCTCAGTCGC			
K54	K54-F	CATTAGCTCAGTGGTGGCT	881	55	Fang et al., 2007
	K54-R	GCTTGACAAACACCATAGCAG			
K57	K57-F	CTCAGGGCTAGAAGTGTCA	1037	55	Fang et al., 2007
	K57-R	CACTAACCCAGAAAGTCGAG			
<i>wcaG</i>	<i>wcaG</i> -F	GGTTGGKTCAGCAATCGTA	169	53	Turton et al., 2010
	<i>wcaG</i> -R	ACTATCCGCCAACCTTTC			
<i>rmpA</i>	<i>rmpA</i> -F	ACTGGGCTACCTCTGCTCA	535	50	Nadasy et al., 2007
	<i>rmpA</i> -R	CTTGCATGAGGCCATTTCA			
<i>magA</i>	<i>magA</i> -F	GGTGCTCTTACATCATTGC	1282	53	Fang et al., 2004
	<i>magA</i> -R	GCAATGGGCCATTGCGTTAG			
Aerobactin	Aerobactin-F	GCATAGGCGGATAACGAACAT	556	55	Yu et al., 2008
	Aerobactin-R	CACAGGGCAATTGCTTACCT			

enzyme and were separated in a Rotaphor System 6.0 instrument (Whatman Biometra). Banding patterns of 23 isolates were analyzed using the UVIBand software program (UVItec Ltd., Cambridge, United Kingdom), and the degree of sequence homology was calculated (unweighted pair group method using average linkages [UPGMA], 0.5% Master Lane). MLST was performed on representative isolates of each clonal type. Seven housekeeping genes of *K. pneumoniae* (*gapA*, *infB*, *mdh*, *pgi*, *phoE*, *rpoB*, and *tonB*) were amplified, sequenced and analyzed (Turton et al., 2007). Allele sequences and sequence types (STs) were analyzed according to the MLST database (<http://pubmlst.org/>).

## Protein Extraction

*K. pneumoniae* isolates were inoculated onto Columbia blood agar (Oxoid, Cambridge, UK) containing 5% sheep blood and incubated 18–24 h at 35°C. Several uniform colonies from fresh plates were re-suspended into 300 µl of distilled water. After addition of 900 µl ethanol, the extraction tube was centrifuged at 12,000 × g for 2 min, the supernatant was then discarded. The bacterial pellet was re-suspended in 50 µl 70% formic acid. Before centrifuging again, 50 µl acetonitrile was added. After 2 min centrifugation at 12,000 × g, 1 µl of supernatant was spotted onto the ground steel target and dried at room temperature. One microliter of alpha-cyano-4-hydroxycinnamic acid (CHCA) was overlaid and dried again.

## MALDI-TOF MS Analysis

MALDI-TOF MS analysis was performed on a Bruker MicroFlex LT mass spectrometer (Bruker Daltonics). Spectra were acquired according to the manufacturer's recommendations, mass range was from 2000 to 20,000 Da and the laser intensity was kept constant. Mass spectra were analyzed by the Biotyper 3.0 software and library (version 3.0, Bruker Daltonics). Identification score criteria used followed those recommended by the manufacturer: a score of >2.000 indicated species-level identification, a score of 1.700–1.999 indicated identification to the genus level, and a score of <1.700 was interpreted as inconclusive.

## Data Analysis

The ClinProTools software (v3.0; Bruker Daltonics) was used for peak analysis. Models were generated using all four available algorithms, genetic algorithm (GA), support vector machine (SVM), supervised neural network (SNN), and quick classifier (QC), followed by comparison to each other. For each model, the recognition capability and cross validation were calculated to demonstrate the sensitivity and specificity of the model, statistical analysis were obtained by the most reliable algorithm model. The receiver operating characteristic (ROC) curves for each of the peaks of interest were obtained from the ClinProTools software. The area under curve (AUC) was used to evaluate the performance of each algorithm.

## RESULTS

### Bacterial Isolates Used in This Study

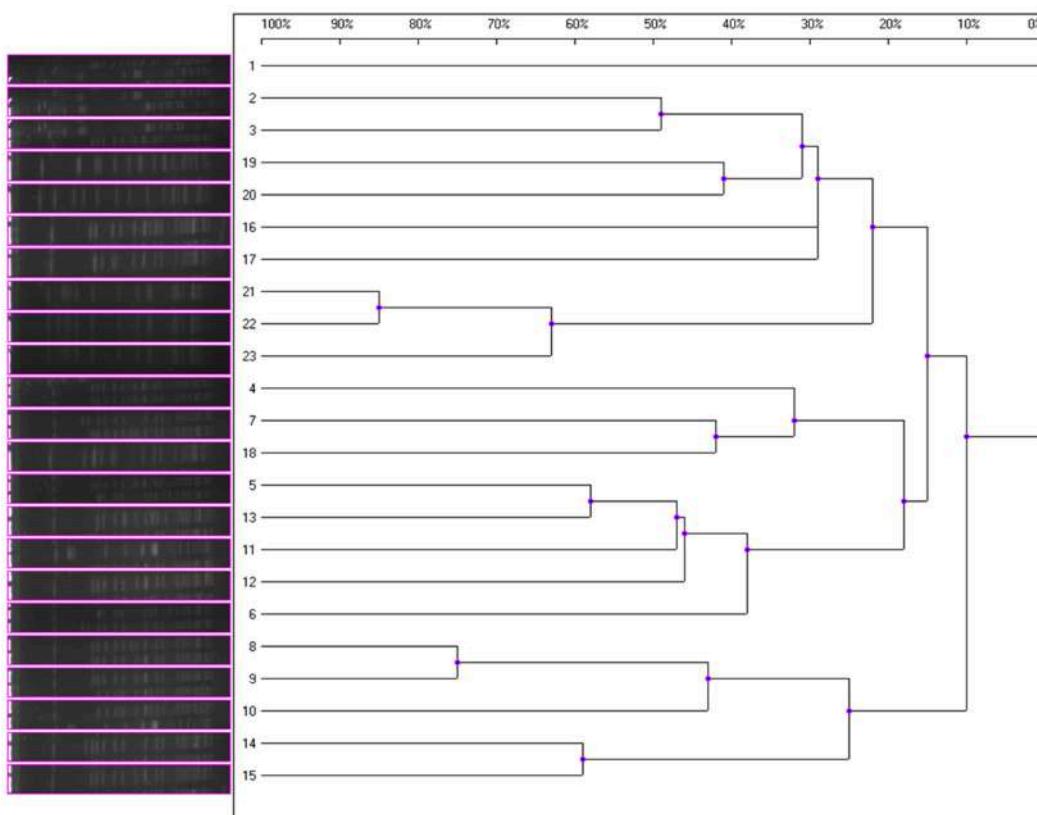
Forty K1hvKP strains were identified from 438 clinical non-repeated *K. pneumoniae* strains isolated from patients in Zhejiang Provincial hospital. Twenty-three out of the 40 K1 hvKP and 20 non-K1 isolates were randomly chosen for the development of MALDI-TOF MS method. The rest of 17 K1 isolates along with another 20 non-K1 isolates were used to evaluate the differentiation power of the MALDI-TOF MS method.

### Molecular Features of K1 and Non-K1 Strains

All 40 K1 hvKP contained K1 capsular gene and at least *magA* and Aerobactin genes. MLST analysis identified four STs including ST23, ST520, ST700, and ST1552 with ST23 the most dominant one (20 ST23 out of a total of 23K1 hvKP). In addition to *magA* and Aerobactin genes, all ST23 K1 hvKP also contained *wcaG* and *rmpA*; one ST700 contained *wcaG* (Table 2). Within the 20 ST23 K1 hvKP strains, only two (21 and 23) were shown to be identical by PFGE analysis (Figure 1). In contrast, the non-K1 strains are found to exhibit a wide range of genetic diversity with 14 different STs being identified in the 20 non-K1 *K. pneumoniae*. None of the non-K1 *K. pneumoniae* harbored any of the four genetic markers suggesting they are not hvKP strains (Table 2). Notably, STs in K1 hvKP and non-K1 *K. pneumoniae* were different suggesting the close association of STs to K1 hvKP in particular ST23.

**TABLE 2 |** Prevalence of ST types and known virulence genes in K1 hvKP and non-K1 *K. pneumoniae* strains.

Capsule type	ST types	No. of isolates	Other virulence genes			
			<i>wcaG</i>	<i>rmpA</i>	<i>magA</i>	Aerobactin
K1	23	20	+	+	+	+
	520	1	−	−	+	+
	700	1	+	−	+	+
	1552	1	−	−	+	+
non-K1	12	1	−	−	−	−
	34	1	−	−	−	−
	35	4	−	−	−	−
	36	1	−	−	−	−
	37	2	−	−	−	−
	138	2	−	−	−	−
	705	1	−	−	−	−
	753	1	−	−	−	−
	875	1	−	−	−	−
	983	1	−	−	−	−
	1411	1	−	−	−	−
	1547	2	−	−	−	−
	1548	1	−	−	−	−
	1551	1	−	−	−	−



**FIGURE 1 |** Magnified dendrogram (representation of hierarchical cluster analysis) of 23 K1hvKP isolates.

## Data Analysis by MALDI-TOF MS

Of the 43 *K. pneumoniae* modeling isolates analyzed, all (100%) were identified correctly as *K. pneumoniae* by MALDI-TOF MS. The models generated using four standard algorithms, GA, SVM, SNN, and QC, for comparing K1 with non-K1 isolates, were shown in Table 3. The use of the QC standard algorithm resulted in the lowest scores, whereas GA standard algorithm exhibited the highest specificity (recognition capability at 100%) when compared to the SVM standard algorithm (recognition capability at 97.83%). Yet SVM gave the highest sensitivity at 83.45% among four models, which is higher than the rate of 73.87% for GA. However, the specificity of SVM is slightly lower than GA being at 97.83%. Overall, SVM and SNN were the most reliable model for differentiating between K1 from classic or non-K1 isolates. The study showed that peaks or integration regions chosen for differentiation of K1 *K. pneumoniae* status by all four of the models were similar. The important peaks identified by the SVM model for K1 were 14, 31, 33, and 34. The peak statistics for these peaks are shown in Table 4.

The low *P*-values for the Anderson-Darling test (PAD) are evidence of the abnormal distribution of the data obtained. Therefore, the *P*-value of the PWK (P-value from combined Wilcoxon rank-sum test and Kruskal-Wallis test) is preferred over the PTTA (*p*-value of *t*-test) (as this is preferable for normally distributed data). The low *P*-values obtained from

PWKW (all were <0.05) indicated that the observed intensity differences of the individual peaks are highly statistically significant (i.e., the lower the *P*-value is, the higher the chance that a respective peak signal is suited to differentiate between the two classes) (Table 4).

The receiver operating characteristic (ROC) curves for each of the peaks of interest generated by the SVM model were also obtained from the ClinProTools software. The area under curve (AUC) can reflect the confidence level of each peak in identifying the sensitivity and specificity of virulent strains and non-K1 strains group, with an AUC of 0.5 representing purely random chance and an AUC of 1 indicating a perfect test (100% sensitivity and specificity; Table 4). All characteristic peaks used for distinguishing virulent from non-K1 strains were >0.8, confirming that these peaks can be used to differentiate K1 from non-K1 *K. pneumoniae* isolates with high accuracy. To demonstrate the differences visually, four representative peaks are shown in Figure 2, in which the spectra of the K1 and non-K1 *K. pneumoniae* isolates were distinguishable.

## Prospective Validation

Another 20 randomly selected non-K1 *K. pneumoniae* and 17 remaining K1-positive hvKP isolates were chosen to perform the prospective verification. The verification result indicated that the

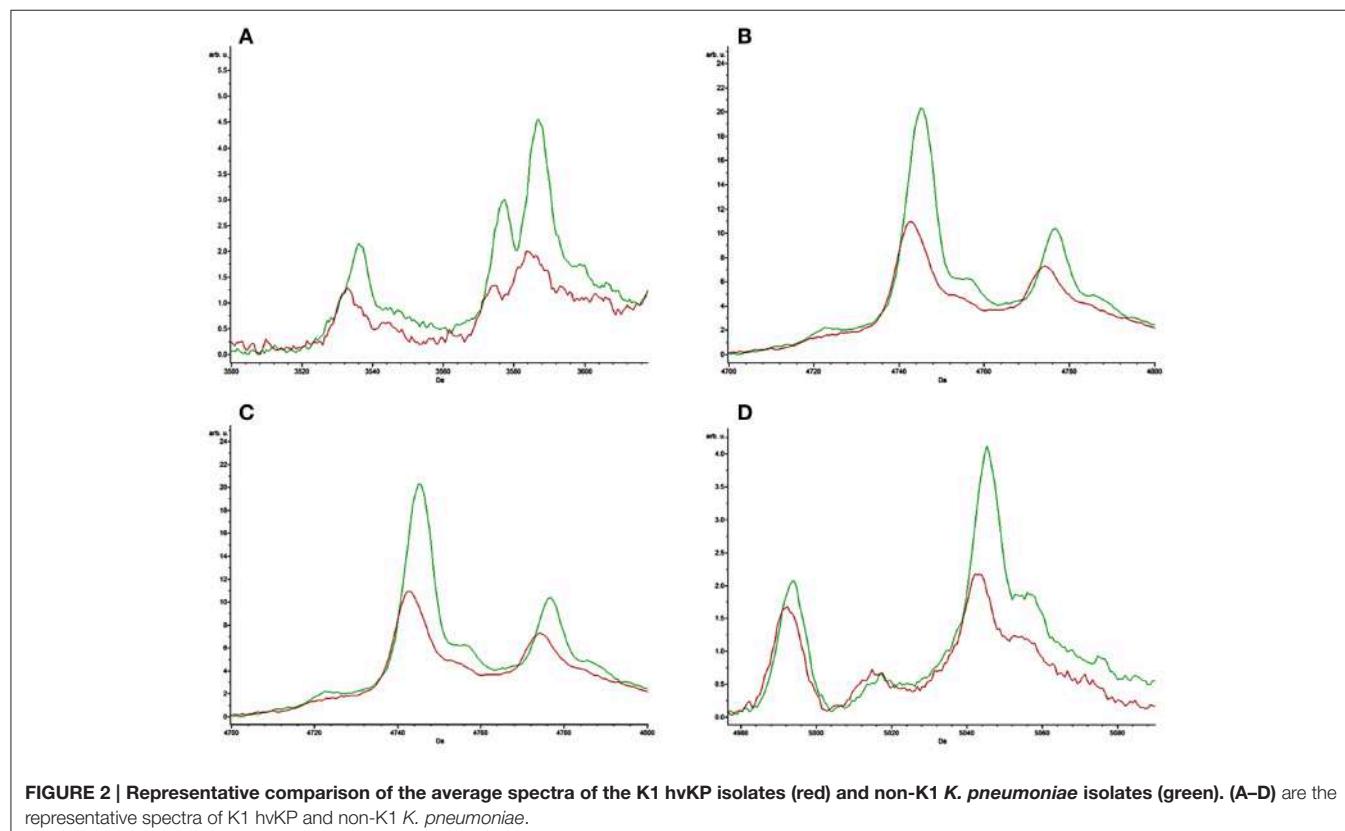
**TABLE 3 | Specificity and sensitivity of different algorithms models for differentiation between K1 and non-K1 *K. pneumoniae* isolates.**

Algorithms model	Assay development		Assay validation			
	Cross validation	Recognition capability	Specificity (true negative)	95% CI of Specificity	Sensitivity (true positive)	95% CI of Sensitivity
GA	100%	73.9%	16/17 (94.1%)	86.5–100%	17/20 (85.0%)	73.5–96.5%
SVM	97.8%	83.5%	16/17 (94.1%)	86.5–100%	18/20 (90.0%)	80.3–99.7%
SNN	100%	81.4%	16/17 (94.1%)	86.5–100%	18/20 (90.0%)	80.3–99.7%
QC	85.7%	70.7%	15/17 (88.3%)	77.9–98.6%	16/20 (80.0%)	67.1–92.9%

**TABLE 4 | ClinProTools peak statistics for the four peaks of interest in both K1 hvKP and non-K1 *K. pneumoniae*\***

Peak	Mass	DAve	PTTA	PWKW	PAD	Ave1	Ave2	SD1	SD2	CV1	CV2
14	3586.58	2.1	0.000339	0.000258	0.0729	3.1	5.2	1.23	1.48	39.71	28.36
31	4744.66	8.27	0.000164	0.000258	0.0843	12.91	21.18	5.41	4.61	41.94	21.74
33	5044.84	1.62	0.00036	0.000819	0.225	3.07	4.69	1.34	0.83	43.6	17.77
34	5148.93	2.59	0.000198	0.000525	0.0729	3.53	6.12	1.51	1.68	42.72	27.49

\*Sort mode, delta average arithmetic; peak, peak index; mass, m/z value; DAve, difference between the maximal and the minimal average peak area/intensity of all classes; PTTA, P-value of t-test; PWKW, P-value of Wilcoxon (preferable for abnormally distributed data); PAD, P-value of Anderson-Darling test (range, 0–1; 0, abnormally distributed; 1, normally distributed); Ave1 and Ave2, peak area/intensity average of class 1 (K1 *K. pneumoniae* isolates) and class 2 (non-K1 *K. pneumoniae* isolates), respectively; SD1 and SD2, standard deviations of the peak area/intensity average of class 1 and class 2, respectively; CV1 and CV2, coefficient of variation (in percentage) of class 1 and class 2, respectively.



**FIGURE 2 | Representative comparison of the average spectra of the K1 hvKP isolates (red) and non-K1 *K. pneumoniae* isolates (green). (A–D)** are the representative spectra of K1 hvKP and non-K1 *K. pneumoniae*.

accuracy of K1 and non-K1 identification was 94.1 and 90.0%, respectively, according to the SVM model suggesting the high accuracy of MALDI-TOF MS method to rapid identification of K1 hvKP (Table 3).

## DISCUSSION

Hospital-acquired *K. pneumoniae* clinical isolates normally exhibit relatively low-level virulence, whereas in most cases of

community-acquired pneumonia, *Klebsiella* isolates are generally highly virulent and known to produce mucoid colonies. Recent studies indicated that hvKP induced liver abscess was a new type of invasive infections (Siu et al., 2012), and that hvKP can not only cause liver abscess, but also metastatic infections such as bacteremia, meningitis, endophthalmitis, and necrosis fasciitis, often resulting in fatality in severe cases. A number of virulence factors have been identified in pathogenic *K. pneumoniae*, including capsular polysaccharide, mucus related gene A (mucoviscosity-associated gene A, *magA*), mucous phenotype A regulator gene (regulator of mucoid phenotype gene A, *rmpA*), and, aerobactin. Capsular polysaccharide is considered one of the major contributive factors of virulence in *K. pneumoniae*, promoting biofilm formation and exhibiting anti-opsonin effect which can help combat the host immune response when expressed in the host body (Cortés et al., 2002). To date, *K. pneumoniae* producing K1 type capsule is the major pathogen causing community-acquired lung infections in Asia (Lin et al., 2015).

Current hvKP detection methods include PCR amplification, multilocus sequence typing (MLST), pulsed field gel electrophoresis (PFGE), and the proteomics approach, which are complicated and time consuming, and demand technical competency. Since hvKP is highly virulent, causing high mobility and mortality among infected patients, a rapid and accurate clinical identification method is urgently needed to guide proper treatment of patients infected by hvKP. MALDI-TOF MS is a revolutionary technique for clinical bacteria identification. It has high power to identify clinical common bacteria, yeasts and fungi and has been shown to be used to identify antimicrobial resistance and virulence gene products. Based on the specific peaks, MALDI TOF mass spectrometry can be used to identify *blaKPC*-positive *K. pneumoniae*, van B positive *Enterococcus faeces* and virulence factors of *S. aureus* such as delta-toxin and PSM-mec (Josten et al., 2014).

## REFERENCES

- Cortés, G., Borrell, N., de Astorza, B., Gómez, C., Sauleda, J., and Alberti, S. (2002). Molecular analysis of the contribution of the capsular polysaccharide and the lipopolysaccharide O side chain to the virulence of *Klebsiella pneumoniae* in a murine model of pneumonia. *Infect. Immun.* 70, 2583–2590. doi: 10.1128/IAI.70.5.2583-2590.2002
- Decré, D., Verdet, C., Emirian, A., Le Gourrierec, T., Petit, J. C., Offenstadt, G., et al. (2011). Emerging severe and fatal infections due to *Klebsiella pneumoniae* in two university hospitals in France. *J. Clin. Microbiol.* 49, 3012–3014. doi: 10.1128/JCM.00676-11
- Fang, C. T., Chuang, Y. P., Shun, C. T., Chang, S. C., and Wang, J. T. (2004). A novel virulence gene in *Klebsiella pneumoniae* strains causing primary liver abscess and septic metastatic complications. *J. Exp. Med.* 199, 697–705. doi: 10.1084/jem.20030857
- Fang, C. T., Lai, S. Y., Yi, W. C., Hsueh, P. R., Liu, K. L., and Chang, S. C. (2007). *Klebsiella pneumoniae* genotype K1: an emerging pathogen that causes septic ocular or central nervous system complications from pyogenic liver abscess. *Clin. Infect. Dis.* 45, 284–293. doi: 10.1086/519262
- Gagnaire, J., Dauwalder, O., Boisset, S., Khau, D., Freydière, A. M., Ader, F., et al. (2012). Detection of *Staphylococcus aureus* delta-toxin production by whole-cell MALDI-TOF mass spectrometry. *PLoS ONE* 7:e40660. doi: 10.1371/journal.pone.0040660
- Hu Yy, C. J., Zhou, H. W., Jiang, Y., and Zhang, R. (2015). Rapid identification of methicillin-resistant *Staphylococcus aureus* and methicillin-sensitive *Staphylococcus aureus* strains by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Chin. J. Microbiol. Immunol.* 1, 42–45. doi: 10.3760/cma.j.issn.0254-5101.2015.01.009
- Josten, M., Dischinger, J., Szekat, C., Reif, M., Al-Sabti, N., Sahl, H. G., et al. (2014). Identification of agr-positive methicillin-resistant *Staphylococcus aureus* harbouring the class A mec complex by MALDI-TOF mass spectrometry. *Int. J. Med. Microbiol.* 304, 1018–1023. doi: 10.1016/j.ijmm.2014.07.005
- Josten, M., Reif, M., Szekat, C., Al-Sabti, N., Roemer, T., Sparbier, K., et al. (2013). Analysis of the matrix-assisted laser desorption ionization-time of flight mass spectrum of *Staphylococcus aureus* identifies mutations that allow differentiation of the main clonal lineages. *J. Clin. Microbiol.* 51, 1809–1817. doi: 10.1128/JCM.00518-13
- Lau, A. F., Wang, H., Weingarten, R. A., Drake, S. K., Suffredini, A. F., Garfield, M. K., et al. (2014). A rapid matrix-assisted laser desorption ionization-time of flight mass spectrometry-based method for single-plasmid tracking in an outbreak of carbapenem-resistant *Enterobacteriaceae*. *J. Clin. Microbiol.* 52, 2804–2812. doi: 10.1128/JCM.00694-14
- Lin, Y. T., Wang, Y. P., Wang, F. D., and Fung, C. P. (2015). Community-onset *Klebsiella pneumoniae* pneumonia in Taiwan: clinical features of the disease and associated microbiological characteristics of isolates from pneumonia and nasopharynx. *Front. Microbiol.* 9:122. doi: 10.3389/fmicb.2015.00122

- Liu, Y. M., Li, B. B., Zhang, Y. Y., Zhang, W., Shen, H., Li, H., et al. (2014). Clinical and molecular characteristics of emerging hypervirulent *Klebsiella pneumoniae* bloodstream infections in mainland China. *Antimicrob. Agents Chemother.* 58, 5379–5385. doi: 10.1128/AAC.02523-14
- Madhava Charyulu, E., Gnanamani, A., and Mandal, A. B. (2012). Identification and discrimination of Methicillin Resistant *Staphylococcus aureus* strains isolated from burn wound sites using PCR and authentication with MALDI-TOF-MS. *Indian J. Microbiol.* 52, 337–345. doi: 10.1007/s12088-011-0245-8
- Martiny, D., Busson, L., Wybo, I., El Haj, R. A., Dediste, A., and Vandenberg, O. (2012). Comparison of the Microflex LT and Vitek MS systems for routine identification of bacteria by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J. Clin. Microbiol.* 50, 1313–1325. doi: 10.1128/JCM.05971-11
- Nadasy, K. A., Domiaty-Saad, R., and Tribble, M. A. (2007). Invasive *Klebsiella pneumoniae* syndrome in North America. *Clin. Infect. Dis.* 45, e25–e28. doi: 10.1086/519424
- Shen, D. X., Wang, J., and Li, D. D. (2013). *Klebsiella pneumoniae* liver abscesses. *Lancet Infect. Dis.* 13, 390–391. doi: 10.1016/S1473-3099(13)70068-9
- Shon, A. S., Bajwa, R. P., and Russo, T. A. (2013). Hypervirulent (hypermucoviscous) *Klebsiella pneumoniae*: a new and dangerous breed. *Virulence* 4, 107–118. doi: 10.4161/viru.22718
- Siu, L. K., Yeh, K. M., Lin, J. C., Fung, C. P., and Chang, F. Y. (2012). *Klebsiella pneumoniae* liver abscess: a new invasive syndrome. *Lancet Infect. Dis.* 12, 881–887. doi: 10.1016/S1473-3099(12)70205-0
- Turton, J. F., Baklan, H., Siu, L. K., Kaufmann, M. E., and Pitt, T. L. (2008). Evaluation of a multiplex PCR for detection of serotypes K1, K2 and K5 in *Klebsiella* sp. and comparison of isolates within these serotypes. *FEMS Microbiol. Lett.* 284, 247–252. doi: 10.1111/j.1574-6968.2008.01208.x
- Turton, J. F., Englender, H., Gabriel, S. N., Turton, S. E., Kaufmann, M. E., and Pitt, T. L. (2007). Genetically similar isolates of *Klebsiella pneumoniae* serotype K1 causing liver abscesses in three continents. *J. Med. Microbiol.* 56, 593–597. doi: 10.1099/jmm.0.46964-0
- Turton, J. F., Perry, C., Elgohari, S., and Hampton, C. V. (2010). PCR characterization and typing of *Klebsiella pneumoniae* using capsular type-specific, variable number tandem repeat and virulence gene targets. *J. Med. Microbiol.* 59, 541–547. doi: 10.1099/jmm.0.015198-0
- Vardakas, K. Z., Matthaiou, D. K., Falagas, M. E., Antypa, E., Koteli, A., and Antoniadou, E. (2015). Characteristics, risk factors and outcomes of carbapenem-resistant *Klebsiella pneumoniae* infections in the intensive care unit. *J. Infect.* 70, 592–599. doi: 10.1016/j.jinf.2014.11.003
- Vila, A., Cassata, A., Pagella, H., Amadio, C., Yeh, K. M., Chang, F. Y., et al. (2011). Appearance of *Klebsiella pneumoniae* liver abscess syndrome in Argentina: case report and review of molecular mechanisms of pathogenesis. *Open Microbiol. J.* 5, 107–113. doi: 10.2174/1874285801105010107
- Yeh, K. M., Lin, J. C., Yin, F. Y., Fung, C. P., Hung, H. C., Siu, L. K., et al. (2010). Revisiting the importance of virulence determinant magA and its surrounding genes in *Klebsiella pneumoniae* causing pyogenic liver abscesses: exact role in serotype K1 capsule formation. *J. Infect. Dis.* 201, 1259–1267. doi: 10.1086/606010
- Yu, W. L., Ko, W. C., Cheng, K. C., Lee, C. C., Lai, C. C., and Chuang, Y. C. (2008). Comparison of prevalence of virulence factors for *Klebsiella pneumoniae* liver abscesses between isolates with capsular K1/K2 and non-K1/K2 serotypes. *Diagn. Microbiol. Infect. Dis.* 62, 1–6. doi: 10.1016/j.diagmicrobio.2008.04.007
- Zhang, Y., Ma, Y., Ye, L., Luo, Y., and Yang, J. (2014). Prevalence and antimicrobial susceptibility of hypervirulent *Klebsiella pneumoniae* isolates in China. *Clin. Infect. Dis.* 58, 1493–1494. doi: 10.1093/cid/ciu110

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Huang, Li, Gu, Fang, Chan, Chen and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Assessment of MALDI-TOF MS as alternative tool for *Streptococcus suis* identification

Marta Pérez-Sancho<sup>1,2</sup>, Ana Isabel Vela<sup>1,3</sup>, Teresa García-Seco<sup>1</sup>, Marcelo Gottschalk<sup>4</sup>, Lucas Domínguez<sup>1,2,3</sup> and José Francisco Fernández-Garayzábal<sup>1,3\*</sup>

<sup>1</sup> Centro de Vigilancia Sanitaria Veterinaria (VISAVENT), Universidad Complutense, Madrid, Spain, <sup>2</sup> Campus de Excelencia Internacional (CEI) Moncloa, Universidad Politécnica de Madrid (UPM), Universidad Complutense de Madrid (UCM), Madrid, Spain, <sup>3</sup> Departamento de Sanidad Animal, Facultad de Veterinaria, Universidad Complutense, Madrid, Spain, <sup>4</sup> Groupe de Recherche sur les Maladies Infectieuses du Porc, Faculté de Médecine Vétérinaire, Université de Montréal, Saint-Hyacinthe, QC, Canada

## OPEN ACCESS

### Edited by:

Andres M. Perez,  
University of Minnesota, USA

### Reviewed by:

Tamas Szakmany,  
Cardiff University, UK  
Markus Kostrzewa,  
Bruker Daltonik GmbH, Germany

### \*Correspondence:

José Francisco  
Fernández-Garayzábal,  
Centro de Vigilancia Sanitaria  
Veterinaria (VISAVENT), Universidad  
Complutense de Madrid, Avenida  
Puerta de Hierro, s/n, Madrid 28040,  
Spain  
garayzab@vet.ucm.es

### Specialty section:

This article was submitted to  
Infectious Diseases, a section of the  
journal Frontiers in Public Health

Received: 25 May 2015

Accepted: 07 August 2015

Published: 21 August 2015

### Citation:

Pérez-Sancho M, Vela Al,  
García-Seco T, Gottschalk M,  
Domínguez L and  
Fernández-Garayzábal JF (2015)  
Assessment of MALDI-TOF MS as  
alternative tool for *Streptococcus suis*  
identification.  
*Front. Public Health* 3:202.  
doi: 10.3389/fpubh.2015.00202

The accuracy of matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) for identifying *Streptococcus suis* isolates obtained from pigs, wild animals, and humans was evaluated using a PCR-based identification assay as the gold standard. In addition, MALDI-TOF MS was compared with the commercial multi-tests Rapid ID 32 STREP system. From the 129 *S. suis* isolates included in the study and identified by the molecular method, only 31 isolates (24.03%) had score values  $\geq 2.300$  and 79 isolates (61.24%) gave score values between 2.299 and 2.000. After updating the currently available *S. suis* MALDI Biotype database with the spectra of three additional clinical isolates of serotypes 2, 7, and 9, most isolates had statistically significant higher score values (mean score: 2.65) than those obtained using the original database (mean score: 2.182). Considering the results of the present study, we suggest using a less restrictive threshold score of  $\geq 2.000$  for reliable species identification of *S. suis*. According to this cut-off value, a total of 125 *S. suis* isolates (96.9%) were correctly identified using the updated database. These data indicate an excellent performance of MALDI-TOF MS for the identification of *S. suis*.

**Keywords:** identification, MALDI-TOF MS, *Streptococcus suis*, PCR, biochemical tests

## Introduction

*Streptococcus suis* is one of the most important pathogens in the swine industry worldwide, causing meningitis and a wide range of diseases, such as, arthritis, endocarditis, pneumonia, and septicemia (1). Furthermore, *S. suis* has been isolated from a range of other mammalian and avian species (2–4). *S. suis* has also been recognized as an emerging human pathogen over the past few years, affecting people in close contact with pigs or pork-derived products (3).

The clinical significance of *S. suis* infections in both human and animal medicine makes necessary to have diagnostic tools able to accurately identify this pathogen. Despite PCR assays for the detection and identification of *S. suis* have been developed (5–7), in many diagnostic laboratories, the identification of this pathogen is still based on bacteriological and biochemical criteria, mainly using commercial multi-test systems that, in general, present controversial results (8). Veterinary diagnostic laboratories can easily identify *S. suis* isolates from clinical cases in pigs, but identification of *S. suis* from healthy pigs can be more difficult due to the presence of other streptococci that are part of the normal tonsillar microflora and are phenotypically similar to *S. suis* (8). Human diagnostic

laboratories can also misidentify *S. suis* with other microorganisms, such as, enterococci, *S. bovis* or viridans group streptococci (8, 9). More recently, MALDI-TOF MS has emerged as a reliable high-throughput tool for microbiological identification. This technique has been demonstrated as a reliable alternative tool for identification of gram-positive bacteria (10), including *Streptococcus* spp. isolates (11–15). However, particular identification of *S. suis* using this approach has not yet been thoroughly reported although this technique could overcome the drawbacks of current routine identification techniques and may contribute to a better understanding of its impact in animal production and public health.

Hence, this study was conducted to evaluate the accuracy of matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) for identifying *S. suis* isolates obtained from pigs, wild animals and humans, using a polymerase chain reaction (PCR)-based identification as the gold standard. In addition, MALDI-TOF MS was compared with the commercial multi-tests Rapid ID 32 STREP system.

## Materials and Methods

### Bacterial Strains

Overall, 129 *S. suis* isolates were used in the study. These include 62 and 64 isolates recovered from swine and wild animals (wild boar, wild rabbit, and Iberian wild goat), respectively, belonging to different serotypes (2 n = 25; 7 n = 5; 9 n = 69; less prevalent serotypes n = 27). Pig isolates were recovered from Spain (n = 47) or Canada (n = 15), while wild animal isolates were all from Spain. All animal isolates from Spain were obtained from the culture collection of VISAVENT Centre (Universidad Complutense de Madrid). Three human serotype 2 isolates were also included in the study. Identification of all isolates was confirmed by a species-specific PCR assay described by Okwumabua et al. (6) which have been considered the gold standard technique in this work. All isolates were grown overnight on Columbia sheep agar plates at 37°C under aerobic conditions.

Isolates with different identification levels based in the Rapid ID 32 STREP system (bioMérieux) were included in order to assess the capability of MALDI-TOF MS technology to overcome the drawbacks detected in biochemical identification. Rapid ID 32 STREP identifications were categorized as follows: species identification (which included acceptable, good, very good, and excellent identification to species level), genus identification (including doubtful identification to species level or acceptable and good identification to genus level), and unreliable identification (which contains unacceptable identification or identified as non-*S. suis*).

### MALDI-TOF MS Analysis

Spectra from each isolate were obtained after ethanol formic acid extraction in accordance with manufacturer's instructions using fresh and pure cultures considered an efficient sample preparation method for gram-positive bacteria (16, 17). One microliter of each isolate extract was spotted onto a 384-spot polished steel target plate, let to dry at room temperature and overlaid with one microliter of α-cyano-4-hydroxy-cinnamic acid (HCCA)

matrix. Data acquisition was performed using a Bruker Daltonics UltrafleXtrem MALDI TOF/TOF equipment and the Biotyper Real Time Classification software v3.1 (Bruker Daltonics, Bremen, Germany).

In order to improve the already available *S. suis* MALDI Biotyper database (including five entries) and generate an updated database (UDB), the MALDI-TOF MS analysis of three additional clinical isolates of serotypes 2, 7, and 9 were performed. The 24 spectra obtained from the eight spots for each of the three strains were analyzed by FlexAnalysis (version 3.0, Bruker Daltonics) according to Rettinger et al. (18). A minimum of 20 accurate spectra were downloaded in MALDI Biotyper (version 3.0, Bruker Daltonics) to create a main spectrum profile (MSP) of each strain according to the manufacturer's suggestions.

### MALDI-TOF MS Identification

All 129 isolates were analyzed by MALDI-TOF MS as described above and the spectra obtained were compared with the original MALDI Biotyper database (ODB) and the UDB. The reliability of the identification using MALDI Biotyper was performed according to the log (score) values calculated by the Biotyper software (version 3.1; 4613 entries) according to manufacturer's parameters: highly probable identification at species level: scores of  $\geq 2.300$ ; secure genus and probable species identification: scores between 2.299 and 2.000; probable genus level identification: score of 1.999–1.700; unreliable identification: scores  $< 1.700$ . For each isolate with scores  $\leq 2.000$ , MALDI-TOF MS analysis and extraction protocol was repeated at least in two independent runs.

### Statistical Analysis

Agreement between the results obtained with both techniques was assessed applying the kappa test. Proportions of positive samples were compared using Z-test (adjust p-values – Bonferroni methods). To evaluate a possible association between MALDI quantitative results (scores) and the categorized API profiles, ANOVA test was used. In addition, the comparison of the usefulness of the ODB and UDB were assessed using the Paired-sample t-test. Analysis of the data was carried out using software SPSS 20 (Statistical Package for the Social Sciences, IBM, New York, USA).

## Results

A total of 31 isolates (24.03%; mean score = 2.348) gave scores values  $\geq 2.300$ , another 79 isolates (61.24%; mean scores = 2.188) gave score values between 2.299 and 2.000 and 19 isolates (14.7%) had score values between 1.999 and 1.748 (**Table 1**). Except in one isolate identified as *Streptococcus pneumoniae* (score value of 1.748), the first identification option by MALDI-TOF MS, irrespective of the score value, was always *S. suis*.

There were differences on the percentage of isolates identified at species level with a score value  $\geq 2.300$  (highly probable species) depending on serotype. Thus, the 42.9% (12/28), 20% (1/5), and 18.8% (13/69) of isolates of serotypes 2, 7, and 9, respectively, had score values  $\geq 2.300$ . These percentages were not significantly higher for any of the analyzed serotypes (Z-test, adjust p-values – Bonferroni methods). The percentage isolates belonging to less prevalent serotypes with score values  $\geq 2.300$  was 19.2%

**TABLE 1 | Comparison of MALDI-TOF MS results depending on database and threshold score used in a panel of 129 *S. suis* isolates recovered from swine, wild animals, and humans.**

Identification level	MALDI-TOF MS		
	ODB <sup>a</sup>		UDB <sup>b</sup>
	No. (%)	Bruker's threshold score No. (%) <sup>c</sup>	Proposed threshold score No. (%) <sup>d</sup>
Highly probable specie	31 (24)	106 (82.2)	125 (96.90)
Probable specie, secure genus	79 (61.2)	19 (14.73)	4 (3.1)
Probable genus	19 (14.7)	4 (3.1)	0 (0)

<sup>a</sup>ODB: original database of MALDI Biotyper (4613 entries with 5 *S. suis* entries; Bruker Daltonics, Germany). Bruker's threshold identification scores: highly probable species identification: scores of  $\geq 2.300$ ; secure genus and probable species identification: scores between 2.299 and 2.000; probable genus level identification: score of 1.999–1.700; unreliable identification: scores <1.700.

<sup>b</sup>UDB: updated database (after inclusion of three new *S. suis* entries of serotypes 2, 7, and 9).

<sup>c</sup>Same Bruker's threshold identification scores as ODB.

<sup>d</sup>Reliable (highly probable and probable) species identification: scores of  $\geq 2.000$ ; secure genus level identification: scores between 1.999 and 1.700; unreliable identification: scores <1.700.

**TABLE 2 | Percentage of *S. suis* isolates with score values  $\geq 2.300$  based on serotype with the original Bruker's MALDI-TOF MS database (ODB; including 5 *S. suis* entries) and the updated database (UDB; after inclusion of three new *S. suis* entries).**

Serotype	%ODB (N/total)	%UDB (N/total)	Fold-increase
2	42.9 (12/28)	89.3 (25/28)	2.1
7	20 (1/5)	100 (5/5)	5
9	18.8 (13/69)	88.4 (61/69)	4.7
Less prevalent	19.2 (5/26)	55.6 (15/27)	2.9

(5/26; **Table 2**). Regarding the host from which *S. suis* was recovered, no significant differences (Z-test, adjust *p*-values – Bonferroni methods) were observed in the percentage of pig (18/62, 29%) and wild animals (13/63, 20.63%) isolates identified as species level with a score value  $\geq 2.300$  by MALDI-TOF MS.

The agreement between Rapid ID 32 STREP and MALDI-TOF MS results (considering isolates with score values of at least 2.300) was slight ( $\kappa = 0.035$ ) in the present study ( $n = 129$ ). Comparison of the MALDI-TOF MS quantitative results (score values) of the three API ID 32 STREP identification categories revealed that those isolates classified as unreliable identification (unacceptable identification or misidentified by the Rapid ID 32 STREP strips) gave significantly lower score values (ANOVA test,  $p < 0.05$ ) than those isolates included in the other two identification categories (identification at species and genus level).

To improve identification of *S. suis*, we constructed spectra containing information on peak masses and peak intensities of three clinical isolates of serotypes 2, 7, and 9, which were used to improve the original reference database (ODB) and create an UDB. After the addition of the three new *S. suis* profiles, most isolates presented statistically significant higher score values (Paired-sample *t*-test,  $p < 0.001$ ) when their spectra were matched

with the UDB (mean score: 2.65) than those obtained using the ODB (mean score: 2.182). These higher score values resulted in an increase in the number (106; 82.2%) of *S. suis* isolates with scores values  $\geq 2.300$  (**Table 1**). Considering serotype, the 89.3% (25/28), 100% (5/5), 88.4% (61/69) and 55.6% (15/27) of isolates of serotypes 2, 7, 9 and less prevalent serotypes, respectively, had score values  $\geq 2.300$  (**Table 2**). The agreement between Rapid ID 32 STREP and MALDI-TOF MS results slightly increased ( $\kappa = 0.286$ ) when the spectra of *S. suis* isolates were matched against the UDB database.

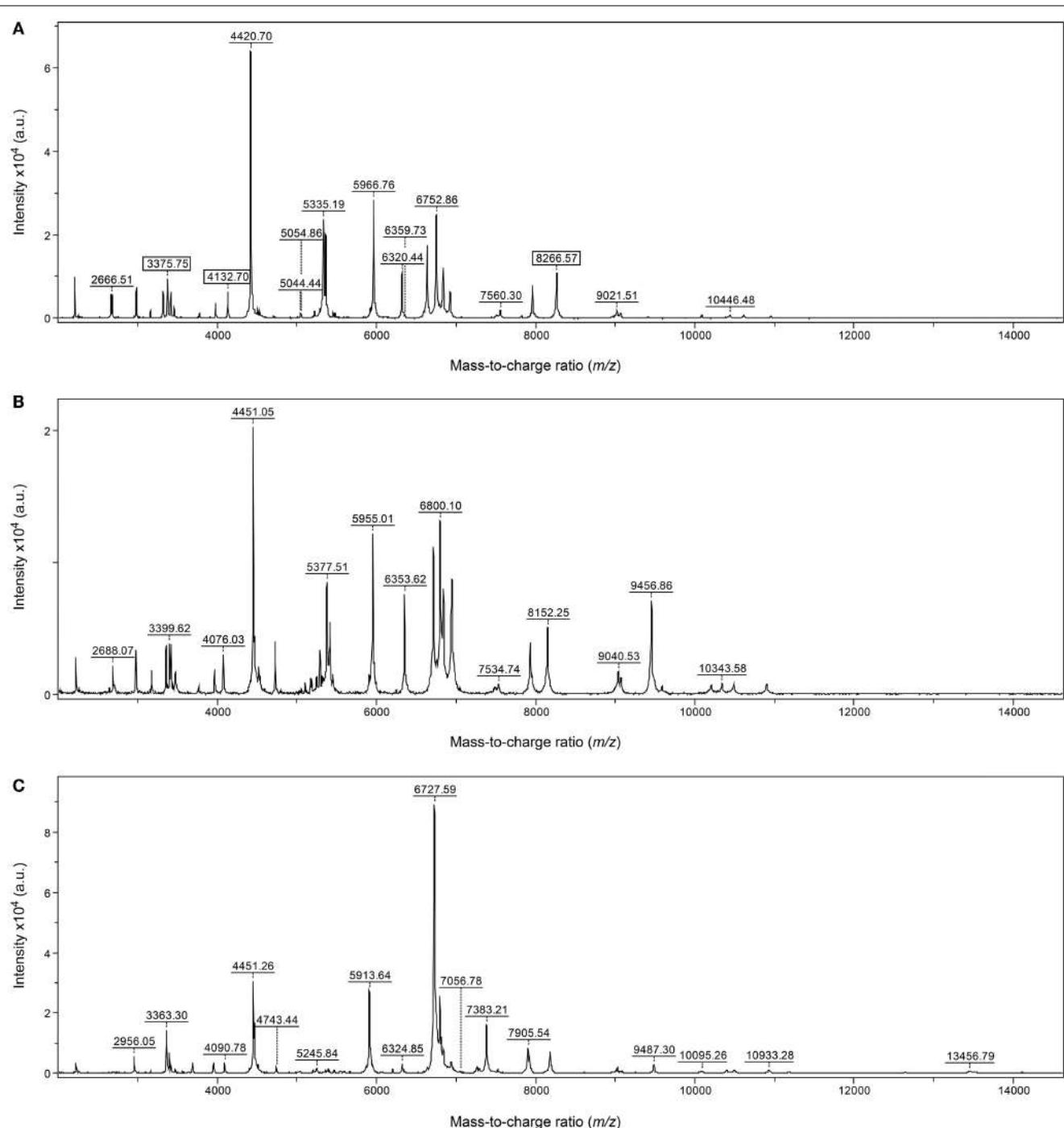
None of the *S. suis* isolates included in the present study was misidentified with *Streptococcus porcinus* and *Streptococcus dysgalactiae* subsp. *equisimilis*, the other two most important pathogens of swine belonging to this genus, included in the ODB. The spectra of *S. suis* presented, regardless their serotype, distinct MS peaks (e.g., *m/z* 3377, 4133, and 8267) which allowed its differentiation from these two streptococcal species by MALDI-TOF MS technique (**Figure 1**).

## Discussion

MALDI-TOF MS is a powerful tool that has attracted worldwide attention for its direct and rapid discrimination and identification of different microorganisms. In the present work, the performance of MALDI-TOF MS system for *S. suis* identification was evaluated using as gold standard a species-specific PCR [this system has been widely used for reliable identification of *S. suis* (5, 19–22)]. These results were compared with the identification results obtained previously using the commercial identification system Rapid ID 32 Strep (bioMérieux). Using the threshold score established by the manufacturer for highly probable species identification (score  $\geq 2.300$ ), the MALDI-TOF MS system was able to identify only a quarter (24.03%) of the *S. suis* isolates. This percentage is relatively low compared with the 63.6% of *S. suis* isolates ( $n = 82$ ) that gave acceptable to excellent identification results using the commercial Rapid ID 32 STREP strips. The 31 *S. suis* isolates properly identified by MALDI-TOF MS were also correctly identified by the Rapid ID 32 STREP system, while the remaining 51 isolates correctly identified by this last system gave score values between 2.000 and 2.299 (considered probable species identification).

Serotype 2 includes most of the clinical pig isolates worldwide (9), but other serotypes, such as, serotypes 7 or 9 are also epidemiologically relevant (8). Serotype 2 showed a higher percentage of isolates with score values  $\geq 2.300$  (12/28; 42.85%) by MALDI-TOF MS system than serotypes 9 (13/69; 18.84%) and 7 (1/5; 20%) (**Table 2**). The relatively low performance of MALDI-TOF MS system for the identification of isolates of the latter serotypes might represent a limitation for the routinely identification of this pathogen.

As the quality and reliability of the identification by MALDI-TOF MS depends on the quality and amount of reference spectra present in the database, the original Bruker's reference database (ODB) was implemented by adding the spectra obtained from three additional clinical isolates of serotypes 2, 7, and 9, creating an updated reference database (UDB). After the implementation of the database, the MALDI TOF analysis of the 129 strains of *S. suis* used in this study resulted in an increase in the number of



**FIGURE 1 | MALDI-TOF MS mass spectra of *S. suis* (A), *S. porcinus* (B), and *S. dysgalactiae* subsp. *equisimilis* (C).** The y axis represents the relative intensities (%) and the x axis represents the  $m/z$  values (masses in Da).

isolates (from 24 to 82.2%; **Table 1**) with scores values  $\geq 2.300$ . This increase was observed in all serotypes but was more evident in isolates of serotypes 7 and 9 in which the percentage of isolates accurately identified increased 5- and 4.7-fold times, respectively (**Table 2**). These results highlight the importance of the implementation of MALDI-TOF MS database to improve the discriminatory power of this identification approach, especially in bacteria with

high genetic heterogeneity as *S. suis* (23) which could exhibit high diverse protein profiles.

Previous studies have suggested the convenience of using less stringent cut-off values to improve the accuracy of MALDI-TOF MS at the species level (24–26). Considering the PCR and MALDI-TOF results of the present study in which all but one of the 129 isolates in our study (99.2%) were identified as *S. suis* by

MALDI-TOF MS as first option irrespective of their score values, we suggest using a less restrictive threshold score of  $\geq 2.000$  for a reliable species identification of *S. suis*. According to this cut-off value, a total of 125 isolates (96.9%) were correctly identified as *S. suis* (**Table 1**) using the UDB, indicating an excellent performance of MALDI-TOF MS.

MALDI TOF system was able to discriminate between *S. suis* and *S. porcinus* and *S. dysgalactiae* subsp. *equisimilis* profiles included in the Biotyper Database. In fact, these bacterial species can also be frequently isolated from diseased-piglets (1). On the other hand, *Streptococcus plurextorum*, *Streptococcus porci* and *Streptococcus porcorum* have been isolated from pigs in the last years (27–29). These species are phylogenetically close related to *S. suis*, but their type strains are not yet available in the MALDI Biotyper database. Therefore, further investigations are required

## References

- Gottschalk M. Streptococcosis. 10th ed. In: Zimmerman J, Karriker J, Ramirez L, Schwartz A, Stevenson K, editors. *Diseases of Swine*. Ames, Iowa: Wiley-Blackwell Publishing (2012). p. 841–55.
- Cruz Colque JI, Devriese LA, Haesbrouck F. Streptococci and enterococci associated with tonsils of cattle. *Lett Appl Microbiol* (1993) **16**:72–4. doi:10.1111/j.1472-765X.1993.tb00346.x
- Gottschalk M, Segura M, Xu J. *Streptococcus suis* infections in humans: the Chinese experience and the situation in North America. *Anim Health Res Rev* (2007) **8**:29–45. doi:10.1017/S1466252307001247
- Salas SI, Lammler C, Devriese LA. Serotypes and putative virulence markers of *Streptococcus suis* isolates from cats and dogs. *Res Vet Sci* (1994) **57**:259–61. doi:10.1016/0034-5288(94)90070-1
- Liu Z, Zheng H, Gottschalk M, Bai X, Lan R, Ji S, et al. Development of multiplex PCR assays for the identification of the 33 serotypes of *Streptococcus suis*. *PLoS One* (2013) **8**(8):e72070. doi:10.1371/journal.pone.0072070
- Okwumabua O, O'Connor M, Shull E. A polymerase chain reaction (PCR) assay specific for *Streptococcus suis* based on the gene encoding the glutamate dehydrogenase. *FEMS Microbiol Lett* (2003) **218**:79–84. doi:10.1111/j.1574-6968.2003.tb11501.x
- Wisselink HJ, Joosten JJ, Smith HE. Multiplex PCR assays for simultaneous detection of six major serotypes and two virulence-associated phenotypes of *Streptococcus suis* in tonsillar specimens from pigs. *J Clin Microbiol* (2002) **40**:2922–9. doi:10.1128/JCM.40.8.2922-2929.2002
- Goyette-Desjardins G, Auger JP, Xu J, Segura M, Gottschalk M. *Streptococcus suis*, an important pig pathogen and emerging zoonotic agent—an update on the worldwide distribution based on serotyping and sequence typing. *Emerg Microbes Infect* (2014) **3**:e45. doi:10.1038/emi.2014.45
- Gottschalk M, Xu J, Calzas C, Segura M. *Streptococcus suis*: a new emerging or an old neglected zoonotic pathogen? *Future Microbiol* (2010) **5**:371–91. doi:10.2217/fmb.10.2
- Barberis C, Almuzara M, Join-Lambert O, Ramirez MS, Famiglietti A, Vay C. Comparison of the Bruker MALDI-TOF mass spectrometry system and conventional phenotypic methods for identification of Gram-positive rods. *PLoS One* (2014) **9**(9):e106303. doi:10.1371/journal.pone.0106303
- Cherkaoui A, Emonet S, Fernández J, Schorderet D, Schrenzel J. Evaluation of matrix-assisted laser desorption ionization-time of flight mass spectrometry for rapid identification of Beta-hemolytic streptococci. *J Clin Microbiol* (2011) **49**:3004–5. doi:10.1128/JCM.00240-11
- Hinse D, Vollmer T, Erhard M, Welker M, Moore ER, Kleesiek K, et al. Differentiation of species of the *Streptococcus bovis/equinus*-complex by MALDI-TOF mass spectrometry in comparison to sodA sequence analyses. *Syst Appl Microbiol* (2011) **34**:52–7. doi:10.1016/j.syapm.2010.11.010
- Ikryannikova LN, Filimonova AV, Malakhova MV, Savinova T, Filimonova O, Ilina EN, et al. Discrimination between *Streptococcus pneumoniae* and *Streptococcus mitis* based on sorting of their MALDI mass spectra. *Clin Microbiol Infect* (2013) **19**:1066–71. doi:10.1111/1469-0691.12113
- Werno AM, Christner M, Anderson TP, Murdoch DR. Differentiation of *Streptococcus pneumoniae* from non-pneumococcal streptococci of the *Streptococcus mitis* group by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J Clin Microbiol* (2012) **50**:2863–7. doi:10.1128/JCM.00508-12
- Woods K, Beighton D, Klein JL. Identification of the '*Streptococcus anginosus* group' by matrix-assisted laser desorption ionization – time-of-flight mass spectrometry. *J Med Microbiol* (2014) **63**:1143–7. doi:10.1099/jmm.0.076653-0
- Drevinek M, Dresler J, Klimentova J, Pisa L, Hubalek M. Evaluation of sample preparation methods for MALDI-TOF MS identification of highly dangerous bacteria. *Lett Appl Microbiol* (2012) **55**:40–6. doi:10.1111/j.1472-765X.2012.03255.x
- Alatoom AA, Cunningham SA, Ihde SM, Mandrekar J, Patel R. Comparison of direct colony method versus extraction method for identification of Gram-positive cocci by use of Bruker Biotyper matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J Clin Microbiol* (2011) **49**:2868–73. doi:10.1128/JCM.00506-11
- Rettinger A, Krupka I, Grunwald K, Dyachenko V, Fingerle V, Konrad R, et al. *Leptospira* spp. strain identification by MALDI TOF MS is an equivalent tool to 16S rRNA gene sequencing and multi locus sequence typing (MLST). *BMC Microbiol* (2012) **12**:185. doi:10.1186/1471-2180-12-185
- Gottschalk M, Lacouture S, Bonifait L, Roy D, Fittipaldi N, Grenier D. Characterization of *Streptococcus suis* isolates recovered between 2008 and 2011 from diseased pigs in Quebec, Canada. *Vet Microbiol* (2013) **162**:819–25. doi:10.1016/j.vetmic.2012.10.028
- Kerdzin A, Akeda Y, Hatrongjit R, Detchawna U, Sekizaki T, Hamada S, et al. *Streptococcus suis* serotyping by a new multiplex PCR. *J Med Microbiol* (2014) **63**:824–30. doi:10.1099/jmm.0.069757-0
- Nga TV, Nghia HD, Tu le TP, Diep TS, Mai NT, Chau TT, et al. Real-time PCR for detection of *Streptococcus suis* serotype 2 in cerebrospinal fluid of human patients with meningitis. *Diagn Microbiol Infect Dis* (2011) **70**:461–7. doi:10.1016/j.diagmicrobio.2010.12.015
- Wang K, Sun X, Lu C. Development of rapid serotype-specific PCR assays for eight serotypes of *Streptococcus suis*. *J Clin Microbiol* (2012) **50**:3329–34. doi:10.1128/JCM.01584-12
- Vela AI, Goyache J, Tarradas C, Luque I, Mateos A, Moreno MA, et al. Analysis of genetic diversity of *Streptococcus suis* clinical isolates from pigs in Spain by pulsed-field gel electrophoresis. *J Clin Microbiol* (2003) **41**:2498–502. doi:10.1128/JCM.41.6.2498-2502.2003
- Bizzini A, Durussel C, Bille J, Greub G, Prod'hom G. Performance of matrix-assisted laser desorption ionization-time of flight mass spectrometry for identification of bacterial strains routinely isolated in a clinical microbiology laboratory. *J Clin Microbiol* (2010) **48**:1549–54. doi:10.1128/JCM.01794-09
- Cherkaoui A, Emonet S, Fernandez J, Schorderet D, Schrenzel J. Evaluation of matrix-assisted laser desorption ionization-time of flight mass spectrometry for rapid identification of b-hemolytic streptococci. *J Clin Microbiol* (2011) **49**:3004–5. doi:10.1128/JCM.00240-11

to compare the spectra profiles among these closely related streptococci.

In summary, MALDI-TOF MS represent a rapid, accurate and cost-saving method and a reliable alternative to PCR-based methods for routinely identification of *S. suis* isolates from both human and animal origins.

## Acknowledgments

The authors thank Almudena Casamayor and Elisa Pulido for their invaluable technical assistance. MPS is recipient of a technical support staff for scientific infrastructures of the Moncloa Campus of International Excellence (Programa CEI09-0019) and Ministerio de Educación, Cultura y Deporte (Orden PRE/1996/2009 de 20 de Julio).

26. Neville SA, Lecordier A, Ziochos H, Chater MJ, Gosbell IB, Maley MW, et al. Utility of matrix-assisted laser desorption ionization-time of flight mass spectrometry following introduction for routine laboratory bacterial identification. *J Clin Microbiol* (2011) **49**:2980–4. doi:10.1128/JCM.00431-11
27. Vela AI, Casamayor A, Sánchez Del Rey V, Domínguez L, Fernández-Garayzábal JF. *Streptococcus plurextorum* sp. nov., isolated from pigs. *Int J Syst Evol Microbiol* (2009) **59**:504–8. doi:10.1099/ijss.0.65636-0
28. Vela AI, Perez M, Zamora L, Palacios L, Domínguez L, Fernández-Garayzábal JF. *Streptococcus porci* sp. nov., isolated from swine sources. *Int J Syst Evol Microbiol* (2010) **60**:104–8. doi:10.1099/ijss.0.011171-0
29. Vela AI, Sanchez V, Mentaberre G, Lavin S, Domínguez L, Fernández-Garayzábal JF. *Streptococcus porcorum* sp. nov., isolated from domestic and wild pigs. *Int J Syst Evol Microbiol* (2011) **61**:1585–9. doi:10.1099/ijss.0.022657-0

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The Guest Associate Editor Andres M. Perez declares that, despite having collaborated with the author Lucas Dominguez, the review process was handled objectively and no conflict of interest exists.

Copyright © 2015 Pérez-Sancho, Vela, García-Seco, Gottschalk, Domínguez and Fernández-Garayzábal. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Evaluation of a High-Intensity Green Fluorescent Protein Fluorophage Method for Drug- Resistance Diagnosis in Tuberculosis for Isoniazid, Rifampin, and Streptomycin

Xia Yu<sup>1</sup>, Yunting Gu<sup>1</sup>, Guanglu Jiang<sup>1</sup>, Yifeng Ma<sup>1</sup>, Liping Zhao<sup>1</sup>, Zhaogang Sun<sup>1</sup>, Paras Jain<sup>2</sup>, Max O'Donnell<sup>3,4</sup>, Michelle Larsen<sup>2</sup>, William R. Jacobs Jr<sup>2,5</sup> and Hairong Huang<sup>1\*</sup>

<sup>1</sup> National Clinical Laboratory on Tuberculosis, Beijing Key Laboratory for Drug-Resistant Tuberculosis Research, Beijing Chest Hospital, Beijing Tuberculosis and Thoracic Tumor Institute, Capital Medical University, Beijing, China, <sup>2</sup> Department of Microbiology and Immunology, Albert Einstein College of Medicine, Bronx, NY, USA, <sup>3</sup> Division of Pulmonary, Allergy, and Critical Care Medicine, Columbia University Medical Center, New York, NY, USA, <sup>4</sup> Department of Epidemiology, Mailman School of Public Health, Columbia University Medical Center, New York, NY, USA, <sup>5</sup> Howard Hughes Medical Institute, Chevy Chase, MD, USA

## OPEN ACCESS

### Edited by:

Julio Alvarez,  
University of Minnesota, USA

### Reviewed by:

Kuangnan Xiong,  
Health Research, Inc., USA  
Raju Mukherjee,  
Indian Institute of Science Education  
and Research, India

### \*Correspondence:

Hairong Huang  
huanghairong@tb123.org

### Specialty section:

This article was submitted to  
Infectious Diseases,  
a section of the journal  
*Frontiers in Microbiology*

Received: 25 August 2015

Accepted: 31 May 2016

Published: 17 June 2016

### Citation:

Yu X, Gu Y, Jiang G, Ma Y, Zhao L, Sun Z, Jain P, O'Donnell M, Larsen M, Jacobs WR Jr and Huang H (2016) Evaluation of a High-Intensity Green Fluorescent Protein Fluorophage Method for Drug- Resistance Diagnosis in Tuberculosis for Isoniazid, Rifampin, and Streptomycin. *Front. Microbiol.* 7:922. doi: 10.3389/fmicb.2016.00922

A novel method for detecting drug resistance in *Mycobacterium tuberculosis* using mycobacteriophage  $\phi^2GFP10$  was evaluated with clinical isolates. The phage facilitates microscopic fluorescence detection due to the high expression of green fluorescence protein which also simplifies the operative protocol as well. A total of 128 clinical isolates were tested by the phage assay for isoniazid (INH), rifampin (RIF), and streptomycin (STR) resistance while conventional drug susceptibility test, by MGIT960, was used as reference. The sensitivities of  $\phi^2GFP10$  assay for INH, RIF, and STR resistance detection were 100, 98.2, and 89.3%, respectively while their specificities were 85.1, 98.6, and 95.8%, respectively. The agreement between phage and conventional assay for detecting INH, RIF, and STR resistance was 92.2, 98.4, and 93.0%, respectively. The  $\phi^2GFP10$ -phage results could be available in 2 days for RIF and STR, while it takes 3 days for INH, with an estimated cost of less than \$2 to test all the three antibiotics. The  $\phi^2GFP10$ -phage method has the potential to be a valuable, rapid and economical screening method for detecting drug-resistant tuberculosis.

**Keywords:** tuberculosis, mycobacteriophage, drug resistance, diagnosis,  $\phi^2GFP10$

## INTRODUCTION

Reduction in transmission of drug resistant tuberculosis (TB) and improved patient management requires timely diagnosis of drug resistant bacilli. Presence of drug-resistant *M. tuberculosis* bacilli is confirmed by genotypic and phenotypic drug susceptibility testing (DST). Genotypic approaches with short turnaround time are based on identification of well-known antibiotic resistance-conferring gene mutations. Despite active research, not all TB drugs can have genotypic drug-susceptibility testing since the genetic basis of resistance may be complex or incompletely characterized (Kruuner et al., 2003; Jain et al., 2011; Cui et al., 2013; Zhang et al., 2015). Phenotypic

assays observe the bacterial response to antibiotics *in vitro* without limitation to any particular antibiotic, allele, or working mechanism although the phenotypic DST is often time consuming. The conventional DST, based on solid medium, takes about 4–6 weeks after the isolation of *M. tuberculosis*, on the other hand BACTEC MGIT960 requires 10–14 days after acquiring a positive culture. Therefore, it is imperative to outline a further rapid, accurate, inexpensive DST method for the diagnosis of drug resistance *M. tuberculosis*.

Phage based methods have been used for drug resistance detection of *M. tuberculosis* since about two decades ago, such as bacteriophage D29 (Wilson et al., 1997) and luciferase reporter phages (LRPs) (Jacobs et al., 1993). A novel, high-intensity mycobacteria-specific fluorophage ( $\phi^2GFP10$ ) was described recently with good results in pre-clinical evaluation of drug resistant tuberculosis (Jain et al., 2012).  $\phi^2GFP10$  is an improved second generation fluorescent reporter phage which expresses *gfp* (green fluorescence protein) under the control of *P<sub>L</sub>* promoter of mycobacterium phage L5 (Guo and Ao, 2012; Jain et al., 2012) and have intensity 100 times brighter than the previous generation of fluorescent reporter phages (Piuri et al., 2009). Unlike the LRPs (Jacobs et al., 1993), the  $\phi^2GFP10$  reporter phage does not require exogenous luciferase substrate and can yield more stable and microscopically detectable intensive fluorescence. The presented study has evaluated the performance of this in-house fluorophage method to determine the rifampicin (RIF), isoniazid (INH), and streptomycin (STR) resistance in a high drug-resistant setting. Here we have evaluated the performance of  $\phi^2GFP10$  phage for detecting drug resistance from *M. tuberculosis* clinical isolates, and also developed a new phage assay method using fluorescent microplate reader.

## MATERIALS AND METHODS

### Strains

A total of 128 clinical *M. tuberculosis* strains were isolated from patients with suspected drug-resistant TB patients visiting the Beijing chest hospital (Beijing, China) from April to June 2014. All of the isolates were identified as *M. tuberculosis* complex (MTBC) strains by performing a growth test on 500 µg/ml p-Nitrobenzoic Acid containing Löwenstein-Jensen medium (Tsukamura and Tsukamura, 1964). The drug susceptibility of the isolates was determined by MGIT960 SIRE kits according to the manufacturer's protocol. The critical concentrations used were as follows: INH (0.1 µg/ml), RIF(1.0 µg/ml), STR(1.0 µg/ml). Ultimately, 61, 56, and 56 were defined as resistant to INH, RIF and STR respectively by MGIT960 system. The laboratory *M. tuberculosis* H37Rv (ATCC 27294) strain was used in all batches as control. Technique round was demonstrated in **Figure 1**.

### Phage Stock Preparation

The fluorophage,  $\phi^2GFP10$  was constructed by Dr William R. Jacobs' laboratory in Albert Einstein University (New York, USA). The stocks were prepared by growing *M. smegmatis* strain (mc<sup>2</sup>155) in the 7H9, containing 10% oleic albumin dextrose catalase (OADC), to an optical density of 1.0 detected by spectrophotometer at 600 nm wavelength. Then 300 µl cell suspension and 200 µl  $\phi^2GFP10$  were mixed and incubated at

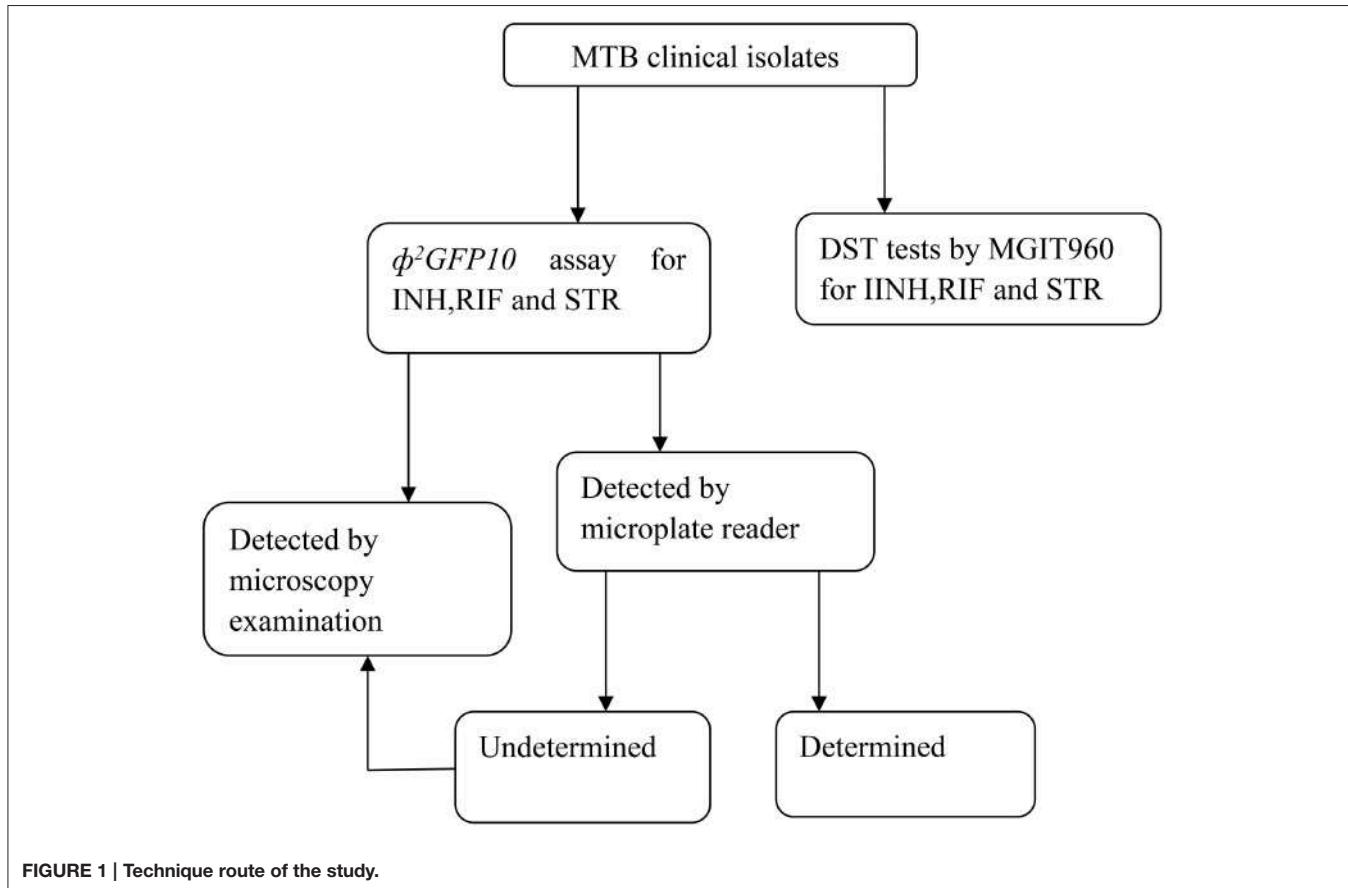
room temperature for 30 min. Subsequently, 3 ml 7H9 containing 0.75% agar was added to the tube, the contents were briefly mixed and poured onto a 100-mm Petri dish containing 7H10-OADC. After the top-agar was solidified, the dish was incubated at 30°C for 48 h. Then 3 ml MP buffer [50 mM Tris (pH 7.6), 150 mM NaCl, 10 mM MgCl<sub>2</sub>, and 2 mM CaCl<sub>2</sub>] was pipetted onto the plates and then incubated on an orbital shaker for 6 h. The buffer was then collected and passed through a 0.22 µm Millipore filter membrane. The phage titers were determined by serial dilution method and the phage density was adjusted to approximately 10<sup>10</sup> plaque forming unit (pfu)/ml for further use. The phage stock was stable when stored at 4°C as no appreciable drop in titer occurred for at least 6 months.

### $\phi^2GFP10$ Assay by Microscopy Examination

INH, RIF, and STR were obtained from Sigma-Aldrich. INH and STR were dissolved in water; RIF was dissolved in DMSO. The stock solutions were 20 mg/ml, 2 mg/ml and 4 mg/ml for INH, RIF, and STR, respectively. All stock solutions were stored at -80°C during experimentation. Fresh colonies on Löwenstein-Jensen media were suspended and homogenized with sterile saline and the cell density was adjusted to 1 McFarland turbidity. A 1 ml aliquot was harvested by centrifugation and the sediment was re-suspended with 250 µl 7H9-OADC in the absence of Tween 80. The desired drugs would be added at this step, when needed, at the following final concentrations: 20 µg/ml for INH, 2 µg/ml for RIF, or 4 µg/ml for STR. The tubes were incubated at 37°C for 24 h, or 48 h specifically for INH tests. From the stock, containing 10<sup>10</sup> pfu fluorophage per ml, 100 µl volume was added to each tube to obtain a multiplicity of infection (MOI) of 100, the tubes were then incubated at 37°C for 16 h. After adding 350 µl of 4% paraformaldehyde in phosphate-buffered saline (PBS) to each tube, the reaction mixtures, left at room temperature for 90 min to ensure killing of the bacteria, were then centrifuged at 12,000 g for 15 min to remove excess phage and media. The pellet was re-suspended with 10 µl 7H9 media. Five micro liter re-suspension was spotted on a glass slide and visualized by a fluorescence microscope. The criterion for resistance with phage assay was the presence of at least one fluorescent bacillus per 50 high power fields (Rondon et al., 2011).

### $\phi^2GFP10$ Assay by Microplate Reader

To detect fluorescence of  $\phi^2GFP10$  in a 96 well format relative light unit (RLU) of fluorescence was measured by a fluorescent microplate reader (Berthold LB970, Germany). The wavelength of excitation and emitted light was 485 and 535 nm respectively and the counting time was 5 s. Each test was performed in triplicate and the mean value was used for interpretation. After removing excess phage and media by centrifuging, the pellet was re-suspended in 200 µl PBS, and then, the RLU of the plate was read. The influence of drug exposure on the fluorescence production of the bacilli was interpreted by the remaining fluorescence rate (RFR), calculated by the following formula: (reaction counts-blank counts)/(positive control counts -blank counts) × 100% ("reaction" indicates the well with drug exposure; "blank" indicates the well without drug and phage; "positive control" indicates the reaction without drug). Precision



of the assay was determined with quintuple measurements of the following three strains: H37Rv which was susceptible to all the three tested drugs, strain 14,301 which was resistant to all the three tested drugs and strain 14,161 which was resistant to INH and STR but susceptible to RIF. The stability of fluorescence was determined by detecting the same samples at 0, 4, and 24 h, respectively. All the reactions were performed in triplicate.

## Statistical Analysis

According to MGIT960 phenotypic outcomes, receiver-operation characteristic (ROC) analysis was performed using SPSS (version 17.0) and used to define a cutoff value for phage plate reader assay. Kappa value was calculated to compare the method between  $\phi^2GFP10$  assay and MGIT960 phenotypic outcomes. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) was calculated from the following URL: <http://vassarstats.net>. The discriminative power of fluorescent microplate reader for drug resistance was analyzed using ROC curve and area under the curve (AUC). The optimal cut-off value was defined as the one with the least  $(1 - \text{sensitivity})^2 + (1 - \text{specificity})^2$  value (Shu et al., 2013).

## RESULTS

### $\phi^2GFP10$ Assay by Microscopy

According to the outcomes of MGIT960, the sensitivity of  $\phi^2GFP10$  assay for detecting INH, RIF and STR resistance were

100, 98.2, and 89.3% respectively whereas the specificities were 85.1, 98.6, and 95.8% respectively. The agreement between phage assay and MGIT960 for detecting INH, RIF and STR resistance was 92.2, 98.4, and 93.0%, respectively (see Table 1). The Kappa coefficient between  $\phi^2GFP10$  assay and MGIT960 phenotypic outcomes for INH, RIF and STR were 0.85, 0.97, and 0.856, respectively.

### $\phi^2GFP10$ Assay by Fluorescent Microplate Reader

In the preliminary validation of the assay among the 3 tested isolates, the stability tests demonstrated that the fluorescence was stable for at least 24 h at 4°C (see Table 2). The coefficient of variation (CV) of RLU among the tested isolates in quintuple ranged from 1.22 to 6.73%, with the mean value at  $4.67 \pm 1.26\%$ . In the validation assays using clinical isolates, in contrast to the phenotypic DST, the AUC under the ROC curve of INH, RIF and STR were 0.957, 0.960, 0.917, and the optimal cutoff value for each drug has been listed in Table 3 and Figure 2. The sensitivity of  $\phi^2GFP10$  assay by plate reader for detecting INH, RIF and STR resistance, according to the cutoff values, were 86.9, 89.3, and 83.9%, respectively, while the specificity for INH, RIF and STR resistance detection were 94.0, 90.3, and 87.5%, respectively. The Kappa coefficient between  $\phi^2GFP10$  assay and MGIT960 phenotypic outcomes for INH, RIF and STR were 0.73, 0.79, and 0.71, respectively. Furthermore, evaluation of the RFR value

**TABLE 1 | Results and performance parameters:  $\phi^2$ GFP10 assay by microscopy vs. MGIT960 system.**

Antibiotics	No. of isolates				% (95% CI)			
	MGIT960 R		MGIT960 S		Total	Sensitivity	Specificity	PPV
	Phage R	Phage S	Phage R	Phage S				
INH	61	0	10	57	128	100.0 (92.6–100.0)	85.1 (73.8–92.2)	85.9 (75.2–92.7)
RIF	55	1	1	71	128	98.2 (89.2–99.9)	98.6 (91.5–99.9)	98.2 (89.2–99.9)
STR	50	6	3	69	128	89.3 (77.4–95.6)	95.8 (87.5–98.9)	94.3 (83.4–98.5)

R, resistant; S, sensitive.

**TABLE 2 | The stability and precision of fluorescent microplate reader assay.**

Strains	Time (h)	The counts of Fluorescent microplate reader (mean±SD)				
		Blank control	Positive control	INH (RFR*)	RIF (RFR*)	STR (RFR*)
H37Rv	0	5476.00 ± 342.39	9902.00 ± 503.95	6926.00 ± 84.44 (32.76)*	6296.00 ± 267.82 (18.53)	6540.00 ± 256.12 (24.04)
	4	5336.00 ± 359.35	9626.00 ± 515.10	6792.00 ± 181.30 (33.94)	6974.00 ± 323.77 (26.53)	6766.00 ± 253.24 (33.33)
	24	5218.00 ± 299.03	10088.00 ± 624.40	6414.00 ± 283.16 (24.60)	6364.00 ± 335.01 (23.53)	6476.00 ± 302.62 (25.83)
14031	0	5144.00 ± 184.07	8650.00 ± 347.20	8058.00 ± 297.19 (83.11)	8688.00 ± 413.55 (108.08)	8314.00 ± 533.32 (90.42)
	4	5140.00 ± 252.59	8410.00 ± 312.17	7992.00 ± 342.15 (87.22)	8302.00 ± 452.46 (97.70)	8352.00 ± 448.74 (98.23)
	24	5318.00 ± 332.60	8558.00 ± 257.04	8230.00 ± 278.57 (89.88)	8470.00 ± 527.4 (97.28)	8516.00 ± 310.85 (98.70)
14161	0	4568.00 ± 273.17	8354.00 ± 417.17	7694.00 ± 327.46 (82.57)	5380.00 ± 185.07 (21.45)	7612.00 ± 353.09 (80.40)
	4	4576.00 ± 298.63	8338.00 ± 363.00	7566.00 ± 458.34 (79.48)	5768.00 ± 206.57 (31.69)	7650.00 ± 401.25 (81.71)
	24	4688.00 ± 80.44	8502.00 ± 378.31	7948.00 ± 382.84 (85.47)	5320.00 ± 257.97 (16.57)	7884.00 ± 477.94 (83.80)

\*Remaining fluorescence rate (RFR) = (reaction -blank)/(positive control-blank) × 100% ("reaction" indicates the well with drug exposure; "blank" indicates the well without drug and phage; "positive control" indicates the reaction without drug).

according to the susceptibility predictability demonstrated that when the RFR is lower than 45% the NPV for INH, RIF and STR were 98.0, 98.2, and 87.5%, respectively. Similarly, when the ratio was higher than 60%, the PPV for INH, RIF and STR were 92.6, 97.9, and 95.3%, respectively. When the ratio was between 45 and 60%, then the outcome should be interpreted as undetermined (see Table 4).

## DISCUSSION

Significant efforts are being made to develop rapid, simple and accurate tests for drug-resistant *M. tuberculosis*. One of such technologies currently been worked at is based on using bacteriophages (Banaiee et al., 2008; Piuri et al., 2009; Rondon et al., 2011; Jain et al., 2012; Sivaramakrishnan et al., 2013). LRP have been used for identifying drug resistance by detecting the luminescence produced by LRP after their infection in *M. tuberculosis*. These assays classify the samples as drug-susceptible if no luminescence is detected in the luminescence drug-containing samples. A meta-analysis showed that the sensitivity and specificity of luciferase reporter phage for rapid detection of RIF resistance in *M. tuberculosis* ranged from 92 to 100% and 89 to 100% (Minion and Pai, 2010). Although this method has shown high sensitivity and high specificity, a relatively expensive luminometer is required. The newly constructed phage  $\phi^2$ GFP10 facilitates the microscopic signal detection since it yields 100-folds higher fluorescence signal per-cell than any previously described reporter phages (Jain et al.,

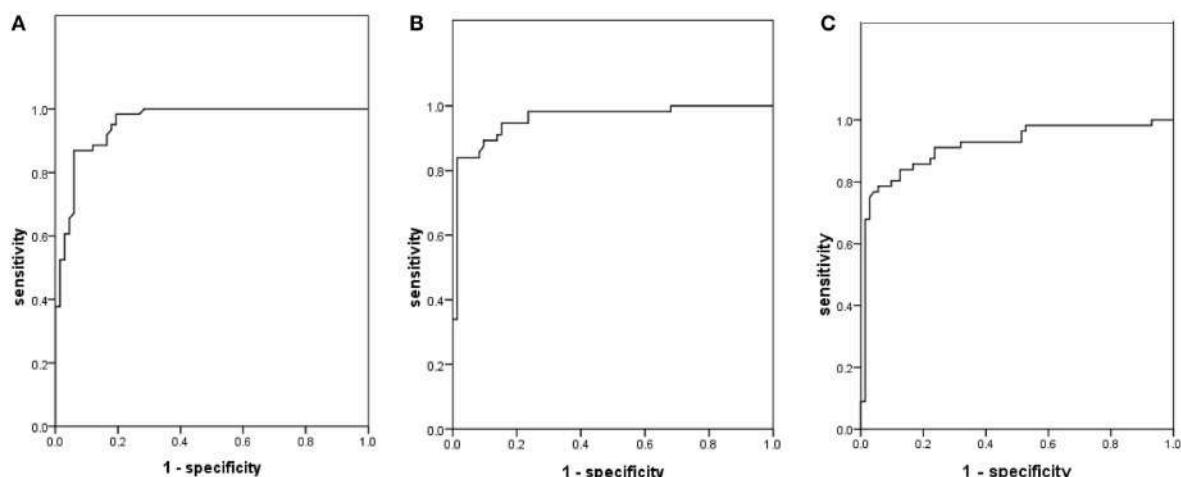
2011). Meanwhile, as the intensive fluorescence increased the contrast between reaction and background, tedious washing steps were not necessary for phage assay, so the  $\phi^2$ GFP10 assay was easier to perform than the other phage assays (Rondon et al., 2011).

In this study, drug resistance detection by  $\phi^2$ GFP10 assay yielding applicable sensitivities and specificities, especially for RIF. A recent report also obtained great consistency between  $\phi^2$ GFP10 assay and GeneXpert for RIF resistance diagnosis (O'Donnell et al., 2015). Fluorescent microscopy may be a limiting factor for  $\phi^2$ GFP10 assay application due to its cost, whereas less expensive and simpler fluorescent microscopes using light emitting diodes (LED) have been applied for smear test of tuberculosis in various clinical laboratories (Marzouk et al., 2013; Reza et al., 2013; Xia et al., 2013) and they could readily be adapted for  $\phi^2$ GFP10 assay. The reagents for fluorophage growth and amplification are inexpensive, safe, and universally available. The total cost for the 3-drug test per strain was less than US \$2. Additionally, fixation of the sample with paraformaldehyde prior to analysis overcomes the substantial biosafety concerns.

In our study, even though a longer drug exposure time of 48 h for INH, the specificity was still lower than those of RIF and STR whose drug exposure time was 24 h. Those observations were consistent with other mycobacteriophage-based assays (Rondon et al., 2011). The anti-TB mechanism of INH involves inhibiting the synthesis of mycolic acid of cell wall (Eltringham et al., 1999), which is a slow process compared with the quickly bactericidal activities of RIF and STR, therefore the bacilli can stay alive

**TABLE 3 | Results and performance parameters: fluorescent microplate reader vs. MGIT960 system.**

Antibiotics	AUC	Cutoff (%)	% (95% CI)			
			sensitivity	specificity	PPV	NPV
INH	0.957	55.00	86.9 (75.2–93.8)	94.0 (84.7–98.1)	93.0 (82.2–97.7)	88.7 (78.5–94.7)
RIF	0.960	53.60	89.3 (77.4–95.6)	90.3 (80.4–95.7)	87.7 (75.7–94.5)	91.5 (81.9–96.5)
STR	0.917	48.65	83.9 (71.2–91.9)	87.5 (77.1–93.8)	83.9 (71.2–91.9)	87.5 (77.1–93.8)

**FIGURE 2 |** Receiver operating characteristic curves of fluorescent microplate reader in detecting drug resistance according to BactecMGIT 960 system. (A) INH; (B) RIF; (C) STR.**TABLE 4 |** Results and performance parameters by sectional outcomes interpretation: fluorescent microplate reader vs. MGIT960 system.

Drug	RFR	No. of isolates	MGIT960 DST outcomes	
			R	S
INH	<45%	51	1	50
	45–60%	23	10	13
	>60%	54	50	4
RIF	<45%	56	1	55
	45–60%	25	9	16
	>60%	47	46	1
STR	<45%	64	8	56
	45–60%	21	7	14
	>60%	43	41	2

R, resistant; S, sensitive; RFR, remaining fluorescence rate.

for quite a while even when they are sensitive to INH. The fluorophage could infect such bacilli and produce fluorescence signal which decreased the specificity of the assay. To enhance the bactericidal activity of INH, we increased the INH concentration for phage assay to 20 µg/ml, which was dramatically higher than those of previous studies (Xia et al., 2013; Zhang et al., 2015). However, the specificity was still lower for INH than those of RIF and STR, which indicating the INH tolerance deviation for clinical strains was universal.

First-line-drug ethambutol (EMB) also affects the cell wall construction in *M. tuberculosis*, and is a bacteriostatic drug rather than a bactericidal drug. We attempted to develop a  $\phi^2GFP10$  assay for EMB, but the pilot test demonstrated that even an exposure concentration of 50 µg/ml could not lead to obvious CFU loss. A previous study reported low concordance (53%) between the phage amplified biologically assay and resistance ratio method for EMB, which they attributed to the bacteriostatic role of EMB both *in vitro* and in a macrophage model (Eltringham et al., 1999). Therefore, the  $\phi^2GFP10$  assay might only be feasible to test the rapid-action bacteriocidal drugs.

Although we acquired favorable sensitivity and specificity for  $\phi^2GFP10$  assay by microscopy, we found that microscopic examination process is experience dependent, and a lot of time is required when handing a batch of samples. So we developed a phage assay to interpret the outcomes simply by fluorescence plate reader, which can read a 96-well plate in 10 min. We monitored the fluorescence diminishing after the bacilli were exposed to the tested drugs. When categorized the RFR lower than 45% as “susceptible,” greater than 60% as “resistant,” while between 45 and 60% as undetermined, plate reader could interpret over 80% of the test strains successfully for all the 3 drugs, with 95.24% (100/105), 98.06% (101/103), and 90.65% (97/107) consistency with MGIT960 system for INH, RIF, and STR susceptibility tests, respectively. Therefore, we suggest interpreting the outcomes in sections to increase the reliability of the assay. For undetermined samples, we

suggest to recheck by microscopy to produce more reliable results.

Like other fluoromycobacteriophages,  $\phi^2$ GFP10 assay could not discriminate between *M. tuberculosis* and other members of MTBC species (Riska et al., 1997). For the settings with high isolation rate of nontuberculous mycobacteria, one should be cautious to report drug resistance without MTBC identification.

In summary, the  $\phi^2$ GFP10 assay is a rapid, simple and inexpensive DST method. In clinical laboratories with limited resources, the  $\phi^2$ GFP10 based assay has the potential to be used for drug resistance diagnosis of tuberculosis.

## AUTHOR CONTRIBUTIONS

HH designed the study. HH and XY wrote the paper. WJ modified the paper. XY and YG performed the  $\phi^2$ GFP10

## REFERENCES

- Banaiee, N., January, V., Barthus, C., Lambrick, M., Roditi, D., Behr, M. A., et al. (2008). Evaluation of a semi-automated reporter phage assay for susceptibility testing of *Mycobacterium tuberculosis* isolates in South Africa. *Tuberculosis (Edinb)* 88, 64–68. doi: 10.1016/j.tube.2007.08.006
- Cui, Z., Wang, J., Lu, J., Huang, X., Zheng, R., and Hu, Z. (2013). Evaluation of methods for testing the susceptibility of clinical *Mycobacterium tuberculosis* isolates to pyrazinamide. *J. Clin. Microbiol.* 51, 1374–1380. doi: 10.1128/jcm.03197-12
- Eltringham, I. J., Wilson, S. M., and Drobnewski, F. A. (1999). Evaluation of a bacteriophage-based assay (phage amplified biologically assay) as a rapid screen for resistance to isoniazid, ethambutol, streptomycin, pyrazinamide, and ciprofloxacin among clinical isolates of *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* 37, 3528–3532.
- Guo, S., and Ao, Z. (2012). Phage in the diagnosis and treatment of tuberculosis. *Front. Biosci.* 17, 2691–2697. doi: 10.2741/4080
- Jacobs, W. R. Jr., Barletta, R. G., Urdan, R., Chan, J., Kalkut, G., Sosne, G., et al. (1993). Rapid assessment of drug susceptibilities of *Mycobacterium tuberculosis* by means of luciferase reporter phages. *Science* 260, 819–822. doi: 10.1126/science.8484123
- Jain, P., Hartman, T. E., Eisenberg, N., O'Donnell, M. R., Kriakov, J., Govender, K., et al. (2012). phi(2)GFP10, a high-intensity fluorophage, enables detection and rapid drug susceptibility testing of *Mycobacterium tuberculosis* directly from sputum samples. *J. Clin. Microbiol.* 50, 1362–1369. doi: 10.1128/jcm.06192-11
- Jain, P., Thaler, D. S., Maiga, M., Timmins, G. S., Bishai, W. R., Hatfull, G. F., et al. (2011). Reporter phage and breath tests: emerging phenotypic assays for diagnosing active tuberculosis, antibiotic resistance, and treatment efficacy. *J. Infect. Dis.* 204(suppl. 4), S1142–S1150. doi: 10.1093/infdis/jir454
- Kruuner, A., Jureen, P., Levina, K., Ghebremichael, S., and Hoffner, S. (2003). Discordant resistance to kanamycin and amikacin in drug-resistant *Mycobacterium tuberculosis*. *Antimicrob. Agents Chemother.* 47, 2971–2973. doi: 10.1128/AAC.47.9.2971-2973.2003
- Marzouk, M., Ferjani, A., Dhaou, M., Ali, M. H., Hannachi, N., and Bouakadida, J. (2013). Comparison of LED and conventional fluorescence microscopy for detection of acid-fast bacilli in an area with high tuberculosis incidence. *Diagn. Microbiol. Infect. Dis.* 76, 306–308. doi: 10.1016/j.diagmicrobio.2013.03.023
- Minion, J., and Pai, M. (2010). Bacteriophage assays for rifampicin resistance detection in *Mycobacterium tuberculosis*: updated meta-analysis. *Int. J. Tuberc. Lung. Dis.* 14, 941–951.
- O'Donnell, M. R., Pym, A., Jain, P., Munsamy, V., Wolf, A., Karim, F., et al. (2015). A novel reporter phage to detect tuberculosis and rifampin resistance in a high-HIV-burden population. *J. Clin. Microbiol.* 53, 2188–2194. doi: 10.1128/jcm.03530-14
- Piuri, M., Jacobs, W. R. Jr., and Hatfull, G. F. (2009). Fluoromycobacteriophages for rapid, specific, and sensitive antibiotic susceptibility testing of *Mycobacterium tuberculosis*. *PLoS ONE* 4:e4870. doi: 10.1371/journal.pone.0004870
- Reza, L. W., Satyanarayna, S., Enarson, D. A., Kumar, A. M., Sagili, K., Kumar, S., et al. (2013). LED-fluorescence microscopy for diagnosis of pulmonary tuberculosis under programmatic conditions in India. *PLoS ONE* 8:e75566. doi: 10.1371/journal.pone.0075566
- Riska, P. F., Jacobs, W. R. Jr., Bloom, B. R., McKittrick, J., and Chan, J. (1997). Specific identification of *Mycobacterium tuberculosis* with the luciferase reporter mycobacteriophage: use of p-nitro-alpha-acetylamin-beta-hydroxy propiophenone. *J. Clin. Microbiol.* 35, 3225–3231.
- Rondon, L., Piuri, M., Jacobs, W. R. Jr., De Waard, J., Hatfull, G. F., and Takiff, H. E. (2011). Evaluation of fluoromycobacteriophages for detecting drug resistance in *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* 49, 1838–1842. doi: 10.1128/jcm.02476-10
- Shu, C. C., Wu, M. F., Hsu, C. L., Huang, C. T., Wang, J. Y., Hsieh, S. L., et al. (2013). Apoptosis-associated biomarkers in tuberculosis: promising for diagnosis and prognosis prediction. *BMC Infect. Dis.* 13:45. doi: 10.1186/1471-2334-13-45
- Sivaramakrishnan, G., Subramanyam, B., C. P., and Kumar, V. (2013). Luciferase reporter phage pHAE85 for rapid detection of rifampicin resistance in clinical isolates of *Mycobacterium tuberculosis*. *Asian Pac. J. Trop. Med.* 6, 728–731. doi: 10.1016/s1995-7645(13)60127-3
- Tsukamura, M., and Tsukamura, S. (1964). Differentiation of *Mycobacterium tuberculosis* and *Mycobacterium bovis* by p-nitrobenzoic acid susceptibility. *Tubercl* 45, 64–65. doi: 10.1016/S0041-3879(64)80091-X
- Wilson, S. M., Al-Suwaidi, Z., Mcnerney, R., Porter, J., and Drobnewski, F. (1997). Evaluation of a new rapid bacteriophage-based method for the drug susceptibility testing of *Mycobacterium tuberculosis*. *Nat. Med.* 3, 465–468. doi: 10.1038/nm0497-465
- Xia, H., Song, Y. Y., Zhao, B., Kam, K. M., O'Brien, R. J., Zhang, Z. Y., et al. (2013). Multicentre evaluation of Ziehl-Neelsen and light-emitting diode fluorescence microscopy in China. *Int. J. Tuberc. Lung. Dis.* 17, 107–112. doi: 10.5588/ijtld.12.0184
- Zhang, X., Liu, L., Zhang, Y., Dai, G., Huang, H., and Jin, Q. (2015). Genetic determinants involved in p-aminosalicylic acid resistance in clinical isolates from tuberculosis patients in northern China from 2006 to 2012. *Antimicrob. Agents Chemother.* 59, 1320–1324. doi: 10.1128/aac.03695-14
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Copyright © 2016 Yu, Gu, Jiang, Ma, Zhao, Sun, Jain, O'Donnell, Larsen, Jacobs and Huang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.**

# Detection of *Mycobacterium tuberculosis* in extrapulmonary biopsy samples using PCR targeting IS6110, *rpoB*, and nested-*rpoB* PCR Cloning

Hossein Meghdadi<sup>1</sup>, Azar D. Khosravi<sup>1,2\*</sup>, Ata A. Ghadiri<sup>3,4</sup>, Amir H. Sina<sup>5</sup> and Ameneh Alami<sup>1</sup>

<sup>1</sup> Department of Microbiology, School of Medicine, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran, <sup>2</sup> Health Research Institute, Infectious and Tropical Diseases Research Center, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran, <sup>3</sup> Department of Immunology, School of Medicine, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran, <sup>4</sup> Cell and Molecular Research Center, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran, <sup>5</sup> Danesh Medical Laboratory, Ahvaz, Iran

## OPEN ACCESS

### Edited by:

Julio Alvarez,  
University of Minnesota, USA

### Reviewed by:

Paras Jain,  
Albert Einstein College  
of Medicine, USA

Li Xu,  
Cornell University, USA

### \*Correspondence:

Azar D. Khosravi,  
Department of Microbiology, School  
of Medicine, Ahvaz Jundishapur  
University of Medical Sciences; Health  
Research Institute, Infectious  
and Tropical Diseases Research  
Center, Ahvaz Jundishapur University  
of Medical Sciences, Golestan,  
61335 Ahvaz, Iran  
azarkhosravi69@gmail.com

### Specialty section:

This article was submitted to  
Infectious Diseases,  
a section of the journal  
*Frontiers in Microbiology*

Received: 11 March 2015

Accepted: 22 June 2015

Published: 03 July 2015

### Citation:

Meghdadi H,  
Khosravi AD, Ghadiri AA, Sina AH  
and Alami A (2015) Detection  
of *Mycobacterium tuberculosis* in  
extrapulmonary biopsy samples using  
PCR targeting IS6110, *rpoB*  
and nested-*rpoB* PCR Cloning.

Front. Microbiol. 6:675.  
doi: 10.3389/fmicb.2015.00675

Present study was aimed to examine the diagnostic utility of polymerase chain reaction (PCR) and nested PCR techniques for the detection of *Mycobacterium tuberculosis* (MTB) DNA in samples from patients with extra pulmonary tuberculosis (EPTB). In total 80 formalin-fixed, paraffin-embedded (FFPE) samples comprising 70 samples with definite diagnosis of EPTB and 10 samples from known non- EPTB on the basis of histopathology examination, were included in the study. PCR amplification targeting IS6110, *rpoB* gene and nested PCR targeting the *rpoB* gene were performed on the extracted DNAs from 80 FFPE samples. The strong positive samples were directly sequenced. For negative samples and those with weak band in nested-*rpoB* PCR, TA cloning was performed by cloning the products into the plasmid vector with subsequent sequencing. The 95% confidence intervals (CI) for the estimates of sensitivity and specificity were calculated for each method. Fourteen (20%), 34 (48.6%), and 60 (85.7%) of the 70 positive samples confirmed by histopathology, were positive by *rpoB*-PCR, IS6110- PCR, and nested-*rpoB* PCR, respectively. By performing TA cloning on samples that yielded weak ( $n = 8$ ) or negative results ( $n = 10$ ) in the PCR methods, we were able to improve their quality for later sequencing. All samples with weak band and 7 out of 10 negative samples, showed strong positive results after cloning. So nested-*rpoB* PCR cloning revealed positivity in 67 out of 70 confirmed samples (95.7%). The sensitivity of these combination methods was calculated as 95.7% in comparison with histopathology examination. The CI for sensitivity of the PCR methods were calculated as 11.39–31.27% for *rpoB*-PCR, 36.44–60.83% for IS6110- PCR, 75.29–92.93% for nested-*rpoB* PCR, and 87.98–99.11% for nested-*rpoB* PCR cloning. The 10 true EPTB negative samples by histopathology, were negative by all tested methods including cloning and were used to calculate the specificity of the applied methods. The CI for 100% specificity of each PCR method were calculated as 69.15–100%. Our results indicated that nested-*rpoB* PCR combined with TA cloning and sequencing is a preferred method for the detection of MTB DNA in EPTB samples with high sensitivity and specificity which confirm the histopathology results.

**Keywords:** *Mycobacterium tuberculosis*, extrapulmonary tuberculosis, polymerase chain reaction, biopsy, TA cloning

## Introduction

Despite the great advances in diagnosis and treatment of tuberculosis (TB), *Mycobacterium tuberculosis* (MTB) is still regarded as a major public health concern (Nahid et al., 2006; Tsara et al., 2009). World Health Organization (WHO), reported that nearly one-third of the global population infected with a member of the MTB complex (MTBC). Approximately 8.6 million new cases and 1.3 million deaths occur annually (World Health Organization, 2013). In most cases, TB affects the lungs, but the disease can potentially influence all organs of the body. In low incidence countries, approximately 15% of the cases are Extrapulmonary TB (EPTB; Hillemann et al., 2011). Conventional detection and identification of MTB, based on combination of microscopic examination and culture has been accepted as the gold standard, however, culture method despite being sensitive, is time-consuming and can takes several weeks (Moure et al., 2011), and direct microscopy and Ziehl-Neelsen staining do not provide sufficient sensitivities (about 20–80%) (Cattamanchi et al., 2010). EP specimens usually contain small number of bacteria, consequently culture and staining for acid-fast bacilli associated with the least sensitivity (Chakravorty et al., 2005). The histological examination of tissue samples is the traditional technique for diagnosis of EPTB which reveals the presence of granulomatous inflammation and caseous necrosis (Almadi et al., 2009). However, this method does not distinguish between EPTB and infections caused by other granulomatous diseases such as nontuberculous mycobacterial (NTM) diseases, sarcoidosis, leprosy and systemic lupus erythematosus (Mehta et al., 2012). The polymerase chain reaction (PCR) technique is a useful tool for rapid diagnosis of MTB in EP samples with high sensitivity and specificity (Supiyaphun et al., 2010). Among EP specimens, formalin-fixed, paraffin-embedded (FFPE) blocks can be used to study the tubercular infections. These samples cannot be cultured and may be unsuitable for PCR, because chemical alterations of the DNA (as degradation of the DNA) reduces the sensitivity of the PCR, but it can be improved by changing the amplification parameters such as using various targets with different sizes (Barcelos et al., 2008). The PCR can be used for a variety of targets such as IS6110, *devR*, *mtp40*, 16S rRNA, *rpoB* gene, etc (Balamurugan et al., 2006; Osores et al., 2006; Yang et al., 2011). In most MTB strains, there are 10 to 16 copies of IS6110 repetitive element (Piersimoni and Scarpa, 2003), though in some strains, particularly strains from Southeast Asia, there is no copy of this sequence or their number is negligible (Lok et al., 2002). So in such strains, the *rpoB* gene can be used as a target for identification of MTBC. *rpoB* encodes the B subunit of RNA polymerase, the *rpoB* nucleotide sequences comprising the Rif<sup>r</sup> region, which is associated with resistance to rifampin (Nisha et al., 2012). Previously Yun et al. (2005) used nested *rpoB*-PCR combined with TA cloning to improve the detection of MTB in joint biopsy sections successfully.

In the present study, The PCR technique targeting the *rpoB* gene and IS6110 fragment and also nested PCR targeting *rpoB* gene were applied to formalin fixed, paraffin-embedded tissue samples from patients suspected of having EPTB. Because of very low number of bacteria in such samples which makes PCR bands

usually weak or undetectable by agarose gel electrophoresis, TA cloning-sequencing was used for PCR confirmation.

## Materials and Methods

### Clinical Specimens Preparation

Eighty FFPE samples were collected from the departments of Pathology, university teaching hospitals, Ahvaz Jundishapur University of Medical Sciences, Iran. The initial proposal of the work was approved in the University high research and ethics combined committee. Seventy samples were belonged to patients with definite diagnosis of EPTB made on the basis of histopathology examination of FFPE tissues as gold standard method showing necrotizing granulomatous, to assess the sensitivity of PCR analyses, and ten samples were from known non-EPTB samples as true negatives to estimate the specificity of PCR technique. The samples type were: lymph node 29, pleural effusion 22, skin 8, breast 6, colon 4, thyroid 4, soft tissue 4, bladder, omentum, and kidney each one sample. The samples belonged to 36 male (45%) and 44 female (55%) patients. From each block, 2–3 sections of 8–10  $\mu\text{m}$  with an average surface area of 1  $\text{cm}^2$  were prepared for DNA extraction.

### DNA Extraction and PCR Amplification

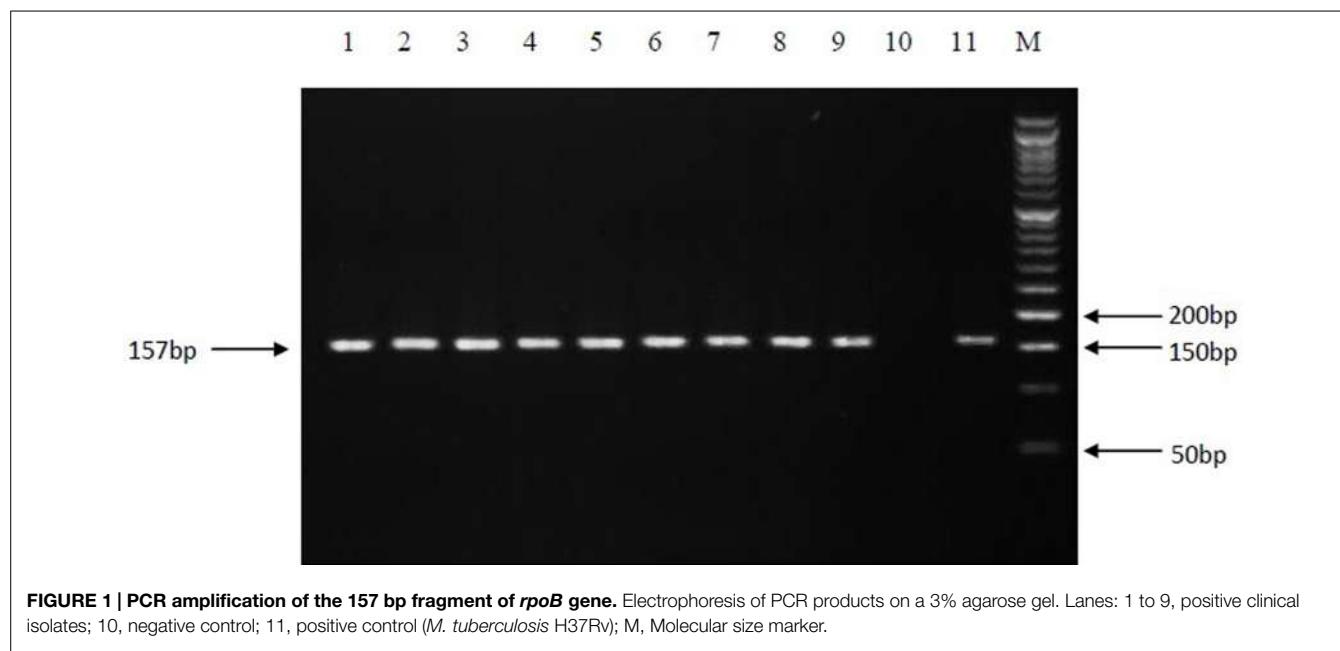
QIAamp DNA FFPE Tissue kit (Qiagen, Germany) was used for DNA extraction according to the manufacturer's instruction. Two different target genes were used in PCR technique.

PCR amplification for *rpoB* gene was performed by formerly reported set of primers [MF, 5'-CGACCACTTCGGCAACCG-3' and MR, 5'-TCGATCGGGCACATCCGG-3'] (Kim et al., 1999), which amplify a fragment of 342-base pair (bp) of the target gene. DNA amplification was performed in a thermocycler nexus gradient (Eppendorf, Germany), in a final volume of 25  $\mu\text{l}$  containing 10x PCR buffer, 1.5 mM Mg Cl<sub>2</sub>, 10 mM dNTPs, 0.5  $\mu\text{M}$  of each primer, 1.5 U super *Taq* polymerase and 5  $\mu\text{l}$  of template DNA. All the reagents were purchased from Roche Company, Germany. The amplification program was consisted of initial denaturation at 95° C for 5 min, followed by 35 cycles of denaturation at 94° C for 30 s, annealing at 60° C for 30 s, extension at 72° C for 45 s, and a final extension at 72° C for 5 min.

PCR amplification for IS6110 fragment was performed by a set of primer [MTB1, 5'-CCTGCGAGCGTAGGCCTCGG-3' and MTB2, 5'-CTCGTCCAGCGCCGCTTCGG-3'] (Eisenach et al., 1990). The primers are specific for MTBC and amplify a 123-bp fragment of repetitive sequence of IS6110 gene. The reaction mixture was the same as prepared for *rpoB* gene and the amplification condition consisted of initial denaturation at 95° C for 4 min, followed by 35 cycles of denaturation at 94° C for 1 min, annealing at 68° C for 2 min, extension at 72° C for 1 min and final extension at 72° C for 7 min. *M. tuberculosis* H37Rv standard strain as positive control and a non-necrotizing granulomatous inflammation FFPE tissue sample and also master mix without DNA sample as two negative controls were included in each PCR run.

### Nested *rpoB* PCR

A nested PCR based on the amplification of the *rpoB* gene of MTB was performed later for samples with inadequate DNA and



**FIGURE 1 | PCR amplification of the 157 bp fragment of *rpoB* gene.** Electrophoresis of PCR products on a 3% agarose gel. Lanes: 1 to 9, positive clinical isolates; 10, negative control; 11, positive control (*M. tuberculosis* H37Rv); M, Molecular size marker.

showed weak bands on the first PCR amplification. The first round PCR was done using the formerly reported outer primers of TB1, 5'-ACGTGGAGGCGATCACACCGCAGACGT-3', and TB2, 5'-TGCACGTCGCGGACCTCCAGCCCCGCA-3', which amplify a region of 205 bp of target gene (Kim et al., 2001). The second round (Nested) PCR used the amplicon of the first round as template and the reported inner primers of TR9, 5'-TCGCCCGATCAAGGAGT-3', and TR8, 5'-TGCACGTCGCGGACCTCCA-3', which amplify a 157 bp fragment (Cheng et al., 2007).

The total reaction volume in the first PCR round was 25  $\mu$ l and contained 0.2  $\mu$ M of each dNTP, 1.5 mM MgCl<sub>2</sub>, 0.5  $\mu$ M of each primers, 1.5 U of Super Taq<sup>TM</sup> DNA polymerase (Roche, Germany), 14  $\mu$ l sterile distilled water and 5  $\mu$ l of Template DNA. The PCR conditions consisted of initial denaturation at 95° C for 4 min, followed by 30 cycles of denaturation at 94° C for 30 s, annealing at 70° C for 30 s, extension at 72° C for 30 s and final extension at 78° C for 5 min. For nested PCR, 2  $\mu$ l of the first round amplicon was diluted (200-fold) and transferred to 48  $\mu$ l of a master mix solution (total reaction was 50  $\mu$ l) containing the same concentrations as described above, except that 36.6  $\mu$ l of sterile distilled water was used. The PCR conditions consisted of initial denaturation at 94° C for 5 min, followed by 28 cycles of denaturation at 94° C for 30 s, annealing at 55° C for 30 s, extension at 72° C for 30 s and final extension at 72° C for 5 min.

The products of each PCR assay were analyzed by gel electrophoresis on 3% agarose (w/vol.) containing 1  $\mu$ g/ml ethidium bromide (Fisher Scientific, USA). Results were recorded using a gel documentation apparatus (AlphaImager Systems, Protein Simple, USA). A 50 bp DNA ladder was used as size marker (Roche, Germany). The PCR products were sequenced for further analysis. SPSS software (SPSS Inc no. 13) was used for data analysis. Receiver-operating-characteristic curves (ROC) were calculated and expressed as areas under the curve, with an asymptotic 95% confidence interval (CI).

### Thymine-Adenine (TA) Cloning

The samples yielding a weak positive results in the nested-*rpoB* PCR, and all negative samples (including the 10 EPTB true negatives by histopathology examination), were subjected to cloning using a TA cloning kit (Invitrogen, USA), according to the supplier instructions. In brief, the ligation reaction was prepared and was transformed into competent *Escherichia coli* TOP10 cells. Recombinant PCR<sup>®</sup> 2.1: *rpoB* clones. The cloning process was confirmed by application a PCR using inner primers (TR9 and TR8) after a DNA extraction step based on simple boiling method. Simultaneously, a few transformed colonies were cultured in Luria-Bertani medium containing ampicillin. The plasmid DNA was then extracted by use of The GF-1 Plasmid DNA Extraction Kit (Vivantis, Malaysia).

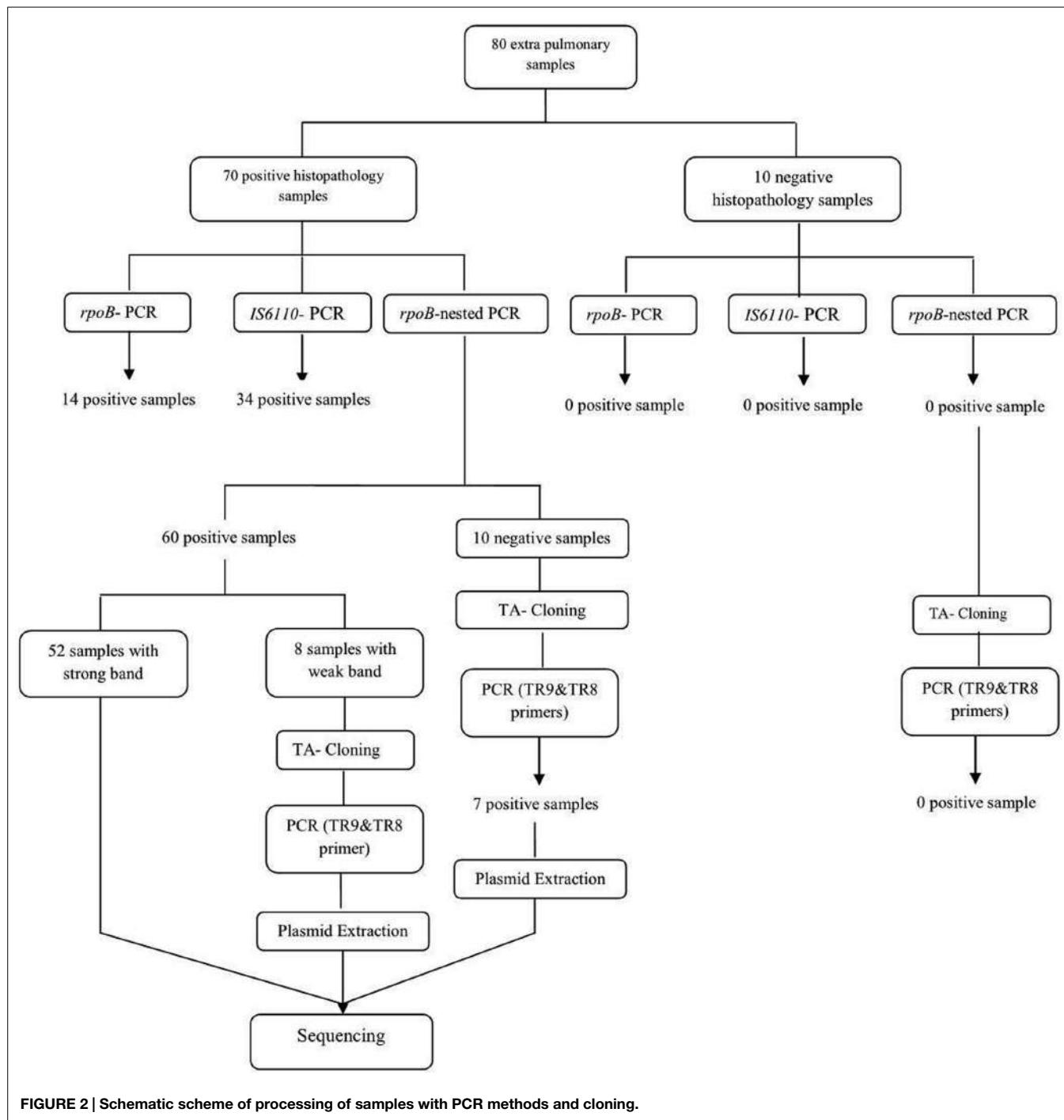
The 95% CI for the estimates of sensitivity and specificity were calculated for each method used according to Clopper-Pearson (Agresti, 2002).

### Nucleotide Sequencing and Sequence Analyses

The nucleotide sequence of the *rpoB* PCR product (157 bp) was determined by using inner primers (TR9 and TR8) and The nucleotide sequences of the purified plasmid by using universal M13 forward and reverse primers which were supplied in the TA cloning kit. Sequences were aligned by using blast to the reference DNA sequence.

## Results

From the total 70 samples confirmed as EPTB positive according to the histopathology examination (gold-standard), 14 samples (20%) were positive by PCR amplification of the *rpoB* gene, while by IS6110-based amplification, 34 samples (48.6%) and by nested-*rpoB* PCR, 60 samples (85.7%) were positive for the presence of MTBC (Figure 1).



**FIGURE 2 |** Schematic scheme of processing of samples with PCR methods and cloning.

By performing TA cloning on samples that yielded weak ( $n = 8$ ) or negative results ( $n = 10$ ) in the PCR methods, we were able to improve their quality for later sequencing. All samples with weak band and 7 out of 10 negative samples, showed strong positive results after cloning. So nested-*rpoB* PCR cloning revealed positivity in 67 out of 70 confirmed samples (95.7%) (Figure 2). The sensitivity of these combination methods was calculated as 95.7% in comparison with histopathology examination. The sequences of cloned plasmid were compared

with MTB H37Rv (Genbank), which 98–100% homology was revealed. In Table 1, the distribution of positive samples according to their origin by application of each PCR method is presented.

The details of samples underwent cloning are presented in Table 2. In certain clinical samples i.e. skin, colon, soft tissue, TA cloning showed twice positivity rate compared to the *rpoB*-nested PCR alone. Three samples (two lymph nodes and one pleural effusion) were negative by nested- *rpoB* PCR with cloning.

**TABLE 1 | Distribution of positive samples according to PCR amplification methods used.**

Clinical sample (No.)	<i>rpoB</i> -based PCR No. (%)	<i>IS6110</i> -based PCR No. (%)	<i>rpoB</i> -nested PCR No. (%)	<i>rpoB</i> -nested PCR with Cloning, No. (%)
Lymph node (25)	10 (40)	15 (60)	22 (88)	23 (92)
Pleural effusion (19)	4 (21)	10 (52.6)	17 (89.5)	18 (94.7)
Skin (8)	–	4 (50)	6 (75)	8 (100)
Breast (5)	–	2 (40)	4 (80)	5 (100)
Colon (4)	–	1 (25)	4 (100)	4 (100)
Soft tissue (4)	–	1 (25)	3 (75)	4 (100)
Thyroid (2)	–	–	2 (100)	2 (100)
Kidney (1)	–	1 (100)	1 (100)	1 (100)
Bladder (1)	–	–	1 (100)	1 (100)
Omentum (1)	–	–	–	1 (100)
Total (70)	14 (20)	34 (48.6)	60 (85.7)	67 (95.7)

**TABLE 2 | The EPTB histopathologic-confirmed samples with weak positive results or negative by PCR methods candidate for TA cloning.**

Clinical sample that were cloned (No.)	Weak positive positive	Undetectable by agarose gel electrophoresis
Lymph node (5)	2	3
Pleural effusion (2)	1	1
Skin (4)	2	2
Breast (2)	–	2
Colon (2)	2	–
Soft tissue (2)	1	1
Omentum (1)	–	1
Total (18)	8	10

**TABLE 3 | True positive and true negative samples detected by each method and sensitivity, specificity and 95% confidence intervals in relation to the histopathology results.**

Methods	True positive No. (%)	Sensitivity	95% confidence interval	True negative No. (%)	Specificity	95% confidence interval
<i>rpoB</i> - PCR	14/70 (20)	20%	(11.39–31.27%)	10/10 (100)	100%	(69.15–100%)
<i>IS6110</i> - PCR	34/70 (48.6)	48.6%	(36.44–60.83%)	10/10(100)	100%	(69.15–100%)
<i>rpoB</i> -nested PCR	60/70 (85.7)	85.7%	(75.29–92.93%)	10/10(100)	100%	(69.15–100%)
Nested PCR-cloning	67/70 (95.7)	95.7%	(87.98–99.11%)	10/10(100)	100%	(69.15–100%)

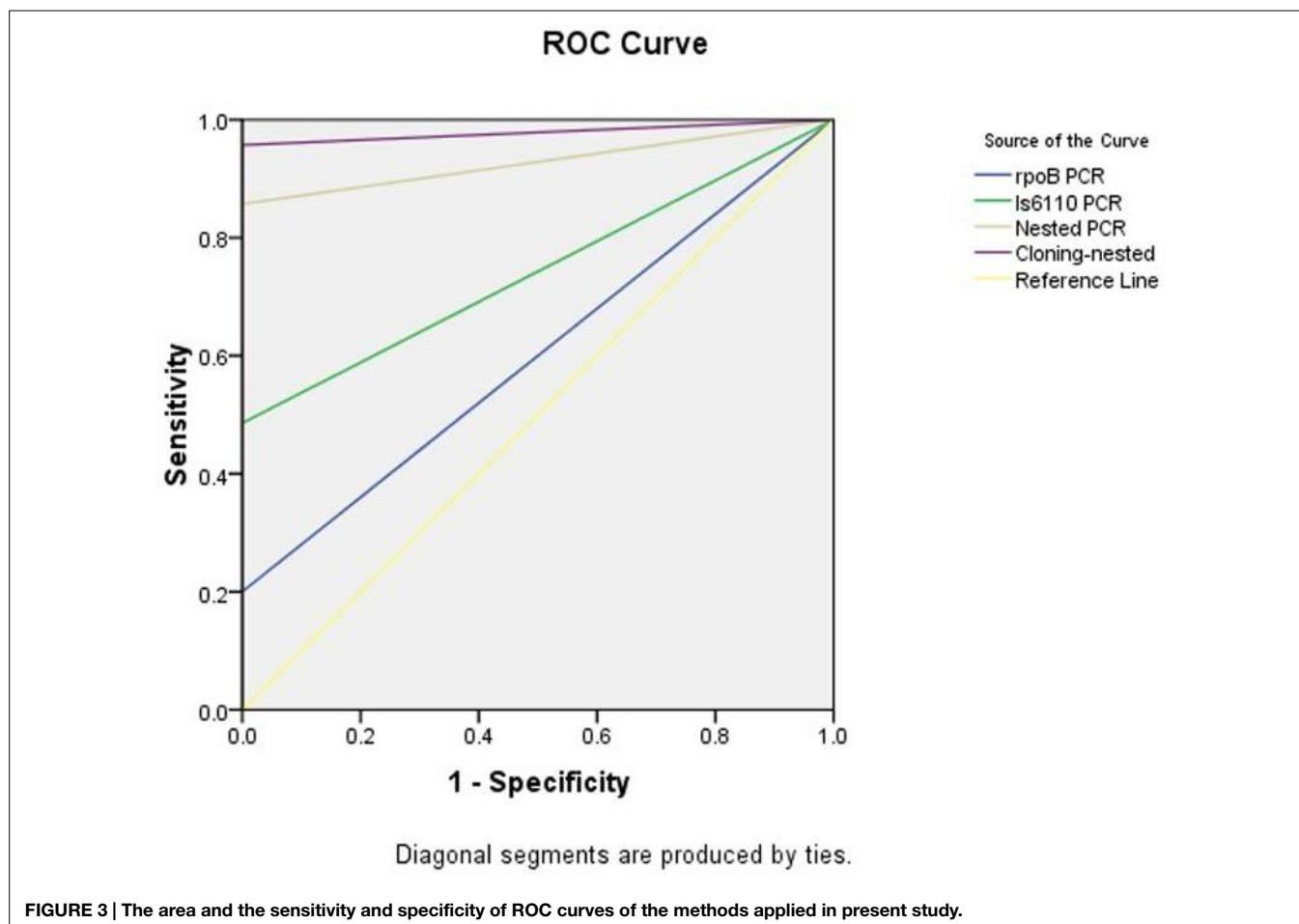
The 10 true negative samples for EPTB, were negative by all tested methods including cloning and were used to calculate the specificity of the applied methods. **Table 3** represents true positive and true negative samples detected by each PCR method and sensitivity, specificity and 95% CI in relation to the histopathology results.

## Discussion

Extra pulmonary tuberculosis is a significant health problem in both developed and developing countries. Literature reviews reveal that it accounts almost one of the three of total cases of TB in the world. In a study enrolled in US on 253229 cases of TB during 14 years, more than 18.7 percent were diagnosed for EPTB (Peto et al., 2009). Moreover, WHO statistical data shows that TB is the main cause of 66,000 death in year which is equivalent to eight death per hour in European countries. In Iran, EPTB accounts for more than 29 percent of all TB cases (World Health

Organization, 2013). Therefore, EPTB cases constitute a large number of TB burden which needs serious attentions of public health authorities in diagnosis and their identification in order to treat them properly.

The major problem is that the diagnosis of the disease in its different clinical presentations, remains a true challenge (Cailhol et al., 2005). The lack of a sensitive, specific and rapid method for the early diagnosis of EPTB poses a difficulty in initiating early therapy (Ong et al., 2004). Fresh clinical samples are desired for laboratory diagnosis of mycobacterial infections. However, in most cases of EPTB, fresh samples are not possible to obtain, so, in such cases, formalin-fixed paraffin-embedded tissue samples are used. These samples comprise technical limitations, including the lack of ability to culture. So the histopathology screening and PCR methods are the valuable diagnostic tools. For PCR assay, choosing one or more appropriate target (s), can be a great way to achieve a high sensitivity. PCR targeting *rpoB*, *IS6110* genes and nested PCR have been generally used to detect MTBC in



**FIGURE 3 |** The area and the sensitivity and specificity of ROC curves of the methods applied in present study.

**TABLE 4 |** The test results variable(s): *rpoB* PCR, IS6110 PCR, Nested PCR, Cloning-nested PCR.

Test result	Asymptotic 95% confidence interval					
	variable(s)	Area	Standard error <sup>a</sup>	Asymptotic Significance <sup>b</sup>	Lower bound	Upper bound
<i>rpoB</i> PCR	0.600	0.083		0.309	0.436	0.764
IS6110 PCR	0.743	0.062		0.013	0.622	0.864
Nested PCR	0.929	0.028		0.000	0.873	0.984
Cloning-nested PCR	0.979	0.015		0.000	0.949	1.008

The PCR methods has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

<sup>a</sup>Under the non-parametric assumption.

<sup>b</sup>Null hypothesis: true area = 0.5.

EP samples (Yun et al., 2005; Huang et al., 2009; Maurya et al., 2012).

In present study from 70 samples have been confirmed by histopathology, 20, 48.6, and 85.7% were positive by *rpoB*-PCR, IS6110-PCR and nested- *rpoB* PCR respectively, and 10 samples (25.7%) were either weak positive or negative with nested- *rpoB* PCR method. In study of Chakravorty et al. (2005), 99 EP specimens were screened for the detection of MTB by PCR targeting *devR* and IS6110 genes and their results showed positivity rates of 66.7 and 83% respectively, and 87.5% for the combination of both assays. In comparison, our findings revealed lower positivity rate by using IS6110 -PCR. In study of Hillemann

et al. (2006), 24 paraffin-embedded tissue specimens were studied for the presence of MTBC and they concluded that the real-time PCR assay exhibits a higher sensitivity (66.7%) for the detection of MTBC DNA compared to an alternative in-house IS6110 PCR (33.3%). As shown in their study, the rate achieved by IS6110- PCR was lower than our results. IS6110 is one of the main targets for the detection of MTB isolates, however, in this study, in comparison to nested *rpoB* PCR, the rate of positivity was lower. This positivity rate difference, may be due to chemical and physical alterations on DNA during fixation and tissue preparation (Barcelos et al., 2008), or is a 0-band strain (Lok et al., 2002).

The *rpoB*-PCR alone showed low sensitivity in this study, but when this target gene was used in nested PCR, we had a high sensitivity of 85.7%. This was in concordance with similar study of Huang et al. (2009) which a rate of 86.1% was reported. By using TA cloning and subsequent sequencing of the cloned plasmid, we gained an overall 95.7% sensitivity compared to histopathology examination leading to construct a ROC curve analysis (Figure 3), which the PCR methods results variable related to curve are presented in Table 4. The nucleic acid sequence analysis revealed a 98–100% homology in comparison with MTB H37Rv reference strain (Genbank). TA cloning showed much higher sensitivity compared to other applied techniques. By cloning, we were able to first, detect seven additional positive samples among the initial 10 negative samples in the PCR methods including nested *rpoB*-PCR and second, improve the quality of eight samples with weak positive results by nested *rpoB*-PCR, allowing its sequencing.

However, despite the advantages of TA cloning, it is a costly and not easy to perform technique and is not applicable for routine use in microbiology laboratories.

There were also some limitations in our study. Unfortunately, we had no access to the fresh or frozen samples and all the specimens were archived FFPE tissues. In such samples, the DNA integrity could vary and the degradation might be increased. In study of Ritis et al. (2000), the nested PCR and conventional methods for the diagnosis of EPTB samples were compared. According to their findings, the positivity rate by nested PCR was 90.9%, while conventional Acid fast microscopy and culture methods showed only 18.2% positivity. The samples in their study were fresh EP samples, which comprise less difficulties in processing and expecting to achieve more sensitivities compared to FFPE tissue samples. However, the results from combination PCR-TA cloning methods in our study, were in concordance to the results of histopathology examination and made us sure

about the true positive rate despite the nature of our tested samples. For a country surrounded with very high incidence and prevalence of TB such as Afghanistan, Pakistan, Azerbaijan, Tajikistan and Iraq, there is an urgent demand to settle down a guideline for appropriate molecular diagnostic methods with high sensitivity and specificity. Khuzestan province is in close contact with Iraqi population which needs more collaborative public health measurement by the authorities in the south-western region of Iran.

Our results indicated that from three PCR assays applied on samples, nested- *rpoB* PCR showed a high detection rate for MTBC DNA. Due to several problems are associated with the detection and identification of MTBC in FFPE tissues, nested-*rpoB* PCR combined with TA cloning-sequencing can be a useful method for the detection of MTBC DNA in samples from EPTB. In this study we gained a true positive rate of 95.7% by this combination compared to histopathology examination (70 samples) with 95% CI for our estimates of sensitivity and specificity. Further investigation with larger number of samples specially fresh biopsies and using more molecular targets are urge to perform in order to evaluate our results.

## Acknowledgments

This work is part of MSc thesis of Hossein Meghdadi. The initial proposal of the work was discussed and approved on session dated September 29th, 2013, in the University high research and ethics combined committee and has been double approved in Infectious and Tropical Diseases Research Center and was financially supported by Research affairs (Grant No. A-1209), Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran. The authors wish to Thank Dr. Amal Saki from statistics department for her assistance in Data interpretation.

## References

- Agresti, A. (2002). *Categorical Data Analysis*, 2nd Edn. New Jersey, NY: John Wiley.
- Almadi, M. A., Ghosh, S., and Aljebreen, A. M. (2009). Differentiating intestinal tuberculosis from Crohn's disease: a diagnostic challenge. *Am. J. Gastroenterol.* 104, 1003–1012. doi: 10.1038/ajg.2008.162
- Balamurugan, R., Venkataraman, S., John, K. R., and Ramakrishn, B. S. (2006). PCR amplification of the IS6110 insertion element of *Mycobacterium tuberculosis* in fecal samples from patients with intestinal tuberculosis. *J. Clin. Microbiol.* 44, 1884–1886. doi: 10.1128/JCM.44.5.1884-1886.2006
- Barcelos, D., Franco, M. F., and Leao, S. C. (2008). Effects of tissue handling and processing steps on PCR for detection of *Mycobacterium tuberculosis* in formalin-fixed paraffin-embedded samples. *Rev. Inst. Med. Trop. Sao Paulo* 50, 321–326. doi: 10.1590/S0036-46652008000600002
- Cailhol, J., Declut, B., and Che, D. (2005). Sociodemographic factors that contribute to the development of extrapulmonary tuberculosis were identified. *J. Clin. Epidemiol.* 58, 1066–1071. doi: 10.1016/j.jclinepi.2005.02.023
- Cattamanchi, A., Davis, J. L., Pai, M., Huang, L., Hopewell, P. C., and Steingart, K. R. (2010). Does bleach processing increase the accuracy of sputum smear microscopy for diagnosing pulmonary tuberculosis? *J. Clin. Microbiol.* 48, 2433–2439. doi: 10.1128/JCM.00208-10
- Chakravorty, S., Sen, M. K., and Tyagi, J. S. (2005). Diagnosis of extrapulmonary tuberculosis by smear, culture, and PCR using universal sample processing technology. *J. Clin. Microbiol.* 43, 4357–4362. doi: 10.1128/JCM.43.9.4357-4362.2005
- Cheng, X., Zhang, J., Yang, L., Xu, X., Liu, J., Yu, W., et al. (2007). A new Multi-PCR-SSCP assay for simultaneous detection of isoniazid and rifampin resistance in *Mycobacterium tuberculosis*. *J. Microbiol. Methods* 70, 301–305. doi: 10.1016/j.mimet.2007.05.002
- Eisenach, K. D., Cave, M. D., Bates, J. H., and Crawford, J. T. (1990). Polymerase chain reaction amplification of a repetitive DNA sequence specific for *Mycobacterium tuberculosis*. *J. Infect. Dis.* 161, 977–981. doi: 10.1093/infdis/161.5.977
- Hillemann, D., Galle, J., Vollme, E., and Richter, E. (2006). Real-time PCR assay for improved detection of *Mycobacterium tuberculosis* complex in paraffin-embedded tissues. *Int. J. Tuberc. Lung Dis.* 10, 340–342.
- Hillemann, D., Rusch-Gerdes, S., Boehme, C., and Richter, E. (2011). Rapid molecular detection of extrapulmonary tuberculosis by the automated GeneXpert MTB/RIF system. *J. Clin. Microbiol.* 49, 1202–1205. doi: 10.1128/JCM.02268-10
- Huang, H. J., Xiang, D. R., Sheng, J. F., Li, J., Pan, X. P., Yu, H. Y., et al. (2009). *rpoB* nested PCR and sequencing for the early diagnosis of tuberculous meningitis and rifampicin resistance. *Int. J. Tuberc. Lung Dis.* 13, 749–754.
- Kim, B. J., Lee, K. H., Park, B. N., Kim, S. J., Park, E. M., Park, Y. G., et al. (2001). Detection of rifampin-resistant *Mycobacterium tuberculosis* in sputa by nested PCR-linked single-strand conformation polymorphism and DNA sequencing. *J. Clin. Microbiol.* 39, 2610–2617. doi: 10.1128/JCM.39.7.2610-2617.2001
- Kim, B. J., Lee, S. H., Lyu, M. A., Kim, S. J., Bai, G. H., Chae, G. T., et al. (1999). Identification of mycobacterial species by comparative sequence analysis of the RNA polymerase gene (*rpoB*). *J. Clin. Microbiol.* 37, 1714–1720.
- Lok, K. H., Benjamin, W. H. Jr., Kimerling, M. E., Pruitt, V., Lathan, M., Razeq, J., et al. (2002). Molecular differentiation of *Mycobacterium tuberculosis* strains without IS6110 insertions. *Emerg. Infect. Dis.* 8, 1310–1313. doi: 10.3201/eid0811.020291

- Maurya, A. K., Kant, S., Nag, V. L., Kushwaha, R., and Dhole, T. N. (2012). Detection of 123 bp fragment of insertion element IS6110 *Mycobacterium tuberculosis* for diagnosis of extrapulmonary tuberculosis. *Indian J. Med. Microbiol.* 30, 182–186. doi: 10.4103/0255-0857.96688
- Mehta, P. K., Raj, A., Singh, N., and Khuller, G. K. (2012). Diagnosis of extrapulmonary tuberculosis by PCR. *FEMS Immunol. Med. Microbiol.* 66, 20–36. doi: 10.1111/j.1574-695X.2012.00987.x
- Moure, R., Munoz, L., Torres, M., Santin, M., Martin, R., and Alcaide, F. (2011). Rapid detection of *Mycobacterium tuberculosis* complex and rifampin resistance in smear-negative clinical samples by use of an integrated real-time PCR method. *J. Clin. Microbiol.* 49, 1137–1139. doi: 10.1128/JCM.01831-10
- Nahid, P., Pai, M., and Hopewell, P. C. (2006). Advances in the diagnosis and treatment of tuberculosis. *Proc. Am. Thorac. Soc.* 3, 103–110. doi: 10.1513/pats.200511-119JH
- Nisha, A., Govindarajan, S., Muthuraj, M., Manupriya, S., Usharani, B., Kamatchyamal, S., et al. (2012). Molecular characterization of *rpoB* gene encoding the RNA polymerase β subunit in rifampin-resistant *Mycobacterium tuberculosis* strains from south India. *Afr. J. Biotechnol.* 11, 3160–3168. doi: 10.5897/AJB10.449
- Ong, A., Creasman, J., Hopewell, P. C., Gonzalez, L. C., Wong, M., Jasmer, R. M., et al. (2004). A molecular epidemiological assessment of extrapulmonary tuberculosis in San Francisco. *Clin. Infect. Dis.* 38, 25–31. doi: 10.1086/380448
- Osores, F., Nolasco, O., Verdonck, K., Arevalo, J., Ferrufino, J. C., Agapito, J., et al. (2006). Clinical evaluation of a 16S ribosomal RNA polymerase chain reaction test for the diagnosis of lymph node tuberculosis. *Clin. Infect. Dis.* 43, 855–859. doi: 10.1086/507536
- Peto, H. M., Pratt, R. H., Harrington, T. A., LoBue, P. A., and Armstrong, L. R. (2009). Epidemiology of extrapulmonary tuberculosis in the United States, 1993–2006. *Clin. Infect. Dis.* 49, 1350–1357. doi: 10.1086/605559
- Piersimoni, C., and Scarparo, C. (2003). Relevance of commercial amplification methods for direct detection of *Mycobacterium tuberculosis* complex in clinical samples. *J. Clin. Microbiol.* 41, 5355–5365. doi: 10.1128/JCM.41.12.5355-5365.2003
- Ritis, K., Tzoanopoulos, D., Speletas, M., Papadopoulos, E., Arvanitidis, K., Kartali, S., et al. (2000). Amplification of IS6110 sequence for detection of *Mycobacterium tuberculosis* complex in HIV-negative patients with fever of unknown origin (FUO) and evidence of extrapulmonary disease. *J. Intern. Med.* 248, 415–424. doi: 10.1046/j.1365-2796.2000.00750.x
- Supiyaphun, P., Tumwasorn, S., Udomsantisuk, N., Keelawat, S., Songsrisanga, W., Prasurthsin, P., et al. (2010). Diagnostic tests for tuberculous lymphadenitis: fine needle aspirations using tissue culture in mycobacteria growth indicator tube and tissue PCR. *Asian Biomed.* 4, 787–792.
- Tsara, V., Serasli, E., and Christaki, P. (2009). Problems in diagnosis and treatment of tuberculosis infection. *Hippokratia* 13, 20–22.
- Yang, S., Zhong, M., Zhang, Y., and Wang, Y. (2011). Rapid detection of *rpoB* and *katG* genes from the sputum of multidrug-resistant *Mycobacterium tuberculosis* by polymerase chain reaction (PCR)-direct sequencing analysis. *Afr. J. Microbiol. Res.* 5, 4519–4523. doi: 10.5897/AJMR11.358
- Yun, Y. J., Lee, K. H., Haihua, L., Ryu, Y. J., Kim, B. J., Lee, Y. H., et al. (2005). Detection and identification of *Mycobacterium tuberculosis* in joint biopsy specimens by *rpoB* PCR cloning and sequencing. *J. Clin. Microbiol.* 43, 174–178. doi: 10.1128/JCM.43.1.174-178.2005
- World Health Organization. (2013). *Global Tuberculosis Report 2013*. Geneva: World Health Organization.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Meghdadi, Khosravi, Ghadiri, Sina and Alami. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Survey and Visual Detection of *Zaire ebolavirus* in Clinical Samples Targeting the Nucleoprotein Gene in Sierra Leone

Huan Li<sup>1</sup>, Xuesong Wang<sup>†</sup>, Wei Liu, Xiao Wei, Weishi Lin, Erna Li, Puyuan Li, Derong Dong, Lifei Cui, Xuan Hu, Boxing Li, Yanyan Ma, Xiangna Zhao\*, Chao Liu\* and Jing Yuan\*

Institute of Disease Control and Prevention, Academy of Military Medical Sciences, Beijing, China

## OPEN ACCESS

### Edited by:

Andres M. Perez,  
University of Minnesota, USA

### Reviewed by:

Ihem Messaoudi,  
Oregon Health and Science  
University, USA  
Hideki Ebihara,  
National Institutes of Health, USA

### \*Correspondence:

Jing Yuan  
yuanjing6216@163.com;  
Chao Liu  
liuchao9588@sina.com;  
Xiangna Zhao  
xnzhao@163.com

<sup>†</sup>These authors are co-first authors.

### Specialty section:

This article was submitted to  
Infectious Diseases,  
a section of the journal  
*Frontiers in Microbiology*

Received: 02 August 2015

Accepted: 12 November 2015

Published: 01 December 2015

### Citation:

Li H, Wang X, Liu W, Wei X, Lin W, Li E, Li P, Dong D, Cui L, Hu X, Li B, Ma Y, Zhao X, Liu C and Yuan J (2015) Survey and Visual Detection of *Zaire ebolavirus* in Clinical Samples Targeting the Nucleoprotein Gene in Sierra Leone. *Front. Microbiol.* 6:1332.  
doi: 10.3389/fmicb.2015.01332

Ebola virus (EBOV) can lead to severe hemorrhagic fever with a high risk of death in humans and other primates. To guide treatment and prevent spread of the viral infection, a rapid and sensitive detection method is required for clinical samples. Here, we described and evaluated a reverse transcription loop-mediated isothermal amplification (RT-LAMP) method to detect *Zaire ebolavirus* using the nucleoprotein gene (NP) as a target sequence. Two different techniques were used, a calcein/Mn<sup>2+</sup> complex chromogenic method and real-time turbidity monitoring. The RT-LAMP assay detected the NP target sequence with a limit of 4.56 copies/μL within 45 min under 61°C, a similar even or increase in sensitivity than that of real-time reverse transcription-polymerase chain reaction (RT-PCR). Additionally, all pseudoviral particles or non- *Zaire* EBOV genomes were negative for LAMP detection, indicating that the assay was highly specific for EBOV. To appraise the availability of the RT-LAMP method for use in clinical diagnosis of EBOV, of 417 blood or swab samples collected from patients with clinically suspected infections in Sierra Leone, 307 were identified for RT-LAMP-based surveillance of EBOV. Therefore, the highly specific and sensitive RT-LAMP method allows the rapid detection of EBOV, and is a suitable tool for clinical screening, diagnosis, and primary quarantine purposes.

**Keywords:** *Zaire* EBOV, RT-LAMP, sensitivity, specificity, rapid detection, prevalence

## INTRODUCTION

The 2014 Ebola virus (EBOV) outbreak was the largest to date in West African countries (Frieden et al., 2014; Hampton, 2014). It spread through direct contact with infected individuals, or by touching the blood, organs, bodily secretions and fluids, or the contaminated clothes of such people (MacNeil and Rollin, 2012). Because the initial symptoms of ebolavirus infection can be confused with those of other febrile illnesses such as endemic malaria (Chertow et al., 2014), and because the infection cannot be detected rapidly in patients living in remote areas (World Health Organization [WHO], 2014)<sup>1</sup>, the numbers of infected people and deaths were the highest yet recorded. To control infection and to prevent further transmission during outbreaks of filoviruses such as EBOV, rapid detection is therefore essential (Grolla et al., 2005).

<sup>1</sup><http://www.who.int/mediacentre/news/ebola/18-november-2014-diagnostics/en/>

Current methods for the detection and diagnosis of EBOV infection include virus isolation, electron microscopy, immunohistochemistry (Zaki et al., 1999), enzyme-linked immunosorbent assay testing (Niikura et al., 2001), reverse transcription-polymerase chain reaction (RT-PCR), serologic testing for IgM/IgG virus-specific antibodies (Ksiazek et al., 1999a; Saito et al., 2001), and point-of-care biosensors (Baca et al., 2015). In general, when the EBOV viral load in the blood gets to a higher case fatality rate, the detection of antigens as a suitable method is used for laboratory diagnosis (Fisher-Hoch et al., 1992; Ksiazek et al., 1999b). Thus, the World Health Organization recommends real-time RT-PCR as the first choice for EBOV diagnosis. However, inhibitors present in crude biological samples can inactivate the *Taq* DNA polymerase used in PCR-based techniques (de Franchis et al., 1988). Moreover, these methods are relatively complex and require specialized instruments. Thus, to complement PCR-based methods, another rapid, simple, and effective assay is needed.

Loop-mediated isothermal amplification (LAMP) is a one-step nucleic acid detection method developed by Notomi et al. (2000) which relies on autocycling strand displacement DNA synthesis. This novel method is highly specific and sensitive, takes advantage of four or six specific primers to recognize six or eight different sequences of the target gene, and is performed under isothermal conditions in less than 1 h using *Bst* DNA polymerase. Furthermore, LAMP is less influenced by inhibitors present in complex samples than standard PCR, which is highly beneficial for clinical specimens such as blood components, sputum, feces, or body fluids (Kaneko et al., 2007). LAMP assays have been widely applied to genetic diagnoses, the detection of epidemic bacteria (Hara-Kudo et al., 2005; Song et al., 2005) and viruses (Okafuji et al., 2005), fetal sex identification (Hirayama et al., 2013), and parasite recognition (Chen et al., 2011; Kong et al., 2012).

Kurosaki et al. (2007) developed a simple reverse transcription loop-mediated isothermal amplification (RT-LAMP) assay for the detection of *Zaire ebolavirus*, targeting the trailer region of the viral genome. However, this method has yet to be tested in clinical samples. The EBOV genome is approximately 19 kb, and encodes the following seven genes, which are flanked by untranslated regions: nucleoprotein (*NP*), viral structural protein (*VSP*)35, *VSP*40, glycoprotein, *VP*30, *VP*24, and RNA-dependent RNA polymerase (Ali and Islam, 2015).

*NP* is highly conserved among all EBOV species currently known, and plays an important role in intracellular events such as replication and transcription of the viral genome, and nucleocapsid formation (Ali and Islam, 2015). It is therefore recommended by the World Health Organization for use as a target gene for the RT-PCR assay.

In the present work, we developed a point-of-care RT-LAMP assay targeted to *NP*. Five sets of primers for the detection of EBOV were designed and used in optimization of the RT-LAMP assay. We also evaluated the specificity and sensitivity of the LAMP method. Finally, 417 blood samples collected from patients with clinically suspected infections were analyzed by RT-LAMP and RT-PCR in clinical diagnosis.

## MATERIALS AND METHODS

### Viruses, RNA Extraction, and Preparation of Templates

Twenty-six genomes of respiratory pathogens including artificial RNAs of Sudan EBOV (Subtype Sudan, strain Gulu), Zaire EBOV and MARV, SARS coronavirus, influenza A H7N9, H1N1, H2N3, human parainfluenza viruses (PIV) type 1/2/3 and 4, adenoviruses (ADV; serotype 3, serotype 5, and serotype 55), respiratory syncytial virus infection RSVA/RSVB, MERS RNA, human metapneumovirus HMPV, human coronavirus HCoV-229E/ HCoV-OC43/HCoV-NL63, and HCoV-HKU1, bocavirus BoV, as well as three respiratory bacterial pathogens such as *Legionella pneumophila* 9135, *Mycobacterium tuberculosis* 005, and *Haemophilus influenzae* ATCC 49247 were used in this study. Total viral RNAs were extracted from 200  $\mu$ l of each culture using a QIAamp viral RNA mini kit (Qiagen, Hilden, Germany). All infectious materials were handled in biosafety level 3 facilities.

### Preparation of Artificial EBOV RNA

Preparation of artificial EBOV RNA was performed as described previously (Watanabe et al., 2004) with modifications. Briefly, 663 kb *NP* fragments were synthesized (Sangon Biotech Co., Ltd., Shanghai, China) and cloned into vector pGEM-3Zf(+) with inverse orientation of the T7 promoter sequence (Promega). *In vitro* transcription of artificial EBOV RNA from *NP* subclones was carried out using 50 U of T7 RNA polymerase (Promega) in a 50- $\mu$ l reaction volume according to the manufacturer's instructions. The RNA concentration was determined by measuring the optical density at 260 nm (OD260), and the RNA purity was determined by calculating the OD260/OD280 absorption ratio (ratios were ensured to be  $>1.8$ ). RNA was then dissolved in 20  $\mu$ L DEPC-treated water, and stored at  $-70^{\circ}\text{C}$  before use.

### Primer Design

Based on the *NP* sequences of strain Mayinga deposited in GenBank (accession no. AF086833), we selected potential target regions and further analyzed the sequence using Primer Explorer V4 software<sup>2</sup> by aligning it with other species of EBOV. We designed specific primer sets for the detection of EBOV in RT-LAMP, with each set including an outer forward primer (F3), an outer backward primer (B3), a forward inner primer (FIP), and a backward inner primer (BIP) linked by a four thymidine spacer (TTTT), which can recognize both sense and anti-sense strands. To accelerate the RT-LAMP reaction, an additional loop primer (LB) was designed. All primers were synthesized commercially (Sangon Biotech Co., Ltd.).

### RT-LAMP Assays

Reverse transcription loop-mediated isothermal amplification reactions were performed using a Loopamp RNA amplification kit (Eiken Chemical Co., Ltd., Tokyo, Japan) in a volume

<sup>2</sup><https://primerexplorer.jp/e/>

of 25  $\mu\text{L}$  according to the manufacturer's protocol. Each reaction included 80 pmol of FIP and BIP, 40 pmol of LB, 10 pmol of F3 and B3, and 2  $\mu\text{L}$  template RNA. The reaction was carried out at 61°C for 60–80 min in dry bath incubators.

Reverse transcription loop-mediated isothermal amplification amplified products were detected by turbidity monitoring as well as visual observation. To assess turbidity, the amount of white magnesium phosphate precipitate produced during the LAMP reaction process was monitored using a Loopamp Real-time Turbidimeter (LA-230; Eiken Chemical Co., Ltd., Tochigi, Japan) recording the reaction curves at 650 nm every 6 s with magnesium ion ( $\text{Mg}^{2+}$ ) in the reaction buffer (Mori et al., 2001). For visual inspection, tubes containing 1  $\mu\text{l}$  of fluorescent calcine were observed by the naked eye and photographed under natural light or UV light at 365 nm. The color changed from orange to green for a positive reaction, while the negative control remained orange.

### Real-time RT-PCR Assays

To illustrate RT-LAMP detection sensitivity, we targeted a region of *NP* using the Liferiver™ EBOV Real Time RT-PCR Kit (Shanghai ZJ BioTech Co., Ltd.) recommended by the World Health Organization. Thermocycler conditions followed the manufacturer's instructions. During the amplification process, the fluorescence intensity of the reporter dye (FAM) and a quencher dye (TAMRA) was recorded to calculate the normalized reporter signal, which is linked to the amount of product amplified. The standard curve was drawn using the *in vitro* transcribed RNA standard to calculate the number of RNA copies in viral RNA extracts. The threshold cycle ( $C_t$  value) refers to the number of amplification cycles for the fluorescence to reach the threshold.

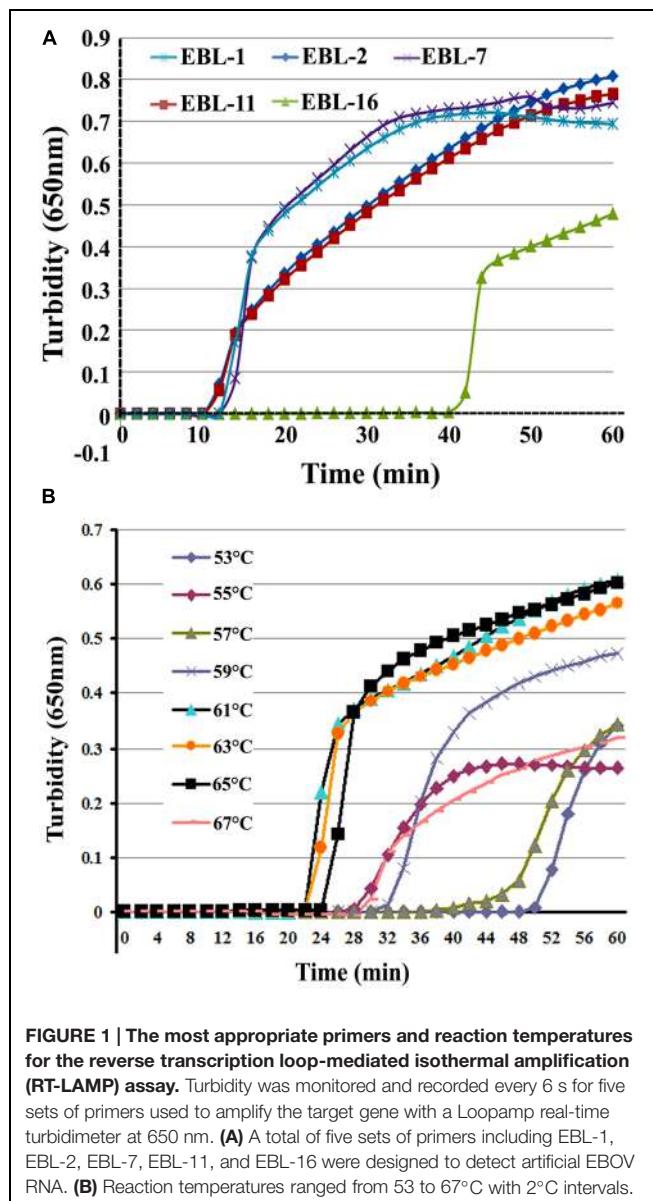
### Clinical Specimens

A total of 417 clinical specimens from whole blood or swabs were collected from patients thought to be infected by EBOV during the outbreak in Sierra Leone, from 2014 to 2015. RT-LAMP assays and real-time RT-PCR were performed simultaneously by the China Mobile Laboratory Testing Team in Freetown, Sierra Leone. Information about the clinical samples is listed in Supplemental Tables S1 and S2. This study was carried out in accordance with the recommendations of the Institute of Disease Control and Prevention, China with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

## RESULTS

### Optimizing the RT-LAMP Assay

A total of five sets of primers were initially designed to detect artificial EBOV RNA using the Real-time Turbidimeter. As shown in **Figure 1A**, the EBL-2 primer set amplified *NP* in the shortest time (~10 min), so this was chosen as the optimal primer set for RT-LAMP detection of EBOV (**Table 1**).



To further optimize the amplification, we compared reaction temperatures ranging from 53 to 67°C with 2°C intervals. The most suitable reaction temperature range was shown to be

**TABLE 1 |** Sequence of primers used for specific amplification of *NP*.

Primer	Type	Sequence(5'-3')
EBL-2F3	Forward outer	GATGGAAGCTACGGCGAA
EBL-2B3	Backward outer	GTGAGGGCCTGGACATT
EBL-2FIP	Forward inner	AGTGTCCCTCGTCGCTCGTAGAG TTACTCGGAAAACGGC
EBL-2BIP	Backward inner	ATCGACCAAGGGTGGACAACAGTTGT TCTGTCTGCCCTCTA
EBL-2LB	Loop backward	AAGAACAGTCAAAGGCCAGC
EBL-2LF	Loop forward	CCAAGTCATCTGGTGATTCT

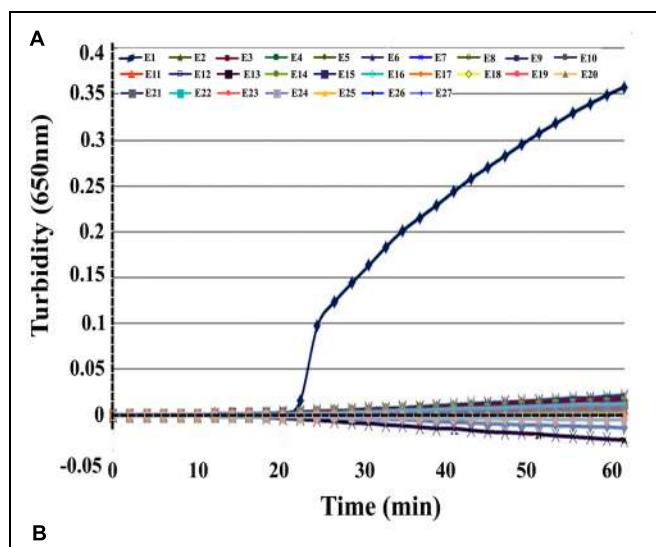
59–65°C (Figure 1B), and 61°C was ultimately chosen as the optimal reaction temperature.

### Specificity of NP Detection by RT-LAMP

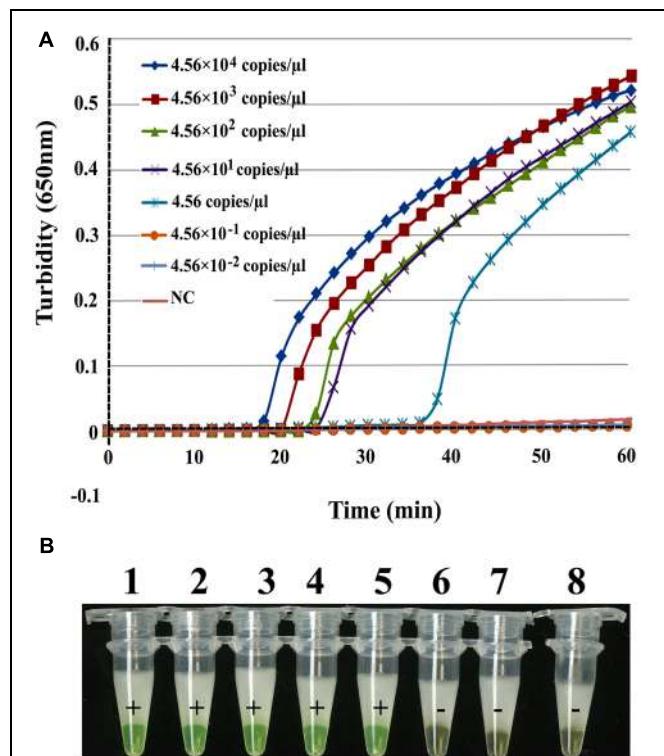
To test the LAMP specificity for NP, we tested 26 non-Zaire EBOV viruses in addition to EBOV itself and *in vitro* transcribed artificial EBOV RNA as the positive control. Figure 2 shows that EBOV RNA was identified positively by RT-LAMP with the EBL-2 primer set using turbidity monitoring and visual observation. All non-Zaire EBOV strains tested negative, including the blank control, indicating that the RT-LAMP method was specific for EBOV.

### Sensitivity of NP Detection by RT-LAMP

To determine the sensitivity of the RT-LAMP assay for EBOV, a series of dilutions were prepared of artificial EBOV RNA



**FIGURE 2 | Specificity of EBOV NP detection by RT-LAMP.** **(A)** Turbidity was monitored and recorded every 6 s by a Loopamp real-time turbidimeter at 650 nm. **(B)** Visual detection using a calcein fluorescent detection reagent. Lane 1, positive control (artificial EBOV RNA); lane 2, negative control (double-distilled water); lane 3, Sudan EBOV (artificial Sudan EBOV RNA); lane 4, MARV (artificial MARV RNA); lane 5, SARS coronavirus; lane 6, H7N9; lane 7, H1N1; lane 8, H2N3; lanes 9–12, human parainfluenza viruses (PIV) 1, 2, 3, and 4; lanes 13–15, adenoviruses (ADV; serotypes 3, 5, and 55); lanes 16 and 17, respiratory syncytial virus infection, RSVA, RSVB; lane 18, MERS RNA; lane 19, human metapneumovirus, HMPV; lane 20, bocavirus, BoV; lanes 21–24, human coronavirus, HCoV-229E, HCoV-OC43, HCoV-NL63, and HCoV-HKU1; lane 25, *Legionella pneumophila* 9135; lane 26, *Mycobacterium tuberculosis* 005; and lane 27, *Haemophilus influenzae* ATCC 49247.



**FIGURE 3 | Comparison of RT-LAMP sensitivities in detecting EBOV NP.** Artificial EBOV RNA was serially diluted 10-fold from 4.56 × 10<sup>4</sup> copies/μL to 4.56 × 10<sup>-2</sup> copies/μL. **(A)** Turbidity was monitored with a Loopamp Realtime Turbidimeter at 650 nm every 6 s. **(B)** The reaction was detected visually using a calcein fluorescent detection reagent. Artificial EBOV RNA concentrations were: tube 1, 4.56 × 10<sup>4</sup> copies/μL; tube 2, 4.56 × 10<sup>3</sup> copies/μL; tube 3, 4.56 × 10<sup>2</sup> copies/μL; tube 4, 4.56 × 10<sup>1</sup> copies/μL; tube 5, 4.56 copies/μL; tube 6, 4.56 × 10<sup>-1</sup> copies/μL; tube 7, 4.56 × 10<sup>-2</sup> copies/μL; tube 8, ddH<sub>2</sub>O.

ranging from 4.56 × 10<sup>4</sup> to 4.56 × 10<sup>-2</sup> copies/μL. As shown in Figure 3A, the times of positivity detection ranged from 18 min for 4.56 × 10<sup>4</sup> copies/μL to 36 min for 4.56 copies/μL of virus RNA by real-time monitoring. Thus, the RT-LAMP detection limit for NP is 4.56 copies/μL of artificial RNA in a 61°C reaction lasting for 60 min. For the visual inspection, all positive reactions changed to green while negative ones remained orange under natural or 365 nm UV light (Figure 3B). These data indicate that the sensitivity of the two detection methods was the same. The detection limit of real-time RT-PCR for NP was 4.56 copies/μL, but this was achieved with a higher C<sub>t</sub> value (C<sub>t</sub> = 41). Thus, we concluded that the sensitivity of the RT-LAMP assay for EBOV was similar or higher than real-time RT-PCR.

### Clinical Sample Detection

The 417 clinical blood or swab samples were simultaneously analyzed by RT-LAMP and real-time RT-PCR. Of these, 307 patients were confirmed to be infected with EBOV, while 106 tested negative (Figure 4 and Table 2). A higher C<sub>t</sub> value (C<sub>t</sub> > 36) was recorded for the remaining four samples by



**FIGURE 4 | Visual detection of RT-LAMP findings of 413 patients with clinically suspected infections in Sierra Leone.** Sample information is listed in Supplemental Tables S1 and S2. A total of 307 positive (A) and 106 negative (B) samples were tested for EBOV infection by real-time RT-PCR and RT-LAMP. NC, negative control (distilled water); PC, positive control (artificial EBOV RNA).

**TABLE 2 | Reverse transcription loop-mediated isothermal amplification (RT-LAMP) and real-time reverse transcription-polymerase chain reaction (RT-PCR) findings of clinical blood or swab samples.**

Results	RT-LAMP	Real-time RT-PCR
Positive samples	307 (Time: 10–30min)	307 ( $C_t \leq 30$ )
Negative samples	106 (no change in color)	106 (no peak)

RT-PCR, and green fluorescence was observed after ~45 min by RT-LAMP, indicating a low level of EBOV RNA. We suggested that these patients should be monitored for 2 weeks in hospital.

## DISCUSSION

Ebola virus has extremely high morbidity and mortality levels in humans, it reemerged and caused an outbreak in Western Africa where 28,476 suspected, probable, and confirmed cases, and 11,298 deaths were reported in Sierra Leone up until Oct. 21 2015 according to Ebola Situation Report from WHO<sup>3</sup>). Although several chemical agents, vaccines, and antibodies inhibit the spread of EBOV in humans

and animals, effective therapies for clinical treatment are scarce.

To combat the increasing incidence of EBOV infections, we developed an RT-LAMP assay specific for EBOV diagnosis using primers spanning the 663 bp NP sequence of the viral genome. We found that the limit of detection for this technique was 4.56 copies/ $\mu$ L, which compares with 13.4 copies/ $\mu$ L for the Real-Time RT-PCR assay using the Liferiver™-EBOV Kit according to the manufacturer's instructions. In comparison, the limit of detection was 10 RNA molecule standards targeting the nucleoprotein gene by RT-PCR reported by Weidmann et al. (2004). In the present work, the RT-LAMP assay showed an equivalent or superior sensitivity to RT-PCR assays, indicating that it is sufficiently sensitive to detect low copy numbers of RNA. Furthermore, the RT-LAMP assay showed a high level of specificity, with no cross-reactivity with other species of EBOV or other viruses. It was also comparable with real-time RT-PCR at confirming cases of EBOV infection in clinical samples.

The RT-LAMP assay can detect the presence of virus in a single step in which reverse transcription and DNA amplification proceed in a single tube at a constant temperature of 61°C. Compared with RT-PCR, the sensitivity of the RT-LAMP assay is far greater in the presence of inhibitors. Moreover, RT-LAMP primers specifically recognize target sequences using

<sup>3</sup><http://apps.who.int/ebola/current-situation/ebola-situation-report-21-october-2015>

five independent target sequence regions, compared with RT-PCR primers that recognize only two independent regions. Therefore, the RT-LAMP assay is more suitable for the rapid detection of *NP* in clinical samples.

## CONCLUSION

We established a rapid and effective visual RT-LAMP assay targeting EBOV *NP*, which we showed to be extremely specific and sensitive in the molecular diagnosis of EBOV infections. It is a reliable tool for the identification of EBOV, so could be used as an alternative method of diagnostic testing at clinical laboratories without the need for special apparatus. Moreover, it can provide accurate results within 1 h, so may be of use in the clinical diagnosis of EBOV in developing countries.

## AUTHOR CONTRIBUTIONS

JY and XZ conceived and designed the experiments. XW and WL performed clinical detection in Sierra Leone. HL, PL, and DD performed and developed the RT-LAMP. XY, EL, PL, DD, DZ, LC, and XH performed the experiments. XZ wrote the manuscript. JY and XZ edited the manuscript.

## REFERENCES

- Ali, M. T., and Islam, M. O. (2015). A highly conserved GEQYQQQLR epitope has been identified in the nucleoprotein of Ebola virus by using an in silico approach. *Adv. Bioinform.* 2015, 278197–278203. doi: 10.1155/2015/278197
- Baca, J. T., Severns, V., Lovato, D., Branch, D. W., and Larson, R. S. (2015). Rapid detection of Ebola virus with a reagent-free, point-of-care biosensor. *Sensors (Basel)* 15, 8605–8614. doi: 10.3390/s150408605
- Chen, R., Tong, Q., Zhang, Y., Lou, D., Kong, Q., Lv, S., et al. (2011). Loop-mediated isothermal amplification: rapid detection of *Angiostrongylus cantonensis* infection in *Pomacea canaliculata*. *Parasit. Vectors* 4, 204. doi: 10.1186/1756-3305-4-204
- Chertow, D. S., Kleine, C., Edwards, J. K., Scaini, R., Giuliani, R., and Sprecher, A. (2014). Ebola virus disease in West Africa-clinical manifestations and management. *N. Engl. J. Med.* 371, 2054–2057. doi: 10.1056/NEJMmp1413084
- de Franchis, R., Cross, N. C., Foulkes, N. S., and Cox, T. M. (1988). A potent inhibitor of Taq polymerase copurifies with human genomic DNA. *Nucleic Acids Res.* 16, 10355. doi: 10.1093/nar/16.21.10355
- Fisher-Hoch, S. P., Perez-Oronoz, P. I., Jackson, E. L., Hermann, L. M., and Brown, B. G. (1992). Filovirus clearance in non-human primates. *Lancet* 340, 451–453. doi: 10.1016/0140-6736(92)91770-9
- Frieden, T. R., Damon, I., Bell, B. P., Kenyon, T., and Nichol, S. (2014). Ebola 2014—New challenges, new global response and responsibility. *N. Engl. J. Med.* 371, 1177–1180. doi: 10.1056/NEJMmp1409903
- Grolla, A., Lucht, A., Dick, D., Strong, J. E., and Feldmann, H. (2005). Laboratory diagnosis of Ebola and Marburg hemorrhagic fever. *Bull. Soc. Pathol. Exot.* 98, 205–209.
- Hampton, T. (2014). Largest-ever outbreak of Ebola Virus disease thrusts experimental therapies, vaccines into spotlight. *JAMA* 312, 987–989. doi: 10.1001/jama.2014.11170
- Hara-Kudo, Y., Yoshino, M., Kojima, T., and Ikeda, M. (2005). Loop-mediated isothermal amplification for the rapid detection of *Salmonella*. *FEMS Microbiol. Lett.* 253, 155–161. doi: 10.1016/j.femsle.2005.09.032
- Hirayama, H., Kageyama, S., Moriyasu, S., Sawai, K., and Minamihashi, A. (2013). Embryo sexing and sex chromosomal chimerism analysis by loop-mediated isothermal amplification in cattle and water buffaloes. *J. Reprod. Dev.* 59, 321–326. doi: 10.1262/jrd.2013-028
- Kaneko, H., Kawana, T., Fukushima, E., and Suzutani, T. (2007). Tolerance of loop-mediated isothermal amplification to a culture medium and biological substances. *J. Biochem. Biophys. Methods* 70, 499–501. doi: 10.1016/j.jbbm.2006.08.008
- Kong, Q. M., Lu, S. H., Tong, Q. B., Lou, D., Chen, R., Zheng, B., et al. (2012). Loop-mediated isothermal amplification (LAMP): early detection of *Toxoplasma gondii* infection in mice. *Parasit. Vectors* 5, 2. doi: 10.1186/1756-3305-5-2
- Ksiazek, T. G., Rollin, P. E., Williams, A. J., Bressler, D. S., Martin, M. L., Swanepoel, R., et al. (1999a). Clinical virology of Ebola hemorrhagic fever (EHF): virus, virus antigen, and IgG and IgM antibody findings among EHF patients in Kikwit, Democratic Republic of the Congo, 1995. *J. Infect. Dis.* 179 (Suppl. 1), S177–S187. doi: 10.1086/514321
- Ksiazek, T. G., West, C. P., Rollin, P. E., Jahrling, P. E., and Peters, C. J. (1999b). ELISA for the detection of antibodies to Ebola viruses. *J. Infect. Dis.* 179 (Suppl. 1), S192–S198. doi: 10.1086/514313
- Kurosaki, Y., Takada, A., Ebihara, H., Grolla, A., Kamo, N., Feldmann, H., et al. (2007). Rapid and simple detection of Ebola virus by reverse transcription-loop-mediated isothermal amplification. *J. Virol. Methods* 141, 78–83. doi: 10.1016/j.jviromet.2006.11.031
- MacNeil, A., and Rollin, P. E. (2012). Ebola and Marburg hemorrhagic fevers: neglected tropical diseases? *PLoS Negl. Trop. Dis.* 6:e1546. doi: 10.1371/journal.pntd.0001546
- Mori, Y., Nagamine, K., Tomita, N., and Notomi, T. (2001). Detection of loop-mediated isothermal amplification reaction by turbidity derived from magnesium pyrophosphate formation. *Biochem. Biophys. Res. Commun.* 289, 150–154. doi: 10.1006/bbrc.2001.5921
- Niikura, M., Ikegami, T., Saito, M., Kurane, I., Miranda, M. E., and Morikawa, S. (2001). Detection of Ebola viral antigen by enzyme-linked immunosorbent assay using a novel monoclonal antibody to nucleoprotein. *J. Clin. Microbiol.* 39, 3267–3271. doi: 10.1128/JCM.39.9.3267-3271

## FUNDING

This work was supported by a grant from the National Natural Science Foundation of China (31370093 and 81201320) to JY and XW, mega-projects of Science and Technology Research of China (Grant 2011ZX10004-001), and a grant from the National High Technology Research and Development Program of China (863 Program; grant no. SS2014AA022210).

## ACKNOWLEDGMENTS

We thank the staff of The China Mobile Laboratory Testing Team in Sierra Leone including Xiushan Zhang, Leili Jia, Chuanfu Zhang, Rongzhang Hao, Shuguang Tian, Xiang Zhao, Wei Wu, Lihua Wang, Ziqian Xu, Xitong Yuan, Ruizhong Jia, Rongtao Zhao, Yong Chen, Wenyi Zhang, Guohui Chang, and Zeliang Chen for the clinical detection of EBOV. We are also grateful to Li Fengjing and Xiao Shengli from Beijing Lanpu Bio-tech Co., Ltd. for technical assistance and helpful discussions.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmich.2015.01332>

- Notomi, T., Okayama, H., Masubuchi, H., Yonekawa, T., Watanabe, K., Amino, N., et al. (2000). Loop-mediated isothermal amplification of DNA. *Nucleic Acids Res.* 28, E63. doi: 10.1093/nar/28.12.e63
- Okafuji, T., Yoshida, N., Fujino, M., Motegi, Y., Ihara, T., Ota, Y., et al. (2005). Rapid diagnostic method for detection of mumps virus genome by loop-mediated isothermal amplification. *J. Clin. Microbiol.* 43, 1625–1631. doi: 10.1128/JCM.43.4.1625–1631.2005
- Saijo, M., Niikura, M., Morikawa, S., Ksiazek, T. G., Meyer, R. F., Peters, C. J., et al. (2001). Enzyme-linked immunosorbent assays for detection of antibodies to Ebola and Marburg viruses using recombinant nucleoproteins. *J. Clin. Microbiol.* 39, 1–7. doi: 10.1128/JCM.39.1.1–7.2001
- Song, T., Toma, C., Nakasone, N., and Iwanaga, M. (2005). Sensitive and rapid detection of *Shigella* and enteroinvasive *Escherichia coli* by a loop-mediated isothermal amplification method. *FEMS Microbiol. Lett.* 243, 259–263. doi: 10.1016/j.femsle.2004.12.014
- Watanabe, S., Watanabe, T., Noda, T., Takada, A., Feldmann, H., Jasenosky, L. D., et al. (2004). Production of novel ebola virus-like particles from cDNAs: an alternative to ebola virus generation by reverse genetics. *J. Virol.* 78, 999–1005. doi: 10.1128/JVI.78.2.999–1005.2004
- Weidmann, M., Muhlberger, E., and Hufert, F. T. (2004). Rapid detection protocol for filoviruses. *J. Clin. Virol.* 30, 94–99. doi: 10.1016/j.jcv.2003.09.004
- World Health Organization [WHO] (2014). *Urgently Needed: Rapid, Sensitive, Safe and Simple Ebola Diagnostic Tests*. Available at: <http://www.who.int/mediacentre/news/ebola/18-november-2014-diagnostics/en/> (accessed March 23, 2015).
- Zaki, S. R., Shieh, W. J., Greer, P. W., Goldsmith, C. S., Ferebee, T., Katshitshi, J., et al. (1999). A novel immunohistochemical assay for the detection of Ebola virus in skin: implications for diagnosis, spread, and surveillance of Ebola hemorrhagic fever. *Commission de Lutte contre les Epidemies a Kikwit. J. Infect. Dis.* 179 (Suppl. 1), S36–S47. doi: 10.1086/514319

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Li, Wang, Liu, Wei, Lin, Li, Li, Dong, Cui, Hu, Li, Ma, Zhao, Liu and Yuan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Corrigendum: Survey and Visual Detection of Zaire Ebolavirus in Clinical Samples Targeting the Nucleoprotein Gene in Sierra Leone

## OPEN ACCESS

### Edited and reviewed by:

Andres M. Perez,  
University of Minnesota, USA

### \*Correspondence:

Xiangna Zhao  
xnzhao@163.com;  
Chao Liu  
liuchao9588@sina.com;  
Jing Yuan  
yuanjing6216@163.com

<sup>†</sup>These authors are co-first authors.

### Specialty section:

This article was submitted to  
Infectious Diseases,  
a section of the journal  
*Frontiers in Microbiology*

Received: 26 May 2016

Accepted: 01 June 2016

Published: 15 June 2016

### Citation:

Li H, Wang X, Liu W, Wei X, Lin W, Li E, Li P, Dong D, Cui L, Hu X, Li B, Ma Y, Zhao X, Liu C and Yuan J (2016) Corrigendum: Survey and Visual Detection of Zaire Ebolavirus in Clinical Samples Targeting the Nucleoprotein Gene in Sierra Leone. *Front. Microbiol.* 7:948. doi: 10.3389/fmicb.2016.00948

Huan Li <sup>†</sup>, Xuesong Wang <sup>†</sup>, Wei Liu, Xiao Wei, Weishi Lin, Erna Li, Puyuan Li, Derong Dong, Lifei Cui, Xuan Hu, Boxing Li, Yanyan Ma, Xiangna Zhao <sup>\*</sup>, Chao Liu <sup>\*</sup> and Jing Yuan <sup>\*</sup>

Institute of Disease Control and Prevention, Academy of Military Medical Sciences, Beijing, China

**Keywords:** Zaire EBOV, RT-LAMP, rapid detection, sensitivity, specificity, prevalence

### A corrigendum on

#### Survey and Visual Detection of Zaire Ebolavirus in Clinical Samples Targeting the Nucleoprotein Gene in Sierra Leone

by Li, H., Wang, X., Liu, W., Wei, X., Lin, W., Li, E., et al. (2015). *Front. Microbiol.* 6:1332. doi: 10.3389/fmicb.2015.01332

### Reason for Corrigendum:

In the original article, the patients' names were inadvertently disclosed in the Supplementary Material. The Supplementary Files have since been updated to protect the patients' privacy. The authors apologize for this mistake.

## AUTHOR CONTRIBUTIONS

JY and XZ conceived and designed the experiments. XW and WL performed clinical detection in Sierra Leone. HL, PL, and DD performed and developed the RT-LAMP. YM, EL, PL, DD, DZ, LC, and XH performed the experiments. XZ wrote the manuscript. JY and XZ edited the manuscript.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Li, Wang, Liu, Wei, Lin, Li, Dong, Cui, Hu, Li, Ma, Zhao, Liu and Yuan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# A Novel Typing Method for *Streptococcus pneumoniae* Using Selected Surface Proteins

Arnau Domenech<sup>1,2†‡</sup>, Javier Moreno<sup>1,2‡</sup>, Carmen Ardanuy<sup>1,2</sup>, Josefina Liñares<sup>1,2</sup>, Adela G. de la Campa<sup>3,4</sup> and Antonio J. Martin-Galiano<sup>3\*</sup>

<sup>1</sup> Servicio de Microbiología, Hospital Universitari de Bellvitge, Universitat de Barcelona, IDIBELL, Barcelona, Spain, <sup>2</sup> CIBER de Enfermedades Respiratorias, Madrid, Spain, <sup>3</sup> Bacterial Genetics, Centro Nacional de Microbiología, Instituto de Salud Carlos III, Majadahonda, Spain, <sup>4</sup> Presidencia, Consejo Superior de Investigaciones Científicas, Madrid, Spain

## OPEN ACCESS

### Edited by:

Andres M. Perez,  
University of Minnesota, USA

### Reviewed by:

Margaret Ip,  
Chinese University of Hong Kong,  
China  
Josselin Noirel,  
Conservatoire National des Arts et  
Métiers, France

### \*Correspondence:

Antonio J. Martin-Galiano  
mgaliano@isciii.es

### †Present Address:

Arnau Domenech,  
Molecular Microbiology, Groningen  
Institute for Biomolecular Sciences  
and Biotechnology, Groningen,  
Netherlands

<sup>‡</sup>These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Infectious Diseases,  
a section of the journal  
Frontiers in Microbiology

Received: 13 November 2015

Accepted: 16 March 2016

Published: 31 March 2016

### Citation:

Domenech A, Moreno J, Ardanuy C,  
Liñares J, de la Campa AG and  
Martin-Galiano AJ (2016) A Novel  
Typing Method for *Streptococcus*  
*pneumoniae* Using Selected Surface  
Proteins. *Front. Microbiol.* 7:420.  
doi: 10.3389/fmicb.2016.00420

The diverse pneumococcal diseases are associated with different pneumococcal lineages, or clonal complexes. Nevertheless, intra-clonal genomic variability, which influences pathogenicity, has been reported for surface virulence factors. These factors constitute the communication interface between the pathogen and its host and their corresponding genes are subjected to strong selective pressures affecting functionality and immunogenicity. First, the presence and allelic dispersion of 97 outer protein families were screened in 19 complete pneumococcal genomes. Seventeen families were deemed variable and were then examined in 216 draft genomes. This procedure allowed the generation of binary vectors with 17 positions and the classification of strains into surfotypes. They represent the outer protein subsets with the highest inter-strain discriminative power. A total of 116 non-redundant surfotypes were identified. Those sharing a critical number of common protein features were hierarchically clustered into 18 surfogroups. Most clonal complexes with comparable epidemiological characteristics belonged to the same or similar surfogroups. However, the very large CC156 clonal complex was dispersed over several surfogroups. In order to establish a relationship between surfogroup and pathogenicity, the surfotypes of 95 clinical isolates with different serogroup/serotype combinations were analyzed. We found a significant correlation between surfogroup and type of pathogenic behavior (primary invasive, opportunistic invasive, and non-invasive). We conclude that the virulent behavior of *S. pneumoniae* is related to the activity of collections of, rather than individual, surface virulence factors. Since surfotypes evolve faster than MLSTs and directly reflect virulence potential, this novel typing protocol is appropriate for the identification of emerging clones.

**Keywords:** diagnosis, emergent clones, genomics, surface proteins, virulence factors

## INTRODUCTION

*Streptococcus pneumoniae*, the pneumococcus, is a prevalent member of the commensal flora of the nasopharynx. This bacterium can turn into a versatile pathogen with the ability to successfully colonize many environments inside the host (Bogaert et al., 2004). Pneumococcus is a major etiological agent of pneumonia, meningitis, sepsis, and otitis media. The chance of suffering a pneumococcal infection is dependent on the age group, lifestyle, and patient

co-morbidities. Different types of disease, symptom severity, and antimicrobial resistance rates associate epidemiologically to different pneumococcal lineages. Thus, a rational classification of isolates would improve patient management. Up to 96 serotypes have been classified according to the immunogenic properties of the polysaccharide capsule. The capsule is an important virulence factor that prevents complement-mediated phagocytosis. However, isolates that have switched their serotype by capsule gene exchange are favored under the selective pressure exerted by the serotype-based vaccines (Brueggemann et al., 2007).

Multilocus Sequence Typing (MLST; Maiden et al., 1998) is a typing method which provides a simplified view of genotypes. It is based on the allelic profiles of seven housekeeping gene fragments (*aroE*, *gdh*, *gki*, *recP*, *spi*, *xpt*, and *ddl*), which render sequence types (ST) grouped into clonal complexes (CC). For instance, ST180 and ST181 share all but one allele. Then, ST181 is a single locus variant of ST180. Both STs are grouped into clonal complex CC180, considering ST180 as the founder. However, intra-clonal variability associated with clinical behavior, e.g., local outbreaks, does exist (Silva et al., 2006; Moschioni et al., 2013). Subclones can emerge either from point mutations, deletions/duplications of key genes, or prophage integrations. However, the major source of evolution in *S. pneumoniae* is genetic recombination, a process facilitated by the natural competence of this bacterium. Recently, the massive sequencing of complete genomes has allowed the analysis of recent variations in alternative genes or genomic accessory regions (Donkor et al., 2012; Browall et al., 2013), which were not detectable by MLST or serotyping. These intra-clonal polymorphisms commonly occur on surface proteins (Croucher et al., 2011; Browall et al., 2013), which constitute the communication interface between pathogen and host. Many of these proteins play a role in virulence (Bergmann and Hammerschmidt, 2006). They typically have modular architectures: a universal cell-wall anchoring domain fused to an outer region that determines functional specificity. This outer region can diverge from strain to strain. This sequence divergence dictates surface protein activity and immunogenicity (Gravekamp et al., 1997). Moreover, “Non-Classical Surface Proteins” (NCSP) have also been reported, such as central metabolism enzymes that exert moonlighting activities when located in the cell wall (Bergmann et al., 2001).

Since isolates that have the same MLST may convey different surface proteins that affect pathogenicity, a new postgenomic typing system is required. In this study, we have developed such a system, termed surfotyping.

## MATERIALS AND METHODS

### Family Selection

The 19 genomes that were analyzed were selected among the 25 complete closed sequences stored at the NCBI FTP site (status: Jan/2014; Supplementary File S1). Surface proteins were identified using profiles and the Pfam domain search function applying gathering thresholds (Finn et al., 2014): choline-binding proteins (CBPs) using PF01473 and LPxTG-anchor proteins using PF00746. Lipoproteins were predicted with PRED-LIPO

(Bagos et al., 2008). NCSPs were obtained from two literature reviews (Bergmann and Hammerschmidt, 2006; Pérez-Dorado et al., 2012).

### Computational Surfotype and MLST Assignment

Surface proteins were identified in draft proteomes by BLAST using representative protein sequences (Supplementary File S2). BLAST was used using thresholds selected from the gold standard of 19 genomes. The combination of identity and BLAST score thresholds (Supplementary File S3) were established in the average point between lowest *bona fide* hits and the highest non-specific hits. Using the existence or absence of BLAST hits, surfotypes were derived as Boolean vectors. A BLAST *p*-value < 0.001 was required in all cases. Draft genomes were typed by MLST using BLAST. Query sequences used were those of the alleles present in the MLST web page (<http://pubmlst.org/spneumoniae/>). Assignment of an allele required a 100% identity over 100% length of the sequence. Subsequent ST and CC assignment was carried out using the information available in the same web page. Draft proteomes were downloaded from the public NCBI ftp site, [ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria\\_DRAFT/](ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria_DRAFT/) (Status 15/09/2014; Supplementary File S4).

### Surfotype Clustering into Surfogroups

The Inter-Surfomic distance (ISD) parameter between two surfotypes (*v* and *w*) was defined as:

$$ISD(v, w) = \frac{-100}{\log_{10} \prod_{i=1}^{17} F_{feat}}$$

where *i* stands for protein family index and *F<sub>feat</sub>* stands for the global frequency of the feature (either presence/absence or full/truncated allele) in the dataset of 19 reference genomes if *v* and *w* features matched or a value of 1 if they mismatch. Surfotypes were hierarchically clustered by their ISDs using the ward method of the *hclust* procedure available in the fast cluster package (Müllner, 2013) of the R-project.

Cluster feature consistency was calculated for every protein as the percentage of cases that match the most prominent feature in the surfogroup, considering the one with the lowest general frequency in case of a tie. Given that protein features have different occurrence, a normalized consistency (NC) for every protein was applied:

$$NC = \frac{\sum_{j=1}^{T_j} \left( \frac{n}{v} \right) \times Fn_j}{T_j}$$

where *j* stands for the cluster index, *T<sub>j</sub>* for the total number of clusters, *n* for the total number of surfotypes considered, *v* for the number of features per protein (*v* = 2 in this work) and *Fn<sub>j</sub>* for the natural frequency for the most prominent feature in the cluster *j*. Theoretically, NC may range from 50% (all features are equally represented in all clusters) to 100% (just one kind of feature is represented in every cluster).

## Experimental Surfotype Assignment by PCR

The 17 genes were screened by PCR in 95 isolates collected from patient attended between 2009 and 2011 at the Hospital Universitari de Bellvitge. These 95 isolates were selected as representatives of the different genotype-serotype combinations. MLST and serotypes were obtained retrospectively from frozen stocks of the isolates. Data was routinely obtained as part of the hospital daily practice. To estimate the relationship between surfogroup and epidemiological data, we assumed that all isolates of each serotype-genotype combination share the same surfogroup. We only considered the 27 cases in which  $\geq 4$  clinical records were available for isolates with the same SG-ST combination. Invasive rates and average patient age of clonal complexes were calculated from 610 clinical isolates collected from non-invasive (acute exacerbation of COPD  $n = 131$ , and non-bacteremic pneumonias  $n = 167$ ) and invasive ( $n = 334$ ) pneumococcal disease. Oligonucleotide sequences were acquired from the literature when dedicated papers for the family were available. Otherwise, they were designed for optimal selectivity using the reference genomes on gene regions identical in all family members. PCR conditions and oligonucleotides utilized in this work are listed in Supplementary File S5. Surfotype profiles were assigned to the pre-existing surfogroup with the most significant  $p$ -value (when  $< 0.05$ ). The  $p$ -value was calculated as the product of the probabilities of the matching features between the profile and the surfogroup signature. Unassigned profiles likewise were screened to the surfotype library, but applying a  $p$ -value threshold of 0.01. The classification performance of all cases ( $n$ ) was quantified through several estimators using true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Accuracy was defined as  $(TP + TN)/n$ ; sensitivity as  $TP/(TP + FN)$ ; specificity as  $TN/(TN + FP)$ ; and precision as  $TP/(TP + FP)$ .

## RESULTS

### Selection of Variable Protein Families

Surface proteins showing the highest variability were identified using a gold standard of 19 complete genomes with different surfogroup-sequence type (SG-ST) combinations (Figure 1A). Despite only 11 out of the 96 pneumococcal serotypes are included in these reference genomes, these serotypes are associated to a vast majority of clinical cases. In addition, they carry the virulence factors described in the literature at molecular and clinical detail (Bergmann and Hammerschmidt, 2006; Pérez-Dorado et al., 2012). Proteins considered for further analysis were those containing either a choline-binding domain; the LPxTG domain; the lipoprotein “lipobox” motif; or being reported as NCSP. The 1599 sequences found in the 19 genomes belonged to 97 homolog families (Supplementary File S6). Up to 75 families were present in most reference strains ( $\geq 16$ ) and their homologs shared a high identity ( $\geq 85\%$ ) over most of the sequence alignment ( $\geq 85\%$ ). These families were considered invariable and were discarded from the analysis (Figures 1B,C). The remaining 22 families showed

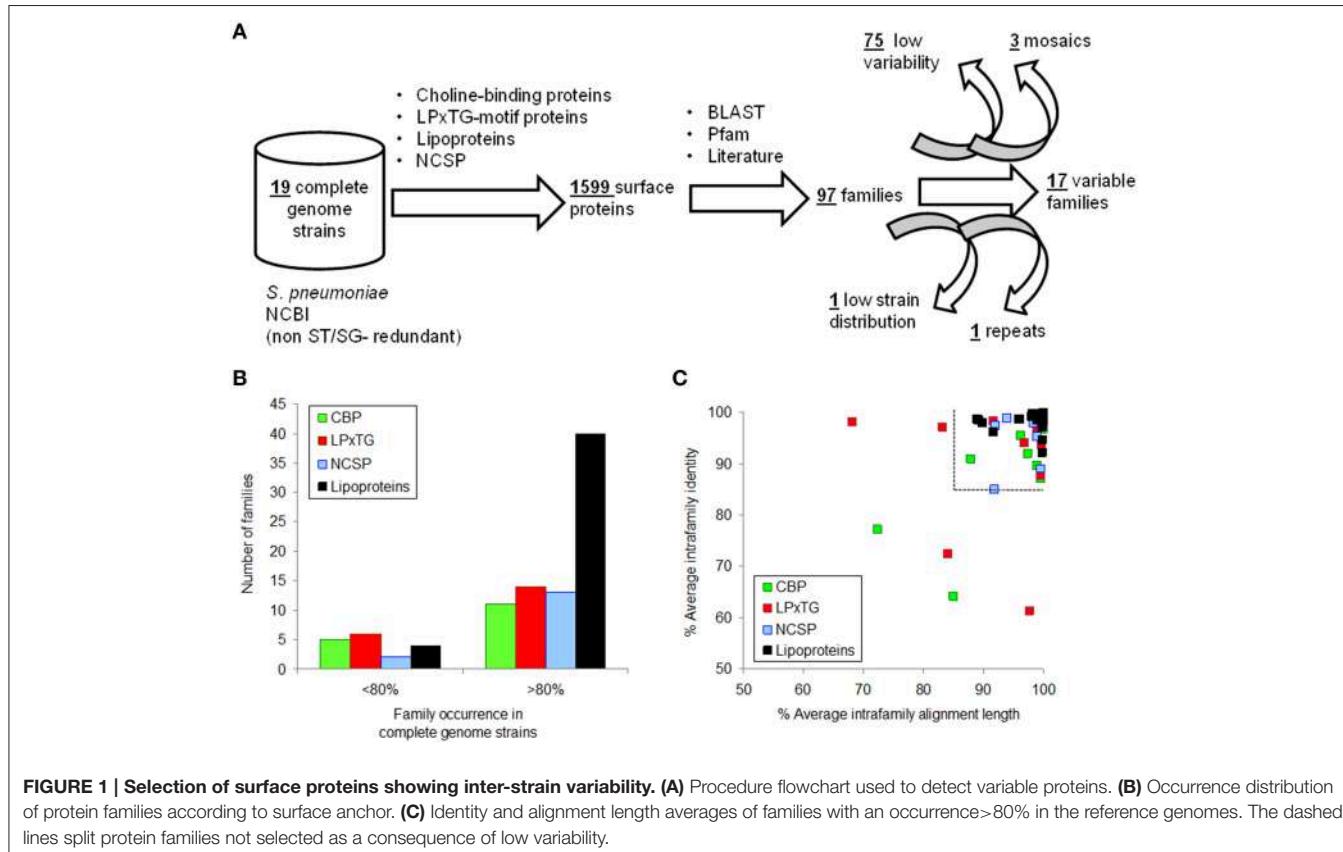
five kinds of disparity: presence versus absence; full versus truncated versions; continuum of number of repeated motifs; high sequence divergence; and domain mosaicism. Five protein families were further rejected. PavB was rejected because the actual number of repeats can be changed due to genome misassembling (Jensch et al., 2010). CbpA, IgA1, and PspA were rejected because their large sequence divergence or mosaicism (Hollingshead et al., 2000; Iannelli et al., 2002; Bek-Thomsen et al., 2012) prevented the direct comparison between variants. Lrp was rejected because was present in just two strains. Finally, 17 families were chosen for typing (Table 1): 15 with a pattern of presence/absence and 2 with a pattern of full/truncation. Many of these proteins are well-documented virulence factors and show particular Pfam domain combinations.

### Construction of Surfotypes and Clustering into Surfogroups

Binary patterns for the set of 17 protein families, denoting their presence/absence or full/truncated versions, may reflect the virulent capacity of clones. Representative protein sequences of every family were used to generate a library. The family members from TIGR4 and R6 strains were preferentially chosen since these isolates have been extensively used to study the molecular virulence of pneumococcus. This library was used to perform a BLAST screening on 216 draft proteomes, which covered 110 known STs (and 21 new) grouped into 31 CCs (plus 19 singletons). A total of unique 116 combinations, called surfotypes, were detected.

The convergence between surfotypes was quantified by the ISD (see Section Materials and Methods), a parameter that also considers the relative occurrence, in the dataset, of each protein feature. An ISD matrix between all unique surfotypes was calculated and then subjected to hierarchical clustering. The resulting clades were validated at progressive levels of granularity, from 1 to 40 clusters, calculating feature NC and clonal complex homogeneity at every level (Figure S1). This allows assessing the similarity between surfotype members.

The quality estimators reached an asymptote with 18 clusters, i.e., 87.5 and 67.6% for intracluster NC and MLST clonal complex homogeneity, respectively. From this point, a lower number of clusters caused spurious isolate cross-classification whereas a higher number results in excessive data partitioning without a substantial increment of cluster purity. These meaningful clusters were termed “surfogroups,” whose members shared a minimal common set of protein attributes that were termed “signatures” (Figure 2). The resultant surfogroups were dominated by clonal complexes whose pathogenic behavior is documented in the literature (Supplementary File S7). We utilized the fact that all surfogroups were dominated by a CC. Only if published data concerning the representative CC were scarce or nonexistent, virulence was supported by data from its commonest serotype or data from secondary (less abundant) CCs in the surfogroup. Primary invasive were those showing high invasive rates (CC217 and CC306 of serotype 1, 6, and CC191 of serotype 7F) or extreme rates or mortality (CC180 of serotype 3) in young adults.



**FIGURE 1 | Selection of surface proteins showing inter-strain variability. (A)** Procedure flowchart used to detect variable proteins. **(B)** Occurrence distribution of protein families according to surface anchor. **(C)** Identity and alignment length averages of families with an occurrence >80% in the reference genomes. The dashed lines split protein families not selected as a consequence of low variability.

Opportunistic invasive show higher carriage rates, although are still invasive in an age/comorbidity dependent manner. This is typical of CCs linked to 19A and 19F serotypes. Finally, non-invasive can cause non-invasive infection (described above) and/or show high carriage rates (CC81 linked to 23F serotype).

These reported epidemiological data are congruent with the hierarchical tree: seven surfogroups were ascribed to primary invasive (highly invasive in healthy population) isolates, seven were ascribed to opportunistic invasive pneumococcal disease (invasive potential in elderly patients and/or with co-morbidities) and 4 were correlated with non-invasive types of the disease.

## Correlation between Surfotyping and MLST

Up to 88 and 96% of isolates with the same ST shared the same surfotype or surfogroup, respectively. The clonal complexes had a more dispersed pattern since only 62 and 84% of strains with the same CC were classified into the same surfotype and surfogroup, respectively (Figure S2A). To obtain further insight into this intra-clonal discrepancy, the analysis was selectively performed on the five most prominent STs ( $\geq 5$  strains) and CCs ( $\geq 10$  strains). These STs were variable at the level of the preferred surfotype (37–86%), but essentially belonged to the same surfogroup (Figure S2B). All these CCs contained 5–7 surfotypes from 1 to 2 surfogroups (Figure S2C), with the exception of CC156, which dispersed into 17 surfotypes and 6 surfogroups.

## Surfotyping of Clinical Isolates

The 17 genetic features were screened by PCR in 95 isolates showing different genotype-serotype combinations (Supplementary File S8). All the isolates but one (98.9%) were reliably assigned to a surfogroup. To correlate surfogroup and epidemiological data, clinical reports recorded were utilized (See Section Materials and Methods; Figure 3A). The rate of primary invasive predictions was higher for those isolates that were, in fact, invasive (as defined as the ratio of invasive samples in the ST-SG combination)  $\geq 0.75$  and patient age  $\leq 68$  years. Opportunistic invasive predictions mainly appeared in the area of the graph covering an invasiveness score of 0.32–0.75 and an invasiveness score of >0.75 combined with patient age >68 years. Finally, non-invasive predictions correlate with isolates with an invasiveness score of <0.32. Using these clinical boundaries, surfogrouping predicted correctly 20 out of 27 tested ST-SG combinations (precision = 74.1%,  $p$ -value = 0.006 Fisher's exact test; Figure 3B).

## DISCUSSION

In this study, we have developed a strategy for formally classifying *S. pneumoniae* using the binary patterns of 17 highly discriminatory outer proteins (Figure 4). This allows for addressing the following issues: (1) to what extent outer protein profiles correlate to the invasive potential of pneumococcal

**TABLE 1 | List of selected surface proteins.**

Symbol <sup>a</sup>	TIGR4 <sup>b</sup>	Description	Pfam architecture <sup>c</sup>
CbpF (CBP)	SP0391	Choline binding protein F	[CB]6
CbpG (CBP)	SP0390	Choline binding protein G	Tripsin-[CB]3
CbpI (CBP)	SP0069	Choline binding protein I	[CB]6
CbpJ (CBP)	SP0378	Choline binding protein J	[CB]8
CbpL(CBP)	SP0667	Choline binding protein L	Excalibur-[CB]7-Lipoprotein_Ltp
DiiA (GPA)	SP1992	Dimorphic invasion-involved protein A	[B02864]1-2-DUF1542-LPxTG
NanC (NC)	SP1326	Neuraminidase C	Sialidase-BNR-BNR_2
NanE (LPP)	SP1330	N-acetylmannosamine-6-P epimerase	NanE
PclA (GPA)	NF	Collagen-like surface-anchored protein	YSIRK-G5-[Collagen]6
PhtA (NC)	SP1175	Pneumococcal histidine triad protein A	[Strep_His_triad]2-B01076-Strep_His_triad
PspC2 (GPA)	NF	Pneumococcal surface protein C	<YSIRK>-RICH-<B16622>-<B9758>-<[B503]1-2>-LPxTG
PsP (GPA)	SP1772	Pneumococcal serine-rich	[B214]10-LPxTG
RrgB (GPA)	SP0463	Ancillary pilus subunit B	Cna_B-LPxTG
SP_1796(LPP)	SP1796	Unknown substrate ABC transporter	SBP_bac_1
SrtD (LPP)	SP0468	Sortase D	Sortase
ZmpC (GPA)	SP0071	Zinc metalloproteinase	YSIRK-B134-B5460-B1438-G5-Peptidase_M26_N-B1656 Peptidase_M26_C
ZmpD (GPA)	NF	Zinc metalloproteinase	B5200-LPxTG-G5

<sup>a</sup>Protein class. CBP, Choline-binding protein; GPA, Gram-positive anchor (LPxTG motif-containing) protein; LPP, lipoprotein; NC, Non-classical surface protein.

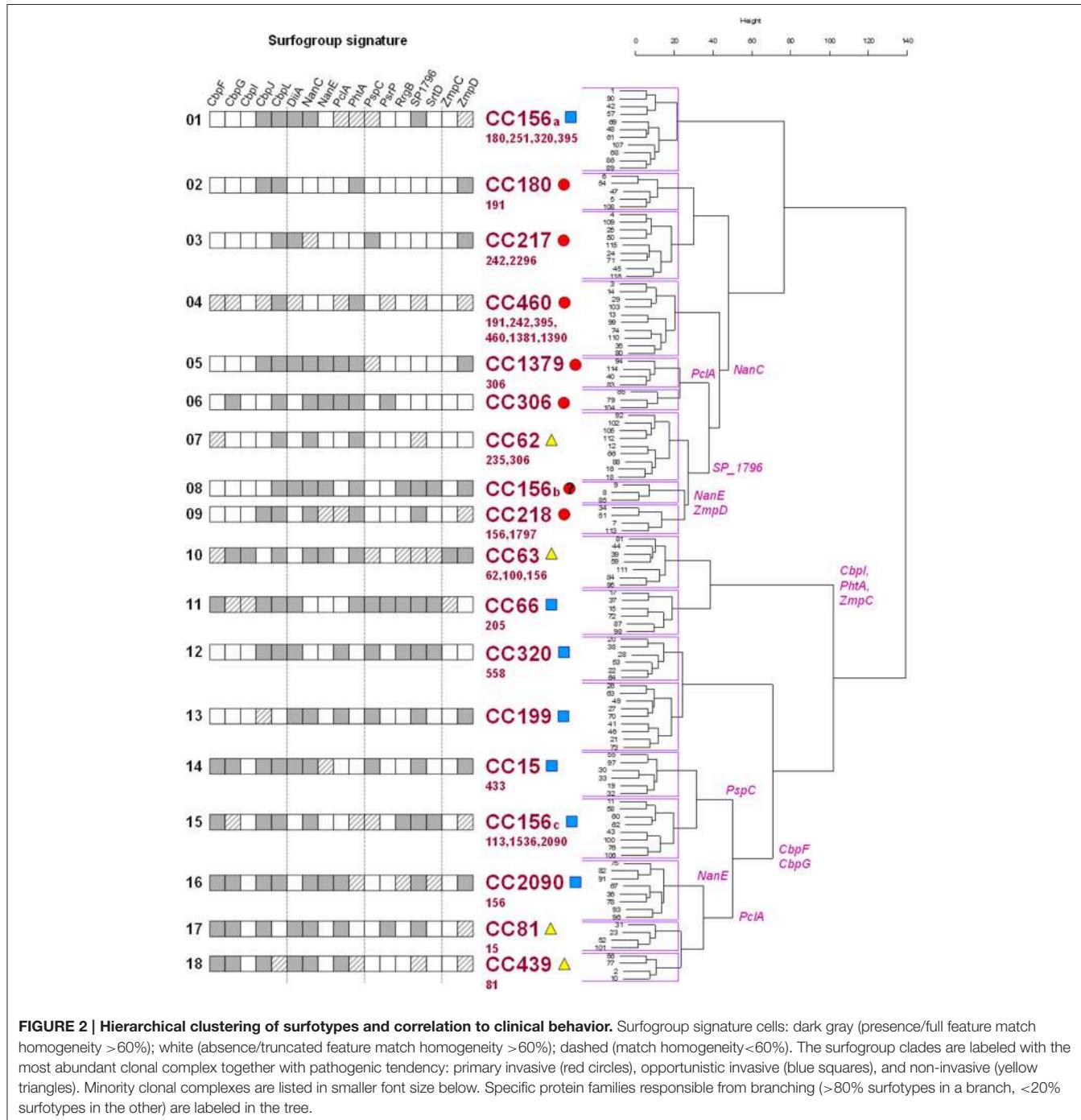
<sup>b</sup>NF, not found in TIGR4 strain.

<sup>c</sup>PfamA and PfamB (those starting by "B") domains are in sequential order. Accessory domains are in angle brackets. Repeated motifs are in square brackets together with the observed number of repeats. CB, Choline-binding motif. LPxTG, Gram-positive anchor containing the "LPxTG" sortase motif. Pfam domains with the lowest E-values were prioritized. Pfam domains (Eval < 0.01) were also considered only if overlapped <50% in length with more significant domains.

clones and, consequently, the potential diagnostic applications of surfotypes; and (2) the relationship between the evolution of the surface proteome and the MLST genes. Despite what other similar studies have been reported (Dagerhamn et al., 2008; Desa et al., 2008; Imai et al., 2011; Browall et al., 2013), our approach is more comprehensive in terms of strain disparity, is focused on accessory surface proteins, and applies new statistical strategies. Surfotyping relies on profiles acquired via PCR screening or genomic sequencing, techniques which may lead to misleading results. Oligonucleotides may not anneal with sufficient affinity to template DNA in the case of a mismatch. Likewise, ORFs targeted in draft genomes might be interrupted by the contig limits and remain spuriously undetected. Nevertheless, these two methodologies complemented each other reasonably well. As illustrative examples, SG1-ST306 isolates, which cause invasive disease in young adults without prior colonization of the nasopharynx, were assigned to the primary invasive surfogroup Sfg06. 15A-ST63 clones, which typically cause acute exacerbations in COPD patients (Domenech et al., 2014), were classified as non-invasive Sfg10. The most remarkable exception was SG3-ST180, which was predicted to be an invasive opportunistic isolate after surfotyping despite being in a non-invasive position. This may be a consequence of the especially thick capsule of type 3, which would affect the activity of some protein determinants and therefore cause misclassification.

Despite the fact that SG-ST combinations are associated with different capacities to colonize human body niches and distinct patient types, current studies have failed to attribute virulent behavior to a single gene (Manso et al., 2014).

Moreover, the contribution of a given gene to virulence seems dependent on other genome regions (Thomas et al., 2011). This is probably because the factors necessary for virulence are relatively redundant (Blomberg et al., 2009). There is evidence to support the idea that pneumococcal virulence is network-based and, therefore, a matter to be understood through the lens of systems biology, as proposed for *Staphylococcus aureus* (Sanchez et al., 2011). These pathofunctional networks may operate by following an orchestrated spatiotemporal pattern that eventually leads to a given clinical outcome. However, inferring explicit relationships between these proteins and disease is far from trivial considering that some of them play unknown or several roles. For instance, CbpG is not only involved in adherence to epithelium, but also in the cleavage of extracellular matrix (Mann et al., 2006). The non-invasive Sfg10 signature contains the sialic acid epimerase NanE, the putative Zn-scavenger PhtA (RiouxB et al., 2011), and ZmpC, which prevents the influx of neutrophils (Surewaard et al., 2013). These three functions combined may favor long-term mucosae disease patterns and be selected for in isolates causing non-bacteremic pneumonia and COPD acute exacerbations. The RrgA and StrD proteins, which are involved in the constitution and location of the adhesive pilus, are present in the opportunistic invasive surfogroups Sfg11, Sfg12, and Sfg15. This observation suggests that many of the discriminatory proteins selected in this work may be involved in long-term persistence and asymptomatic colonization. These processes have to be maintained until the infection is favored by particular host conditions. In this light, Sfg02, Sfg03, and Sfg04 surfogroups harbor the lowest number of surface proteins in



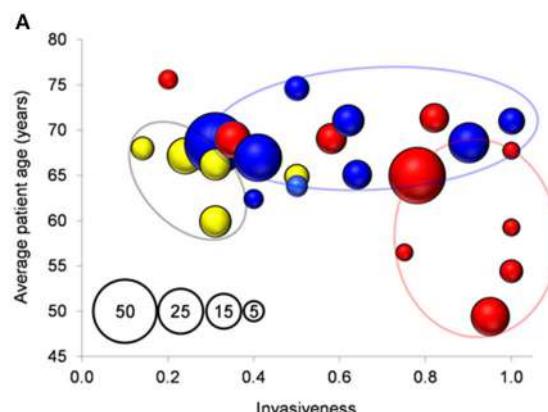
**FIGURE 2 | Hierarchical clustering of surfotypes and correlation to clinical behavior.** Surfogroup signature cells: dark gray (presence/full feature match homogeneity >60%); white (absence/truncated feature match homogeneity >60%); dashed (match homogeneity <60%). The surfogroup clades are labeled with the most abundant clonal complex together with pathogenic tendency: primary invasive (red circles), opportunistic invasive (blue squares), and non-invasive (yellow triangles). Minority clonal complexes are listed in smaller font size below. Specific protein families responsible from branching (>80% surfotypes in a branch, <20% surfotypes in the other) are labeled in the tree.

the dataset, even though they are related to a primary invasive phenotype. Isolates belonging to these surfogroups have short colonization periods, and therefore would require less adhesive factors.

A relevant factor that could interfere to surfotyping is the introduction of pneumococcal conjugate vaccines. Some degree of co-evolution between the gene pools encoding the capsule and the accessory surfome could be expected, which together may largely determine the pathogenic behavior of a given

pneumococcal lineage. In this light, the detection of surfogroups in capsular types in which they were not previously reported may be synonym of potential emergent clone generated by capsule switching and should be tracked.

MLST and surfotyping methods are conceptually different (Table 2). MLST genes evolve at a slow pace, making them appropriate for reconstructing the phylogeny of the species. MLST is based on the analysis of SNPs, which should not have a noticeable influence on protein function. In

**B**

Pathogenic behavior	Accuracy	Sensitivity	Specificity
Primary invasive	85.2%	66.0%	100.0%
Opportunistic invasive	77.8%	80.0%	76.5%
Non-invasive	85.2%	80.0%	86.4%

**FIGURE 3 | Correlation between surfogroup and clinical isolates.**

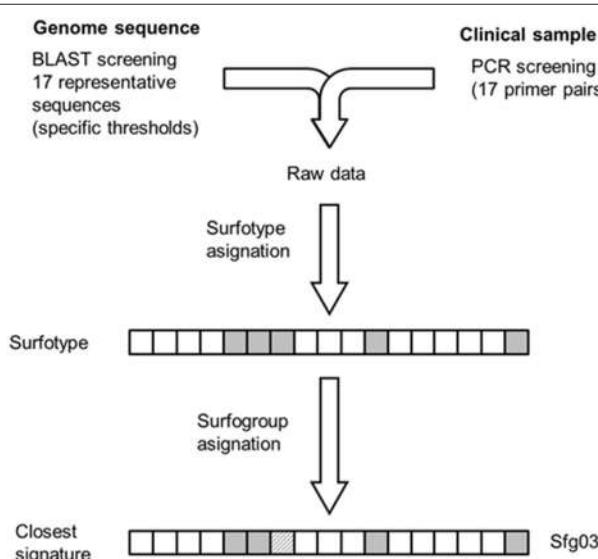
(A) Each bubble represents a unique SG-ST combination. Bubble size (see pattern in the inset): number of clinical isolates. Bubbles are colored according to type of pathogenicity after surfogroup prediction according to **Figure 2**. (B) Measures of the classification performance.

**TABLE 2 | Essential differences between MLST and Surfotyping.**

Topic	MLST	Surfotyping
Number of genes	Fixed (seven)	Variable (species-specific)
Type of variability	SNPs	Presence/absence, distant allelic variants (many residue changes), large insertions/deletions, mosaicism
Gene evolution rate	Slow	Fast
Protein location	Cytoplasm	Cell wall
Protein role	Housekeeping	Virulence
Correlation to pathogenicity	Indirect association	Direct causality
Protein structural nature	Globular	Disordered regions, tandem repeats, anchor modules
Gene phyletic dispersion	Universal	Species-specific

emerging, highly-virulent clones. The 116 non-redundant surfotypes found describe a continuous 17th dimensional space, in which surfogroups can be observed as dense zones enriched in strains causing similar diseases. Genes that do not contribute sufficiently to the life-style phenotype may eventually be lost. This would explain why these genes are therefore absent from the signature. Surfogrouping merge some clonal complexes. However, large clonal complexes can be located in separate surfogroups. In particular, the giant CC156 surfogroup, whose size is a consequence of eBURST clustering collapse, reached a mere 40% surfogroup homogeneity. Meanwhile, the CC156 lineages derived from MLST-96 technique (Moschioni et al., 2013) are reasonably associated with different surfogroups.

Our results support a model of a complex association between pneumococcal surface factors and disease. The pathogen-host interaction would not behave according to a lock-and-key paradigm respect to their target molecules but as a bunch of keys for an array of locks. Pneumococci use its highly recombinogenic capacity as if they were “slot machines” whose winning feature combinations provide a higher efficiency for a given virulent scenario. This work provides a first report of the combinations that may be useful for predicting disease progression.

**FIGURE 4 | Methodological scheme for surfotype and surfogroup assignment of test isolates.** Raw data derived from either sequencing or PCR is processed into a 17-mer Boolean vector (presence-full or absence-truncation). Assignment of surfotypes to the surfogroups showed in **Figure 2** can be done through feature-by-feature comparison against the surfogroup signatures.

contrast, surfotyping prioritizes functions, which are subjected to strong selective pressure in terms of adaptation to defined pathogenic scenarios. Thus, surfotyping may be instrumental in detecting subtle genetic changes that could lead to

## ETHICAL STANDARDS

Written or oral informed consent was not required, because the source of the bacterial isolates was anonymized.

## AUTHOR CONTRIBUTIONS

AD and JM carried out the experimental work. CA, JL, AD, and AM designed and analyzed the assays. AM carried out the computational work and wrote the manuscript.

## FUNDING

This work was supported by a Miguel Servet contract from the Spanish Ministry of Health to AM, Plan Nacional de I+D+I of the Ministry of Science and Innovation (BIO2011-25343, BIO2014-555462-R, SAF2012-39444-C02), Fondo de Investigaciones Sanitarias de la Seguridad Social (PI11/00763) and Fondo Europeo de Desarrollo Regional (FEDER). CIBER Enfermedades Respiratorias is an initiative of the Instituto de Salud Carlos III.

## REFERENCES

- Bagos, P. G., Tsirigos, K. D., Liakopoulos, T. D., and Hamodrakas, S. J. (2008). Prediction of lipoprotein signal peptides in Gram-positive bacteria with a Hidden Markov Model. *J. Proteome Res.* 7, 5082–5093. doi: 10.1021/pr08012c
- Bek-Thomsen, M., Poulsen, K., and Kilian, M. (2012). Occurrence and evolution of the paralogous zinc metalloproteases IgA1 protease, ZmpB, ZmpC, and ZmpD in *Streptococcus pneumoniae* and related commensal species. *MBio* 3:e00303-12. doi: 10.1128/mBio.00303-12
- Bergmann, S., and Hammerschmidt, S. (2006). Versatility of pneumococcal surface proteins. *Microbiology* 152, 295–303. doi: 10.1099/mic.0.28610-0
- Bergmann, S., Rohde, M., Chhatwal, G. S., and Hammerschmidt, S. (2001). alpha-Enolase of *Streptococcus pneumoniae* is a plasmin(ogen)-binding protein displayed on the bacterial cell surface. *Mol. Microbiol.* 40, 1273–1287. doi: 10.1046/j.1365-2958.2001.02448.x
- Blomberg, C., Dagerhamn, J., Dahlberg, S., Browall, S., Fernebro, J., Albiger, B., et al. (2009). Pattern of accessory regions and invasive disease potential in *Streptococcus pneumoniae*. *J. Infect. Dis.* 199, 1032–1042. doi: 10.1086/597205
- Bogaert, D., de Groot, R., and Hermans, P. W. (2004). *Streptococcus pneumoniae* colonisation: the key to pneumococcal disease. *Lancet Infect. Dis.* 4, 144–154. doi: 10.1016/S1473-3099(04)00938-7
- Browall, S., Norman, M., Tangrot, J., Galanis, I., Sjostrom, K., Dagerhamn, J., et al. (2013). Intralclonal variations among *Streptococcus pneumoniae* isolates influence likelihood of invasive disease in children. *J. Infect. Dis.* 209, 377–388. doi: 10.1093/infdis/jit481
- Brueggemann, A. B., Pai, R., Crook, D. W., and Beall, B. (2007). Vaccine escape recombinants emerge after pneumococcal vaccination in the United States. *PLoS Pathog.* 3:e168. doi: 10.1371/journal.ppat.0030168
- Croucher, N. J., Harris, S. R., Fraser, C., Quail, M. A., Burton, J., van der Linden, M., et al. (2011). Rapid pneumococcal evolution in response to clinical interventions. *Science* 331, 430–434. doi: 10.1126/science.1198545
- Dagerhamn, J., Blomberg, C., Browall, S., Sjöström, K., Morfeldt, E., and Henriques-Normark, B. (2008). Determination of accessory gene patterns predicts the same relatedness among strains of *Streptococcus pneumoniae* as sequencing of housekeeping genes does and represents a novel approach in molecular epidemiology. *J. Clin. Microbiol.* 46, 863–868. doi: 10.1128/JCM.01438-07
- Desa, M. N., Sekaran, S. D., Vadivelu, J., and Parasakthi, N. (2008). Distribution of CBP genes in *Streptococcus pneumoniae* isolates in relation to vaccine types, penicillin susceptibility and clinical site. *Epidemiol. Infect.* 136, 940–942. doi: 10.1017/S0950268807009363
- Domenech, A., Ardanuy, C., Tercero, A., García-Somoza, D., Santos, S., and Liñares, J. (2014). Dynamics of the pneumococcal population causing acute exacerbations in COPD patients in a Barcelona hospital (2009–12): comparison with 2001–04 and 2005–08 periods. *J. Antimicrob. Chemother.* 69, 932–939. doi: 10.1093/jac/dkt476
- Donkor, E. S., Stabler, R. A., Hinds, J., Adegbola, R. A., Antonio, M., and Wren, B. W. (2012). Comparative phylogenomics of *Streptococcus pneumoniae* isolated from invasive disease and nasopharyngeal carriage from West Africans. *BMC Genomics* 13:569. doi: 10.1186/1471-2164-13-569
- Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2014). Pfam: the protein families database. *Nucleic Acids Res.* 42, D222–D230. doi: 10.1093/nar/gkt1223
- Gravekamp, C., Kasper, D. L., Michel, J. L., Kling, D. E., Carey, V., and Madoff, L. C. (1997). Immunogenicity and protective efficacy of the alpha C protein of group B streptococci are inversely related to the number of repeats. *Infect. Immun.* 65, 5216–5221.
- Hollingshead, S. K., Becker, R., and Briles, D. E. (2000). Diversity of PspA: mosaic genes and evidence for past recombination in *Streptococcus pneumoniae*. *Infect. Immun.* 68, 5889–5900. doi: 10.1128/IAI.68.10.5889-5900.2000
- Iannelli, F., Oggioni, M. R., and Pozzi, G. (2002). Allelic variation in the highly polymorphic locus pspC of *Streptococcus pneumoniae*. *Gene* 284, 63–71. doi: 10.1016/S0378-1119(01)00896-4
- Imai, S., Ito, Y., Ishida, T., Hirai, T., Ito, I., Yoshimura, K., et al. (2011). Distribution and clonal relationship of cell surface virulence genes among *Streptococcus pneumoniae* isolates in Japan. *Clin. Microbiol. Infect.* 17, 1409–1414. doi: 10.1111/j.1469-0919.2010.03446.x
- Jensch, I., Gámez, G., Rothe, M., Ebert, S., Fulde, M., Somplatzki, D., et al. (2010). PavB is a surface-exposed adhesin in *Streptococcus pneumoniae* contributing to nasopharyngeal colonization and airways infections. *Mol. Microbiol.* 77, 22–43. doi: 10.1111/j.1365-2958.2010.07189.x
- Maiden, M. C., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., et al. (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. U.S.A.* 95, 3140–3145. doi: 10.1073/pnas.95.6.3140
- Mann, B., Orihuela, C., Antikainen, J., Gao, G., Sublett, J., Korhonen, T. K., et al. (2006). Multifunctional role of choline binding protein G in pneumococcal pathogenesis. *Infect. Immun.* 74, 821–829. doi: 10.1128/IAI.74.2.821-829.2006
- Manso, A. S., Chai, M. H., Atack, J. M., Furi, L., De Ste, C. M., Haigh, R., et al. (2014). A random six-phase switch regulates pneumococcal virulence via global epigenetic changes. *Nat. Commun.* 5, 5055. doi: 10.1038/ncomms6055
- Moschioni, M., Lo, S. M., Crisafulli, G., Torricelli, G., Guidotti, S., Muzzi, A., et al. (2013). Sequence analysis of 96 genomic regions identifies distinct evolutionary lineages within CC156, the largest *Streptococcus pneumoniae* clonal complex in the MLST database. *PLoS ONE* 8:e61003. doi: 10.1371/journal.pone.0061003
- Müllner, D. (2013). Fastcluster: fast hierarchical, agglomerative clustering routines for R and python. *J. Stat. Softw.* 53, 1–18. doi: 10.18637/jss.v053.i09
- Pérez-Dorado, I., Galan-Bartual, S., and Hermoso, J. A. (2012). Pneumococcal surface proteins: when the whole is greater than the sum of its parts. *Mol. Oral Microbiol.* 27, 221–245. doi: 10.1111/j.2041-1014.2012.00655.x
- Rioux, S., Neyt, C., Di, P. E., Turpin, L., Charland, N., Labbé, S., et al. (2011). Transcriptional regulation, occurrence and putative role of the Phf family of *Streptococcus pneumoniae*. *Microbiology* 157, 336–348. doi: 10.1099/mic.0.042184-0
- Sanchez, C. J., Hinojosa, C. A., Shivshankar, P., Hyams, C., Camberlein, E., Brown, J. S., et al. (2011). Changes in capsular serotype alter the surface exposure of pneumococcal adhesins and impact virulence. *PLoS ONE* 6:e26587. doi: 10.1371/journal.pone.0026587

## ACKNOWLEDGMENTS

We thank Tahl Zimmerman for English correction.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2016.00420>

- Silva, N. A., McCluskey, J., Jefferies, J. M., Hinds, J., Smith, A., Clarke, S. C., et al. (2006). Genomic diversity between strains of the same serotype and multilocus sequence type among pneumococcal clinical isolates. *Infect. Immun.* 74, 3513–3518. doi: 10.1128/IAI.00079-06
- Surewaard, B. G., Trzcinski, K., Jacobino, S. R., Hansen, I. S., Vughs, M. M., Sanders, E. A., et al. (2013). Pneumococcal immune evasion: ZmpC inhibits neutrophil influx. *Cell. Microbiol.* 15, 1753–1765. doi: 10.1111/cmi.12147
- Thomas, J. C., Figueira, M., Fennie, K. P., Laufer, A. S., Kong, Y., Pichichero, M. E., et al. (2011). *Streptococcus pneumoniae* clonal complex 199: genetic diversity and tissue-specific virulence. *PLoS ONE* 6:e18649. doi: 10.1371/journal.pone.0018649

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Domenech, Moreno, Ardanuy, Liñares, de la Campa and Martin-Galiano. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Prevalence of *Escherichia coli* Virulence Genes in Patients with Diarrhea and a Subpopulation of Healthy Volunteers in Madrid, Spain

Adriana Cabal<sup>1,2</sup>, María García-Castillo<sup>3,4</sup>, Rafael Cantón<sup>3,4</sup>, Christian Gortázar<sup>2</sup>, Lucas Domínguez<sup>1</sup> and Julio Álvarez<sup>5\*</sup>

<sup>1</sup> VISAVER Health Surveillance Centre, Universidad Complutense, Madrid, Spain, <sup>2</sup> SaBio-IREC (CSIC-UCLM-JCCM), Ciudad Real, Spain, <sup>3</sup> Servicio de Microbiología, Hospital Universitario Ramón y Cajal, Madrid, Spain, <sup>4</sup> Servicio de Microbiología, Instituto Ramón y Cajal de Investigación Sanitaria, Madrid, Spain, <sup>5</sup> Department of Veterinary Population Medicine, College of Veterinary Medicine, University of Minnesota, St. Paul, MN, USA

## OPEN ACCESS

### Edited by:

Paul D. Brown,  
University of the West Indies, Jamaica

### Reviewed by:

Mirjam Kooistra-Smid,  
University Medical Center Groningen,  
Netherlands  
Marquita Vernescia Gittens-St.Hilaire,  
University of the West Indies,  
Barbados

### \*Correspondence:

Julio Alvarez  
jalvarez@umn.edu

### Specialty section:

This article was submitted to  
Infectious Diseases,  
a section of the journal  
*Frontiers in Microbiology*

Received: 21 January 2016

Accepted: 18 April 2016

Published: 02 May 2016

### Citation:

Cabal A, García-Castillo M, Cantón R, Gortázar C, Domínguez L and Álvarez J (2016) Prevalence of *Escherichia coli* Virulence Genes in Patients with Diarrhea and a Subpopulation of Healthy Volunteers in Madrid, Spain. *Front. Microbiol.* 7:641.  
doi: 10.3389/fmicb.2016.00641

Etiological diagnosis of diarrheal diseases may be complicated by their multi-factorial nature. In addition, *Escherichia coli* strains present in the gut can occasionally harbor virulence genes (VGs) without causing disease, which complicates the assessment of their clinical significance in particular. The aim of this study was to detect and quantify nine VGs (*stx1*, *stx2*, *eae*, *aggR*, *ehxA*, *invA*, *est*, *elt* and *bfpA*) typically present in five *E. coli* enteric pathotypes [enterohaemorrhagic *E. coli* (EHEC), enterotoxigenic *E. coli* (ETEC), enteropathogenic *E. coli* (EPEC), enteroaggregative *E. coli* (EAEC), and enteroinvasive *E. coli* (EIEC)] in fecal samples collected from 49 patients with acute diarrhea and 32 healthy controls from Madrid, Spain. In addition, the presence of four serotype-related genes (*wzxO104* and *fliC<sub>H4</sub>*, *rfbO157*, and *fliC<sub>H7</sub>*) was also determined. Presence of target genes was assessed using a quantitative real-time PCR assay previously developed, and the association of presence and burden of VGs with clinical disease and/or other risk factors was explored. Prevalence of *ehxA* [typically associated with Shigatoxin producing *E. coli* (STEC) and (EPEC), *invA* (EIEC), and the *rfbO157*+*fliC<sub>H7</sub>* (STEC)] combination were significantly ( $p < 0.02$ ) higher in the diarrheic group, while the *wzxO104*+*fliC<sub>H4</sub>* combination was significantly ( $p = 0.014$ ) more prevalent in the control group. On the other hand, *eae* was detected in more than 90% of the individuals in both patient and control populations, and it was not associated with *bfpA*, suggesting the absence of typical EPEC. No significant differences in the quantitative values were detected for any VG among study groups, but the difference in the load of *aggR* (EAEC) and *invA* in the patients with respect to the controls was close to the significance, suggesting a potential role of these VGs in the clinical signs observed when they are present at high levels.

**Keywords:** *E. coli*, diarrhea, virulence genes, pathotypes, prevalence

## INTRODUCTION

*Escherichia coli* are commensal bacteria living in the intestinal tract of animals and humans. Some innocuous strains can incorporate virulence genes (VGs) by lateral gene transfer that may allow them to cause intestinal and extraintestinal disease (Ochman et al., 2000). There are six pathotypes than can produce intestinal disease in humans: STEC (Shiga-toxin-producing *E. coli*, including EHEC – enterohaemorrhagic *E. coli*), EPEC (Enteropathogenic *E. coli*), ETEC (Enterotoxigenic *E. coli*), EIEC (Enteroinvasive *E. coli*), EAEC (Enteropathogenic *E. coli*), and DAEC (Diffusely adherent *E. coli*; Kaper, 2005). *E. coli* strains are classified into these pathotypes depending on the presence/absence of several combinations of VGs. In EPEC, the virulence machinery is based on the carriage of *eae*, *tir*, and other proteins required for causing attachment and effacement (A/E) lesions and they are encoded on a chromosomal pathogenicity island known as the locus for enterocyte effacement (LEE; Kaper et al., 2004). Typical EPEC (tEPEC) strains also possess the EAF plasmid, which encodes for the bundle-forming pili (*bfpA* gene) and the *perABC* genes that regulate *eae* expression. Atypical EPEC (aEPEC) strains lack the EAF plasmid (Johnson and Nolan, 2009), and can affect children in both developed and developing countries, and together with EHEC and EAEC strains, are considered an emerging pathogen (Trabulsi et al., 2002; Huang et al., 2004). EAEC strains possess a pAA plasmid for fimbriae production, which contains the *aggR* transcriptional activator and its regulated genes (Kaper et al., 2004; Johnson and Nolan, 2009). In addition, there is a cluster of atypical EAEC strains that lacks *aggR* and may not produce diarrhea in humans. EHEC infection can lead to hemorragic colitis (HC) and in the most severe cases hemolytic-uremic syndrome (HUS; Paton and Paton, 1998), which is often associated with serotypes O157:H7 and, recently, O104:H4. Their virulence is mainly due to two types of Shiga-like toxins, *Stx1* and *Stx2*, but also to *eae* (intimin) and *ehxA* (enterohaemolysin). EIEC carries an invasion plasmid (*pInv*) which encodes the *vir* regulon, which is key for intestinal dysentery. ETEC has been reported as well as the causative agent of the traveler's diarrhea in patients who traveled to developing countries (Gascon et al., 1998). ETEC pathogenicity is mainly determined by the production of heat-stable (ST) and labile-stable (LT) enterotoxins. In patients with an EIEC or ETEC infection, symptoms usually can be resolved without any complication while in severe cases of EHEC, 10% of the patients can develop HUS (Nataro and Kaper, 1998). New pathotypes may also emerge due to the genetic recombination of some of these VGs, as demonstrated in the 2011 outbreak in Germany caused by a new O104:H4 EAEC/EHEC strain that affected 3,816 people and caused 54 deaths (Frank et al., 2011). In Spain, only one HUS STEC case and one non-HUS case due to the infection with the EAEC/EHEC strain were reported (Mora et al., 2011).

In both developing and developed countries, the causative agents of diarrhea may vary. In developed countries, diarrhea caused by *E. coli* (attributed to EPEC, ETEC, and other pathotypes) in children accounts for less than 0.4% of the cases (Fletcher et al., 2013) while the detection rate of

*E. coli* in adults (>12 years) is even lower. In Spain, EPEC, STEC, and EAEC are the *E. coli* pathotypes most commonly isolated in clinical samples (Blanco et al., 2006). However, there is limited knowledge about the occurrence and quantitative burden of *E. coli* VGs in healthy individuals, which could help to evaluate results from clinically affected individuals.

The objective of this study was to evaluate if there were differences in the presentation and amount of VGs in feces from patients with diarrhea and a subpopulation of healthy individuals potentially exposed to certain *E. coli* pathotypes (veterinary students and laboratory staff) using a direct quantitative real-time PCR. Four serotype-related genes associated with the O157:H7 and O104:H4 serotypes (previously associated with outbreaks of importance in public health) were also investigated by qPCR. Additionally, we evaluated the association between certain individual characteristics and the presence and burden of VGs in the healthy population to identify factors promoting exposure.

## MATERIALS AND METHODS

### Study Population

Feces from a total of 81 individuals from two different populations were collected. First, 49 samples from individuals belonging to a clinically affected population ("cases") were selected randomly among those received during a 1-month period at the Ramón y Cajal Hospital (Madrid, Spain) from patients suffering from acute diarrhea (aqueous to mucoid) of an unknown origin. Information on the gender and age of the patients and the origin of the patient (hospitalization area, emergency room or primary health center) were recovered from the clinical chart from all but three patients. Samples from 30 female and 16 male patients were included in the study, with ages ranging from 1 to 94 years old. Thirty-three went to their Primary Health Center, 10 went directly to the Hospital and 3 were admitted in the Emergency room. None of the patients had received antimicrobial treatment before their samples were collected.

The healthy population ("controls") consisted of 20 veterinary students at the University Complutense and 12 members of the VISAVENT Health Surveillance Center staff (20 females and 12 males with ages ranging between 18 and 44 years) that volunteered to participate in the project. None of them presented any gastrointestinal symptoms when the samples were collected. Information on the gender, eating habits, traveling history, previous intestinal illnesses, and antimicrobial therapy were collected using a questionnaire (available upon request). This study was authorized by the Ethics Committee for Clinical Research from the Hospital Ramon y Cajal (Reference 098/12).

### VG Detection

Gene selection was made based on a previous study (Cabal et al., 2013) so that it included the most representative VGs from the intestinal pathotypes of *E. coli* and four genes associated with relevant serotypes for public health (Supplementary Table S1).

Three grams of the fecal samples were mixed individually with 3 ml of PBS (1/2 dilution) and homogenized vigorously. Four-hundred milligrams of the resulting solution were used for DNA extraction, which was performed using a commercial kit (Qiagen DNA stool mini kit) according to the manufacturer's instructions. Then, the amplification of the targets was performed using a real-time PCR assay (qPCR) as previously described (Cabal et al., 2015). Briefly, for detection of VGs, a conventional PCR was first performed using specific primers (Supplementary Table S2) for control strains (see below) to obtain specific PCR products, which were ten-fold diluted in order to generate the standard curves. For the serotype related genes, 10-fold dilutions of STEC control strains were performed and further used for building the standards needed for the qPCR. Then, qPCR analysis of the clinical samples was carried out using primers and probes described in Supplementary Tables S1 and S3, and quantification was achieved using the standard curves generated as explained below. The final gene copy number per milligram of sample was estimated taking into account the initial weight of the stool, the dilution factor and the DNA volume added to the qPCR.

The following *E. coli* control strains were included in all qPCR reactions as positive controls and were also used to generate the standard curves: STEC O157:H7 (strain CNM 2686/03 positive to *stx1*, *sxt2*, *eae*, and *ehxA*), a typical EPEC (strain CNM764, positive to *bfpA*, and *eae*), provided by Dr. Silvia Herrera León (National Center of Microbiology, Institute of Health Carlos III, Madrid, Spain), ETEC (strain H10407, positive to *elt*, and *est*), one EIEC (strain 1280 positive to *invA*), and the EAEC/EHEC outbreak strain (LB226692 positive to *aggR* and *stx2*), provided by Dr. Martina Bielaszewska (Institute of Hygiene, Münster University, Münster, Germany).

## Statistical Analysis

The proportion of positive samples and the quantitative results for each gene in the case and control groups were compared using Pearson's and Fisher's Exact  $\chi^2$  tests and Mann–Whitney tests, respectively. In addition, results of the PCR analysis among the controls were also compared depending on their individual characteristics. Calculations were performed using SPSS V20.0 (IBM, Chicago, IL, USA).

## RESULTS

Qualitative results obtained in case and control samples are depicted in **Table 1**. The most prevalent VGs among patients were *eae* ( $n = 45$ ; 91.8%), *aggR* ( $n = 31$ ; 63.3%), and *invA* ( $n = 40$ ; 81.6%), while *eae* ( $n = 30$ ; 93.8%), and *aggR* ( $n = 16$ ; 50%) were the most commonly found VGs in the controls (**Table 1**). *EhxA* and *bfpA* were both moderately represented in patients ( $n = 20$ ; 40.8% and  $n = 14$ ; 28.6%) and controls ( $n = 5$ ; 15.6% and  $n = 10$ ; 31.3%), respectively. *Stx1* was infrequent in both groups ( $n = 2$ ; 4.1% in patients and  $n = 1$ ; 3.1% in controls). *Stx2* was only detected in controls ( $n = 1$ , 3.1%) while *est* was only found in patients ( $n = 1$ ; 2%). Interestingly, the *elt* gene was not found in any case or control sample. *InvA*, *ehxA*, and the simultaneous presence of *rfbO157* with *fliCH7* were significantly more prevalent

( $p < 0.05$ ) in the diarrheal patients than in healthy controls, while the opposite was true for the simultaneous detection of both *wzxO104* and *fliCH4* (**Table 1**).

No significant differences between age or sex and detection of VGs were detected in any of the two groups. Analysis of the information collected in the survey of the control population revealed that a significantly higher proportion of this group was positive for the *rfbO157* gene among those taking antibiotics in comparison with those who did not ( $p = 0.01$ ). Marginally significant differences between the presence of *ehxA* and the consumption of antibiotics was also observed (higher in those taking antibiotics,  $p = 0.057$ ). Also, proportion of controls positive to *aggR* and *bfpA* was marginally significantly higher ( $p = 0.1$  and  $p = 0.08$ , respectively) among those with travel history (**Table 2**). No significant differences in the proportion of positive samples to any other VG and the variables under study were detected.

In the quantitative analysis, only VGs that were present in at least five samples (*ehxA*, *eae*, *aggR*, *bfpA*, *rfbO157*, *fliCH7*, and *fliCH4*) were considered (**Table 3**). Within these analyses, no significant differences between the quantitative values found in each group were detected, except for the *fliCH4* gene [with a significantly ( $p < 0.007$ ) higher number of gene copies per milligram of sample in the control group]. Marginally significant differences were found for *aggR* and *invA*, present at higher quantities in the patients ( $p = 0.07$  and 0.084, respectively). In healthy controls, the maximum gene copy number for a given VG per milligram of feces was always under  $10e^{+04}$ . In contrast, for the patients, there was a maximum of  $10e^{+07}$  gene copy number. If a tentative cut-off value was established so that patients with gene copy numbers per milligram of feces  $> 10e^{+03}$  (Cabal et al., 2015) were considered positive for that VG, only four patients would be positive for VGs from pathogenic *E. coli* compared to one control. Those four patients with copy numbers above  $10e^{+03}$  for at least one VG were positive to *eae/aggR/est*, *aggR/invA*, *eae/bfpA*, and *eae*, respectively.

**TABLE 1 |** Percentage and 95% confidence intervals (CI) of the proportion of samples positive to each molecular target in patients (cases) and healthy volunteers (controls).

Molecular target	Cases ( $n = 49$ )	Controls ( $n = 32$ )	<i>p</i>
<i>stx1</i>	4.1 (1.1–13.7)	3.1 (0.55–15.7)	0.66
<i>stx2</i>	0.0 (0.0–7.27)	3.1 (0.55–15.7)	0.40
<i>eae</i>	91.8 (80.8–96.8)	93.8 (79.9–98.3)	0.56
<i>ehxA</i>	40.8 (28.2–54.8)	15.6 (6.86–31.8)	<b>0.014</b>
<i>elt</i>	0.0 (0.0–7.27)	0 (0.0–10.7)	–
<i>est</i>	2.04 (0.36–10.7)	0 (0.0–10.7)	0.61
<i>invA</i>	81.6 (68.6–90.0)	6.25 (1.73–20.1)	<b>&lt;0.001</b>
<i>bfpA</i>	28.6 (17.8–42.4)	31.3 (18.0–48.6)	0.49
<i>aggR</i>	63.3 (49.3–75.3)	50.0 (33.6–66.4)	0.17
O157+H7 <sup>a</sup>	53.1 (39.4–66.3)	18.8 (8.89–35.3)	<b>0.002</b>
O104+H4 <sup>a</sup>	2.04 (0.36–10.7)	18.8 (8.89–35.3)	<b>0.014</b>

<sup>a</sup>Indicates simultaneous detection in a given sample of genes linked to the somatic and flagellar antigens (*rfbO157*, *fliCH7*, and *wzxO104*, *fliCH4*).

Significant *p*-values are highlighted in bold.

**TABLE 2 | Characteristics of the control population and proportion of positive samples to the virulence genes (VGs) for which marginally significant ( $p < 0.1$ ) differences were found.**

Variable	Response	N (%)	<i>ehxA</i> <sup>a</sup>	<i>rfbO157</i> <sup>a</sup>	<i>bfpA</i> <sup>a</sup>	<i>aggR</i> <sup>a</sup>
Received antibiotic treatment in the last 3 months	Yes	7 (21, 9%)	42, 8%	71, 4%		
	No	25 (78, 1%)	8%	16%		
Foreign travel in the last 3 months	Yes	4 (12,5%)			75%	100%
	No	28 (87, 5%)			25%	42,8%
Gastrointestinal problems in the last 3 months	Yes	11 (34, 4%)				
	No	21 (65, 6%)				
Frequent contact with animals	Yes	27 (84, 4%)				
	No	5 (15, 6%)				
Vegetarian	Yes	1 (3, 1%)				
	No	31 (96, 9%)				
Regular consumption of probiotics	Yes	23 (71, 9%)				
	No	9 (28, 1%)				

<sup>a</sup>Percentage of positive samples among the controls that were positive to each virulence factor.

**TABLE 3 | Quantitative values (median) for VGs present in at least five samples in patients and controls.**

Gene	Patients		Controls		<i>p</i> -Value
	Median values (n) <sup>a</sup>	Samples > 1.00e <sup>+03</sup> (n = 5) <sup>b</sup>	Median values (n) <sup>a</sup>	Samples > 1.00e <sup>+03</sup> (n = 5) <sup>b</sup>	
<i>eae</i>	0.82 (45)	4	0.59 (30)	1	0.88
<i>ehxA</i>	0.22 (20)	0	0.24 (5)	0	0.92
<i>aggR</i>	1.41 (31)	2	0.87 (16)	0	0.07
<i>bfpA</i>	1.72 (14)	1	4.56 (10)	0	0.62
<i>invA</i>	43.5 (40)	1	8.04 (2)	0	0.084
<i>rfbO157</i>	0.50 (30)	0	0.84 (9)	0	0.27
<i>fliC<sub>H7</sub></i>	1.21 (42)	0	3.06 (16)	0	0.13
<i>fliC<sub>H4</sub></i>	182.4 (29)	0	14.760 (23)	0	0.007

<sup>a</sup>Median values are expressed in gene copies per milligram of feces. <sup>b</sup>Number of positive individuals when gene copy number per milligram of feces is equal or above 1.00e<sup>+03</sup>.

## DISCUSSION

In this study, we estimated the qualitative and quantitative (gene copy number per milligram of sample) abundance of VGs in the feces of two groups of individuals, healthy volunteers and patients with diarrhea. Overall, the prevalence of VGs in both groups was high. In fact, while the presence of VGs in the feces of patients with diarrhea was expected to some extent, healthy controls carried VGs as well (Stephan et al., 2000; Fujihara et al., 2009). Humans have been proven to act sometimes as mere carriers of these VGs without being necessarily related with intestinal disease (Jenkins et al., 2007) since they may need to coexist with other VGs and eventually in an adequate microorganisms to produce clinical symptoms (Rosenshine et al., 1996; Barletta et al., 2011). The high prevalence observed may be due in part to the analytical approach adopted here, since qPCR-based direct detection of DNA extracted from fecal samples may yield a higher analytic sensitivity than isolation of *E. coli* followed by PCR-detection.

The *aggR* gene was observed at high prevalence in both patients (63.3%) and controls (50%), in contrast with a similar study by Samie et al. (2007) that described higher prevalence of *aggR* in the patient group. However, the higher numbers of

gene copies in positive case samples compared to the controls, which was close to the statistical significance (Table 3), was in line with the findings of that study, suggesting that that *aggR* may contribute to the patient symptoms (Samie et al., 2007).

The *wzxO104/fliC<sub>H4</sub>* combination, recently associated with EAEC (Bielaszewska et al., 2011), was more common in healthy individuals (18.8%) than in patients (2.04%), with most of these positive control samples being positive also to *aggR* (4/6 and 66.7%), suggesting asymptomatic carriage of EAEC and/or EAEC/EHEC as previously reported (Mathewson et al., 1985; Balabanova et al., 2013).

In contrast, the prevalence and quantities of *bfpA*, which may indicate presence of tEPEC, was similar in both patients and healthy individuals, suggesting a limited pathogenic role as previously proposed (Wang et al., 2013). Interestingly, the marginally significant differences found in the controls who traveled abroad for *bfpA* but also for *aggR* may indicate that EPEC and EAEC are the pathotypes most probably acquired during international travels, as previously described (Tangden et al., 2010; Laaveri et al., 2014). Detection of other tEPEC virulence markers in *bfpA*-positive samples could help in elucidating the real role of *bfpA* in disease pathogenesis.

In the case of *eae*, its large prevalence (>90%) in *bfpA*-negative fecal samples could indicate the presence of aEPEC. In addition, the fact that *eae* frequencies were similar in both groups was in agreement with previous studies (Fujihara et al., 2009; Ghosh and Ali, 2010; Schmidt, 2010), and therefore it also questions its potential pathogenic role in our study population.

The low prevalence found for *stx*-genes was in agreement with that reported by Pradel et al. (2000), which was also similar for both patients and controls. Still, the high frequencies obtained for *eae*, *ehxA*, and *rfbO157/fliC<sub>H7</sub>* genes (particularly among case samples) may indicate the presence of O157:H7 strains despite the low prevalence obtained for the Shiga toxins. It has been shown that both EPEC and EHEC O157:H7 strains can exist in patients with HUS or diarrhea (Ferdous et al., 2015), and a whole genome analysis approach would be needed to differentiate them.

In the case of EIEC, represented in this study by *invA*, the high prevalence obtained for patients together with a higher gene copy number in this group could indicate that *invA* is directly related with pathogenicity, thus explaining its absence (or presence at very low levels) in healthy controls (Labbé and García, 2013). Bruijnesteijn van Coppenraet et al. (2015) also found EIEC-typical genes in a significantly higher proportion of cases compared with a control population, while no differences were observed for STEC-associated genes.

Lastly, detection rates for *est* and/or *elt*, usually associated with the ETEC pathotype and therefore with Traveler's diarrhea, were low among diarrheal patients, although interpretation of these results is not possible due to the lack of information on the patient's travel history (Qadri et al., 2005; Rivera et al., 2013). Among the variables included in the survey performed in the control population, we only observed a significant association between *ehxA* or *rfbO157* and antibiotic intake (Table 2). A marginally significant association between *aggR* or *bfpA* and foreign travel was seen in agreement with previous reports (Evans and Evans, 1996; Jiang et al., 2002).

Comparison of results obtained in the case and control populations obtained here must be performed carefully since, while cases were sampled randomly among those submitted to a large hospital in Madrid, our control population was severely biased toward a very specific subpopulation (veterinary students and laboratory staff) that could be exposed to certain risks (i.e., STEC of animal origin), and may thus not be representative of the general population. Interestingly, prevalence of VGs associated with STEC (i.e., *stx1*, and *stx2*) was very low in both cases and controls (Table 1), thus suggesting a non-differential degree of exposure to sources of this particular pathotype. Another

## REFERENCES

- Balabanova, Y., Klar, S., Deleré, Y., Wilking, H., Faber, M. S., Lassen, S. G., et al. (2013). Serological evidence of asymptomatic infections during *Escherichia coli* O104:H4 outbreak in Germany in 2011. *PLoS ONE* 8:e73052. doi: 10.1371/journal.pone.0073052
- Barletta, F., Ochoa, T. J., Mercado, E., Ruiz, J., Ecker, L., Lopez, G., et al. (2011). Quantitative real-time polymerase chain reaction for enteropathogenic *Escherichia coli*: a tool for investigation of asymptomatic versus symptomatic infections. *Clin. Infect. Dis.* 53, 1223–1229. doi: 10.1093/cid/cir730

limitation was the limited size of our sample; even though matching by age and sex could have helped to increase the power of the study, this information was not available when controls were recruited and was thus not possible.

In summary, our strategy allowed the assessment of the prevalence and abundance of VGs associated with diarrheagenic *E. coli* in healthy volunteers and also in patients with diarrhea. In some cases both groups were carriers of the same VGs, highlighting the fact that not all VGs might be linked with pathogenicity or may need additional association with other VG to produce diarrhea. Moreover, the VG quantification could help to determine whether a VG is crucial for the onset of the diarrhea or not and may facilitate the isolation of positive colonies. Although some individual characteristics were marginally associated with increased odds of being positive for certain VGs in the control population, the limited sample size used here prevents extracting definitive conclusions.

## AUTHOR CONTRIBUTIONS

RC, CG, LD, and JA conceived and designed the study. AC and MG-C collected the samples and performed the laboratory analyses. AC and JA analyzed the data and drafted the first version of the manuscript. All authors revised the manuscript critically and contributed to the final version.

## FUNDING

This work is a contribution to EU FP7 ANTIGONE (project number 278976). JA was a recipient of a Sara Borrell postdoctoral contract (CD11/00261, Ministerio de Ciencia e Innovación).

## ACKNOWLEDGMENTS

Authors would like to thank all people who volunteered to participate in the present study. Authors are grateful to Dr. K. VanderWaal for her help editing the final manuscript.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2016.00641>

- Bielaszewska, M., Mellmann, A., Zhang, W., Kock, R., Fruth, A., Bauwens, A., et al. (2011). Characterisation of the *Escherichia coli* strain associated with an outbreak of haemolytic uraemic syndrome in Germany, 2011: a microbiological study. *Lancet Infect. Dis.* 11, 671–676. doi: 10.1016/S1473-3099(11)70165-7
- Blanco, M., Blanco, J. E., Dahbi, G., Alonso, M. P., Mora, A., Coira, M. A., et al. (2006). Identification of two new intimin types in atypical enteropathogenic *Escherichia coli*. *Int. Microbiol.* 9, 103–110.
- Bruijnesteijn van Coppenraet, L. E., Dullaert-de Boer, M., Ruijs, G. J., van der Reijden, W. A., van der Zanden, A. G., Weel, J. F., et al. (2015). Case-control comparison of bacterial and protozoan microorganisms associated with

- gastroenteritis: application of molecular detection. *Clin. Microbiol. Infect.* 21, 592.e9–592.e19. doi: 10.1016/j.cmi.2015.02.007
- Cabal, A., Geue, L., Gomez-Barrero, S., Barth, S., Barcena, C., Hamm, K., et al. (2015). Detection of virulence-associated genes characteristic of intestinal *Escherichia coli* pathotypes, including the enterohemorrhagic/enteroaggregative O104:H4, in bovines from Germany and Spain. *Microbiol. Immunol.* 59, 433–442. doi: 10.1111/1348-0421.12275
- Cabal, A., Gomez-Barrero, S., Porrero, C., Barcena, C., Lopez, G., Canton, R., et al. (2013). Assessment of virulence factors characteristic of human *Escherichia coli* pathotypes and antimicrobial resistance in O157:H7 and non-O157:H7 isolates from livestock in Spain. *Appl. Environ. Microbiol.* 79, 4170–4172. doi: 10.1128/AEM.00537-13
- Evans, D. J., and Evans, D. G. (1996). “*Escherichia coli* in diarrheal disease,” in *Medical Microbiology*, 4th Edn, ed. S. Baron (Galveston, TX: University of Texas Medical Branch at Galveston).
- Ferdous, M., Zhou, K., Mellmann, A., Morabito, S., Croughs, P. D., De Boer, R. F., et al. (2015). Is shiga toxin-negative *Escherichia coli* O157:H7 enteropathogenic or enterohemorrhagic *Escherichia coli*? Comprehensive molecular analysis using whole-genome sequencing. *J. Clin. Microbiol.* 53, 3530–3538. doi: 10.1128/JCM.01899-15
- Fletcher, S. M., McLaw, M. L., and Ellis, J. T. (2013). Prevalence of gastrointestinal pathogens in developed and developing countries: systematic review and meta-analysis. *J. Public Health Res.* 2, 42–53. doi: 10.4081/jphr.2013.e9
- Frank, C., Werber, D., Cramer, J. P., Askar, M., Faber, M., An Der Heiden, M., et al. (2011). Epidemic profile of Shiga-toxin-producing *Escherichia coli* O104:H4 outbreak in Germany. *N. Engl. J. Med.* 365, 1771–1780. doi: 10.1056/NEJMoa1106483
- Fujihara, S., Arikawa, K., Aota, T., Tanaka, H., Nakamura, H., Wada, T., et al. (2009). Prevalence and properties of diarrheagenic *Escherichia coli* among healthy individuals in Osaka City, Japan. *Jpn. J. Infect. Dis.* 62, 318–323.
- Gascon, J., Vargas, M., Quinto, L., Corachan, M., Jimenez De Anta, M. T., and Vila, J. (1998). Enteragggregative *Escherichia coli* strains as a cause of traveler's diarrhea: a case-control study. *J. Infect. Dis.* 177, 1409–1412. doi: 10.1086/517826
- Ghosh, P. K., and Ali, A. (2010). Isolation of atypical enteropathogenic *Escherichia coli* from children with and without diarrhoea in Delhi and the National Capital Region, India. *J. Med. Microbiol.* 59, 1156–1162. doi: 10.1099/jmm.0.014530-0
- Huang, D. B., Okhuyzen, P. C., Jiang, Z. D., and Dupont, H. L. (2004). Enteragggregative *Escherichia coli*: an emerging enteric pathogen. *Am. J. Gastroenterol.* 99, 383–389. doi: 10.1099/jmm.0.014530-0
- Jenkins, C., Chart, H., Willshaw, G. A., Cheasty, T., and Tompkins, D. S. (2007). Association of putative pathogenicity genes with adherence characteristics and fimbrial genotypes in typical enteragggregative *Escherichia coli* from patients with and without diarrhoea in the United Kingdom. *Eur. J. Clin. Microbiol. Infect. Dis.* 26, 901–906. doi: 10.1007/s10096-007-0388-z
- Jiang, Z. D., Greenberg, D., Nataro, J. P., Steffen, R., and Dupont, H. L. (2002). Rate of occurrence and pathogenic effect of enteragggregative *Escherichia coli* virulence factors in international travelers. *J. Clin. Microbiol.* 40, 4185–4190. doi: 10.1128/JCM.40.11.4185-4190.2002
- Johnson, T. J., and Nolan, L. K. (2009). Pathogenomics of the virulence plasmids of *Escherichia coli*. *Microbiol. Mol. Biol. Rev.* 73, 750–774. doi: 10.1128/MMBR.00015-09
- Kaper, J. B. (2005). Pathogenic *Escherichia coli*. *Int. J. Med. Microbiol.* 295, 355–356. doi: 10.1016/j.ijmm.2005.06.008
- Kaper, J. B., Nataro, J. P., and Moseley, H. L. (2004). Pathogenic *Escherichia coli*. *Nat. Rev. Microbiol.* 2, 123–140. doi: 10.1038/nrmicro818
- Laaveri, T., Pakkanen, S. H., Antikainen, J., Riutta, J., Mero, S., Kirveskari, J., et al. (2014). High number of diarrhoeal co-infections in travellers to Benin, West Africa. *BMC Infect. Dis.* 14:81. doi: 10.1186/1471-2334-14-81
- Labbé, R. G., and García, S. (2013). *Guide to Foodborne Pathogens*. Chichester: John Wiley & Sons Inc.
- Mathewson, J. J., Johnson, P. C., Dupont, H. L., Morgan, D. R., Thornton, S. A., Wood, L. V., et al. (1985). A newly recognized cause of travelers' diarrhea: enteroadherent *Escherichia coli*. *J. Infect. Dis.* 151, 471–475. doi: 10.1093/infdis/151.3.471
- Mora, A., Herrerra, A., Lopez, C., Dahbi, G., Mamani, R., Pita, J. M., et al. (2011). Characteristics of the Shiga-toxin-producing enteragggregative *Escherichia coli* O104:H4 German outbreak strain and of STEC strains isolated in Spain. *Int. Microbiol.* 14, 121–141. doi: 10.2436/20.1501.01.142
- Nataro, J. P., and Kaper, J. B. (1998). Diarrheagenic *Escherichia coli*. *Clin. Microbiol. Rev.* 11, 142–201.
- Ochman, H., Lawrence, J. G., and Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* 405, 299–304. doi: 10.1038/35012500
- Paton, J. C., and Paton, A. W. (1998). Pathogenesis and diagnosis of Shiga toxin-producing *Escherichia coli* infections. *Clin. Microbiol. Rev.* 11, 450–479.
- Pradel, N., Livrelli, V., De Champs, C., Palcoux, J. B., Reynaud, A., Scheutz, F., et al. (2000). Prevalence and characterization of Shiga toxin-producing *Escherichia coli* isolated from cattle, food, and children during a one-year prospective study in France. *J. Clin. Microbiol.* 38, 1023–1031.
- Qadri, F., Svenssonholm, A. M., Faruque, A. S., and Sack, R. B. (2005). Enterotoxigenic *Escherichia coli* in developing countries: epidemiology, microbiology, clinical features, treatment, and prevention. *Clin. Microbiol. Rev.* 18, 465–483. doi: 10.1128/CMR.18.3.465-483.2005
- Rivera, F. P., Medina, A. M., Aldasoro, E., Sangil, A., Gascon, J., Ochoa, T. J., et al. (2013). Genotypic characterization of enterotoxigenic *Escherichia coli* strains causing traveler's diarrhea. *J. Clin. Microbiol.* 51, 633–635. doi: 10.1128/JCM.02572-12
- Rosenshine, I., Ruschikowski, S., and Finlay, B. B. (1996). Expression of attaching/effacing activity by enteropathogenic *Escherichia coli* depends on growth phase, temperature, and protein synthesis upon contact with epithelial cells. *Infect. Immun.* 64, 966–973.
- Samie, A., Obi, C. L., Dillingham, R., Pinkerton, R. C., and Guerrant, R. L. (2007). Enteragggregative *Escherichia coli* in Venda, South Africa: distribution of virulence-related genes by multiplex polymerase chain reaction in stool samples of human immunodeficiency virus (HIV)-positive and HIV-negative individuals and primary school children. *Am. J. Trop. Med. Hyg.* 77, 142–150.
- Schmid, M. A. (2010). LEEways: tales of EPEC, ATEC and EHEC. *Cell. Microbiol.* 12, 1544–1552. doi: 10.1111/j.1462-5822.2010.01518.x
- Stephan, R., Ragettli, S., and Untermann, F. (2000). Prevalence and characteristics of verotoxin-producing *Escherichia coli* (VTEC) in stool samples from asymptomatic human carriers working in the meat processing industry in Switzerland. *J. Appl. Microbiol.* 88, 335–341. doi: 10.1046/j.1365-2672.2000.00965.x
- Tangden, T., Cars, O., Melhus, A., and Lowdin, E. (2010). Foreign travel is a major risk factor for colonization with *Escherichia coli* producing CTX-M-type extended-spectrum beta-lactamases: a prospective study with Swedish volunteers. *Antimicrob. Agents Chemother.* 54, 3564–3568. doi: 10.1128/AAC.00220-10
- Trabulsi, L. R., Keller, R., and Tardelli Gomes, T. A. (2002). Typical and atypical enteropathogenic *Escherichia coli*. *Emerg. Infect. Dis.* 8, 508–513. doi: 10.3201/eid0805.010385
- Wang, L., Wakushima, M., Aota, T., Yoshida, Y., Kita, T., Maehara, T., et al. (2013). Specific properties of enteropathogenic *Escherichia coli* isolates from diarrheal patients and comparison to strains from foods and fecal specimens from cattle, swine, and healthy carriers in Osaka City, Japan. *Appl. Environ. Microbiol.* 79, 1232–1240. doi: 10.1128/AEM.03380-12

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Cabal, García-Castillo, Cantón, Gortázar, Domínguez and Álvarez. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Phenotypic and Genetic Heterogeneity in *Vibrio cholerae* O139 Isolated from Cholera Cases in Delhi, India during 2001–2006

Raikamal Ghosh<sup>1</sup>, Naresh C. Sharma<sup>2</sup>, Kalpataru Halder<sup>3</sup>, Rupak K. Bhadra<sup>3</sup>, Goutam Chowdhury<sup>1</sup>, Gururaja P. Pazhani<sup>1</sup>, Sumio Shinoda<sup>4</sup>, Asish K. Mukhopadhyay<sup>1</sup>, G. Balakrish Nair<sup>5</sup> and Thadavarayan Ramamurthy<sup>5\*</sup>

<sup>1</sup> Division of Bacteriology, National Institute of Cholera and Enteric Diseases, Kolkata, India, <sup>2</sup> Maharishi Valmiki Infectious Diseases Hospital, Delhi, India, <sup>3</sup> Infectious Diseases and Immunology Division, Council of Scientific and Industrial Research-Indian Institute of Chemical Biology, Kolkata, India, <sup>4</sup> Collaborative Research Center of Okayama University for Infectious Diseases in India, National Institute of Cholera and Enteric Diseases, Kolkata, India, <sup>5</sup> Center for Human Microbial Ecology, Translational Health Science and Technology Institute, Faridabad, India

## OPEN ACCESS

### Edited by:

Andres M. Perez,  
University of Minnesota, USA

### Reviewed by:

Paras Jain,  
Albert Einstein College of Medicine,  
USA

Daniela Ceccarelli,  
Wageningen University and Research  
Centre, Netherlands

Christopher John Grim,  
United States Food and Drug  
Administration, USA

### \*Correspondence:

Thadavarayan Ramamurthy  
tramu@thsti.res.in

### Specialty section:

This article was submitted to  
Infectious Diseases,  
a section of the journal  
*Frontiers in Microbiology*

Received: 01 October 2015

Accepted: 27 July 2016

Published: 09 August 2016

### Citation:

Ghosh R, Sharma NC, Halder K, Bhadra RK, Chowdhury G, Pazhani GP, Shinoda S, Mukhopadhyay AK, Nair GB and Ramamurthy T (2016) Phenotypic and Genetic Heterogeneity in *Vibrio cholerae* O139 Isolated from Cholera Cases in Delhi, India during 2001–2006. *Front. Microbiol.* 7:1250. doi: 10.3389/fmicb.2016.01250

Incidence of epidemic *Vibrio cholerae* serogroup O139 has declined in cholera endemic countries. However, sporadic cholera caused by *V. cholerae* O139 with notable genetic changes is still reported from many regions. In the present study, 42 *V. cholerae* O139 strains isolated from 2001 to 2006 in Delhi, India, were retrospectively analyzed to understand their phenotype and molecular characteristics. The majority of isolates were resistant to ampicillin, furazolidone and nalidixic acid. Though the integrative conjugative element was detected in all the O139 isolates, the 2004–2006 isolates remained susceptible to co-trimoxazole, chloramphenicol, and streptomycin. Cholera toxin genotype 1 was present in the majority of the O139 isolates while few had type 3 or a novel type 4. In the cholera toxin encoding gene (*ctx*) restriction fragment length polymorphism, the majority of the isolates harbored three copies of CTX element, of which one was truncated. In this study, the *ctx* was detected for the first time in the small chromosome of *V. cholerae* O139 and one isolate harbored 5 copies of CTX element, of which 3 were truncated. The ribotype BII pattern was found in most of the O139 isolates. Three *V. cholerae* O139 isolated in 2001 had a new ribotype BVIII. Pulsed-field gel electrophoresis analysis revealed clonal variation in 2001 isolates compared to the 2004–2006 isolates. Molecular changes in *V. cholerae* O139 have to be closely monitored as this information may help in understanding the changing genetic features of this pathogen in relation to the epidemiology of cholera.

**Keywords:** *V. cholerae* O139, ribotypes, CT genotype, CTX prophage, PFGE

## INTRODUCTION

The aquatic bacterium *Vibrio cholerae* is the causative agent of cholera or cholera-like diarrhea in humans. Of the 206 serogroups identified in this species (Yamai et al., 1997), the serogroups O1 and O139 are responsible for global cholera epidemics. *V. cholerae* serogroup O1 is further divided into two biotypes, classical and El Tor and each has two distinct serotypes, Inaba and Ogawa. The

classical biotype was associated with cholera in first six pandemics (Sack et al., 2004). The current 7th cholera pandemic is represented by *V. cholerae* O1 El Tor biotype, which became dominant from 1961 and gradually replaced the classical biotype from the global cholera scenario. *V. cholerae* O139 serogroup emerged in 1992 by replacing the El Tor biotype in the Indian subcontinent and spread to more than 14 countries in the following years (Nair et al., 1994a; Siddique et al., 1996; Ramamurthy et al., 2003). Emergence of *V. cholerae* O139 serogroup was thought to be the beginning of the 8th cholera pandemic considering the rapid spread of the pathogen (Nair et al., 1994b). However, after causing large cholera epidemics in 1993, the serogroup O139 disappeared abruptly from the endemic scenario ensuing resurgence of *V. cholerae* O1 El Tor biotype in cholera endemic regions (Sharma et al., 1997). Until late 1999, there has been periodic shift between El Tor and O139 in India and Bangladesh (Basu et al., 2000; Faruque et al., 2003a). In 2008, the incidence of *V. cholerae* O139 in China was 32% among cholera cases (WHO, 2009) and continued until 2012 (Zhang et al., 2014).

In *V. cholerae* O139, changes in the antimicrobial susceptibility patterns and arrangement of genetic elements, especially the organization of ribosomal RNA operons, location, and arrangement of cholera toxin prophages (CTXΦ) were reported during its emergence on several occasions (Sharma et al., 1997; Faruque et al., 2003a; Nandi et al., 2003; Chatterjee et al., 2007; Ghosh et al., 2008). Initial genetic analysis showed that emergence of *V. cholerae* O139 may be due to the insertion of a novel 35-kb *wbf* gene that encodes O139-somatic (O) antigen in a *V. cholerae* serogroup O22 strain or due to the loss of a 22-kb *wbe* region in a *V. cholerae* O1 that encodes the O1 antigen (Yamasaki et al., 1999). The whole genome sequence analysis by Chun et al. (2009) confirmed the above finding, i.e., substitution of the gene cluster coding for the O139 antigen took place by horizontal gene transfer but not the deletion.

Based on the amino acid changes, the B-subunits of CT have been designated into several CT-genotypes or *ctxB* alleles (Safa et al., 2008; Raychoudhuri et al., 2009). CT genotyping (*ctxB* allele) can be made using Mismatch amplification mutation assay (MAMA) PCR (Morita et al., 2008). CT genotype 1 is reported in strains of the classical biotype worldwide and in US Gulf Coast, genotype 2 is found in El Tor biotype strains from Australia, and genotype 3 is prevalent in El Tor biotype from the 7th pandemic and the Latin American epidemic strains (Olsvik et al., 1993). *V. cholerae* O1 El Tor isolates that produces classical CT is a newly emerged trait, which is said to be associated with the severity of the illness (Siddique et al., 2010) with a large number of cholera outbreaks (Nair et al., 2006; Safa et al., 2008; Raychoudhuri et al., 2009). CT encoding genes of O1 and O139 serogroups is carried by a filamentous CTXΦ, which is known to use the toxin-coregulated pili (TCP) as its receptor (Waldor and Mekalanos, 1996). *V. cholerae* O139 harboring CTX<sup>class</sup>Φ and CTX<sup>calc</sup>Φ has been described based on the difference in the sequence of *rstR* that encodes for the repressor protein of the CTXΦ (Faruque et al., 2003a; Bhattacharya et al., 2006; Raychoudhuri et al., 2010).

This study was undertaken to understand the phenotype and genetic changes of *V. cholerae* O139 isolated from sporadic

hospitalized cholera cases in Delhi during 2001–2006. The outcome of this study may be useful to comprehend the epidemiology of *V. cholerae* O139.

## MATERIALS AND METHODS

### Bacterial Strains

*V. cholerae* O139 was isolated from cholera patients admitted at the Maharishi Valmiki Infectious Diseases Hospital, Delhi. Between 2001 and 2006, 42 isolates individually isolated strains were included in this study (Table 1). *V. cholerae* O1 569B (classical biotype), N16961 (El Tor biotype), and SG-24 (serogroup O139) were used as reference strains. In the pulsed-field gel electrophoresis (PFGE), *Salmonella enterica* serotype Braenderup strain H9812 was used as the molecular size standard.

### Bacteriology and Serotyping

*V. cholerae* isolates were grown on thiosulphate-citrate-bile salt-sucrose (TCBS) agar (Eiken, Tokyo, Japan) at 37°C for 16–18 h. Typical sucrose fermenting yellow colonies were further streaked on Luria agar (LA, Difco, Detroit, MD, USA) and subsequently used in the rapid biochemical identification (Nair et al., 1987). Presumptively identified *V. cholerae* isolates were further confirmed by oxidase test and confirmed serologically by slide agglutination test using O1 and O139 monoclonal antibodies prepared at the National Institute of Cholera and Enteric Diseases, Kolkata, India (Garg et al., 1994; Ramamurthy et al., 1995).

### Antimicrobial Susceptibility

Antimicrobial susceptibility testing was performed using commercially available disks (Difco) following the Clinical and Laboratory Standard Institute guidelines (CLSI, 2014). The concentration of antibiotics in the disc was as follows: ampicillin (10 µg), chloramphenicol (30 µg), co-trimoxazole (sulfamethoxazole/trimethoprim, 1.25/23.45 µg), ciprofloxacin (5 µg), furazolidone (100 µg), norfloxacin (10 µg), gentamycin (10 µg), nalidixic acid (30 µg), neomycin (30 µg), streptomycin (10 µg), tetracycline (30 µg), and erythromycin (15 µg). Except for furazolidone, the minimal inhibitory concentrations (MICs) of antibiotics (ampicillin, chloramphenicol, erythromycin, nalidixic acid, streptomycin, sulfamethoxazole/trimethoprim) were determined by E-test (AB bioMérieux, Solna, Sweden).

### Extraction of Chromosomal DNA

Modified method of Murray and Thompson (1980) was used for *V. cholerae* genomic DNA extraction.

### Polymerase Chain Reaction (PCR) Assay

Multiplex PCRs were used for the detection of *rfl* genes encoding the somatic antigen of O139/O1, CT encoding gene (*ctxA*), and biotypes based on the allelic difference in the *tcpA* gene (Keasler and Hall, 1993; Hoshino et al., 1998). Simplex PCR assays with specific primers were made for the detection of *rstR* alleles (Bhattacharya et al., 2006). MAMA-PCR was made to detect the presence of *ctxB* alleles (CT genotypes) as described

**TABLE 1 | Phenotypic and genetic characteristics of *V. cholerae* O139 isolates.**

Isolate	Year	<i>rstR</i>	<i>rstR</i>	<i>rstR</i>	<i>ctxB*</i>	<i>ctxB*</i>	Ribotype	<i>ctx</i> copy	Antibiogram		MIC ( $\mu\text{g/ml}$ )				
		classical	El Tor	cal	El Tor	classical			A	Na	S	Co	C	E	
2	2001	–	+	–	–	+	B1	1	ACoS	3	–	64	4	–	–
4	2001	–	+	–	–	+	B1	1	ACoFzS	4	–	128	4	–	–
21	2001	–	+	–	–	+	BII	1	Fz	–	–	–	–	–	–
36	2001	–	+	–	–	+	B1	1	CCoS	–	–	64	12	2	–
37	2001	+	+	+	+	+	BVIII	ND	AFzNaS	3	>256	128	–	–	–
46	2001	+	+	+	+	+	BVIII	2 + 3 TRN	AFzNaSE	4	>256	>256	–	–	1.5
103	2001	+	+	+	+	+	BVIII	ND	NaS	–	>256	64	–	–	–
174	2001	–	+	–	–	+	BII	1	ACFzS	3	–	64	–	4	–
262	2001	–	+	–	–	+	BII	1	AFzS	4	–	128	–	–	–
274	2001	–	+	–	–	+	B1	1	ACFzS	4	–	>256	–	4	–
280	2001	–	+	–	–	+	BII	ND	AFzS	3	–	>256	–	–	–
X	2001	–	+	+	+	+	BII	ND	ACCoFzNaS	3	16	64	4	4	–
3653	2004	–	+	+	–	+	BII	2 + 1TRN	AFzNa	4	>256	–	–	–	–
3686	2004	–	+	+	–	+	BII	2 + 1TRN	AFzNa	6	>256	–	–	–	–
3705	2004	–	+	+	–	+	BII	2 + 1TRN	AFzNa	6	16	–	–	–	–
3710	2004	–	+	+	–	+	BII	2 + 1TRN	AFzNa	4	>256	–	–	–	–
3711	2004	–	+	+	–	+	BII	2 + 1TRN	AFzNaE	12	16	–	–	–	1
3712	2004	–	+	+	–	+	BII	2 + 1TRN	AFzNa	6	>256	–	–	–	–
3719	2004	–	+	+	–	+	BII	2 + 1TRN	FzNa	–	16	–	–	–	–
3722	2004	–	+	+	–	+	BII	2 + 1TRN	AFzNa	4	16	–	–	–	–
3736	2004	–	+	+	–	+	BII	2 + 1TRN	AFzNa	6	16	–	–	–	–
3784	2004	–	+	+	–	+	BII	2 + 1TRN	AFzNa	4	>256	–	–	–	–
3786	2004	–	+	+	–	+	BII	ND	AFzNa	6	>256	–	–	–	–
3791	2004	–	+	+	–	+	BII	ND	AFzNa	4	16	–	–	–	–
3795	2004	–	+	+	–	+	BII	2 + 1TRN	AFzNa	4	12	–	–	–	–
3796	2004	–	+	+	–	+	BII	2 + 1TRN	ANa	4	>256	–	–	–	–
3799	2004	–	+	+	–	+	BII	2 + 1TRN	AFzNa	4	16	–	–	–	–
3822	2004	–	+	+	–	+	BII	ND	FzNa	–	16	–	–	–	–
3848	2004	–	+	+	–	+	BII	2 + 1TRN	AFzNa	4	8	–	–	–	–
8/15	2004	–	+	+	–	+	BII	ND	NaE	–	16	–	–	–	1.5
24/6	2004	–	+	+	–	+	BII	ND	ANa	4	>256	–	–	–	–
12/17	2004	–	+	–	–	+	BII	ND	A	4	–	–	–	–	–
OS-227	2004	–	+	+	–	+	BII	2 + 1TRN	FzNa	–	64	–	–	–	–
5037/05	2005	–	+	+	–	+	BII	2 + 1TRN	AFzNa	4	24	–	–	–	–
130/06	2006	–	+	+	–	+	BII	2 + 1TRN	FzNa	–	16	–	–	–	–
4602/06	2006	–	+	+	–	+	BII	2 + 1TRN	AFzNa	6	>256	–	–	–	–
5340/06	2006	–	+	+	–	+	BII	2 + 1TRN	AFzNa	8	>256	–	–	–	–
5801/06	2006	–	+	+	–	+	BII	2 + 1TRN	AFzNa	6	24	–	–	–	–
5932/06	2006	–	+	+	–	+	BII	2 + 1TRN	AFzNa	4	>256	–	–	–	–
6080/06	2006	–	+	+	–	+	BII	2 + 1TRN	AFzNa	6	16	–	–	–	–
6120/06	2006	–	+	+	–	+	BII	2 + 1TRN	AFzNa	3	64	–	–	–	–
6127/06	2006	–	+	+	–	+	Bl	2 + 1TRN	AFzNa	4	3	–	–	–	–

\*As identified by MAMA-PCR. Abbreviations; ND, not done; TRN, truncated gene; A, ampicillin; C, chloramphenicol; Co, co-trimoxazole; Fz, furazolidone; E, erythromycin; Na, nalidixic acid; S, streptomycin.

All the isolates had ICE, floR, str, and dfr genes are present in the respective chloramphenicol, streptomycin, co-trimoxazole resistant *V. cholerae* O139 isolates of 2001.

previously (Morita et al., 2008). Location of CTX prophage in chromosome II was confirmed by PCR using published methods (Maiti et al., 2006). To confirm the presence of integrative conjugative element (ICE) that carries the SXT element, two sets of primers were used in this study. Primers 10SF13 (5'-TTGTGGTGGAAAGAGGGTG-3'), SXT-13 (5'-CCAACAAAG AACAGTTGACTC-3'), and ORF-16 (5'-CATCTACCACTT CATAGGCAGG-3'), YND-2 (5'-CAGCTTAACTCACCAAGG AC-3') were designed using conserved right and left terminal ends of the ICE, respectively. In addition, *floR*, *str*, and *dfr* genes encoding chloramphenicol, streptomycin, co-trimoxazole resistance was identified using published methods (Hochhut et al., 2001). In these PCRs, *V. cholerae* 569B, N16961, and SG-24 were used as reference strains. PCR assays were performed using an automated thermocycler (Gene Amp PCR system 9700, Applied Biosystems, Foster City, CA).

## DNA Sequencing

The 460 bp region of *ctxB* gene was amplified by PCR from eight representative isolates of *V. cholerae* O139 covering all the years (Olsvik et al., 1993). The amplified product was purified using a PCR purification kit (Qiagen, Hilden, Germany) and used directly as a template for nucleotide sequencing. Both the strands of DNA were sequenced with BigDye terminator cycle sequencing kit using an automated sequencer ABI 3700 (Applied Biosystems). The nucleotide and amino acid sequences were compared with the sequences available in the GenBank. The nucleotide sequence data generated with five representative isolates of *V. cholerae* O139 were submitted to the GenBank with accession numbers from GQ892075 to GQ892079.

## Ribotyping

A 7.5-kb *BamH*1 (Fermentas, Waltham, MA, USA) fragment of plasmid pKK3535 containing the 16S and 23S rRNA genes of *Escherichia coli* was used as a rRNA probe (Brosius et al., 1981). Standard *V. cholerae* ribotyping was followed in this study (Faruque et al., 2000). Instead of radioisotope, we used chemiluminescent dye (Gene Images Alkaphos direct labeling and detection system, Amersham Biosciences, UK) in the DNA hybridization analysis.

## *ctxA* RFLP

Restriction enzymes *Hind*III, *Pst*I, and *Bgl*II (Fermentas) were used for the digestion of *V. cholerae* O139 chromosomal DNA and immobilized on nylon membranes (Amersham International). The CT encoding gene (*ctxA*) probe was a 540-bp *Xba*I-*Cla*I (Fermentas) fragment cloned into the plasmid pKTN901 using *Eco*R1 linkers (Kaper et al., 1988). The 267-bp *cep* probe was derived from *Eco*R1 (Fermentas) digested pSC01 plasmid.

## Pulsed-Field Gel Electrophoresis (PFGE)

PFGE of *V. cholerae* O139 was performed as described previously for *V. cholerae* O1 (Cooper et al., 2006). PFGE profiles were analyzed using the BioNumerics version 4.0 software (Applied Maths, Sint Martens Latem, Belgium). The tagged image file formats were normalized by using the universal *S. enterica* serotype Braenderup (H9812) size standard on each gel against the reference in the database. In the dendrogram analysis, the PFGE profiles were matched using the Dice coefficient and unweighted pair group method using arithmetic averages (UPGMA). Clustering of PFGE profiles was made using 1.5% band position tolerance window and 1.5% optimization.

## RESULTS

### Identification

Conventional serology and multiplex PCRs employed in this study confirmed all the isolates as *V. cholerae* O139.

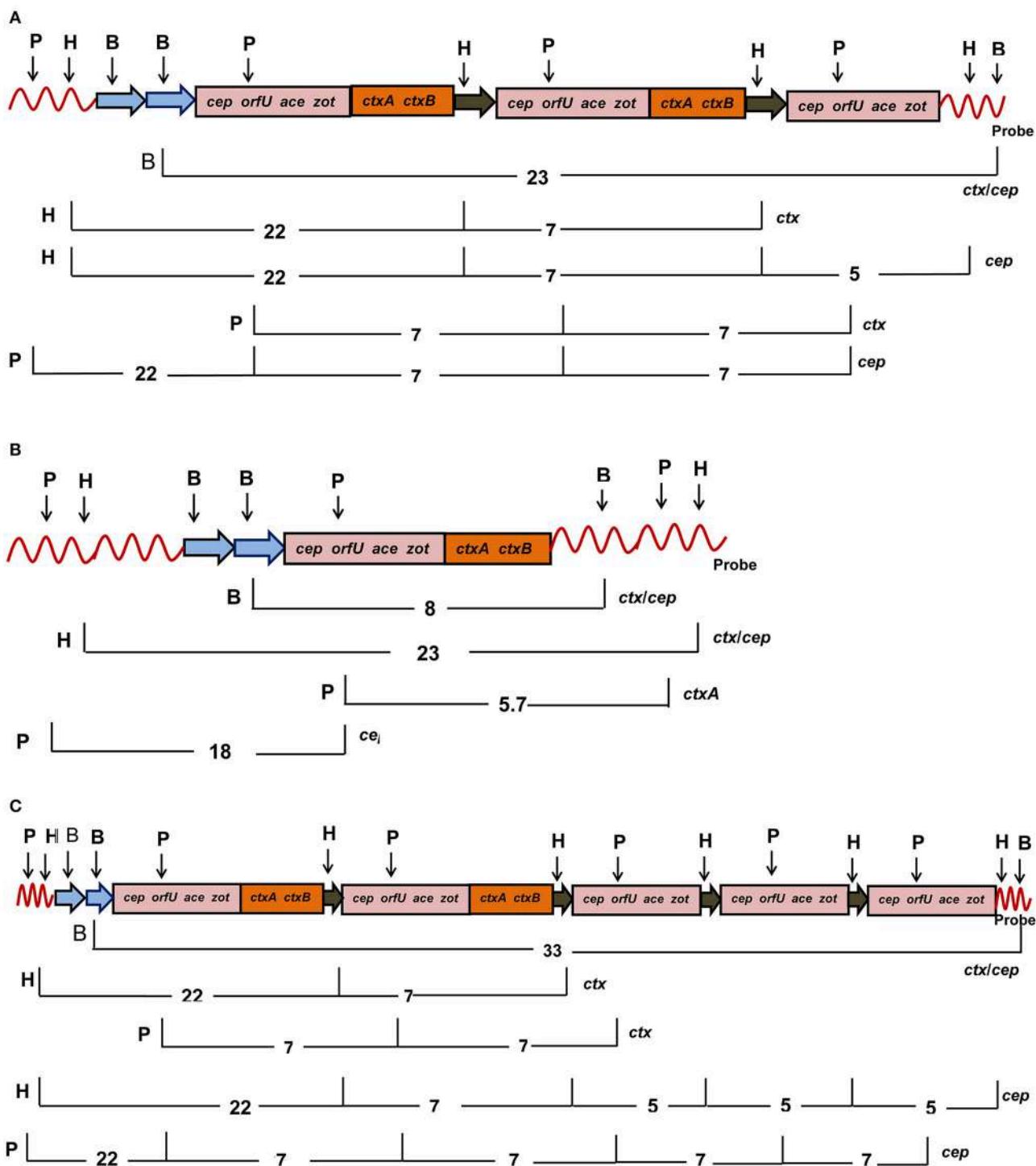
### Antimicrobial Susceptibility

In the antimicrobial susceptibility testing by disc diffusion assay, more than 60% of the *V. cholerae* O139 isolates were resistant to ampicillin, furazolidone, and nalidixic acid displaying the antibiogram as AFzNa (Table 1). The susceptibility pattern of *V. cholerae* O139 isolated during 2001 differed from the rest of the study period by displaying resistance to chloramphenicol, co-trimoxazole, and streptomycin. During the same year, 66% of the isolates were susceptible to nalidixic acid. However, in the subsequent years (2004–2006), all most all the isolates were resistant to ampicillin, furazolidone, and nalidixic acid (Table 1). For neomycin, 23 isolates showed reduced susceptibility and 19

**TABLE 2 |** CT genotypes of *V. cholerae* O139 based on DNA sequences of *ctxB*.

<i>V. cholerae</i> [Isolate No.] (year of isolation)	No. of isolates	Nucleotide at position				Amino acid at position			CT genotype
		83	115	138	203	28	39	46	
<i>V. cholerae</i> O1 [569B], classical	1	A	C	T	C	D	H	F	T
<i>V. cholerae</i> O1 El Tor, [N16961]	1	A	T	T	T	D	Y	F	I
<i>V. cholerae</i> O139 [37, 46, 103] (2001)	3	C	C	T	C	A	H	F	T
<i>V. cholerae</i> O139 [3722] (2004)	1	C	C	T	C	A	H	F	T
<i>V. cholerae</i> O139 [3705] (2004)	1	A	C	T	C	D	H	F	T
<i>V. cholerae</i> O139 [5037] (2005)	1	A	C	T	C	D	H	F	T
<i>V. cholerae</i> O139 [6080, 6127] (2006)	2	C	C	T	C	A	H	F	T

\*New CT genotype.



**FIGURE 1 |** Mapping of CTX genetic element consisting of *cep*, *orfU*, *ace*, *zot*, and *ctxAB* genes in *V. cholerae* O139 after digestion of chromosomal DNA with *Hind*III (H), *Pst*I (P), *Bgl*II (B), and probed with *ctxA*. **(A)** Eighteen *V. cholerae* O139 isolates had two intact copies (pink and brown) and one truncated CTX elements (pink). **(B)** Six *V. cholerae* O139 isolates harbored single intact copy of the CTX element (pink and brown). **(C)** The isolate 46 had two intact copies (pink and brown) and three copies of the truncated (pink) CTX elements. Restriction sites are abbreviated as B, H, and P for *Bgl*II, *Hind*III, and *Pst*I, respectively. List of probes used for hybridization was given at the right hand side. The number denotes length of the DNA fragments in Kb after digestion with the marked restriction enzymes on the left hand side.

remained susceptible (data not shown). The MIC values varied considerably for ampicillin (4–12 µg/ml), co-trimoxazole (4–12 µg/ml), nalidixic acid (3 to >256 µg/ml), and streptomycin (64 to >256 µg/ml). MIC for chloramphenicol (2–4 µg/ml) and erythromycin (1–1.5 µg/ml) remained low (**Table 1**).

## Analysis of Virulence Loci, ICE and Antimicrobial Resistance Encoding Genes

The O139 isolates uniformly harbored *ctxA* with an El Tor allele of *tcpA*. In the MAMA-PCR, all the isolates were identified as CT genotype 1. In addition, four isolates (37, 46, 103, and X) collected in 2001 exhibited CT genotype 3 (**Table 1**). The amplified *ctxB* gene from eight isolates was directly sequenced. The deduced amino acid sequence analysis identified heterogeneity in the B subunit of CT. Some of the 2004 and 2005 isolates had aspartic acid (D), histidine (H), phenylalanine (F), and threonine (T) at positions 28, 39, 46, and 68, respectively in the CtxB, which is similar to the CT genotype 1 of the *V. cholerae* O1 classical 569B strain (**Table 2**). However, the isolates representing 2001, 2004, and 2006 had amino acids alanine (A), H, F, T at positions 28, 39, 46, 68, respectively, which has been classified as CT genotype 4. This genotype was described in our previous report as genotype 5 with *V. cholerae* O139 isolates from Bangladesh (Bhuiyan et al., 2009). Subsequently, this was corrected in our publication in 2010 (Raychoudhuri et al., 2010).

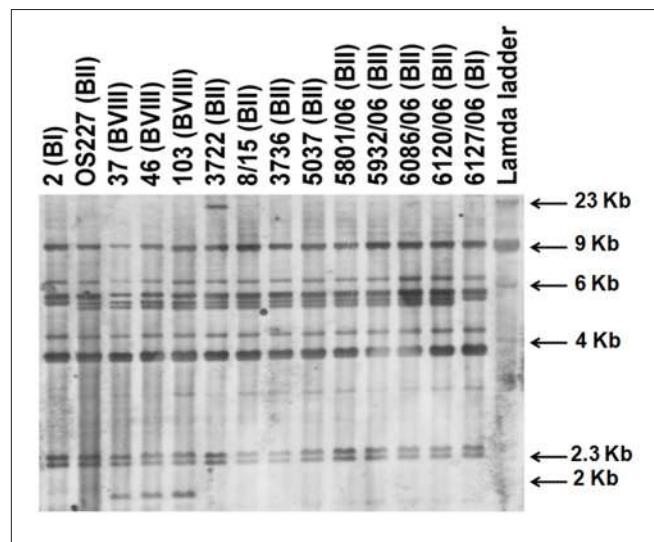
About 80% of the isolates possessed more than one allele of *rstR*, one being the El Tor type (*rstR<sup>ET</sup>*) and the other with *rstR<sup>calc</sup>* type. Interestingly, three 2001 isolates (37, 46, and 103) carried all the three *rstR* alleles, i.e., *rstR<sup>Cl</sup>*, *rstR<sup>ET</sup>*, and *rstR<sup>Calc</sup>*. These isolates belonged to a new ribotype (**Table 1**). ICE was present in all the *V. cholerae* O139 isolates as confirmed by two sets of primers. *V. cholerae* O139 isolated in 2001 that were resistant to chloramphenicol, streptomycin and co-trimoxazole respectively harbored *floR*, *str*, and *dfr* genes.

## *ctxA* RFLP

Twenty four *V. cholerae* O139 isolated during 2004–2006 displayed two tandemly arranged copies of intact CTX prophages with *cep*, *orfU*, *ace*, *zot*, and *ctxAB* as a 23 Kb fragment (**Figure 1A**). These CTX prophages were closely bordered with a 5 Kb truncated prophage (without *ctxAB*) as detected by the *cep* probe (**Table 1, Figure 1A**). Seven *V. cholerae* O139 isolated in 2001 had a single copy of CTX prophage as detected by 8 Kb *ctx/cep* probes (**Table 1, Figure 1B**). One isolate harbored two entire copies of CTX prophages as detected by *ctx* probe along with 3 truncated phages that were detected as three 5 Kb fragments by *cep* probe (**Table 1, Figure 1C**). Mapping could not be accomplished for 10 isolates with the applied strategy in this study.

## Chromosomal Location of CTX Prophages

Three of the 2001 isolates (37, 46, and 103) carried CTX prophages on both the chromosomes, which were confirmed by PCR with specific primers for chromosome I and II of *V. cholerae* (Maiti et al., 2006). In the rest of the *V. cholerae* O139 isolates, the CTX prophages remained in chromosome 1. To our knowledge,



**FIGURE 2 |** Ribotyping of *V. cholerae* O139 isolates after digestion of chromosomal DNA with *Bgl*I. Letter in parentheses against each isolate indicate the ribotype number.

this is the first report indicating the presence of CTX prophages on chromosome II in *V. cholerae* O139.

## Ribotyping

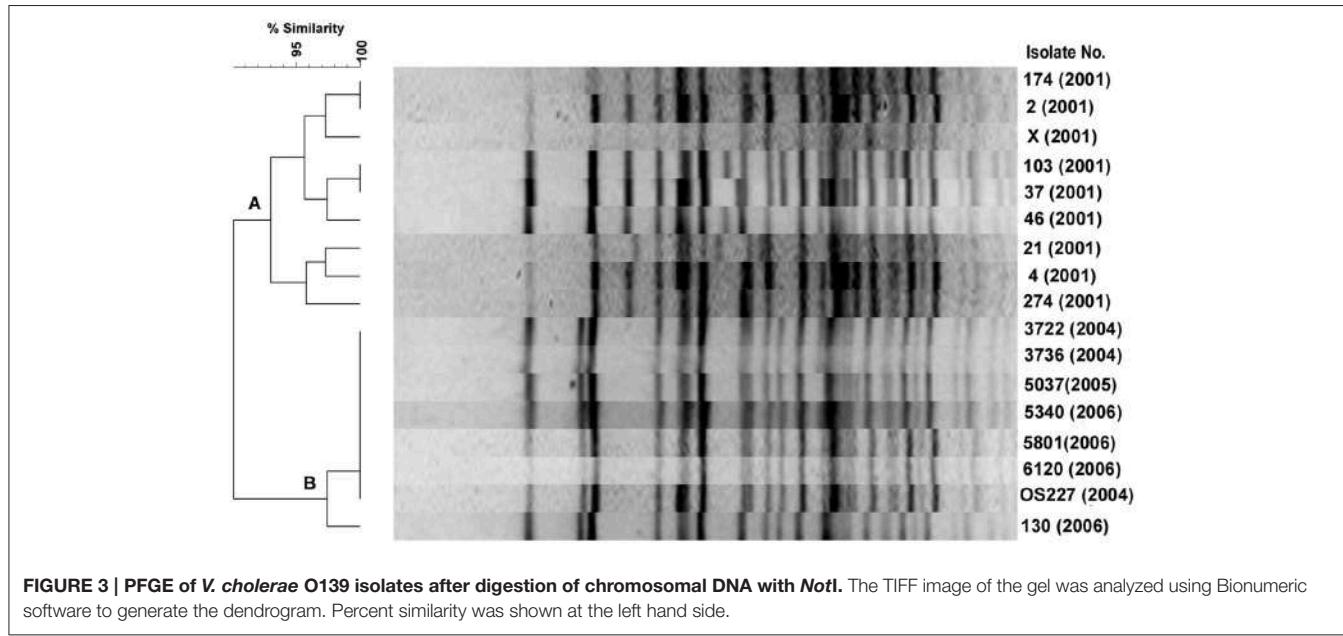
*V. cholerae* O139 isolates exhibited three different ribotypes (**Table 1, Figure 2**). Ribotype BII was predominant in 34 isolates, while 5 isolates exhibited BI ribotype. All the isolates of 2004–2005 exhibited ribotype BII pattern (**Table 1**). Ribotype patterns of 2001 isolates had mixture of BI (with 4 isolates) and BII (with 5 isolates). Interestingly, three isolates (37, 46, and 103) identified in 2001 exhibited a new ribotype pattern (**Table 1, Figure 2**). These three isolates had an extra DNA band around the 2-Kb region (**Figure 2**). This could be the new ribotype BVIII of *V. cholerae* O139.

## Pulsed-Field Gel Electrophoresis (PFGE)

Among the 9 2001 isolates, 8 different PFGE profiles were identified demonstrating the diversity of their genomes (**Figure 3**, cluster A). However, 3 isolates of 2001 belongs to ribotype BVIII were closely related in the PFGE. *V. cholerae* O139 isolated during 2004–2006 had similar PFGE profiles (**Figure 3**, cluster B), but diverged from the other isolates of 2001. A consistent correlation existed in both ribotyping and PFGE methods as most of the isolates having BII ribotype pattern were placed in clusters B. In addition, the dendrogram displayed subtypes among *V. cholerae* O139 isolates with the BII and BVIII ribotypes at about 97% similarity level (**Figure 3**).

## DISCUSSION

One of the phenotypic markers used in the epidemiology of cholera is the antimicrobial susceptibility patterns. In this study, *V. cholerae* O139 isolates were resistant to ampicillin, furazolidone, and nalidixic acid, a trend observed in majority



of the *V. cholerae* O1 serotype Inaba isolated during 2004–2005 from different parts of India (Dutta et al., 2006). The O139 isolates identified in 1992 were resistant to chloramphenicol, co-trimoxazole, and streptomycin (Mukhopadhyay et al., 1998). The reemerged *V. cholerae* O139 during 1996–1997 in India and Bangladesh showed susceptibility toward co-trimoxazole (Mitra et al., 1998; Faruque et al., 2003a).

In *V. cholerae* O1 and O139, mobile ICE that carried antimicrobial resistance genes in the variable region expressed resistance to chloramphenicol, co-trimoxazole, and streptomycin (Hochhut et al., 2001). In this study, ICE was detected in all the O139 isolates. However, only some of the 2001 isolates were resistant to chloramphenicol, streptomycin, and co-trimoxazole and harbored *floR*, *str*, and *dfr*. These resistance-encoding genes were not present in other isolates in the ICE variable region. Early studies conducted during the emergence of *V. cholerae* O139 in India showed a trend of resistance to neomycin (Mukhopadhyay et al., 1998). In this study, the O139 isolates were either susceptible or showed reduced susceptibility to neomycin. As seen in previous reports, all the *V. cholerae* O139 isolates remained susceptible to norfloxacin, tetracycline, and ciprofloxacin, which are used in the treatment of cholera (Basu et al., 2000).

The CT genotype of *V. cholerae* O1 El Tor isolates from many countries has changed from CT genotype 3 to 1 (Safa et al., 2008; Raychoudhuri et al., 2009) and such changes were detected in strains associated with large cholera outbreaks in India and Bangladesh (Kumar et al., 2009; Nguyen et al., 2009; Taneja et al., 2009). CT genotype 4 has closest homology to CT genotype 1 with a difference of only single nucleotide (nucleotide cytosine instead of adenine) at position 83 (Raychoudhuri et al., 2010). Overall, our finding matches with the observation made in *V. cholerae* O139 isolated during 1998, 2000, and 2002 from Bangladesh and Kolkata, respectively (Bhuiyan et al.,

2009; Raychoudhuri et al., 2010). Compared to El Tor, the hybrid isolates with CT genotype 4 have caused larger cholera outbreaks with more severe clinical symptoms (Kumar et al., 2009; Nguyen et al., 2009; Taneja et al., 2009; Siddique et al., 2010).

Epidemiologically, the CTXΦ appear to be very important as they show the genetic changes among *V. cholerae* O1/O139 that emerged during different periods (Faruque et al., 2000; Qu et al., 2003). In the *ctxA* RFLP analysis, three prophages were encountered in different years. The unusual genetic features of the three 2001 isolates of *V. cholerae* O139 includes identification of the new ribotype BVIII pattern, the presence of three *rstR* allele types, CTX prophages of the classical type, and integration of CTX prophage in both the chromosomes. Epidemiologically, the new ribotypes of *V. cholerae* O1/O139 has been identified along with changes in the CTX prophage or *rstR* allele (Faruque et al., 1997). Considering several genetic events in the past, it has been inferred that the *V. cholerae* O139 may have multiple origins with different progenitors (Faruque et al., 2003b; Garg et al., 2003; Qu et al., 2003).

Genesis of *V. cholerae* O1 El Tor from the classical biotype, the emergence of the serogroup O139, and existence of El Tor that produces classical CT suggests that the *V. cholerae* is in a continuous state of adaptability, resulting in generation of new serogroups and/or new variants of the same serogroup. Our results suggest that the genome of *V. cholerae* O139 is dynamic and has undergone several changes since its emergence in 1992. Continuous surveillance and proper monitoring of *V. cholerae* O139 are however needed to detect subtle genetic changes in the genomes and its implications in its epidemiology, pathogenesis and persistence. Future studies should focus on epigenetic studies to find answers to the question as to why the O139 serogroup has disappeared from cholera endemic regions despite several genetic changes.

## AUTHOR CONTRIBUTIONS

RG, NS, KH, GC, and GP isolated and identified the pathogens, performed phenotypic characterization and all the genetic analysis. RB, AM, SS, GN, and TR conceived the idea analyzed the data and wrote the manuscript. All authors were involved in the compilation of the report and approved the final version.

## REFERENCES

- Basu, A., Garg, P., Datta, S., Chakraborty, S., Bhattacharya, T., Khan, A., et al. (2000). *Vibrio cholerae* O139 in Calcutta, 1992–1998: incidence, antibiograms, and genotypes. *Emerg. Infect. Dis.* 6, 139–147. doi: 10.3201/eid0602.000206
- Bhattacharya, T., Chatterjee, S., Maiti, D., Bhadra, R. K., Takeda, Y., Nair, G. B., et al. (2006). Molecular analysis of the *rstR* and *orfU* genes of the CTX prophages integrated in the small chromosomes of environmental *Vibrio cholerae* non-O1, non-O139 strains. *Environ. Microbiol.* 8, 526–634. doi: 10.1111/j.1462-2920.2005.00932.x
- Bhuiyan, N. A., Nusrin, S., Alam, M., Morita, M., Watanabe, H., Ramamurthy, T., et al. (2009). Changing genotypes of cholera toxin (CT) of *Vibrio cholerae* O139 in Bangladesh and description of three new CT genotypes. *FEMS Immunol. Med. Microbiol.* 57, 136–141. doi: 10.1111/j.1574-695X.2009.00590.x
- Brosius, J., Ullrich, A., Raker, M. A., Gray, A., Dull, T. J., Gutell, R. R., et al. (1981). Construction and fine mapping of recombinant plasmids containing the *rrnB* ribosomal RNA operon of *E. coli*. *Plasmid* 6, 112–118. doi: 10.1016/0147-619X(81)90058-5
- Chatterjee, S., Ghosh, K., Raychoudhuri, A., Pan, A., Bhattacharya, M. K., Mukhopadhyay, A. K., et al. (2007). Phenotypic and genotypic traits and epidemiological implication of *Vibrio cholerae* O1 and O139 strains in India during 2003. *J. Med. Microbiol.* 56, 824–832. doi: 10.1099/jmm.0.46982-0
- Chun, J., Grim, C. J., Hasan, N. A., Lee, J. H., Choi, S. Y., Haley, B. J., et al. (2009). Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic *Vibrio cholerae*. *Proc. Natl. Acad. Sci. U.S.A.* 106, 15442–15447. doi: 10.1073/pnas.0907787106
- CLSI (2014). *Performance Standards for Antimicrobial Susceptibility Testing: Twenty-Fourth Informational Supplement*. CLSI Document M100-S24. Wayne, PA: Clinical and Laboratory Standards Institute.
- Cooper, K. L., Luey, C. K., Bird, M., Terajima, J., Nair, G. B., Kam, K. M., et al. (2006). Development and validation of a PulseNet standardized pulsed-field gel electrophoresis protocol for subtyping of *Vibrio cholerae*. *Foodborne Pathog. Dis.* 3, 51–58. doi: 10.1089/fpd.2006.3.51
- Dutta, B., Ghosh, R., Sharma, N. C., Pazhani, G. P., Taneja, N., Raychowdhuri, A., et al. (2006). Spread of cholera with newer clones of *Vibrio cholerae* O1 El Tor, serotype Inaba, in India. *J. Clin. Microbiol.* 44, 3391–3393. doi: 10.1128/JCM.00632-06
- Faruque, S. M., Ahmed, K. M., Siddique, A. K., Zaman, K., Alim, A. R., and Albert, M. J. (1997). Molecular analysis of toxigenic *Vibrio cholerae* O139 Bengal strains isolated in Bangladesh between 1993 and 1996: evidence for emergence of a new clone of the Bengal vibrios. *J. Clin. Microbiol.* 35, 2299–2306.
- Faruque, S. M., Chowdhury, N., Kamruzzaman, M., Ahmad, Q. S., Faruque, A. S., Salam, M. A., et al. (2003a). Reemergence of epidemic *Vibrio cholerae* O139, Bangladesh. *Emerg. Infect. Dis.* 9, 1116–1122. doi: 10.3201/eid0909.020443
- Faruque, S. M., Sack, D. A., Sack, R. B., Colwell, R. R., Takeda, Y., and Nair, G. B. (2003b). Emergence and evolution of *Vibrio cholerae* O139. *Proc. Natl. Acad. Sci. U.S.A.* 100, 1304–1309. doi: 10.1073/pnas.0337468100
- Faruque, S. M., Saha, M. N., Asadulghani, Bag, P. K., Bhadra, R. K., Bhattacharya, S. K., et al. (2000). Genomic diversity among *Vibrio cholerae* O139 strains isolated in Bangladesh and India between 1992 and 1998. *FEMS Microbiol. Lett.* 184, 279–284. doi: 10.1111/j.1574-6968.2000.tb09027.x
- Garg, P., Aydianian, A., Smith, D. J., Glenn, M. Jr., Nair, G. B., and Stine, O. C. (2003). Molecular epidemiology of O139 *Vibrio cholerae*: mutation, lateral gene transfer, and founder flush. *Emerg. Infect. Dis.* 9, 810–814. doi: 10.3201/eid0907.020760
- Garg, S., Ramamurthy, T., Mukhopadhyay, A. K., Deb, B. C., Nair, G. B., Shimada, T., et al. (1994). Production and cross-reactivity patterns of a panel of high affinity monoclonal antibodies to *Vibrio cholerae* O139 Bengal. *FEMS Immunol. Med. Microbiol.* 8, 293–298. doi: 10.1111/j.1574-695X.1994.tb00455.x
- Ghosh, R., Nair, G. B., Tang, L., Morris, J. G., Sharma, N. C., Ballal, M., et al. (2008). Epidemiological study of *Vibrio cholerae* using variable number of tandem repeats. *FEMS Microbiol. Lett.* 288, 196–201. doi: 10.1111/j.1574-6968.2008.01352.x
- Hochhut, B., Lotfi, Y., Mazel, D., Faruque, S. M., Woodgate, R., and Waldor, M. K. (2001). Molecular analysis of the antibiotic resistance gene clusters in the *Vibrio cholerae* O139 and O1 SXT constins. *Antimicrob. Agents Chemother.* 45, 2991–3000. doi: 10.1128/AAC.45.11.2991-3000.2001
- Hoshino, K., Yamasaki, S., Mukhopadhyay, A. K., Chakraborty, S., Basu, A., Bhattacharya, S. K., et al. (1998). Development and evaluation of a multiplex PCR assay for rapid detection of toxigenic *Vibrio cholerae* O1 and O139. *FEMS Immunol. Med. Microbiol.* 20, 201–207. doi: 10.1111/j.1574-695X.1998.tb01128.x
- Kaper, J. B., Morris, J. G. Jr., and Nishibuchi, M. (1988). “DNA probes for pathogenic *Vibrio* species,” in *DNA Probes for Infectious Diseases*, ed F. C. Tenover (Boca Raton, FL: CRC Press), 65–77.
- Keasler, S. P., and Hall, R. H. (1993). Detecting and biotyping *Vibrio cholerae* O1 with multiplex polymerase chain reaction. *Lancet* 341, 1661. doi: 10.1016/0140-6736(93)90792-F
- Kumar, P., Jain, M., Goel, A. K., Bhaduria, S., Sharma, S. K., Kamboj, D. V., et al. (2009). A large cholera outbreak due to a new cholera toxin variant of the *Vibrio cholerae* O1 El Tor biotype in Orissa, Eastern India. *J. Med. Microbiol.* 58, 234–238. doi: 10.1099/jmm.0.002089-0
- Maiti, D., Das, B., Saha, A., Nandy, R. K., Nair, G. B., and Bhadra, R. K. (2006). Genetic organization of pre-CTX and CTX prophages in the genome of an environmental *Vibrio cholerae* non-O1, non-O139 strain. *Microbiology* 152, 3633–3641. doi: 10.1099/mic.0.2006/000117-0
- Mitra, R., Basu, A., Dutta, D., Nair, G. B., and Takeda, Y. (1998). Resurgence of *Vibrio cholerae* O139 Bengal with altered antibiogram in Calcutta, India. *Lancet* 348, 1181. doi: 10.1016/S0140-6736(05)65326-3
- Morita, M., Ohnishi, M., Arakawa, E., Bhuiyan, N. A., Nusrin, S., Alam, M., et al. (2008). Development and validation of a mismatch amplification mutation PCR assay to monitor the dissemination of an emerging variant of *Vibrio cholerae* O1 biotype El Tor. *Microbiol. Immunol.* 52, 314–317. doi: 10.1111/j.1348-0421.2008.00041.x
- Mukhopadhyay, A. K., Basu, A., Garg, P., Bag, P. K., Ghosh, A., Bhattacharya, S. K., et al. (1998). Molecular epidemiology of reemergent *Vibrio cholerae* O139 Bengal in India. *J. Clin. Microbiol.* 36, 2149–2152.
- Murray, M. G., and Thompson, W. F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* 8, 4321–4325. doi: 10.1093/nar/8.19.4321
- Nair, G. B., Bhattacharya, S. K., and Deb, B. C. (1994b). *Vibrio cholerae* O139 Bengal: the eighth pandemic strain of cholera. *Indian J. Public Health* 38, 33–36.
- Nair, G. B., Misra, S., Bhadra, R. K., and Pal, S. C. (1987). Evaluation of the multitest medium for rapid presumptive identification of *Vibrio cholerae* from environmental sources. *Appl. Environ. Microbiol.* 53, 1203–1205.
- Nair, G. B., Qadri, F., Holmgren, J., Svenssonholm, A. M., Safa, A., Bhuiyan, N. A., et al. (2006). Cholera due to altered El Tor strains of *Vibrio cholerae* O1 in Bangladesh. *J. Clin. Microbiol.* 44, 4211–4213. doi: 10.1128/JCM.01304-06
- Nair, G. B., Ramamurthy, T., Bhattacharya, S. K., Mukhopadhyay, A. K., Garg, S., Bhattacharya, M. K., et al. (1994a). Spread of *Vibrio cholerae* O139 Bengal in India. *J. Infect. Dis.* 169, 1029–1034. doi: 10.1093/infdis/169.5.1029
- Nandi, S., Maiti, D., Saha, A., and Bhadra, R. K. (2003). Genesis of variants of *Vibrio cholerae* O1 biotype El tor: role of the CTX $\delta$ d array and its position in the genome. *Microbiology* 149, 89–97. doi: 10.1099/mic.0.25599-0
- Nguyen, B. M., Lee, J. H., Cuong, N. T., Cho, S. Y., Hien, N. T., Anh, D. D., et al. (2009). Cholera outbreaks caused by an altered *Vibrio cholerae* O1 El Tor

## FUNDING

This work supported in part by the Japan Agency for Medical Research and Development, Ministry of Education, Culture, Sports, Science and Technology, Japan, Council of Scientific and Industrial Research, and Indian Council of Medical Research, New Delhi, India.

- biotype strain producing classical cholera toxin B in Vietnam in 2007 to 2008. *J. Clin. Microbiol.* 47, 1568–1571. doi: 10.1128/JCM.02040-08
- Olsvik, O., Wahlberg, J., Petterson, B., Uhlén, M., Popovic, T., Wachsmuth, I. K., et al. (1993). Use of automated sequencing of polymerase chain reaction-generated amplicons to identify three types of cholera toxin subunit B in *Vibrio cholerae* O1 strains. *J. Clin. Microbiol.* 31, 22–25.
- Qu, M., Xu, J., Ding, Y., Wang, R., Liu, P., Kan, B., et al. (2003). Molecular epidemiology of *Vibrio cholerae* O139 in China: polymorphism of ribotypes and CTX elements. *J. Clin. Microbiol.* 41, 2306–2310. doi: 10.1128/JCM.41.6.2306-2310.2003
- Ramamurthy, T., Garg, S., and Nair, G. B. (1995). Monoclonal antibodies against Ogawa-specific and Ogawa-Inaba common antigenic determinants of *Vibrio cholerae* & their diagnostic utility. *Indian J. Med. Res.* 101, 10–12.
- Ramamurthy, T., Yamasaki, S., Takeda, Y., and Nair, G. B. (2003). *Vibrio cholerae* O139 Bengal: odyssey of a fortuitous variant. *Microbes Infect.* 5, 329–344. doi: 10.1016/S1286-4579(03)00035-2
- Raychoudhuri, A., Mukherjee, P., Ramamurthy, T., Nandy, R. K., Takeda, Y., Nair, G. B., et al. (2010). *rstr* alleles of *Vibrio cholerae* O139 strains isolated from Kolkata over a decade. *FEMS Microbiol. Lett.* 303, 107–115. doi: 10.1111/j.1574-6968.2009.01856.x
- Raychoudhuri, A., Patra, T., Ghosh, K., Ramamurthy, T., Nandy, R. K., Takeda, Y., et al. (2009). Classical *ctxB* in *Vibrio cholerae* O1, Kolkata, India. *Emerg. Infect. Dis.* 15, 131–132. doi: 10.3201/eid1501.080543
- Sack, D. A., Sack, R. B., Nair, G. B., and Siddique, A. K. (2004). Cholera. *Lancet* 363, 223–233. doi: 10.1016/S0140-6736(03)15328-7
- Safa, A., Sultana, J., Dac Cam, P., Mwansa, J. C., and Kong, R. Y. (2008). *Vibrio cholerae* O1 hybrid El Tor strains, Asia and Africa. *Emerg. Infect. Dis.* 14, 987–988. doi: 10.3201/eid1406.080129
- Sharma, C., Maiti, S., Mukhopadhyay, A. K., Basu, A., Basu, I., Nair, G. B., et al. (1997). Unique organization of the CTX genetic element in *Vibrio cholerae* O139 strains which reemerged in Calcutta, India, in September 1996. *J. Clin. Microbiol.* 35, 3348–3350.
- Siddique, A. K., Akram, K., Zaman, K., Mutsuddy, P., Eusof, A., and Sack, R. B. (1996). *Vibrio cholerae* O139: how great is the threat of a pandemic? *Trop. Med. Int. Health* 1, 393–398. doi: 10.1046/j.1365-3156.1996.d01-54.x
- Siddique, A. K., Nair, G. B., Alam, M., Sack, D. A., Huq, A., Nizam, A., et al. (2010). El Tor cholera with severe disease: a new threat to Asia and beyond. *Epidemiol. Infect.* 138, 347–352. doi: 10.1017/S0950268809990550
- Taneja, N., Mishra, A., Sangar, G., Singh, G., and Sharma, M. (2009). Outbreaks caused by new variants of *Vibrio cholerae* O1 El Tor, India. *Emerg. Infect. Dis.* 15, 352–354. doi: 10.3201/eid1502.080943
- Waldor, M. K., and Mekalanos, J. J. (1996). Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* 272, 1910–1914. doi: 10.1126/science.272.5270.1910
- WHO (2009). Cholera: global surveillance summary, 2008. *Wkly. Epidemiol. Rec.* 84, 309–324.
- Yamai, S., Okitsu, T., Shimada, T., and Katsume, Y. (1997). Distribution of serogroups of *Vibrio cholerae* non-O1 non-O139 with specific reference to their ability to produce cholera toxin, and addition of novel serogroups. *Kansenshogaku Zasshi* 71, 1037–1045. doi: 10.11150/kansenshogakuzasshi1970.71.1037
- Yamasaki, S., Shimizu, T., Hoshino, K., Ho, S. T., Shimada, T., Nair, G. B., et al. (1999). The genes responsible for O-antigen synthesis of *Vibrio cholerae* O139 are closely related to those of *Vibrio cholerae* O22. *Gene* 237, 321–332. doi: 10.1016/S0378-1119(99)00344-3
- Zhang, P., Zhou, H., Diao, B., Li, F., Du, P., Li, J., et al. (2014). Molecular surveillance reveals the prevalence of *Vibrio cholerae* O139 isolates in China from 1993 to 2012. *J. Clin. Microbiol.* 52, 1146–1152. doi: 10.1128/JCM.03354-13

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Ghosh, Sharma, Halder, Bhadra, Chowdhury, Pazhani, Shinoda, Mukhopadhyay, Nair and Ramamurthy. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Prevalence and genetic diversity of clinical *Vibrio parahaemolyticus* isolates from China, revealed by multilocus sequence typing scheme

Dongsheng Han<sup>1†</sup>, Hui Tang<sup>2†</sup>, Chuanli Ren<sup>1</sup>, Guangzhou Wang<sup>1</sup>, Lin Zhou<sup>1</sup> and Chongxu Han<sup>1\*</sup>

<sup>1</sup> Department of Clinical Microbiology, Clinical Medical Examination Center, Northern Jiangsu People's Hospital, Yangzhou, China, <sup>2</sup> Department of Biobank, Northern Jiangsu People's Hospital, Yangzhou, China

## OPEN ACCESS

### Edited by:

Julio Alvarez,  
University of Minnesota, USA

### Reviewed by:

Ronald Paul Rabinowitz,  
University of Maryland School of  
Medicine, USA

Li Xu,  
Cornell University, USA

### \*Correspondence:

Chongxu Han,  
Department of Clinical Microbiology,  
Clinical Medical Examination Center,  
Northern Jiangsu People's Hospital,  
Nantongxi Road, Yangzhou 225001,  
China  
huisheng113905@163.com

<sup>†</sup>These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Infectious Diseases, a section of the  
journal Frontiers in Microbiology

Received: 06 February 2015

Accepted: 24 March 2015

Published: 09 April 2015

### Citation:

Han D, Tang H, Ren C, Wang G, Zhou L and Han C (2015) Prevalence and genetic diversity of clinical *Vibrio parahaemolyticus* isolates from China, revealed by multilocus sequence typing scheme. *Front. Microbiol.* 6:291. doi: 10.3389/fmicb.2015.00291

The population structure of clinical *Vibrio parahaemolyticus* isolates spreading in China remains undefined. We brought 218 clinical isolates from the pubMLST database originating from different regions of China collected since the year of 1990, analyzed by multilocus sequence typing (MLST), to elucidate the prevalence and genetic diversity of *V. parahaemolyticus* circulating in Chinese population. The MLST scheme produced 137 sequence types (STs). These STs were clustered into six clonal complexes (CCs), six doublets, and 91 singletons, exhibiting a high level of genetic diversity. However, less diversity was displayed on the peptide level: only 46 different peptide sequence type (pST) were generated, with pST2 (44.0%, 96/218) and pST1 (15.1%, 33/218) the predominant. Further analysis confirmed all the pSTs belong to a single complex founded by pST1, pST2, pST3, and pST4. *recA* presented the highest degree of nucleotide diversity (0.026) and the largest number of variable sites (176) on the nucleotide level. *pyrC* was the most diverse locus on the peptide level, possessing the highest percentage of variable sites (9.2%, 15/163). Significant linkage disequilibrium with the alleles was detected when the Standardized Index of Association ( $I_A^S$ ) was calculated both for the entire isolates collection ( $0.7169, P < 0.01$ ) and for the 137 STs ( $I_A^S = 0.2648, P < 0.01$ ). In conclusion, we provide an overview of prevalence and genetic diversity of clinical *V. parahaemolyticus* spreading in Chinese population using MLST analysis. The results would offer genetic evidences for uncovering the microevolution relationship of *V. parahaemolyticus* populations.

**Keywords:** *Vibrio parahaemolyticus*, multilocus sequence typing, phylogenetic analysis, clonal complex, peptide sequence types

## Introduction

*Vibrio parahaemolyticus* is a leading cause of food-borne outbreaks and acute gastroenteritis throughout the world, especially in coastal countries and regions. Consumption of raw or undercooked seafood is the major route of transmission for *V. parahaemolyticus* infection (Pal and Das, 2010). *V. parahaemolyticus* infection is caused by diverse serotypes; however, since the pandemic serovar O3:K6 emerged in Asia in 1996, it was confirmed as the predominant cause of outbreaks of

*V. parahaemolyticus* infection on a global scale (Okuda et al., 1997; Bag et al., 1999; Chowdhury et al., 2000; Nair et al., 2007). In recent years, at least 21 pandemic serotypes were identified as being associated with the outbreaks of *V. parahaemolyticus* infection (Nair et al., 2007).

*V. parahaemolyticus* is the first leading cause of food-borne outbreaks and bacterial infectious diarrhea in China, especially in the southeast coastal area (Lin et al., 2011). During 2000–2009, a multicentric surveillance for *V. parahaemolyticus* diarrhea at 11 provinces (Beijing, Jiangsu, Shanghai, Zhejiang, et al.) was conducted by China CDC, they collected stool or rectal swab specimens from 79,075 diarrhea patients to detected the prevalence ratio of *V. parahaemolyticus*, and found the average ratio of *V. parahaemolyticus* was 3.11% in these diarrhea patients. Studies also confirmed clinical isolates in China typically correspond to the pandemic serovar O3:K6 and several serovariants of it (e.g., O1:KUT, O1:K56, and O4:K68) (Chao et al., 2009; Yu et al., 2011; Fan et al., 2013; Shi et al., 2013; Li et al., 2014). So it's critical to clarify the prevalence and genetic diversity of this pathogen circulating in a particular population for minimizing both the risk of infection and economical burden. For the molecular genetic studies of *V. parahaemolyticus*, a number of molecular typing techniques have been developed and applied (Marshall et al., 1999; Gonzalez-Escalona et al., 2008; Chen et al., 2012). Multilocus sequence typing (MLST) of *V. parahaemolyticus* was developed by González-Escalona et al in 2008 (Gonzalez-Escalona et al., 2008). Gonzalez-Escalona's own and a number of subsequent studies demonstrated MLST was a powerful tool with a high resolution rate in identification of clonal complexes (CCs) of *V. parahaemolyticus* population and in understanding the processes leading to the emergence and spread of pathogenic isolates (Harth et al., 2009; Yu et al., 2011; Ellis et al., 2012; Banerjee et al., 2014).

Although several studies already used MLST analysis to study the genetic diversity of *V. parahaemolyticus* in China in recent years, they were restricted to specific regions (Chao et al., 2009; Han et al., 2012; Fan et al., 2013; Shi et al., 2013), focused exclusively on pandemic pathogenic isolates (Chao et al., 2011; Yan et al., 2011) or were based on a limited isolate number (Yu et al., 2011). In the present study, we brought 218 Chinese clinical isolates from the pubMLST database (<http://pubmlst.org/vparahaemolyticus/>) into our analyses. These clinical isolates were collected mostly from the provinces of Beijing, Jiangsu, Zhejiang, Fujian, Guangdong, and Taiwan. With these isolates, we aimed to elucidate the prevalence and genetic diversity of clinical *V. parahaemolyticus* isolates circulating in Chinese population. We would investigate the sequence/peptide polymorphisms of the isolates and analyze the probable evolutionary relationships among the isolates. The differences in regard to CC and sequence type (ST)/peptide sequence type (pST) affiliation in the analyzed isolates were considered. We provide a broader overview of the genetic population structure of clinical *V. parahaemolyticus* from China, and predict that the results will provide genetic evidences for uncovering the microevolution relationship among different isolates and might be conducive to the early warning and prevention of *V. parahaemolyticus* infection.

## Materials and Methods

### Sampling of *V. parahaemolyticus* Isolates

A total of 218 clinical *V. parahaemolyticus* isolates from Chinese patients with acute gastroenteritis were selected as the research subject of this study, they were all available in the pubMLST database (<http://pubmlst.org/vparahaemolyticus/>). These isolates were both temporally (collected from 1990 to November 2014) and geographically (collected from the provinces of Beijing, Jiangsu, Zhejiang, Fujian, Guangdong, and Taiwan) diverse (see Table S1 in the Supplemental Material).

### MLST Analysis and Phylogenetic Analysis

There is a PCR protocol of internal fragments of the seven house-keeping genes [*recA*(729bp), *dnaE*(557bp), *gyrB*(592bp), *dtdS*(458bp), *pntA*(430bp), *pyrC*(493bp), and *tmaA*(423bp)] on *V. parahaemolyticus* pubMLST website (<http://pubmlst.org/vparahaemolyticus/>), the allele designations and STs of all the 218 isolates had been determined. Based on the related STs, all the isolates were subdivided into CCs or groups by goeBURST analysis using Phyloviz software (<http://www.phyloviz.net>). We also implemented “population snapshot” analysis on the basis of STs and pSTs by using goeBURST. Isolates that shared 100% identity in six of the seven loci with at least one other member of the group, the single locus variants (SLVs), were assigned to a single CC. The primary founder of a CC, SLVs, double locus variants (DLVs), and singletons were defined as in previous study (Han et al., 2014).

When a nucleotide sequence was translated in frame, a peptide sequence could be obtained, in other words, each nucleotide sequence correspond to a unique peptide sequence, an individual isolate contains a unique ST as well as a pST. So translating the in-frame nucleotide sequences into peptide sequences allows a phylogenetic analysis based on pSTs. In this study, the assignment of pSTs of the analyzed isolates to CCs was carried out as previously (Theethakaew et al., 2013; Urmersbach et al., 2014) and predicted also by goeBURST algorithm.

Minimum-evolution (ME) trees for the in-frame concatenated sequences (*recA-dnaE-gyrB-dtdS-pntA-pyrC-tmaA*) of each (p)ST were constructed by Mega 5 software, genetic distance of the analyzed isolates was estimated by the Kimura two-parameter model, as did in the other study (Han et al., 2014).

### Population Genetic Analysis

DnaSP V5 was used to calculate the following parameters: the number of alleles, the number of polymorphic sites and nucleotide diversity( $\pi$ ), for evaluating the varying degrees of the loci in our selected isolates (Librado and Rozas, 2009). START V2 was implemented to calculate the ratio of non-synonymous-to-synonymous substitutions (dN/dS) through the Nei and Gojobori method (Jolley et al., 2001). dN/dS < 1 indicates that the relative gene was mainly affected by purifying selection during the population evolution, dN/dS = 1 indicates neutral selection and dN/dS > 1 indicates positive selection. The value of Standardized Index of Association ( $I_A^S$ ) was calculated by START2, in order to access the population structure of *V. parahaemolyticus*.  $I_A^S = 0$  indicates alleles are in linkage equilibrium

(alleles are independently distributed at all loci analyzed) and recombination occurred frequently (Gonzalez-Escalona et al., 2008).

## Results

### Diversity of Sequence Types (STs)

The data on diversities of the seven loci in the 218 *V. parahaemolyticus* isolates are showed in **Table 1**. All these analyzed isolates resulted in 137 unique STs by applying MLST analysis. Among these STs, Individual STs were mostly recovered once (118 STs), ST3 was most frequent (52 isolates), 18 STs was constituted of between two and five isolates (see Table S1 in the Supplemental Material). When the geographical subsets were considered, the number of different STs was high in Guangdong, Jiangsu, Taiwan and Shanghai (**Table 2**).

The number of alleles observed for each locus ranged from 51 (*pntA* and *tmaA*) to 79 (*gyrB*). *gyrB* possessed the most number of alleles (79 alleles) but only 62 (10.47%) variable sites. The nucleotide diversity ranged from 0.011(*pntA*) to 0.026(*recA*). The dN/dS ratio for every locus was close to zero (**Table 1**), this suggests that the housekeeping genes are mainly affected by purifying selection during the evolutionary process.

### Diversity of Peptide Sequence Types (pSTs)

A total of 46 different pSTs were obtained from the analyzed isolates (see Table S1 in the supplemental material), occurred with a frequency of 0.5 to 44.0%. pST2 (44.0%, 96/218) and pST1 (15.1%, 33/218) were the two predominant pSTs. One particular pST could be comprised of numerous STs, in this study, we found that pST1 were translated by the nucleotide sequences of 27 different STs (ST62, ST91, ST120, etc.), pST2 by 38 STs (ST3, ST189, ST345, etc.), pST3 by 7 STs (ST328, ST332, ST444, etc.), pST4 by 9 STs (ST8, ST224, ST301, etc.), and the other pSTs by four or less STs.

On pST level, the proportion of allele ONE was more than 90.0% for most of the seven loci, except for *dnaE* (allele ONE accounting for 35.8%) and *pyrC* (allele ONE accounting for 73.4%) (see Table S1 in the Supplemental Material). The individual loci possessed 2 (*gyrB*) to 16 (*pyrC*) unique alleles. *pyrC* possessed the highest percentage of variable sites (9.2%, 15/163) (**Table 1**).

### Clonal Relationships of the Collected Isolates

In this study, the calculated  $I_A^S$  value was 0.7169 ( $P < 0.01$ ) for all of the 218 analyzed isolates, that is to say, the alleles in the seven housekeeping genes are in linkage disequilibrium. When we calculated the  $I_A^S$  value repeatedly, using one isolate to

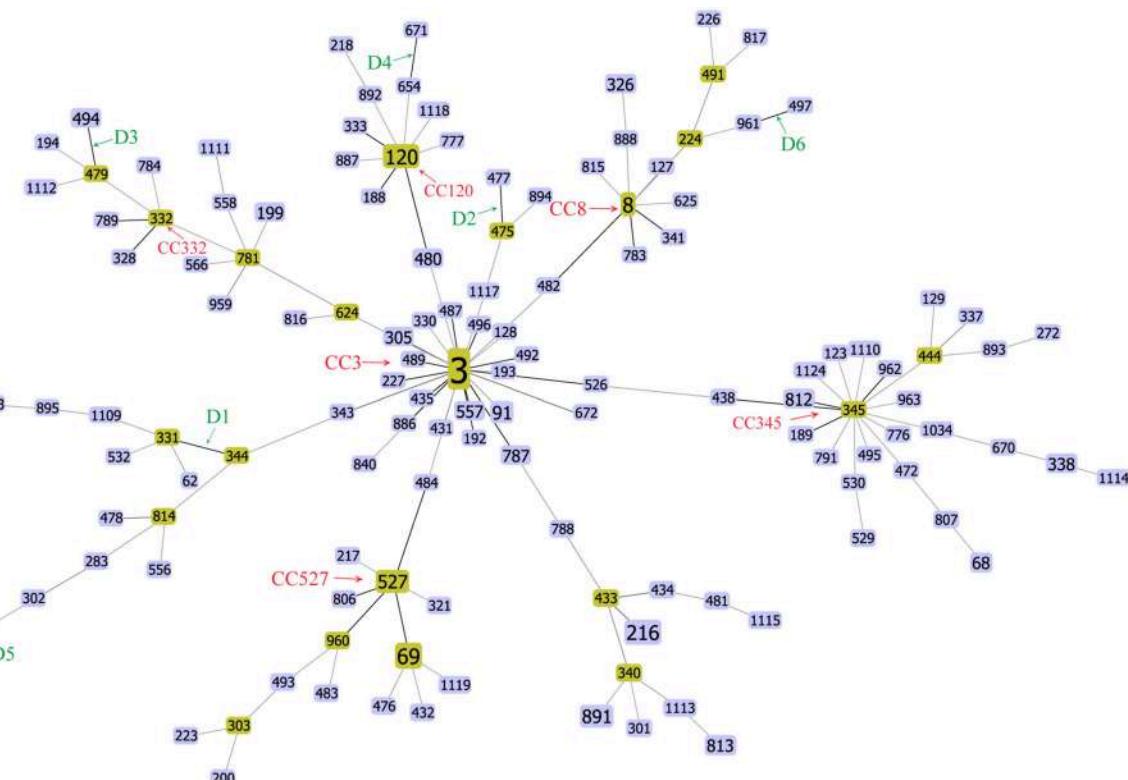
**TABLE 1 | Diversities of the seven loci in the 218 *V. parahaemolyticus* isolates.**

Locus	Fragment size		No of alleles		No (%) of variable sites		nucleotide diversity ( $\pi$ )	dN/dS
	ST	pST	ST	pST	ST	pST	ST	ST
<b>I</b>								
<i>dnaE</i>	557bp	185aa	67	6	63 (11.31)	5 (2.70)	0.012	0.028
<i>gyrB</i>	592bp	197aa	79	2	62 (10.47)	–(–)*	0.015	0.001
<i>recA</i>	729bp	242aa	75	4	176 (24.14)	19 (7.85)	0.026	0.017
<b>II</b>								
<i>dtdS</i>	458bp	152aa	68	3	47 (10.26)	2 (1.32)	0.023	0.002
<i>pntA</i>	430bp	143aa	51	12	40 (9.30)	10 (6.99)	0.011	0.031
<i>pyrC</i>	493bp	163aa	62	16	52 (10.55)	15 (9.20)	0.013	0.066
<i>tmaA</i>	423bp	141aa	51	7	42 (9.93)	6 (4.26)	0.013	0.013

\*For the occurring alleles no variable sites could be determined.

**TABLE 2 | Geographic distribution of the 218 *V. parahaemolyticus* isolates.**

Geographical subsets	No. of isolates	No. of STs	Mainly STs	No. of pSTs	Mainly pSTs
Guangdong	51	33	ST3, ST305, ST480, etc.	16	pST1,pST2,pST11, pST116, etc.
Jiangsu	37	22	ST3,ST8,ST120, etc.	14	pST1,pST2,pST4, pST26, etc.
Taiwan	30	11	ST3,ST8, ST120, etc.	7	pST1,pST2,pST4, pST31, etc.
Shanghai	17	17	ST283,ST301,ST302, etc.	6	pST1,pST2,pST3, pST4, etc.
zhejiang	5	5	ST654,ST670,ST671, etc.	3	pST1,pST2,pST16
Beijing	4	4	ST431,ST432,ST478,ST479	3	pST1,pST2,pST163
Hong Kong	3	2	ST3,ST91	2	pST1,pST2
Fujian	2	2	ST963,ST1034	2	pST2, pST197
Unkown	69	53	ST3,ST69,ST216, etc.	21	pST1,pST2,pST3, pST4, etc.
ALL isolates	218	137	ST3,ST8, ST120, etc.	46	pST1,pST2,pST3, pST4, etc.



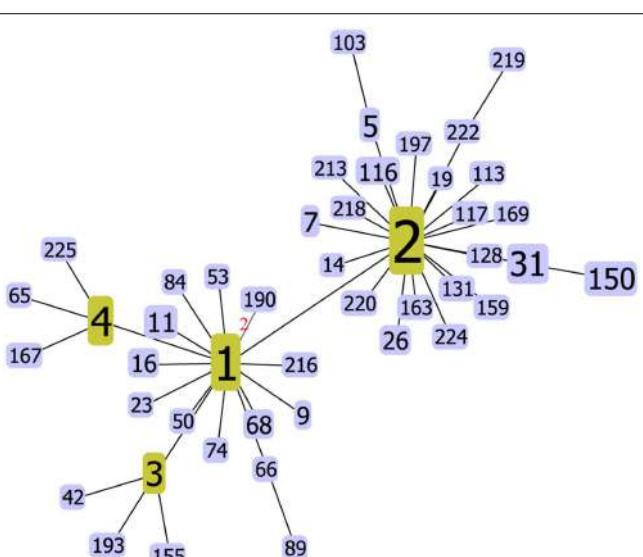
**FIGURE 1 | FullMST of *V. parahaemolyticus* in China based on nucleotide sequence.** Six clonal complexes (CC3, CC8, C120, CC332, CC345, and CC527), six doublets (D1–D6), and 91 singletons were

identified. All connections were drawn. STs that are SLVs of each other are connected by black lines. STs that differ in two or more alleles are connected via light gray lines.

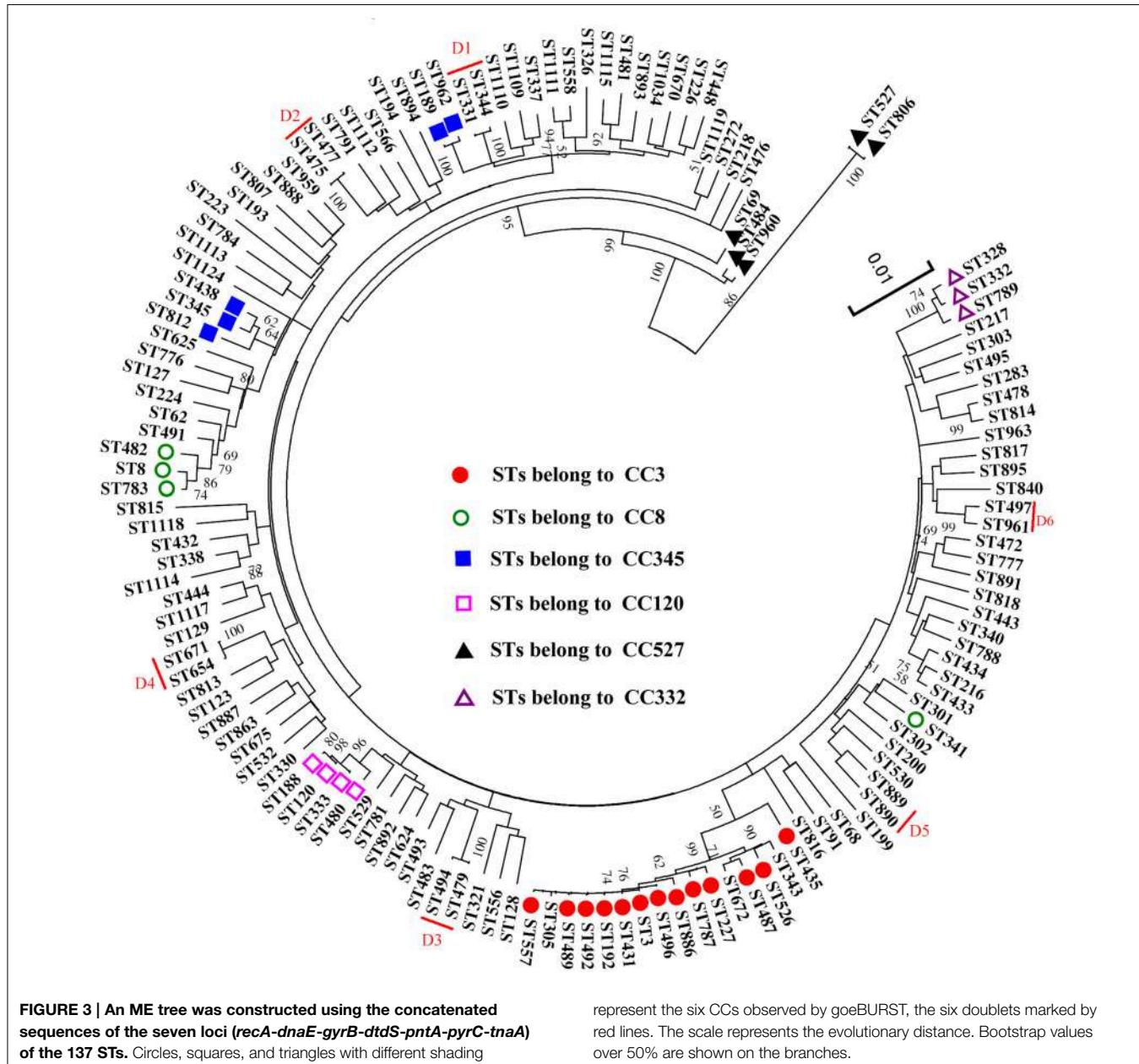
represent each of the 137 STs, it's 0.2648 ( $P < 0.01$ ). Although the value represented a markedly decrease from 0.7169, these alleles are in linkage disequilibrium. This indicates a nonrandom distribution of alleles in the *V. parahaemolyticus* population in general.

### Identification of Clonal Complexes

The goeBURST algorithm used in our study resolved the 137 STs into six CCs (CC3, CC8, C120, CC332, CC345, and CC527) and six doublets (D1–D6). The remaining 91 STs were singletons (Figure 1). CC3 was the most prevalent CC, including 66 isolates with 13 STs. However, neither the relationships between the 91 singletons themselves nor the relationships between the singletons and the defined CCs or doublets could be deduced here. It suggested that goeBURST had limited utility on nucleotide level for identifying related isolates. To counter this, we implemented a “population snapshot” analysis by using goeBURST on the basis of pSTs. The result showed that only pST190 differed in more than one allele to all the other 45 pSTs, leading to a single complex founded by pST1, pST2, pST3, and pST4 (Figure 2). So the relationship among the isolates appears more closely when analyzed on peptide level than on nucleotide level.



**FIGURE 2 | *V. parahaemolyticus* ‘population snapshot’ in China based on peptide sequence.** The number of different alleles is indicated in the case of DLVs, all other pSTs are SLVs. In fact, only pST190 is a DLV of pST1, all other pSTs are SLVs of pST1, pST2, pST3, or pST4.

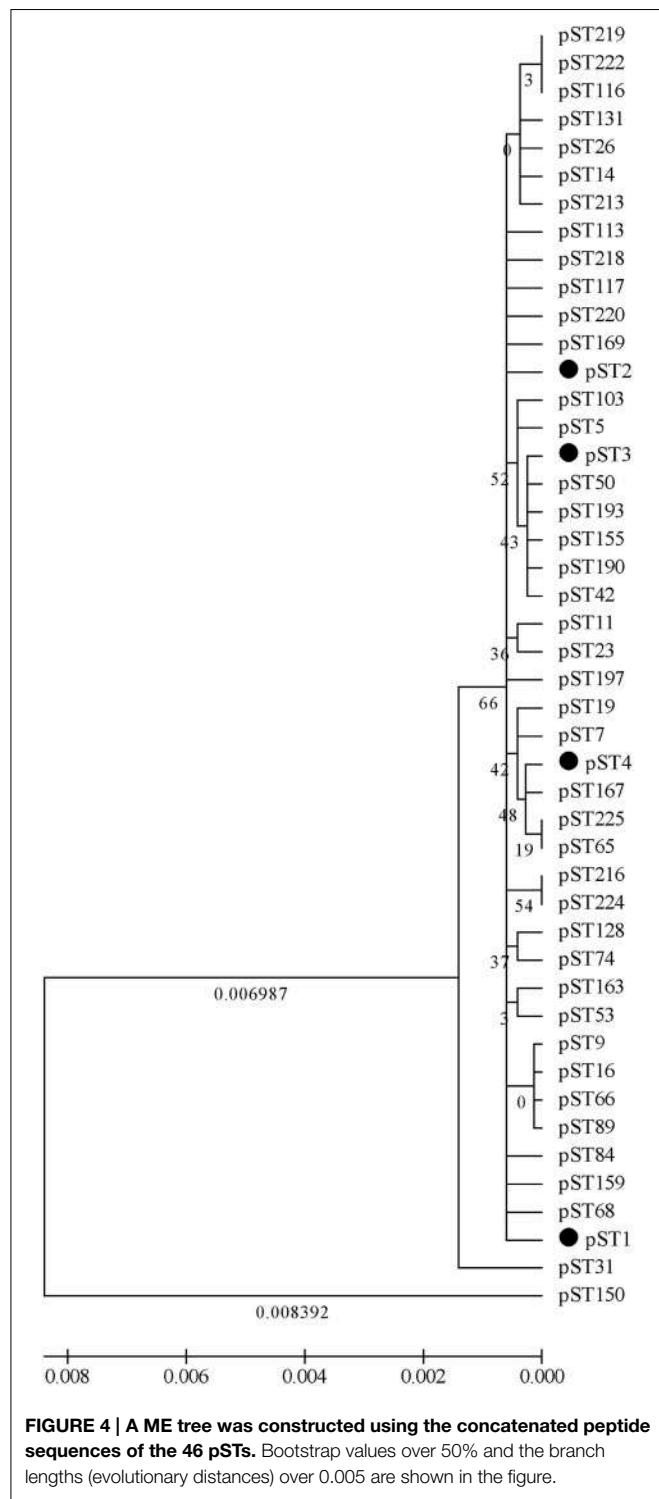


## Phylogenetic Analysis

The result of a phylogenetic analysis using ME tree on nucleotide level revealed a high genetic diversity among the 218 analyzed isolates (Figure 3). The isolates belong to the same CCs and Doublets (D1–D6) in the goeBURST analysis were also clustered together in the ME tree, except for the isolates of CC345 and CC527. CC345 was divided into two different clusters, ST438 and ST812 exhibited a relatively closer evolutionary distance to ST345 (the ancestral type of CC345) than ST962 and ST189. When compared the number of single nucleotide polymorphisms (SNPs), we found ST962 and ST189 showed differences of 25 SNPs (found in *gyrB* allele), and 24 SNPs (found in *recA* allele) from ST345 respectively, whereas ST438 and ST812 only differs from ST345

by 9 SNPs (found in *recA* allele) and 15 SNPs (found in *recA* allele), respectively. Similarly, the difference between SNP numbers had an important impact on the classification of CC527, ST806 (only one SNP different from ST527) showed a closer evolutionary relationships to ST527 than ST69 (139 SNPs differs), ST484 (136 SNPs differs), and ST960 (135 SNPs differs). Thus, a phylogenetic analysis based on the nucleotide sequences provides a better resolution and elucidated some genetic relationships among isolates that were not resolved by goeBURST analysis.

An ME tree analysis was also conducted based on the concatenated sequences of peptide sequences among the 46 pSTs (Figure 4). The low bootstrap values (all <70%) indicated that



the topology of the ME tree was poorly supported. However, the longest branch length (evolutionary distance) was only 0.008392 (found between ST150 and the cluster of other pSTs); the genetic similarity of all the clusters of the pSTs was higher than 99.99%. Thus, the 46 analyzed pSTs belong to a single CC when a phylogenetic analysis carried out on peptide level.

## Discussion

In the present study, we analyzed the prevalence and extent of genetic diversity of *V. parahaemolyticus* among 218 clinical isolates collected in different regions of China. The diversity of the *V. parahaemolyticus* isolates was analyzed on nucleotide as well as peptide level. It is clear that multiple sequence/peptide types are contributed to human infection in China, and the genetic relationship among the isolates appears more closely on peptide level than on nucleotide level.

The observed alleles, variable sites, dN/dS-value, nucleotide diversity ( $\pi$ ),  $I_A^S$ -value and (p)STs of our isolates were similar to those derived from a collection of global isolates in the pubMLST database, which were presented in a study conducted by Urmersbach et al. (2014). This reveals a high diversity of clinical isolates collection in Chinese population.

The low dN/dS ratios (all close to zero) obtained for all the seven genes indicating purifying selection for all the loci, as shown by others (Turner et al., 2013). It means synonymous substitutions were dominant in the nucleotide sequences, which do not alter amino acid sequences. This finding could explain the fact that the numbers of different alleles per locus were reduced and the loci were dominated by a single allele when analyzed on peptide level.

*V. parahaemolyticus* population was extremely genetic diverse in China. A total of 137 STs were identified in this study. ST3 was the most common ST, in agreement with the finding of a previous study on the basis of a global clinical isolates collection (Han et al., 2014). As we know, ST3 has been a sequence type of *V. parahaemolyticus* with an international distribution. Numerous reports revealed that it was widely distributed and played an important role in *V. parahaemolyticus* infection in China (Yu et al., 2011; Han et al., 2012; Fan et al., 2013; Shi et al., 2013). However, when carried out an analysis based on the peptide sequences, the diversity decreased. Only 46 pSTs were generated, and pST1, pST2, pST3, and pST4 were the most prevalent pSTs, which also the predominant pSTs in the pubMLST database (Urmersbach et al., 2014).

Eight CCs and 11 doublets have been identified when we analyzed the clinical isolates with a global collection in the pubMLST dataset (Han et al., 2014). In this study, six CCs and six doublets were found in the “population snapshot” of the 137 STs. CC3, a global pandemic clone of *V. parahaemolyticus* (Martinez-Urtaza et al., 2004, 2005; Ansaruzzaman et al., 2005), was also the most prevalent CC in China, being comprised of 69 isolates with 13 different STs in this study, posing a significant public health threat. However, the “population snapshot” of pSTs consists of only one unique CC, with pST1, pST2, pST3, and pST4 being the ancestral types at the same time. Other pSTs might have arisen from the four types by genetic drift associated with genetic changes (Osorio et al., 2012).

As discussed above, when we analyzed the genetic relationships among different isolates based on STs, goeBURST algorithm showed a decreased ability in identifying the related genotypic relationships due to the high degree of allelic diversity. Relationships are reliable only for identical or closely related isolates. When isolates are more distantly related, such as the 91

singletons, little information can be gained about their relationships. However, when the goeBURST analysis was implemented on the basis of pSTs, all the isolates were classified into one unique CC. This result maybe more representative of the real relationships among the isolates on the phenotypic level. Using pSTs instead of STs might be more efficient in reaching a reliable identification of related isolates.

Previous study confirmed that linkage disequilibrium could be observed when a recent, more epidemic clone arises (Smith et al., 1993). The calculated  $I_A^S$  value was 0.7169 ( $P < 0.01$ ) for all of the analyzed isolates in this study, suggesting that all alleles in the seven housekeeping genes were in linkage disequilibrium or were non-randomly distributed. However, even the analysis was repeated using one isolate to represent each of the 137 STs, which would weaken the influence of the potential pandemic isolates in the data set, these alleles are still in linkage disequilibrium ( $I_A^S = 0.2648$ ,  $P < 0.01$ ). These results indicate that a non-randomly distribution of alleles in the *V. parahaemolyticus* population in general, even though recombination might be occurring in different subtypes (Gonzalez-Escalona et al., 2008). These observations are also typical for epidemic populations (Ellis et al., 2012; Theethakaew et al., 2013; Turner et al., 2013), thus to some extent, our data support the hypothesis that the population structure of *V. parahaemolyticus* follows the epidemic model of clonal expansion (Yu et al., 2011).

The results of goeBURST and ME tree shared a high similarity in the identifying of CCs. In the MLST scheme, the isolates of four CCs (CC3, CC8, C120, and CC332) and the six doublets were also clustered together in the ME tree, indicating that they were genetically exclusive complexes or groups. However, the isolates of CC345 and CC527 were resolved into different clusters. It could be explained by the different approaches used in ME tree and goeBURST algorithm. The ME tree is sequence-based, all sequences with fewer differences could be clustered together. goeBURST algorithm is allelic profile-based, only the SLVs were assigned to a single CC or groups. So the ME tree seems to be more suitable for analyzing genetic relationships of *V. parahaemolyticus* populations (Yan et al., 2011).

An ME tree was also constructed for phylogenetic analysis of *V. parahaemolyticus* on peptide level. All the 46 analyzed pSTs are grouped into a single CC, with the genetic similarity of all the clusters of the pSTs higher than 99.99%. A similar result has been observed by Osorio et al. who aimed to deduce putative ancestral relationships between different *Brachyspira hyodysenteriae* isolates (Osorio et al., 2012).

In this study, isolates without a clear source or STs were screened out. This may has some influence on the genetic characteristics of *V. parahaemolyticus* in China in general.

However, our findings would facilitate the researchers in this field to understand the population structure of *V. parahaemolyticus* in China. As the *V. parahaemolyticus* PubMLST database is not mandatory for uploading laboratory data, not all the research data of *V. parahaemolyticus* in the world (including China) have been completely uploaded into this public database. Only when a new ST is discovered, the researcher will sent the isolate information to the manager of the database to get identification for a new subtype. Therefore, the pubMLST database contains all the STs of *V. parahaemolyticus* all around the world, but does not contain all the discovered isolates. Here, we recommend that the database should be made some mandatory improvement, for example, when a new subtype is uploaded, the corresponding biological characteristics of isolates (sample source, resistance, serotype, virulence gene, etc.) should be also updated. We believe this will enable the pubMLST database to play a more important role in studies on infection and molecular epidemiology of *V. parahaemolyticus*.

In summary, we provide an overview of prevalence and genetic diversity of clinical *V. parahaemolyticus* spreading in Chinese population using MLST analysis. We implemented the identification of CCs and phylogenetic analysis both on nucleotide level and on peptide level. The results in this study will provide genetic evidences for uncovering the microevolution relationship among different pathogenic isolates of *V. parahaemolyticus*. The pubMLST database provides a platform for the comprehensive analysis of genetic relationships of *V. parahaemolyticus*. With the growing number of the uploaded isolates, more molecular biological characteristics of *V. parahaemolyticus* in china and other counties would be revealed.

## Acknowledgments

The authors thank Dr. Narjol Gonzalez-Escalona, who is from FDA, USA, for development of the *V. parahaemolyticus* MLST scheme. This study was jointly supported by grant NO.81400899 from the National Natural Science Foundation of China, grant NO.YZ2014061 from the Science and Technology Foundation Program of Yangzhou city, and a research project (NO.yzucms201425) of Northern Jiangsu People's Hospital, Yangzhou, China.

## Supplementary Material

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fmicb.2015.00291/abstract>

## References

- Ansaruzzaman, M., Lucas, M., Deen, J. L., Bhuiyan, N. A., Wang, X. Y., Safa, A., et al. (2005). Pandemic serovars (O3:K6 and O4:K68) of *Vibrio parahaemolyticus* associated with diarrhea in Mozambique: spread of the pandemic into the African continent. *J. Clin. Microbiol.* 43, 2559–2562. doi: 10.1128/JCM.43.6.2559-2562.2005

- Bag, P. K., Nandi, S., Bhadra, R. K., Ramamurthy, T., Bhattacharya, S. K., Nishibuchi, M., et al. (1999). Clonal diversity among recently emerged isolates of *Vibrio parahaemolyticus* O3:K6 associated with pandemic spread. *J. Clin. Microbiol.* 37, 2354–2357.
- Banerjee, S. K., Kearney, A. K., Nadon, C. A., Peterson, C. L., Tyler, K., Bakouche, L., et al. (2014). Phenotypic and genotypic characterization of Canadian clinical isolates of *Vibrio parahaemolyticus* collected from

- 2000 to 2009. *J. Clin. Microbiol.* 52, 1081–1088. doi: 10.1128/JCM.03047-13
- Chao, G., Jiao, X., Zhou, X., Yang, Z., Huang, J., Pan, Z., et al. (2009). Serodiversity, pandemic O3:K6 clone, molecular typing, and antibiotic susceptibility of foodborne and clinical *Vibrio parahaemolyticus* isolates in Jiangsu, China. *Foodborne Pathog. Dis.* 6, 1021–1028. doi: 10.1089/fpd.2009.0295
- Chao, G., Wang, F., Zhou, X., Jiao, X., Huang, J., Pan, Z., et al. (2011). Origin of *Vibrio parahaemolyticus* O3:K6 pandemic clone. *Int. J. Food Microbiol.* 145, 459–463. doi: 10.1016/j.ijfoodmicro.2011.01.022
- Chen, W., Xie, Y., Xu, J., Wang, Q., Gu, M., Yang, J., et al. (2012). Molecular typing of *Vibrio parahaemolyticus* isolates from the middle-east coastline of China. *Int. J. Food Microbiol.* 153, 402–412. doi: 10.1016/j.ijfoodmicro.2011.12.001
- Chowdhury, N. R., Chakraborty, S., Ramamurthy, T., Nishibuchi, M., Yamasaki, S., Takeda, Y., et al. (2000). Molecular evidence of clonal *Vibrio parahaemolyticus* pandemic isolates. *Emerg. Infect. Dis.* 6, 631–636. doi: 10.3201/eid0606.000612
- Ellis, C. N., Schuster, B. M., Striplin, M. J., Jones, S. H., Whistler, C. A., and Cooper, V. S. (2012). Influence of seasonality on the genetic diversity of *Vibrio parahaemolyticus* in New Hampshire shellfish waters as determined by multilocus sequence analysis. *Appl. Environ. Microbiol.* 78, 3778–3782. doi: 10.1128/AEM.07794-11
- Fan, Y. Y., Zhu, M., Shang, X. R., Wang, M., Huang, Y. F., Gu, H. T., et al. (2013). Virulence characteristics and multilocus sequence type of *Vibrio parahaemolyticus* isolated from clinic(in Chinese). *Chin. J. Epidemiol.* 36, 548–552. doi: 10.1371/journal.pone.0107371
- Gonzalez-Escalona, N., Martinez-Urtaza, J., Romero, J., Espejo, R. T., Jaykus, L. A., and DePaola, A. (2008). Determination of molecular phylogenetics of *Vibrio parahaemolyticus* isolates by multilocus sequence typing. *J. Bacteriol.* 190, 2831–2840. doi: 10.1128/JB.01808-07
- Han, D., Tang, H., Lu, J., Wang, G., Zhou, L., Min, L., et al. (2014). Population structure of clinical *Vibrio parahaemolyticus* from 17 coastal countries, determined through multilocus sequence analysis. *PLoS ONE* 9:e107371. doi: 10.1371/journal.pone.0107371
- Han, D. S., Chen, X., Li, Z. J., Zheng, S. F., Yu, F., and Chen, Y. (2012). Prevalence and multi -locus sequence typing of *Vibrio parahaemolyticus* associated with acute diarrhea in Zhejiang (in Chinese). *J. Environ. Health* 29, 917–919.
- Harth, E., Matsuda, L., Hernandez, C., Rioseco, M. L., Romero, J., Gonzalez-Escalona, N., et al. (2009). Epidemiology of *Vibrio parahaemolyticus* outbreaks, southern Chile. *Emerg. Infect. Dis.* 15, 163–168. doi: 10.3201/eid1502.071269
- Jolley, K. A., Feil, E. J., Chan, M. S., and Maiden, M. C. (2001). Sequence type analysis and recombinational tests (START). *Bioinformatics* 17, 1230–1231. doi: 10.1093/bioinformatics/17.12.1230
- Li, Y., Xie, X., Shi, X., Lin, Y., Qiu, Y., Mou, J., et al. (2014). *Vibrio parahaemolyticus*, Southern Coastal Region of China, 2007–2012. *Emerg. Infect. Dis.* 20, 685–688. doi: 10.3201/eid2004.130744
- Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452. doi: 10.1093/bioinformatics/btp187
- Lin, X. H., Ran, L., Ma, L., Wang, Z. J., and Feng, Z. J. (2011). Analysis on the cases of infectious diarrhea (other than cholera, dysentery, typhoid, and paratyphoid) reported in China in 2010(in Chinese). *Chin. J. Food Hyg.* 23, 385–389.
- Marshall, S., Clark, C. G., Wang, G., Mulvey, M., Kelly, M. T., and Johnson, W. M. (1999). Comparison of molecular methods for typing *Vibrio parahaemolyticus*. *J. Clin. Microbiol.* 37, 2473–2478.
- Martinez-Urtaza, J., Lozano-Leon, A., DePaola, A., Ishibashi, M., Shimada, K., Nishibuchi, M., et al. (2004). Characterization of pathogenic *Vibrio parahaemolyticus* isolates from clinical sources in Spain and comparison with Asian and North American pandemic isolates. *J. Clin. Microbiol.* 42, 4672–4678. doi: 10.1128/JCM.42.10.4672-4678.2004
- Martinez-Urtaza, J., Simental, L., Velasco, D., DePaola, A., Ishibashi, M., Nakaguchi, Y., et al. (2005). Pandemic *Vibrio parahaemolyticus* O3:K6, Europe. *Emerg. Infect. Dis.* 11, 1319–1320. doi: 10.3201/eid1108.050322
- Nair, G. B., Ramamurthy, T., Bhattacharya, S. K., Dutta, B., Takeda, Y., and Sack, D. A. (2007). Global dissemination of *Vibrio parahaemolyticus* serotype O3:K6 and its serovariants. *Clin. Microbiol. Rev.* 20, 39–48. doi: 10.1128/CMR.00025-06
- Okuda, J., Ishibashi, M., Hayakawa, E., Nishino, T., Takeda, Y., Mukhopadhyay, A. K., et al. (1997). Emergence of a unique O3:K6 clone of *Vibrio parahaemolyticus* in Calcutta, India, and isolation of isolates from the same clonal group from Southeast Asian travelers arriving in Japan. *J. Clin. Microbiol.* 35, 3150–3155.
- Osorio, J., Carvajal, A., Naharro, G., La, T., Phillips, N. D., Rubio, P., et al. (2012). Dissemination of clonal groups of *Brachyspira hyodysenteriae* amongst pig farms in Spain, and their relationships to isolates from other countries. *PLoS ONE* 7:e39082. doi: 10.1371/journal.pone.0039082
- Pal, D., and Das, N. (2010). Isolation, identification and molecular characterization of *Vibrio parahaemolyticus* from fish samples in Kolkata. *Eur. Rev. Med. Pharmacol. Sci.* 14, 545–549.
- Shi, X. L., Wang, Y., Hu, Q. H., Li, Y. H., Lin, Y. M., Qiu, Y. Q., et al. (2013). Molecular characterization of *Vibrio parahaemolyticus* collected from human infections in Shenzhen, between 2002 and 2008(in Chinese). *Chin. J. Epidemiol.* 34, 609–613.
- Smith, J. M., Smith, N. H., O'Rourke, M., and Spratt, B. G. (1993). How clonal are bacteria? *Proc. Natl. Acad. Sci. U.S.A.* 90, 4384–4388.
- Theethakaew, C., Feil, E. J., Castillo-Ramirez, S., Aanensen, D. M., Suthienkul, O., Neil, D. M., et al. (2013). Genetic relationships of *Vibrio parahaemolyticus* isolates from clinical, human carrier, and environmental sources in Thailand, determined by multilocus sequence analysis. *Appl. Environ. Microbiol.* 79, 2358–2370. doi: 10.1128/AEM.03067-12
- Turner, J. W., Paranjpye, R. N., Landis, E. D., Biryukov, S. V., Gonzalez-Escalona, N., Nilsson, W. B., et al. (2013). Population structure of clinical and environmental *Vibrio parahaemolyticus* from the Pacific northwest coast of the United States. *PLoS ONE* 8:e55726. doi: 10.1371/journal.pone.0055726
- Urmersbach, S., Alter, T., Koralage, M. S., Sperling, L., Gerdts, G., Messelhausser, U., et al. (2014). Population analysis of *Vibrio parahaemolyticus* originating from different geographical regions demonstrates a high genetic diversity. *BMC Microbiol.* 14:59. doi: 10.1186/1471-2180-14-59
- Yan, Y., Cui, Y., Han, H., Xiao, X., Wong, H. C., Tan, Y., et al. (2011). Extended MLST-based population genetics and phylogeny of *Vibrio parahaemolyticus* with high levels of recombination. *Int. J. Food Microbiol.* 145, 106–112. doi: 10.1016/j.ijfoodmicro.2010.11.038
- Yu, Y., Hu, W., Wu, B., Zhang, P., Chen, J., Wang, S., et al. (2011). *Vibrio parahaemolyticus* isolates from southeastern Chinese coast are genetically diverse with circulation of clonal complex 3 isolates since 2002. *Foodborne Pathog. Dis.* 8, 1169–1176. doi: 10.1089/fpd.2011.0865

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Han, Tang, Ren, Wang, Zhou and Han. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Sero-Prevalence and Genetic Diversity of Pandemic *V. parahaemolyticus* Strains Occurring at a Global Scale

Chongxu Han<sup>1†</sup>, Hui Tang<sup>2</sup>, Chuanli Ren<sup>1</sup>, Xiaoping Zhu<sup>1</sup> and Dongsheng Han<sup>1\*†</sup>

<sup>1</sup> Clinical Medical Examination Center, Northern Jiangsu People's Hospital, Yangzhou, China, <sup>2</sup> Experimental Research Center, Northern Jiangsu People's Hospital, Yangzhou, China

## OPEN ACCESS

### Edited by:

Julio Alvarez,  
University of Minnesota, USA

### Reviewed by:

Biao Kan,

National Institute for Communicable Disease Control and Prevention, China

Kai Zhou,

University Medical Center Groningen,  
Netherlands

### \*Correspondence:

Dongsheng Han  
hands1103@163.com

<sup>†</sup>These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Infectious Diseases,  
a section of the journal  
Frontiers in Microbiology

Received: 18 October 2015

Accepted: 05 April 2016

Published: 22 April 2016

### Citation:

Han C, Tang H, Ren C, Zhu X and  
Han D (2016) Sero-Prevalence and  
Genetic Diversity of Pandemic  
*V. parahaemolyticus* Strains Occurring  
at a Global Scale.  
Front. Microbiol. 7:567.  
doi: 10.3389/fmicb.2016.00567

Pandemic *Vibrio parahaemolyticus* is an emerging public health concern as it has caused numerous gastroenteritis outbreaks worldwide. Currently, the absence of a global overview of the phenotypic and molecular characteristics of pandemic strains restricts our overall understanding of these strains, especially for environmental strains. To generate a global picture of the sero-prevalence and genetic diversity of pandemic *V. parahaemolyticus*, pandemic isolates from worldwide collections were selected and analyzed in this study. After a thorough analysis, we found that the pandemic isolates represented 49 serotypes, which are widely distributed in 22 countries across four continents (Asia, Europe, America and Africa). All of these serotypes were detected in clinical isolates but only nine in environmental isolates. O3:K6 was the most widely disseminated serotype, followed by O3:KUT, while the others were largely restricted to certain countries. The countries with the most abundant pandemic serotypes were China (26 serotypes), India (24 serotypes), Thailand (15 serotypes) and Vietnam (10 serotypes). Based on MLST analysis, 14 sequence types (STs) were identified among the pandemic strains, nine of which fell within clonal complex (CC) 3. ST3 and ST305 were the only two STs that have been reported in environmental pandemic strains. Pandemic ST3 has caused a wide range of infections in as many as 16 countries. Substantial serotypic diversity was mainly observed among isolates within pandemic ST3, including as many as 12 combinations of O/K serotypes. At the allele level, the *dtdS* and *pntA*, two loci that perfectly conserved in CC3, displayed a degree of polymorphism in some pandemic strains. In conclusion, we provide a comprehensive understanding of sero-prevalence and genetic differentiation of clinical and environmental pandemic isolates collected from around the world. Although, further studies are needed to delineate the specific mechanisms by which the pandemic strains evolve and spread, the findings in this study are helpful when seeking countermeasures to reduce the spread of *V. parahaemolyticus* in endemic areas.

**Keywords:** *Vibrio parahaemolyticus*, multilocus sequence typing, pandemic clone, gastroenteritis, phylogenetic analysis

## INTRODUCTION

*Vibrio parahaemolyticus*, an organism with a high genetic diversity, has emerged as a pathogen causing acute gastroenteritis with a worldwide distribution. Before 1996, *V. parahaemolyticus* infections usually exhibited a localized distribution and were linked to diverse serotypes (e.g., O2:K3, O3:K6, O4:K8) (Wong et al., 2000). In February 1996, the pandemic O3:K6 serotype emerged, resulting in an inexplicable increase of *V. parahaemolyticus* gastroenteritis in Kolkata city, India (Okuda et al., 1997). This unique serotype subsequently quickly spread into coastal regions of southern Asia (Bag et al., 1999), America (Martinez-Urtaza et al., 2004), Africa (Ansaruzzaman et al., 2005), and Europe (Martinez-Urtaza et al., 2005) and caused numerous outbreaks within a few years (Okuda et al., 1997; Chowdhury et al., 2000a; Nair et al., 2007). Such widespread occurrence of a single serotype of *V. parahaemolyticus* had not been previously reported.

All the pandemic O3:K6 strains share the following specific genetic markers: positivity for the thermostable direct hemolysin(*tdh*) gene, negativity for the TDH-related hemolysin(*trh*) gene and positivity for a *toxRS/new* gene, which can be amplified via a specific PCR method known as “GS-PCR” (Matsumoto et al., 2000; Chao et al., 2011; de Jesús Hernández-Díaz et al., 2015). To our surprise, in recent years, some new serotypes [e.g., O4:K68, O1:K25, O1:KUT(untypable)] have been detected that exhibit identical genotypes and molecular profiles to the pandemic O3:K6 serotype (Chang et al., 2000; Bhuiyan et al., 2002). These serotypes may diverge from the pandemic O3:K6 serotype in alteration of the O and/or K antigens and are referred to as “serovariants” of the pandemic O3:K6 serotype (Chowdhury et al., 2000b; Matsumoto et al., 2000). Currently, all of the pandemic serotypes are grouped as belonging to the “O3:K6 pandemic clone.” Through 2007, a total of 22 serotypes had been reported to belong to this clone (Nair et al., 2007).

Many surveys have shown that pandemic *V. parahaemolyticus* serovariants can be identified not only in clinical samples (Li et al., 2014; Pazhani et al., 2014; Ueno et al., 2016), but also in seafood and other environmental samples (Arakawa et al., 1999; Vuddhakul et al., 2000; Deepanjali et al., 2005; Quilici et al., 2005; Chao et al., 2009; Caburlotto et al., 2010), indicating that the pandemic strains have established ecological niches in many regions, resulting in a heightened perception of the threat to the public health of the local population. An accurate description of the distribution and spread of the pandemic strains is important for understanding the epidemiology of this pathogen and preventing outbreaks and sporadic illnesses. However, after G. Balakrish Nair and colleagues reviewed the global dissemination of pandemic *V. parahaemolyticus* serotype O3:K6 and its serovariants in 2007 (Nair et al., 2007), few studies have specifically focused on the pandemic *V. parahaemolyticus* on a global scale, especially concerning the worldwide dissemination of environmental strains. Therefore, it would be beneficial to integrate and update the available scientific data on pandemic *V. parahaemolyticus*.

The establishment of a multilocus sequence typing (MLST) scheme for *V. parahaemolyticus* has enhanced our knowledge of the population structure and genetic diversity of *V. parahaemolyticus* (Gonzalez-Escalona et al., 2008). Previous studies based on MLST assay have shown that the increasing prevalence of clonal complex 3 (CC3) has become an ongoing public health concern (Gonzalez-Escalona et al., 2008; Haendiges et al., 2014; Han et al., 2015), and most pandemic strains have been identified as belonging to CC3 (Chen et al., 2016). Thus, clarifying the genetic diversity among the pandemic strains will aid in the selection of preventative strategies targeting pandemic strain infections.

In this study, we collected data on pandemic strains mainly from the pubMLST database (<http://pubmlst.org/vparahaemolyticus>) and previous studies, in an effort to generate a comprehensive overview of the spread of clinical and environmental pandemic *V. parahaemolyticus* strains occurring over wide geographic areas since the emergence of this clone. Furthermore, through MLST phylogenetic analysis, we determined the genetic diversity of the pandemic clone to provide a holistic understanding of the microevolution of pandemic strains.

## MATERIALS AND METHODS

### Datasets Utilized in the Present Study

A total of 267 representative clinical and environmental *V. parahaemolyticus* isolates with pandemic genetic marks (*toxRS/new+*, *tdh+*, and *trh-*) were selected as the research subject of this study, among which 263 isolates came from the literature and four from the pubMLST database (<http://pubmlst.org/vparahaemolyticus/>). To identify relevant publications, we conducted a comprehensive search of the US National Library of Medicine PubMed database and the Elsevier, Springer, and China National Knowledge Infrastructure databases for all relevant studies using combinations of the following terms: “*Vibrio parahaemolyticus*,” “pandemic clone,” “pandemic strains,” and “O3:K6 clone” (until July 1, 2015). Additional eligible studies were identified from references cited in the relevant articles. The full text of each potentially relevant paper was scrutinized and a total of 263 isolates with pandemic genetic marks (*toxRS/new+*, *tdh+*, and *trh-*) were finally extracted from 39 papers. Details on the individual isolates are summarized in Additional file 1: Table S1.

### Multilocus Sequence Typing Analysis

To determine the genetic diversity of the pandemic strains through MLST analysis, another file containing 185 isolates were analyzed (see Additional file 2: Table S2). These isolates included 95 pandemic isolates from Table S1 and 90 strains that are non-pandemic but belong to CC3 from the pubMLST database. The sequence types (STs) of the 185 isolates were compared using eBURST V3 (<http://eburst.mlst.net/>) (Feil et al., 2004). Additionally, a “population snapshot” analysis of the entire *V. parahaemolyticus* population was also implemented based on the total pubMLST dataset, which illustrated the

**TABLE 1 | Presence of clinical and environmental pandemic serovariants of *V. parahaemolyticus* occurring at a global scale.**

Serotype	Country (year of isolation)
<b>CLINICAL(49 SEROTYPES):</b>	
O1:K25	Bangladesh(1999), China(1998, 2005–2012), India(2004–2010), Thailand(1999–2010), Vietnam(1998–1999), Japan (1998)
O1:K26	China(2007)
O1:K30	India(2011)
O1:K33	India(2002)
O1:K36	China(2006–2012)
O1:K38	India(2001)
O1:K41	Thailand(1998–1999, 2002), Vietnam(1998–1999)
O1:K5	China(2007, 2009)
O1:K56	China(2008), India(2011), Vietnam(1998–1999)
O1:KUT*	Bangladesh(1998, 2000), China(2003, 2005–2012), India(1998, 2001), Peru(2005, 2007), Thailand(2001–2010)
O10:K60	China(2010–2012), Japan(2013), India(2012)
O10:KUT	Mexico(2004–2010)
O11:K36	China(2007)
O2:K3	China(2010–2012), Thailand(2000), India(2002)
O2:K4	India(2005)
O3:K25	China(2007)
O3:K29	China(2007), Thailand(2002–2003), Mexico(2011–2013)
O3:K3	China(2010–2012)
O3:K30	Peru(2007)
O3:K46	Thailand(2001, 2004)
O3:K5	India(2004)
O3:K58	Peru(1998–1999)
O3:K59	Chile(2007)
O3:K6	Angola(1999), Bangladesh(1998), Brazil(2002), Chile(1998, 2007), China(1996–2012), Ecuador(1999), France(2004), India(1996), Indonesia(1997), Italy(2007–2008), Japan(1996, 1998), Korea(1997–1998), Laos(1997), Mexico(2004–2013), Mozambique(2004), Peru(1996–2003), Russia(2012), Singapore(1996, 1998), Thailand(1996–1997, 2000–2010), Vietnam(1997), Spain(2004), USA(1998, 2012)
O3:K6,59	Chile(2007)
O3:K68	China(2006), Peru(1998)
O3:K75	Vietnam(1998–1999)
O3:K8	China(2009–2011)
O3:KUT	Brazil(2002), China(2010–2012), India(2003–2004), Peru(2007), Spain(2004), Mexico(2004–2010), Thailand(2004, 2006–2010)
O4:K10	India(2004)
O4:K12	Chile(2004), Thailand(1998–1999), Mexico(2004–2010), Vietnam(1998–1999)
O4:K13	India(2010)
O4:K25	India(2005)
O4:K4	India(2004), Thailand(2005)
O4:K48	China(2005–2008)
O4:K55	India(2001)
O4:K68	Bangladesh(1998), China(1999, 2003, 2005–2012), India(1998–1999), Mozambique(2004), Singapore(1998), Vietnam(1998), Thailand(1999–2001, 2003–2005)
O4:K8	China(2006–2010), Thailand(2006–2010), Vietnam(1998–1999)
O4:K9	Thailand(2006–2010)

(Continued)

**TABLE 1 | Continued**

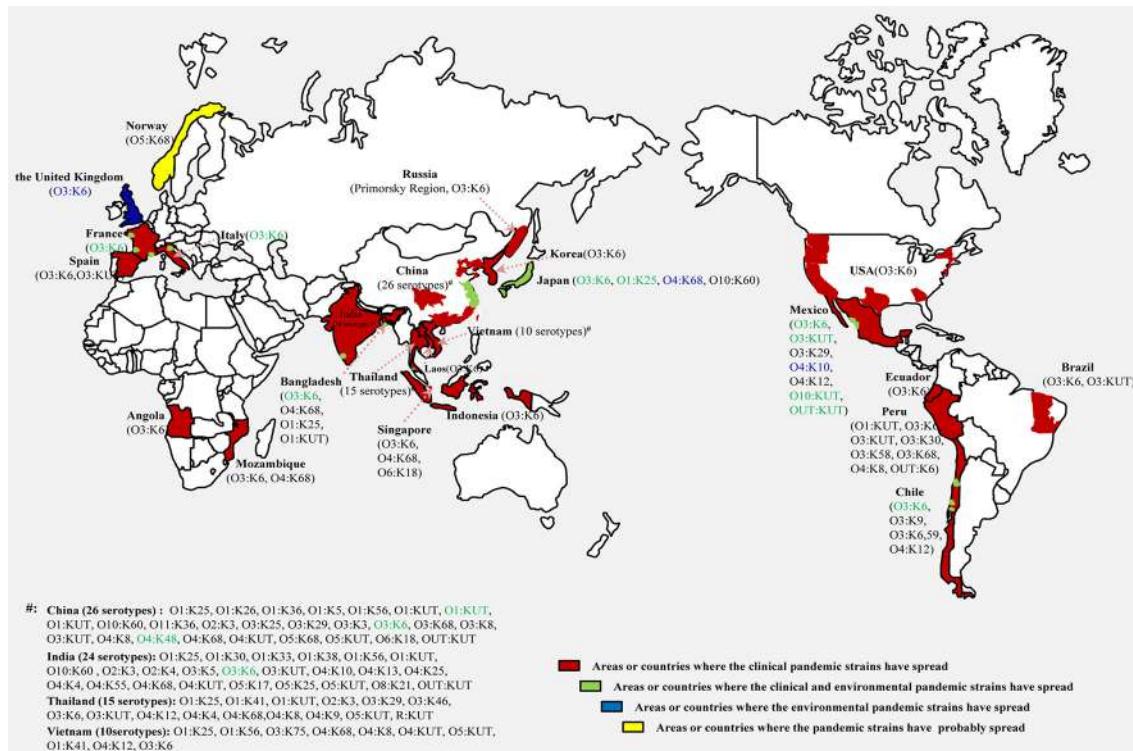
Serotype	Country (year of isolation)
O4:KUT	China(2006–2007), India(2009), Vietnam(1998–1999)
O5:K17	India(2002)
O5:K25	India(2002)
O5:K68	China(2007–2012), Norway(2002) <sup>#</sup>
O5:KUT	China(2010–2012), India(2004), Vietnam(1998–1999), Thailand(2003)
O6:K18	China(2005), Singapore(1998)
O8:K21	India(2006)
OUT:K6	Peru(1998)
O10:KUT	Mexico(2011–2013)
OUT:KUT	China(2007), India(2003–2004), Mexico(2004–2013)
<b>ENVIRONMENT (9 SEROTYPES):</b>	
O3:K6	Chile(2008–2009), China(2005–2008), Italy(2007), France(1997, 1998, 1999), Mexico(2004–2013), the UK(2012), Bangladesh(2007), Japan(1998, 2000, 2001), India(2002)
O3:KUT	Mexico(2011–2013)
O4:K48	China(2005–2008)
O1:KUT	China(2005–2008)
O1:K25	Japan(2001)
O4:K10	Mexico(2012)
O4:K68	Japan(2001)
O10:KUT	Mexico(2004–2010)
OUT:KUT	Mexico(2004–2010)

<sup>\*</sup>UT, untyped.<sup>#</sup>There was no sufficient evidences to determine the isolate of pandemic O5:K68 was really collected from Norway in 2002.

population differentiation of the whole population. Clonal complexes were defined conservatively as a cluster of STs in an eBURST diagram, in which all STs were linked as single-locus variants (SLVs, two STs differing from each other at a single locus) to at least one other ST (Feil et al., 2004). The singleton STs corresponded to STs differing from all the others by three or more of the seven loci (Esteves et al., 2015).

## Genetic Diversity and Phylogenetic Analysis

The diversity of the seven loci in the pandemic isolates was revealed by DnaSP V5 (<http://www.ub.edu/dnasp/>) with respect to the following parameters: the number of alleles, number (%) of polymorphic sites, nucleotide diversity (per site) and Tajima's D value. The purpose of Tajima's D test is to distinguish housekeeping genes evolving randomly ("neutrally") vs. those evolving under a non-random process (Tajima, 1989). A  $P > 0.05$  indicates that the target gene is evolving randomly and that mutations in the gene have no effect on the fitness and survival of an organism (Tajima, 1989; Ferreira et al., 2008). A minimum-evolution (ME) tree for the concatenated sequences of each ST of the 185 isolates was generated using Mega 5 software with the Kimura two-parameter model to estimate genetic distances. The statistical support for the



nodes in the ME tree was assessed through 1000 bootstrap resamplings.

## RESULTS

### Global Spread of Pandemic Serovariants

According to a detailed review, a total of 49 pandemic serotypes from 22 countries across four continents (Asia, Europe, America, and Africa) were identified. All of these serotypes were detected in clinical isolates but only nine in environmental isolates. O3:K6 was the most widely disseminated serotype, and patients in all 22 countries had been infected with this subtype at some point in time. O3:KUT was the second most widely distributed serotype. Several serotypes, such as O1:K25, O1:KUT, and O4:K68, also exhibited multi-country distributions but were mainly restricted to Southeast Asia (Table 1). The sources of environmental pandemic isolates were diverse, mainly including shellfish, oyster, clam, and shrimp, sediment and seawater samples collected in nine countries (see in Additional file 1: Table S1).

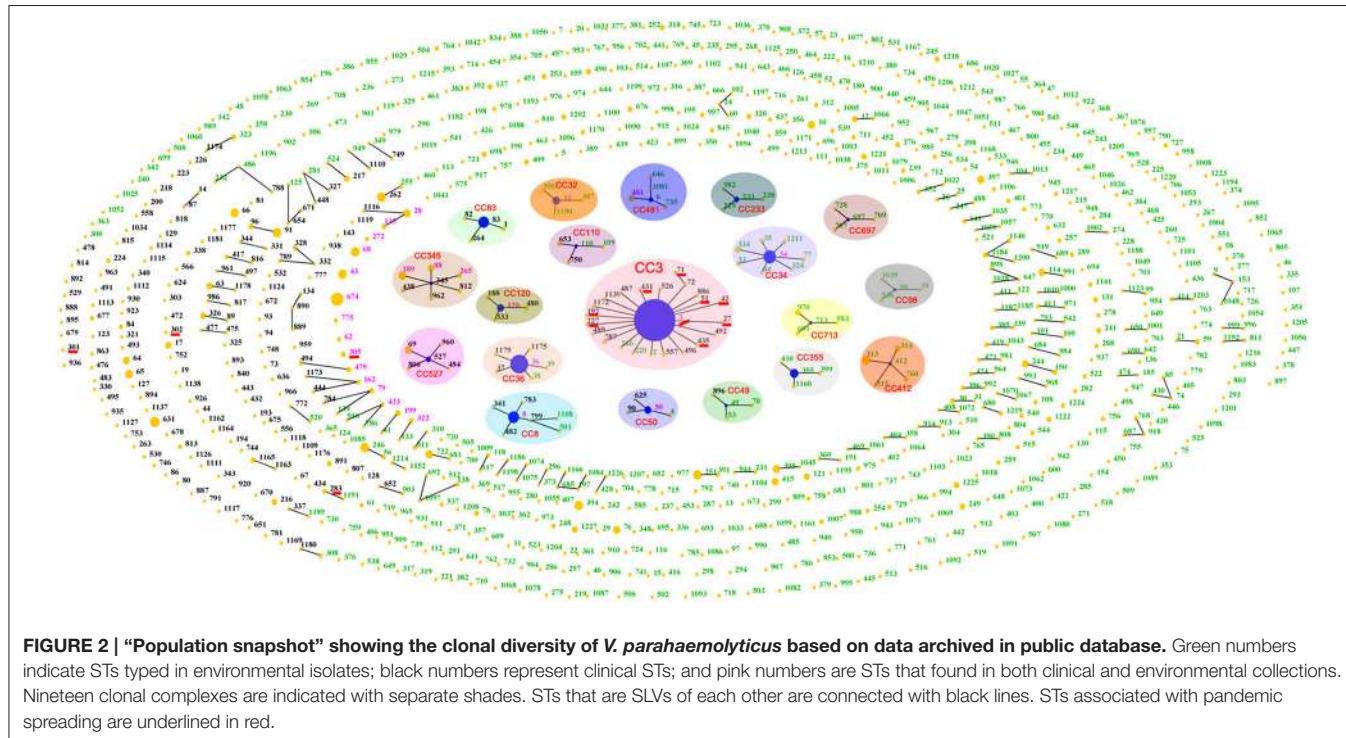
A comprehensive map of the dissemination of the clinical and environmental pandemic serotypes on a global scale was generated (Figure 1). The serotypes of the pandemic clone were highly abundant and variable in coastal regions of China, India, Thailand and Vietnam. It was notable that most of the environmental pandemic serotypes present in a certain country were also detected in patients from that country. O3:K6 was the

typical serotype. Four environmental serotypes (O3:K6, O3:KUT, O10:KUT, and OUT:KUT) in Mexico were also found spread in its local population.

### Widely Dispersed Clones of *V. parahaemolyticus* and Genetic Differentiation of the Pandemic Isolates

Until August 2015, a total of 954 STs had been identified in the *V. parahaemolyticus* pubMLST database, approximately two-thirds of which were detected in environmental isolates, while less than one-third came from clinical isolates, and only 26 were present both in environmental and clinical isolates. The total population displayed 19 CCs as well as some doublets and numerous singletons (Figure 2). CC3 was the most prevalent CC, being comprised of 18 STs with no less than 15 serotypes (Table 2).

After thoroughly analyzing the sequence data for the 185 isolates (in Additional file 2: Table S2), we found that the pandemic strains exhibited 14 STs, only two of which (ST3 and ST305) had ever been identified in environmental samples (Figure 3C). China was the country with the most pandemic STs (10 STs). Nine of these 14 pandemic STs could be classified into CC3, among which, ST3 was the only pandemic ST that had caused a wide range of infections in as many as 16 countries (Table 2, Figure 3A). ST305 and ST672 were DLVs of ST3 but were not members of CC3 because there was



no ST in CC3 could act as their SLV. The other three STs (ST283, ST301, and ST302) originating from coastal areas of China were identified as singletons with no relationship to CC3.

## The Association between Pandemic STs and Serotypes

The Minimum spanning tree of the 14 pandemic STs resulting from the MLST analysis showed a substantial serotypic diversity among isolates within ST3, but not among isolates of the other 13 STs (Figure 3B). Specifically, pandemic ST3 comprised isolates of 12 serotypes (O1:K25, O1:K36, O1:KUT, O3:K6, O3:K25, O3:K30, O3:K58, O3:K68, O3:KUT, O4:K68, O5:K68, and O11:K36). ST305 included isolates belonging to O1:K25 and O1:KUT serotypes. The remaining STs were consisted of a single serotype, respectively (Table 2, Figure 3B). From another perspective, the pandemic O3:K6 serotype was shared by six different STs (ST3, ST27, ST42, ST71, ST435, and ST672). Other serotypes were clustered in no more than two different pandemic STs, respectively (Table 2).

## Genetic Diversity of the Pandemic Isolates

The data on the nucleotide and allelic diversity of the pandemic isolates are summarized in Table 3. The highest percentage of polymorphic sites was detected in *dtdS* (5.46%). Nucleotide diversity ranged from 0.01082 (*pyrC*) to 0.02926 (*dtdS*). *dtdS* and *pntA* were perfectly conserved in CC3 (the allele types were *dtdS4* and *pntA29*), but in the pandemic isolates, five different alleles were detected for each of the two genes; the number of SNPs was 25(5.46%) for *dtdS* and 10(2.33%) for *pntA* (Table 3).

## Phylogenetic Analysis of Pandemic Isolates

Phylogenetic analysis may provide a better resolution and elucidate some phylogenetic relationships among CCs or singletons that are not observed or resolved using goeBURST. Therefore, an ME tree representing the concatenated sequences of the seven housekeeping gene fragments in the 185 isolates is shown in Figure 4. In the goeBURST analysis, five pandemic STs (ST305, S672, ST301, ST302, and ST283) were not grouped into CC3. However, in the ME tree analysis, ST305 and ST672 were clustered together with STs of CC3, and only ST301, ST302, and ST283 exhibited relatively greater evolutionary distances from STs in CC3. In fact, the number of SNPs in the seven alleles of these last STs was greater than in ST305 and ST672 when compared with the STs of CC3.

## DISCUSSION

In previous studies, we successfully made extensive descriptions of strains from a global clinical collection and from Chinese patients, respectively, exhibiting a highly degree of genetic diversity and a complicated population structure of *V. parahaemolyticus* in general (Han et al., 2014, 2015). In this study, we elucidated the sero-prevalence and genetic differentiation of the pandemic clone, which has becoming an emerging public health concern (Martinez-Urtaza et al., 2010; Velazquez-Roman et al., 2012, 2014; Powell et al., 2013; Li et al., 2014; Pazhani et al., 2014). The results will be useful in uncovering the microevolution relationships among pandemic *V. parahaemolyticus* strains. Serotyping is

**TABLE 2 | Sequence types, allele profiles, and geographic locations of CC3 and pandemic *V. parahaemolyticus* clone.**

Clone	ST	Allele types							Collection countries (serotypes)		
		CC3	Pandemic	dnaE	dtdS	gyrB	pntA	pyrC	recA	ttaA	clinical
Yes	No	2	11 <sup>&amp;</sup>	4	4	29	4	19	22	—	China(O3:K6)
Yes	No	72	3	4	4	29	4	4	22	Thailand(O3:KUT)	—
Yes	No	220	3	4	4	29	4	15	22	—	China(O3:K6)
Yes	No	266	3	4	4	29	4	19	78	—	China(O1:K33)
Yes	Unknown	557	3	4	4	29	4	19	152	China(Unknown)	—
Yes	Unknown	787	3	4	4	29	48	19	22	China(O4:K68)	—
Yes	Unknown	886	51	4	4	29	4	19	22	China(Unknown)	—
Yes	Unknown	1139	3	4	415	29	4	19	22	Mexico(O3:K6)	—
Yes	Unknown	1172	3	4	23	29	4	19	22	Chile(Unknown)	—
Yes	Yes	3	3	4	4	29	4	19	22	*	#
Yes	Yes	27	17	4	4	29	4	19	22	Korea(O3:K6)	—
Yes	Yes	42	22	4	4	29	4	19	22	USA(O3:K6)	—
Yes	Yes	51	29	4	4	29	4	19	22	Bangladesh(O3:KUT)	—
Yes	Yes	71	3	4	4	29	4	4	20	Ecuador(O3:K6)	—
Yes	Yes	192	3	4	126	29	4	19	22	China(O1:K26)	—
Yes	Yes	227	3	4	4	29	22	19	22	China(Unknown)	—
Yes	Yes	431	3	4	225	29	4	19	22	China(Unknown)	—
Yes	Yes	435	3	4	4	29	4	31	22	China(O3:K6)	—
No	Yes	283	27	84	127	139	54	124	37	China(O4:K8)	—
No	Yes	301	140	167	136	151	50	135	17	China(O4:KUT)	—
No	Yes	302	27	106	127	152	54	124	101	China(O4:KUT)	—
No	Yes	305	3	147	4	93	4	19	22	China(O1:K25)	China(O1:KUT)
No	Yes	672	1	4	147	29	4	19	22	China(O3:K6)	—

\*Bangladesh(O1:K25,O1:KUT,O3:K6,O4:K68), Chile(O3:K6), China(O1:K25, O1:K36,O1:KUT,O11:K36,O3:K25,O3:K6,O3:K68,O4:K68), Ecuador(O3:K6), India(O1:KUT,O3:K6,O4:K68), Indonesia(O3:K6), Japan(O1:K25,O3:K6), Korea(O3:K6), Mexico(O3:K6), Mozambique(O3:K6,O4:K68), Norway(O5:K68), Peru(O1:KUT,O3:K30,O3:K58,O3:K6,O3:KUT), Singapore(O3:K6,O4:K68), Spain(O3:K6), Thailand(O1:K25,O3:KUT,O4:K68), USA(O3:K6).

<sup>#</sup>Chile(O3:K6), China(O1:KUT). These isolates belong to the pandemic clone. Others can't be determined as pandemic isolates but typed as ST3 were not listed here, see them in Additional file 2: Table S2.

<sup>&</sup>Compared with ST3, the changed allele types in other STs in bold.

**TABLE 3 | Sequence analysis of the seven loci studies for the isolates of CC3 and pandemic clone.**

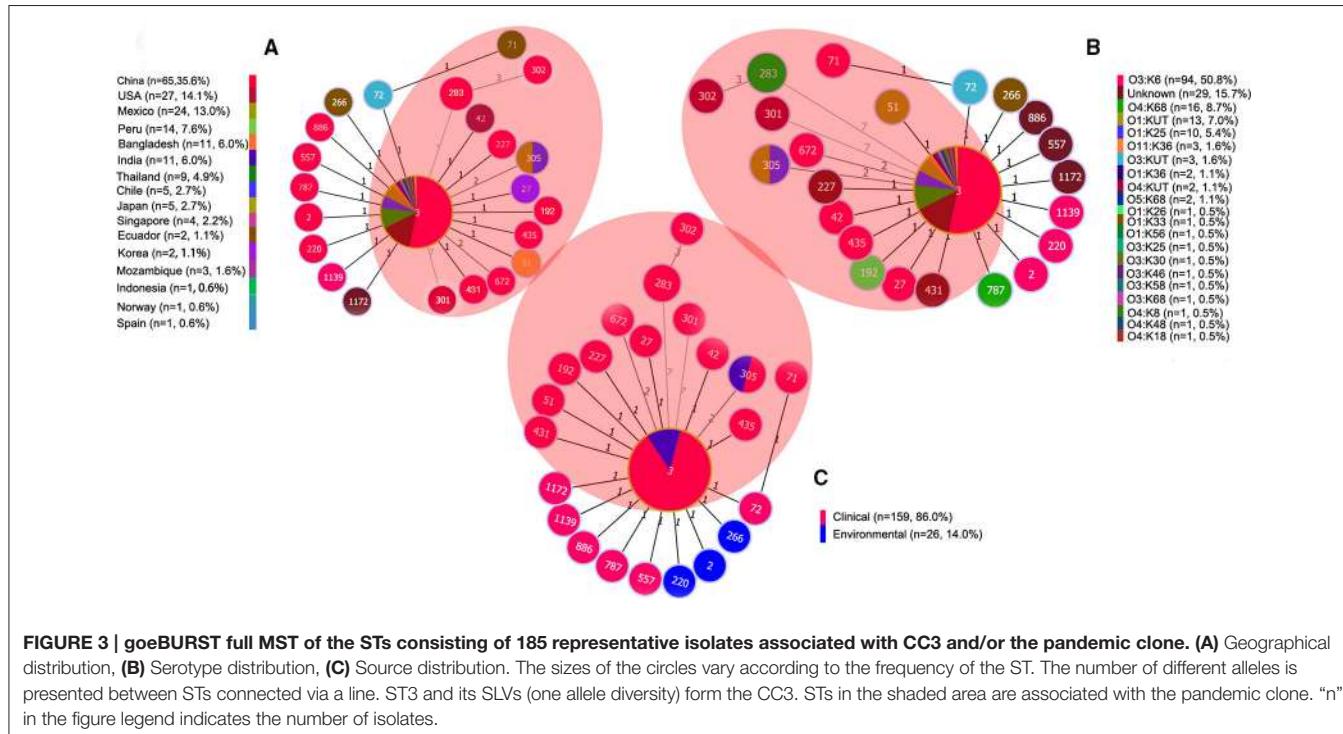
Locus	Fragment size	No of alleles		Number (%) of polymorphic sites		Nucleotide diversity (per site)		Tajima's D (P-Value)	
		CC3	Pandemic Clone	CC3	Pandemic Clone	CC3	Pandemic Clone	CC3	Pandemic Clone
dnaE	557	6	7	18 (3.23)	19 (3.41)	0.01352	0.01368	-0.27603 ( $P > 0.10$ )	-0.09825 ( $P > 0.10$ )
gyrB	592	5	6	10(5.92)	18(3.04)	0.00676	0.01261	-1.19267 ( $P > 0.10$ )	-0.32862 ( $P > 0.10$ )
recA	729	4	5	28(3.84)	26(3.57)	0.01920	0.02058	-0.86044 ( $P > 0.10$ )	1.50445 ( $P > 0.10$ )
dtdS	458	1	5	—	25(5.46)	—	0.02926	—	0.86832 ( $P > 0.10$ )
pntA	430	1	5	—	10 (2.33)	—	0.01209	—	0.59633 ( $P > 0.10$ )
pyrC	493	3	4	8(1.62)	10 (2.03)	0.01082	0.01082	—	-0.22234 ( $P > 0.10$ )
ttaA	423	4	5	11 (2.60)	15 (3.55)	0.01300	0.01608	-0.83741 ( $P > 0.10$ )	-0.40617 ( $P > 0.10$ )

Four or more sequences are needed to compute the Tajima's test.

the primary basis of the classification of *V. parahaemolyticus* strains. Pandemic strains exhibit rapidly changing their serotypes (Nair et al., 2007). From 1996 to 2007, 22 pandemic serotypes were identified (Nair et al., 2007). In the present study, as many as 49 serotypes identified to date in investigations

conducted by different laboratory groups around the world could be confirmed as being associated with the pandemic clone.

Several lines of evidence have been presented in support of the hypothesis that these new serotypes might have emerged



from the pandemic O3:K6 strains through replacement of the putative O and K antigen gene clusters (Okura et al., 2008; Harth et al., 2009). In the present study, as many as 12 combinations of O/K serotypes were grouped in pandemic ST3, demonstrating a remarkably high degree of serotypic diversity among the pandemic isolates and suggesting that the O- and K-antigen encoding loci are subject to exceptionally high rates of recombination in isolates with the same genotype (Gavilan et al., 2013; Theethakaew et al., 2013). Herein, we agree that the high frequency of alterations in the O and/or K antigens is a significant biological characteristic of pandemic *V. parahaemolyticus* strains, which might be an important means of survival in the face of changing external environments and host immunological resistance.

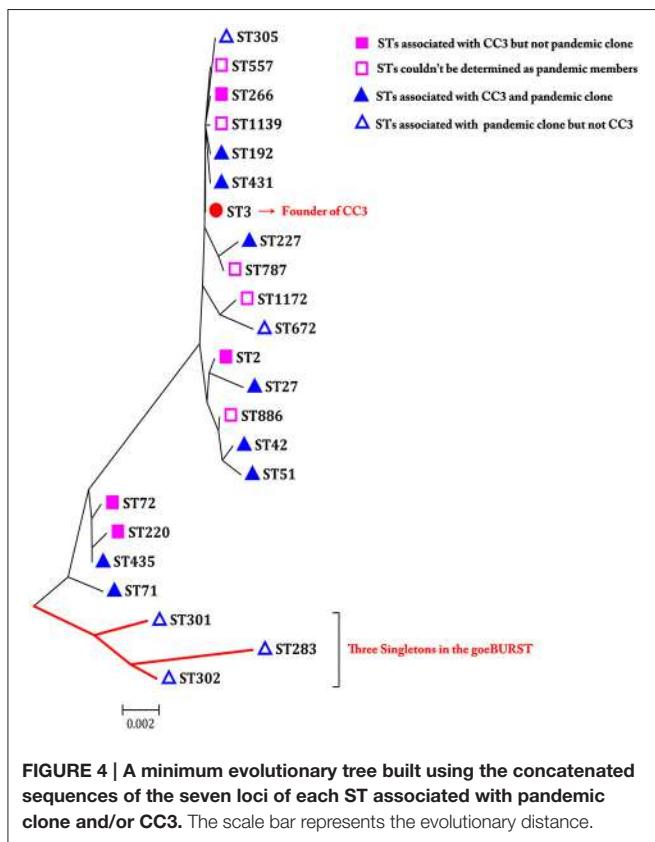
Regional persistence of the clinical pandemic O3:K6 serotype has been identified in coastal areas of many countries, such as Mexico(2004–2015) (Velazquez-Roman et al., 2012; de Jesús Hernández-Díaz et al., 2015), Peru(2007) (Gil et al., 2007), Chile(2007–2009) (Cabello et al., 2007; Garcia et al., 2009), China(2007–2012) (Zhang et al., 2013; Li et al., 2014), India(2001–2012) (Pazhani et al., 2014), and Thailand(2006–2010) (Thongjun et al., 2013). However, it is not obvious what specific factors conferred upon this serotype the ability to disseminate around world. Some environmental conditions (e.g., seawater temperature, PH or salinity effects) affecting survival and the unique pathogenic potential of pandemic O3:K6 strains vs. other strains have been compared, but the specific advantage of pandemic O3:K6 strains over other strains remains unclear (Wong et al., 2000; Yeung et al., 2002). Further, investigations should focus on revealing

the routes and mechanisms of the rapid spread of the pandemic clone.

In addition to the O3:K6 serotype, other pandemic serotypes have been isolated in both clinical and environmental samples from some certain countries, such as O1:KUT and O4:K48 in China (Chao et al., 2009), O1:K25 in Japan (Hara-Kudo et al., 2003) and O3:KUT, O10:KUT, and OUT:OUT in Mexico (Velazquez-Roman et al., 2012; de Jesús Hernández-Díaz et al., 2015). Although, the specific relationships between environmental serotypes and those leading to illnesses have not been determined, it is important to first understand epidemic situation of these serotypes through active surveillance.

In the present study, we showed that the population structure of *V. parahaemolyticus* was extremely genetically diverse based on the successful identification of 19 CCs and a large number of singletons, in agreement with previous findings (Han et al., 2015). Over half of the pandemic STs belonged to CC3 according to goeBURST analysis. The *dtdS* and *pntA* genes were found to be perfectly conserved throughout the evolution of CC3, whereas they presented some degree of polymorphism in pandemic strains. In our analysis, none of the values of Tajima’s D was significantly different from zero ( $P > 0.10$ ), suggesting that the housekeeping genes of the pandemic strains evolve under a random process (“neutrally”) and are subject to low selective pressure. The similar conclusion was obtained in studies based on the entire *V. parahaemolyticus* population (Theethakaew et al., 2013).

According to the available data, 64.3% of the STs (9/14) of the pandemic clones were isolated from China, suggesting that this country represents an important reservoir for the emergence of



**FIGURE 4 |** A minimum evolutionary tree built using the concatenated sequences of the seven loci of each ST associated with pandemic clone and/or CC3. The scale bar represents the evolutionary distance.

novel pandemic strains. If a global network for the prevention and control of *V. parahaemolyticus* infection is established in the future, the coastal regions of China should be recognized as important monitoring points. Three special STs (ST283, ST301, and ST302) typed in pandemic isolates originating from China were identified as singletons presenting distant relationships with other STs of the pandemic clone in this study. However, in a study by Chen et al. (2012), the corresponding strains were clustered together with other pandemic strains based on other molecular typing methods, such as enterobacterial repetitive intergenic consensus sequence PCR (ERIC-PCR) and sequence analysis of the *gyrB* gene. Thus, it can be observed that current molecular typing methods, including MLST, could lead to controversial results, making it difficult to draw conclusions, although such methods have been confirmed to provide a high level of resolution and information for elucidating the evolution of the *V. parahaemolyticus* clonal complex (Chen et al., 2012). Therefore, to accurately portray the relationships among strains at the molecular level, combined use of different molecular typing techniques with better discrimination could be considered in epidemiological investigations of *V. parahaemolyticus*. Whole genome sequencing (WGS), a powerful typing method with a robust differentiation ability for characterizing related isolates, is another outstanding alternative for analyzing the evolution and population structure of *V. parahaemolyticus* (Haendiges et al., 2015).

Invalid data in the pubMLST database were one problem restricting our analysis in this study. As of 15th July 2015, a total of 1844 records of isolates had been deposited, but definite STs were only available for 1700. Moreover, information on the corresponding biological characteristics of many uploaded isolates, such as sample sources, regions, drug sensitivity, serotypes and virulence genes was deficient. This lack of information is not conducive to conducting further epidemiologic and etiologic analyses of *V. parahaemolyticus* at a global scale. In this study, for five STs belonging to CC3 (ST557, ST787, ST886, ST1139, and ST1172), it could not be determined whether they were associated with pandemic clone, because of missing of *toxRS/new* gene and/or *tdh* gene sequences. As MLST assays play an important role in studies on the molecular epidemiology of *V. parahaemolyticus*, we recommend that the researchers uploaded their data on isolates as accurately and completely as possible.

In summary, the present study provides novel information on the abundance and prevalence of pandemic *V. parahaemolyticus* based on the analysis of clinical and environmental isolates from a worldwide collection. We showed that the regional persistence of pandemic O3:K6 has been established in coastal areas of many countries. The presence and persistence of pandemic *V. parahaemolyticus* strains, and especially the continuous appearance of environmental pandemic strains, is a matter of concern for public health authorities. We analyzed the genetic diversity of the pandemic clone to provide a comprehensive understanding of the microevolutionary relationships between pandemic strains. The answers to some unresolved questions about the pandemic clone, such as the advantage of pandemic O3:K6 over other strains and the mechanisms underlying the spread of strains with pandemic genetic marks, remain speculative, and require further investigations.

## AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: DH, HT, CH. Performed the experiments: DH, CR. Analyzed the data: DH, HT, XZ. Contributed reagents/materials/analysis tools: DH, CH. Wrote the paper: DH.

## ACKNOWLEDGMENTS

This study was jointly supported by grant NO.81400899 from the National Natural Science Foundation of China, grant NO.YZ2014061 from the Science & Technology Foundation Program of Yangzhou city, and two research projects (NO.yzucms201425 and NO.yzucms201403) of Northern Jiangsu People's Hospital, Yangzhou, China.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2016.00567>

## REFERENCES

- Ansaruzzaman, M., Lucas, M., Deen, J. L., Bhuiyan, N. A., Wang, X. Y., Safa, A., et al. (2005). Pandemic serovars (O3:K6 and O4:K68) of *Vibrio parahaemolyticus* associated with diarrhea in Mozambique: spread of the pandemic into the African continent. *J. Clin. Microbiol.* 43, 2559–2562. doi: 10.1128/JCM.43.6.2559-2562.2005
- Arakawa, E., Murase, T., Shimada, T., Okitsu, T., Yamai, S., and Watanabe, H. (1999). Emergence and prevalence of a novel *Vibrio parahaemolyticus* O3:K6 clone in Japan. *Jpn. J. Infect. Dis.* 52, 246–247.
- Bag, P. K., Nandi, S., Bhadra, R. K., Ramamurthy, T., Bhattacharya, S. K., Nishibuchi, M., et al. (1999). Clonal diversity among recently emerged strains of *Vibrio parahaemolyticus* O3:K6 associated with pandemic spread. *J. Clin. Microbiol.* 37, 2354–2357.
- Bhuiyan, N. A., Ansaruzzaman, M., Kamruzzaman, M., Alam, K., Chowdhury, N. R., Nishibuchi, M., et al. (2002). Prevalence of the pandemic genotype of *Vibrio parahaemolyticus* in Dhaka, Bangladesh, and significance of its distribution across different serotypes. *J. Clin. Microbiol.* 40, 284–286. doi: 10.1128/JCM.40.1.284-286.2002
- Cabello, F. C., Espejo, R. T., Hernandez, M. C., Rioseco, M. L., Ulloa, J., and Vergara, J. A. (2007). *Vibrio parahaemolyticus* O3:K6 epidemic diarrhea, Chile, 2005. *Emerg. Infect. Dis.* 13, 655–656. doi: 10.3201/eid1304.061152
- Caburlotto, G., Gennari, M., Ghidini, V., Tafi, M., and Lleo, M. M. (2010). Serological and molecular characterization of *Vibrio parahaemolyticus* marine strains carrying pandemic genetic markers. *ISME J.* 4, 1071–1074. doi: 10.1038/ismej.2010.34
- Chang, B., Yoshida, S., Miyamoto, H., Ogawa, M., Horikawa, K., Ogata, K., et al. (2000). A unique and common restriction fragment pattern of the nucleotide sequences homologous to the genome of vf33, a filamentous bacteriophage, in pandemic strains of *Vibrio parahaemolyticus* O3:K6 O4:K68, and O1:K untypeable. *FEMS Microbiol. Lett.* 192, 231–236. doi: 10.1111/j.1574-6968.2000.tb09387.x
- Chao, G., Jiao, X., Zhou, X., Yang, Z., Huang, J., Pan, Z., et al. (2009). Serodiversity, pandemic O3:K6 clone, molecular typing, and antibiotic susceptibility of foodborne and clinical *Vibrio parahaemolyticus* isolates in Jiangsu, China. *Foodborne Pathog. Dis.* 6, 1021–1028. doi: 10.1089/fpd.2009.0295
- Chao, G., Wang, F., Zhou, X., Jiao, X., Huang, J., Pan, Z., et al. (2011). Origin of *Vibrio parahaemolyticus* O3:K6 pandemic clone. *Int. J. Food Microbiol.* 145, 459–463. doi: 10.1016/j.ijfoodmicro.2011.01.022
- Chen, W., Xie, Y., Xu, J., Wang, Q., Gu, M., Yang, J., et al. (2012). Molecular typing of *Vibrio parahaemolyticus* isolates from the middle-east coastline of China. *Int. J. Food Microbiol.* 153, 402–412. doi: 10.1016/j.ijfoodmicro.2011.12.001
- Chen, Y., Chen, X., Yu, F., Wu, M., Wang, R., Zheng, S., et al. (2016). Serology, virulence, antimicrobial susceptibility and molecular characteristics of clinical *Vibrio parahaemolyticus* strains circulating in southeastern China from 2009 to 2013. *Clin. Microbiol. Infec.* 22, 258.e9–258.e16. doi: 10.1016/j.cmi.2015.11.003
- Chowdhury, N. R., Chakraborty, S., Eampokalap, B., Chaicumpa, W., Chongsanguan, M., Moolasart, P., et al. (2000a). Clonal dissemination of *Vibrio parahaemolyticus* displaying similar DNA fingerprint but belonging to two different serovars (O3:K6 and O4:K68) in Thailand and India. *Epidemiol. Infect.* 125, 17–25. doi: 10.1017/S0950268899004070
- Chowdhury, N. R., Chakraborty, S., Ramamurthy, T., Nishibuchi, M., Yamasaki, S., Takeda, Y., et al. (2000b). Molecular evidence of clonal *Vibrio parahaemolyticus* pandemic strains. *Emerg. Infect. Dis.* 6, 631–636. doi: 10.3201/eid0606.000612
- Deepanjali, A., Kumar, H. S., Karunasagar, I., and Karunasagar, I. (2005). Seasonal variation in abundance of total and pathogenic *Vibrio parahaemolyticus* bacteria in oysters along the southwest coast of India. *Appl Environ. Microbiol.* 71, 3575–3580. doi: 10.1128/AEM.71.7.3575-3580.2005
- de Jesús Hernández-Díaz, L., Leon-Sicairos, N., Velazquez-Roman, J., Flores-Villaseñor, H., Guadron-Llanos, A. M., Martinez-Garcia, J. J., et al. (2015). A pandemic *Vibrio parahaemolyticus* O3:K6 clone causing most associated diarrhea cases in the Pacific Northwest coast of Mexico. *Front. Microbiol.* 6:221. doi: 10.3389/fmicb.2015.00221
- Esteves, K., Mosser, T., Aujoulat, F., Hervio-Heath, D., Monfort, P., and Jumas-Bilak, E. (2015). Highly diverse recombining populations of *Vibrio cholerae* and *Vibrio parahaemolyticus* in French Mediterranean coastal lagoons. *Front. Microbiol.* 6:708. doi: 10.3389/fmicb.2015.00708
- Feil, E. J., Li, B. C., Aanensen, D. M., Hanage, W. P., and Spratt, B. G. (2004). eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *J. Bacteriol.* 186, 1518–1530. doi: 10.1128/JB.186.5.1518-1530.2004
- Ferreira, J. M., Oliveira, L. R. D., Wynen, L., Bester, M. N., Guinet, C., Moraes-Barros, N., et al. (2008). Multiple origins of vagrant Subantarctic fur seals: a long journey to the Brazilian coast detected by molecular markers. *Polar Biol.* 31, 303–308. doi: 10.1007/s00300-007-0358-z
- Garcia, K., Torres, R., Uribe, P., Hernandez, C., Rioseco, M. L., Romero, J., et al. (2009). Dynamics of clinical and environmental *Vibrio parahaemolyticus* strains during seafood-related summer diarrhea outbreaks in southern Chile. *Appl. Environ. Microbiol.* 75, 7482–7487. doi: 10.1128/AEM.01662-09
- Gavilan, R. G., Zamudio, M. L., and Martinez-Urtaza, J. (2013). Molecular epidemiology and genetic variation of pathogenic *Vibrio parahaemolyticus* in Peru. *PLoS Negl. Trop. Dis.* 7:e2210. doi: 10.1371/journal.pntd.0002210
- Gil, A. I., Miranda, H., Lanata, C. F., Prada, A., Hall, E. R., Barreno, C. M., et al. (2007). O3:K6 serotype of *Vibrio parahaemolyticus* identical to the global pandemic clone associated with diarrhea in Peru. *Int. J. Infect. Dis.* 11, 324–328. doi: 10.1016/j.ijid.2006.08.003
- Gonzalez-Escalona, N., Martinez-Urtaza, J., Romero, J., Espejo, R. T., Jaykus, L. A., and DePaola, A. (2008). Determination of molecular phylogenetics of *Vibrio parahaemolyticus* strains by multilocus sequence typing. *J. Bacteriol.* 190, 2831–2840. doi: 10.1128/JB.01808-07
- Haendiges, J., Rock, M., Myers, R. A., Brown, E. W., Evans, P., and Gonzalez-Escalona, N. (2014). Pandemic *Vibrio parahaemolyticus*, Maryland, USA, 2012. *Emerg. Infect. Dis.* 20, 718–720. doi: 10.3201/eid2004.130818
- Haendiges, J., Timme, R., Allard, M. W., Myers, R. A., Brown, E. W., and Gonzalez-Escalona, N. (2015). Characterization of *Vibrio parahaemolyticus* clinical strains from Maryland (2012–2013) and comparisons to a locally and globally diverse *V. parahaemolyticus* strains by whole-genome sequence analysis. *Front. Microbiol.* 6:125. doi: 10.3389/fmicb.2015.00125
- Han, D., Tang, H., Lu, J., Wang, G., Zhou, L., Min, L., et al. (2014). Population structure of clinical *Vibrio parahaemolyticus* from 17 coastal countries, determined through multilocus sequence analysis. *PLoS ONE* 9:e107371. doi: 10.1371/journal.pone.0107371
- Han, D., Tang, H., Ren, C., Wang, G., Zhou, L., and Han, C. (2015). Prevalence and genetic diversity of clinical *Vibrio parahaemolyticus* isolates from China, revealed by multilocus sequence typing scheme. *Front. Microbiol.* 6:291. doi: 10.3389/fmicb.2015.00291
- Hara-Kudo, Y., Sugiyama, K., Nishibuchi, M., Chowdhury, A., Yatsuyanagi, J., Ohtomo, Y., et al. (2003). Prevalence of pandemic thermostable direct hemolysin-producing *Vibrio parahaemolyticus* O3:K6 in seafood and the coastal environment in Japan. *Appl. Environ. Microbiol.* 69, 3883–3891. doi: 10.1128/AEM.69.7.3883-3891.2003
- Harth, E., Matsuda, L., Hernandez, C., Rioseco, M. L., Romero, J., Gonzalez-Escalona, N., et al. (2009). Epidemiology of *Vibrio parahaemolyticus* outbreaks, southern Chile. *Emerg. Infect. Dis.* 15, 163–168. doi: 10.3201/eid1502.071269
- Li, Y., Xie, X., Shi, X., Lin, Y., Qiu, Y., Mou, J., et al. (2014). *Vibrio parahaemolyticus*, Southern Coastal Region of China, 2007–2012. *Emerg. Infect. Dis.* 20, 685–688. doi: 10.3201/eid2004.130744
- Martinez-Urtaza, J., Bowers, J. C., Trinanes, J., and DePaola, A. (2010). Climate anomalies and the increasing risk of *Vibrio parahaemolyticus* and *Vibrio vulnificus* illnesses. *Food Res. Int.* 43, 1780–1790. doi: 10.1016/j.foodres.2010.04.001
- Martinez-Urtaza, J., Lozano-Leon, A., DePaola, A., Ishibashi, M., Shimada, K., Nishibuchi, M., et al. (2004). Characterization of pathogenic *Vibrio parahaemolyticus* isolates from clinical sources in Spain and comparison with Asian and North American pandemic isolates. *J. Clin. Microbiol.* 42, 4672–4678. doi: 10.1128/JCM.42.10.4672-4678.2004
- Martinez-Urtaza, J., Simental, L., Velasco, D., DePaola, A., Ishibashi, M., Nakaguchi, Y., et al. (2005). Pandemic *Vibrio parahaemolyticus* O3:K6, Europe. *Emerg. Infect. Dis.* 11, 1319–1320. doi: 10.3201/eid1108.050322
- Matsumoto, C., Okuda, J., Ishibashi, M., Iwanaga, M., Garg, P., Rammamurthy, T., et al. (2000). Pandemic spread of an O3:K6 clone of *Vibrio parahaemolyticus* and emergence of related strains evidenced by arbitrarily primed PCR and toxRS sequence analyses. *J. Clin. Microbiol.* 38, 578–585.

- Nair, G. B., Ramamurthy, T., Bhattacharya, S. K., Dutta, B., Takeda, Y., and Sack, D. A. (2007). Global dissemination of *Vibrio parahaemolyticus* serotype O3:K6 and its serovariants. *Clin. Microbiol. Rev.* 20, 39–48. doi: 10.1128/CMR.00025-06
- Okuda, J., Ishibashi, M., Hayakawa, E., Nishino, T., Takeda, Y., Mukhopadhyay, A. K., et al. (1997). Emergence of a unique O3:K6 clone of *Vibrio parahaemolyticus* in Calcutta, India, and isolation of strains from the same clonal group from Southeast Asian travelers arriving in Japan. *J. Clin. Microbiol.* 35, 3150–3155.
- Okura, M., Osawa, R., Tokunaga, A., Morita, M., Arakawa, E., and Watanabe, H. (2008). Genetic analyses of the putative O and K antigen gene clusters of pandemic *Vibrio parahaemolyticus*. *Microbiol. Immunol.* 52, 251–264. doi: 10.1111/j.1348-0421.2008.00027.x
- Pazhani, G. P., Bhowmik, S. K., Ghosh, S., Guin, S., Dutta, S., Rajendran, K., et al. (2014). Trends in the epidemiology of pandemic and non-pandemic strains of *Vibrio parahaemolyticus* isolated from diarrheal patients in Kolkata, India. *PLoS Neglect. Trop. D.* 8:e2815. doi: 10.1371/journal.pntd.0002815
- Powell, A., Baker-Austin, C., Wagley, S., Bayley, A., and Hartnell, R. (2013). Isolation of pandemic *Vibrio parahaemolyticus* from UK water and shellfish produce. *Microb. Ecol.* 65, 924–927. doi: 10.1007/s00248-013-0201-8
- Quilici, M. L., Robert-Pillot, A., Picart, J., and Fournier, J. M. (2005). Pandemic *Vibrio parahaemolyticus* O3:K6 spread, France. *Emerg. Infect. Dis.* 11, 1148–1149. doi: 10.3201/eid1107.041008
- Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123, 585–595.
- Theethakaew, C., Feil, E. J., Castillo-Ramirez, S., Aanensen, D. M., Suthienkul, O., Neil, D. M., et al. (2013). Genetic relationships of *Vibrio parahaemolyticus* isolates from clinical, human carrier, and environmental sources in Thailand, determined by multilocus sequence analysis. *Appl. Environ. Microb.* 79, 2358–2370. doi: 10.1128/AEM.03067-12
- Thongjum, J., Mittraparp-arthon, P., Yingkajorn, M., Kongreung, J., Nishibuchi, M., and Vuddhakul, V. (2013). The trend of *Vibrio parahaemolyticus* infections in Southern Thailand from 2006 to 2010. *Trop. Med. Health* 41, 151–156. doi: 10.2149/tmh.2013-06
- Ueno, H., Tomari, K., Kikuchi, K., Kobori, S., and Miyazaki, M. (2016). The first report of *Vibrio parahaemolyticus* O10:K60 in Japan, a new combination of O and K serotypes strain isolated from a gastroenteritis patient. *Jpn. J. Infect. Dis.* 69, 28–32. doi: 10.7883/yoken.JJID.2014.538
- Velazquez-Roman, J., León-Sicairos, N., de Jesus Hernández-Díaz, L., and Canizalez-Roman, A. (2014). Pandemic *Vibrio parahaemolyticus* O3:K6 on the American continent. *Front. Cell. Infect. Mi.* 3:110. doi: 10.3389/fcimb.2013.00110
- Velazquez-Roman, J., Leon-Sicairos, N., Flores-Villasenor, H., Villafana-Rauda, S., and Canizalez-Roman, A. (2012). Association of Pandemic *Vibrio parahaemolyticus* O3:K6 present in the coastal environment of Northwest Mexico with cases of recurrent Diarrhea between 2004 and 2010. *Appl. Environ. Microb.* 78, 1794–1803. doi: 10.1128/AEM.06953-11
- Vuddhakul, V., Chowdhury, A., Laohaprerthisan, V., Pungrasamee, P., Patararungroong, N., Thianmontri, P., et al. (2000). Isolation of a pandemic O3:K6 clone of a *Vibrio parahaemolyticus* strain from environmental and clinical sources in Thailand. *Appl. Environ. Microb.* 66, 2685–2689. doi: 10.1128/AEM.66.6.2685-2689.2000
- Wong, H. C., Liu, S. H., Wang, T. K., Lee, C. L., Chiou, C. S., Liu, D. P., et al. (2000). Characteristics of *Vibrio parahaemolyticus* O3:K6 from Asia. *Appl. Environ. Microb.* 66, 3981–3986. doi: 10.1128/AEM.66.9.3981-3986.2000
- Yeung, P. S., Hayes, M. C., DePaola, A., Kaysner, C. A., Kornstein, L., and Boor, K. J. (2002). Comparative phenotypic, molecular, and virulence characterization of *Vibrio parahaemolyticus* O3:K6 isolates. *Appl. Environ. Microb.* 68, 2901–2909. doi: 10.1128/AEM.68.6.2901-2909.2002
- Zhang, H., Sun, S., Shi, W., Cui, L., and Gu, Q. (2013). Serotype, virulence, and genetic traits of foodborne and clinical *Vibrio parahaemolyticus* isolates in Shanghai, China. *Foodborne Pathog. Dis.* 10, 796–804. doi: 10.1089/fpd.2012.1378

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Han, Tang, Ren, Zhu and Han. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Evaluation of the risk factors contributing to the African swine fever occurrence in Sardinia, Italy

**Beatriz Martínez-López<sup>1\*</sup>, Andres M. Perez<sup>2</sup>, Francesco Feliziani<sup>3</sup>, Sandro Rolesu<sup>4</sup>, Lina Mur<sup>5</sup> and José M. Sánchez-Vizcaíno<sup>5</sup>**

<sup>1</sup> Center for Animal Disease Modeling and Surveillance, Department of Medicine & Epidemiology, School of Veterinary Medicine, University of California, Davis, Davis, CA, USA, <sup>2</sup> Department of Veterinary Population Medicine, College of Veterinary Medicine, University of Minnesota, Saint Paul, MN, USA, <sup>3</sup> Istituto Zooprofilattico Sperimentale dell'Umbria e delle Marche, Perugia, Italy, <sup>4</sup> Osservatorio Epidemiologico Veterinario Regionale, Istituto Zooprofilattico Sperimentale della Sardegna, Cagliari, Italy, <sup>5</sup> Animal Health Department and Centro de Vigilancia Sanitaria Veterinaria, Veterinary School, Complutense University of Madrid, Madrid, Spain

## OPEN ACCESS

### Edited by:

Evangelos Giamarellos-Bourboulis,  
University of Athens Medical School,  
Greece

### Reviewed by:

Mohammad Mohseni Sajadi,  
University of Maryland School  
of Medicine, USA  
Efthymia Giannitsioti,  
ATTIKON University General Hospital,  
Greece

### \*Correspondence:

Beatriz Martínez-López,  
Center for Animal Disease Modeling  
and Surveillance, Department  
of Medicine & Epidemiology, School  
of Veterinary Medicine, University  
of California, Davis, One Shields  
Avenue, 2415A Tupper Hall, Davis,  
CA 95616, USA  
beamartinezlopez@ucdavis.edu

### Specialty section:

This article was submitted to  
Infectious Diseases, a section of the  
journal Frontiers in Microbiology

**Received:** 19 January 2015

**Accepted:** 29 March 2015

**Published:** 14 April 2015

### Citation:

Martínez-López B, Perez AM,  
Feliziani F, Rolesu S, Mur L and  
Sánchez-Vizcaíno JM (2015)  
Evaluation of the risk factors  
contributing to the African swine fever  
occurrence in Sardinia, Italy.  
*Front. Microbiol.* 6:314.  
doi: 10.3389/fmicb.2015.00314

This study assesses the relation between hypothesized risk factors and African swine fever virus (ASFV) distribution in Sardinia (Italy) after the beginning of the eradication program in 1993, using a Bayesian multivariable logistic regression mixed model. Results indicate that the probability of ASFV occurrence in Sardinia was associated to particular socio-cultural, productive and economical factors found in the region, particularly to large number of confined (i.e., closed) farms (most of them backyard), high road density, high mean altitude, large number of open fattening farms, and large number of pigs per commune. Conversely, large proportion of open farms with at least one census and large proportion of open farms per commune, were found to be protective factors for ASFV. Results suggest that basic preventive and control strategies, such as yearly census or registration of the pigs per farm and better control of the public lands where pigs are usually raised, together with enhanced efforts of outreach and communication with pig producers should help in the success of the eradication program for ASF in the Island. Methods and results presented here will inform decision making to better control and eradicate ASF in Sardinia and in all those areas with similar management and epidemiological conditions.

**Keywords:** Bayesian model, risk-based surveillance, eradication program, spatial epidemiology, backyard pigs

## Introduction

African swine fever (ASF) is caused by the infection with a complex and large DNA virus (ASFV) of the *Asfarviridae* family (Salas, 1999; Dixon et al., 2005). ASF is a haemorrhagic disease of pigs in which clinical signs depend on the virulence of the ASFV isolate, dose, and route of infection and host (domestic or wild pig). Clinical presentation may vary from a hyperacute form, with almost 100% mortality after 4–7 days post-infection, to a chronic or an asymptomatic form in which most of animals may survive and become carriers (Sánchez-Vizcaíno et al., 2012).

African swine fever is, arguably, one of the most difficult to control and economically devastating viral diseases of pigs. Difficulty to control ASF and the economic impact of the disease are consequence of a multiplicity of factors, including the role played by soft ticks (*Ornithodoros* genus) on disease transmission, lack of an effective vaccine to prevent ASF infection, long persistence of

the virus in the environment and in pig products, presence of asymptomatic and carrier animals, and severe restrictions to the international trade of pigs and their products imposed to regions in which the disease is known or suspected to be present (Sánchez-Vizcaíno et al., 2012).

After the first description of the disease by Montgomery (1921) in Africa, the disease rapidly spread during 1960s and 1970s, from Sub-Saharan African countries into Europe, and Central and South America. The ASFV also recently spread into Georgia, Armenia, Azerbaijan, and the Russian Federation, which are countries that were believed to be ASF-free prior to 2007 (Food and Agriculture Organization [FAO], 2009). ASF has been eradicated from the Americas and Western Europe, with the exception of the Mediterranean island of Sardinia, where the disease has been endemic since 1978. Although a rigorous European Union (EU)-supported ASFV eradication program has been in place in Sardinia since 1993 (EU Official Bulletin no. L 116, 1990), ASFV outbreaks are still reported on an annual basis in the island or even have increased in the last years in some areas. The ASF eradication program for Sardinia has been continuously evolving, but the basics are found on the Council Decision of April 25, 1990 (EU Official Bulletin no. L 116, 1990) which comprises measures for (i) a rapid elimination of ASF outbreaks (i.e., immediate killing, destruction, disinfection, and compensation of owners after an outbreak, etc.), (ii) surveillance and protection of pig farms (i.e., serological testing of pigs in high risk areas, epidemiological outbreaks investigations, pre-movement tests, sampling of wild pigs, etc.), (iii) identification of pigs and pig farms, and (iv) construction of facilities for sanitary control. Notice that the ASF eradication program in Sardinia is similar to other ASF eradication programs conducted, for example, in countries such as Spain or Portugal, although in those two countries ASF has been successfully eradicated.

Presence of ASF-infected territories in Sardinia seems not to pose a major threat to the pig industry of EU-free regions, given that only one outbreak was caused in northern Italy in 1983, by illegal introduction of pork from Sardinia (Mannelli et al., 1998). However, its presence inflicts severe economic losses, not only for EU and Italy, because of the funds invested in ASF control and eradication programs, but also for local producers due to the trade restrictions and the depreciation of their pig products. In studies conducted in the 1990s it has been hypothesized that the ASF eradication program in Sardinia has failed because of a variety of risk factors that characterize the traditional pig management practices of farmers in the island, including, for example, extensive premises with nil or insufficient biosecurity measures in place, illegal trade and production of pigs and pig, presence of wild boars, and use of waste to feed pigs (Wilkinson, 1984; Mannelli et al., 1997, 1998). Conversely, there is no evidence that soft ticks play any role in the maintenance or spread of ASFV-infection in Sardinia (Ruiu et al., 1989). Noteworthy, no peer-review recent studies have been published assessing association between ASFV infection and epidemiological factors that could influence disease occurrence in Sardinia after more than 15 years (1993–2011) of ASFV eradication program. Such knowledge will contribute to drive the allocation of human and financial resources to support the eradication of ASF in the only region

of the EU that remains endemic for the disease. Moreover, the pig production and epidemiological situation in Sardinia may be comparable to those observed in the regions that have been recently infected in the Caucasus region and Russian Federation, and for that reason, this study may contribute to better prevent and control the disease also in those areas with similar management and epidemiological conditions (Food and Agriculture Organization [FAO], 2009).

In the study here, Bayesian modeling was used to explore the nature and extent of the association between ASF occurrence in Sardinia from 1993 to 2009, and hypothesized risk factors for the disease. Results will provide quantitative knowledge on the spatial distribution of ASF in Sardinia and on factors that have influenced occurrence of ASF in Sardinia since 1993. Such knowledge will help to improve the effectiveness of ASF-eradication program in Sardinia and in other ASF-infected territories.

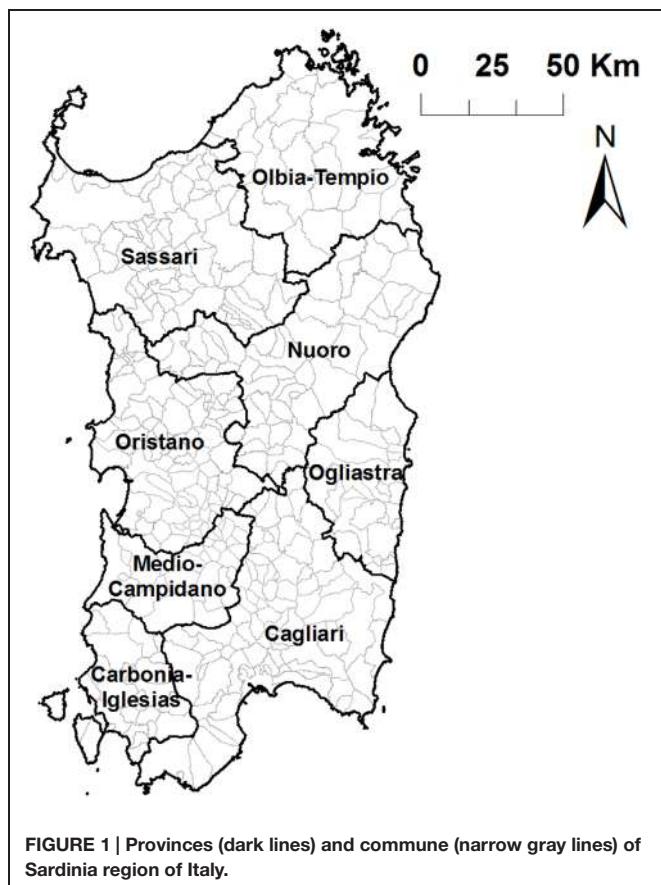
## Materials and Methods

### Unit of Analysis and Data Collection

The unit of analysis here was the smallest administrative division in Sardinia for which ASFV outbreaks were reported, which is referred to as commune. The commune is the basic administrative unit of Italy and it may be considered the third level of the country's administrative organization, where the first and second administrative aggregations are the region and the province, respectively. Sardinia, which is the second largest island in the Mediterranean Sea, is one of the 20 regions of Italy and it is divided in eight provinces and 376 communes. The mean area of a commune in Sardinia is 64 Km<sup>2</sup> (**Figure 1**).

An extensive review of the existing data on the distribution of potential risk factors for ASFV outbreak occurrence in Sardinia was conducted to ensure the completeness and reliability of the study. Firstly, potential risk factors for ASFV occurrence in Sardinia were identified searching in “MEDLINE” and “ISI-Web of Knowledge” databases by using a combination of English and Italian keywords such as of “African swine fever,” “peste suina africana,” “risk,” “Sardinia,” “Italy,” “epidemiology,” “cinghiale,” etc. As a result of this literature review, 153 references were selected; and the most updated and reliable information data on the identified risk factors were gathered per commune from national and international sources (**Table 1**). Moreover, selection of risk factors was validated during an expert opinion elicitation (including the administration of a paper questionnaire) conducted on May 20th, 2010 in Oristano, Sardinia, in which more than 25 local veterinary authorities involved in the current ASF control and eradication program were asked to identify the main risk factors for ASF occurrence in the island.

For each commune, information on the number of ASFV outbreaks, pig and human demographics, pig movements and environmental and geographic factors was collected. Specifically, the number of ASFV outbreaks reported per commune from January 1993 to March 2009 was provided by the Istituto Zooprofilattico Sperimentale della Sardegna and by the Istituto Zooprofilattico Sperimentale dell’Umbria e delle Marche. Each farm confirmed as ASFV infected was considered an ASFV outbreak. For each



ASF outbreak information regarding the unique identifier and type of farm, farm location, number of pigs present on farm, number of pigs with ASF compatible clinical symptoms, number of death pigs, the day of ASFV detection on farm, day of ASFV confirmation by laboratory, and day of farm depopulation were available.

Data regarding demographics, production type and pig movements per commune in 2009 were obtained from the Italian National Register (Anagrafe Nazionale Zootecnica – Statistiche, 2011). Briefly, just consider that pig production in Sardinia is characterized by traditional pig farming systems, with none or low biosecurity measures in place, either in closed/confined farms (65.4%) or in open/extensive areas (34.6%). Most (77%) of pig farms in Sardinia have less than 25 pigs on farm and more than 41% of the total number of farms are classified as farms for familiar consumption.

The Corine Land Use (2000), from the European Environmental Agency, was used to estimate the areas suitable for the presence/absence of wild boar populations. All agroforestry areas, bushes and cork oak plantations (CLC\_codes: 311–313 and 321–324) were selected as areas with potential presence of wild boar populations, following Rolesu et al. (2007). Other variables such as size of the human population, density of roads, or water areas (i.e., rivers, lakes, or inland waterways), density of grazing lands, and mean altitude were used as proxies for the estimation of the potential feeding of pigs with waste food,

legal, or illegal pig trade and the presence of illegal free range pigs, respectively. For example, mean altitude per commune was computed based on detailed ( $20 \text{ m}^2$ ) raster maps provided by the Istituto Zooprofilattico Sperimentale della Sardegna using the zonal statistic function on ArcMap 9.2 (ESRI®). Size of human population per commune in 2009 was obtained from the Istituto Nazionale di Statistica [ISTAT] (2011). Roads, water areas and mean altitude were obtained from the Regional Geographical Service (Servizio Informativo e Cartografico Regionale, Regione Sardegna, 2011). All available information was organized and stored in a Microsoft® Office Access 2007 database.

Although sources and data used for this study were, to the best of the authors' knowledge, the most complete and reliable information available, information regarding pig farms and pig population is imperfect, as it has been highlighted in the Italian National Register (Anagrafe Nazionale Zootecnica – Statistiche, 2011). Moreover, it should be noticed that although data regarding ASF outbreaks corresponded to 1993–2009, covariate information available to us was collected only in 2009. Because a wide year-to-year fluctuations in the traditional pig demographics, movements and husbandry system in Sardinia was considered unlikely as suggested by Mannelli et al. (1998), the assumption that information in 2009 was representative of the entire period was considered an acceptable simplification of our model.

## Model Approach

A Bayesian multivariable logistic regression mixed model was used to quantify the strength of the association between ASFV-infected communes and epidemiological factors hypothesized to influence the presence of ASF in Sardinia. The response variable was whether or not the commune reported ASF outbreaks from 1993 to 2009 (yes, no). Candidate variables to fit the prediction were each of the 29 epidemiological factors for which information was collected and their second-order interactions (Table 1). Candidate variables were alternative modeled as standardized continuous variables or categorized as binomial variables (above, below the median). After identifying a strong spatial structure in the data using a Bernoulli scan statistic model (SaTScan v9.1.1.), we decided to include in the model spatially structured ( $S_i$ ) and unstructured ( $U_i$ ) random effects to account for unmeasured factors that had some spatial structure and that were randomly distributed, respectively. Non-informative Normal priors of the form  $(0, 4)$  were used for the intercept and the regression coefficients. A non-informative normal prior with mean = 0 and precision =  $\text{dgamma}(0.5, 0.001)$  was assumed for  $U_i$ . An intrinsic Gaussian autoregressive (CAR) structure was used to model  $S_i$ , where the prior distribution of each  $S_i$  was conditional on the value of the response variable in an adjacent commune (i.e., those communes sharing common boundaries). The precision for  $S_i$  was assumed to be  $\text{dgamma}(0.5, 0.001)$ .

The model was fitted using WinBUGS v.1.4. (Spiegelhalter et al., 2002), with 20,000 iterations, after the first 1,000 samples were discarded as burn-in. Model was built by using a manual forward selection process (i.e., introducing one variable at a time). The model with the lowest deviance information criterion (DIC) value was considered the one that best fitted the data (Spiegelhalter et al., 2002). R2WinBUGS package (Sturtz

et al., 2005) in R-language software (R Development core team, 2011) was used to run the models in WinBUGS 1.4. The final model was assessed using autocorrelation plots to visually verify absence of autocorrelation in the predictions. Convergence of the model was also checked by using Gelman–Rubin Plots. The standard deviations (SD) of  $S_i$  and  $U_i$  were used to compare the variability of structured and unstructured random effects, as a proxy to estimate whether epidemiological factors not included in the model were spatially correlated or not. Maps were generated using ArcMap 9.2 (ESRI®).

## Results

African swine fever outbreaks in Sardinia were spatially clustered, as indicated by the two significant clusters found

with the Bernoulli scan statistic model (**Figure 2**) and by the much larger value of the DIC when the spatially structured random effect  $S_i$  was removed from the model (DIC = 408.96), compared with the model with  $S_i$  implemented on it (DIC = 267.00). The monthly number of ASF reported outbreaks from 1993 and 2009 is presented in **Figure 3**.

The model that best fitted the data (DIC = 220.88) included 14 variables; however, only four of them (namely, the proportion of open farms with at least one census, the number of confined/closed farms, the density of roads and, the mean altitude) were significantly ( $P < 0.05$ ) associated with ASF status (**Table 2**). The number of open fattening farms, the total number of pigs, and the proportion of open farms were marginal significant ( $P < 0.1$ ). The spatial distribution of significant variables is shown in **Figure 4**.

**TABLE 1 | List of variables used in the Bayesian model for African swine fever in Sardinia.**

Variable	Mean (min, max)*	Standardized	Binomial (Codification 0/1 using median)	Source
(1) The number of open fattening farms	4 (0, 61)	YES	na	A
(2) The number of open farms for self-consumption	17 (0, 146)	YES	YES	A
(3) The number of open farrowing farms	20 (0, 366)	YES	YES	A
(4) The number of open farms with no specification of the production type.	3 (0, 46)	YES	na	A
(5) The number of confined/closed farms	23 (0, 153)	YES	YES	A
(6) The total number of open farms	44 (0, 421)	YES	YES	A
(7) The number of sows	143 (0, 864)	na	YES	A
(8) The number of boars	37 (0, 511)	YES	YES	A
(9) The total number of pigs	226 (0, 912)	YES	YES	A
(10) The average number of farms with incoming or outgoing pig movements per month	1 (0, 28)	YES	YES	A
(11) The average number of pigs introduced into farms per month	10 (0, 433)	YES	YES	A
(12) The average number of pigs leaving farms per month	77 (0, 5886)	YES	YES	A
(13) The average number of pigs sent to the slaughterhouse per month	73 (0, 5884)	YES	YES	A
(14) The average number of new born pigs per month	5 (0, 273)	YES	na	A
(15) The area of the Commune (Km <sup>2</sup> )	64 (3, 545)	YES	na	B
(16) The mean altitude (meters)	663 (305, 1991)	YES	YES	B
(17) The density of water areas	0.32 (0, 6.09)	YES	YES	B
(18) The density of roads	0.27 (0, 0.83)	YES	YES	B
(19) The proportion of areas suitable for wild boar presence (%)	50 (0, 99)	YES	YES	C
(20) The proportion of open farms with at least one census (i.e., farms declaring the number of pigs on farm at least once per year; %)	80 (2, 100)	YES	YES	A
(21) The number of male people	2176 (50, 72684)	YES	YES	D
(22) The number of female people	2263 (44, 84267)	YES	YES	D
(23) The number of people (human population)	4439 (94, 156951)	YES	YES	D
(24) The mean farm size (number of pigs/number of farms)	4 (0, 29)	YES	YES	A
(25) Pig density (number of pigs/Km <sup>2</sup> )	5 (0, 39)	YES	YES	A
(26) Open farm density (number of farms/Km <sup>2</sup> )	1 (0, 5)	YES	YES	A
(27) Proportion of open farms (number of open farms/total farms) (%)	69 (0, 100)	YES	YES	A
(28) Density of grazing areas in the commune (Km <sup>2</sup> )	1432 (0, 27806)	YES	YES	B
(29) Proportion of area used as grazing areas in the commune (density of grazing areas in the commune Km <sup>2</sup> / total area of the commune in Km <sup>2</sup> ) (%)	20 (0, 75)	YES	YES	B

\*descriptive statistics of the original (i.e., not standardized) variable.

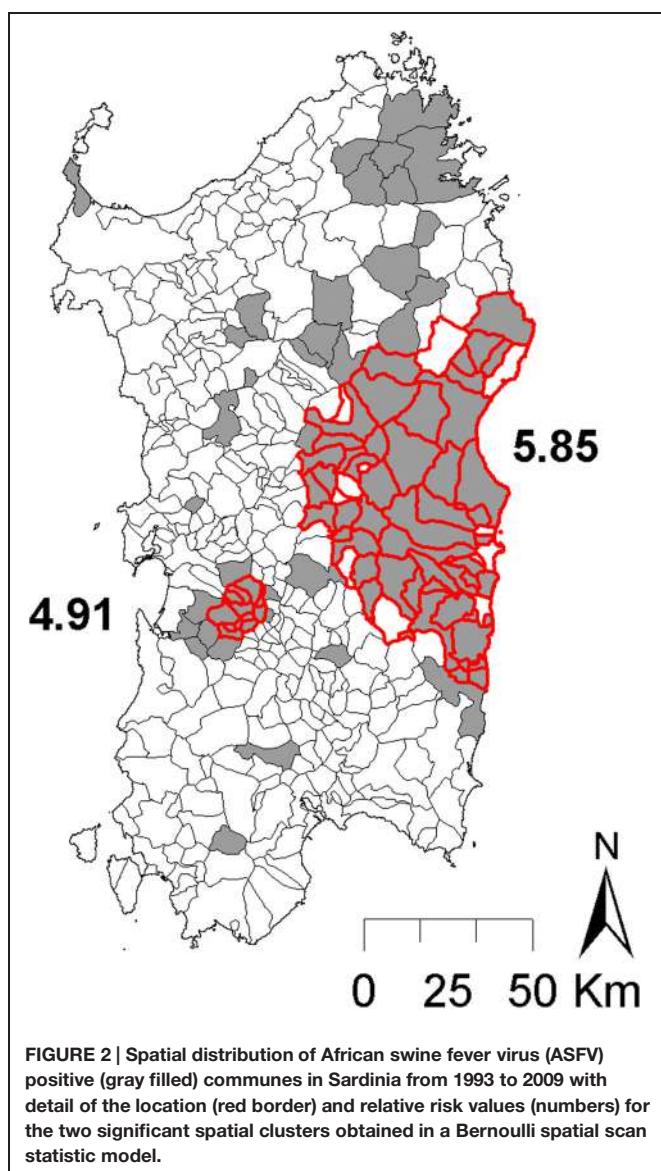
na = Not applicable

A = Anagrafe Nazionale Zootecnica – Statistiche (2011).

B = Servizio Informativo e Cartografico Regionale, Regione Sardegna (2011).

C = Corine Land Use (2000) and Rolesu et al. (2007).

D = Istituto Nazionale di Statistica [ISTAT] (2011).



The posterior probability of ASFV occurrence in Sardinia predicted by the fitted model resembled the spatial structure of the disease observed in the data, with areas at highest risk concentrated in Nuoro, Ogliastra, Olbia-Tempio, and Oristano provinces (**Figure 5**).

The SD of  $S_i$  (3.497) was substantially larger than SD of  $U_i$  ( $4.77 \cdot 10^{-6}$ ) in the final model (**Figure 6**), which indicates that effects not included into the model were spatially structured.

Two interactions were also found to contribute to the ASF prediction in the final model (**Table 2**). Specifically, in communes where the total number of pigs was greater than the median, an increase in the density of roads was associated with an increase in the predicted probability of ASFV occurrence in Sardinia. Similarly, communes where the total number of pigs was smaller than the median of all communes, an increase in the mean altitude was associated with an increase in the predicted probability of ASFV.

Convergence of the model was obtained after the first 200 iterations and no problems with autocorrelation were identified in the autocorrelation plots for the posterior inferences (data not shown).

## Discussion

Eradication of ASF in Sardinia seems to be far of being accomplished. From 1993 to 1998, the number of ASF outbreaks, which were mainly concentrated in the province of Nuoro, started to decrease, reaching minimum levels in 1999 but, in 2005 a large number of outbreaks were declared in Oristano province and in other areas outside the historical risk area of Nuoro, which emphasized the need to improve eradication efforts. The study presented here was aimed to better understand the nature and extent of epidemiological risk factors for ASFV occurrence in Sardinia in the last 15 years, after the beginning of the eradication plan in 1993.

An unexpected result of the model indicated that the number of confined (closed) farms was associated with high risk of occurrence of ASFV. This relation may be either the effect of a high detection and/or notification of outbreaks in those types of farms, or the consequence of risky management practices, particularly in small-size back yard closed farms, such as the use of waste containing pork products to feed pigs. These results may be in agreement with the hypothesis presented in previous studies by Mannelli et al. (1997), suggesting that the use of garbage to feed pigs was the main way of ASF infection in small farms in Sardinia. It should be noticed that most of the confined farms in Sardinia are small size with poor or no biosecurity measures that favors the access of pigs to contaminated waste food or the contact between infected and susceptible individuals. Only a limited number of those confined farms are large size and have an intensive type of production and high biosecurity level.

The number of open fattening farms was also considered to be a risk factor (OR = 1.49) for the occurrence of ASFV. This may be associated to a greater stocking density, higher/more frequent introduction of pigs on farm or other husbandry practices related to this production system. However, this result may also reflect one of the hypotheses suggested by regional authorities (Sardegna Salute, 2011) regarding the intentional introduction of the disease into some areas. The system of farmer's compensation after an ASF-outbreak in Italy, which is regulated by the Law n.218/88 (available for example at: [http://www.codima.info/trunk/nor\\_file\\_39\\_l218-88.pdf](http://www.codima.info/trunk/nor_file_39_l218-88.pdf)), essentially provides to the breeder the 100% of the value of slaughtered animals, calculated on the basis of the pig market of Modena (Borsa Merci di Modena, 2011). The value of the slaughtered animals, considered reasonable in relation to animals kept in industrial breeding farms, in some cases may be excessive for Sardinian pigs, bred using non-professional type of farming, and mainly when referring to fattening pigs. In such cases, the possibility to receive an extra-profit from the compensation has been suggested to be an incentive to intentionally introduce the disease in their herds for speculative purposes.

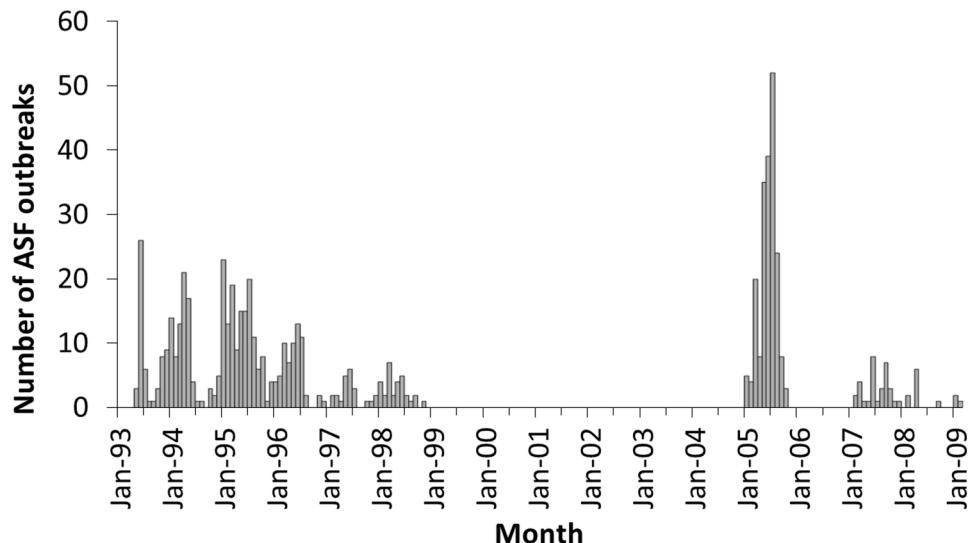


FIGURE 3 | Epidemic curve of the ASF notified outbreaks in Sardinia from 1993 to 2009.

It is also noteworthy that communes in which the proportion of areas suitable for wild boar presence, the number of open farms for self-consumption, the number of farms with incoming or outgoing pig movements, the herd size, or the pig density was above the median of all communes were at low risk (although not significant) for ASFV, as indicated by ORs values <1 (**Table 2**). These results suggest, that the role that wild boars play in causing ASF outbreaks is not crucial in Sardinia, which is in agreement with previous studies (Laddomada et al., 1994; Mannelli et al., 1997, 1998). Actually, it has been suggested that wild boars are often

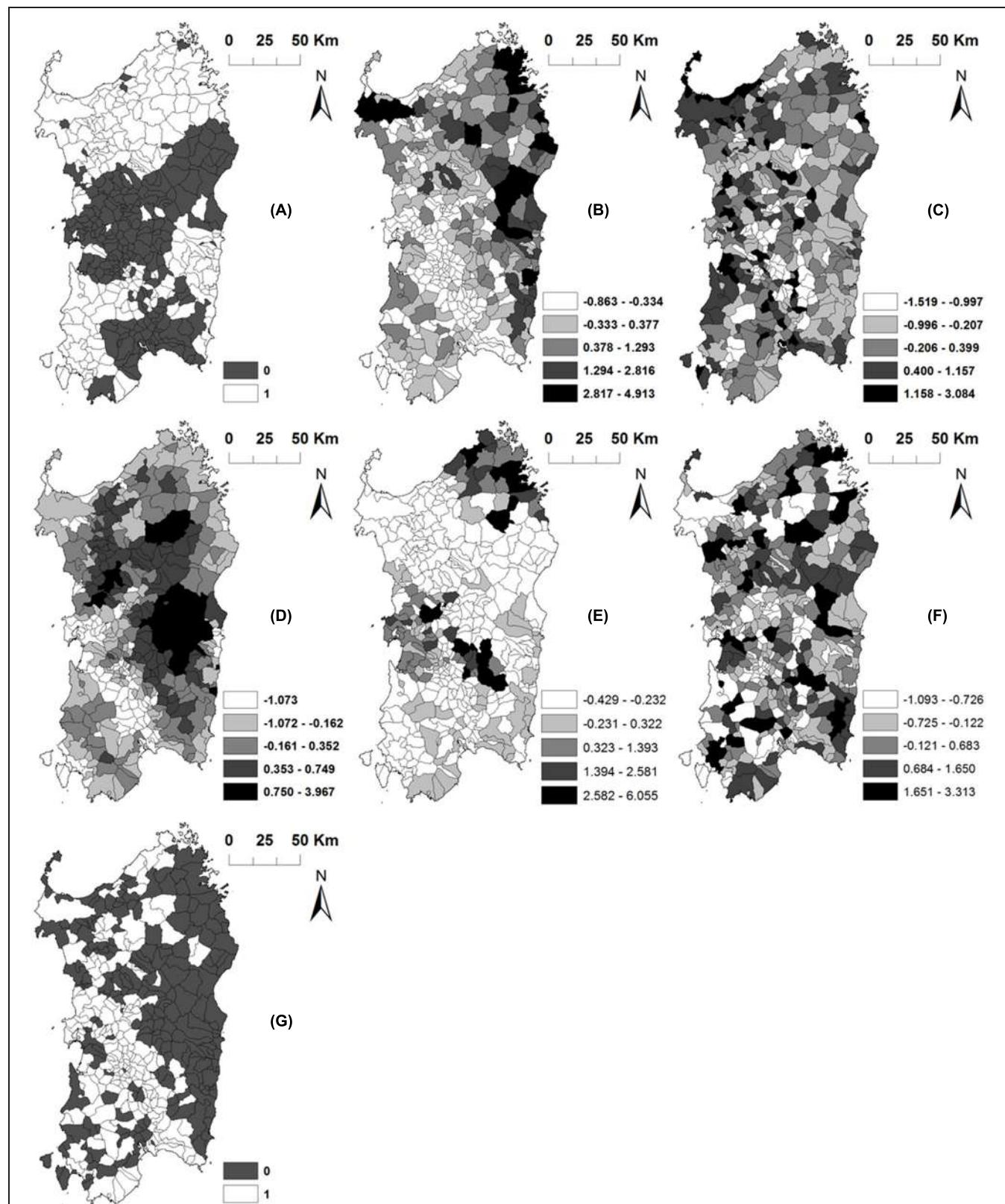
infected from domestic pigs in open grazing areas. This hypothesis is supported by the progressive disappearance of classical swine fever and ASF in wild boar only by controlling the grazing of domestic pigs, and by the time sequence of the reports of disease, which occur first in domestic pigs and after in wild boars. However, the potential role of wild boars as reservoir of the disease during short time periods is largely unknown and it has been suggested to be conditioned by several factors such as population dynamics, hunting pressure, and climatic factors such as fires and drought (Sardegna Salute, 2011). Therefore, wild

**TABLE 2 | Beta coefficients and Odds ratios for the best fitting model.**

Variable	Beta		OR	
	Median	95% PI	Median	95% PI
The proportion of open farms with at least one census (B)	-0.84 <sup>i,ii</sup>	[-1.63, -0.08]	0.43 <sup>i,ii</sup>	[0.20, 0.93]
The number of confined/closed farms (S)	0.79 <sup>i,ii</sup>	[0.21, 1.42]	2.21 <sup>i,ii</sup>	[1.23, 4.13]
Open farm density (number of farms/ Km <sup>2</sup> ) (B)	-0.61	[-1.41, 0.15]	0.55	[0.24, 1.16]
The density of roads (S)	0.52 <sup>i,ii</sup>	[0.03, 1.05]	1.69 <sup>i,ii</sup>	[1.03, 2.84]
The mean altitude (S)	0.70 <sup>i,ii</sup>	[0.15, 1.27]	2.01 <sup>i,ii</sup>	[1.16, 3.55]
The proportion of areas suitable for wild boar presence (B)	-0.25	[-1.09, 0.55]	0.78	[0.34, 1.73]
The number of open fattening farms (S)	0.40 <sup>ii</sup>	[-0.07, 0.90]	1.49 <sup>ii</sup>	[0.93, 2.47]
The total number of pigs (S)	0.45 <sup>ii</sup>	[-0.10, 1.03]	1.56 <sup>ii</sup>	[0.91, 2.80]
The number of open farms for self-consumption (S)	-0.09	[-0.60, 0.43]	0.91	[0.55, 1.54]
The number of farms with incoming or outgoing pig movements (B)	-0.29	[-1.05, 0.45]	0.75	[0.35, 1.58]
The total number of open farms (B)	-0.36	[-1.17, 0.41]	0.69	[0.31, 1.51]
Proportion of open farms (number of open farms/total farms; B)	-0.69 <sup>ii</sup>	[-1.50, 0.08]	0.50 <sup>ii</sup>	[0.22, 1.08]
The mean farm size (number of pigs/number of farms; B)	-0.38	[-1.18, 0.42]	0.69	[0.31, 1.52]
Pig density (number of pigs/Km <sup>2</sup> ; B)	-0.31	[-1.09, 0.47]	0.74	[0.34, 1.60]
The total number of pigs (S) × The density of roads (S)	0.29	[-0.18, 0.78]	1.33	[0.84, 2.19]
The total number of pigs (S) × The mean altitude (S)	0.06	[-0.37, 0.52]	1.06	[0.69, 1.68]

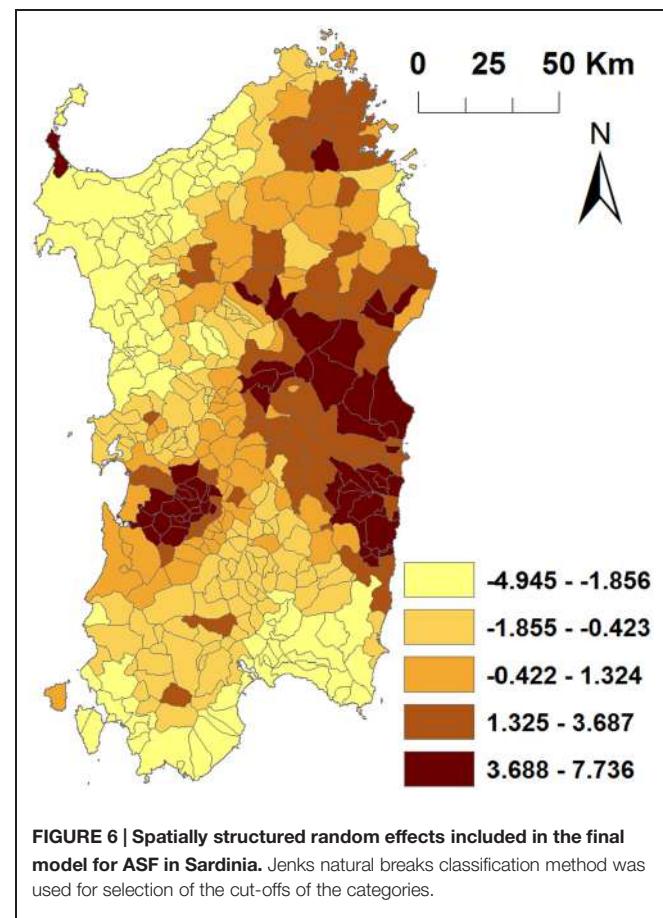
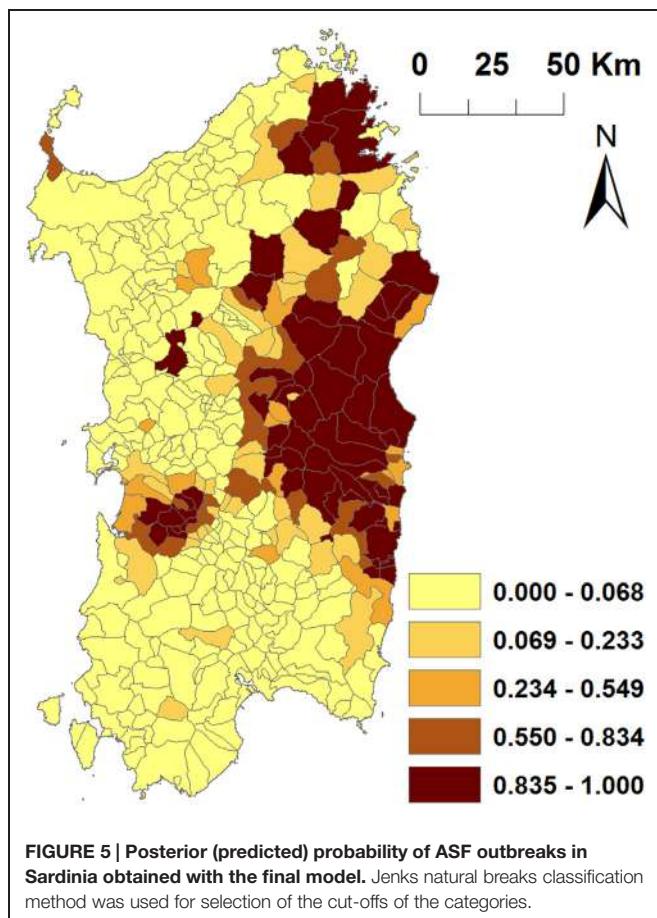
B, Binomial variable (codification 0/1 using median); S, Standardized variable.

<sup>i</sup>Significant coefficients of the best fitting model using the 95% prediction interval and <sup>ii</sup>the 90% prediction interval.



**FIGURE 4 | Spatial distribution of the significant variables included in the final model.** Maps represent: **(A)** the proportion of open farms with at least one census, **(B)** the number of confined (closed) farms, **(C)** the density of

roads, **(D)** the mean altitude, **(E)** the number of open fattening farms, **(F)** the total number of pigs and **(G)** the proportion of open farms. Jenks natural breaks classification method was used for categorization of quantitative variables **(B-F)**.



boards demographics should not be neglected in the design of an effective eradication program in areas where extensive farming is practiced.

The association between the total number of pigs, the density of roads and the ASFV occurrence per commune may be explained, at least in part, by a higher risk of infection thought vehicles or other transport-associated fomites in those areas. Other possible explanation may be the potential illegal trade of pigs, in agreement with some speculations (Sardegna Salute, 2011). Similarly, the association between ASFV occurrence, small number of pigs and high altitude per commune may have three alternative explanations. One potential explanation is that there is a longer persistence of the virus in the environment in those mountainous areas. A second one may be that not easily accessible areas located at high altitudes have low biosecurity measures and less control by veterinary authorities, which increases the risk for ASFV infection. Alternatively, another explanation may be that those remote areas are more likely to introduce illegal pigs (not declared/not censed). In any case, it is interesting to note that those remote areas with low number of pigs are at higher risk for ASF than other areas of the Island. This result highlights the importance of increasing efforts not only to control the population of pigs (reducing the illegal pigs) but also to regulate the contact among pig populations in those territories. In this regard, it could be very important to increase the

surveillance in the “*terre pubbliche*” or “*pascoli comunali*,” which are open areas that farmers can freely use to allocate and feed their animals (including pigs), and which may promote the ASFV transmission. Those open areas not only have low biosecurity and are difficult to disinfect and control, in case of ASF infection, but also are places where pigs may easily contact with swill feeding or waste coming from picnics and other celebrations there.

Conversely, the proportion of open farms with at least one census in the commune was the most significant protective factor ( $OR = 0.43$ ). Notice that censored farms are those in which the number of pigs on farm has been provided either by the farmer or by the veterinary authority. Consequently, communes with most of their farms censored (i.e., controlled by the veterinary authority) are less likely to have ASF outbreaks than not censored (i.e., unsupervised) farms. This relation may suggest that communes with a large proportion of not censored farms are more likely to have illegal pigs and/or risky management practices, and, in consequence, more ASF outbreaks than communes with a large proportion of censored farms. This result highlights the importance that relatively simple measures, such as a compulsory registration of the pig population, may have in the control ASF in Sardinia. Results suggest that such measure may, for example, help to reduce the number of farms holding illegal pigs, and consequently, risky farm practices that promote ASF spread.

Counter-intuitively, farms for self-consumption as well as areas with high pig density and large herd sizes and intense trade/movement of animals, seem to have less risk of ASFV occurrence than the background risk of the region. This finding may reflect that those types of farms have less contact with infected domestic pigs or pig products as a result of less trade, in the case of farms for self-consumption, or higher biosecurity measures, in the case of large/industrial pig farms.

In addition, unmeasured risk factors are likely associated with ASF in Sardinia as it has been suggested by the *S* and *U* components of the model. Spatial structure random effect (*S*) was considerably higher than unstructured effect (*U*). It means that these unmeasured factors are specifically spatial located, mainly in east part of the Island, and could be related, at least in part, with characteristic cultural and local pig husbandry management associated to specific areas. These factors could be associated with the illegal presence of pigs or pig trade, or the use of waste containing raw pork to feed of pig, which are factors difficult to estimate and for which not much information is available.

Unfortunately, information on the value of covariates prior to 2009 was not available to us and for that reason, an assumption of this study is that values recorded in 2009 are representative of the entire period. Also, certain data such as the number of pigs or number of pig farms were aggregated to the commune level and that some variables such as roads or water areas were polylines and had also to be transformed to the commune level. In general, aggregating data may lead to potential ecological bias. This study used the smallest unit of analysis to reduce this potential ecological bias and to facilitate the decision making process in the control and eradication of the disease. In this regard, commune was assumed to be the best level of aggregation as it is relatively small (median = 64 km<sup>2</sup>) and it is the unit of analysis and policy making used in the ASF eradication plan (Decreto assoreggiene sanità 6 luglio 2010, n. 33/1302, 2011).

## Conclusion

Future eradication programs should reinforce the compulsory registration of the pig population in Sardinia and the

surveillance and control measures in those “*terre pubbliche*” or “*pascoli comunali*” used by pigs. Those measures will help to reduce the number and trade of illegal pigs and to minimize the ASFV transmission in the Island. Results presented here suggest that it is fundamental to adapt the ASF preventive and control strategies considering all risk factors as well as the socio-cultural, productive and economical conditions of the region, in order to eradicate ASF from Sardinia and to achieve the ultimately goal of eradicating the disease from the EU and other regions of the world.

## Author Contributions

BL collected and processed the data from Anagrafe Nazionale Zootecnica – Statistiche, performed analyses and drafted the manuscript. AP provided assistance on the model development, supervised the analysis and gave major advices about the methodology and interpretation of results. FF and SR contributed with some of the data, provided expertise and technical considerations regarding the ASF eradication program in Sardinia and assisted on the interpretation and discussion of results. LM provided assistance on the data collection and processing and together with JS-V assisted on the study design and discussion of results. All authors read, reviewed, and approved the final manuscript.

## Acknowledgments

This work has been developed under the ASFRisk project (FP7 KBBE-2007-1-3-05, Grant Agreement n° 211691) and ASForce project (FP7 2007-2013, Grant Agreement n° 311931). Authors would like to acknowledge all personnel from the Istituto Zooprofilattico Sperimentale della Sardegna and Istituto Zooprofilattico Sperimentale dell’Umbria e delle Marche and Oristano colleagues. Authors wish to dedicate this manuscript to Dr. Cristiana Patta.

## References

- Anagrafe Nazionale Zootecnica – Statistiche. (2011). Available at: [http://statistiche.izs.it/portal/page?\\_pageid=73,12918&\\_dad=portal&\\_schema=PORTAL](http://statistiche.izs.it/portal/page?_pageid=73,12918&_dad=portal&_schema=PORTAL) [accessed March 21, 2011].
- Borsa Merci di Modena. (2011). Available at: <http://www.borsamercimodena.it/> [accessed September 11, 2011].
- Corine Land Use. (2000). Available at: <http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2000-clc2000-seamless-vector-database> [accessed March 14, 2011].
- Decreto assoreggiene sanità 6 luglio 2010, n. 33/1302. (2011). Decreto attuativo del Piano di eradicazione della Peste Suina Africana e della Peste Suina Classica. Anno 2010. Available at: [http://www.anmvioggi.it/files/DECRETO%20ASSESSORE%20IGIENE%20SANITA%27%206%20luglio%202010,%20n.%2033\\_1302.pdf](http://www.anmvioggi.it/files/DECRETO%20ASSESSORE%20IGIENE%20SANITA%27%206%20luglio%202010,%20n.%2033_1302.pdf) [accessed September 2, 2011].
- Dixon, L. K., Escribano, J. M., Martins, C., Rock, D. L., Salas, M. L., and Wilkinson, P. J. (2005). “Asfarviridae,” in *Virus Taxonomy: Eighth Report of the ICTV*, eds
- C. M. Fauquet, M. A. Mayo, J. Maniloff, U. Desselberger, and L. A. Ball . (London: Elsevier Academic Press), 135–143.
- Food and Agriculture Organization [FAO]. (2009). *African Swine Fever Spread in the Russian Federation and the Risk for the Region*. In *Empres Watch*, 2009. Available at: <ftp://ftp.fao.org/docrep/fao/012/ak718e/ak718e00.pdf> [accessed September 5, 2011].
- Istituto Nazionale di Statistica [ISTAT]. (2011). Available at: [www.istat.it/en/](http://www.istat.it/en/) [accessed March 9, 2011].
- Laddomada, A., Patta, C., Oggiano, A., Caccia, A., Ruiu, A., Cossu, P., et al. (1994). Epidemiology of classical swine fever in Sardinia: a serological survey of wild boar and comparison with African swine fever. *Vet. Rec.* 134, 183–187. doi: 10.1136/vr.134.8.183
- Mannelli, A., Sotgia, S., Patta, C., Oggiano, A., Carboni, A., Cossu, P., et al. (1998). Temporal and spatial patterns of African swine fever in Sardinia. *Prev. Vet. Med.* 35, 297–306. doi: 10.1016/S0167-5877(98)00063-4
- Mannelli, A., Sotgia, S., Patta, C., Sarria, A., Madrau, P., Sanna, L., et al. (1997). Effect of husbandry methods on seropositivity to African swine fever virus

- in Sardinian swine herds. *Prev. Vet. Med.* 32, 235–241. doi: 10.1016/S0167-5877(97)00026-3
- Montgomery, R. E. (1921). On a form of swine fever occurring in British East Africa (Kenya Colony). *J. Comp. Pathol.* 34, 159–191. doi: 10.1016/S0368-1742(21)80031-4
- R Development core team. (2011). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Rolesu, S., Alois, D., Ghironi, A., Oggiano, N., Oggiano, A., Puggioni, G., et al. (2007). Geographic information systems: a useful tool to approach African swine fever surveillance management of wild pig populations. *Vet. Ital.* 43, 463–467.
- Ruiu, A., Cossu, P., and Patta, C. (1989). Ricerca di zecche del genere Ornithodoros e di altri artropodi in allevamenti suini ed in cinghiali della Provincia di Nuoro. [Search for ticks of the genus Ornithodoros and other arthropods in pig farms and in wild boars in the Province of Nuoro]. *Atti della Soc. Ital. Sci. Vet. Proc. Soc. Ital. Vet. Sci.* 43, 1378–1391.
- Salas, M. (1999). “African swine fever virus (Asfarviridae),” in *Encyclopedia of Virology*, 2nd Edn, eds A. Granoff, R. G. Webster (London: Academic Press), 30–38.
- Sánchez-Vizcaíno, J. M., Mur, L., and Martínez-López, B. (2012). African swine fever: an epidemiological update. *Transbound. Emerg. Dis.* 59, 27–35. doi: 10.1111/j.1865-1682.2011.01293.x
- Sardegna Salute. (2011). Available at: <http://www.sardegnasalute.it/index.php?xsl=313&s=36929&v=2&c=3268> [accessed September 1, 2011].
- Servizio Informativo e Cartografico Regionale, Regione Sardegna. (2011). Available at: <http://www.sardegnaterritorio.it/> [accessed July 6, 2011].
- Spiegelhalter, D. J., Thomas, A., Best, N., and Lunn, D. (2002). *WinBUGS User Manual. Version 1.4*. Cambridge: MRC Biostatistics Unit.
- Sturtz, S., Ligges, U., and Gelman, A. (2005). R2WinBUGS: a package for running WinBUGS. *J. Stat. Softw.* 12, 1–16.
- Wilkinson, P. J. (1984). The persistence of African swine fever in Africa and the Mediterranean. *Prev. Vet. Med.* 2, 71–82. doi: 10.1016/0167-5877(84)90050-3

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Martínez-López, Perez, Feliziani, Rolesu, Mur and Sánchez-Vizcaíno. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Applications of Bayesian Phylodynamic Methods in a Recent U.S. Porcine Reproductive and Respiratory Syndrome Virus Outbreak

**Mohammad A. Alkhamis**<sup>1,2\*</sup>, **Andres M. Perez**<sup>1</sup>, **Michael P. Murtaugh**<sup>3</sup>, **Xiong Wang**<sup>1,3</sup> and **Robert B. Morrison**<sup>1</sup>

<sup>1</sup> Department of Veterinary Population Medicine, College of Veterinary Medicine, University of Minnesota, St. Paul, MN, USA

<sup>2</sup> Environmental and Life Sciences Research Center, Kuwait Institute for Scientific Research, Kuwait City, Kuwait,

<sup>3</sup> Department of Veterinary and Biomedical Sciences, University of Minnesota, St. Paul, MN, USA

## OPEN ACCESS

### Edited by:

Jörg Linde,

Leibniz-Institute for Natural Product Research and Infection Biology - Hans-Knoell-Institute, Germany

### Reviewed by:

Hein Min Tun,

University of Manitoba, Canada

Sebastian Mueller,

University of Cambridge, UK

### \*Correspondence:

Mohammad A. Alkhamis

malkahmi@umn.edu

### Specialty section:

This article was submitted to

Infectious Diseases,

a section of the journal

Frontiers in Microbiology

**Received:** 25 August 2015

**Accepted:** 14 January 2016

**Published:** 02 February 2016

### Citation:

Alkhamis MA, Perez AM,

Murtaugh MP, Wang X and

Morrison RB (2016) Applications of

Bayesian Phylodynamic Methods in a

Recent U.S. Porcine Reproductive

and Respiratory Syndrome Virus

Outbreak. *Front. Microbiol.* 7:67.

doi: 10.3389/fmicb.2016.00067

Classical phylogenetic methods such as neighbor-joining or maximum likelihood trees, provide limited inferences about the evolution of important pathogens and ignore important evolutionary parameters and uncertainties, which in turn limits decision making related to surveillance, control, and prevention resources. Bayesian phylodynamic models have recently been used to test research hypotheses related to evolution of infectious agents. However, few studies have attempted to model the evolutionary dynamics of porcine reproductive and respiratory syndrome virus (PRRSV) and, to the authors' knowledge, no attempt has been made to use large volumes of routinely collected data, sometimes referred to as big data, in the context of animal disease surveillance. The objective of this study was to explore and discuss the applications of Bayesian phylodynamic methods for modeling the evolution and spread of a notable 1-7-4 RFLP-type PRRSV between 2014 and 2015. A convenience sample of 288 ORF5 sequences was collected from 5 swine production systems in the United States between September 2003 and March 2015. Using coalescence and discrete trait phylodynamic models, we were able to infer population growth and demographic history of the virus, identified the most likely ancestral system (root state posterior probability = 0.95) and revealed significant dispersal routes (Bayes factor > 6) of viral exchange among systems. Results indicate that currently circulating viruses are evolving rapidly, and show a higher level of relative genetic diversity over time, when compared to earlier relatives. Biological soundness of model results is supported by the finding that sow farms were responsible for PRRSV spread within the systems. Such results cannot be obtained by traditional phylogenetic methods, and therefore, our results provide a methodological framework for molecular epidemiological modeling of new PRRSV outbreaks and demonstrate the prospects of phylodynamic models to inform decision-making processes for routine surveillance and, ultimately, to support prevention and control of food animal disease at local and regional scales.

**Keywords:** Bayesian phylodynamics, PRRSV, RFLP type 1-7-4, ORF5 gene, molecular surveillance

## INTRODUCTION

Porcine Reproductive and Respiratory Syndrome (PRRS) is, arguably, the most important swine disease in the United States due to the continuous emergence of new outbreaks that cause severe economic losses (Neumann et al., 2005; Holtkamp et al., 2013). Type 2 PRRSV, which is endemic in North America, was discovered in 1989 in the U.S., although the earliest serological evidence was found in eastern Canada (Benfield et al., 1992; Zimmerman, 2003; Murtaugh et al., 2010). PRRSV is a single-stranded, enveloped RNA virus that belongs to the *Arteriviridae* family (Benfield et al., 1992). Its genome consists of nine open reading frames (ORF) that code seven structural proteins and 14 non-structural proteins (Dokland, 2010). ORF5 encodes a major envelope surface glycoprotein (GP5) with high genetic diversity, thus has been widely used in molecular epidemiology studies of PRRSV (Kapur et al., 1996; Shi et al., 2010; Brar et al., 2015).

PRRSV transmission is rapid and can occur through direct and indirect contact (Dea et al., 2000; Cho et al., 2007). Emerging PRRSV strains are capable of spreading over long distances, referred to as distance-independent dispersal, as a result of aerosol transmission, animal movements, and use or movement of contaminated semen, equipment, or trucks (Shi et al., 2010, 2013). The combination of varied transmission routes and absence of regulated control and prevention activities makes virus control or elimination, at both local and regional levels, extremely challenging (Corzo et al., 2010; Rowland and Morrison, 2012). Hence, intensifying efforts toward designing effective and efficient surveillance programs, with the long-term goal of eliminating the disease, must be prioritized to minimize the current impact of the PRRSV on the US swine industry (Perez et al., 2015).

Since the 1980's, the U.S. Department of Agriculture has conducted extensive surveillance activities for swine diseases using classical statistical sampling methods that can account for imperfect diagnostic testing (Cameron and Baldock, 1998). However, current disease surveillance activities do not fully account for modern swine production systems in which pigs are spatially separated by age or production stage, or for pathogens that evolve rapidly (Rowland and Morrison, 2012; Perez et al., 2015).

In the past few decades, many studies investigated the molecular epidemiology of PRRSV, due to its high potential for mutation and recombination (Martín-Valls et al., 2014). Some studies focused on establishing associations between the evolutionary features of PRRSV and epidemiological characteristics of outbreaks in different geographical levels (Goldberg et al., 2000; Shi et al., 2010, 2013; Yoon et al., 2013; Nguyen et al., 2014; Rosendal et al., 2014). Others discriminated between novel and preexisting strains to model viral spread and maintenance within affected populations (Larochelle et al., 2003; Tun et al., 2011; Alonso et al., 2013; Brito et al., 2014; Chen et al., 2015). Whether the studies used classical phylogenetic methods to either genotype newly emerging PRRSV strains on the basis of restriction fragment length polymorphism (RFLP) patterns, or assessed correlations between the similarities of nucleotide sequences and other epidemiologic features,

they typically ignored uncertainties associated with estimates of phylogenetic relationships, temporal factors, and spatial factors (Suchard et al., 2001). Furthermore, they examined the temporal and spatial dynamics of the virus isolates in separate methodological settings, and attempted to draw conclusions from the outputs of both epidemiological and evolutionary analytical methods (Suchard et al., 2001). Therefore, many methodological approaches previously used to study PRRSV have ignored that evolutionary and epidemiological dynamics of rapidly evolving pathogens like PRRSV occur on approximately the same time-scale, and thus, they must be studied in a unified methodological setting in order to be properly understood and to prevent biased conclusions, subsequently improving the related decision making processes (Pybus et al., 2013). The field of phylodynamics aims to model, in a Bayesian statistical framework, the joint evolutionary, and epidemiological characteristics of rapidly evolving pathogens using analytical methods from the well-established field of phylogenetics (Grenfell et al., 2004). This approach uses important evolutionary parameters of rapidly evolving pathogens as random variables, and assigns a specified prior probability distribution for each parameter to infer their corresponding posterior probability distribution (Lemey et al., 2009). Thus, such Bayesian framework provides powerful analytical tools capable of accounting for uncertainties in the evolutionary parameters, including the pathogen phylogeny, population demographics, size, and history of dispersal between geographical regions and hosts (Lemey et al., 2009).

Bayesian phylodynamic models have recently become well-established tools for studying the evolution of many infectious viral diseases. However, only a few studies have modeled the evolutionary dynamics of PRRSV (Tun et al., 2011; Shi et al., 2013; Brito et al., 2014; Nguyen et al., 2014; Chaikhumwang et al., 2015). Such studies revealed the potential of phylodynamic methods in answering many long-standing questions on the molecular epidemiology and evolution of PRRSV. Furthermore, the method has previously been applied in a research context, rather than for routine surveillance of field data intended to support disease prevention and control. Such implementation is challenging because of the complexity and size of the data being analyzed. Data with these features, sometimes referred to as big data, requires special procedures for preparation and analysis.

The objective of this study was to demonstrate the application of Bayesian phylodynamic models to data routinely collected by swine production systems to support a near real-time early warning surveillance system for PRRSV and, potentially, other food animal viruses. The method was applied to the spread of a virulent RFLP 1-7-4 type PRRSV between 2014 and 2015 in the U.S. A discrete-trait phylodynamic model was adopted to estimate both the geographical history of viral migration and the movement of the virus among age groups of pigs. Our study provides quantitative estimates of mechanisms that lead to the emergence, spread and maintenance of the RFLP 1-7-4 PRRSV family throughout the U.S. It further illustrates the prospects of the Bayesian approach in improving the decision making process related to reducing the impact of PRRS on the national swine industry with the long-term goal of successful control and prevention.

## MATERIALS AND METHODS

### Sequence Data

Complete PRRS ORF5 nucleotide sequences ( $n = 6774$ ) from field isolates obtained between January 1998 and April 2015 were provided by five independent swine production systems in the U.S. with metadata on the date of isolation, system code (A, B, C, D, and E) and type of farm (farrow to wean or farrow to feeder sow farms and growing pig farms) from which the sequences were isolated (Table S1). Sequences were deposited in Genbank with accession numbers KT902023-KT905410 and KU501407-KU504248. The data were shared under agreement that identity and location of participants and their respective farms was confidential. Sequencing was performed according to the procedures in use at the time in various veterinary diagnostic laboratories or in private laboratories on a fee-for-service basis.

### Preliminary Phylogenetic Analysis

The complete sequence database was manually validated for presence of a complete ORF5 and absence of ambiguous nucleotides then aligned using MUSCLE version 3.8 (Edgar, 2004). A maximum likelihood (ML) phylogenetic analysis was performed in MEGA6, resulting in identification of a cluster of 288 sequences that were further studied. The sequence file was re-aligned using MUSCLE, and adjusted manually using amino-acid translation method implemented in Mesquite version 3.01 (Maddison and Maddison, 2011), to ensure that the protein-coding region of ORF5 remained in frame. Sequences with 100% nucleotide identity were removed (34%) from the subsequent analyses. While using Recombination Detection Program version 3 (RDP3), no homologous recombination was detected in the remaining sequences (Martin et al., 2010). For this analytical approach, it is important to select the substitution model that best describes the specific virus. For example, it was found that for some Dengue viruses, the mixed substitution model best fit the data (Drummond and Rambaut, 2007). That may, however, not be true for PRRSV. Thus, the best fitting partitioning scheme and nucleotide substitution model were selected using the Bayesian Information Criterion (BIC) implemented in PartitionFinder V 1.1 (Lanfear et al., 2012). Finally, maximum-likelihood estimates of the phylogeny under the selected mixed-substitution model were used to assess the degree of topological (in)congruence, in which 100 non-parametric bootstrap replicate searches were performed using RAxML version 8 (Stamatakis, 2014).

### Divergence-Time, Growth Rate, and Population Size Estimation

Divergence time was estimated using the relaxed molecular-clock model with GTR+ $\Gamma_4$  mixed-substitution, which was selected based on the results of PartitionFinder analysis mentioned above, implemented in BEAST v 1.8 (Drummond and Rambaut, 2007). To estimate divergence time and viral growth rate within each production system, we assessed the fit of the sequence data to five node-age coalescent priors, namely, (1) constant population size assuming that the population growth rate is zero (Griffiths and Tavaré, 1994); (2) exponential growth assuming that the population growth rate is fixed over time (Griffiths and Tavaré,

1994); (3) expansion growth assuming that the population growth rate increases over time (Griffiths and Tavaré, 1994); (4) logistic growth assuming that the population growth rate decreases over time (Griffiths and Tavaré, 1994); and (5) piecewise-constant Bayesian skyline coalescent model (BS) assuming effective population size is experiencing episodic stepwise change over time (Drummond et al., 2005). For each node-age model, we compared the uncorrelated exponential (UCED) and the log-normal (UCLN) relaxed clock branch-rate prior models, to assess whether our sequence data had a substitution rate on adjacent branches that sampled from either shared exponential or log-normal distributions, respectively. Isolation dates of the sequences were used to calibrate divergence-time estimates. We first estimated the marginal likelihood for each of the 10 candidate phylodynamic models from the resulting posterior samples using the posterior simulation-based analog of Akaike's information criterion (AICM; Raftery et al., 2007), which were estimated using Tracer version 1.6 (Suchard et al., 2001; Rambaut et al., 2014). The AICMs and their Monte Carlo standard errors (SE) were calculated using 1000 replicates. Bayes factor (BF) comparisons indicated that the sequence data followed a population expansion growth with a UCED branch-rate model, which provided the best fit for ORF5 (BF > 25 for the log marginal likelihood) among parametric models (Table S2). However, the BF comparison was not significant when the expansion model was compared against the BS coalescent tree prior (Table S2). Hence, the BS coalescent tree prior model with a UCED branch-rate was used to estimate changes in the effective population size through time (File S2; Minin et al., 2008).

We used the Markov Chain Monte Carlo (MCMC) algorithms implemented in BEAST to estimate the joint posterior probability distributions of the model parameters. For each MCMC simulation, we run  $3 \times 10^8$  cycles, which was thinned by sampling every 10,000 cycles. Two replicate MCMC simulations were carried out to aid in assessing simulation performance. We used Tracer to evaluate convergence of each candidate model by estimating effective sample sizes (ESS) for each posterior parameter. Hence, our ESS evaluations suggested that the MCMC algorithms requires the removal of the first 10% of the samples (the "burn-in") to provide reliable approximations of the posterior probability densities for each estimated parameter. We used Tree Annotator to summarize the posterior results in form of maximum clade credibility (MCC) trees. A BS plot was generated to infer effective population size (EPS) of the virus between 2001 and 2015, in which the EPS is defined as the relative genetic diversity (NeT), where Ne is the effective population size and T is the generation time (Minin et al., 2008).

### Estimation of Viral Dispersal History between Regional Systems

Geographical location was incorporated as described elsewhere (Lemey et al., 2009). Briefly, We reconstructed the phylogeny of the virus by incorporating discrete traits (i.e., systems), to describe the dispersal evolution of PRRSV epidemic among those selected systems. We used the continuous-time Markov model implemented in BEAST to model the dispersal history among systems as discrete states, which comprised a number

of non-zero transition rates identified by a Bayesian stochastic search variable selection (BSSVS) approach (Lemey et al., 2009). Furthermore, we investigated directionality of the geographical dispersal of the virus among systems by assessing the fit of the data to two candidate discrete trait models (Table S2), including both symmetric and asymmetric models with irreversible and reversible transitions, respectively. Here, the symmetric model with irreversible transitions indicate that the directional spread of the virus between two systems ( $A \rightarrow B$  or/and  $B \rightarrow A$ ) is insignificant, while the asymmetric model with reversible transition indicate that the directional spread between two systems ( $A \rightarrow B$  or/and  $B \rightarrow A$ ) is significant. To reconstruct the history of viral migration between discrete system areas, we used the coalescent Gaussian Markov Random field (GMRF) Bayesian Skyride model as a prior on the node times in the tree and a mean-one exponential prior for the rate parameters of the candidate models, while we used the same remaining parameters described in the above analyses (e.g., substitution model, UCLN, and UCED branch-rate models). Similarly, we estimated the marginal-likelihoods in order to compute the BFs to select among the candidate models (e.g., UCLN symmetric vs. UCED Asymmetric; Table S2; File S3). We used FigTree version 1.4 (Rambaut, 2012) to plot the summarized MCC consensus tree with the root state posterior probabilities (RSPP) of systems areas. Here, the RSPP is defined as the posterior probability of transition from one discrete trait to another mapped onto the interior nodes of the phylogeny of the virus, in which a discrete trait with a high RSPP indicate that trait as the likely ancestral trait of the given phylogeny. Finally, we used SPREAD version 1.0.6 (Bielejec et al., 2011) to identify non-zero transition rates between discrete traits (i.e., significant dispersal routes among systems). We used a BF cutoff = 6 to assess the strength and significance of transition rates between discrete geographic system areas. Because actual centroids of the site locations were confidential, relationally correct, anonymous latitude and longitude locations were placed in Alaska and a keyhole markup language (KML) file was generated to visualize regional migration of the virus.

## Modeling Viral Transmission in a System

Evolutionary movement between farm types (a proxy for production type), in which farm type were classified as sow herd (e.g., farrow to wean and farrow to feeder sow), and all other farms (e.g., finisher and nursery). A discrete-trait model was used for farm type (sow herd, other farms) to infer the history of PRRSV migration between farm types through time. The number of non-zero transition rates in the model was estimated using BSSVS. The relative strength of transition rates (e.g., sow farms → all other farms) was estimated using Bayes factors (BFs). We estimated the ancestral states (farm type) at internal nodes of the tree under a composite phylogenetic model that included the above detailed analyses. We used FigTree to plot the MCC consensus tree with the RSPP of the discrete trait and we assessed the strength of transition rates between states (farm types) using the BF comparisons implemented in SPREAD similar to the above analyses. Similarly, the use of the asymmetric or symmetric discrete trait models allowed us to assess the

strength and significance of directionality between farm types (e.g., sow farms → other farms, or/and farms → sow farms; File S4).

## Uncertainty and Statistical Analysis of Discrete-Trait Mappings

We used the Kullback–Leibler divergence (KL) statistic to quantify the magnitude of phylogenetic uncertainty in the discrete-trait estimates of the RSPP (for regional systems and farm type; Kullback and Leibler, 1951). KL statistics were calculated for each selected tree using the Razavi function (Razavi, 2008) implemented in Matlab v 2013a (MathWorks, 2012) to measure the departure between prior and the corresponding posterior probability distributions for a given phylodynamic parameter (i.e., in this case the RSPPs). A large KL-value ( $KL > 1$ ) indicates that the prior provided sufficient information for estimating the posterior parameters. Finally, we calculated the parsimony score (PS) and the association index (AI) statistics to assess the hypothesis that a taxon with a given trait (farm type or regional system) is more likely to share that trait with adjoining taxa in the MCC tree than would be expected by chance. The AI and PS statistics were calculated using Bayesian Tip-Significance Testing (BaTS) software version 1.0 (Parker, 2008). Significant AI and PS statistics indicate that our selected trait did have a significant role in shaping the posterior phylogeny of the sequence data.

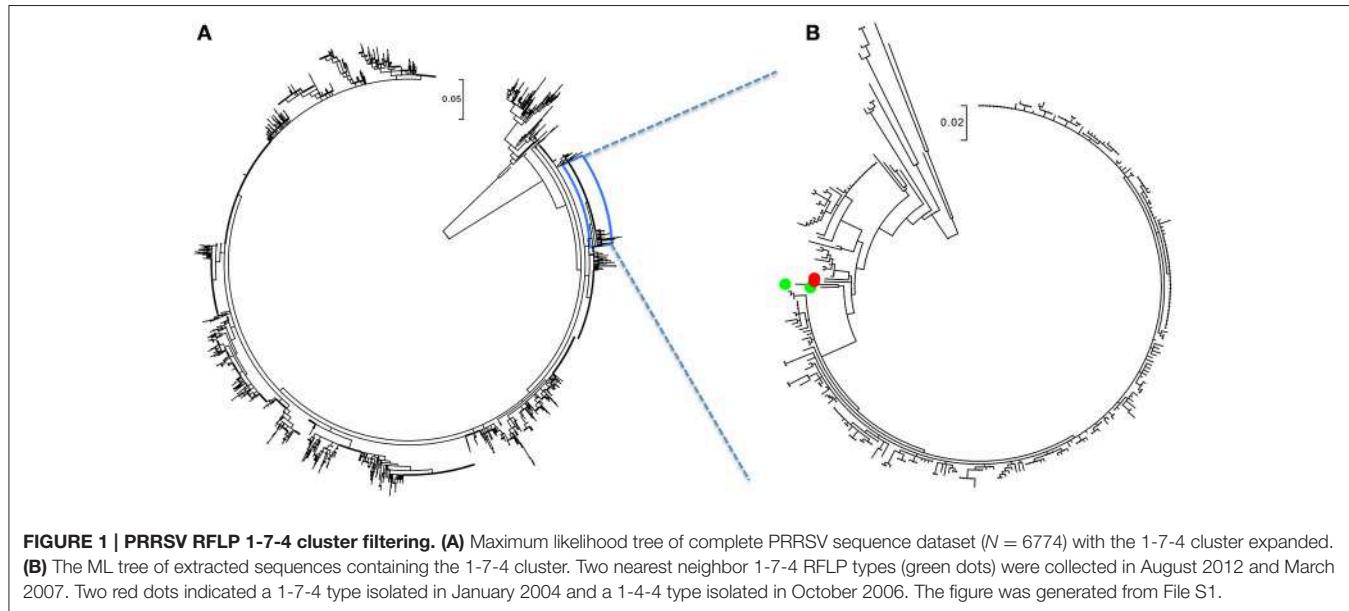
## RESULTS AND DISCUSSION

### Preliminary Phylogenetic Analysis

The ML analysis was performed to screen-out unrelated strains and because, although RFLP nomenclature is typically used to refer to PRRSV strains, the RFLP method is not an accurate discriminator of phylogenetic relations. As a result of the ML analysis, a total of 288 sequences, with isolation dates between September 2003 and March 2015, formed a phylogenetic branch shown in **Figure 1**. Within the branch a single, monophyletic clade of 241 sequences obtained in 15 months, between January 2014 and March 2015, stood out (**Figure 1**, Table S1). Those 241 were identified by the dominant RFLP-type, 1-7-4, whereas the other 45 genetically related strains belonged to a number of other RFLP types. Two nearest neighbor 1-7-4 RFLP types (depicted as green dots in **Figure 1**, Table S1) were collected in August 2012 and March 2007, whereas the two red dots indicated a 1-7-4 type isolated in January 2004 and a 1-4-4 type isolated in October 2006 (**Figure 1**, Table S1).

### Divergence-Time, Growth Rate, and Population Size

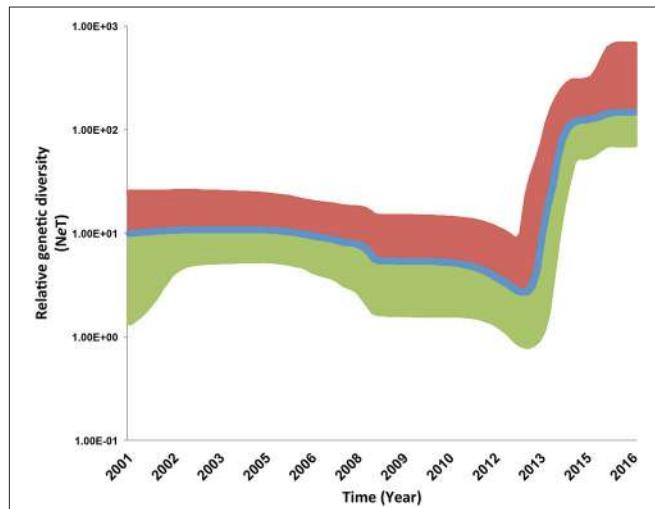
For the PRRSV ORF5 sequence dataset isolated between September 2003 and March 2015, the BF comparisons significantly favored the parametric expansion node-age coalescent model, indicating that the population size of the current 1-7-4 type PRRSVs was under rapid increase with an estimated mean growth rate of 1.02 (95% highest posterior density, HPD, from 0.59 to 1.46) and mean evolutionary



rate of  $3.27 \times 10^{-3}$ /site/year (95% HPD from  $2.37 \times 10^{-3}$  to  $4.27 \times 10^{-3}$ ), which lays within the range of previously estimated evolutionary rates for PRRSVs isolated from different geographical locations and period of times (Forsberg, 2005; Nguyen et al., 2014; Chaikhumwong et al., 2015). However, analysis of the virus population dynamics revealed a distinct continuous increase in the genetic diversity of the virus in March 2015, with no signs of population decline. This corresponds to the current increase of PRRSV incidence throughout the regional production systems in the U.S. (Figure 2). Our findings suggest the rate, or speed, at which the number of PRRSVs in the population increased, sometimes referred to as growth rate, was higher, compared to earlier phylogenetic relatives. This higher growth rate may suggest expanding diversity, and an unusual continuous increase in the relative genetic diversity over time, compared to those earlier phylogenetic relatives. That finding may be attributed to an evolutionary drift that resulted from either continuous circulation or maintenance within the production region, or recombination events with field viruses migrated from other production regions (Wang et al., 2015). An earlier study also suggested that this expanding diversity behavior of newly emerging strains is attributed to environmental factors associated with the continuous changes in swine husbandry practices rather than intrinsic factors within the host species (Murtaugh et al., 2010). The estimated divergence time for this sequence dataset was September 1996 (95% HPD, July 1986–December 2001), which completely overlaps with the TMRCA of sequences isolated from system C (Table 1). The youngest divergence time estimated for the viruses isolated from system A was August 2009 (95% HPD, December 2007–August 2011; Table 1).

## Viral Dispersal History between Regional Systems

The asymmetric variants of the discrete-trait model did not achieve full convergence, even after increasing the number of



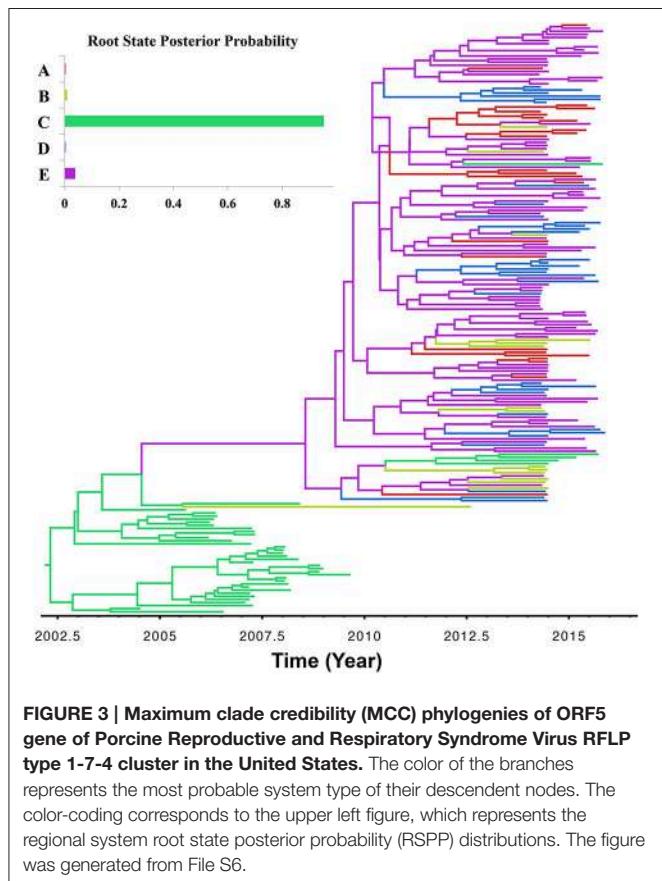
**FIGURE 2 | Bayesian Skyline plots (BSP) illustrating temporal changes in the relative genetic diversity of Porcine Reproductive and Respiratory Syndrome Virus RFLP type 1-7-4 isolated between September 2003 and March 2015 in the United States estimated from the ORF5 gene sequences.** Line plots summarize estimates of the effective population size ( $N_{eT}$ ), a measure of genetic diversity, for ORF5 gene segment; the shaded regions correspond to the 95% HPD (Upper, Red; Median, Blue; Lower, Green).

MCMCs to  $1 \times 10^{10}$  cycles; and therefore were discarded from the subsequent analyses. Our BF comparisons suggested that the symmetric UCED branch-rate model had the largest log-marginal likelihood ( $BF > 25$ ), and hence, was chosen as the best fitting phylodynamic model for ORF5 gene regions (Table S2). This result suggested that unidirectional spread of the virus between systems, when designated as origin and destination, had no significant role in the evolution of the currently circulating PRRSV. Figure 3 shows the ORF5 RSPP along with the time-scaled MCC tree (Figure S1). We also

**TABLE 1 |** Estimated TMRCAs for the PRRSV ORF5 gene sequences isolated within each system.

System	Mean TMRCA*	Lower 95% HPD	Upper 95% HPD
A	Aug-2009	Dec-2007	Aug-2011
B	Aug-2006	Nov-2003	Aug-2009
C	Sep-1996	Jul-1989	Dec-2001
D	Jul-2009	Jan-2007	Aug-2011
E	Jul-2009	Jan-2007	Aug-2011
Overall	Sep-1996	Jul-1989	Dec-2001

\*Time to the most recent common ancestor (TMRCA).



generated a KML file to demonstrate the temporal dynamics and spatial diffusion of the virus between systems (File S5). System C was strongly supported as the most likely regional system of origin for the currently circulating RFLP type 1-7-4 with a substantially large RSPP of 0.95. Divergence-time estimates under the discrete trait model indicated that the viral dispersal event from system C was initiated in September of 2000 (95% HPD, July 1999–December 2002). Significant ( $BF > 6$ ) nonzero rates for the dispersal routes between systems are summarized in **Table 2**. Our results suggest that the most significant routes of virus exchange were estimated exclusively between system C and all other remaining systems. Interestingly, no significant routes of viral exchange were estimated between

**TABLE 2 |** Bayes factor (BF) tests for non-zero transition rates between system type states.

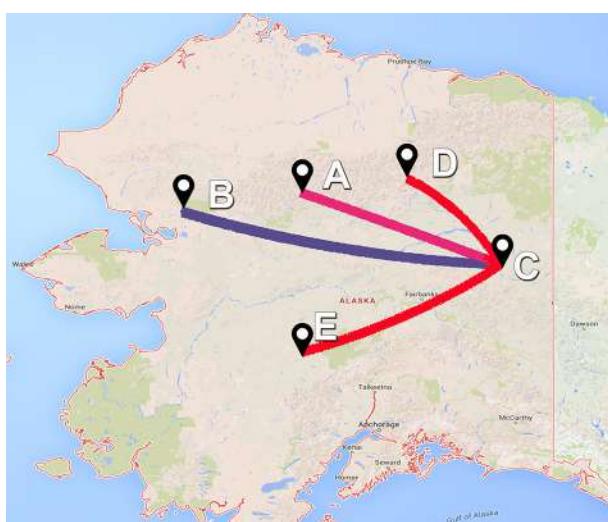
BF	Between
30529	C E
30529	C D
1695	C A
9	C B

BF-values  $>6$  indicate significant rates of directional exchange between systems.

systems other than C (**Figure 4**). Uncertainty and statistical analyses for validating the fit of the sequence data to the selected discrete-trait phylodynamic models are summarized in **Table 3**. The KL-value suggests that the data under the selected discrete phylodynamic model was able to generate RSPPs that are substantially different from the underlying priors and thus the posterior tree is statistically robust. Furthermore, the AI and PS tests rejected the null hypothesis of no association between sampled system and the structure of the phylogeny ( $P < 0.05$ ). This strongly suggests that the geographical distribution of swine systems are indeed having a significant role in shaping the phylogeny of endemic and newly emerging PRRSV in the US. This role mainly relies on the characteristics of the hog transportation network between systems (Shi et al., 2013; Thanapongtharm et al., 2014; Brar et al., 2015).

## Virus Transmission Patterns in a System

There were two reasons for exploring the role of farm type in viral transmission, (1) to demonstrate how discrete traits may be incorporated in the analysis, and (2) to test the biological soundness of the model results, given that one would expect PRRSV spread to occur mostly from sow farms into other types of farms, following the natural flow of animals. BF comparisons indicated that the Asymmetric UCED branch-rate model with reversible transitions provided the best fit for ORF5 gene regions (Bayes factor  $> 25$  for the log marginal likelihood; Table S2). Sow farms were the most likely ancestral farm type for the currently circulating type 1-7-4 PRRSV (RSPP = 0.95; **Figure 5**; Figure S2). Our divergence-time estimates suggest PRRSV originated in sow farms approximately in September of 1999 (95% HPD, July 1997–December 2001), and that it was maintained and circulated in sow farms until now. Only one significant nonzero rate transmission route was observed exclusively ( $BF > 6$ ) from sow farms to all other farms. However, most branches of the MCC tree under the farm type phylodynamic model were weakly supported (Branch rate posterior probability  $< 0.6$ ). In addition, the low KL-value under the farm type model was substantially less robust, when compared to the systems model (**Table 3**). This is because, small KL divergence statistic values between any prior and posterior probability distributions indicate that the data contain little information regarding the value of the selected parameter, and therefore, its posterior probability distribution will be similar to the corresponding prior probability distribution (Lemey et al., 2009). Furthermore, the AI and PS test failed to reject the null hypothesis of no association between farm type and the structure of the phylogeny ( $P > 0.05$ ). This is expected



**FIGURE 4 | Bayes factor (BF) test for significant non-zero rates in ORF5 gene of Porcine Reproductive and Respiratory Syndrome Virus RFLP type 1-7-4 cluster in the United States.** Only rates supported by a BF greater than six are indicated. The color of lines correspond to the probability of the inferred transmission rates. Blue and red line gradients indicate relatively weak to strong support, respectively. Site locations for the five systems (A–E) were anonymous and therefore latitude and longitude locations were placed in Alaska.

because the typical structure of swine farms in the US is farrow-to-weaning, which in turn segregates breeding pigs from growing pigs, and thus, makes sow farms more likely as sources of virus spread through pig movement than growing pigs sites from which most pigs go to market (Jeong et al., 2014). Therefore, and as suggested elsewhere, sow farms are more likely sources of transmission and maintenance of newly emerging PRRSVs (Kwong et al., 2013).

## Value to Participants, the Swine Industry, and Society

The veterinarians and pork producers who voluntarily share the disease status and location of their farms are vanguards in food production. By doing so, the individual participant or a particular farm risks being identified, either correctly, or incorrectly, as being a source of virus evolution and spread to other farms. And yet, the nature and structure of the swine industry is much more responsible for pathogen movement than any individual farm. That is, weaned pigs are transported away from the sow farm to allow pathogens such as PRRSV to be more effectively eliminated from the sow herd. This means that growing pigs must be transported to nursery and finishing sites and by doing so, pathogens are also conveniently moved around the country. Secondly, health is difficult to maintain if growing pig sites become too large. Therefore, we have a distributed system of growing pig sites which also lends to pathogens being moved around the country. Finally, farms might be using live virus vaccination in the short term to reduce clinical impact and aid in the elimination of field virus in the long term. So it is a bold decision to share data for a study such as this. There is a greater

good being pursued by these industry leaders. They voluntarily share their premises identities and pathogen status in the interests of national disease control such that we might detect emerging pathogens earlier than otherwise and take actions accordingly. Work such as reported in this paper is on the cusp of a new era of disease control.

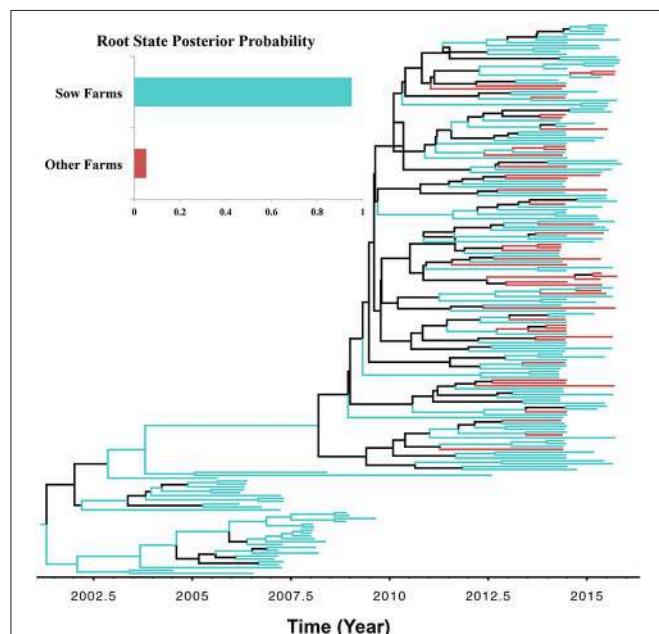
## Considerations and Future Applications of the Method

The methodological approach presented here entailed several compromises, including: (1) imprecise epidemiological information related to the discrete traits investigated, and (2) incomplete and biased sampling of PRRSV ORF5 sequences. For the first, we demonstrate the impact of the accuracy and availability of epidemiological information on the MCC trees (Figures 3, 5) and their posterior inferences. This impact on the performance of phylodynamic models has been discussed elsewhere (Chaikhumwang et al., 2015). However, this issue is chronic in the context of surveillance data and almost impossible to avoid in practical reality of animal disease surveillance (Perez et al., 2011). Therefore, rigorous analysis of a selected Bayesian phylodynamic model (i.e., assessing fit and uncertainty) is essential before deriving conclusions from their posterior inferences. For the second, inferences under the phylodynamic models assume that we have either a complete or random sample of sequence data. In the present case, this requires that the PRRSV sequences were collected randomly with respect to time (between 1999 and 2015) with their corresponding epidemiological information. Like most phylogenetic studies, our data were from a convenience sample and might suffer from strongly biased samples. The impact of these departures from random sampling on the estimates is difficult to quantify (Alkhamis et al., 2015). However, our study is based on all available sequence data from our participants for the ORF5 gene associated with the currently circulating RFLP type 1-7-4 epidemic in the US, and therefore reflects our best understanding based on the available data. It is worth noting that despite the unequal number of sequences obtained from different systems (Table S1), our posterior inferences for dispersal of the virus between system was not biased toward systems with more included sequences, such as E ( $n = 120$ ) and D ( $n = 55$ ), when compared to C ( $n = 52$ ). This constitutes an example for the utility and robustness of such methods in the context of molecular surveillance of swine diseases.

Bayesian phylodynamic models have not yet been widely accepted as a resource by veterinary agencies to support disease surveillance, control and prevention strategies. This is attributed by part to the intensive computational requirements of the methods presented here. For example, we were unable to assess the topology of 6774 sequences using BEAST due to the lack of sufficient computational resources. Instead, we used the traditional ML method to help in identifying the key cluster of interest, while reducing the computational requirements of the Bayesian analyses used to address our main hypotheses. That said, computational resources are in continuous improvement in terms of speed and cost, and therefore, in the near future the presented analytical pipeline can be completely transformed

**TABLE 3 | Uncertainty and statistical analyses for assessing the fit of the viral data to the discrete phylodynamic models.**

Model	KL-value <sup>a</sup>	Mean values	AI <sup>b</sup>	95% CI	P-values	PS <sup>c</sup>	95% CI	P-values
Systems	2.9	Observed	14.54	(13.28, 15.66)	0.02*	124.8	(117.0, 132.0)	0.01*
		Null	16.32	(15.09, 17.45)		132.75	(127.70, 136.83)	
Production Type	0.83	Observed	9.9	(8.76, 11.04)	0.61	73.9	(68.0, 80.0)	0.69
		Null	9.63	(8.53, 10.84)		72.7	(68.80, 75.94)	

<sup>a</sup>Kullback–Leibler (KL) divergence.<sup>b</sup>Association index (AI).<sup>c</sup>Parsimony score (PS).\*Statistically significant (*p*-value < 0.05).

**FIGURE 5 | Maximum clade credibility (MCC) phylogenies of ORF5 gene of Porcine Reproductive and Respiratory Syndrome Virus RFLP type 1-7-4 cluster in the United States.** The color of the branches represents the most probable farm type of their descendent nodes. The color-coding corresponds to the upper left figure, which represents the production type root state posterior probability (RSPP) distributions. Black branches in the tree indicate posterior probability < 0.60. The figure was generated from File S7.

to Bayesian statistical framework. However, previous use of such methods on avian influenza and the Ebola epidemics demonstrated the ability of phylodynamic methods to shed novel insights into the evolutionary epidemiology of infectious diseases and provide support for decisions regarding animal and public health (Lam et al., 2012; Pybus et al., 2013; Alizon et al., 2014). Our phylodynamic analyses of a PRRSV ORF5 sequence dataset and associated epidemiological information, in an endemic country like the U.S., were in agreement with previous inferences about the demographic histories and population growth patterns of viral lineages and sub-lineages of the virus in the U.S. (Shi et al., 2010). Bayesian phylodynamic models show one remarkable improvement compared to traditional methods, namely, they make use of associated epidemiological information, such as time

and place of isolation, to infer genetic relations. The inclusion of information on nucleotide substitution schemes obtained from the data, allowing for different model assumptions to assess the degree of genetic relatedness under time-scaled phylogenies, has provided a robust strategy, for example, to distinguish between potentially related PRRSV strains detected in air samples and swine farms in high and low swine density regions (Brito et al., 2014). In the analysis here, we incorporated time of prior isolation to reconstruct the phylogenetic dendrogram, hence, making use of temporal distances to infer genetic relations. This approach can help to shed further light on several evolutionary and epidemiological characters of endemic PRRSV. Furthermore, extended phylodynamic models can provide insights on the ancestral origins of the outbreak between swine systems (e.g., the ancestral system or herd type) and spatio-temporal progression of an epidemic. These inferences could be used, for example, to identify viral dispersion routes that correspond with transportation patterns involving high PRRSV risk.

## CONCLUSION

Classical phylogenetic methods such as neighbor-joining or maximum likelihood trees, provide limited inferences about the evolution of important pathogens and ignore important evolutionary parameters and uncertainties, which in turn limits decision making related to surveillance, control, and prevention resources. However, in this study, we illustrated the applications and potential of phylodynamic methods as tools for molecular surveillance of food animal viruses by assessing the evolution of newly emerging PRRSVs in the U.S. We analyzed different epidemiological and evolutionary aspects of a recently collected ORF5 gene sequence dataset. Using coalescence and discrete trait phylodynamic models, we obtained a phylogeny adjusted for many important epidemiological parameters such as space, time, and host type. Furthermore, we were able to (1) infer population growth and demographic history of the virus, which aids in assessing the magnitude of epidemic progression; (2) identified the most likely ancestral system, which aids in guiding risk-based surveillance activities; and (3) modeled viral transmission patterns between systems and farm types, which sheds important insights about viral transmission dynamics between and within swine herds. Accordingly, incorporating phylodynamic analyses as a standard tool for the molecular surveillance of swine diseases might support the development of more effective

economically rational policy decisions for the control of PRRSV in high-risk systems. However, investments must be mobilized toward improving genomic databases and building efficient bioinformatics and computational infrastructures, which are the base requirements for the field of applied phylogenetics (Scotch et al., 2011; Scotch and Mei, 2013).

## AUTHOR CONTRIBUTIONS

MA formulated the Bayesian models and was primarily responsible for report and manuscript preparation; AP provided interpretation on the use of epidemiological models, collaborated in the design of the analytical model, and assisted in manuscript preparation; MM contributed with the interpretation of results related with PRRSV genetic dynamics and manuscript preparation and editing; XW helped with data preparation and management; RM conceived the study, was responsible for communication with the industry and supervision of the entire

project, provided insight on the implementation of results at the field level, and assisted in manuscript preparation.

## ACKNOWLEDGMENTS

This study was funded in part by the University of Minnesota MnDrive and College of Veterinary Medicine Population Systems grant programs, by the National Pork Board, and by Boehringer Ingelheim Vetmedica, Inc. We also acknowledge the five anonymous systems and their herd veterinarians who graciously provided sequences and related data that made the study possible.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2016.00067>

## REFERENCES

- Alizon, S., Lion, S., Murall, C. L., and Abbate, J. L. (2014). Quantifying the epidemic spread of *Ebola virus* (EBOV) in Sierra Leone using phylogenetics. *Virulence* 5, 825–827. doi: 10.4161/21505594.2014.976514
- Alkhamis, M. A., Moore, B. R., and Perez, A. M. (2015). Phylogenetics of H5N1 highly pathogenic avian influenza in Europe, 2005–2010: potential for molecular surveillance of new outbreaks. *Viruses* 7, 3310–3328. doi: 10.3390/v7062773
- Alonso, C., Murtaugh, M. P., Dee, S. A., and Davies, P. R. (2013). Epidemiological study of air filtration systems for preventing PRRSV infection in large sow herds. *Prev. Vet. Med.* 112, 109–117. doi: 10.1016/j.prevetmed.2013.06.001
- Benfield, D. A., Nelson, E., Collins, J. E., Harris, L., Goyal, S. M., Robison, D., et al. (1992). Characterization of swine infertility and respiratory syndrome (SIRS) virus (isolate ATCC VR-2332). *J. Vet. Diagn. Invest.* 4, 127–133. doi: 10.1177/104063879200400202
- Bielejec, F., Rambaut, A., Suchard, M. A., and Lemey, P. (2011). SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. *Bioinformatics* 27, 2910–2912. doi: 10.1093/bioinformatics/btr481
- Brar, M. S., Shi, M., Murtaugh, M. P., and Leung, F. C. (2015). Evolutionary diversification of type 2 porcine reproductive and respiratory syndrome virus. *J. Gen. Virol.* 96, 1570–1580. doi: 10.1099/vir.0.000104
- Brito, B., Dee, S., Wayne, S., Alvarez, J., and Perez, A. (2014). Genetic diversity of PRRS virus collected from air samples in four different regions of concentrated swine production during a high incidence season. *Viruses* 6, 4424–4436. doi: 10.3390/v6114424
- Cameron, A. R., and Baldock, F. C. (1998). A new probability formula for surveys to substantiate freedom from disease. *Prev. Vet. Med.* 34, 1–17. doi: 10.1016/S0167-5877(97)00081-0
- Chaikhunwong, P., Tantituvanont, A., Tripipat, T., Tipsombatboon, P., Piriayapongsa, J., and Nilubol, D. (2015). Dynamics and evolution of highly pathogenic porcine reproductive and respiratory syndrome virus following its introduction into a herd concurrently infected with both types 1 and 2. *Infect. Genet. Evol.* 30, 164–174. doi: 10.1016/j.meegid.2014.12.025
- Chen, J. Z., Peng, J. M., Bai, Y., Wang, Q., Liu, Y. M., Zhang, Q. Y., et al. (2015). Characterization of two novel porcine reproductive and respiratory syndrome virus isolates with deletions in the GP2 gene. *Vet. Microbiol.* 176, 344–351. doi: 10.1016/j.vetmic.2015.01.018
- Cho, J. G., Deen, J., and Dee, S. A. (2007). Influence of isolate pathogenicity on the aerosol transmission of Porcine reproductive and respiratory syndrome virus. *Can. J. Vet. Res.* 71, 23–27.
- Corzo, C. A., Mondaca, E., Wayne, S., Torremorell, M., Dee, S., Davies, P., et al. (2010). Control and elimination of porcine reproductive and respiratory syndrome virus. *Virus Res.* 154, 185–192. doi: 10.1016/j.virusres.2010.08.016
- Dea, S., Gagnon, C. A., Mardassi, H., Pirzadeh, B., and Rogan, D. (2000). Current knowledge on the structural proteins of porcine reproductive and respiratory syndrome (PRRS) virus: comparison of the North American and European isolates. *Arch. Virol.* 145, 659–688. doi: 10.1007/s007050050662
- Dokland, T. (2010). The structural biology of PRRSV. *Virus Res.* 154, 86–97. doi: 10.1016/j.virusres.2010.07.029
- Drummond, A. J., and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214. doi: 10.1186/1471-2148-7-214
- Drummond, A. J., Rambaut, A., Shapiro, B., and Pybus, O. G. (2005). Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* 22, 1185–1192. doi: 10.1093/molbev/msi103
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Forsberg, R. (2005). Divergence time of porcine reproductive and respiratory syndrome virus subtypes. *Mol. Biol. Evol.* 22, 2131–2134. doi: 10.1093/molbev/msi208
- Goldberg, T. L., Hahn, E. C., Weigel, R. M., and Scherba, G. (2000). Genetic, geographical and temporal variation of porcine reproductive and respiratory syndrome virus in Illinois. *J. Gen. Virol.* 81, 171–179. doi: 10.1099/0022-1317-81-1-171
- Grenfell, B. T., Pybus, O. G., Gog, J. R., Wood, J. L., Daly, J. M., Mumford, J. A., et al. (2004). Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* 303, 327–332. doi: 10.1126/science.1090727
- Griffiths, R. C., and Tavaré, S. (1994). Ancestral inference in population. *Genetics* 9, 307–319. doi: 10.1214/ss/1177010378
- Holtkamp, D. J., Kliebenstein, J. B., Neumann, E. J., Zimmerman, J. J., Rotto, H. F., Yoder, T. K., et al. (2013). Assessment of the economic impact of porcine reproductive and respiratory syndrome virus on United States pork producers. *J. Swine Health Prod.* 21, 72–84. Available online at: <https://www.aasv.org/shap/issues/v21n2/v21n2p72.pdf>
- Jeong, J., Aly, S. S., Cano, J. P., Polson, D., Kass, P. H., and Perez, A. M. (2014). Stochastic model of porcine reproductive and respiratory syndrome virus control strategies on a swine farm in the United States. *Am. J. Vet. Res.* 75, 260–267. doi: 10.2460/ajvr.75.3.260
- Kapur, V., Elam, M. R., Pawlovich, T. M., and Murtaugh, M. P. (1996). Genetic variation in porcine reproductive and respiratory syndrome virus isolates in the midwestern United States. *J. Gen. Virol.* 77(Pt 6), 1271–1276. doi: 10.1099/0022-1317-77-6-1271
- Kullback, S., and Leibler, R. (1951). On information and sufficiency. *Ann. Math. Stat.* 22, 79–86. doi: 10.1214/aoms/1177729694

- Kwong, G. P., Poljak, Z., Deardon, R., and Dewey, C. E. (2013). Bayesian analysis of risk factors for infection with a genotype of porcine reproductive and respiratory syndrome virus in Ontario swine herds using monitoring data. *Prev. Vet. Med.* 110, 405–417. doi: 10.1016/j.prevetmed.2013.01.004
- Lam, T. T., Hon, C. C., Lemey, P., Pybus, O. G., Shi, M., Tun, H. M., et al. (2012). Phylogenetics of H5N1 avian influenza virus in Indonesia. *Mol. Ecol.* 21, 3062–3077. doi: 10.1111/j.1365-294X.2012.05577.x
- Lanfear, R., Calcott, B., Ho, S. Y., and Guindon, S. (2012). Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol. Biol. Evol.* 29, 1695–1701. doi: 10.1093/molbev/mss020
- Larochelle, R., D'Allaire, S., and Magar, R. (2003). Molecular epidemiology of porcine reproductive and respiratory syndrome virus (PRRSV) in Quebec. *Virus Res.* 96, 3–14. doi: 10.1016/S0168-1702(03)00168-0
- Lemey, P., Rambaut, A., Drummond, A. J., and Suchard, M. A. (2009). Bayesian phylogeography finds its roots. *PLoS Comput. Biol.* 5:e1000520. doi: 10.1371/journal.pcbi.1000520
- Maddison, W. P., and Maddison, D. R. (2011). *Mesquite Version 2.75: A Modular System for Evolutionary Analysis*. Available online at: <http://mesquitemproject.org> (Accessed February 20, 2014).
- Martin, D. P., Lemey, P., Lott, M., Moulton, V., Posada, D., and Lefevre, P. (2010). RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* 26, 2462–2463. doi: 10.1093/bioinformatics/btq467
- Martín-Valls, G. E., Kvistgaard, L. K., Tello, M., Darwich, L., Cortey, M., Burgara-Estrella, A. J., et al. (2014). Analysis of ORF5 and full-length genome sequences of porcine reproductive and respiratory syndrome virus isolates of genotypes 1 and 2 retrieved worldwide provides evidence that recombination is a common phenomenon and may produce mosaic isolates. *J. Virol.* 88, 3170–3181. doi: 10.1128/JVI.02858-13
- MathWorks. (2012). *Matlab 2013a*. Natick, MA.
- Minin, V. N., Bloomquist, E. W., and Suchard, M. A. (2008). Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol. Biol. Evol.* 25, 1459–1471. doi: 10.1093/molbev/msn090
- Murtaugh, M. P., Stadejek, T., Abrahante, J. E., Lam, T. T., and Leung, F. C. (2010). The ever-expanding diversity of porcine reproductive and respiratory syndrome virus. *Virus Res.* 154, 18–30. doi: 10.1016/j.virusres.2010.08.015
- Neumann, E. J., Kliebenstein, J. B., Johnson, C. D., Mabry, J. W., Bush, E. J., Seitzinger, A. H., et al. (2005). Assessment of the economic impact of porcine reproductive and respiratory syndrome on swine production in the United States. *J. Am. Vet. Med. Assoc.* 227, 385–392. doi: 10.2460/javma.2005.227.385
- Nguyen, V. G., Kim, H. K., Moon, H. J., Park, S. J., Chung, H. C., Choi, M. K., et al. (2014). A Bayesian phylogeographical analysis of type 1 porcine reproductive and respiratory syndrome virus (PRRSV). *Transbound. Emerg. Dis.* 61, 537–545. doi: 10.1111/tbed.12058
- Parker, J. (2008). *BaTS Version 1.0: Bayesian Tip-Significant Testing*. Available online at: <http://evolve.zoo.ox.ac.uk/Evolve/BaTS.html>; <http://mesquitemproject.org> (Accessed January 5, 2015).
- Perez, A., Alkhamis, M., Carlsson, U., Brito, B., Carrasco-Medanic, R., Whedbee, Z., et al. (2011). Global animal disease surveillance. *Spat. Spatiotemporal Epidemiol.* 2, 135–145. doi: 10.1016/j.sste.2011.07.006
- Perez, A. M., Davies, P. R., Goodell, C. K., Holtkamp, D. J., Mondaca-Fernández, E., Poljak, Z., et al. (2015). Lessons learned and knowledge gaps about the epidemiology and control of porcine reproductive and respiratory syndrome virus in North America. *J. Am. Vet. Med. Assoc.* 246, 1304–1317. doi: 10.2460/javma.246.12.1304
- Pybus, O. G., Fraser, C., and Rambaut, A. (2013). Evolutionary epidemiology: preparing for an age of genomic plenty. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 368, 20120193. doi: 10.1098/rstb.2012.0193
- Raftery, A., Newton, M., Satagopan, J., and Krivitsky, P. (2007). *Estimating the Integrated Likelihood via Posterior Simulation Using the Harmonic Mean Identity*. Oxford: Oxford University Press.
- Rambaut, A. (2012). *FigTree v.1.4.2: Tree Figure Drawing Tool*. Available online at: <http://tree.bio.ed.ac.uk/software/figtree> (Accessed August 11, 2014).
- Rambaut, A., Suchard, M. A., Xie, D., and Drummond, A. J. (2014). *Tracer v1.6*. Available online at: <http://beast.bio.ed.ac.uk/Tracer>
- Razavi, N. (2008). *Kullback-Leibler Divergence*. Available online at: <http://www.mathworks.com/matlabcentral/fileexchange/20688-kullback-leibler-divergence> (Accessed March 14, 2014).
- Rosendal, T., Dewey, C., Friendship, R., Wootton, S., Young, B., and Poljak, Z. (2014). Association between PRRSV ORF5 genetic distance and differences in space, time, ownership and animal sources among commercial pig herds. *Transbound. Emerg. Dis.* doi: 10.1111/tbed.12253. [Epub ahead of print].
- Rowland, R. R., and Morrison, R. B. (2012). Challenges and opportunities for the control and elimination of porcine reproductive and respiratory syndrome virus. *Transbound. Emerg. Dis.* 59(Suppl. 1), 55–59. doi: 10.1111/j.1865-1682.2011.01306.x
- Scotch, M., and Mei, C. (2013). Phylogeography of swine influenza H3N2 in the United States: translational public health for zoonotic disease surveillance. *Infect. Genet. Evol.* 13, 224–229. doi: 10.1016/j.meegid.2012.09.015
- Scotch, M., Sarkar, I. N., Mei, C., Leaman, R., Cheung, K. H., Ortiz, P., et al. (2011). Enhancing phylogeography by improving geographical information from GenBank. *J. Biomed. Inform.* 44(Suppl. 1), S44–S47. doi: 10.1016/j.jbi.2011.06.005
- Shi, M., Lam, T. T., Hon, C. C., Murtaugh, M. P., Davies, P. R., Hui, R. K., et al. (2010). Phylogeny-based evolutionary, demographical, and geographical dissection of North American type 2 porcine reproductive and respiratory syndrome viruses. *J. Virol.* 84, 8700–8711. doi: 10.1128/JVI.02551-09
- Shi, M., Lemey, P., Singh Brar, M., Suchard, M. A., Murtaugh, M. P., Carman, S., et al. (2013). The spread of type 2 Porcine Reproductive and Respiratory Syndrome Virus (PRRSV) in North America: a phylogeographic approach. *Virology* 447, 146–154. doi: 10.1016/j.virol.2013.08.028
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Suchard, M. A., Weiss, R. E., and Sinsheimer, J. S. (2001). Bayesian selection of continuous-time Markov chain evolutionary models. *Mol. Biol. Evol.* 18, 1001–1013. doi: 10.1093/oxfordjournals.molbev.a003872
- Thanapongtharm, W., Linard, C., Pamaronan, N., Kawkalong, S., Noimoh, T., Chanachai, K., et al. (2014). Spatial epidemiology of porcine reproductive and respiratory syndrome in Thailand. *BMC Vet. Res.* 10:174. doi: 10.1186/s12917-014-0174-y
- Tun, H. M., Shi, M., Wong, C. L., Ayudhya, S. N., Amonsin, A., Thanawonguwech, R., et al. (2011). Genetic diversity and multiple introductions of porcine reproductive and respiratory syndrome viruses in Thailand. *Virology* 38, 164. doi: 10.1186/1743-422X-8-164
- Wang, X., Marthaler, D., Rovira, A., Rossow, S., and Murtaugh, M. P. (2015). Emergence of a virulent porcine reproductive and respiratory syndrome virus in vaccinated herds in the United States. *Virus Res.* 210, 34–41. doi: 10.1016/j.virusres.2015.07.004
- Yoon, S. H., Kim, H., Kim, J., Lee, H. K., Park, B., and Kim, H. (2013). Complete genome sequences of porcine reproductive and respiratory syndrome viruses: perspectives on their temporal and spatial dynamics. *Mol. Biol. Rep.* doi: 10.1007/s11033-013-2802-1. [Epub ahead of print].
- Zimmerman, J., Yoon, K.-J., and Neumann, E. (2003). *PRRS Compendium Producer Edition*. Des Moines, IA: National Pork Board. Available online at: <http://www.pork.org/production-topics/swine-health/2003-prrs-compendium-producer-edition/>

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Alkhamis, Perez, Murtaugh, Wang and Morrison. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Advances and Limitations of Disease Biogeography Using Ecological Niche Modeling

Luis E. Escobar<sup>1,2\*</sup> and Meggan E. Craft<sup>1</sup>

<sup>1</sup> Veterinary Population Medicine, College of Veterinary Medicine, University of Minnesota, St. Paul, MN, USA, <sup>2</sup> Minnesota Aquatic Invasive Species Research Center, University of Minnesota, St. Paul, MN, USA

## OPEN ACCESS

**Edited by:**

Yuji Morita,

Aichi Gakuin University, Japan

**Reviewed by:**

Pelayo Acevedo,

Instituto de Investigación en Recursos Cinegéticos (UCLM-CSIC-JCCM),

Spain

Mike Taylor,

University of Auckland, New Zealand

**\*Correspondence:**

Luis E. Escobar

lescobar@umn.edu

**Specialty section:**

This article was submitted to

Infectious Diseases,

a section of the journal

Frontiers in Microbiology

**Received:** 31 January 2016

**Accepted:** 15 July 2016

**Published:** 05 August 2016

**Citation:**

Escobar LE and Craft ME (2016) Advances and Limitations of Disease Biogeography Using Ecological Niche Modeling. *Front. Microbiol.* 7:1174.  
doi: 10.3389/fmicb.2016.01174

Mapping disease transmission risk is crucial in public and animal health for evidence based decision-making. Ecology and epidemiology are highly related disciplines that may contribute to improvements in mapping disease, which can be used to answer health related questions. Ecological niche modeling is increasingly used for understanding the biogeography of diseases in plants, animals, and humans. However, epidemiological applications of niche modeling approaches for disease mapping can fail to generate robust study designs, producing incomplete or incorrect inferences. This manuscript is an overview of the history and conceptual bases behind ecological niche modeling, specifically as applied to epidemiology and public health; it does not pretend to be an exhaustive and detailed description of ecological niche modeling literature and methods. Instead, this review includes selected state-of-the-science approaches and tools, providing a short guide to designing studies incorporating information on the type and quality of the input data (i.e., occurrences and environmental variables), identification and justification of the extent of the study area, and encourages users to explore and test diverse algorithms for more informed conclusions. We provide a friendly introduction to the field of disease biogeography presenting an updated guide for researchers looking to use ecological niche modeling for disease mapping. We anticipate that ecological niche modeling will soon be a critical tool for epidemiologists aiming to map disease transmission risk, forecast disease distribution under climate change scenarios, and identify landscape factors triggering outbreaks.

**Keywords:** spatial epidemiology, prediction, fundamental niche, infectious disease, risk map

## INTRODUCTION

Human history has been shaped by information captured in maps. Concepts such as disease occurrence, epidemics, and outbreaks implicitly have a geographic context. In fact, early stages of epidemiology attempted to understand disease occurrence linking disease cases (e.g., human cholera) with environmental features (e.g., a street pump) in a spatial perspective (Koch and Denike, 2009). Understanding and anticipating the “where” of an outbreak may be a valuable tool for effective public health interventions (Frieden, 2013) as well as for animal health. Thus, disease mapping is key in understanding and anticipating disease occurrence and generating visual tools for decision makers.

Ecology has been proposed as an additional discipline to assist the understanding of why a disease is present in a specific place, but is absent in another (Peterson, 2008). Epidemiology and ecology share goals: the World Health Organization [WHO] (2015) defines epidemiology as “*...the study of the distribution and determinants of health-related states or events...*” while Krebs (1972) defines ecology as the study of the distribution and abundance of species. Terminology for ecology and epidemiology is similar as both disciplines attempt to study and understand the distribution of organisms and their abundance. Such organisms may include plants, animals, or even parasites (i.e., pathogenic or not). At this point, both fields can complement each other; ecology for example has grown through analytical methods and conceptual bases, on the other hand epidemiology has developed an impressive data compilation (Anderson, 1991). Unfortunately, both fields usually work in isolation (Manlove et al., 2016).

Disease biogeography is an emerging field aiming to study the geography of diseases including pathogens, vectors, reservoirs, and susceptible hosts. Disease biogeography links ecology and epidemiology by applying analytical tools from distributional ecology for the study of epidemics. In modern ecology the concept of “niche” is key, thus, we cannot talk about disease biogeography without talking about the ecological niches of a pathogen. This review was inspired by two key publications in disease biogeography: “*Natural nidality of transmissible diseases, with special reference to the landscape epidemiology of zoonanthroposes*” by the Russian Academician Pavlovsky (1966) and “*Biogeography of diseases: A framework for analysis*” by the American Professor Peterson (2008). These pioneers clearly define ideas, terms, and examples of the field of disease biogeography. Pavlovsky and Peterson were the first in developing the use of the ecological niche approach to the study of infectious disease systems, using a diversity of diseases and scenarios to explain why diseases are not distributed at random and that some environmental factors may explain their occurrence in time and space at coarse (i.e., fundamental niche) or local scales (i.e., disease nidus or realized niche). While Pavlovsky’s contribution was observational in nature, Peterson developed a conceptual and methodological framework to develop and interpret quantitative analyses on the biogeography of diseases. Their work can be seen as a theoretical base for applied landscape epidemiology and spatial epidemiology. We aim to provide an overview of current tools and important steps when mapping diseases using ecological niche modeling approaches. This review is intended to serve as an introductory guide for epidemiologists and researchers not familiar with ecological niche modeling techniques.

## DISEASE BIOGEOGRAPHY AS A NEW PARADIGM IN EPIDEMIOLOGY

Here we will refer to the agents responsible for causing infectious disease as parasites, including micro- and macro-parasites (Hatcher and Dunn, 2011). Some of these parasites may not cause disease in the host (e.g., non-pathogenic strains of *Vibrio cholerae*). In addition to the study of parasites, epidemiologists

could be interested also in the vectors and reservoirs (Estrada-Peña et al., 2014) to understand how the parasites are dispersed and maintained in the landscape, respectively. Once ecological features of parasites are defined, their geographic distribution can be expressed in the form of maps, usually in the form of disease risk maps (Peterson, 2008). In this review we will explore the field of ecological niche modeling for understanding disease distribution and posterior disease mapping.

In the late 20th century, in response to the limited understanding of disease dynamics from a biological perspective, a new paradigm was proposed in epidemiology: eco-epidemiology (Susser and Susser, 1996a). Eco-epidemiology is based on the need to understand infectious diseases using an ecological approach to help anticipate disease distribution. Although almost three decades have passed since this idea was formally proposed, there are still ambiguities in this approach. While Susser and Susser (1996b) correctly state that disease systems include a set of interconnected environmental factors, they did not define and delimit the eco-epidemiology concept *per se*, which could be a cause of the slow adoption of this term more broadly in epidemiology and ecology (only ~200 articles in 20 years with this term were found in PubMed). The initial description of eco-epidemiology includes the *Chinese boxes* idea, stating that disease systems are a set of factors with a coherent hierarchy structure, thus, alteration in the parasite system may cause disease only if factors at higher levels of the structure are affected (Susser and Susser, 1996b). However, from a biological perspective, this approach seems inaccurate considering that in nature, when considering parasites in ecosystems, there is not a chain-like configuration. Instead, parasites in natural communities interact with several species in the food web, thus, parasite systems appear as part of an interconnected network of species (Hudson et al., 2006). This was exemplified by Pavlovsky (1966) who used a series of examples on the study of infectious diseases including yellow fever (*Flavivirus*), plague (*Yersinia pestis*), tularemia (*Francisella tularensis*), and leishmaniasis (*Leishmania* spp.) to demonstrate the complexities in the biological interactions of parasite systems. Later, Peterson (2008) proposed disease biogeography as the branch of biology related to the geography of infectious diseases; disease biography aims to identify the factors associated with disease occurrence allowing us to understand and potentially predict epidemics.

## THE ECOLOGICAL NICHE OF PARASITES

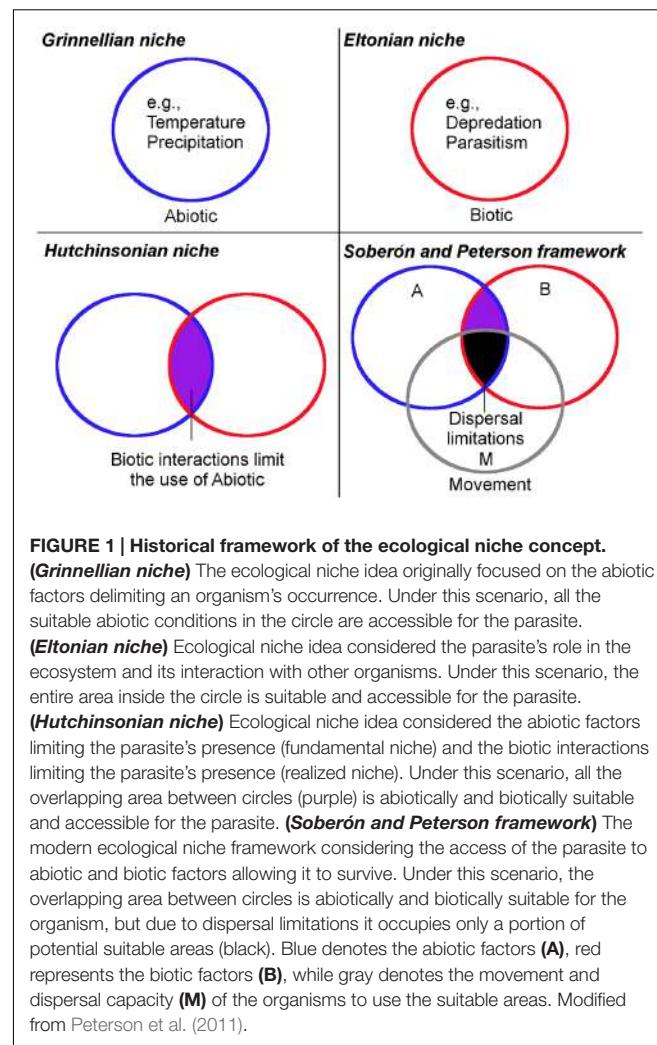
### The Term Niche

Infectious diseases are, by definition, the complex association between at least two organisms: pathogen and host. Infectious disease requires the presence of all key actors in a disease transmission system (e.g., parasite, vector, susceptible host; Peterson, 2006b). Identifying the environmental factors which allow the presence of one of these actors in the disease system elucidates the ecology and geography of a specific infectious disease (Peterson, 2007). Recognizing patterns of species distributions and identifying the specific environmental

requirements of species to persist in the long term has been extensively studied in ecology with an empirical and theoretical body that supports the study of distributional ecology of organisms. However, disease biogeography is just one component of multifaceted questions in understanding the ecology of parasite transmission, and has been poorly addressed in biodiversity research (Peterson et al., 2011).

A disease system may include a pathogen species, a vector species, and a host, or it may be more complex including a vast number of competent vectors and host species in the same locality, sharing the environmental conditions suitable for them, sharing their *ecological niches* (Soberón, 2007). Furthermore, the parasites' ecological niche is linked to its geographic distribution (Soberón and Nakamura, 2009). Establishing environmental variables able to track areas of potential distribution of diseases was proposed in epidemiology in the mid-20th century, using rudimentary techniques to correlate parasites with environmental factors. It was found that diseases do not occur randomly in space; hence the concept of infectious disease *nidality* was defined in epidemiology as the feature of an infectious disease to be constrained under specific environmental conditions (Pavlovsky, 1966). The words *niche* and *nidality* have the root word *nidus* which means nest. The association between the parasite's suitable environments and geographic ranges is the base of the ecological niche modeling field. If the environmental factors suitable for a parasite are available outside the known range, an epidemic may appear in a novel (suitable) area. This phenomenon is well known in invasion biology (Peterson and Vieglais, 2001), while in epidemiology the invasion by parasites is known as *pathogen pollution* (Anderson et al., 2004) but has not been explored quantitatively in detail.

The term "niche" was used in ecology by Grinnell (1917) to refer to the combination of environmental factors present in a species' range; he discussed how these factors may restrict the distribution of species. Even though the word niche was first used for a bird, the concept was successfully adopted by ecologists and is today a key concept in ecology. However, the niche concept suffered from ambiguity and incorrect use (Godsoe, 2010; McInerny and Etienne, 2012; Warren, 2012). In fact, the concept had four main stages before its current definition (Figure 1). Grinnell (1917) used the term niche to refer to the environmental factors required by a species for its distribution (*Grinnellian niche*; Figure 1). Then years later Charles S. Elton defined niche as the role of a species in an ecosystem and its interactions with other species (*Eltonian niche*; Figure 1). Grinnellian and Eltonian definitions clearly were based on different points of view. G. Evelyn Hutchinson attempted to reduce the ambiguity of the niche concept, differentiating it as the fundamental niche and realized niche (*Hutchinsonian niche*; Figure 1). The fundamental niche was proposed as a hypervolume of environmental variables that allow the species to exist without immigration, while the realized niche incorporates the idea of the portion of the fundamental niche actually used by the species due to negative (e.g., competition) or positive (e.g., facilitation) biological interactions with other organisms (Bruno et al., 2003). Finally, Soberón and Peterson (2005) defined the niche concept using the **BAM** framework and a body of empirical and theoretical



**FIGURE 1 | Historical framework of the ecological niche concept.**

**(Grinnellian niche)** The ecological niche idea originally focused on the abiotic factors delimiting an organism's occurrence. Under this scenario, all the suitable abiotic conditions in the circle are accessible for the parasite.

**(Eltonian niche)** Ecological niche idea considered the parasite's role in the ecosystem and its interaction with other organisms. Under this scenario, the entire area inside the circle is suitable and accessible for the parasite.

**(Hutchinsonian niche)** Ecological niche idea considered the abiotic factors limiting the parasite's presence (fundamental niche) and the biotic interactions limiting the parasite's presence (realized niche). Under this scenario, all the overlapping area between circles (purple) is abiotically and biotically suitable and accessible for the parasite. **(Soberón and Peterson framework)** The modern ecological niche framework considering the access of the parasite to abiotic and biotic factors allowing it to survive. Under this scenario, the overlapping area between circles is abiotically and biotically suitable for the organism, but due to dispersal limitations it occupies only a portion of potential suitable areas (black). Blue denotes the abiotic factors (**A**), red represents the biotic factors (**B**), while gray denotes the movement and dispersal capacity (**M**) of the organisms to use the suitable areas. Modified from Peterson et al. (2011).

background explaining this framework. Thus, the current ecological niche term refers to the environmental conditions in which a species can maintain populations in the long term without need of immigration. The species, however, may not use its entire niche due to biological or dispersal limitations. According to Soberón and Peterson (2005), a missing component in previous definitions of niche was the dispersal capability and movement potential of species to reach suitable areas (BAM framework of Soberón and Peterson; Figure 1). They suggest that a species may have a broad fundamental niche, but may be unable to use it entirely, due to biogeographic limitations (e.g., mountains, rivers, oceans acting as barriers). All the different niche concepts reflect the considerable debate to define the term; now a clear and delimited definition of a niche is available and employed in ecological niche modeling (Warren, 2012).

## The “BAM” Framework in Disease Systems

The BAM diagram incorporates the dispersal capacity of parasites when describing their niches. Dispersal abilities of parasites

are key to understanding disease distributions. For example, chikungunya virus (*Alphavirus*) was absent in the Americas prior to 2013 due to an ocean functioning as natural barrier between Asia and the Americas (Van Bortel et al., 2014). However, global movement of humans allowed dispersal of chikungunya into South, Central, and North America, as well as the Caribbean, and resulted in a successful virus invasion. This invasion demonstrates that the Americas have environmental conditions falling inside the ecological niche of chikungunya, but the virus was not previously present in the Americas due to dispersal limitations. Once dispersal limitations are overcome or the natural population dynamics change, pandemics may appear (Pavlovsky, 1966; Hatcher and Dunn, 2011). To understand which areas have the potential for disease dispersal under an ecological niche approach, the **BAM** diagram is a powerful tool that helps to (i) understand ecological processes, (ii) design studies, and (iii) interpret experimental, virtual, or field research (Soberón and Peterson, 2005; Peterson, 2008). Peterson et al. (2011) discuss in detail the modern use of the ecological niche concept using examples and mathematical terms. Here we briefly describe the **BAM** framework.

### **“B”**

**B** refers to biotic factors shaping the distribution of the parasite (i.e., biocenose *sensu* Pavlovsky, 1966, or binomic *sensu* Hutchinson; Soberón, 2007). This is a critical component in the ecological niche of parasites considering that biotic interactions between hosts and vectors may promote or limit parasite occurrence even in environmentally suitable (i.e., **A**) and accessible areas (i.e., **M**). Some biotic factors such as host nutrition, density, behavior (e.g., cultural practice of kissing dead bodies that may promote Ebola infections; Woźniak-Kosek et al., 2015), and co-infections may benefit parasite presence. On the other hand, some biotic factors may limit parasite occurrence including host immunity (e.g., acquired immunity to *V. cholerae* after an epidemic will reduce cases in the next cholera outbreak; Koelle et al., 2005) and behavior (e.g., use of protective measures to avoid sexually transmitted diseases). The biotic component is critical to understand the ecology of parasite transmission, and their effects are evident when developing studies at fine geographic scales.

Pavlovsky (1966) studied the occurrence of parasites using an ecological niche approach and proposes the concept of “*micronidus*.” The *micronidus* is the term for the biotic factors indispensable for the parasite’s cycle at a very fine scale; such factors, however, should not affect estimations of the ecological niche at coarse scales (Pavlovsky, 1966). The *micronidus* simply occupies a specific portion of the parasite’s niche. Peterson et al. (2011) identify these factors acting at a local scale and also ignore them when modeling a species’ ecological niche at a coarse scale with successful predictions. Empirical evidence suggests that the *micronidus* may not be key for the parasites’ ecological niche in general terms (Maher et al., 2010), such as when considering climate conditions. Assuming that local biological interactions are meaningless at coarse scales is referred to as the “*Eltonian Noise Hypothesis*” (Peterson et al., 2011). The idea behind this hypothesis is that biotic interactions at the individual level (e.g.,

host immunity) or the microhabitat required by a specific phase in the parasite cycle (e.g., the humidity in the host burrow required during the metamorphosis of *Phlebotomus* vectors) may play a minor role when estimating the parasites’ niche at coarse scales. Furthermore, biotic interactions are important only when studying diseases at very fine spatial scales such as when studying transmission dynamics within a population (Peterson et al., 2011). This series of evidence and assumptions supports the idea of mapping diseases based on climatic variables or other environmental features.

In fact, when estimating the ecological niche of a generalist parasite, it is evident that climate conditions are crucial for parasite establishment and biotic interactions may be ignored at coarse scales. For instance, plague bacterium occurs in consistent and measurable climatic conditions; in other words, the environmental signature allows us to predict plague occurrence in North American wild mammals with no information about biological interactions (Maher et al., 2010). For example, Maher et al. (2010) suggest that, after assessing 72 plague reservoirs, plague occurs under predictable spatial and environmental situations, and host species involved in the transmission cycle are less relevant to maintain the parasite permanence than climate. Parasites with broad niches (i.e., generalist species) may maintain disease cycles under diverse environmental conditions and consequently may affect a broad range of taxa (e.g., plague, influenza (*Influenzavirus A*), leptospira (*Leptospira* spp.); Pavlovsky, 1966; Tong et al., 2012). Parasites that use different species of hosts and vectors are termed *polyhostal* and *polyvectored* respectively (Pavlovsky, 1966) and can be modeled including all the actors in the system or based on disease cases only (Peterson, 2007).

Biological interactions between species at very fine scales are complex. In co-infections, two parasite species within a host may even interact; one parasite may limit the presence of other parasite species (Pavlovsky, 1966). For example, after *in vivo* experiments of multiple inoculations of the parasites *Brucella suis* and *Coxiella burnetii* in guinea pigs, Mika et al. (1959) suggested that one parasite species may show apparent competition-like interactions. In such experiments, infected guinea pigs showed faster recovery or even unapparent infections of *C. burnetii* when the *B. suis* was present, as opposed to those with single infections—suggesting that the presence of *B. suis* is protective for *C. burnetii*. This mechanism is recognized and used in the poultry industry through the use of non-pathogenic bacteria to promote competitive exclusion against pathogenic strains of *Salmonella* spp. (Revolledo et al., 2006).

### **“A”**

The **A** factor on the **BAM** diagram represents the abiotic conditions limiting survival of parasite populations in the long term (i.e., geobiocenose *sensu* Pavlovsky, 1966, or scenopoetic *sensu* Hutchinson; Soberón, 2007). The area of overlap between the abiotic factors **A** with biotic factors **B** denotes which factors allow for parasite presence (Figure 1). Examples of abiotic conditions critical for parasite survival may include temperature and humidity (e.g., bacterial diseases in plants), solar radiation (e.g., viruses outside the host), and soil chemistry (e.g., fungi).

Thus, abiotic environmental factors allowing for the occurrence of parasites in a specific location can be measured at diverse spatial scales.

### **M**

Parasite occurrence may be constrained due to limited dispersal abilities (e.g., short movements of soft ticks) and biogeographic barriers (e.g., oceans). This key concept is termed movement or **M** in the **BAM** framework and considers the geographic accessibility of organisms. Changes in **M** affect the parasite's distribution and can be expressed as limitations of accessibility (e.g., dengue virus was absent from Pascua Island due to the isolation of this island, despite the vector being abundant; Perret et al., 2003) or an increased potential of accessibility (e.g., ballast waters increased dispersal of *V. cholerae*; McCarthy and Khambaty, 1994).

Identifying the environments suitable for an organism (**A**) is feasible at different spatial scales from petri dishes to continental extents (e.g., *V. cholerae*; see Huq et al., 1984 and Lobitz et al., 2000). Characterization of the biotic component (**B**), allowing for a parasite presence or absence, is much more complicated due to temporal-spatial dynamics and complexities of biotic interactions. In ecological niche models of large taxa (e.g., birds) the **B** component is usually neglected, ignoring biotic interactions in view of the robustness in predicting species occurrence, and based on the assumption that biological interactions are indistinguishable at coarse scales (Peterson et al., 2011). For parasites, however, their strong dependence on other organisms (i.e., the host) makes considering biotic interactions important in estimating their areas of occurrence. Thus, the more we understand about the natural history of a disease, the better the ecological niche estimation and interpretation of model outputs; however, modeling such complexities could be a challenge.

Parasites occur naturally with animals and plants and have key roles in ecosystems. The presence of a parasite in a host does not necessarily represent disease. In fact, there is growing evidence that parasite diversity is an indicator of ecosystem health (Hudson et al., 2006). When anthropogenic perturbations alter parasite cycles or communities, disease outbreaks can appear (Hatcher and Dunn, 2011). Parasites are usually considered negative in the context of human populations. In fact, a considerable amount of literature related to the distribution of parasites is developed only under disturbed/epidemic events; therefore limited knowledge exists about parasites in natural and non-disturbed conditions. Human societies should maintain pristine areas as reserves, including the greatest variety of biomes possible, to understand parasite ecology for epidemic prevention purposes (Pavlovsky, 1966).

## FROM DISEASE REPORTS TO DISEASE MAPS

The environmental space that a parasite is occupying (i.e., the existent fundamental niche; Peterson et al., 2011) can be expressed in terms of geography. By identifying the suitable environmental conditions for a parasite, we can identify the areas

in which a parasite can maintain populations in the long-term; this helps to understand the geographic distribution of parasites (Peterson, 2006b). Due to this link between the niche and the distribution of species, the terms ecological niche modeling and species distribution modeling are often used as interchangeable terms. However, niches are characterized in an environmental dimension, while geographic distributions are the expression of the ecological niche in the geography (Warren, 2012).

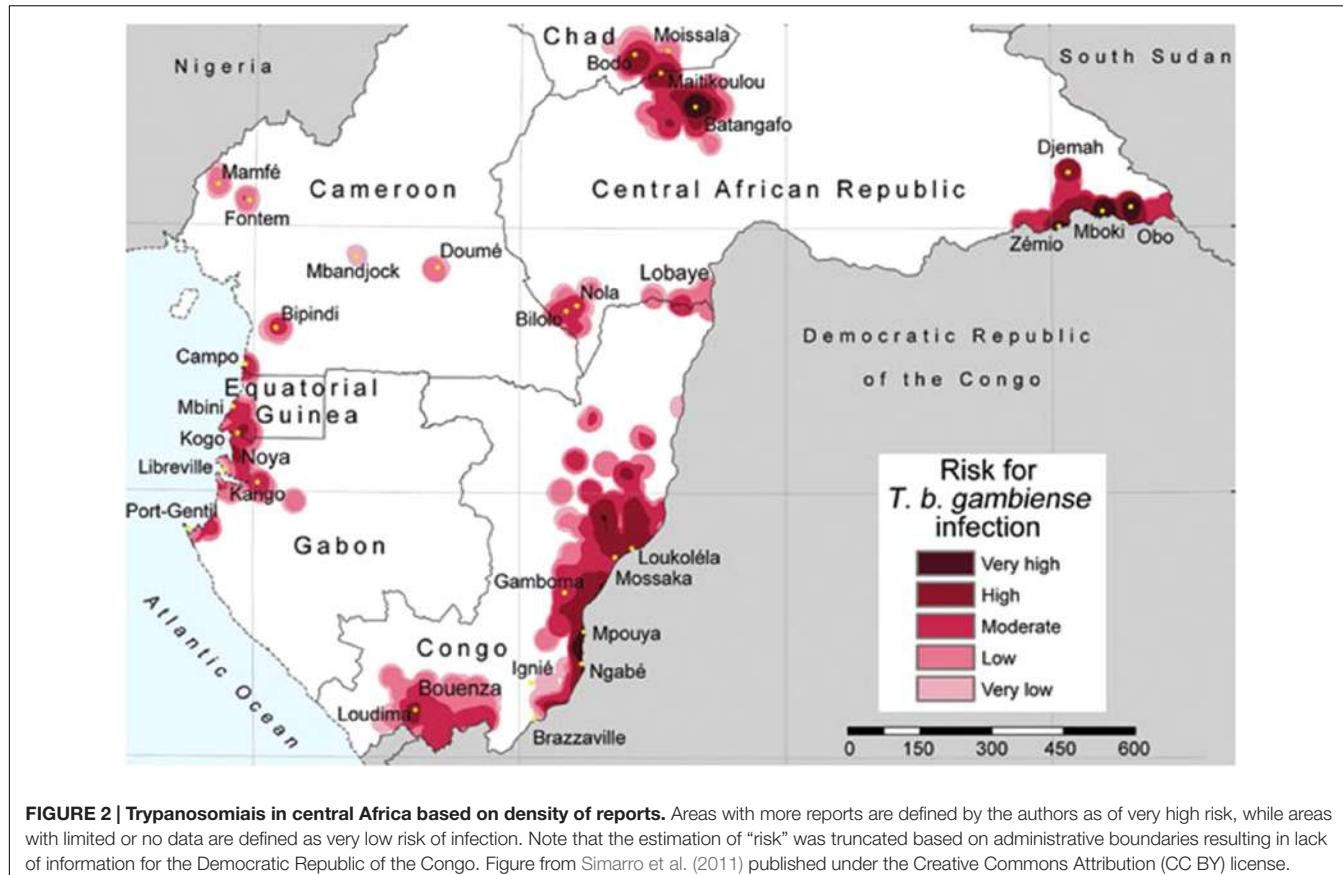
## Current Methods for Disease Mapping

Techniques to show disease distribution include choropleth maps (i.e., coloration of political/administrative units with colors according to categories of incidence, prevalence, or risk rates established *a priori*) and proportional symbols (i.e., symbols like circles with sizes classified according to predefined disease occurrence categories). These methods are data descriptive and easy to read and interpret, but fail to anticipate the parasite occurrence in areas where no data are available.

Analytical tools have improved since early disease mapping in the 19th century (Carpenter, 2011). However, when comparing John Snow's historical 1854 cholera map with current disease mapping based on density analyses with novel tools and software (e.g., Le Comber et al., 2011), it is evident that mapping approaches in epidemiology have not substantially improved. Indeed, epidemiology is dominated by studies using spatial density of cases, spatial interpolations of reports, and geographic distances to identify areas of potential disease-transmission risk (Auchincloss et al., 2012). These epidemiological techniques are a powerful source of information to show patterns of disease surveillance and reporting effort; however, they have several limitations in predicting disease risk. In fact, because spatial interpolation maps base estimations on available geographic coordinates solely, they should be considered as surveillance-effort maps instead of disease-transmission risk maps.

Using maps based solely on spatial interpolation to identify disease risk could be challenging. Mapping methods using spatial distances usually base their analyses on straight lines of geographic or even Euclidean distances (Auchincloss et al., 2012), which may fail to capture the biological realism of disease systems. Thus, spatial interpolation and cluster analyses are in essence data driven and prone to sampling bias effects. Spatial interpolations may attribute low parasite occurrence to an area with no data. This type of analysis is thus prone to miss areas of high disease risk because of surveillance gaps. For example, poor countries with limited epidemiological surveillance may appear healthier due to zero (i.e., lack of) cases reported, however, the real situation may include high disease incidence.

For example, recent research proposed risk levels of human Trypanosomiasis, a vector-borne protozoan parasite (Simarro et al., 2011), where risk estimations were based on the spatial density of human cases reported. To define close and distant cases authors proposed a 30 km radius—a pragmatic value neglecting the biology of the vector. In the Trypanosomiasis study, areas with high number of reports are defined as of “very-high risk,” while areas with no data are simply ignored by the model and assumed in the “very-low” or no risk categories (Figure 2). Additionally, the risk estimation was restricted to administrative



boundaries, failing to include the natural history of this disease caused by tsetse flies of the *Glossina* genus. This error was later replicated with the same data and method by Simarro et al. (2012) adding new countries. Both publications resulted in two isolated studies that did not provide a complete history on the biogeography of this disease in central Africa, but more importantly the studies proposed no risk in broad areas where no information is available, which may result in an incomplete or incorrect message to public health authorities.

Spatial interpolations for disease mapping, including kernel smoothing and kriging, attempt to describe biological mechanisms driving parasite spread among populations, but are strongly biased by surveillance effort (e.g., if most data were collected close to roads; Kadmon et al., 2004). Because disease maps may be used to guide surveillance and disease management (Stevens and Pfeiffer, 2011), controlling, or at least recognizing, sampling bias is critical. Additionally, spatial interpolations based solely on geographic coordinates assume that the landscapes where parasites occur are environmentally homogeneous, failing to provide explanations for the environmental processes and landscape variables triggering or limiting outbreaks.

## Spatial Interpolation vs. Environmental Interpolation

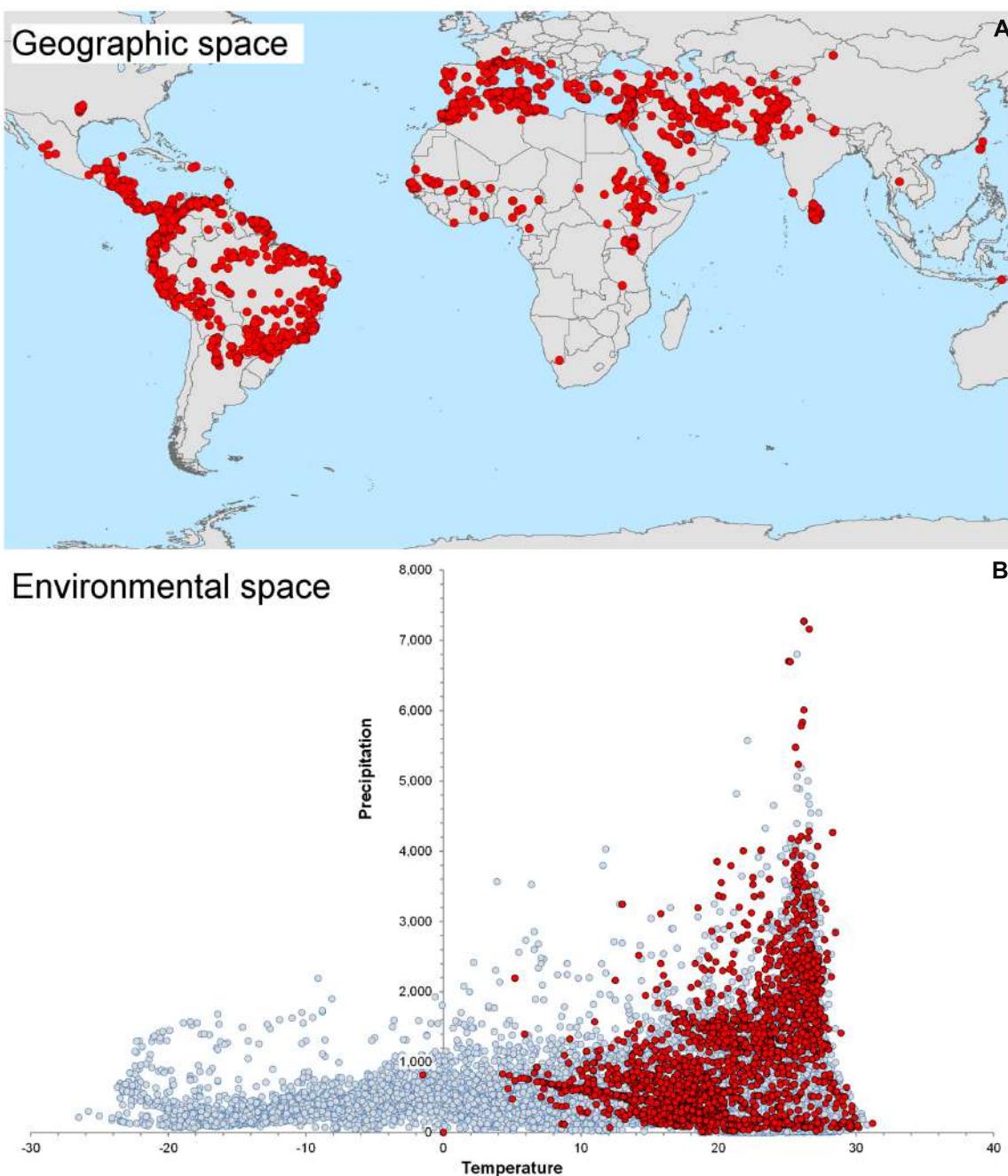
To improve on the limitations of disease maps based on density and distance of geographic coordinates solely, ecologists started

linking environmental variables with disease occurrence. Thus, environmental interpolations can be an alternative to spatial interpolations (Peterson, 2014). Environmental interpolations are the core approach in ecological niche modeling and include two main characteristics. The first is descriptive; ecological niche models attempt to identify the environment associated with the parasite's occurrence in the field or via laboratory experiments of physiological tolerance to specific environmental variables (Peterson et al., 2011). The second characteristic is predictive, searching across areas of interest to identify environmental combinations similar to those where the parasite occurs. Thus, while geographic interpolations occur in the geographic space, environmental interpolation is developed in multidimensional environmental scenarios. By using ecological niche modeling techniques we gain knowledge on the association of organisms with environmental variables of interest, contributing to our understanding of the parasite's ecology and geographic distribution (Peterson, 2006b). Additionally, with the knowledge obtained from few observations, inferences may allow us to identify areas environmentally suitable for the parasite in areas without reports available (Peterson et al., 2004).

We highlight the differences between spatial and environmental interpolation using data from the global burden of cutaneous leishmaniasis (Pigott et al., 2014a). The 6,426 cutaneous leishmaniasis occurrences were plotted in the geographic space using latitude and longitude as coordinates and

then were plotted in the environmental space using temperature and precipitation as coordinates (**Figure 3**). This procedure reduced the original 6,426 geographic coordinates to 1,964 single coordinates in environmental dimensions, allowing identification of the environmental space used by the species. We then modeled

the potential areas for the occurrence of this vector-borne disease using spatial and environmental interpolations (**Figure 4**). First we developed maps based on simple geographic interpolation using a density kernel estimation that identifies areas with high or low number of occurrences under a specified radius.



**FIGURE 3 | Global distribution of cutaneous leishmaniasis.** **(A)** Visualization of 6,426 cutaneous leishmaniasis occurrences (red points) in the geographic space. **(B)** Distribution of leishmaniasis in the environmental space. Some occurrences have identical environmental values and therefore resulted in 1,954 single occurrences in the environmental space (red points). Notice the diversity of environments available across the globe (gray points) and the consistent, narrow, predictable environmental space occupied by the disease. The environmental space (gray points) was generated using 10,000 random points globally to capture values of temperature (x axis) and precipitation (y axis). Data obtained from Hijmans et al. (2005) and Pigott et al. (2014a).

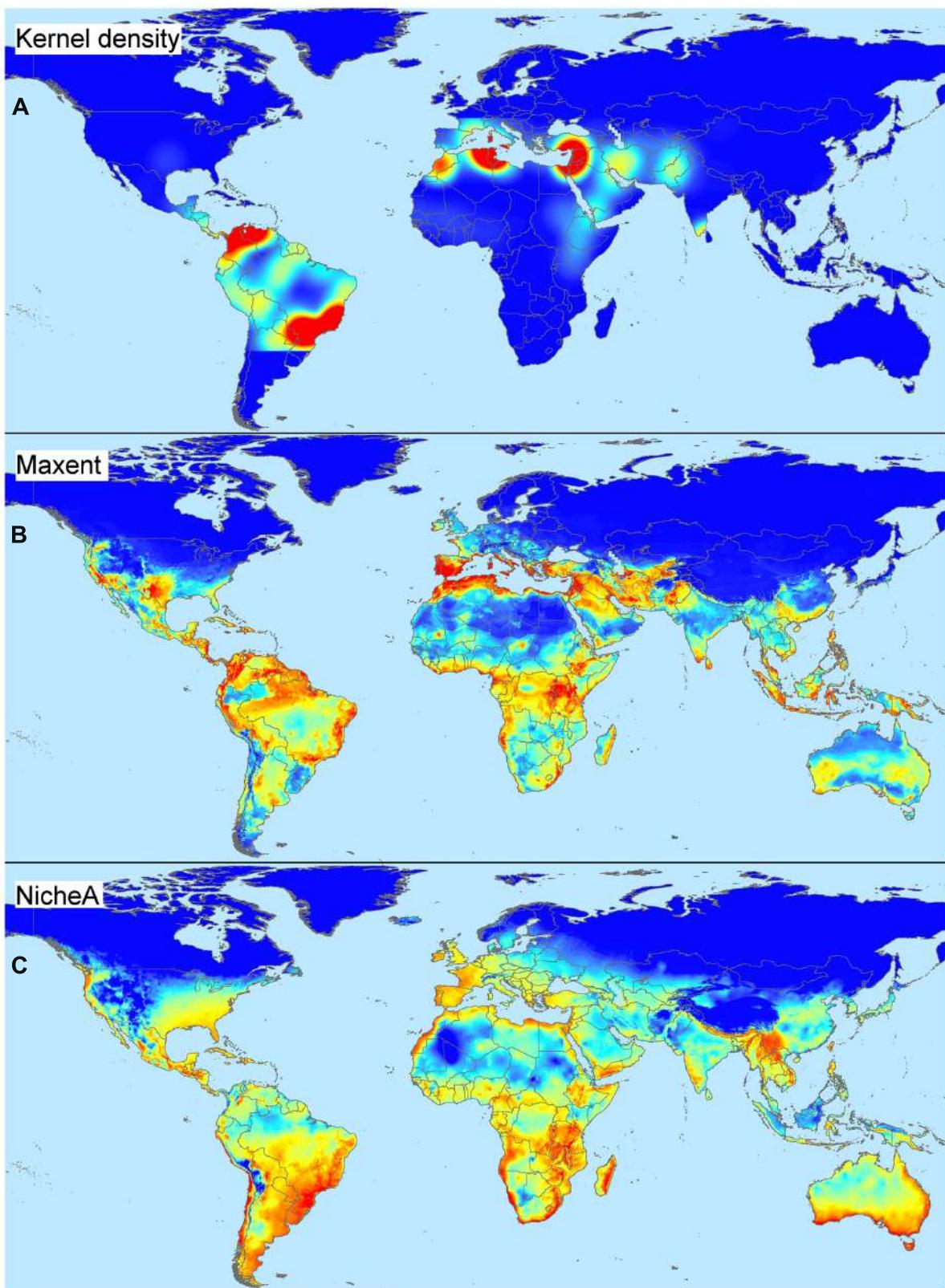


FIGURE 4 | Continued

**FIGURE 4 | Spatial and environmental interpolations of cutaneous leishmaniasis.** **(A)** Kernel density estimation based on leishmaniasis occurrences from **Figure 3A** falling in a pre-determined occurrence-distance radius, no environmental conditions are considered. Model constructed using default parameters in ArcGIS 10.2 (ESRI, 2015). Notice that areas proposed as of risk (red) are overfitted to locations with abundant disease reports, while areas with low or no reports are denoted as of low importance (see similarity with map in **Figure 3A**). The model does a poor job of predicting areas where data are absent (e.g., truncated areas in South America). **(B)** Environmental suitability index based on environmental similarity with sites where leishmaniasis was reported; environmental conditions are considered in this model. Model constructed using default in Maxent 3.3.3.k (Phillips et al., 2006) and bioclimatic variables Bio1 – Bio7 and Bio10 – Bio17 cell size of ~4.5 km (Hijmans et al., 2005). Notice that suitable areas for disease occurrence are predicted in areas with a lack of data. Areas of high suitability, however, mirror the areas with abundant disease reports; this is a form of model overfitting based on environmental values. **(C)** Environmental suitability index based on distance to the niche centroid; environmental conditions are considered. Model constructed using default parameters in NicheA 3.0 with the same environmental variables as above (Qiao et al., 2015a). Notice that suitable areas are estimated in the environmental space thereby reducing spatial overfitting; predictions do not reflect the number of occurrences, but their position in an environmental cloud provides the highest values to points in central areas of the environmental space and low values to most external disease reports (see distribution in the environmental space in **Figure 3B**). Estimations of low (blue) and high (red) represent values ranging between 0 – 7.25 (Kernel density), 0 – 0.86 (Maxent), and -1 – 0.99 (NicheA). Data obtained from Hijmans et al. (2005) and Pigott et al. (2014a).

Then we modeled the ecological niche of the disease using two different methods: Maxent, that identifies the association between occurrences and environmental variables weighted by the number of occurrences, and NicheA, that identifies the environmental space occupied by occurrences and weights the occurrences based on their position in the environmental space, thereby mitigating the effect of oversampled areas (**Figure 4**). In this exploration, spatial interpolations were restricted to denote high values only in areas with adequate data. The ecological niche models from Maxent and NicheA, based on environmental interpolations, found areas suitable for potential leishmaniasis occurrence even in areas with gaps of surveillance.

Ecological niche modeling is now commonly practiced in ecology and there are a number of sophisticated niche modeling tools available to analyze a wide range of datasets. Modeling techniques that link parasite occurrence with environmental variables include: (i) those requiring presence-only data like Bioclimatic Envelop Algorithm (BIOCLIM), Ecological Niche Factor Analysis (ENFA), and Niche Analyst (NicheA); (ii) regression models requiring presence plus true absences or pseudoabsence data as Boosting Regression Trees (BRT), Classification and Regression Trees (CART), Generalized Linear Models (GLM), Generalized Additive Models (GAM), and Random Forest (RF); and (iii) algorithms requiring presence-background data including Maximum Entropy (Maxent),  $n$ -dimensional hypervolume (Blonder et al., 2014), and Genetic Algorithm for Rule-set Production (GARP). There is no “best algorithm” that fits with all study case configurations. Instead, several algorithms must be assessed in each study case to identify those performing well under the specific conditions of the disease system and available data (Qiao et al., 2015b). These modeling algorithms have been explained and discussed with more detail elsewhere (Elith et al., 2006; Franklin, 2009; Peterson et al., 2011).

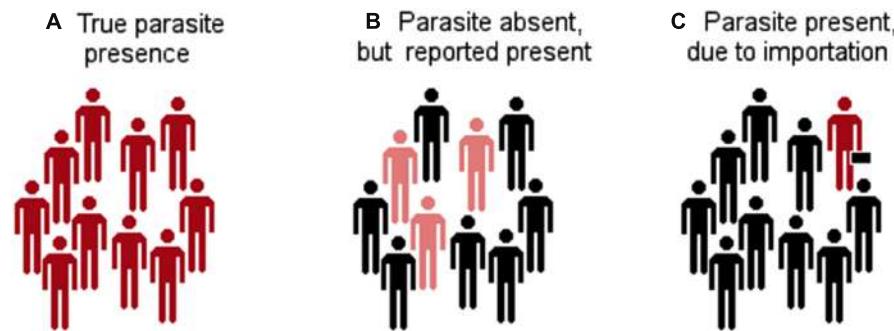
## REPORTS OF DISEASE PRESENCE

Ecological niche modeling generally needs records of sites where the parasite is present to link the parasite's occurrence with the environmental features chosen by the researcher. Presence records are critical and need to be accurate in terms of parasite identification and geolocation. However, reports of parasite presence may include some level of uncertainty (**Figure 5**). For example, the parasite may be present in a site and it could be

correctly identified and georeferenced. But in some instances the parasite may be reported as present in a location when in reality the parasite is absent. This could be due to incorrect diagnostic tests. This occurs with parasites that have sympatric and taxonomically close species with similar morphological or immunological characteristics. Another confounding factor is the report of the presence of a parasite in a site where, in fact, suitable conditions do not exist. This may occur in situations where the parasite was translocated by the host. For example, a human infected with Ebola in Africa can move to the Arctic in less than 24 h, in this simple case, reporting the disease detection in the Arctic may generate estimations that do not resemble the parasites' niche. Thus, using reports of parasite's presence including errors of identification and site of infection will generate inaccurate risk models. Additionally, georeferencing accuracy is an issue of disease mapping that deserves critical attention (Auchincloss et al., 2012), but has been neglected when modeling the potential distribution of infectious disease (but see Nakazawa et al., 2010; Peterson and Samy, 2016).

Algorithms could be calibrated using parasite, vector, or reservoir occurrences plus environmental information. Data on reservoir occurrence may come from the researcher's fieldwork, scientific literature, natural history museum collections, official public health agencies, and laboratories. Data for vectors exist but are scarce compared to data for vertebrate reservoirs (Peterson, 2014). Data for parasites are scarce and can be generated by the researcher or can be obtained from health agencies, scientific literature, or online repositories like Healthmap<sup>1</sup>, but need considerable data cleaning to reduce errors and uncertainty (Peterson, 2014). Georeferencing error in occurrence points and distance between them is also informative when determining the environmental variables required for ecological niche modeling. No magic recipe exists to establish the ideal or minimum number of reports for ecological niche model calibration; it simply depends on the research question, study design, the environmental variables considered, and data available. Something to keep in mind is to avoid the modeler's spatial-bias (i.e., bias implicit when thinking in the geographic space neglecting environmental dimensions; see **Figure 3**). For example, the number of occurrences used for model calibration could be numerous in the geographic space, but may be meaningless when considered in the environmental space. To

<sup>1</sup>[www.healthmap.org](http://www.healthmap.org)



**FIGURE 5 | Types of parasite presence reports.** **(A)** The parasite was identified in a host population and it is reported as present. **(B)** There are incorrect reports of parasite occurrence due to error in the identification of other parasites that generate cross reactivity to the diagnostic test. **(C)** Reports of parasite occurrence but the infections occurred in a different location to that included in the site of diagnosis (e.g., from a traveler). Red: infected with the parasite. Pink: not infected with the parasite but seropositive. Black: not infected.

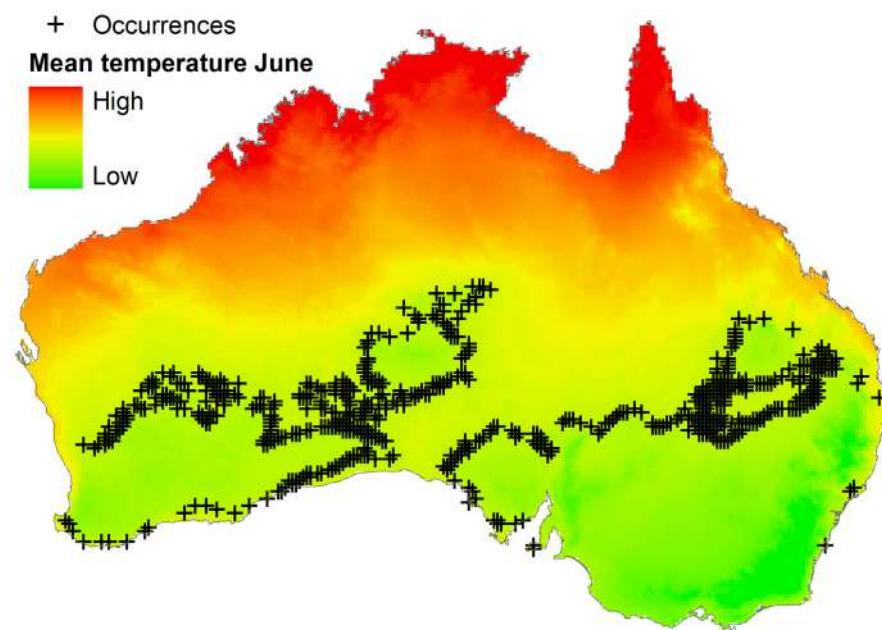
show this, more than 600 occurrence points were generated for a virtual parasite in mainland Australia, but when such points are considered in environmental terms, for example the mean temperature in June, all points have the same mean temperature value (i.e., 12.5°C; **Figure 6**). Thus, these abundant points from a geographic perspective represent a single point in environmental terms.

More points are always better during model calibration for more informed and less variable forecasts (Escobar et al., 2013), but a balance should exist between the number of geographic occurrences and the environmental representativeness of them. More important than the number of occurrences is their quality. Several studies have utilized all the available occurrence data of species in original format for model calibration without careful data curation (e.g., Brito-Hoyos et al., 2013; Bárcenas-Reyes et al., 2015). Indeed, failing to reduce pseudoreplicates (i.e., non-independent samples), and the consequent overrepresentation of environmental conditions, could produce models that simply reflect biases in the surveillance effort. **Figure 7** shows how a single environmental value could be overrepresented in a model due to sampling bias. In this example, it is evident how different data curation approaches and assumptions could vary in the use of occurrences from the original 45 occurrences to one occurrence per pixel or even a single pixel to summarize the same information. However, species found consistently in the same environmental space, with occurrences frequently falling in the same conditions, could be a classic case of an endemic specialist species of narrow niche (e.g., **Figure 6**). Thus, it is critical to differentiate between species with narrow niches and narrow niches resulting from sampling bias. For example, a forecast of bat-borne rabies in cattle in Mexico suffered model overfitting, resulting in the estimations of narrow areas with predictions of “high transmission risk” and areas with gaps of surveillance predicted of “low” risk (Bárcenas-Reyes et al., 2015). However, a reanalysis of the same data removing pseudorePLICATE occurrences and redundant variables, showed broad areas that were now predicted to be at risk of rabies occurrence in cattle (**Figure 8**).

## REPORTS OF DISEASE ABSENCE

Early models to link parasites with landscape features included logistic regressions. Logistic regressions, however, require the identification of locations with the presence and absence of the parasite. Most parasite presence data may be accurate in terms of taxonomic identification and georeferencing due to modern diagnostic methods and global positioning system devices; the correct identification of parasite’s absence may be, on the other hand, uncertain or incorrect (Peterson, 2014). Using incorrect absence data for model calibration may reduce the model fit by including locations where the parasite is or may be present but is reported as absent. A parasite may be reported as absent in a specific location due to many reasons (**Figure 9**). The parasite may be present in the host population, but it was simply undetected by the researcher (MacKenzie et al., 2002); the parasite may be present but it was eradicated recently; or the parasite could be present but biogeographic barriers do not allow it to use the suitable areas (**Figure 1**). Thus, calibrating ecological niche models of parasites, vectors, or hosts using absence data may fail to correctly capture the environmental signature of the target species.

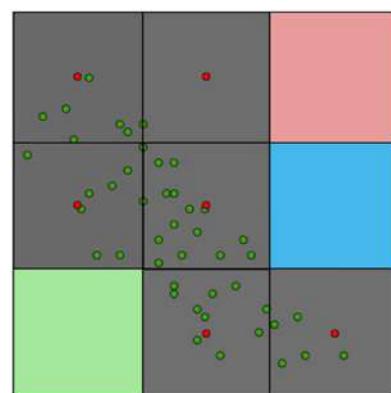
Because absence data are hard to collect from the field, some approaches create dummy absence data sets in order to provide regression models with the absence data required for calibration. Some of these approaches include the random generation of virtual absences across the study area; such virtual absences are termed *pseudoabsences* and lack biological meaning (Lobo et al., 2007). To mitigate the error implicit in models requiring absence data, new algorithms are available for mapping diseases using only robust reports of parasites’ occurrence including presence-only and presence-background algorithms. Descriptions of such techniques have received broad attention and are broadly accepted by the scientific community (Franklin, 2009; Peterson et al., 2011; Pliscott and Fuentes, 2011).



**FIGURE 6 | Distribution of occurrences based on a geographic dimension.** Occurrences of a virtual parasite ( $n = 636$ ) are dispersed across Australia (black crosses). Mean temperature for June is variables in the country (background colors), occurrences however, do not have environmental independence: all points fall in sites with identical temperature values (i.e.,  $12.5^{\circ}\text{C}$ ). Temperature data obtained from <http://www.worldclim.org/> (Hijmans et al., 2005).

## ENVIRONMENTAL VARIABLES USED IN ECOLOGICAL NICHE MODELING

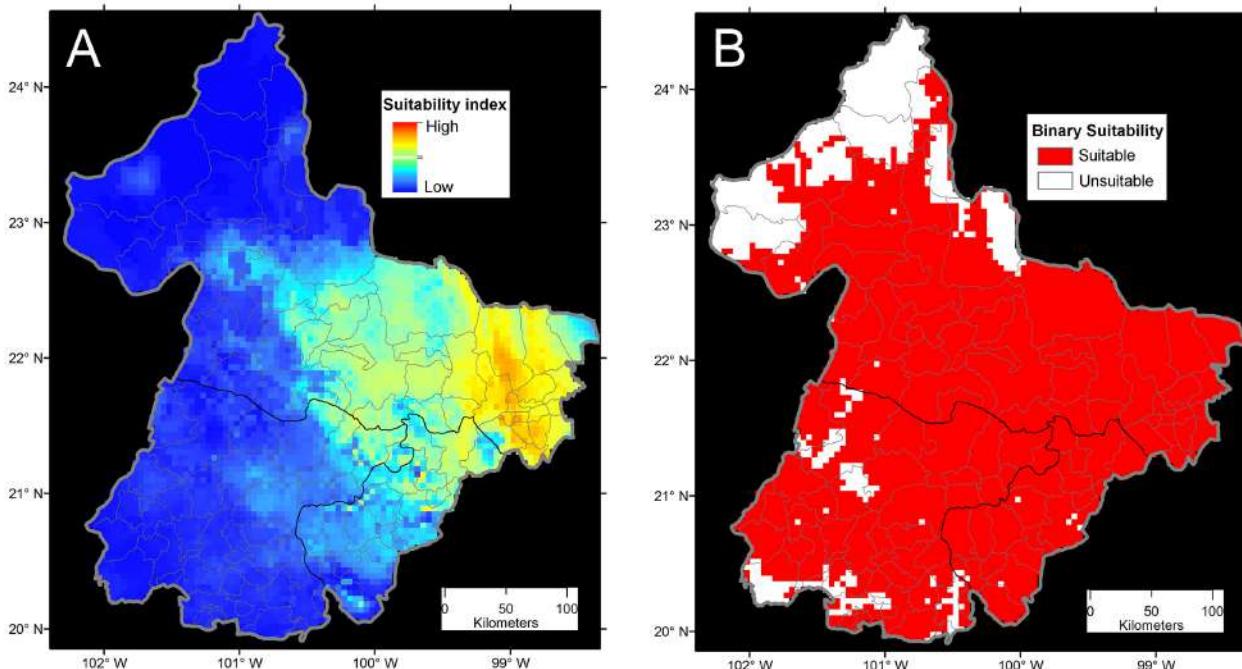
Different environmental variables exist at diverse spatial and temporal resolutions (Figure 10). The environmental variables selected should respond to the specific scientific question and should consider the parasite's biogeography, the spatial scale, the availability of parasite occurrence data, and the spatial and temporal match between occurrences and environmental variables. For global disease maps with considerable georeferencing error, environmental variables may include climate data, while for models at medium scale (i.e., continental-country size) with good referencing accuracy, remote sensing data could be a valuable source of environmental information to inform models (Peterson et al., 2011). At more fine scales (e.g., forest, host's body) variables may not exist requiring their development by the modeler. Potential variables at a fine scale (i.e., a forest, cropland, town block) are typically not available, but drones capturing land reflectance are a promising tool to generate fine scale environmental grids with spatial resolutions at a centimeter scale (Figure 11). At a very fine scale (i.e., the host) environmental variables could include features of host's skin, temperature range in the host surface, epithelia type, among others of crucial importance for the parasite to survive and maintain populations. However, biotic interactions (B from the BAM diagram; Figure 1) also should be considered at this microscale. Research on the distribution of parasites at the fine scale is still a neglected field and challenges include our limited understanding of competition, mutualism, and facilitation among parasites in the host and host



**FIGURE 7 | Example of occurrences represented in geographic and environmental terms.** Original reports of disease (green points) can be an overrepresentation of environmental conditions associated with sampling bias. The study area (grid) may require a resampling strategy to obtain only one report per environmental cell to mitigate model overfitting in oversampled areas (red points). In this study area four environmental values are present: gray, pink, blue, and green. A more strict modeling approach (e.g., Qiao et al., 2016) would require only one point per environmental value. Thus, in this example, only one value representing the occupied environment (i.e., gray) should be considered for modeling purposes. Source (Escobar and Peterson, 2013).

immunity and behavior during novel or recurrent infections or co-infections.

The spatial scale must also be considered during variable selection. Climate data are often required at broad scales to



**FIGURE 8 | Ecological niche model comparisons of bat-borne rabies in cattle in Mexico.** (A) The continuous model reflects the sampling effort in terms of numbers of rabies cases available (not shown) and high model overfitting from using all the 19 bioclimatic variables from Worldclim (Adapted from Bárcenas-Reyes et al., 2015). Risk was defined in terms of Maxent suitability index as areas of high (red) and low risk (blue). (B) Binary ecological niche model with reduced occurrences from filtering pseudoreplicated cases. This model was calibrated using principal components for a reduced number of variables. In this case, risk was defined as the presence of suitable conditions for rabies occurrence (red) with an omission error of 5%. From Hijmans et al. (2005), Bárcenas-Reyes et al. (2015), and Escobar (2016).

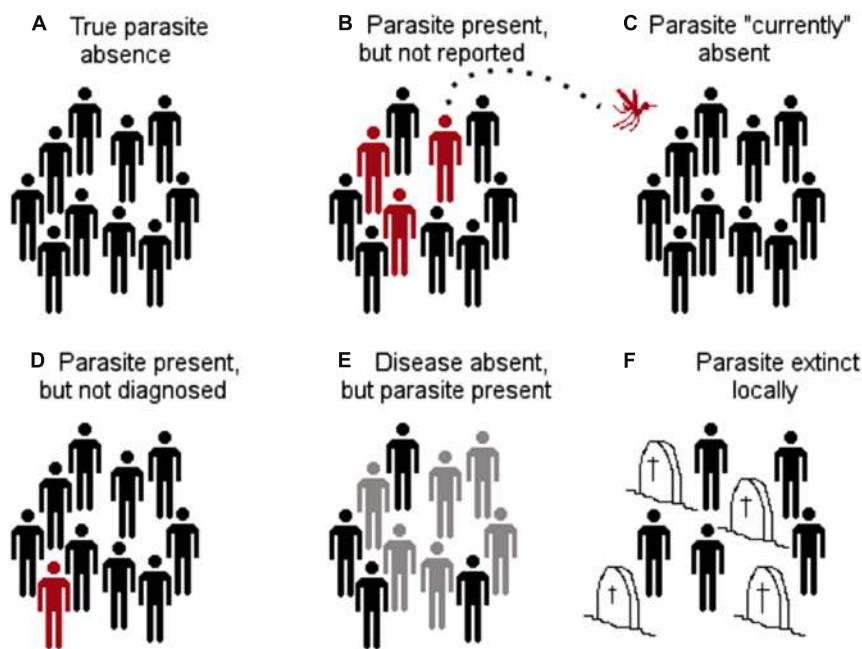
capture environmental signatures of species across space. Disease maps generated at a local scale calibrated using climate data may fail to clearly identify patterns of parasite occurrence, as these data could be highly spatially autocorrelated at narrow extents (Peterson and Nakazawa, 2007). For example, a study of the spatial epidemiology of bat-borne rabies demonstrated that using climate as environmental space could be too coarse to explain the spatial distribution of vampire rabies in small countries (Escobar and Peterson, 2013). However, the *status quo* in ecological niche models in epidemiology is a default utilization of the 19 bioclimatic variables of Worldclim (Hijmans et al., 2005). In fact, Worldclim bioclimatic variables have high correlation (Figure 12), resulting in model overfitting and redundant information in the models (Peterson et al., 2011). The limits between spatial scales are fuzzy; more research in this arena is necessary given that most of the literature in ecological niches is based on one robust set of climatic variables with a decade of use (Hijmans et al., 2005). Thus, attention is needed to identify the spatial scale considered in each study (for more details in variable selection and data sources see Peterson et al., 2011; Peterson, 2014).

## Scale, Scale, Scale

Conceptual, methodological, or philosophical disputes may emerge when discussing the factors defining a parasite's ecological

niche. For instance, it is well known that in ecology debates arise when scientific questions are addressed at different spatial scales (Levin, 1992). In spatial epidemiology, debates dealing with different spatial scales also occur (Astorga et al., 2015b). For instance the environmental variables and assumptions required to describe a parasite's ecological niche are highly dependent on the spatial scale.

The study of parasites may be expressed under diverse spatial scales. For example, at a micro scale rabies virus affects nervous tissues, in fact, rabies is diagnosed using samples from the brainstem and cerebellum in view of the high replication and detection of virus in these organs (Rupprecht et al., 2002). Thus, rabies virus is not distributed at random in the host, instead it occurs in specific tissues. Consequently, the environmental requirements of rabies may be tractable at this tiny scale. At a larger scale, rabies virus has a taxonomic signature, and even with a diversity of potential hosts, the virus can be perpetuated only in mammalian hosts, mainly Carnivora and Chiroptera. Thus, rabies is not distributed at random among all taxa and the physiological features of hosts have an environmental pattern for the virus that may be tractable (Gough and Jorgenson, 1976). Rabies virus can also have a landscape level signature. In bat-borne rabies, the environmental features required by the virus is a combination of soil, vegetation, and moisture requirements defining the ideal habitat to find an infected host (Escobar et al., 2013). Finally, at a more coarse scale, the potential distribution



**FIGURE 9 | Different configurations of parasite absence records.** **(A)** True parasite absence as a consequence of inhabitable conditions. E.g., absence of Tanapox virus in the Antarctic region due to the lack of suitable environmental conditions and absence of reservoirs and vectors (Monroe et al., 2014). **(B)** False absence: parasite present but not detected. E.g., in situations in which countries lack adequate epidemiological surveillance systems (e.g., Dominican Republic reports bat-borne rabies, while Haiti does not report this virus; both countries are located on the same island (Escobar et al., 2015a). **(C)** False absence: parasite currently absent, but environmental conditions are suitable and vectors and reservoirs are present. The parasite is likely to invade if translocated from an infected region to the free/suitable area (e.g., Chikungunya virus was absent in the Americas in early 2013, but was introduced later that year; Van Bortel et al., 2014). Thus, models considering the Americas as unsuitable due to the absence of the virus failed to consider the disease biogeography. **(D)** False absence: the parasite is present but not diagnosed due to low abundance or low sensitivity of diagnostic tests (e.g., Chikungunya virus may be present in some locations but due to the symptoms is referred as dengue infection). **(E)** False absence: the parasite is present in the host population, but there is no evidence of disease (e.g., host population has high immunity to avoid symptoms, however, shedding exists). **(F)** False absence: the parasite was or will be present, but at the time of the survey it was absent (e.g., suitable conditions for the parasite occurrence exist, but the parasite was extinguished due to control measures or the susceptible host population was removed by the parasite). Red: infected with the parasite. Gray: infected with the parasite but without symptoms. Black: not infected. Headstones: removed by the parasite.

of rabies may be inferred at continental scales due to patterns of host occurrence under climate conditions (Kim et al., 2014).

Scale complexities also occur for other pathogens like *Pseudogymnoascus destructans* causing the white-nose syndrome in bats. The fungus affects the hairless skin of hibernacula bats of North America; again not distributed at random on the skin. This parasite has been found mainly in bats, and six bat species appear to be the most susceptible to the disease (Blehert et al., 2009). At a larger scale, conditions inside a cave may offer variations in humidity, temperature, wind speed, and substrate type that differ in the level of suitability for the fungus growth. At a different scale, the disease can also be tracked at landscape level including soil, climate, and landscape features associated with caves where the disease occurs (Flory et al., 2012). At a more continental scale, ecological niche modeling can be employed to understand patterns of distribution and invasion of this parasite based on climate conditions (Escobar et al., 2014).

Cholera for instance has a special affinity to the intestinal epithelia (Harris et al., 2012), thus, the environmental features *in vivo* can also be tracked. At another scale, pH, salinity, and temperature are associated with *V. cholerae* occurrence

(Huq et al., 1984). At a coarser scale, *in situ* environmental requirements of this bacterium helped to predict its distribution in seawater environments at a global scale (Escobar et al., 2015c).

In a model of rabies in livestock in three states of Mexico, climatic data showed high homogeneity among neighbor cells as a result of the interpolative nature of these data (Escobar, 2016). In other words, climate data failed to capture fine scale patterns of environmental variability, suggesting that the study area was too small to calibrate an ecological niche model based on climate solely. When remote sensing (e.g., land surface temperature) and climate data (i.e., precipitation) were considered, the environmental conditions across the study area were more heterogeneous, capturing more information. However, at the same spatial resolution (1 km), remote sensing data summarizing primary productivity provided more landscape details (e.g., identification of water bodies; Figure 10). Thus, one should be aware of incorrect comparisons between models at different scales. In basic ecology, errors in conclusions associated with comparisons at different scales is termed “the Beale fallacy,” and has not been proposed until recently (Escobar et al., 2013; Peterson, 2014). One model calibrated at landscape scale may

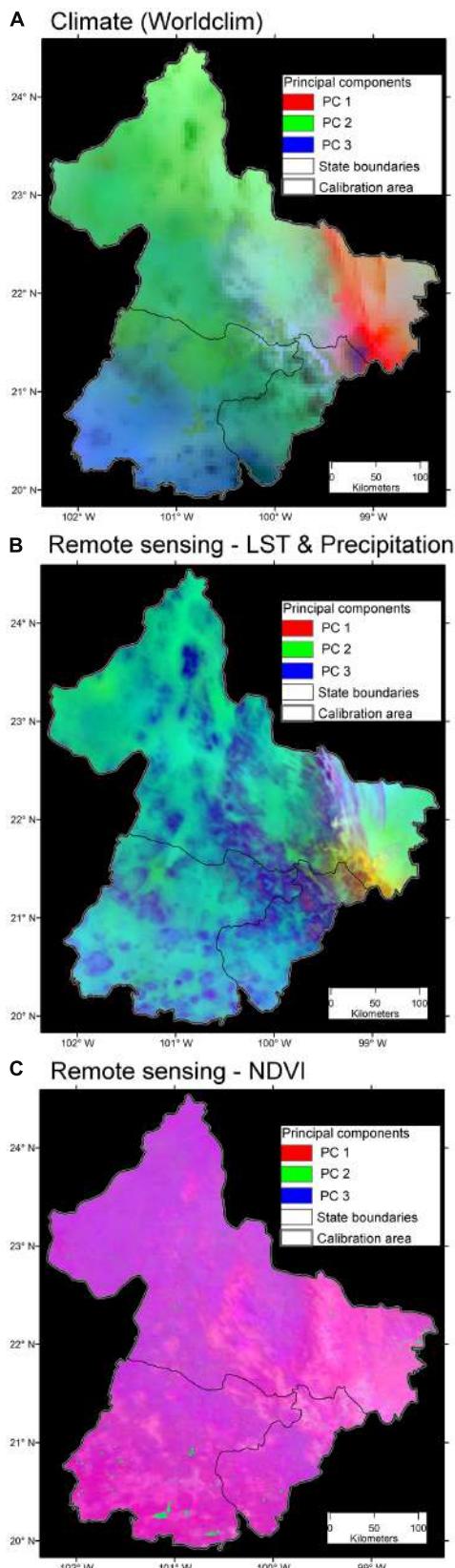


FIGURE 10 | Continued

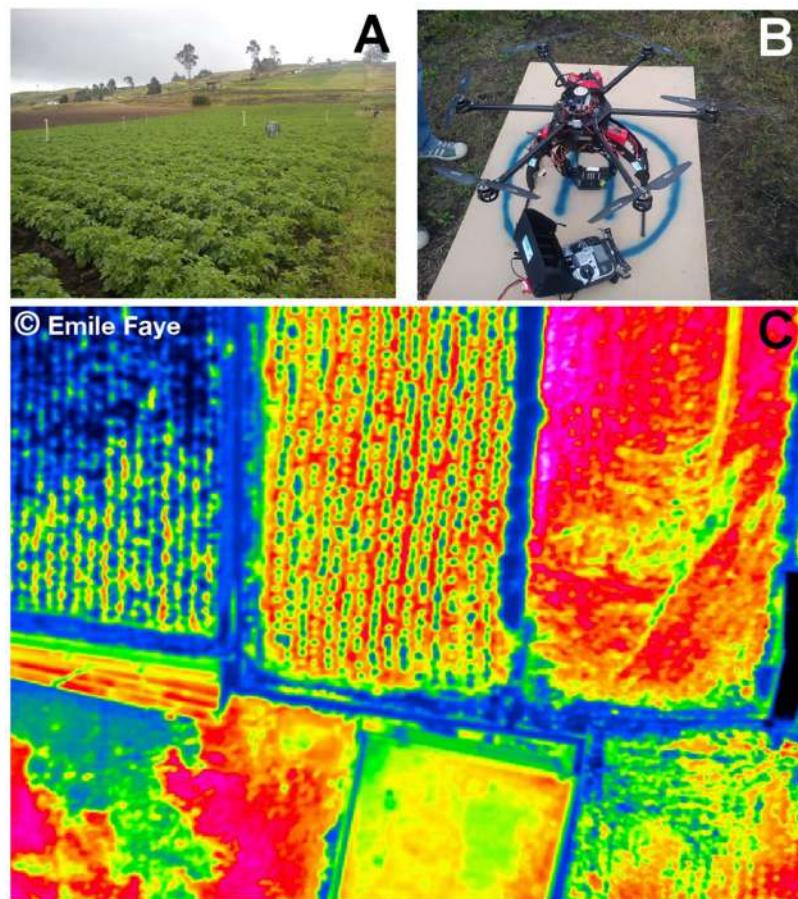
**FIGURE 10 | Variability of three environmental data sets.** **(A)** Bioclimatic variables at ~5 km resolution. Note broad areas with homogeneous environmental values (i.e., high spatial autocorrelation; continuous areas of similar color). The first three components accumulated 99.44% of the overall variance from the 19 variables. Data obtained from Hijmans et al. (2005). **(B)** Land surface temperature (LST) from satellite imagery and precipitation at ~1 km. Note the increased heterogeneity of environmental values across the study area. The first three components accumulated 97.79% of the overall variance from 19 variables. Data obtained from <http://worldgrids.org/> (Hengl et al., 2015). **(C)** Vegetation index, NDVI, from the AVHRR sensor at ~1 km resolution. Note the high detail of information, even capturing areas with water bodies in the south (blue). The first three components accumulated 81.32% of the overall variance from 19 variables, thus, this data set provided less correlation between variables. Data obtained from UMD (2001). Principal components 1 (red), 2 (green) and 3 (blue) from the original variables. Source (Escobar, 2016).

have a different pattern than a model developed at continental scale. Defining *a priori* the study area extent is also a crucial step during the study design of ecological niche models. Different study areas can generate different ecological niche model results (Barve et al., 2011). For example, Peterson and Samy (2016) recently proposed that a detailed selection of the study areas, to map Ebola in Africa, could be more informative, realistic, and robust than a model calibrated in the entire continent (Pigott et al., 2014b). Therefore, the study area extent should be strongly supported by ecological, instead of political, pragmatic, or administrative, reasons.

## PARASITE, VECTOR, RESERVOIR: WHAT TO MODEL IN THE DISEASE SYSTEM?

Ecological niche modeling is a useful tool to understand the ecology of diseases caused by novel or poorly understood parasites (e.g., Ebola and Marburg viruses; Peterson et al., 2004). Researchers may need to identify the ecological factors driving epidemics (Bhatt et al., 2013), propose potential vector species in a disease system of unknown vectors (e.g., candidate vectors for Chagas disease in Brazil; Gurgel-Gonçalves et al., 2012), or to identify the best candidate species to be the reservoir of an emergent parasite (e.g., candidate reservoirs for Tanapox virus in equatorial Africa; Monroe et al., 2014). Thus, ecological niche models can be calibrated using parasites, vectors, or reservoir occurrences. We also could use reports of human or animal disease for modeling as they summarize the entire disease system (in ecological niche modeling termed black-box models *sensu* Peterson, 2007; Figure 13).

Once a parasite's ecological niche has been characterized, this information can be used to anticipate suitable areas for the parasite outside the known range or in the future. This approach was described and patented by Peterson and Vieglais (2001), and today it is applied in spatial epidemiology to identify potential areas for epidemics (Peterson et al., 2004, 2014; Zhu and Peterson, 2014). Using a parasite's niche to identify novel areas of potential spread is based on the assumption that its ecological niche will remain consistent

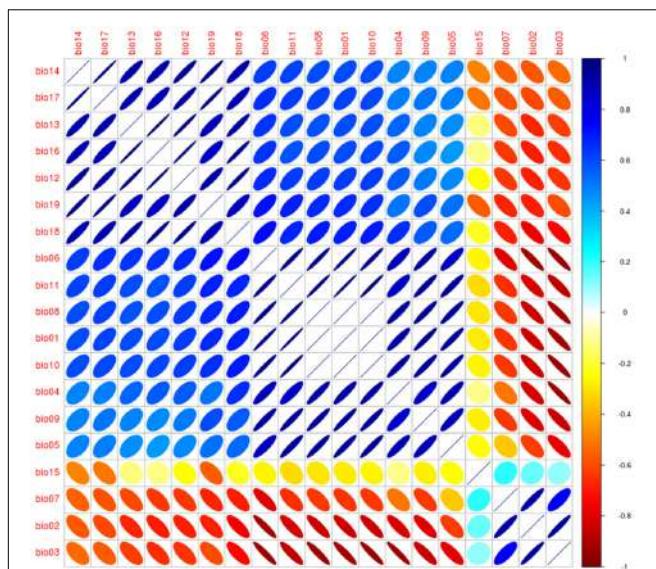


**FIGURE 11 | Remote sensing data of temperature collected using drone technology. (A)** Potato crop field in Ecuador where thermal images were collected. **(B)** RC-drone PIXTIM, DJI Wookong-M autopilot with GPS receptor and 12 min flight autonomy used for data collection. **(C)** Resulting temperature image at 2 cm spatial resolution. Images courtesy: Danilo Yanez and Emile Faye – IRD.

through time. In simple terms, the ecological niche will not evolve. Empirical evidence supports the idea that ecological niche will remain consistent (Peterson et al., 1999; Peterson, 2006a, 2011; Warren et al., 2008). In fact, it is considered that a parasite's niche remains constant even if strains change in virulence. For example, *Toxoplasma gondii* strains may increase in virulence after passage through animals; the niche in abiotic terms, however, remains (Pavlovsky, 1966). Abiotic changes in ecological niches at coarse scales are rare (Soberón and Peterson, 2011; Petitpierre et al., 2012). Ecological niche models usually suggest that diseases such as malaria (*Plasmodium* spp.; Peterson, 2009), leishmaniasis (Peterson and Shaw, 2003), and cholera (Escobar et al., 2015c) would increase their distribution under current climate change trends. How parasites adapt to novel environmental conditions and changes in virulence deserves future research, and experimental studies covering a long generational time of the parasite, more than "human" time, are necessary to understand niche evolution and changes in environmental tolerances of parasites. Such studies may be feasible in some taxa in view of their short generation time (e.g., bacteria).

## RISK MAPS: WHAT IS RISK? HOW DO WE MAP IT?

Diseases could be a complex combination between the parasite's abundance and strain, vector abundance and activity, and host immunity and force of infection. In fact, even when all the actors required in a disease system are present in a site, the disease may be absent, for example due to hosts with high immunity. Furthermore, defining the disease risk spatially is complex. Current literature, however, is crowded with the use of "risk" without a clear definition of the risk. Authors can overuse this concept in the title of manuscripts even when factors associated with risk are not contemplated in the study, making it difficult to identify literature related to mapping disease risk. When mapping disease risk, we suggest that risk should be quantifiable and defined for every study case, specifying if risk is proposed as: (i) the density of previous disease cases; (ii) the suitable areas for the occurrence of parasite, vector, or reservoir (Figure 14); or (iii) the factors associated with the susceptibility and vulnerability of the population of interest (e.g., low immunity, lack of health care, human behavior that facilitates transmission).



**FIGURE 12 | Correlalogram from the Worldclim variables.** Correlation between variables was explored for the 19 Worldclim bioclimatic variables obtained from Hijmans et al. (2005) as in Escobar (2016). Size of ellipsoids denotes high (broad) and low (narrow) variability, direction of ellipsoids represents positive (right) or negative (left) association, colors denote correlation coefficient values to identify high negative (red), low (yellow), and high positive (blue). Analysis was developed using NicheToolBox (Osorio-Olvera, 2016).

In ecological niche models, risk areas can be considered as areas with suitable environmental conditions for the occurrence of the parasite, vectors, and/or reservoirs (Figure 14). Risk delimitation in environmental terms can be complemented with factors acting at local scales. We strongly promote the use of the “disease-transmission risk” or “parasite-exposure risk” concepts, considering that even when the parasite is present in a population, disease *per se* may be absent (i.e., asymptomatic hosts) making the use of the term “disease risk” a strong assumption of exposure, infection, and symptomatology. Novel parasite discovery should not be considered as a true report of a pathogen (e.g., Bai et al., 2011), simply because recently discovered parasites may not be pathogenic. In the same context, a parasite found inside an arthropod does not prove that the arthropod acts as vector transmitting the parasite. But in both cases, risk may be “assumed” in terms of the potential of the parasite or the arthropod to participate in a disease system due to similarities (i.e., taxonomical, morphological, behavioral) with known pathogens or vectors. Noise also appears in reports of risk from emerging diseases. Emergent diseases could be hidden in the past but appear in modern times due to social circumstances, like an increment in surveillance effort, better diagnostic methods, or perhaps the entry of a new susceptible population to the parasite’s niche. Under this scenario, even when the risk of infection was always present in the population, there was no consideration of risk. In summary, the term risk must be defined in each study, as it is context dependent and because its assumptions and features change according to the population of interest. For

public health, for example, risk could be generalized to: “No people, no risk.”

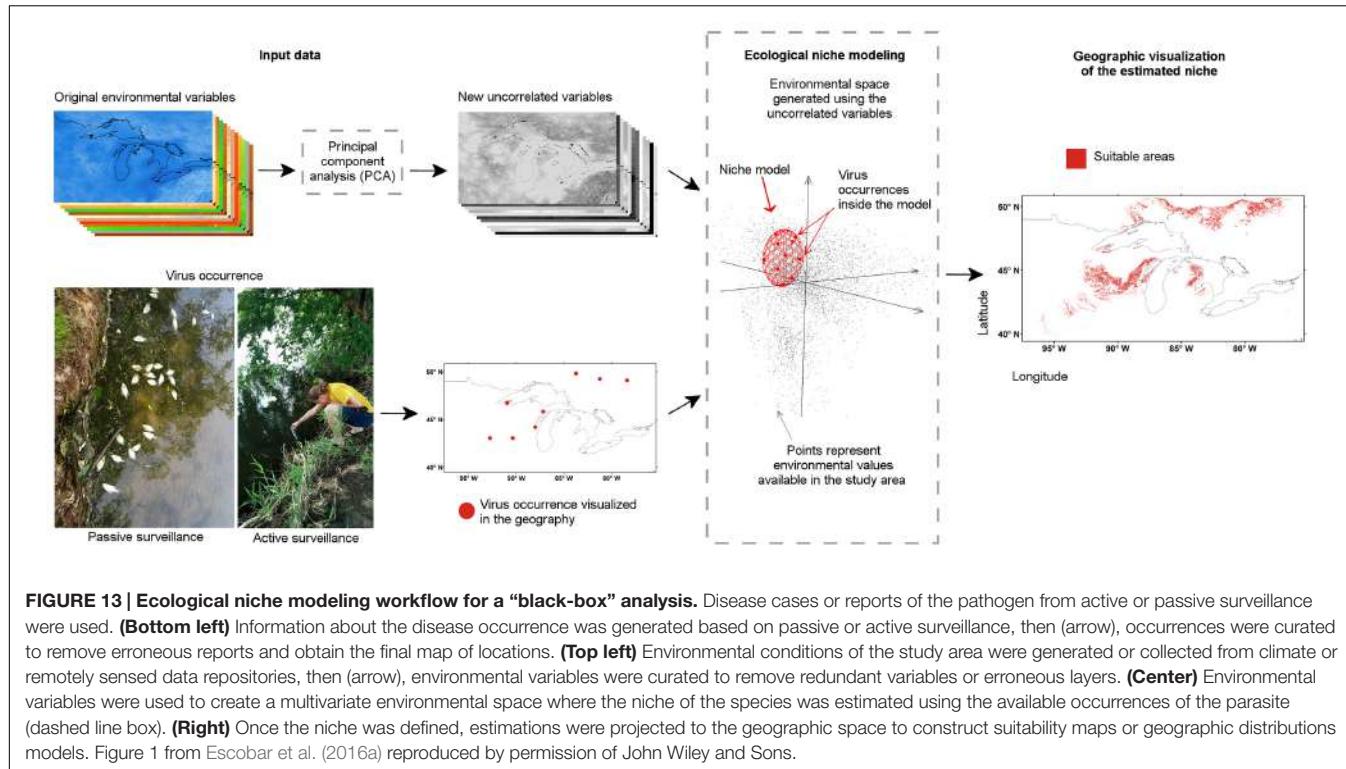
## Adding Risk Factors to Suitability Maps

Suitability maps of parasites could be complemented with information on factors associated with the facilitation of their transmission, including biotic interactions (Figure 14). Recently, Anderson (2016) discussed the positive influences that information resembling parasitehosts interactions could have in ecological niche models, resulting in more detailed, place- and time-dependent, realistic predictions. For example, Samy et al. (2014) modeled the potential occurrence of Mycetoma disease, an infectious skin and bone disease that had been linked to the presence of trees from the *Acacia* genus. Mycetoma models were more accurate when *Acacia* records, a plausible tree reservoir, were added to the model. Also, Astorga et al. (2015a) used an ecological niche model of bat-borne rabies in Chile and, as a post processing step, added a dog-density surface to refine the predictive map and incorporate a risk dimension in terms of potential spillover of rabies from bats to dogs. The result was a risk map with an assertive and more informative forecast of bat rabies spillover events in dogs. Another example of supplementing an ecological niche model with variables of potential risk includes the use of air passenger flow between countries in view of the robustness of air traffic to explain epidemic spread (Brockmann and Helbing, 2013). Passenger flow via air transportation complemented a niche model for the identification of potential areas for chikungunya virus occurrence in countries across the Americas (Escobar et al., 2016b). Finally, a recent study used ecological niche modeling enriched with human density data and nighttime light satellite imagery to successfully estimate areas for human rabies transmission (Escobar et al., 2015b).

## EVALUATION OF DISEASE MAPS

Spatial epidemiologists should acknowledge the effort of ecologists in developing tools, conceptual bases, and variables for ecological niche modeling. The body of literature for these methods has been, however, inspired mainly from the fields of ornithology and biological conservation (Franklin, 2009; Peterson et al., 2011). In this regard, disease mapping is different in limitations and assumptions because parasites occur in complex systems incorporating several species and because incorrect predictions may have negative implications in human and animal health. Models developed to guide public health interventions should have an intense and robust validation process before publication.

Contrasting with the large debate on algorithm performance and software development in ecological niche modeling for modeling biodiversity (Elith et al., 2006; Fitzpatrick et al., 2013; Blonder et al., 2014; Qiao et al., 2015b), little attention has been paid to the critical step of model evaluation (but see Muscarella et al., 2014). Nowadays, the gold standard test for ecological niche model performance uses the area under the curve (AUC) of the receiver operating characteristic (ROC) metric. AUC



ROC *sensu stricto* employs parasite’s presence and absence for model evaluation. As described above, true absences are hard to obtain. To solve this problem, a common practice is to use virtual absences to feed the ROC metric, resulting in questionable evaluations (Lobo et al., 2007; Peterson et al., 2008; Golicher et al., 2012). This may be acceptable when modeling the potential distribution of, for example, an endangered plant or other non-lethal organism. But if the goal of an ecological niche model is to anticipate the potential distribution of Ebola or rabies viruses, models require deep assessment avoiding artificial data. Additionally, modelers use the AUC ROC metric to evaluate model predictions based on the points employed during model calibration, which may be not challenging for the algorithm, considering that points used to create the model are used to validate it, thereby lacking statistical independence (Hurlbert, 1984). The AUC ROC metric also fails to identify those models that over predict the potential areas for disease occurrence (which is not too bad) or those models that under estimate the areas in which suitable conditions for the parasite exist, but the model simply neglects them (which is dangerous for virulent parasites) (see Lobo et al., 2007).

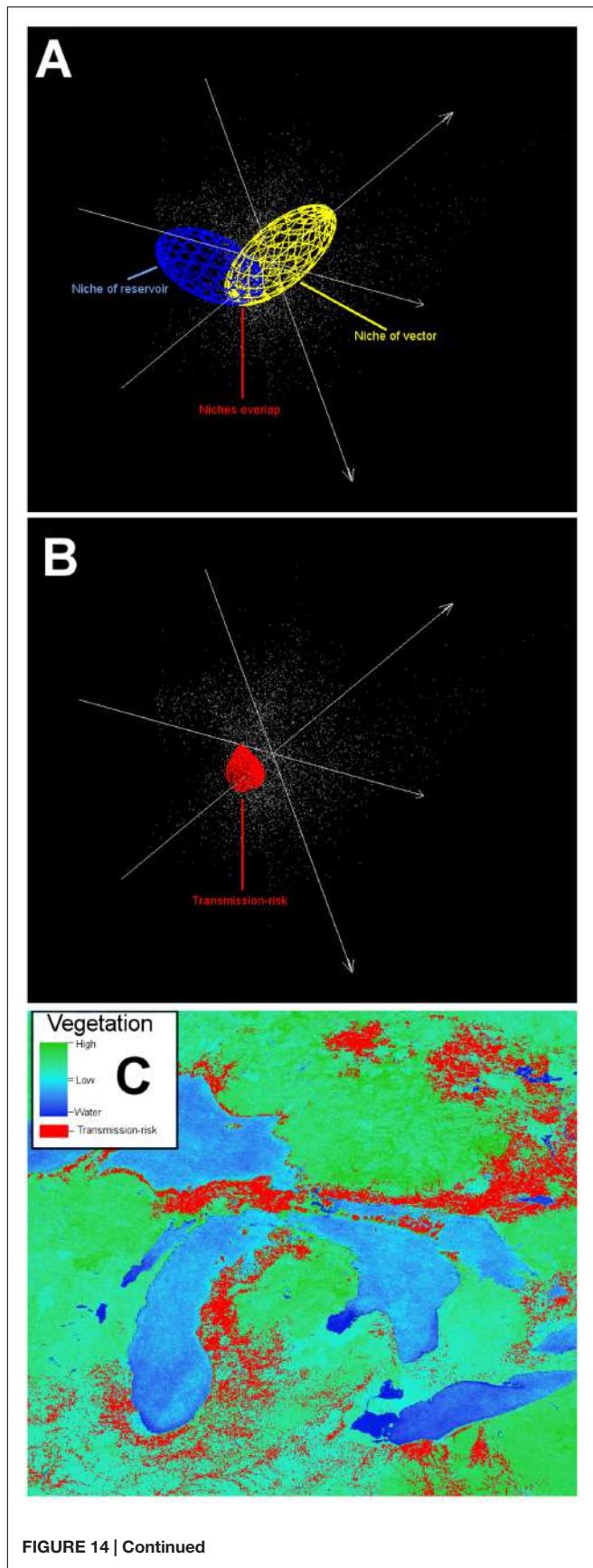
In a recent study, researchers used ecological niche modeling and a detailed set of *Aedes aegypti* and *A. albopictus* occurrences to determine the spatial limits of dengue fever and chikungunya at global scale (Kraemer et al., 2015). The model resulted in important inferences about the potential distribution of these vectors. The model evaluation process was based on a metric requiring presence and absence data, but absence data were not available. To solve this, authors generated their own absence data via random points across the world, reducing the robustness

of the test. The study generated models that mainly reflect the areas with reports (i.e., high model overfit). In this study, Chile, for example, was predicted unsuitable for *A. aegypti*, but the current lack of vector reports in this country is the result of aggressive efforts of authorities for vector eradication and active epidemiological surveillance. *A. aegypti*, however, has been recorded recently, again, in Chile in Arica and Camarones with reports including adult female mosquitoes and larvae (Instituto de Salud Pública, 2016).

Three alternative metrics can be employed to evaluate presence-only ecological niche models including Akaike information criterion (Warren and Seifert, 2011), cumulative binomial probability test (CBP; i.e., identifies if models are predicting occurrences better than by random using an independent set of occurrence data not employed during model calibration) (Peterson et al., 2011), and Partial ROC, a new metric incorporating both ROC AUC and CBP approaches (Peterson et al., 2008; Peterson, 2012; Escobar et al., 2013). Nevertheless, more efforts are needed to test the abilities of these and new metrics to discriminate among different model hypotheses.

## FINAL REMARKS

In this manuscript we describe how ecology, especially the parasite’s ecological niche, is key to understanding the biogeography of disease systems. Ecological niche models of parasites may help us to respond to ecological and distributional questions related to epidemic potential. However, epidemiology



**FIGURE 14 | Example of an ecological niche modeling application including a vector and a reservoir.** The ecological niche of a complex disease system may include a vector and a reservoir. Identifying the potential environmental overlap between the vector and the reservoir can inform potential areas for the effective pathogen occurrence. This scenario is more informative than the black-box approach, because more information is available (i.e., data about the vector and reservoir). **(A)** Ecological niche model of a vector (yellow ellipsoid) and a reservoir (blue ellipsoid). **(B)** Niche overlap between vector and reservoir (red polyhedron) denotes areas of potential pathogen cycle. **(C)** Projection of overlapping niches into the geographic space (red areas). Data obtained from <http://worldgrids.org/> (Hengl et al., 2015).

has largely failed to adopt the conceptual bases that help to correctly design and interpret ecological niche models for disease mapping.

The use of ecological niche modeling methods for disease mapping should be based on a clear understanding of the **BAM** framework and its diversity of plausible configurations (Peterson, 2008; Saupe et al., 2012). Strikingly, click-and-run software dominates the ecological niche modeling practice and users argue that their selection of the method was “because [it] had been validated in peer-review publications,” showing that modelers basically develop predictions without a clear understanding of the process (Joppa et al., 2013). In epidemiology, several ecological niche models have been generated through a “recipe” without a clear justification of the study area extent and inclusion criteria for the data employed (e.g., Abedi-Astaneh et al., 2015; Ali Hanafi-Bojd et al., 2015; Hanafi-Bojd et al., 2015; Gholamrezaei et al., 2016). This practice has been criticized (Anderson, 2014) and more detailed study designs have been encouraged (Peterson, 2014).

Five main questions have been identified for the study design of ecological niche modeling of diseases (Escobar, 2016): (i) Which occurrences to use and why? (e.g., pathogen or reservoir, occurrence inclusion criteria) (ii) Where to calibrate the models and why? (i.e., study area extent) (iii) Which variables should be employed and why? (iv) What algorithms will be explored and why? and (v) How models will be evaluated and why? (e.g., evaluations based on information theory or independent data sets). These questions could help to guide early stages of study designs and could be a helpful tool for readers and reviewers aiming to differentiate between good and incomplete research. Answers to these questions must be based on the research question, the empirical data available, and the natural history of the disease. Here we have explained how the ecological niche of parasites could be studied at different spatial scales, but parameters and variables required need to be generated at very fine scales. The results of the models could be used to map areas of potential transmission risk. Risk is a complex term, but to facilitate its utilization in spatial epidemiology, it should be defined and quantified clearly in each study case. Control and eradication of diseases demand first an understanding of its niche to interrupt the system on any stage or component. Ecological niche modeling shows a promising future in modern epidemiology, but their usefulness lays on the quantitative robustness and biological realism of their products.

## AUTHOR CONTRIBUTIONS

LE conceived and designed the idea of the manuscript, performed the analyses, and wrote the paper; MC co-wrote the paper.

## FUNDING

LEE was funded by the Minnesota Environment and Natural Resources Found and MEC was funded by the National Science Foundation (DEB-1413925), the University of Minnesota's Academic Health Center Seed Grant and the Office of the

Vice President for Research, and the Cooperative State Research Service, US Department of Agriculture, under Projects Nos. MINV-62-044 and 62-051.

## ACKNOWLEDGMENTS

LEE thanks Danilo Yáñez and Emile Faye for images provided. LEE thanks the Student English Program Support and Nick Fountain-Jones for their input in an early stage of this manuscript. Authors thank PA and MT for their kind suggestions, edit, and their time as reviewers.

## REFERENCES

- Abedi-Astaneh, F., Akhavan, A. A., Shirzadi, M. R., Rassi, Y., Yaghoobi-Ershadi, M. R., Hanafi-Bojd, A. A., et al. (2015). Species diversity of sand flies and ecological niche model of *Phlebotomus papatasii* in central Iran. *Acta Trop.* 149, 246–253. doi: 10.1016/j.actatropica.2015.05.030
- Ali Hanafi-Bojd, A., Yaghoobi-Ershadi, M. R., Haghdoost, A. A., Akhavan, A. A., Rassi, Y., Karimi, A., et al. (2015). Modeling the distribution of cutaneous leishmaniasis vectors (Psychodidae: Phlebotominae) in Iran: a potential transmission in disease prone areas. *J. Med. Entomol.* 52, 557–565. doi: 10.1093/jme/tjv058
- Anderson, P. K., Cunningham, A. A., Patel, N. G., Morales, F. J., Epstein, P. R., and Daszak, P. (2004). Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends Ecol. Evol.* 19, 535–544. doi: 10.1016/j.tree.2004.07.021
- Anderson, R. (1991). Populations and infectious diseases: ecology or epidemiology? *J. Anim. Ecol.* 60, 1–50. doi: 10.2307/5443
- Anderson, R. P. (2014). Modeling niches and distributions: It's not just "click, click, click," in *I Simposio de Biogeografía: Actualidad y Retos* (Puebla: Congreso Nacional de Mastozoología), 11–27.
- Anderson, R. P. (2016). When and how should biotic interactions be considered in models of species niches and distributions? *J. Biogeogr.* (in press). doi: 10.1111/jbi.12825
- Astorga, F., Escobar, L. E., Poo-Munoz, D. A., and Medina-Vogel, G. (2015a). Dog ownership, abundance and potential for bat-borne rabies spillover in Chile. *Prev. Vet. Med.* 118, 397–405. doi: 10.1016/j.prevetmed.2015.01.002
- Astorga, F., Poo-Muñoz, D. A., Escobar, L. E., and Medina-Vogel, G. (2015b). In response to: "Increased dog population and potential for bat-borne rabies spillover in Chile in response to 'Dog management, abundance and potential for bat-borne rabies spillover in Chile' by Astorga et al. [Prev. Vet. Med. 118:397–405]." *Prev. Vet. Med.* 120, 248–249. doi: 10.1016/j.prevetmed.2015.04.002
- Auchincloss, A. H., Gebreab, S. Y., Mair, C., and Diez Roux, A. V. (2012). A review of spatial methods in epidemiology, 2000–2010. *Annu. Rev. Public Health* 33, 107–122. doi: 10.1146/annurev-publichealth-031811-124655
- Bai, Y., Kosoy, M., Recuenco, S., Alvarez, D. A., Moran, D., Turmelle, A., et al. (2011). *Bartonella* spp. in Bats. *Guatemala. Emerg. Infect. Dis.* 17, 1269–1272. doi: 10.3201/eid1707.101867
- Bárcenas-Reyes, I., Loza-Rubio, E., Zendejas-Martínez, H., and Luna-Soria, H. (2015). Epidemiological trends in bovine paralytic rabies in central Mexico, 2001–2013. *Rev. Panam. Salud Pública* 38, 396–402.
- Barve, N., Barve, V., Jiménez-Valverde, A., Lira-Noriega, A., Maher, S. P., Peterson, A. T., et al. (2011). The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecol. Modell.* 222, 1810–1819. doi: 10.1016/j.ecolmodel.2011.02.011
- Bhatt, S., Gething, P. W., Brady, O. J., Messina, J. P., Farlow, A. W., Moyes, C. L., et al. (2013). The global distribution and burden of dengue. *Nature* 496, 504–507. doi: 10.1038/nature12060
- Blahert, D. S., Hicks, A. C., Behr, M., Meteyer, C. U., Berlowski-Zier, B. M., Buckles, E. L., et al. (2009). Bat white-nose syndrome: an emerging fungal pathogen? *Science* 323, 227. doi: 10.1126/science.1163874
- Blonder, B., Lamanna, C., Violette, C., and Enquist, B. J. (2014). The n-dimensional hypervolume. *Glob. Ecol. Biogeogr.* 23, 595–609. doi: 10.1111/geb.12146
- Brito-Hoyos, D. M., Sierra, E. B., and Álvarez, V. (2013). Geographic distribution of wild rabies risk and evaluation of the factors associated with its incidence in Colombia, 1982–2010. *Rev. Panam. Salud Pública* 33, 8–14. doi: 10.1590/S1020-49892013000100002
- Brockmann, D., and Helbing, D. (2013). The hidden geometry of complex, network-driven contagion phenomena. *Science* 342, 1337–1342. doi: 10.1126/science.1245200
- Bruno, J. F., Stachowicz, J. J., and Bertness, M. D. (2003). Inclusion of facilitation into ecological theory. *Trends Ecol. Evol.* 18, 119–125. doi: 10.1016/S0169-5347(02)00045-9
- Carpenter, T. E. (2011). The spatial epidemiologic (r)evolution: a look back in time and forward to the future. *Spat. Spatiotemporal. Epidemiol.* 2, 119–124. doi: 10.1016/j.sste.2011.07.002
- Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., et al. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29, 129–151. doi: 10.1111/j.2006.0906-7590.04596.x
- Escobar, L. E. (2016). Ecological niche modeling: five crucial questions. *Pan. Am. J. Public Health* (in press).
- Escobar, L. E., Kurath, G., Escobar-Dodero, J., Craft, M. E., and Phelps, N. B. D. (2016a). Potential distribution of the viral haemorrhagic septicæmia virus in the great lakes region. *J. Fish Dis.* (in press). doi: 10.1111/jfd.12490
- Escobar, L. E., Lira-Noriega, A., Medina-Vogel, G., and Peterson, A. T. (2014). Potential for spread of white-nose fungus (*Pseudogymnoascus destructans*) in the Americas: using maxent and NicheA to assure strict model transference. *Geospat. Health* 11, 221–229. doi: 10.4081/gh.2014.19
- Escobar, L. E., and Peterson, A. T. (2013). Spatial epidemiology of bat-borne rabies in Colombia. *Pan Am. J. Public Health* 34, 135–136.
- Escobar, L. E., Peterson, A. T., Favi, M., Yung, V., and Medina-Vogel, G. (2015a). Bat-borne rabies in Latin America. *Rev. Inst. Med. Trop. São Paulo* 57, 63–72. doi: 10.1590/S0036-46652015000100009
- Escobar, L. E., Peterson, A. T., Favi, M., Yung, V., Pons, D. J., and Medina-Vogel, G. (2013). Ecology and geography of transmission of two bat-borne rabies lineages in Chile. *PLoS Negl. Trop. Dis.* 7:e2577. doi: 10.1371/journal.pntd.0002577
- Escobar, L. E., Peterson, A. T., Papeş, M., Favi, M., Yung, V., Restif, O., et al. (2015b). Ecological approaches in veterinary epidemiology: mapping the risk of bat-borne rabies using vegetation indices and night-time light satellite imagery. *Vet. Res.* 46:92. doi: 10.1186/s13567-015-0235-7
- Escobar, L. E., Qiao, H., and Peterson, A. T. (2016b). Forecasting Chikungunya spread in the Americas via data-driven, empirical approaches. *Parasit. Vectors* 9, 112. doi: 10.1186/s13071-016-1403-y
- Escobar, L. E., Ryan, S. J., Stewart-Ibarra, A. M., Finkelstein, J. L., King, C. A., Qiao, H., et al. (2015c). A global map of suitability for coastal *Vibrio cholerae* under current and future climate conditions. *Acta Trop.* 149, 202–211. doi: 10.1016/j.actatropica.2015.05.028
- ESRI (2015). *ArcGIS Desktop: Release 10.2*. Redlands, CA: Environmental Systems Research Institute.
- Estrada-Peña, A., Ostfeld, R. S., Peterson, A. T., Poulin, R., and de la Fuente, J. (2014). Effects of environmental change on zoonotic disease risk: an ecological primer. *Trends Parasitol.* 30, 205–214. doi: 10.1016/j.pt.2014.02.003
- Fitzpatrick, M. C., Gotelli, N. J., and Ellison, A. M. (2013). MaxEnt versus maxlike: empirical comparisons with ant species distributions. *Ecosphere* 4, 1–15. doi: 10.1890/ES13-00066.1

- Flory, A. R., Kumar, S., Stohlgren, T. J., and Cryan, P. M. (2012). Environmental conditions associated with bat white-nose syndrome mortality in the north-eastern United States. *J. Appl. Ecol.* 49, 680–689. doi: 10.1111/j.1365-2664.2012.02129.x
- Franklin, J. (2009). *Mapping Species Distributions: Spatial Inference and Prediction*. New York, NY: Cambridge University Press.
- Frieden, T. R. (2013). Government's role in protecting health and safety. *N. Engl. J. Med.* 368, 1857–1859. doi: 10.1056/NEJMmp1303819
- Gholamrezaei, M., Mohebali, M., Hanafi-Bojd, A. A., Sedaghat, M. M., and Shirzadi, M. R. (2016). Ecological niche modeling of main reservoir hosts of zoonotic cutaneous leishmaniasis in Iran. *Acta Trop.* 160, 44–52. doi: 10.1016/j.actatropica.2016.04.014
- Godsoe, W. (2010). I can't define the niche but I know it when I see it: a formal link between statistical theory and the ecological niche. *Oikos* 119, 53–60. doi: 10.1111/j.1600-0706.2009.17630.x
- Golicher, D., Ford, A., Cayuela, L., and Newton, A. (2012). Pseudo-absences, pseudo-models and pseudo-niches: pitfalls of model selection based on the area under the curve. *Int. J. Geogr. Informat. Sci.* 8816, 1–15. doi: 10.1080/13658816.2012.719626
- Gough, P. M., and Jorgenson, R. D. (1976). Rabies antibodies in sera of wild birds. *J. Wildl. Dis.* 12, 392–395. doi: 10.7589/0090-3558-12.3.392
- Grinnell, J. (1917). The niche-relationships of the California Thrasher. *Auk* 34, 427–433. doi: 10.2307/4072271
- Gurgel-Gonçalves, R., Galvão, C., Costa, J., and Peterson, A. T. (2012). Geographic distribution of chagas disease vectors in Brazil based on ecological niche modeling. *J. Trop. Med.* 2012, 1–15. doi: 10.1155/2012/705326
- Hanafi-Bojd, A. A., Rassi, Y., Yaghoobi-Ershadi, M. R., Haghdoost, A. A., Akhavan, A. A., Charrahy, Z., et al. (2015). Predicted distribution of visceral leishmaniasis vectors (Diptera: Psychodidae; Phlebotominae) in Iran: a niche model study. *Zoonoses Public Health* 62, 644–654. doi: 10.1111/zph.12202
- Harris, J. B., LaRocque, R. C., Qadri, F., Ryan, E. T., and Calderwood, S. B. (2012). Cholera. *Lancet* 379, 2466–2476. doi: 10.1016/S0140-6736(12)60436-X
- Hatcher, M. J., and Dunn, A. M. (2011). *Parasites in Ecological Communities from Interactions to Ecosystems*. New York, NY: Cambridge University Press, 464.
- Hengl, T., Kilibarda, M., Carvalho-Ribeiro, E. D., and Reuter, H. I. (2015). Worldgrids – A public Repository and a WPS for Global Environmental Layers. *WorldGrids*. Available at: <http://worldgrids.org/doku.php?id=about&rev=1427534899> [accessed January 1, 2015].
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25, 1965–1978. doi: 10.1002/ijoc.1276
- Hudson, P., Dobson, A. P., and Lafferty, K. D. (2006). Is a healthy ecosystem one that is rich in parasites? *Trends Ecol. Evol.* 21, 381–385. doi: 10.1016/j.tree.2006.04.007
- Huq, A., West, P. A., Small, E. B., Huq, M. I., and Colwell, R. R. (1984). Influence of water temperature, salinity, and pH on survival and growth of toxigenic *Vibrio cholerae* serovar O1 associated with live copepods in laboratory microcosms. *Appl. Environ. Microbiol.* 48, 420–424.
- Hurlbert, S. H. (1984). Pseudoreplication and the design of ecological field experiments. *Ecol. Monogr.* 54, 187–211. doi: 10.2307/1942661
- Instituto de Salud Pública (2016). *Ministerio de Salud Confirma Hallazgo de Mosquito Aedes Aegypti en Arica*. Available at: <http://www.ispch.cl/noticia/23228> [accessed June 10, 2016].
- Joppa, L. N., McInerny, G., Harper, R., Salido, L., Takeda, K., O'Hara, K., et al. (2013). Troubling trends in scientific software use. *Science* 340, 814–815. doi: 10.1126/science.1231535
- Kadmon, R., Farber, O., and Danin, A. (2004). Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. *Ecol. Appl.* 14, 401–413. doi: 10.1890/02-5364
- Kim, B. I., Blanton, J. D., Gilbert, A., Castrodale, L., Hueffer, K., Slate, D., et al. (2014). A conceptual model for the impact of climate change on fox rabies in Alaska, 1980–2010. *Zoonoses Public Health* 61, 72–80. doi: 10.1111/zph.12044
- Koch, T., and Denike, K. (2009). Crediting his critics' concerns: remaking John Snow's map of Broad Street cholera, 1854. *Soc. Sci. Med.* 69, 1246–1251. doi: 10.1016/j.socscimed.2009.07.046
- Koelle, K., Rodó, X., Pascual, M., Yunus, M., and Mostafa, G. (2005). Refractory periods and climate forcing in cholera dynamics. *Nature* 436, 696–700. doi: 10.1038/nature03820
- Kraemer, M. U. G., Sinka, M. E., Duda, K. A., Mylne, A. Q. N., Shearer, F. M., Barker, C. M., et al. (2015). The global distribution of the arbovirus vectors *Aedes aegypti* and *Aedes albopictus*. *eLife* 4, 1–18. doi: 10.7554/eLife.08347
- Krebs, C. J. (1972). *Ecology: The Experimental Analysis of Distribution and Abundance*. New York, NY: Longman Higer Education.
- Le Comber, S. C., Rossmo, D. K., Hassan, A. N., Fuller, D. O., and Beier, J. C. (2011). Geographic profiling as a novel spatial tool for targeting infectious disease control. *Int. J. Health Geogr.* 10, 35–43. doi: 10.1186/1476-072X-10-35
- Levin, S. A. (1992). The problem of pattern and scale in ecology: the Robert H. MacArthur Award Lecture. *Ecology* 73, 1943–1967.
- Lobitz, B., Beck, L., Huq, A., Wood, B., Fuchs, G., Faruque, A. S., et al. (2000). Climate and infectious disease: use of remote sensing for detection of *Vibrio cholerae* by indirect measurement. *Proc. Natl. Acad. Sci. U.S.A.* 97, 1438–1443. doi: 10.1073/pnas.97.4.1438
- Lobo, J. M., Jiménez-Valverde, A., and Real, R. (2007). AUC: a misleading measure of the performance of predictive distribution models. *Glob. Ecol. Biogeogr.* 17, 145–151. doi: 10.1111/j.1466-8238.2007.00358.x
- MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, A. A., and Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology* 83, 2248–2255. doi: 10.1890/0012-9658(2002)083[2248:ESORWD]2.0.CO;2
- Maher, S. P., Ellis, C., Gage, K. L., Enscore, R. E., and Peterson, A. T. (2010). Range-wide determinants of plague distribution in North America. *Am. J. Trop. Med. Hygiene* 83, 736–742. doi: 10.4269/ajtmh.2010.10-0042
- Manlove, K. R., Walker, J. G., Craft, M. E., Huyvaert, K. P., Joseph, M. B., Miller, R. S., et al. (2016). "One Health" or three? Publication silos among the one health disciplines. *PLoS Biol.* 14:e1002448. doi: 10.1371/journal.pbio.1002448
- McCarthy, S. A., and Khambaty, F. M. (1994). International dissemination of epidemic *Vibrio cholerae* by cargo ship ballast and other nonpotable waters. *Appl. Environ. Microbiol.* 60, 2597–2601.
- McInerny, G. J., and Etienne, R. S. (2012). Stitch the niche - a practical philosophy and visual schematic for the niche concept. *J. Biogeogr.* 39, 2103–2111. doi: 10.1111/jbi.12032
- Mika, L. A., Pirsich, J. B., and Pollock, N. L. (1959). Studies on mixed infections. V. Effect of stressor compounds in combined infection with *Brucella suis* and *Coxiella burnetii*. *J. Bacteriol.* 77, 189–193.
- Monroe, B. P., Nakazawa, Y. J., Reynolds, M. G., and Carroll, D. S. (2014). Estimating the geographic distribution of human Tanapox and potential reservoirs using ecological niche modeling. *Int. J. Health Geogr.* 13:34. doi: 10.1186/1476-072X-13-34
- Muscarella, R., Galante, P. J., Soley-Guardia, M., Boria, R. A., Kass, J. M., Uriarte, M., et al. (2014). ENMeval: an R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods Ecol. Evol.* 5, 1198–1205. doi: 10.1111/2041-210X.12261
- Nakazawa, Y., Williams, R. A. J., Peterson, A. T., Mead, P. S., Kugeler, K. J., and Petersen, J. M. (2010). Ecological niche modeling of *Francisella tularensis* subspecies and clades in the United States. *Am. J. Trop. Med. Hyg.* 82, 912–918. doi: 10.4269/ajtmh.2010.09-0354
- Osorio-Olvera L. (2016). *NicheToolbox: A Web Tool for Exploratory Data Analysis and Niche Modeling*. Available at: <http://shiny.conabio.gob.mx:3838/nichetoolb2> [accessed January 1, 2016].
- Pavlovsky, E. N. (1966). *Natural Nidality of Transmissible Diseases, with Special Reference to the Landscape Epidemiology of Zooanthropones*, eds F. K. Plous and N. D. Levine (trans.). Urbana: University of Illinois Press.
- Perret, C., Abarca, K., Ovalle, J., Ferrer, P., Godoy, P., Olea, A., et al. (2003). Dengue-1 virus isolation during first dengue fever outbreak on Easter Island, Chile. *Emerg. Infect. Dis.* 9, 1465–1467. doi: 10.3201/eid0911.020788
- Peterson, A. T. (2006a). Conservatism of ecological niche characteristics in North American plant species over the Pleistocene-to-Recent transition. *J. Biogeogr.* 33, 1779–1789. doi: 10.1111/j.1365-2699.2006.01612.x
- Peterson, A. T. (2006b). Ecologic niche modeling and spatial patterns of disease transmission. *Emerg. Infect. Dis.* 12, 1822–1826. doi: 10.3201/eid1212.060373
- Peterson, A. T. (2007). Ecological niche modelling and understanding the geography of disease transmission. *Vet. Ital.* 43, 393–400.
- Peterson, A. T. (2008). Biogeography of diseases: a framework for analysis. *Naturwissenschaften* 95, 483–491. doi: 10.1007/s00114-008-0352-5

- Peterson, A. T. (2009). Shifting suitability for malaria vectors across Africa with warming climates. *BMC Infect. Dis.* 9:59. doi: 10.1186/1471-2334-9-59
- Peterson, A. T. (2011). Ecological niche conservatism: a time-structured review of evidence. *J. Biogeogr.* 38, 817–827. doi: 10.1111/j.1365-2699.2010.02456.x
- Peterson, A. T. (2012). Niche modeling: model evaluation. *Biodivers Inform.* 8:41. doi: 10.17161/bi.v8i1.4300
- Peterson, A. T. (2014). *Mapping Disease Transmission Risk: Enriching Models Using Biology and Ecology*. Baltimore: Johns Hopkins University Press.
- Peterson, A. T., Bauer, J. T., and Mills, J. N. (2004). Ecologic and geographic distribution of Filovirus disease. *Emerg. Infect. Dis.* 10, 40–47. doi: 10.3201/eid1001.030125
- Peterson, A. T., Moses, L. M., and Bausch, D. G. (2014). Mapping transmission risk of Lassa Fever in West Africa: the importance of quality control, sampling bias, and error weighting. *PLoS ONE* 9:e100711. doi: 10.1371/journal.pone.0100711
- Peterson, A. T., and Nakazawa, Y. (2007). Environmental data sets matter in ecological niche modelling: an example with *Solenopsis invicta* and *Solenopsis richteri*. *Glob. Ecol. Biogeogr.* 17, 135–144. doi: 10.1111/j.1466-8238.2007.00347.x
- Peterson, A. T., Papeş, M., and Soberón, J. (2008). Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecol. Model.* 213, 63–72. doi: 10.1016/j.ecolmodel.2007.11.008
- Peterson, A. T., and Samy, A. M. (2016). Geographic potential of disease caused by Ebola and Marburg viruses in Africa. *Acta Trop.* doi: 10.1016/j.actatropica.2016.06.012 [Epub ahead of print].
- Peterson, A. T., and Shaw, J. (2003). Lutzomyia vectors for cutaneous leishmaniasis in southern Brazil: ecological niche models, predicted geographic distributions, and climate change effects. *Int. J. Parasitol.* 33, 919–931. doi: 10.1016/S0020-7519(03)00094-8
- Peterson, A. T., Soberón, J., Pearson, R. G., Anderson, R. P., Martínez-Meyer, E., Nakamura, M., et al. (2011). *Ecological Niches and Geographic Distributions*. Princeton, NJ: Princeton University Press.
- Peterson, A. T., Soberón, J., and Sánchez-Cordero, V. (1999). Conservatism of ecological niches in evolutionary time. *Science* 285, 1265–1267. doi: 10.1126/science.285.5431.1265
- Peterson, A. T., and Vieglais, D. A. (2001). Predicting species invasions using ecological niche modeling: new approaches from bioinformatics attack a pressing problem. *Bioscience* 51, 363–371. doi: 10.1641/0006-3568(2001)051[0363:PSIEN]2.0.CO;2
- Petitpierre, B., Kueffer, C., Broennimann, O., Randin, C., Daehler, C., and Guisan, A. (2012). Climatic niche shifts are rare among terrestrial plant invaders. *Science* 335, 1344–1348. doi: 10.1126/science.1215933
- Phillips, S. J., Anderson, R. P., and Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecol. Model.* 190, 231–259. doi: 10.1016/j.ecolmodel.2005.03.026
- Pigott, D. M., Golding, N., Messina, J. P., Battle, K. E., Duda, K. A., Balard, Y., et al. (2014a). Global database of leishmaniasis occurrence locations, 1960–2012. *Sci. Data* 1:140036. doi: 10.1038/sdata.2014.36
- Pigott, D. M., Golding, N., Mylne, A., Huang, Z., Henry, A. J., Weiss, D. J., et al. (2014b). Mapping the zoonotic niche of Ebola virus disease in Africa. *eLife* 3, 366–378. doi: 10.7554/eLife.04395
- Pliscoff, P., and Fuentes, T. (2011). Modelación de la distribución de especies y ecosistemas en el tiempo y en el espacio: una revisión de las nuevas herramientas y enfoques disponibles. *Rev. Geogr. Norte Gd.* 48, 61–79. doi: 10.4067/S0718-34022011000100005
- Qiao, H., Peterson, A. T., Campbell, L. P., Soberón, J., Ji, L., and Escobar, L. E. (2016). NicheA: creating virtual species and ecological niches in multivariate environmental scenarios. *Ecography* (in press). doi: 10.1111/ecog.01961
- Qiao, H., Soberón, J., Escobar, L. E., Campbell, L., and Peterson, A. T. (2015a). NicheA. Version 3.0.1. Available at: <http://nichea.sourceforge.net/> [accessed 1 December, 2015]
- Qiao, H., Soberón, J., and Peterson, A. T. (2015b). No silver bullets in correlative ecological niche modelling: insights from testing among many potential algorithms for niche estimation. *Methods Ecol. Evol.* 6, 1126–1136. doi: 10.1111/2041-210X.12397
- Revolledo, L., Ferreira, A. J. P., and Mead, G. C. (2006). Prospects in *Salmonella* control: competitive exclusion, probiotics, and enhancement of avian intestinal immunity. *J. Appl. Poult. Res.* 15, 341–351. doi: 10.1093/japr/15.2.341
- Rupprecht, C. E., Hanlon, C. A., and Hemachudha, T. (2002). Rabies re-examined. *Lancet Infect. Dis.* 2, 327–343. doi: 10.1016/S1473-3099(02)00287-6
- Samy, A. M., van de Sande, W. W. J., Fahal, A. H., and Peterson, A. T. (2014). Mapping the potential risk of *Mycetoma* infection in Sudan and South Sudan using ecological niche modeling. *PLoS Negl. Trop. Dis.* 8:e3250. doi: 10.1371/journal.pntd.0003250
- Saupe, E. E., Barve, V., Myers, C. E., Soberón, J., Barve, N., Hensz, C. M., et al. (2012). Variation in niche and distribution model performance: the need for a priori assessment of key causal factors. *Ecol. Model.* 23, 11–22. doi: 10.1016/j.ecolmodel.2012.04.001
- Simarro, P. P., Cecchi, G., Franco, J. R., Paone, M., Diarra, A., Ruiz-Postigo, J. A., et al. (2012). Estimating and mapping the population at risk of sleeping sickness. *PLoS Negl. Trop. Dis.* 6:e1859. doi: 10.1371/journal.pntd.0001859
- Simarro, P. P., Cecchi, G., Franco, J. R., Paone, M., Fèvre, E. M., Diarra, A., et al. (2011). Risk for human African trypanosomiasis, Central Africa, 2000–2009. *Emerg. Infect. Dis.* 17, 2322–2324. doi: 10.3201/eid1712.110921
- Soberón, J. (2007). Grinnellian and Eltonian niches and geographic distributions of species. *Ecol. Lett.* 10, 1115–1123. doi: 10.1111/j.1461-0248.2007.01107.x
- Soberón, J., and Nakamura, M. (2009). Niches and distributional areas: concepts, methods, and assumptions. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19644–19650. doi: 10.1073/pnas.0901637106
- Soberón, J., and Peterson, A. T. (2005). Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Inform.* 2, 1–10.
- Soberón, J., and Peterson, A. T. (2011). Ecological niche shifts and environmental space anisotropy: a cautionary note. *Rev. Mex. Biodiversity* 82, 1348–1355.
- Stevens, K. B., and Pfeiffer, D. U. (2011). Spatial modelling of disease using data- and knowledge-driven approaches. *Spat. Spatiotemporal Epidemiol.* 2, 125–133. doi: 10.1016/j.sste.2011.07.007
- Susser, M., and Susser, E. (1996a). Choosing a future for epidemiology: I. Eras and paradigms. *Am. J. Public Health* 86, 668–673. doi: 10.2105/AJPH.86.5.668
- Susser, M., and Susser, E. (1996b). Choosing a future for epidemiology: II. From black box to Chinese boxes and eco-epidemiology. *Am. J. Public Health* 86, 674–677. doi: 10.2105/AJPH.86.5.674
- Tong, S., Li, Y., Rivailler, P., Conrardy, C., Alvarez, D., Chen, L.-M., et al. (2012). A distinct lineage of influenza A virus from bats. *Proc. Natl. Acad. Sci. U.S.A.* 109, 5–10. doi: 10.1073/pnas.1116200109
- UMD (2001). AVHRR NDVI Data Set. Available at: <http://glcf.umd.edu/data/> [acceded January 1, 2015].
- Van Bortel, W., Dorleans, F., Rosine, J., Blateau, A., Rousseau, D., Matheus, S., et al. (2014). Chikungunya outbreak in the Caribbean region, December 2013 to March 2014, and the significance for Europe. *Eurosurveillance* 19:20759. doi: 10.2807/1560-7917.ES2014.19.13.20759
- Warren, D. L. (2012). In defense of "niche modeling." *Trends Ecol. Evol.* 27, 497–500. doi: 10.1016/j.tree.2012.03.010
- Warren, D. L., Glor, R. E., and Turelli, M. (2008). Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution* 62, 2868–2883. doi: 10.1111/j.1558-5646.2008.00482.x
- Warren, D. L., and Seifert, S. N. (2011). Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecol. Appl.* 21, 335–342. doi: 10.1890/10-1171.1
- World Health Organization [WHO] (2015). *Epidemiology*. Available at: <http://www.who.int/topics/epidemiology/en/> [accessed January 1, 2015].
- Woźniak-Kosek, A., Kosek, J., Mierzewski, J., and Rapiejko, P. (2015). Progress in the diagnosis and control of ebola disease. *Adv. Exp. Med. Biol.* 857, 19–24. doi: 10.1007/5584\_2015\_123
- Zhu, G., and Peterson, A. T. (2014). Potential geographic distribution of the novel avian-origin influenza A (H7N9) virus. *PLoS ONE* 9:e93390. doi: 10.1371/journal.pone.0093390

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Escobar and Craft. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

