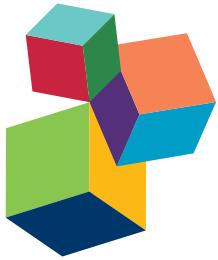


GENOMICS AND EFFECTOMICS OF THE CROP KILLER *XANTHOMONAS*

EDITED BY: Nicolas Denancé, Thomas Lahaye and Laurent D. Noël

PUBLISHED IN: Frontiers in Plant Science and Frontiers in Microbiology





frontiers

Frontiers Copyright Statement

© Copyright 2007-2016 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 978-2-88919-902-0

DOI 10.3389/978-2-88919-902-0

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

GENOMICS AND EFFECTOMICS OF THE CROP KILLER *XANTHOMONAS*

Topic Editors:

Nicolas Denancé, INRA, Institut de Recherche en Horticulture et Semences (IRHS), UMR 1345, F-49071 6 Beaucouzé, France

Thomas Lahaye, University of Tübingen, ZMBP – General Genetics, D-72076 Tübingen, Germany

Laurent D. Noël, INRA, Laboratoire des Interactions Plantes Micro-organismes (LIPM), UMR 441, F-31326 9 Castanet-Tolosan, France; CNRS, Laboratoire des Interactions Plantes Micro-organismes (LIPM), UMR 2594, F-31326 11 Castanet-Tolosan, France



Xanthomonas citri pv. *mangiferaeindicae* causing mango bacterial canker on susceptible *Mangifera indica* leaves.

Image by Lionel Gagnevin, CIRAD, France.

extracellular polysaccharides, plant cell wall-degrading enzymes, adhesins and secreted effectors. In most xanthomonads, type III secretion (T3S) system and secreted effectors (T3Es) are essential to bacterial pathogenicity through the inhibition of plant immunity or the induction of plant susceptibility (*S*) genes, as reported for Transcription Activation-Like (TAL) effectors. Yet, toxins can also be major virulence determinants in some xanthomonads while nonpathogenic *Xanthomonas* species do live in sympatry with plant without any T3S systems nor T3Es.

In a context of ever increasing international commercial exchanges and modifications of the climate, monitoring and regulating pathogens spread is of crucial importance for food security. A deep knowledge of the genomic diversity of *Xanthomonas* spp. is required for scientists to properly identify strains, to help preventing future disease outbreaks and to achieve knowledge-informed sustainable disease resistance in crops.

This Research Topic published in the ‘Plant Biotic Interactions’ section of *Frontiers in Plant Science* and *Frontiers in Microbiology* aims at illustrating several of the recent achievements of the *Xanthomonas* community. We collected twelve manuscripts dealing with comparative genomics or T3E repertoires, including five focusing on TAL effectors which we hope will contribute to advance research on plant pathogenic bacteria.

Citation: Denancé, N., Lahaye, T., Noël, D. L., eds. (2016). Genomics and Effectomics of the Crop Killer *Xanthomonas*. Lausanne: Frontiers Media. doi: 10.3389/978-2-88919-902-0

Table of Contents

- 06 Editorial: Genomics and Effectomics of the Crop Killer Xanthomonas**
Nicolas Denancé, Thomas Lahaye and Laurent D. Noël
- 1. Xanthomonas Genomics and Effectomics**
- 08 Comparative Genomics of Pathogenic and Nonpathogenic Strains of Xanthomonas arboricola Unveil Molecular and Evolutionary Events Linked to Pathoadaptation**
Sophie Cesbron, Martial Briand, Salwa Essakhi, Sophie Gironde, Tristan Boureau, Charles Manceau, Marion Fischer-Le Saux and Marie-Agnès Jacques
- 27 What makes Xanthomonas albilineans unique amongst xanthomonads?**
Isabelle Pieretti, Alexander Pesic, Daniel Petras, Monique Royer, Roderich D. Süßmuth and Stéphane Cociancich
- 34 Genome sequencing reveals a new lineage associated with lablab bean and genetic exchange between Xanthomonas axonopodis pv. phaseoli and Xanthomonas fuscans subsp. fuscans**
Valente Aritua, James Harrison, Melanie Sapp, Robin Buruchara, Julian Smith and David J. Studholme
- 52 Phylogenomics of Xanthomonas field strains infecting pepper and tomato reveals diversity in effector repertoires and identifies determinants of host specificity**
Allison R. Schwartz, Neha Potnis, Sujan Timilsina, Mark Wilson, José Patané, Joaquim Martins Jr., Gerald V. Minsavage, Douglas Dahlbeck, Alina Akhunova, Nalvo Almeida, Gary E. Vallad, Jeri D. Barak, Frank F. White, Sally A. Miller, David Ritchie, Erica Goss, Rebecca S. Bart, João C. Setubal, Jeffrey B. Jones and Brian J. Staskawicz
- 69 Xanthomonas Whole Genome Sequencing: Phylogenetics, Host Specificity and Beyond**
Alice Boulanger and Laurent D. Noël
- 71 Comparative genomics of a cannabis pathogen reveals insight into the evolution of pathogenicity in Xanthomonas**
Jonathan M. Jacobs, Céline Pesce, Pierre Lefevre and Ralf Koebnik
- 84 Interactions of Xanthomonas type-III effector proteins with the plant ubiquitin and ubiquitin-like pathways**
Suayib Üstün and Frederik Börnke
- 90 The Xanthomonas effector XopJ triggers a conditional hypersensitive response upon treatment of *N. benthamiana* leaves with salicylic acid**
Suayib Üstün, Verena Bartetzko and Frederik Börnke

2. Transcription activator like effectors from rice-infecting *Xanthomonads*

- 101 *TAL effectors and activation of predicted host targets distinguish Asian from African strains of the rice pathogen *Xanthomonas oryzae* pv. *oryzicola* while strict conservation suggests universal importance of five TAL effectors***
Katherine E. Wilkins, Nicholas J. Booher, Li Wang and Adam J. Bogdanove
- 116 *QueTAL: a suite of tools to classify and compare TAL effectors functionally and phylogenetically***
Alvaro L. Pérez-Quintero, Léo Lamy, Jonathan L. Gordon, Aline Escalon, Sébastien Cunnac, Boris Surek and Lionel Gagnevin
- 128 *MorTAL Kombat: the story of defense against TAL effectors through loss-of-susceptibility***
Mathilde Hutin, Alvaro L. Pérez-Quintero, Camilo Lopez and Boris Surek
- 144 *Corrigendum: MorTALKombat: the story of defense against TAL effectors through loss-of-susceptibility***
Mathilde Hutin, Alvaro L. Pérez-Quintero, Camilo Lopez and Boris Surek
- 145 *TAL effectors and the executor R genes***
Junli Zhang, Zhongchao Yin and Frank White
- 154 *Decision tools for bacterial blight resistance gene deployment in rice-based agricultural ecosystems***
Gerbert S. Dossa, Adam Sparks, Casiana Vera Cruz and Ricardo Oliva



Editorial: Genomics and Effectomics of the Crop Killer *Xanthomonas*

Nicolas Denancé^{1*}, Thomas Lahaye² and Laurent D. Noël^{3,4}

¹ Institut National de la Recherche Agronomique, Institut de Recherche en Horticulture et Semences (IRHS), UMR 1345, Beaucaire, France, ² University of Tübingen, Centre for Plant Molecular Biology (ZMBP) - General Genetics, Tübingen, Germany, ³ Institut National de la Recherche Agronomique, Laboratoire des Interactions Plantes Micro-organismes (LIPM), UMR 441, Castanet-Tolosan, France, ⁴ Centre National de la Recherche Scientifique, Laboratoire des Interactions Plantes Micro-organismes (LIPM), UMR 2594, Castanet-Tolosan, France

Keywords: immunity, resistance, susceptibility, transcription activator like (TAL) effector, type III effector, Xop

The Editorial on the Research Topic

Genomics and Effectomics of the Crop Killer *Xanthomonas*

Phytopathogenic bacteria of the *Xanthomonas* genus cause severe diseases on hundreds of host plants, including economically important crops, such as bean, cabbage, cassava, citrus, hemp, pepper, rice, sugarcane, tomato, or wheat. Diseases occurring in nature comprise bacterial blight, canker, necrosis, rot, scald, spot, streak, or wilt. *Xanthomonas* spp., are distributed worldwide, and pathogenic and non-pathogenic strains are essentially found in association to plants. Some phytopathogenic strains are emergent or re-emergent and, consequently, dramatically impact agriculture, economy, and food safety. During the last decades, massive efforts were undertaken to decipher *Xanthomonas* biology. So far, more than 100 complete or draft genomes from diverse *Xanthomonas* species have been sequenced (<http://www.xanthomonas.org>), thus providing powerful tools to study genetic determinants triggering pathogenicity and adaptation to plant habitats. *Xanthomonas* spp., employ an arsenal of virulence factors to invade its host, including extracellular polysaccharides, plant cell wall-degrading enzymes, adhesins, and secreted effectors. In most xanthomonads, type III secretion (T3S) system and secreted effectors (T3Es) are essential to bacterial pathogenicity through the inhibition of plant immunity or the induction of plant susceptibility (S) genes, as reported for Transcription Activation-Like (TAL) effectors. Yet, toxins can also be major virulence determinants in some xanthomonads while non-pathogenic *Xanthomonas* species do live in sympatry with plant without any T3S systems nor T3Es.

In a context of ever increasing international commercial exchanges and modifications of the climate, monitoring and regulating pathogens spread is of crucial importance for food security. A deep knowledge of the genomic diversity of *Xanthomonas* spp., is required for scientists to properly identify strains, to help preventing future disease outbreaks and to achieve knowledge-informed sustainable disease resistance in crops.

This research topic published in the “Plant Biotic Interactions” section of Frontiers in Plant Science and Frontiers in Microbiology aims at illustrating several of the recent achievements of the *Xanthomonas* community. We collected twelve manuscripts dealing with comparative genomics or T3E repertoires, including five focusing on TAL effectors which we hope will contribute to advance research on plant pathogenic bacteria.

OPEN ACCESS

Edited and reviewed by:

Felice Cervone,
Sapienza University of Rome, Italy

*Correspondence:

Nicolas Denancé
ndenance@angers.inra.fr

Specialty section:

This article was submitted to
Plant Biotic Interactions,
a section of the journal
Frontiers in Plant Science

Received: 09 December 2015

Accepted: 15 January 2016

Published: 02 February 2016

Citation:

Denancé N, Lahaye T and Noël LD
(2016) Editorial: Genomics and
Effectomics of the Crop Killer
Xanthomonas. *Front. Plant Sci.* 7:71.
doi: 10.3389/fpls.2016.00071

XANTHOMONAS GENOMICS AND EFFECTOMICS

Five papers of this topic deal with genome sequencing and comparative analyses on various *Xanthomonas* species highlighting commonalities and differences in several key players of

bacterial virulence such T3Es. Comparative genomics of pathogenic and non-pathogenic strains of *Xanthomonas arboricola* isolated from walnut tree identified a correlation between absence of a T3S system or a limited repertoire of T3Es with the capacity to live epiphytically and asymptotically on walnut (Cesbron et al.). Pieretti et al. reviewed the latest findings on *Xanthomonas albilineans*, an atypical *Xanthomonas* species pathogenic on sugarcane which has experienced a drastic genome erosion (including the lack of *hrp* T3S and type VI secretion systems) but acquired the capacity of producing albidin, a phytotoxin responsible of the disease symptoms. Sequencing and comparative analysis of 26 strains of *Xanthomonas*-infecting bean species (*Xanthomonas axonopodis* pv. *phaseoli* and *Xanthomonas fuscans* subsp. *fuscans*) resulted in the description of a new genetic lineage for lablab bean isolates that is closely related to the soybean pathogen *Xanthomonas axonopodis* pv. *glycines* and identified a core set of T3Es and genetic determinants that would explain the rise of the *X. fuscans* species (Aritua et al.). Schwartz et al. provide a genomic description of more than 60 *Xanthomonas* strains (species *perforans*, *gardneri*, and *euvescicatoria*) causing bacterial spot on peppers and tomatoes in the US. Their predicted repertoire of T3Es is a key resource for the rational design of disease resistance strategies targeting conserved T3Es among these three species. Jacobs et al. describe the sequencing, assembly, and annotation of two new *Xanthomonas* strains isolated from *Cannabis sativa* in distinct locations. Surprisingly, both strains lack a T3S system and T3Es, although regulatory genes of the T3S system are present. The authors performed a genome-wide analysis of a number of gene families, including those related to virulence, motility and gene regulation, from which they propose a stepwise evolution of pathogenicity.

Additionally, two articles of this topic focused on T3Es. Üstün and Börnke provide an up-to-date mini-review on the role of *Xanthomonas* type III effectors in perturbing host ubiquitin and ubiquitin-like pathways during plant colonization by the pathogen and discuss examples of positive and negative relationships between bacterial effectors and plant machinery. Üstün et al. describe the hypersensitive response observed in *N. benthamiana*-expressing the T3E XopJ upon exogenous application of the plant hormone salicylic acid and analyzed the role of several defense-related genes as well as the proteasome subunit RPT6 in this response.

TRANSCRIPTION ACTIVATOR-LIKE EFFECTORS FROM RICE-INFECTING XANTHOMONADS

Five papers of this research topic report on the recent advances in the rice-*Xanthomonas* pathosystem, which has major economic importance worldwide. The outcome of these interactions relies to a large extent on the repertoires of transcriptional TAL effectors that activate the expression of host *S* genes to promote bacterial growth and fitness. A first study by Wilkins et al. investigated the diversity of rice transcriptional

response to ten distinct *Xanthomonas oryzae* pv *oryzicola* strains causing bacterial leaf streak. Combining these RNA sequencing approaches, TAL effector repertoires and *in silico* predictions of TAL effector binding sites in rice genome yielded a number of novel candidate *S* genes which now await further experimental test. The continuous improvement of DNA sequencing technologies has also caused a continuous increase in the number of TAL effector genes. The article by Pérez-Quintero et al. describes a suite of algorithms that provide a tool kit to efficiently compare numerous TAL effector genes and establish phylogenetic relationships.

Two reviews report on plant immune systems that mediate resistance to xanthomonads that contain TAL effectors: the article by Hulin et al. reviews resistance mechanisms that rely on the suppression of TAL effector activity and/or host *S* gene activation. The review by Zhang et al. focusses on TAL effector-dependent transcriptional activation of a structurally and functionally unique class of plant *R* genes. In parallel, a rationalized framework for *R* gene deployment in rice fields against *X. oryzae* pv. *oryzae* (causal agent of bacterial leaf blight) is presented in Dossa et al. to help improve *R* gene durability and limit disease outbreaks.

We hope that this special focus issue on *Xanthomonas* genomics and effectomics will highlight some of the important progress achieved by this community over the last years and serve as an inspiration for further studies exploring the mechanisms underlying plant pathogen interactions.

AUTHOR CONTRIBUTIONS

ND initiated this research topic and drafted the editorial manuscript. ND, TL, and LN handled manuscripts during the reviewing process. TL and LN revised the editorial manuscript. All authors read and approved the final editorial manuscript.

ACKNOWLEDGMENTS

We thank all the authors for their contribution to this research topic, as well as the reviewers who accepted to evaluate the manuscripts published here. LN is part of the LABEX TULIP (ANR-10-LABX-41 and ANR-11-IDEX-0002-02) and is supported by grants from the Agence Nationale de la Recherche (XANTHOMIX ANR-2010-GENM-013-02, XOPAQUE ANR-10-JCJC-1703-01, and CROptAL ANR 14-CE19-0002-01). TL is supported by grants from the DFG (LA1338) and the Two Blades Foundation.

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 Denancé, Lahaye and Noël. 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.



OPEN ACCESS

Edited by:

Thomas Lahaye,
Ludwig-Maximilians-University
Munich, Germany

Reviewed by:

Liliana Maria Cano,
North Carolina State University, USA
Matthew James Moscou,
The Sainsbury Laboratory, UK

***Correspondence:**

Sophie Cesbron
sophie.cesbron@angers.inra.fr

†Present Address:

Salwa Essakhi,
GDEC - Genetics, Diversity
Ecophysiology of Cereals INRA, UMR
1095, Lempdes, France;
Sophie Gironde,
Terrena, Ancenis, France;
Charles Manceau,
ANSES - French Agency for Food,
Environmental and Occupational
Health & Safety, Plant Health
Laboratory, Angers, France

[#]These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Plant Biotic Interactions,
a section of the journal
Frontiers in Plant Science

Received: 29 May 2015

Accepted: 27 November 2015

Published: 22 December 2015

Citation:

Cesbron S, Briand M, Essakhi S,
Gironde S, Boureau T, Manceau C,
Fischer-Le Saux M and Jacques M-A
(2015) Comparative Genomics of
Pathogenic and Nonpathogenic
Strains of *Xanthomonas arboricola*
Unveil Molecular and Evolutionary
Events Linked to Pathoadaptation.
Front. Plant Sci. 6:1126.
doi: 10.3389/fpls.2015.01126

Comparative Genomics of Pathogenic and Nonpathogenic Strains of *Xanthomonas arboricola* Unveil Molecular and Evolutionary Events Linked to Pathoadaptation

Sophie Cesbron^{1*}, Martial Briand¹, Salwa Essakhi^{1†}, Sophie Gironde^{1†}, Tristan Boureau², Charles Manceau^{1†}, Marion Fischer-Le Saux^{1‡} and Marie-Agnès Jacques^{1†}

¹ INRA, UMR 1345 Institut de Recherche en Horticulture et Semences, Beaucozé, France, ² Université d'Angers, UMR 1345 Institut de Recherche en Horticulture et Semences, Angers, France

The bacterial species *Xanthomonas arboricola* contains plant pathogenic and nonpathogenic strains. It includes the pathogen *X. arboricola* pv. *juglandis*, causing the bacterial blight of *Juglans regia*. The emergence of a new bacterial disease of *J. regia* in France called vertical oozing canker (VOC) was previously described and the causal agent was identified as a distinct genetic lineage within the pathovar *juglandis*. Symptoms on walnut leaves and fruits are similar to those of a bacterial blight but VOC includes also cankers on trunk and branches. In this work, we used comparative genomics and physiological tests to detect differences between four *X. arboricola* strains isolated from walnut tree: strain CFBP 2528 causing walnut blight (WB), strain CFBP 7179 causing VOC and two nonpathogenic strains, CFBP 7634 and CFBP 7651, isolated from healthy walnut buds. Whole genome sequence comparisons revealed that pathogenic strains possess a larger and wider range of mobile genetic elements than nonpathogenic strains. One pathogenic strain, CFBP 7179, possessed a specific integrative and conjugative element (ICE) of 95 kb encoding genes involved in copper resistance, transport and regulation. The type three effector repertoire was larger in pathogenic strains than in nonpathogenic strains. Moreover, CFBP 7634 strain lacked the type three secretion system encoding genes. The flagellar system appeared incomplete and nonfunctional in the pathogenic strain CFBP 2528. Differential sets of chemoreceptor and different repertoires of genes coding adhesins were identified between pathogenic and nonpathogenic strains. Besides these differences, some strain-specific differences were also observed. Altogether, this study provides valuable insights to highlight the mechanisms involved in ecology, environment perception, plant adhesion and interaction, leading to the emergence of new strains in a dynamic environment.

Keywords: *Juglans regia*, vertical oozing canker, bacterial blight, ICE, copper resistance

INTRODUCTION

Xanthomonads are bacteria associated to plants and are commonly pathogens of plants (Vauterin et al., 2000). These bacteria can infect a wide host range and cause diseases on more than 124 monocot species and 268 dicot species including cereals, solanaceous and brassicaceous plants, stone and nut fruit trees (Hayward, 1993; Vauterin et al., 2000). Symptoms and plant parts affected are diverse, however each strain is characterized by a narrow host range. This has led to the definition of the pathovar concept. A pathovar is a group of strains responsible for the same disease on the same host range (Dye et al., 1980).

X. arboricola comprises pathogenic strains distributed in different pathovars (Fischer-Le Saux et al., 2015). The most economically important pathovars in *X. arboricola* are pathovars *pruni*, *corylina*, and *juglandis*, which affect stone and nut fruit trees. *X. arboricola* pv *juglandis* is the causal agent of walnut blight (WB), a serious disease of Persian (English) walnut. It causes necrosis on leaves, catkins, twigs, and fruits, and can induce important crop losses. A few years ago, a new genetic lineage was identified within *X. arboricola* pv *juglandis* as the causal agent of a new disease called vertical oozing canker (VOC) (Hajri et al., 2010). Nonpathogenic *X. arboricola* strains were also isolated from walnut tree during surveys of French orchards. These strains are unable to cause any disease on walnut tree and other plant species (Essakhi et al., 2015). Such xanthomonads, nonpathogenic strains on their host of isolation, have already been isolated from a range of different plants (Vauterin et al., 1996; Vandroemme et al., 2013a; Triplett et al., 2015). Within *X. arboricola*, nonpathogenic strains from *Juglans regia* and from *Fragaria × ananassa* are phylogenetically diverse and do not cluster according to their host of isolation contrary to pathogenic strains from pathovars *pruni*, *corylina*, and *juglandis* (Vandroemme et al., 2013a; Essakhi et al., 2015; Fischer-Le Saux et al., 2015).

Type three effectors (T3Es) secreted in host plant cells via the type three secretion system (T3SS) play a basic role in pathogenicity and host specificity of xanthomonads (Hajri et al., 2009). It was previously shown that strains causing WB and VOC diseases differ by their T3E repertoires, which is composed of 17 T3Es (Hajri et al., 2012). The strains causing VOC could be differentiated from other strains within the pathovar *juglandis* by the absence of *xopAH* and the presence of *xopB* and *xopAI*, the latter being specific to VOC strains within *X. arboricola*. In contrast, some genetic lineages of nonpathogenic strains are devoided of *hrp/hrc* genes encoding the T3SS and possess three T3Es at the most (Vandroemme et al., 2013a; Essakhi et al., 2015). Other nonpathogenic strains possess T3SS genes and seven T3Es genes (*xopR*, *xvrBs2*, *avrXccA1*, *xopA*, *xopF1*, *hrpW*, *hpaA*) among the 18 analyzed T3Es. These results indicate that *X. arboricola* is a model of choice to study the evolutionary events that lead to the emergence of epidemic populations and to decipher the molecular determinants of virulence. Comparative genomic analysis among *Xanthomonas* are useful to identify the distinct gene contents related to virulence, to reveal new features and to explain the differing pathogenic processes (Ryan et al., 2011).

In this report, we present genomic comparisons of four *X. arboricola* strains isolated from walnut tree that are representative of the bacterial diversity encountered on *J. regia* and that were previously analyzed by MLSA, MLVA, and T3Es repertoire (Hajri et al., 2012; Essakhi et al., 2015; Fischer-Le Saux et al., 2015). Strain CFBP 2528 (the type strain of the species), which causes WB, strain CFBP 7179, which causes VOC, both included in the pathovar *juglandis* and two nonpathogenic strains, CFBP 7651 and CFBP 7634, isolated from healthy walnut buds and representing two genetic lineages of nonpathogenic strains with and without *hrp/hrc* genes coding the T3SS respectively were chosen (Essakhi et al., 2015). The aim of this work is to identify differences between pathogenic and nonpathogenic strains, in order to unveil mechanisms of emergence of pathogenic strains. Based on genomic results, phenotypic tests were conducted in an attempt to link genomic content to phenotypic features.

MATERIALS AND METHODS

Bacterial Strains

Bacterial strains used in this study are listed in Table S1. Strains of *X. arboricola* were obtained from the International Center for Microbial Resources, French Collection for Plant-associated Bacteria, (CIRM-CFBP), INRA, Angers, France (<http://www.angers.inra.fr/cfbp/>) or isolated from buds of healthy walnuts in the two main walnut-growing areas in France (Rhône-Alpes region in the southeast and Périgord in the southwest). Bacterial strains were routinely grown at 27°C on TSA medium (3 g/l trypton soya broth; 10 g/l agar) for 24–48 h.

Genomic DNA Isolation, Sequencing, and Annotation

Genomic DNAs from the strains CFBP 2528, CFBP 7179, CFBP 7634, and CFBP 7651 were isolated and purified using the Qiagen's genome DNA isolation kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The Genomic DNA quality and quantity were assessed on an agarose gel and using a NanoDrop ND-1000 spectrophotometer (the NanoDrop Technologies, Wilmington, DE). Libraries with an average insert size of 250 bp and 3 kb (mate-pair libraries) respectively were sequenced using the Illumina HiSeq 2000 platform (GATC Biotech, Germany). Paired-end reads were assembled in contigs using SOAPdenovo 1.05 (Li et al., 2010) and Velvet 1.2.02 (Zerbino and Birney, 2008). Then contigs were scaffolded with LYNX (Gouzy, unpublished data) using mate-pair read information. Annotation was performed using EuGene-PP using similarities with known protein sequences (Sallet et al., 2014). A probably non-exhaustive list of known T3Es that were previously identified in various pathogenic bacteria genus (*Xanthomonas*, *Pseudomonas*, *Ralstonia*, *Erwinia*, *Escherichia*, *Salmonella*) was used to screen for homologs of these effectors in the four *X. arboricola* genomes using tBLASTN and BLASTP. Sequences displaying high sequence similarity (observed with % of length and % identity) with any previously described T3E were searched. We also searched the presence or absence of T3E genes screened

by Hajri et al. (2012). The two models “T4SEpre_bpBAac” and “T4SEpre_psAac” of the T4SEpre package (Wang et al., 2014) were used to predict type four effectors (T4E) from the four *X. arboricola* genomes. Type six secretion system (T6SS) genes from *Xanthomonas campestris* pv *vesicatoria* 85-10, *Xanthomonas fuscans* subsp. *fuscans* 4834R, *Xanthomonas oryzae* pv *oryzae* PXO99A, were used for BLASTN against the four *X. arboricola* genomes.

Genome Accession Numbers

The *X. arboricola* genome sequences of strains CFBP 2528, CFBP 7179, CFBP 7634, and CFBP 7651 have been deposited at DDBJ/EMBL/GenBank under accession no. JZEF00000000, JZEG00000000, JZEH00000000, and JZEI00000000, respectively.

Genomic Comparisons

Identification of orthologous groups between genomes was achieved by orthoMCL V2.0.9 analyses on predicted full-length protein (Li et al., 2003). OrthoMCL clustering analyses were performed using the following parameters: *P*-value Cut-off = 1×10^{-5} ; Percent Match Cut-off = 80; MCL Inflation = 1.5; Maximum Weight = 316. We modified OrthoMCL analysis by using $-F'm\ S'$ option during the BLASTP pre-process. From results were defined unique CDSs, corresponding to CDSs present only in one copy in one genome, and groups of orthologs that corresponded to CDSs present in one copy in at least two genomes. The main part of comparative analyses of genomes and figures were deduced from their distribution. Furthermore, genomes contained CDSs that were present at least in two copies (paralogs) in one or more genomes. Groups of homologs referred to groups of orthologs having paralogs. Venn diagram were obtained using the R package `<< VennDiagram >>` 1.6.5. Chromosomal rearrangements were explored using a script adapted from the R package `<< genoPlotR >>` 0.8.2 (Guy et al., 2010). A circular representation of the orthoMCL analysis was generated with the CGView tool (Grant and Stothard, 2008).

Phylogeny

Average Nucleotide Identity (ANI) analysis was performed as in Scortichini et al. (2013). The Composition Vector Tree (CVTree) tool (Xu and Hao, 2009) was used to build a phylogenetic tree with the four genomes sequenced in this study and eight *X. arboricola* genomes available in public databases (four *X. arboricola* pv *pruni* strains: MAFF 301427, MAFF 301420, MAFF 311562, Xap 33; one *X. arboricola* pc *corylina* strain: NCCB 100457; one *X. arboricola* pv *juglandis* strain: NCPPB 1447; two *X. arboricola* pv *celebensis*: NCPPB 1630, NCPPB 1832). The *X. campestris* pv *campestris* ATCC 33913 strain was used as an outgroup.

Prophages Detection

PhiSpy algorithm was used to find prophages sequences on the four genomes (Akhter et al., 2012).

Copper Resistance

Bacterial suspensions were standardized to 1×10^8 CFU/ml then spotted (10 μ l) in triplicate on CYE-glycerol medium (casitone

1 g/L; yeast extract 0.35 g/L; glycerol 2 ml/L, agar 12 g/L; pH = 7.2), a low nutrient medium with limited copper ion binding capacity (Zevenhuizen et al., 1979), supplemented with Cu⁺⁺: 0 (control), 4, 8, 16, 32, 64 μ g/ml. Copper was brought under CuSO₄, 5H₂O form. Cultures were incubated at 28°C for 72 h. The minimum inhibitory concentration (MIC) in Cu⁺⁺ that prevented colony growth was recorded. Strains able to grow on 32 μ g/ml or greater were considered copper resistant (Gardan et al., 1993).

Design of PCR Primers for Analysis of Copper Resistance Genes

Primer pairs were designed with Primer3 (Koressaar and Remm, 2007; Untergasser et al., 2012) on 11 genes including copper resistance genes (Table S2) located on a CFBP 7179 strain-specific cluster. PCR assays were performed in 20 μ l volumes containing 62.5 μ M dNTP, 0.125 μ M each primer (Table S2), 4 μ l of GoTaq 5 X buffer, 0.3 U/ μ l of GoTaq polymerase, and 5 μ l of a boiled bacterial suspension (1×10^7 CFU/ml). PCR conditions were 3 min at 94°C; followed by 35 cycles of 30 s at 94°C, 30 s at annealing temperature specific of each primer pair, an elongation time adapted to amplicon size at 72°C; and ended with 10 min at 72°C. PCR amplifications were performed in duplicate for each strain.

Pectinase Assays

Bacterial suspensions were standardized to 3×10^8 CFU/ml then spotted (20 μ l) in triplicate on plates. Pectate lyase and Polygalacturonase activities were determined on agar (12 g/L) plates containing polygalacturonic acid (5 g/L) (Sigma) as substrate in Tris-HCl 0.05 M, pH = 8.6 supplemented with CaCl₂ 0.5 mM or in citrate-NaOH 0.1 M, pH = 5, respectively. After 1 week incubation at 28°C, the plates were flooded with Cetyl TrimethylAmmonium Bromide (CTAB) (Eurobio) 1% overnight. Pectate Lyase and Polygalacturonase activities appeared as translucent halos around colonies. Pectin methyl esterase activity was determined on agar (12 g/L) plates containing pectin from citrus (5 g/L; 85% esterified; Sigma) as substrate in citrate-Na₂HPO₄ 0.1 M, pH = 6.4. Plates were flooded with malic acid 0.1 M (Sigma) during 1 h and then stained with ruthenium red 0.02% (Sigma) overnight. Pectin Methyl Esterase activity appeared as a dark red halo surrounding the colonies. All tests were done twice.

Motility Tests

Strain motility was tested in soft-agar assays as detailed in Darrasse et al. (2013). *Xanthomonad* strains were grown at 28°C up to 12 days in MOKA (yeast extract 4 g/l; casamino acids 8 g/l; KH₂PO₄ 2 g/l; MgSO₄.7H₂O 0.3 g/l) and TSA 10% (tryptone soja 3 g/l) medium containing 2 g/l agar. A drop (10 μ l) of a 1×10^8 cfu/ml suspension was deposited in the middle of the plate and the plates were imaged at 5 days. Two independent experiments with three replicates each were conducted.

Design of PCR Primers for Analysis of Flagellar Cluster

To analyse the flagellar cluster diversity, primers developed by Darrasse et al. (2013) to amplify *fliM*, *fliE*, *fliC*, and *fliG* were used in the same conditions. Primer pairs to amplify *fliQ* and *fliB* were designed on CFBP 7179 genome sequence (Table S2).

RESULTS

General Features of the Genome Sequences

General features of the genomes sequenced are summarized in **Table 1**. The sequencing yielded about 57–71.6 million reads giving approximately 583- to 733-fold theoretical genome coverage. The assemblies had a total length comprised between 4.93 and 5.16 MB with the lowest numbers of scaffolds obtained for the nonpathogenic strains (**Table 1**). The G+C contents of CDS ranged from 35 to 77% with an average varying between 65.92 (CFBP 2528) and 66.04% (CFBP 7634) (**Table 1**). Annotation of the genome sequences revealed between 4141 (CFBP 7634) and 4399 (CFBP 7179) putative protein-coding sequences (CDSs), 1 (CFBP 7634) to 12 (CFBP 2528) pseudogenes, 52 or 53 tRNA, and one rRNA operon. We noticed that an extra 16S rRNA gene was present in CFBP 7179. The two 16S rRNA copies exhibited less than 80% identity. The four genome sequences of *X. arboricola* totalized 5126 ortholog groups and specific CDSs of each strain (**Figures 1, 2**). Of those sequences, 3383 (66% of the CDSs) have been assigned to putative functions based on homology with other known proteins and functional domain analysis. No extrachromosomal plasmid have been detected in any of the four strains.

Phylogenomic Relationships Among Completely Sequenced *Xanthomonas* Strains

We used the CVTree tool to study phylogeny of whole genome sequences from *X. arboricola* strains available in public databases. The tree obtained by this algorithm (**Figure 3**) showed that stone and nut fruit tree pathogens clustered according to pathovar classification and shared the same phylogenetic origin. The nonpathogenic strains were included in a different clade together with *X. arboricola* pv *celebensis*, a pathovar of minor incidence (Fischer-Le Saux et al., 2015). The ANI values were all above 95, ranging from 96.4 to 96.7, except between strains CFBP 2528 and CFBP 7179, for which the value was higher (99.2).

A Differential Repertoire of Insertion Sequences (IS) Elements is Observed Between Pathogenic and Nonpathogenic Strains

In their most basic form, IS elements consist of a single gene coding for a site-specific recombinase (called a transposase) and short terminal inverted repeat sequences that are recognized by the transposase. CDSs corresponding to transposases were found

to be scattered over the different scaffolds. The number of IS elements (Table S3) was strikingly different between pathogenic and nonpathogenic strains, indeed a total of 45 (CFBP 2528), 42 (CFBP 7179), and only four (CFBP 7634 and CFBP 7651) IS elements were found in the chromosomes of these strains.

Most IS elements in our *X. arboricola* strains belonged to the IS3 and IS4 families (Table S3). IS 200-like and IS111A/IS1328/IS1533 were encountered in the four sequenced strains. BLASTP searches led to the identification of IS 200-like in other *Xanthomonas* species such as *X. hortorum*, *X. citri*, *X. gardneri*, and *X. campestris*. IS21 as well as Tn3 transposase were only found in CFBP 7179 and in no other *Xanthomonas* species; these IS were related to *Stenotrophomonas maltophilia* and *Pseudomonas aeruginosa* species by BLASTP searches. Mu-transposase was only found in the two pathogenic strains and showed only 92% identity with protein encountered in *X. campestris*.

TABLE 1 | General features of the four draft genome assemblies.

| General features | CFBP 2528 | CFBP 7179 | CFBP 7634 | CFBP 7651 |
|----------------------------|-----------|----------------|-----------|-----------|
| Size (pb) | 5,089,543 | 5,161,669 | 4,935,785 | 5,032,592 |
| G+C, % | 65.92 | 65.93 | 66.04 | 65.98 |
| N50 of contig size (pb) | 1,295,080 | 933,664 | 3,063,886 | 4,192,770 |
| Number of scaffolds | 8 | 14 | 4 | 6 |
| Average coverage | 607 | 733 | 661 | 583 |
| CDS, total | 4325 | 4399 | 4141 | 4229 |
| tRNA | 53 | 52 | 52 | 53 |
| number of ribosomal operon | 1 | 1 ^a | 1 | 1 |
| Pseudogenes | 12 | 9 | 1 | 3 |

^aPresence of an extra 16S rRNA gene.

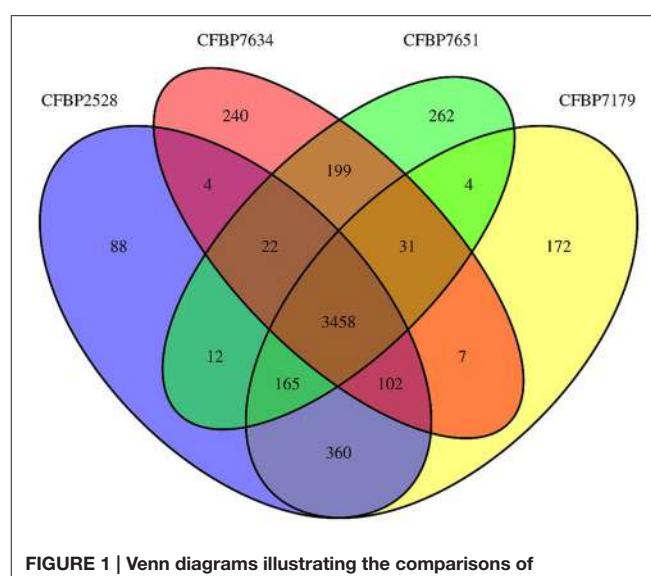
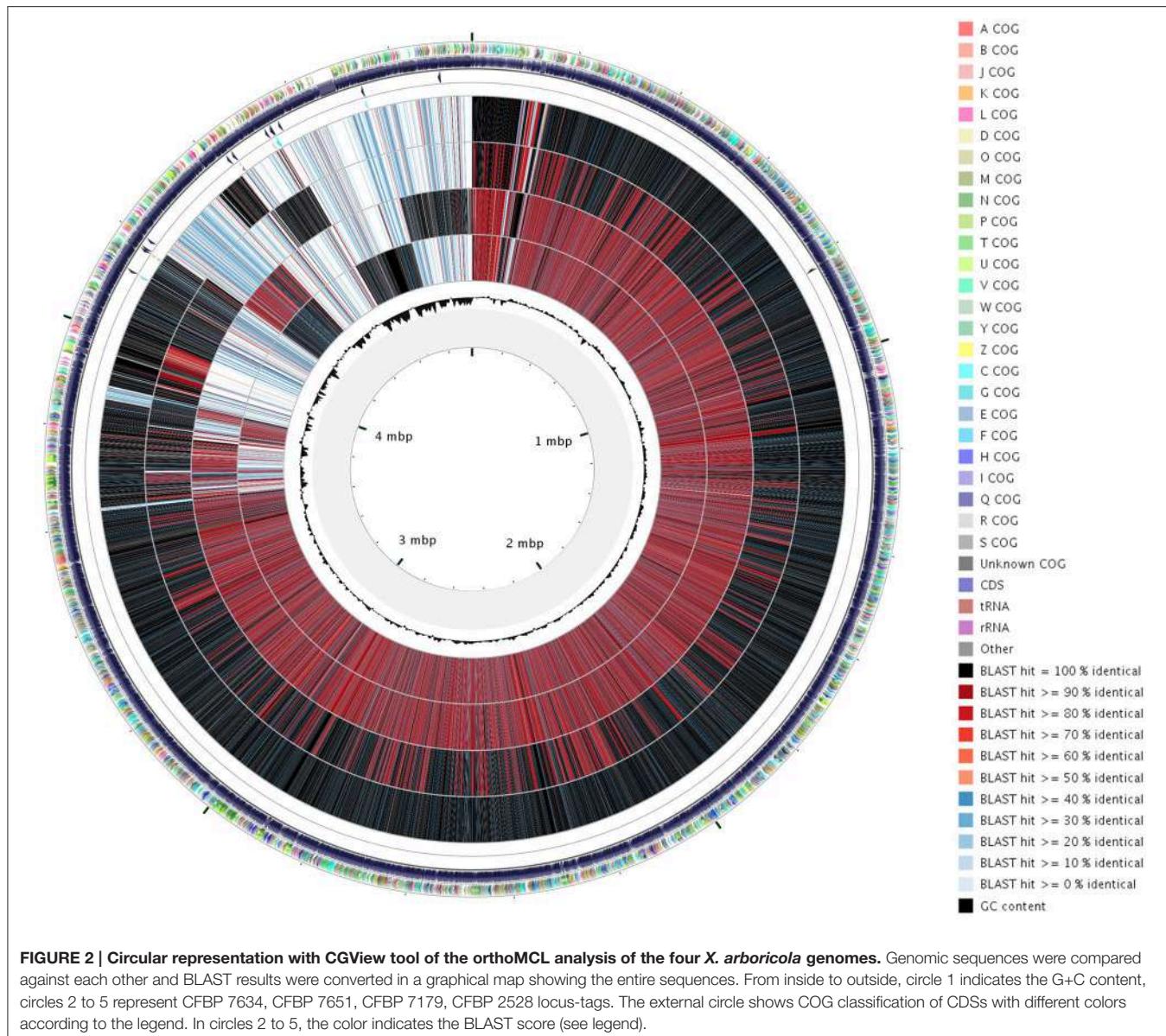


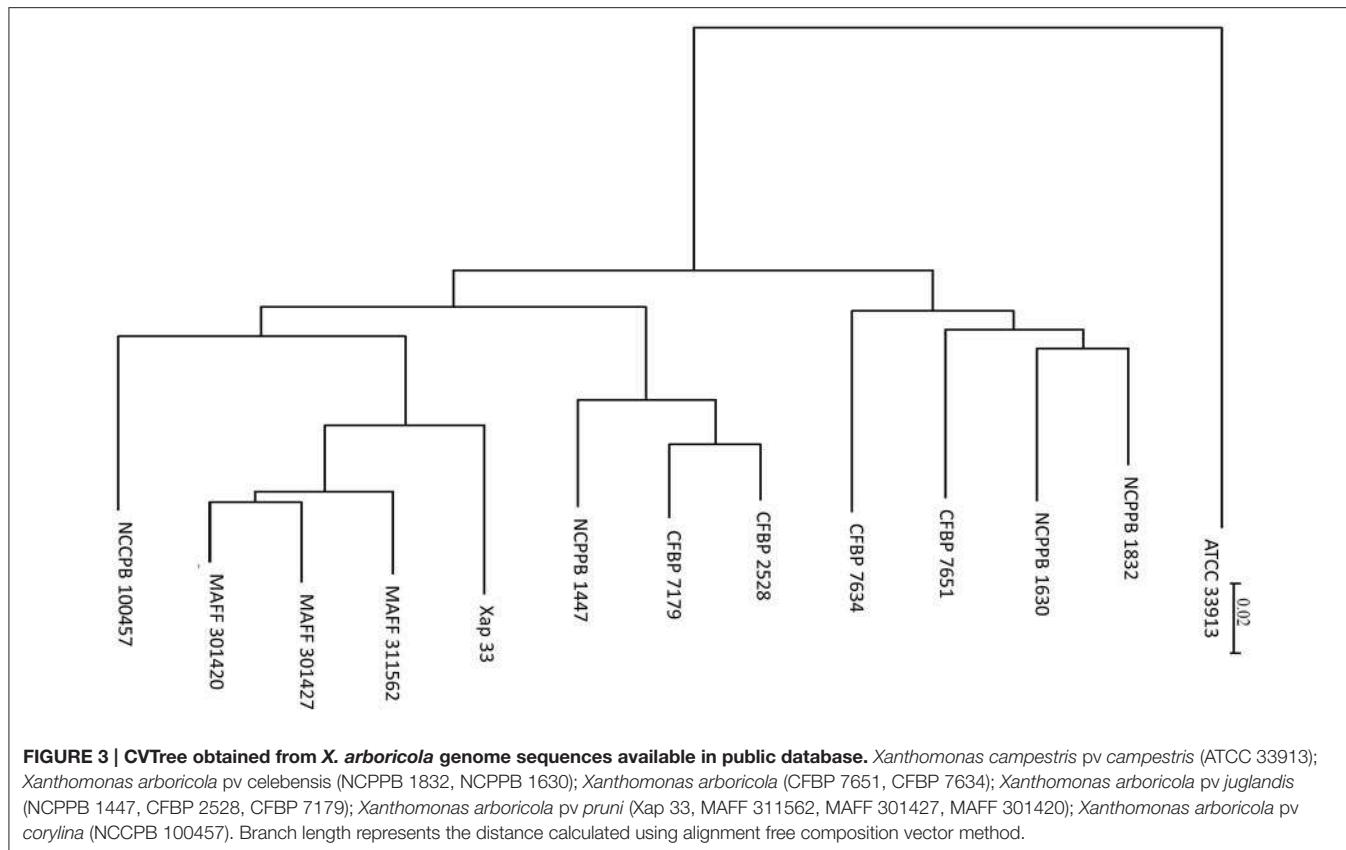
FIGURE 1 | Venn diagrams illustrating the comparisons of *Xanthomonas arboricola* genomes.



An Integrative and Conjugative Element (ICE) Specific of CFBP 7179 Not Yet Described in *Xanthomonas* Triggers Copper Resistance

A genomic island (GI) of 94.8 kb (104 CDSs) was identified specifically in the CFBP 7179 genome sequence (Table S3). This island contained CDSs predicted to be involved in integration and conjugation (integrase/recombinase, pilus formation, excisionase) and was flanked by tRNA^{Gly} attachment sites, one being adjacent to an integrase gene. These features are characteristic of what has been termed the “backbone” of integrative and conjugative elements (ICEs) (Burrus and Waldor, 2004). This ICE found in *X. arboricola* pv *juglandis* CFBP 7179 will be referred here to as Xaj-ICE. The most striking

feature of this GI was the similarity with GI found in bacteria from different genera. Most CDS (102 out of 104) of Xaj-ICE showed high identity (100% identity on 100% of the length) with genes from *P. aeruginosa* strains and *S. maltophilia* strain D457, which belongs to the *Xanthomonadaceae* family. Among the genes located by this ICE, we found CDSs that are predicted to affect the phenotype of pathogens since they are involved in copper resistance (*copA*, *copB*, *copC*, *copD*, *copF*, *copG*, *copK*), in acriflavin resistance and in detoxification (arsenate reductase, mercuric reductase, mercury scavenger protein and mercuric transport protein). Homologs of *copA* and *copB* were also found elsewhere in the four genome sequences and were highly conserved when compared with those of *X. arboricola* pv *pruni* (96% identity; 100% similarity) and other *Xanthomonas* species. These homologs showed the best identity/similarity score by



BLASTN with *copA* and *copB* genes described by Lee (1994) in *X. arboricola* pv *juglandis*.

To determine if the *cop* genes found in *copABCDEFGK* cluster were strain specific and were correlated with copper resistance, we searched by PCR for 11 genes dispersed all over the Xaj-ICE (including copper resistance genes) in the four sequenced strains and 57 additional *X. arboricola* pv *juglandis* strains initially used by Hajri et al. (2010). We tested these strains for copper resistance on CYE medium supplemented with different concentrations in Cu⁺⁺. For most strains, signals at the expected size were generated indicating that these strains should harbor the entire *copABCDEFGK* cluster. Moreover, these strains were shown to be copper resistant (Table 2). However, no signals were obtained for some PCRs in 24 strains, indicating that some genes should be missing. In these strains, no resistance to copper was observed, except for seven strains (CFBP 1022; 12573; 12580; 12582; 12680; 12707; 12714).

Other Mobile Genetic Elements (MGEs) are also Differential Between Pathogenic and Nonpathogenic Strains

Other MGEs such as prophages or integrases were also examined (Table S3). Four prophages were detected in the pathogenic strain CFBP 2528 instead of one or two in the three other strains. No prophage was shared between pathogenic and nonpathogenic strains. A higher number of integrases was found in pathogenic

strains (11–15 integrases per genome) than in the nonpathogenic strains (seven in each genome). The integron described by Gillings et al. (2005) and Barionovi and Scorticini (2008) in pathovars *pruni* and *juglandis* of *X. arboricola*, was localized in the four genomes downstream of the acid dehydratase gene, *ilvD* (Gillings et al., 2005). The integrase gene *intI* should be functional in CFBP 7179 and was degenerated in CFBP 2528. This gene was also degenerated in CFBP 7634 but has retained an integrase domain. *intI* was absent in CFBP 7651 genome. The cassettes of this integron were all different in the four strains and were mostly composed of genes coding hypothetical proteins.

One Hemolysin is Specific of Pathogenic Strains

Hemolysins are toxins secreted via the type I secretion system (T1SS). Homologous CDSs (XARJCFBP 2528_b07940 and XARJCFBP 7179_a04560) coding for a hypothetical protein with a hemolysin BL-binding component (IPR008414 domain) were identified in both pathogenic strains, CFBP 2528 and CFBP 7179. The protein encoded from this CDS showed 98% identity by BLASTP with protein encountered in *X. campestris*. In nonpathogenic strains no CDS was found at the same location. The genes adjacent to the CDS encoding this hypothetical protein were conserved in the four strains. Other CDS encoding proteins linked to hemolysin secretion were only present in pathogenic strains;

TABLE 2 | Copper resistance and PCR results on flagellar genes and on Xaj-ICE.

| Strain numbers | Flagellar genes | | | | | Xaj-ICE | Copper resistance |
|----------------|-------------------------|-------------|-------------|-------------|-------------|---------|-------------------|
| | <i>flgB</i> | <i>flgE</i> | <i>fleQ</i> | <i>fliE</i> | <i>fliM</i> | | |
| 1 | CFBP 878 | + | + | — | — | — | 16 |
| 2 | CFBP 1022 | + | + | + | + | — | 32 |
| 3 | CFBP 2528T = ATCC 49083 | + | + | — | + | — | 8 |
| 4 | CFBP 2564 | + | + | + | + | — | 8 |
| 5 | CFBP 2568 | + | + | + | + | — | 8 |
| 6 | 12572 | + | + | + | + | + | 32 |
| 7 | 12573 | + | + | + | + | — | 32 |
| 8 | 12574 | + | + | + | + | + | 32 |
| 9 | 12575 | + | + | + | + | + | 32 |
| 10 | 12576 | + | + | + | + | + | 64 |
| 11 | 12577 | + | + | + | + | — | 16 |
| 12 | 12578 | + | + | + | + | — | 16 |
| 13 | 12579 | + | + | + | + | + | 64 |
| 14 | 12580 | + | + | + | + | — | 32 |
| 15 | 12581 | + | + | + | + | — | 8 |
| 16 | 12582 | + | + | + | + | — | 32 |
| 17 | 12583 | + | + | + | + | + | 32 |
| 18 | 12584 | + | + | + | + | + | 32 |
| 19 | 12585 | + | + | + | + | + | 32 |
| 20 | 12586 | + | + | + | + | + | 32 |
| 21 | 12587 | + | + | + | + | + | 64 |
| 22 | 12588 | + | + | + | + | + | 32 |
| 23 | 12589 | + | + | + | + | — | 32 |
| 24 | 12590 | + | + | + | + | — | 32 |
| 25 | 12591 | + | + | + | + | + | — |
| 26 | 12592 | + | + | + | + | + | 32 |
| 27 | 12680 | + | + | + | + | — | 32 |
| 28 | 12681 | + | + | + | + | + | 32 |
| 29 | 12707 | + | + | + | + | — | 32 |
| 30 | 12709 | + | + | + | + | + | 32 |
| 31 | 12710 | + | + | + | + | — | 16 |
| 32 | 12711 | + | + | + | + | + | 64 |
| 33 | 12712 | + | + | + | + | + | 64 |
| 34 | 12713 | + | + | + | + | + | 32 |
| 35 | 12714 | + | + | + | + | — | 32 |
| 36 | 12762 | + | + | + | + | + | 32 |
| 37 | CFBP7179 | + | + | + | + | + | 32 |
| 38 | 12764 | + | + | + | + | — | 64 |
| 39 | 12765 | + | + | + | + | + | 32 |
| 40 | 12766 | + | + | + | + | + | 32 |
| 41 | 12767 | + | + | + | + | + | 64 |
| 42 | 12768 | + | + | + | + | + | 32 |
| 43 | 12769 | + | + | + | + | + | 32 |
| 44 | 12770 | + | + | + | + | + | 32 |
| 45 | 12771 | + | + | + | + | + | 32 |
| 46 | 12772 | + | — | + | + | + | 32 |
| 47 | 12773 | + | + | + | + | + | 32 |
| 48 | 12774 | + | + | + | + | + | 32 |
| 49 | 12775 | + | + | + | + | + | 64 |

(Continued)

TABLE 2 | Continued

| Strain numbers | Flagellar genes | | | | | Xaj-ICE | Copper resistance |
|----------------|-----------------|-------------|-------------|-------------|-------------|---------|-------------------|
| | <i>flgB</i> | <i>flgE</i> | <i>fleQ</i> | <i>fliE</i> | <i>fliM</i> | | |
| 50 | 12776 | + | + | + | + | + | 32 |
| 51 | 12777 | + | + | + | + | - | 16 |
| 52 | 12778 | + | + | + | + | + | 32 |
| 53 | 12779 | + | + | + | + | - | 16 |
| 54 | 12780 | + | + | + | + | - | 16 |
| 55 | 12781 | + | + | + | + | + | 64 |
| 56 | 12782 | + | + | + | + | + | 64 |
| 57 | 12783 | + | + | + | + | - | 8 |
| 58 | 12784 | + | + | + | + | - | 16 |
| 59 | 12785 | + | + | + | + | - | 16 |
| 60 | CFBP 7651 | + | + | + | + | - | 16 |
| 61 | CFBP 7634 | + | + | + | + | - | 16 |

they possessed HlyB or HlyD domains (XARJCFBP 2528_a06990 and XARJCFBP 7179_b04000; XARJCFBP 2528_d04670 and XARJCFBP 7179_e04740, respectively). These domains are found in ABC transporter (HlyB) and membrane fusion protein (HlyD) from T1SS (Kanonenberg et al., 2013).

Plant Cell Wall-Degrading Enzymes (PCWDEs) are Active in Nonpathogenic Strains

Orthologs of most CWDEs described in Potnis et al. (2011) and Darrasse et al. (2013) were identified in the four *X. arboricola* genomes except that no orthologs of *xyn30A*, *xynC*, and *cbhA* (encoding 1,4-β cellobiosidase) were found (Table S4).

A differential repertoire of Type 2-secreted degrading enzymes with various activities (peptidases, pectinesterase, pectate lyase, xylosidase...) was identified between pathogenic and nonpathogenic strains (Table S4). On the one hand, homologs of XFF4834R_chr16290 (putative aminopeptidase), XCC0121 (AAM39440, pectinesterase), and of XCC0122 (AAM39441, pectate lyase), were observed in the two nonpathogenic strains. In the two pathogenic strains, only fragments of XFF4834R_chr16290 and XCC0122 were identified and no remnants of XCC0121 were observed. On the other hand, homologs of XFF4834R_chr05470 (coding a putative secreted protease), of XFF4834R_chr25520 (coding a xylosidase), and of XFF4834R_chr23760 (coding a putative pectate lyase), were observed only in pathogenic strains. The putative pectate lyase was observed in pathogenic strains near a peptidase trypsin-like gene. Both were observed in pathogenic strains instead of a pectinesterase gene (GROUPORTHO4194) in nonpathogenic strains. Homologous genes of XFF4834R_chr11410 (encoding for a putative rhamnogalacturonase B) and XFF4834R_chr03290 (encoding a putative glycoside hydrolase) were present in all strains except in CFBP 7179.

We compared the pectinase activities of the four strains using plate assays for pectate lyase, polygalacturonase, and pectin methyl esterase. Only the two nonpathogenic strains showed pectate lyase and pectin methyl esterase activities (Figure 4). No polygalacturonase activity was detected for any strain.

T3SS is Absent in the Nonpathogenic Strain CFBP 7634 and T3Es Repertoire is Reduced in Nonpathogenic Strains

Genomic comparisons of the T3SSs revealed that among conserved genes, the six *hrp* genes (*hrpF*, *hrpW*, *hrpD6*, *hrpB1*, *hrpB4*, and *hrpB7*), the six *hpa* genes (*hpa1*, *hpa2*, *hpa3*, *hpaA*, *hpaB*, *hpaC*) and the 11 *hrc* genes (*hrcC*, *hrcD*, *hrcJ*, *hrcL*, *hrcN*, *hrcQ*, *hrcR*, *hrcS*, *hrcT*, *hrcU*, *hrcV*) were present in genomes of CFBP 2528, CFBP 7179, and CFBP 7651 and were all absent in the genome of CFBP 7634. The genomic comparisons revealed a high synteny for these clusters in the three strains, with the *hrpF* locus followed by the *hrp/hrc* cluster. The sequences flanking the *hrp*-island were the *ltaE* gene at the upstream of the *hrpF* peninsula and *trpG* at the downstream of the *hrp/hrc* cluster (Figure 5).

In the nonpathogenic strain CFBP 7634, the *hrp-hrc* region and the *hrpF* peninsula were absent (Figure 5). In fact instead of the *hrp* cluster a region of 8 kb containing genes coding for ATP-dependent restriction enzymes, such as the *hsdR*, *hsdS*, and *prrC* coding an anticodon nuclease was found between *trpG* and *ltaE*. These genes compose the typeI Restriction-Modification system known in *Escherichia coli* to be involved in phage defense mechanism (Makarova et al., 2013). A BLAST research with alternative T3SSs observed in other bacteria did not lead to the identification of other T3SS (Araki et al., 2006; Diallo et al., 2012). In CFBP 7651, the *hrp/hrc* cluster was followed by a specific region of about 16.7 kb. The first 9 kb are highly identical to regions found in *X. c. pv. raphani*, *X. c. pv. campestris*. This region contained genes encoding a putative xylanase-like, a glycosidase, a

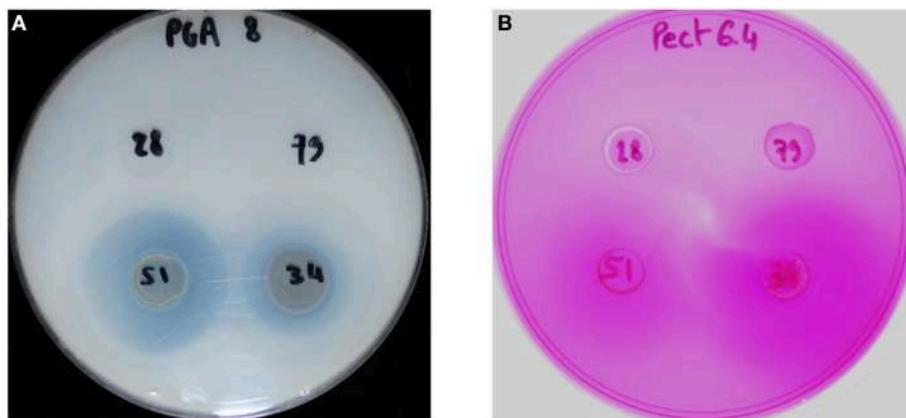


FIGURE 4 | Detection of pectate lyase activity (A) and pectin methyl esterase activity (B) from the strains CFBP 2528 (28), CFBP 7179 (79), CFBP 7651 (51), CFBP 7634 (34). The translucent halo (A) and slightly violet (B) surrounding the colony indicates the activity.

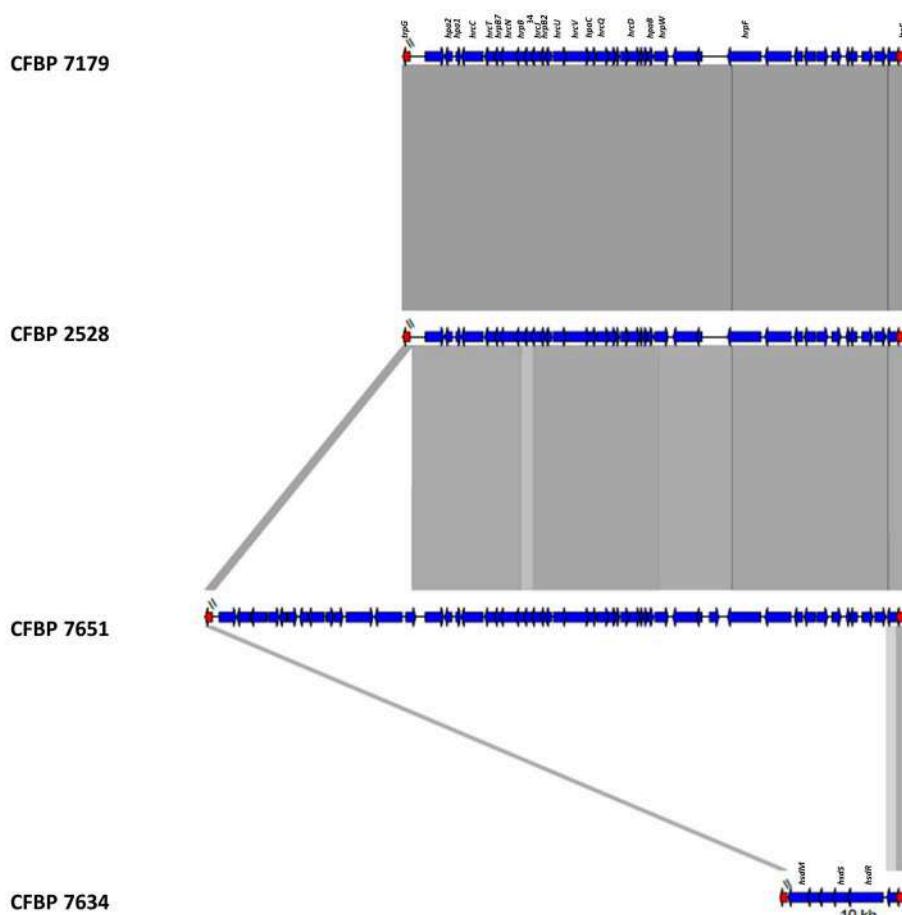


FIGURE 5 | Schematic representation of the T3SS locus of the four *X. arboricola* strains.

methyl-accepting chemotaxis protein (MCP), an oxydoreductase, a transcriptional regulator and a monooxygenase. The last 7.7 kb had BLASTN hits with *Methylobacterium extorquens* and

BLASTX hits with dehydrogenase and epimerase implicated in cell envelope biogenesis, two transcription regulators and a sodium/dicarboxylate symporter.

Examination of the Surrounding Regions of T3Es Provides Clues Relative to their Mechanism of Acquisition

Orthologous sets of 24 and 25 genes were predicted in the two pathogenic strains CFBP2528 and CFBP7179, respectively (Table S5). Among these two sets, 17 T3Es genes were already identified by PCR by Hajri et al. (2012) in *X. arboricola* pv *juglandis* and two (*xopAI* and *xopG*) were not previously identified, probably because of a high diversity in their sequence preventing their amplification by PCR. Two other T3E genes might be present (*xopAA*, *xopAB*), although the percentage of length was low (74 and 66% respectively). The other genes (*awr4*, *sfrJ*, and *xopAR*) were not searched by Hajri et al. (2012) and Essakhi et al. (2015). These three genes were predicted in pathogenic and nonpathogenic strains.

Based on the genome sequences, *xopAI* and *xopB* were identified only in CFBP 7179 and not in CFBP 2528. *xopAI*, was close to an IS4 like in CFBP 7179. *xopB* was close to an integrase and an IS4/5. The presence of an integrase close to *xopB* or an IS4 like close to *xopAI* suggested that these T3Es were probably acquired by lateral gene transfer (LGT) in CFBP 7179. In contrast, *xopAH* was identified in CFBP 2528 and not in CFBP 7179 genome sequences as reported by Essakhi et al. (2015) by PCR. Surrounding regions of *xopAH* were the same between strains. This suggests that *xopAH* was probably acquired by homologous recombination in CFBP 2528.

Other T3E coding genes such as *xopN*, *xopX*, *xopZ*, *xopQ*, *xopK*, *xopV*, *xopL*, *avrXccA2* were scattered in the different scaffolds and were either integrated between genes (*xopN*, *xopX*, *xopZ*) that were shared between the four strains, either found in place of a gene shared by the nonpathogenic strains (*xopL*, *xopV*), or associated to other genes that were not present in the nonpathogenic strains (with transposases for *xopK* or without for *xopQ*).

The Flagellar System is Not Functional in the Type Strain CFBP 2528 and the 22-Amino-Acid Flagellin Epitope is Different in the Pathogenic Strains

Annotation of the flagellar system reveals that a group of nine contiguous genes was lacking in CFBP 2528 compared to CFBP 7179 genome and genomes of nonpathogenic strains. This group of missing genes included *fliS*, a secretion chaperone for the flagellar filament protein *FliC*, and *rpoN*, the sigma factor 54 (σ^{54}) regulating the flagellar system (Figure 6). No swimming motility was observed for CFBP 2528 in a soft agar-assay (Figure 7). To determine if this event could be observed in other *X. arboricola* pv *juglandis* strains consensus primer pairs were used for PCR-amplification of genes dispersed over the flagellar cluster. A collection of *X. arboricola* pv *juglandis* strains (Hajri et al., 2010) was used. Signals at the expected sizes were obtained suggesting a complete flagellar cluster in all strains, excepted in CFBP878, which gave no signal with *fleQ*, *fliE*, and *fliM* primers (Table 2). We also compared the N-terminal *FliC* sequences with the flagellin conserved domain Flg22, which is known as a major pathogen-associated molecular pattern (PAMP), activating host

defense responses (Felix et al., 1999; Navarro et al., 2004; Shi et al., 2015). Nonpathogenic strains possessed the conserved Flg22 epitope whereas CFBP 2528 and CFBP 7179 had a different peptide, with a polymorphism in 7 amino acids (Figure 8). One of these six residues—aspartic acid (D)—has been shown in *X. campestris* pv *campestris* to be critical for elicitation activity in Arabidopsis (Sun et al., 2006). Its replacement by valine (V) in *X. campestris* pv *campestris* eliciting strain suppress the elicitation activity. In our *X. arboricola* pv *juglandis* (pathogenic) strains, the D residue is replaced by a V one (Figure 8).

The Four Strains Share a Different Repertoire of Genes Encoding a Type IV Secretion System (T4SS) and Type IV Effectors (T4Es)

In *X. arboricola*, the T4SS encoding genes are approximately organized as in *X. citri* subsp. *pv citri* (Jacob et al., 2014) considering the fact that some proteins are not conserved (Figure 9). *VirB3* gene is absent in CFBP 2528. The protein encoded by this gene is thought to be involved in the production of the inner-membrane pore. *VirB5* is lacking in CFBP 7651 in consequence of genomic rearrangements in this region and deletion of several genes including *virB5*. This gene encodes a pilus-tip adhesin. Additional genes coding proteins predicted to be involved in conjugative transfer were identified in the nonpathogenic strain, CFBP 7634. For instance, a set of CDS encoding proteins showed more than 92% identity by BLASTP with *TrbB*, *TrbC*, *TrbD*, *TrbE*, *TrbJ*, *TrbL*, *TrbF*, *TrbG*, *TrbI* from other *Xanthomonas* such as *X. gardneri* for the best score. These CDS are embedded in an MGE starting with an integrase/recombinase (XARJCFBP7634_b09150) and containing phagic genes, outer membrane efflux protein encoding genes, transcription regulator and pirin genes. A similar MGE also containing *Trb* genes was identified in the other nonpathogenic strain, CFBP 7651. This MGE also starts with a recombinase (XARJCFBP7651_a21800) but these two arrays of T4SS and MGE encoding genes are localized in different regions in the chromosomes.

So far, no T4E were described in *Xanthomonas*. Identification of T4Es based on T4Es already described in other species is difficult because of the expected low sequence similarity. The number of newly discovered effectors is increasing, but only in a limited number of species (e.g., *Legionella* or *Helicobacter*). According to Wang et al. (2014), amino-acid composition and amino-acid specific positions in C-termini of T4E sequences can be used to predict T4Es. The two models “T4SEpre_bpBAac” and “T4SEpre_psAac” of the T4SEpre package (Wang et al., 2014) were used here to predict T4Es in the four *X. arboricola* genomes. Only locus tags predicted by both models were retained (Table S6), as advised by Wang et al. (2014) to limit the false positive results. We observed that in a same orthologous group of T4Es, some of them were predicted by both models whereas others were predicted by the T4SEpre_psAac model alone (in bold and italic in the Table S6). The number of predicted T4Es was higher in pathogenic strains (17 in CFBP 2528 and 18 in CFBP 7179) than in CFBP 7634 (14 predicted T4Es) and in

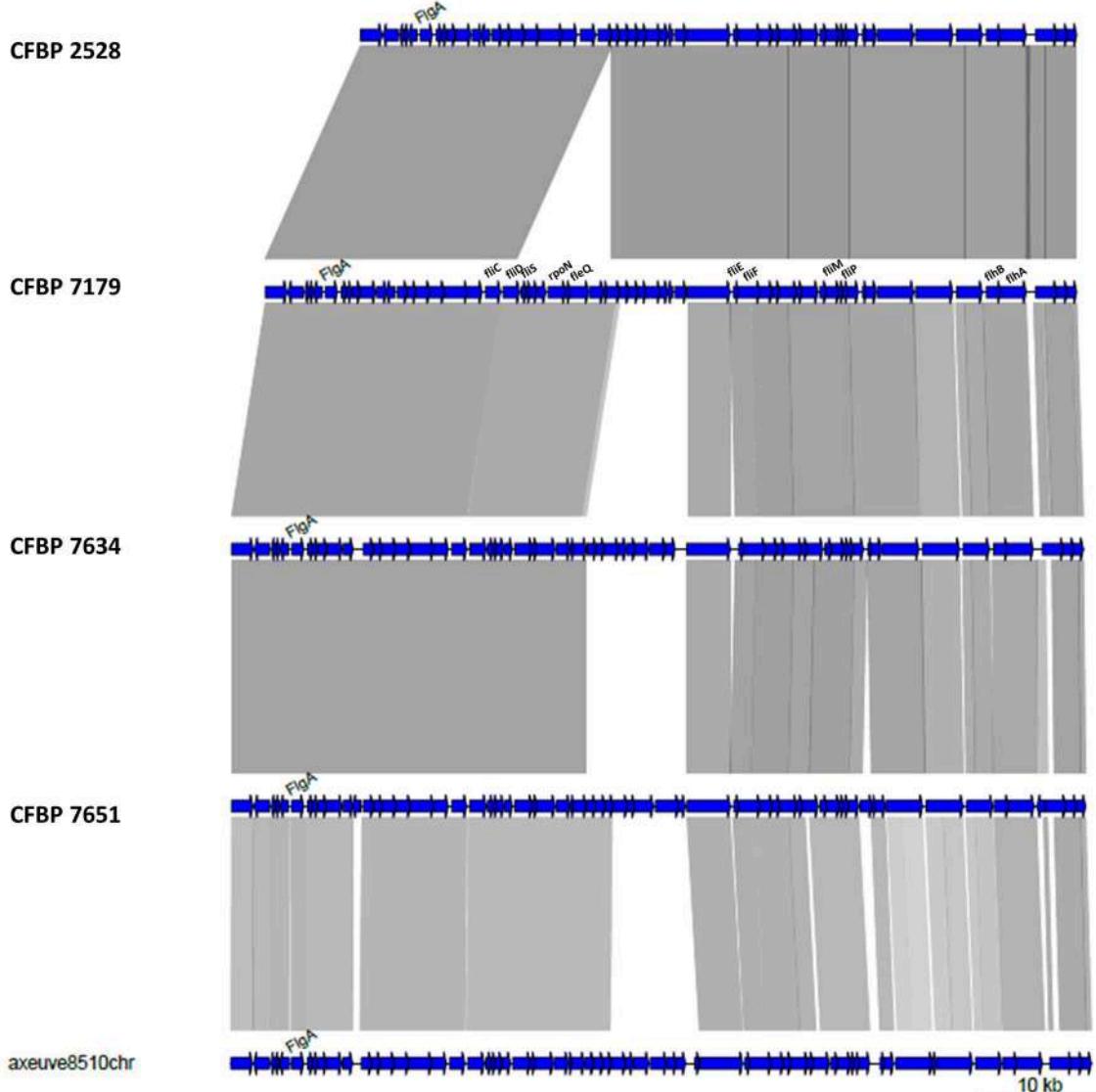


FIGURE 6 | Schematic representation of the flagellar gene cluster of the four *X. arboricola* strains.

CFBP 7651 (10 predicted T4Es). Predicted T4Es specific to the pathogenic strains were localized in region corresponding to mobile genetic elements (near transposases) or in region with low GC% that were probably acquired by LGT. Among them, besides hypothetical proteins, one putative T4E is a cytochrome c-type subunit, as predicted by Wang et al. (2014) in *Salmonella*, and another one is a transcription repressor DNA-binding protein.

Pathogenic Strain Genomes Encodes Specifically Two Non-Fimbrial Adhesins, FhaB and YadA-Like

Bacterial attachment to the host surface is mediated by adhesins that are non-fimbrial (autotransporters; filamentous

haemagglutinin-like proteins) or fimbrial (including type IV pili) adhesins, and both can contribute to virulence (Soto and Hultgren, 1999; Darsonval et al., 2009; Das et al., 2009; Gottig et al., 2009). The repertoires of genes coding non-fimbrial adhesins varied between the pathogenic strains and the nonpathogenic strains.

Two adhesin encoding genes were identified only in pathogenic strains. *yadA*-like and *fhaB* (Table S7). A *yadA*-like CDS was specifically found in pathogenic strains (GROUPORTHO3996), with the predicted domains serralysin-like metalloprotease C-terminal, trimeric autotransporter adhesin, YadA-like C-terminal. The predicted protein was 772-aa-long protein. This *yadA* sequence was associated in the two pathogenic strains with three other CDS encoding a S8 peptidase, a histidine kinase and a signal transduction response regulator,

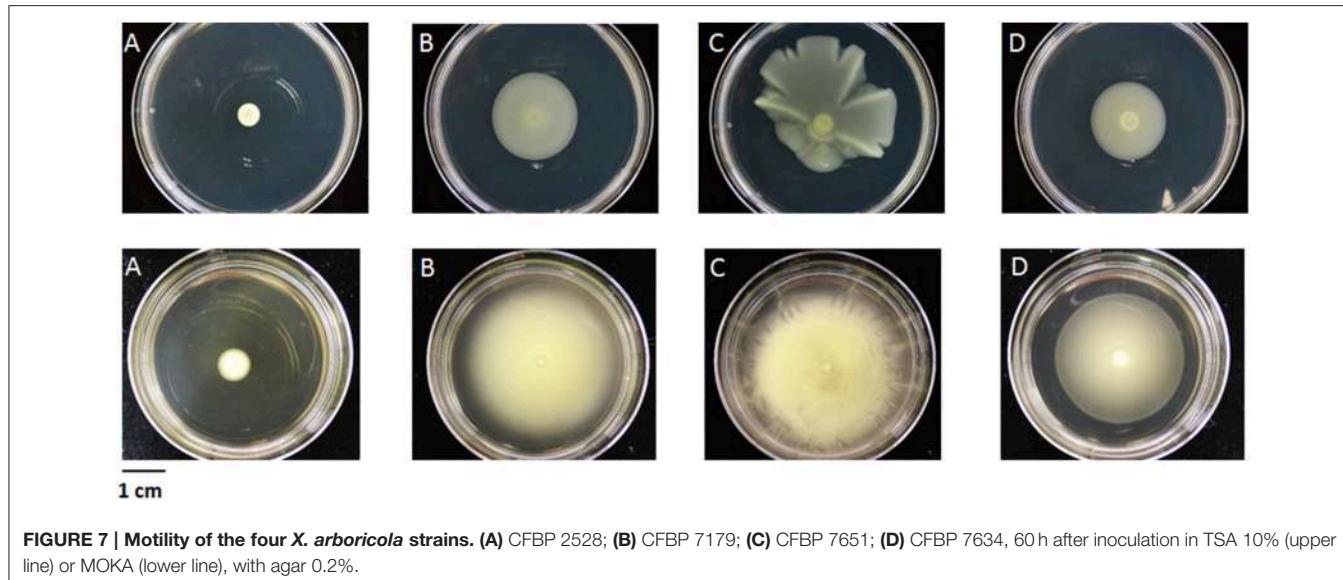


FIGURE 7 | Motility of the four *X. arboricola* strains. (A) CFBP 2528; **(B)** CFBP 7179; **(C)** CFBP 7651; **(D)** CFBP 7634, 60 h after inoculation in TSA 10% (upper line) or MOKA (lower line), with agar 0.2%.

| | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
|-----------|------------------------------|----------|----------|--------------|-----------------------|----|------------|----|
| CFBP 7634 | MAQVINTNVMSLNAQRNLNTSSASMSTS | I | DRLSSGLR | I | SERFTTQIRGLDVASRNANDG | I | SLAQTAEGAI | |
| CFBP 7651 | MAQVINTNVMSLNAQRNLNTSSASMSTS | I | DRLSSGLR | I | SERFTTQIRGLDVASRNANDG | I | SLAQTAEGAI | |
| Flg22 | ----- | ----- | DRLSSGLR | I | SERFTTQIRGLDVASRNANDG | I | SLAQTAEGAI | |
| CFBP 2528 | MAQVINTNVMSLNAQRNLNTSSSLALSI | DQLSSGKR | I | TSAFDAAAGGAI | AERFTTQIRGLDVASRNANDG | I | LSQTAEGAI | |
| CFBP 7179 | MAQVINTNVMSLNAQRNLNTSSSLALSI | DQLSSGKR | I | TSAFDAAAGGAI | AERFTTQIRGLDVASRNANDG | I | LSQTAEGAI | |

FIGURE 8 | Multiple alignment between the flagellin FliC N-terminal regions of the 4 *X. arboricola* strains and the flg22 peptide. The “D” to “V” switch between nonpathogenic and pathogenic strains is shown with the red box.

from the *cheY*-like family. In the corresponding regions of the nonpathogenic strains, five CDS encoding proteins with unknown function and a transcriptional regulator, one *tonB*-dependent receptor precursor, and one nuclease were found. One homolog of *fhab* was predicted (GROUPORTHO3859) to encode a 4308- and 4034-aa-long protein in pathogenic strains, CFBP 2528 and CFBP 7179, respectively. This protein had a N-terminal filamentous hemagglutinin domain and filamentous hemagglutinin repeats, and had also a N-terminal pectate lyase. This encoded protein showed 86% identity and 99% similarity with *FhaB* of *X. fuscans*, *Xanthomonas axonopodis* or *X. citri*. This CDS was near a predicted sequence of *fhaC* coding a hemolysin activation/secretion protein also specific to pathogenic strains. *FhaC* contained a polypeptide-transport-associated (POTRA) domain in N terminus. Two autotransporters, the monomeric, *XadA*, and the trimeric *YapH*, were encoded in the four genomes.

The four genomes also harbored several clusters that are predicted to be involved in the biogenesis of type IV pilus. Type IV pili (T4p) are surface filaments involved in different functions, such as twitching motility, adhesion, biofilm formation, natural transformation, pathogenicity, and immune escape (Mattick, 2002; Craig et al., 2004; Nudleman and Kaiser, 2004). The filament is composed of a major pilin PilA plus the minor pilins PilE, PilV, PilW, PilX, and FimU encountered in *P. aeruginosa* and in the type IVa system, which is the system

encountered in *Xanthomonas* (Burrows, 2012; Dunger et al., 2014). Recently, the organization of *pil* genes have been described in *X. citri* subsp *citri* (Dunger et al., 2014). In CFBP 7179 genome sequence, there was no predicted protein for PilA in the cluster (Table S8). Minor pilin PilX was absent in CFBP 7651 and PilV was absent in the pathogenic strains.

No T6SS was identified after BLASTN with known T6SS genes from Xcv85-10, Xff4834R, Xoo PXO99A against the four *X. arboricola* genomes.

One Chemosensor is Specific of Pathogenic Strains

For chemotaxis sensors, i.e., MCPs, slightly differential repertoires were observed between pathogenic and nonpathogenic strains (Table S9). Two MCPs (GROUPORTHO78, GROUPORTHO3512) were present in CFBP 2528, CFBP 7179, CFBP 7651 but absent in CFBP 7634 probably following genomic rearrangements leading to the deletion of the gene in CFBP 7634. Two other MCPs (GROUPORTHO4206, GROUPORTHO4237) were present in CFBP 7651 and CFBP 7634. Moreover, additional MCP were specific of each strain. One homologous CDS encoding a MCP was present in each of the two pathogenic strains (XARJCFBP2528_d01320; 733 aa and XARJCFBP7179_e01340, 733 aa) but was absent in the two nonpathogenic strains. In CFBP

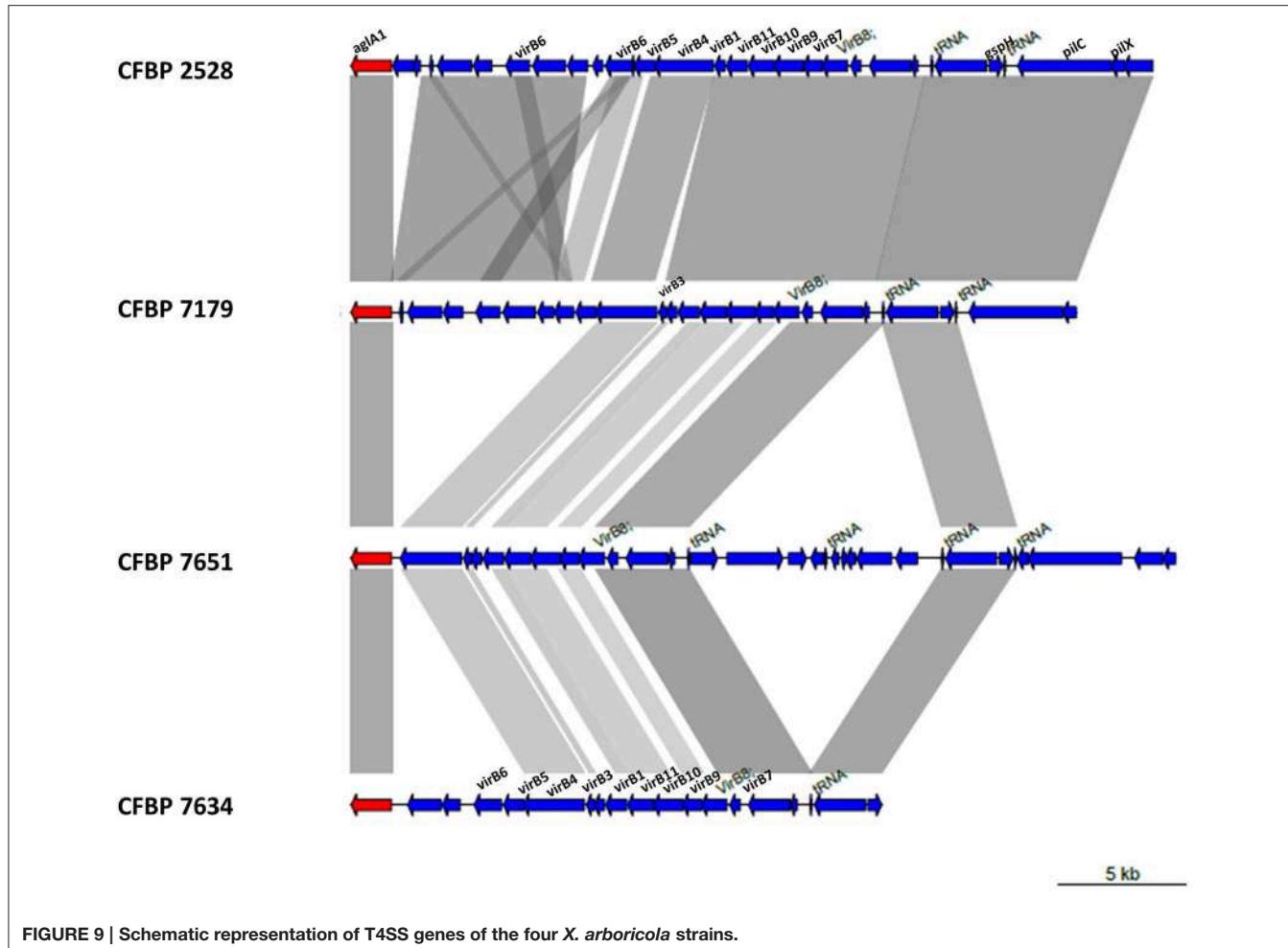


FIGURE 9 | Schematic representation of T4SS genes of the four *X. arboricola* strains.

7634 genes encoding for an integrase (XARJCFBP7634_b11370) and phagic proteins were found at the same location. In CFBP 7651, two CDS were predicted (XARJCFBP7651_a35490 and XARJCFBP7651_a35500), which each corresponded to the C-terminal and N-terminal fragments of this MCP identified in the pathogenic strains. This MCP was therefore nonfunctional in CFBP 7651.

The **Figure 10** highlights common and differential features of the four strains.

DISCUSSION

Features of the Genome Sequencing

Each genome length was very similar to other strains belonging to this species (Caballero et al., 2013; Vandroemme et al., 2013a). It is noticeable that smallest *X. arboricola* genomes corresponded to nonpathogenic strains lacking the *hrp*-island (this study and Vandroemme et al., 2013a). The high GC content was a common characteristic of most genera within the *Xanthomonadaceae* family (Saddler and Bradbury, 2005). ANI values were all above the threshold of the species level (Konstantinidis and Tiedje, 2005) and supports the grouping of the four strains within *X. arboricola*.

This is in accordance with previous study (Essakhi et al., 2015).

Mobile Genetics Elements IS

The IS number observed in pathogenic strains was similar to the one found in *X. campestris* pv *raphani* (Bogdanove et al., 2011). But the number and the diversity of IS elements can be larger in other *Xanthomonas* such as *X. oryzae* (Salzberg et al., 2008) with up to 245 elements distributed in six families, or such as *Xanthomonas fragariae* with up to 420 elements representing at least seven families (Vandroemme et al., 2013b). The IS families are distinct across *Xanthomonas* pathovars and species (Bogdanove et al., 2011). Most IS elements in our *X. arboricola* strains belonged to the IS3 and IS4 families (Table S3), which are also common families in *X. oryzae* and *X. campestris* genomes (Bogdanove et al., 2011). Similarly to *X. arboricola* genomes, in *X. campestris* pv *vesicatoria* and *X. axonopodis* pv *citri*, the IS3 family is also highly abundant, whereas in *X. campestris* pv *campestris* and in *X. fuscans* subsp. *fuscans*, most IS elements belong to the IS5 family (Thieme et al., 2005; Darrasse et al., 2013). IS21 and Tn3 transposase which

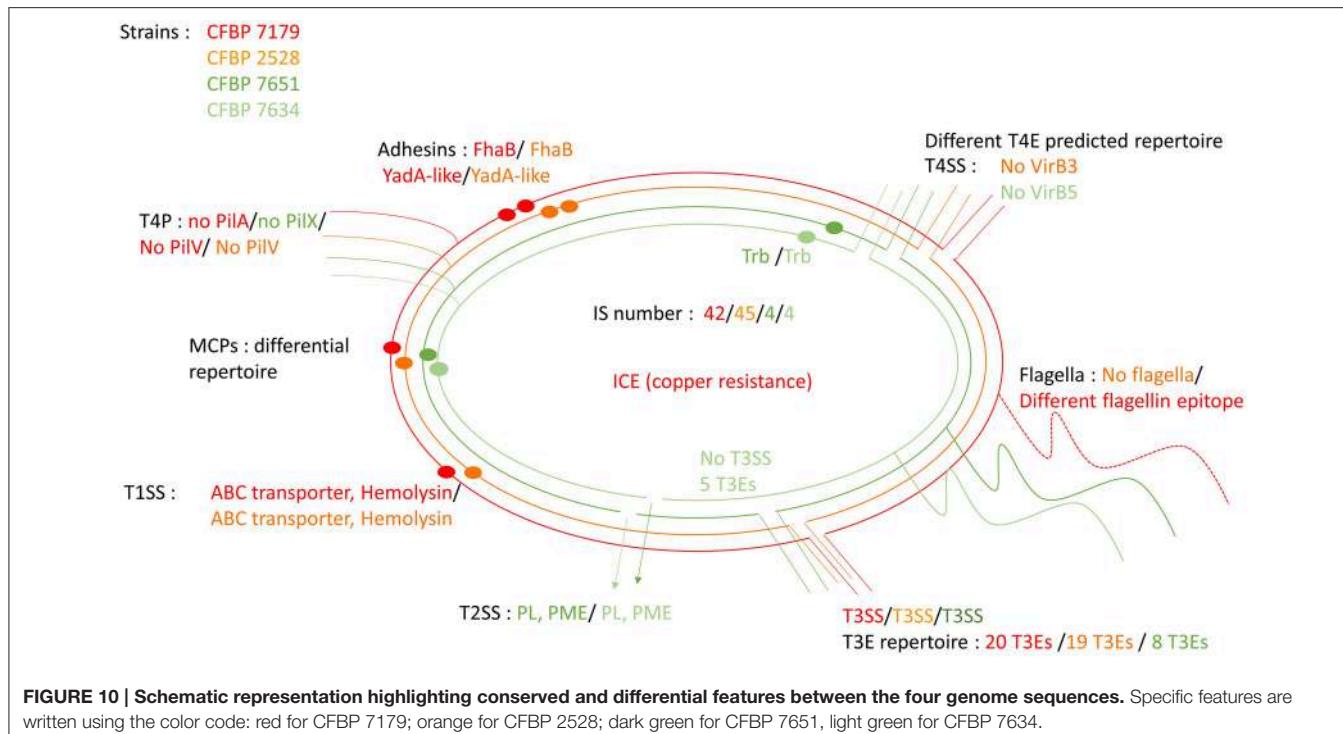


FIGURE 10 | Schematic representation highlighting conserved and differential features between the four genome sequences. Specific features are written using the color code: red for CFBP 7179; orange for CFBP 2528; dark green for CFBP 7651, light green for CFBP 7634.

are unique to CFBP 7179 in *Xanthomonas* genus, are located on an ICE (see below). Previous genomic and genetic studies have established that ISs are a major and powerful force in genome evolution. The presence of multiple copies of an IS in a genome can trigger intragenomic homologous recombination, resulting in genome rearrangements (inversions) or deletions of the intervening genomic region (Salzberg et al., 2008), and interruption of genes, operons, or transcriptional signals (Schneider and Lenski, 2004; Darrasse et al., 2013). Organisms harboring ISs are thus subject to a variety of mechanisms that enhance genomic plasticity. In the two pathogenic strains ISs or transposases were found in the vicinity of several accessory genes like T3Es linked to host specificity. The two nonpathogenic strains have 10 times less ISs than their pathogenic counterpart. A similar observation was previously reported for *Yersinia pestis* compared to *Yersinia pseudotuberculosis*, its ancestral species (Parkhill et al., 2001) and IS expansion was found to be linked to niche specialization in several bacteria (Mira et al., 2006). Genome sequencing has revealed that some genomes contained large numbers of ISs, while others had none at all but Touchon and Rocha (2007) found no association between IS frequency and pathogenicity. ISs can be transferred between genomes by LGT mechanisms (Frost et al., 2005). Iranzo et al. (2014) suggested that the LGT rate might be determined by the bacterial ecological niches. But the abundance of IS copies could be driven by duplication-deletion mechanism (Iranzo et al., 2014). It would be interesting to compare a higher number of genomes in the *X. arboricola* species to confirm the IS number differences between pathogenic and nonpathogenic strains and to try to reveal if mechanisms are used by IS to choose a target (Siguier et al., 2014).

ICE and Copper Resistance

We showed that strain CFBP 7179 and 36 *X. arboricola* pv. *juglandis* strains harbor an ICE with copper resistant genes and are actually copper resistant. Seven strains showed copper resistance without positive *cop* gene detection by PCR. We can't rule out that these strains present sequence variations at primer sites preventing their amplification but a mechanism for copper resistance independent of ICE located *cop* genes could exist in these strains. Indeed, Gardan et al. (1993) have previously shown that copper resistance in strain CFBP 1022 (one of the seven Xaj-ICE negative strains) was linked to the presence of a plasmid of 111 kb. Consequently, we can hypothesize that copper resistance in strains lacking the ICE is associated to a plasmid absent in genomes that were sequenced. In *X. campestris* pv. *campestris*, only plasmid-borne *cop* genes are essential for copper resistance. Nonetheless, homologs of these plasmid-borne copper resistance genes are present in the chromosomes of copper-sensitive and -resistant *Xanthomonas* (Behlau et al., 2011). In *X. axonopodis* pv. *vesicatoria*, expression of *copAB* cluster (putative copper binding proteins) is regulated by CopL, and the corresponding gene is located immediately upstream of *copAB* (Voloudakis et al., 2005). No homolog of *copL* was found in our four genomes which suggests that in *X. arboricola*, *copAB* regulation may be different. Copper is widely used in agriculture but the efficacy of copper is now reduced by the occurrence of copper-resistant strains in *Xanthomonas* (Behlau et al., 2011; Araújo et al., 2012) or *Pseudomonas* (Nakajima et al., 2002) species. The presence of this ICE in CFBP 7179 represents an example of probably environmental driven expansion of a bacterial genome because of a high selective pressure due to the extensive use of copper. Indeed Xaj-ICE has

been only retrieved in strains responsible for recent epidemics in France (Hajri et al., 2010). The acquisition of this element probably conferred a selective advantage to these strains. Most CDS of the Xaj-ICE showed high identity with genes from *P. aeruginosa* strains and *S. maltophilia* strain D457, which belongs to the *Xanthomonadaceae* family. It suggests that the Xaj-ICE should have been taken up by lateral transmission into the CFBP 7179 genome from a different genus donor strain. Transfers between distantly related genomes exist even if genome sequences dissimilarity is a barrier to LGT (Popa et al., 2011). It is interesting to notice that Xaj-ICE is the first ICE detected in *Xanthomonas* to date.

Other MGEs

The higher number in prophage in pathogenic strains could suggest a higher sensitivity to it but it should be assessed with other genomes. According to the evolutionary scheme proposed by Gillings et al. (2005) for the integron, we can hypothesize that genetic rearrangements in the cassettes of this integron could have accompanied niche specialization of our strains after the loss of activity of the integrase. How do the cassette arrays determine the ecological niche of each strain? This is currently unknown because most genes carried by the integron code for proteins with unknown activities.

Secretion Systems

T1SS

Hemolysins are of great importance for the pathogenesis in the host organism (Kanonenberg et al., 2013). CDSs linked to hemolysin secretion or hemolysin were only present in pathogenic strains. We hypothesized that these CDS were acquired by the common ancestor of the *X. arboricola* pv. *juglandis* strains by recombination. To our knowledge, no *Xanthomonas* hemolysin mutant is described to date: it would be interesting to realize functional analyses to study the role of these CDSs in pathogenicity.

T2SS

The absence of a *cbhA* ortholog in the four *X. arboricola* genome sequences is in agreement with their known inability to colonize xylem vessels. The gene *cbhA* is conserved in the xylem-invading *Xanthomonas* species (*X. albilineans*, *X. oryzae* pv. *oryzae*, *X. campestris* pv. *campestris*, *X. campestris* pv. *vasculorum*, and *X. campestris* pv. *musacearum*), but is missing in the non-vascular *Xanthomonas* species (*X. oryzae* pv. *oryzicola*, *X. axonopodis* pv. *citri*, *X. axonopodis* pv. *vesicatoria*) (Pieretti et al., 2012). The *cbhA* gene was also shown to contribute to virulence of the xylem-invading pathogen *Ralstonia solanacearum* (Liu et al., 2005). Putative aminopeptidase and pectate lyase encoding CDSs were present in the nonpathogenic strains, while derived in pathogenic strains: we hypothesized that these fragments represented remnants of the genes present in the common ancestor of pathogenic and nonpathogenic strains. Similarly, as homologs of pectinesterase gene present in nonpathogenic strains, are found in genomes of pathogenic strains belonging to other *Xanthomonas* species, the most parsimonious hypothesis will be in favor of the loss of this gene in the common ancestor of the pathogenic strains CFBP 2528 and CFBP 7179.

Putative pectate lyase and xylosidase homologs were only detected in pathogenic strains and seem to have been acquired in pathogenic strains by LGT. PCWDEs are carbohydrate-active enzymes that have been classified in different families based on homology criteria (<http://www.cazy.org/>, Cantarel et al., 2009). Pectin methylesterase (PME) catalyzes de-esterification of pectin to make substrates available for subsequent action by polygalacturonase and pectate lyase. These enzymes act in concert in pectin degradation. The ability to degrade pectin may facilitate pathogen invasion into the cells of host plants and is useful for pathogens in term of virulence (Hugouvieux-Cotte-Pattat et al., 2014). Although our *in vitro* tests may not detect all pectinase activities, pectinase activities were only observed for the nonpathogenic strains. It has obviously no role in disease process, but may participate in nutrient uptake by these bacteria *in planta*. Vorhölter et al. (2012) shown that oligogalacturonides generated by pectate lyase activity in a pathogenic interaction involving *X. campestris* pv. *campestris*, could elicit plant defense reactions. The two pathogenic strains which have a putative pectate lyase but undetectable PL activity, could have evolved to avoid production of PAMPs by PL activity.

T3SS

The *hrp-hrc* region and the *hrpF* peninsula were absent in the nonpathogenic strain CFBP 7634 however five T3E genes were retrieved in its genome. Given that HrpF functions as a translocon of effector proteins into the host cell (Rossier et al., 2000; Büttner and Bonas, 2002), we can assume that CFBP 7634 T3Es could not be translocated into plant cells. Previous studies showed that mutation of the *hrpF* locus of *X. oryzae* pv. *oryzicola* strain resulted in the loss of pathogenicity in rice and the inability to induce HR in non-host tobacco (Zou et al., 2006). Similarly, mutations in *hrpF* of *X. c. pv. vesicatoria* strain or *X. axonopodis* pv. *glycines* strain resulted in strains that were nonpathogenic in host plants and unable to elicit race-specific HRs (Rossier et al., 2000; Kim et al., 2003).

T3Es

Pathogenic strains presented a repertoire of T3Es, which was moderately large in comparison to other xanthomonads (Hajri et al., 2009). The presence of *awr4*, *sfrJ* and *xopAR* highlighted the limits of the PCR compared to genome sequencing. *sfrJ* is secreted through SPI2 in *Salmonella* (Cordero-Alba et al., 2012). However, these authors suggested a SPI2 independent role in environment as *sfrJ* is also present in a commensal *E. coli* strain devoided of T3SS. Differential T3Es in pathogenic strains (*xopAH* in CFBP 2528; *xopAI* and *xopB* in CFBP 7179) were probably acquired by different mechanisms (homologous recombination or LGT). Transcription activator-like (TAL) effectors were not detected in the four genome sequences but HiSeq technology is not the method of choice to detect these TAL effectors, because of internal repeats. However, Hajri et al. (2012) only detected *avrBs3* by PCR in the pathovar *corylina*.

Flagella

CFBP 2528 is impaired in motility because of a loss of *fliS* and *rpoN* CDSs. Mutant affected in *fliS* still produce functional flagella in *Salmonella* (Yokoseki et al., 1995) whereas *rpoN*

mutant has been shown to loose motility in *X. oryzae* pv *oryzae* (Tian et al., 2014). Darrasse et al. (2013) also reported a lack of motility in other *X. arboricola* strains, such as pathovar *corylina* type-strain CFBP1159. We suggest that the modifications observed for the flagellin epitope in both pathogenic strains affect flagellin perception *in planta* and could prevent recognition of pathogenic strains at an early stage of infection as already observed for other *Xanthomonas* (Sun et al., 2006). This could be a mechanism of evolution to avoid PAMP-triggered immunity as previously suggested for other *Xanthomonas* strains (Jacobs et al., 2015).

T4SS and T4Es

The T4SS translocates DNA and proteins to bacterial or eukaryotic target cells by a direct cell-to-cell contact (Christie et al., 2014). A *virB* cluster was found in the four *X. arboricola* strains. *VirB3* gene is absent in CFBP 2528. The protein encoded by this gene seems to be essential for pilus assembly and substrate translocation (Guglielmini et al., 2014). *VirB5* is lacking in CFBP 7651. This gene encodes a pilus-tip adhesin that could initiate contact with host cells (Backert et al., 2008). An additional T4SS locus with *Trb* genes was found in nonpathogenic strains in different regions in the chromosomes, near a recombinase, suggesting independent LGT events in these two nonpathogenic strains. As no T4E were described so far in *Xanthomonas*, we used the method of Wang et al. (2014) to predict T4Es. Among 10 to 18 T4Es detected *in silico*, a transcription repressor DNA-binding protein was found in pathogenic strains: this latter predicted T4E is perhaps interesting as T4Es can manipulate host pathways for a survival strategy (Hubber and Roy, 2010). Nevertheless, these proteins need further experimental validation analysis. It should be noted that in *X. citri* subsp. *citri* T4SS is not induced under infection conditions (Jacob et al., 2014).

Adhesion and Chemotaxis

Adhesion

One homolog of *fhaB* was detected only in pathogenic strains. *FhaB* seems to be involved in the colonization of both the leaf surface and the apoplast in *X. citri* subsp. *citri* (Gottig et al., 2009). However, other non fimbrial adhesins (*XadA* and *YapH*) are encoded in the four genomes. The absence of CDSs encoding *PilA* in CFBP 7179, *PilX* in CFBP 7651 and *PilV* in both pathogenic strains suggests that T4p biogenesis is probably impaired in these strains (Nguyen et al., 2015). It could be interesting to conduct *in vivo* adhesion analysis in order to observe behaviors of the strains.

Chemotaxis

The repertoire of MCPs was different between strains with MCPs specific of pathogenic or of nonpathogenic strains. MCPs, which are cell membrane-bound chemoreceptors, are involved in the detection of molecules such as attractant or repellent. Subsequent movement of the cell through flagellar motility allow bacteria to go toward or away from perceived molecules (Vladimirov and Sourjik, 2009). Ability to specifically detect a molecule could allow the pathogenic strains to colonize environment that could remain inaccessible for nonpathogenic strains. This suggested

different chemotaxis properties. Characterizing repertoires of MCPs in a large collection of strains and functional analyses would be interesting to further study the role of the different MCPs in the plant colonization.

CONCLUSION

Differences between the two pathogenic strains, CFBP 2528 and CFBP 7179, and the two nonpathogenic strains, CFBP 7634 and CFBP 7651, all isolated from the same plant species, i.e., walnut, concerns a full range of functions involved in ability to colonize plants from sensing of the environment and to cross-talk with the immune system. Several non-fimbrial adhesins and one hemolysin may allow pathogenic strains to adhere or aggregate more efficiently than nonpathogenic strains to plant tissues or to form more stable or resistant biofilms. Differential repertoires of PCWDEs between pathogens and commensals could allow the colonization of separate niches. A larger repertoire of T3Es in pathogens may be an efficient means to interfere with plant immune system allowing ingress and multiplication inside plant tissues, but also can significantly contribute to growth. One chemoreceptor was specifically identified in pathogenic strains and might allow pathogens to differentially perceive the environment. We highlighted a larger set of various mobile genetic elements in pathogen genomes and different genome organizations, which were driven by recombination events or horizontal transfers. We propose that these events were closely related to bacteria encountered in their physical environment rather than to phylogenetically related bacteria. From these genome comparisons it is not possible to answer the question of the origin of these strains. Do pathogenic strains evolve from nonpathogenic ancestor through acquisition of pathogenesis-associated genes or in contrast do nonpathogenic strains evolve from pathogenic ones through the loss of energetically costly functions? Bacteria live in interaction with their biotic environment and evolve in dynamic microbial communities, which may act as reservoir of genes and also favor loss of genes by providing mobile genetic elements. A better understanding of emergence of pathotypes and finally diseases may arise from the deciphering of whole microbial communities.

ACKNOWLEDGMENTS

This project has been financed by Direction Générale de l'Armement (REI project # 2010 34007) and SFR QUASAV (PATHOCOM project). We thank the Collection Française de Bactéries associées aux Plantes (CIRM-CFBP), INRA, Angers, France, for providing *X. arboricola* strains. We thank Jérôme Gouzy and Sébastien Carrère for performing automatic annotation of the genomes. We thank Céline Rousseau for adapting script from R package.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2015.01126>

REFERENCES

- Akhter, S., Aziz, R. K., and Edwards, R. A. (2012). PhiSpy: A novel algorithm for finding prophages in bacterial genomes that combines similarity-and composition-based strategies. *Nucleic Acids Res.* 40, 1–13. doi: 10.1093/nar/gks406
- Araki, H., Tian, D., Goss, E. M., Jakob, K., Halldorsdottir, S. S., Kreitman, M., et al. (2006). Presence/absence polymorphism for alternative pathogenicity islands in *Pseudomonas viridisflava*, a pathogen of Arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* 103, 5887–5892. doi: 10.1073/pnas.0601431103
- Araújo, E. R., Pereira, R. C., Ferreira, M. A. S. V., Fitopatología, D., Brasília, U., De Hortaliças, E., et al. (2012). Sensitivity of xanthomonads causing tomato bacterial spot to copper and streptomycin and *in vivo* infra-specific competitive ability in *Xanthomonas perforans* resistant and sensitive to copper. *J. Plant Pathol.* 94, 79–87. doi: 10.4454/jppfa.2012.004
- Backert, S., Fronzes, R., and Waksman, G. (2008). VirB2 and VirB5 proteins: specialized adhesins in bacterial type-IV secretion systems? *Trends Microbiol.* 16, 409–413. doi: 10.1016/j.tim.2008.07.001
- Barionovi, D., and Scorticchini, M. (2008). Integron variability in *Xanthomonas arboricola* pv. *juglandis* and *Xanthomonas arboricola* pv. *pruni* strains. *FEMS Microbiol. Lett.* 288, 19–24. doi: 10.1111/j.1574-6968.2008.01315.x
- Behlau, F., Canteros, B. I., Minsavage, G. V., Jones, J. B., and Graham, J. H. (2011). Molecular characterization of copper resistance genes from *Xanthomonas citri* subsp. *citri* and *Xanthomonas alfalfa* subsp. *citrumelonis*. *Appl. Environ. Microbiol.* 77, 4089–4096. doi: 10.1128/AEM.03043-10
- Bogdanove, A. J., Koebnik, R., Lu, H., Furutani, A., Anguoli, S. V., Patil, P. B., et al. (2011). Two new complete genome sequences offer insight into host and tissue specificity of plant pathogenic *Xanthomonas* spp. *J. Bacteriol.* 193, 5450–5464. doi: 10.1128/JB.05262-11
- Burrows, L. L. (2012). Twitching motility: type IV pili in action. *Annu. Rev. Microbiol.* 66, 493–520. doi: 10.1146/annurev-micro-092611-150055
- Burrus, V., and Waldor, M. K. (2004). Shaping bacterial genomes with integrative and conjugative elements. *Res. Microbiol.* 155, 376–386. doi: 10.1016/j.resmic.2004.01.012
- Büttner, D., and Bonas, U. (2002). Port of entry—the type III secretion translocon. *Trends Microbiol.* 10, 186–192. doi: 10.1016/S0966-842X(02)02331-4
- Caballero, J. I., Zerillo, M. M., Snelling, J., and Boucher, C. (2013). Genome sequence of *Xanthomonas arboricola* pv. *corylina*, isolated from Turkish Filbert in Colorado. *Genome Announc.* 1, 2012–2013. doi: 10.1128/genomea.00246-13
- Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The Carbohydrate- Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* 37, D233–D238. doi: 10.1093/nar/gkn663
- Christie, P. J., Whitaker, N., and González-Rivera, C. (2014). Mechanism and structure of the bacterial type IV secretion systems. *Biochim. Biophys. Acta* 1843, 1578–1591. doi: 10.1016/j.bbampcr.2013.12.019
- Cordero-Alba, M., Bernal-Bayard, J., and Ramos-Morales, F. (2012). SrfJ, a *Salmonella* type III secretion system effector regulated by PhoP, RcsB, and IolR. *J. Bacteriol.* 194, 4226–4236. doi: 10.1128/JB.00173-12
- Craig, L., Pique, M. E., and Tainer, J. A. (2004). Type IV pilus structure and bacterial pathogenicity. *Nat. Rev. Microbiol.* 2, 363–378. doi: 10.1038/nrmicro885
- Darrasse, A., Carrère, S., Barbe, V., Boureau, T., Arrieta-Ortiz, M. L., Bonneau, S., et al. (2013). Genome sequence of *Xanthomonas fuscans* subsp. *fuscans* strain 4834-R reveals that flagellar motility is not a general feature of xanthomonads. *BMC Genomics* 14:761. doi: 10.1186/1471-2164-14-761
- Darsonval, A., Darrasse, A., Durand, K., Bureau, C., Cesbron, S., and Jacques, M.-A. (2009). Adhesion and fitness in the bean phyllosphere and transmission to seed of *Xanthomonas fuscans* subsp. *fuscans*. *Mol. Plant Microbe Interact.* 22, 747–757. doi: 10.1094/MPMI-22-6-0747
- Das, A., Rangaraj, N., and Sonti, R. V. (2009). Multiple adhesin-like functions of *Xanthomonas oryzae* pv. *oryzae* are involved in promoting leaf attachment, entry, and virulence on rice. *Mol. Plant Microbe Interact.* 22, 73–85. doi: 10.1094/MPMI-22-1-0073
- Diallo, M. D., Monteil, C. L., Vinatzer, B. A., Clarke, C. R., Glaux, C., Guilbaud, C., et al. (2012). *Pseudomonas syringae* naturally lacking the canonical type III secretion system are ubiquitous in nonagricultural habitats, are phylogenetically diverse and can be pathogenic. *ISME J.* 6, 1325–1335. doi: 10.1038/ismej.2011.202
- Dunger, G., Guzzo, C. R., Andrade, M. O., Jones, J. B., and Farah, C. S. (2014). *Xanthomonas citri* subsp. *citri* type IV Pilus is required for twitching motility, biofilm development, and adherence. *Mol. Plant Microbe Interact.* 27, 1132–1147. doi: 10.1094/MPMI-06-14-0184-R
- Dye, D. W., Bradbury, J., Goto, M., Hayward, A. C., Lelliott, R. A., and Schröth, M. N. (1980). International standards for naming pathovars of phytopathogenic bacteria and a list of pathovar names and pathotype strains. *Rev. Plant Pathol.* 59, 153–168.
- Essakhi, S., Cesbron, S., Fischer-Lesaux, M., Bonneau, S., Jacques, M. A., and Manceau, C. (2015). Phylogenetical and Variable-Number Tandem-Repeat analysis identified nonpathogenic lineages within *Xanthomonas arboricola* lacking the canonical Type III secretion system. *Appl. Environ. Microbiol.* 81, 5395–5410. doi: 10.1128/AEM.00835-15
- Felix, G., Duran, J. D., Volko, S., and Boller, T. (1999). Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. *Plant J.* 18, 265–276. doi: 10.1046/j.1365-313X.1999.00265.x
- Fischer-Le Saux, M., Bonneau, S., Essakhi, S., Manceau, C., and Jacques, M. A. (2015). Aggressive emerging pathovars of *Xanthomonas arboricola* represent widespread epidemic clones that are distinct from poorly pathogenic strains, as revealed by multilocus sequence typing. *Appl. Environ. Microbiol.* 81, 4651–4668. doi: 10.1128/AEM.00050-15
- Frost, L. S., Leplae, R., Summers, A. O., and Toussaint, A. (2005). Mobile genetic elements: the agents of open source evolution. *Nat. Rev. Microbiol.* 3, 722–732. doi: 10.1038/nrmicro1235
- Gardan, L., Brault, T., and Germain, E. (1993). Copper resistance of *Xanthomonas campestris* pv. *juglandis* in French walnut orchards and its association with conjugative plasmids. *Acta Hort.* 311, 259–265. doi: 10.17660/ActaHortic.1993.311.33
- Gillings, M. R., Holley, M. P., Stokes, H. W., and Holmes, A. J. (2005). Integrins in *Xanthomonas*: a source of species genome diversity. *Proc. Natl. Acad. Sci. U.S.A.* 102, 4419–4424. doi: 10.1073/pnas.0406620102
- Gottig, N., Garavaglia, B. S., Garofalo, C. G., Orellano, E. G., and Ottado, J. (2009). A filamentous hemagglutinin-like protein of *Xanthomonas axonopodis* pv. *citri*, the phytopathogen responsible for citrus canker, is involved in bacterial virulence. *PLoS ONE* 4:e4358. doi: 10.1371/journal.pone.0004358
- Grant, J. R., and Stothard, P. (2008). The CGView Server: a comparative genomics tool for circular genomes. *Nucleic Acids Res.* 36, 181–184. doi: 10.1093/nar/gkn179
- Guglielmini, J., Nérón, B., Abby, S. S., Garcillán-Barcia, M. P., La Cruz, D. F., and Rocha, E. P. C. (2014). Key components of the eight classes of type IV secretion systems involved in bacterial conjugation or protein secretion. *Nucleic Acids Res.* 42, 5715–5727. doi: 10.1093/nar/gku194
- Guy, L., Kultima, J. R., and Andersson, S. G. E. (2010). genoPlotR: Comparative gene and genome visualization in R. *Bioinformatics* 26, 2334–2335. doi: 10.1093/bioinformatics/btq413
- Hajri, A., Brin, C., Hunault, G., Lardeux, F., Lemaire, C., Manceau, C., et al. (2009). A «repertoire for repertoire» hypothesis: Repertoires of type three effectors are candidate determinants of host specificity in *Xanthomonas*. *PLoS ONE* 4:e6632. doi: 10.1371/journal.pone.0006632
- Hajri, A., Meyer, D., Delort, F., Guillaumès, J., Brin, C., and Manceau, C. (2010). Identification of a genetic lineage within *Xanthomonas arboricola* pv. *juglandis* as the causal agent of vertical oozing canker of Persian (English) walnut in France. *Plant Pathol.* 59, 1014–1022. doi: 10.1111/j.1365-3059.2010.02362.x
- Hajri, A., Pothier, J. F., Fischer-Le Saux, M., Bonneau, S., Poussier, S., Manceau, C., et al. (2012). Type three effector gene distribution and sequence analysis provide new insights into the pathogenicity of plant-pathogenic *Xanthomonas arboricola*. *Appl. Environ. Microbiol.* 78, 371–384. doi: 10.1128/AEM.06119-11
- Hayward, A. C. (1993). “The hosts of *Xanthomonas*,” in *Xanthomonas*, eds J. G. Swings and E. L. Civerolo (London: Chapman and Hall), 1–119.
- Hubber, A., and Roy, C. R. (2010). Modulation of host cell function by *Legionella pneumophila* type IV effectors. *Annu. Rev. Cell Dev. Biol.* 26, 261–283. doi: 10.1146/annurev-cellbio-100109-104034
- Hugouvieux-Cotte-Pattat, N., Condemeine, G., and Shevchik, V. E. (2014). Bacterial pectate lyases, structural and functional diversity. *Environ. Microbiol. Rep.* 6, 427–440. doi: 10.1111/1758-2229.12166

- Iranzo, J., Gómez, M. J., López de Saro, F. J., and Manrubia, S. (2014). Large-scale genomic analysis suggests a neutral punctuated dynamics of transposable elements in bacterial genomes. *PLoS Comput. Biol.* 10:e1003680. doi: 10.1371/journal.pcbi.1003680
- Jacob, T. R., De Laia, M. L., Moreira, L. M., Gonçalves, J. F., Carvalho, F. M. D. S., Ferro, M. I. T., et al. (2014). Type IV secretion system is not involved in infection process in citrus. *Int. J. Microbiol.* 2014:763575. doi: 10.1155/2014/763575
- Jacobs, J. M., Pesce, C., Lefevre, P., and Koebnik, R. (2015). Comparative genomics of a cannabis pathogen reveals insight into the evolution of pathogenicity in *Xanthomonas*. *Front. Plant Sci.* 6:431. doi: 10.3389/fpls.2015.00431
- Kanonenberg, K., Schwarz, C. K., and Schmitt, L. (2013). Type I secretion systems—a story of appendices. *Res. Microbiol.* 164, 596–604. doi: 10.1016/j.resmic.2013.03.011
- Kim, J. G., Park, B. K., Yoo, C. H., Jeon, E., Oh, J., and Hwang, I. (2003). Characterization of the *Xanthomonas axonopodis* pv. *glycines* Hrp pathogenicity island. *J. Bacteriol.* 185, 3155–3166. doi: 10.1128/JB.185.10.3155-3166.2003
- Konstantinidis, K. T., and Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2567–2572. doi: 10.1073/pnas.0409727102
- Koressaar, T., and Remm, M. (2007). Enhancements and modifications of primer design program Primer3. *Bioinformatics* 23, 1289–1291. doi: 10.1093/bioinformatics/btm091
- Lee, Y. A., Hendson, M., Panopoulos, N. J., and Schroth, M. N. (1994). Molecular cloning, chromosomal mapping, and sequence analysis of copper resistance genes from *Xanthomonas campestris* pv. *juglandis*: homology with small blue copper proteins and multicopper oxidase. *J. Bacteriol.* 176, 173–188.
- Li, L., Stoeckert, C. J. Jr., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189. doi: 10.1101/gr.1224503
- Li, Y., Hu, Y., Bolund, L., and Wang, J. (2010). State of the art de novo assembly of human genomes from massively parallel sequencing data. *Hum. Genomics* 4, 271–277. doi: 10.1186/1479-7364-4-4-271
- Liu, H., Zhang, S., Schell, M. A., and Denny, T. P. (2005). Pyramiding unmarked deletions in *Ralstonia solanacearum* shows that secreted proteins in addition to plant cell-wall degrading enzymes contribute to virulence. *Mol. Plant Microbe Interact.* 18, 1296–1305. doi: 10.1094/MPMI-18-1296
- Makarova, K. S., Wolf, Y. I., and Koonin, E. V. (2013). Comparative genomics of defense systems in archaea and bacteria. *Nucleic Acids Res.* 41, 4360–4377. doi: 10.1093/nar/gkt157
- Mattick, J. S. (2002). Type IV pili and twitching motility. *Annu. Rev. Microbiol.* 56, 289–314. doi: 10.1146/annurev.micro.56.012302.160938
- Mira, A., Pushker, R., and Rodriguez-Valera, F. (2006). The Neolithic revolution of bacterial genomes. *Trends Microbiol.* 14, 200–206. doi: 10.1016/j.tim.2006.03.001
- Nakajima, M., Goto, M., and Hibi, T. (2002). Similarity between Copper Resistance Genes from *Pseudomonas syringae* pv. *actinidiae* and *P. syringae* pv. *tomato*. *J. Gen. Plant Pathol.* 68, 68–74. doi: 10.1007/PL00013056
- Navarro, L., Zipfel, C., Rowland, O., Keller, I., Robatzek, S., Boller, T., et al. (2004). The Transcriptional Innate Immune Response to Flg22. Interplay and overlap with Avr gene-dependent defense responses and bacterial pathogenesis. *Plant Physiol.* 135, 1113–1128. doi: 10.1104/pp.103.036749
- Nguyen, Y., Sugiman-Marangos, S., Harvey, H., Bell, S. D., Charlton, C. L., Junop, M. S., et al. (2015). *Pseudomonas aeruginosa* minor pilins prime Type IVa Pilus assembly and promote surface display of the PilY1 adhesin. *J. Biol. Chem.* 290, 601–611. doi: 10.1074/jbc.M114.616904
- Nudelman, E., and Kaiser, D. (2004). Pulling together with type IV pili. *J. Mol. Microbiol. Biotechnol.* 7, 52–62. doi: 10.1159/000077869
- Parkhill, J., Wren, B. W., Thomson, N. R., Titball, R. W., Holden, M. T. G., Prentice, M. B., et al. (2001). Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* 413, 523–527. doi: 10.1038/35097083
- Pieretti, I., Royer, M., Barbe, V., Carrere, S., Koebnik, R., Couloux, A., et al. (2012). Genomic insights into strategies used by *Xanthomonas albilineans* with its reduced artillery to spread within sugarcane xylem vessels. *BMC Genomics* 13:658. doi: 10.1186/1471-2164-13-658
- Popa, O., Hazkani-Covo, E., Landan, G., Martin, W., and Dagan, T. (2011). Directed networks reveal genomic barriers and DNA repair bypasses to lateral gene transfer among prokaryotes. *Genome Res.* 21, 599–609.
- Potnis, N., Krasileva, K., Chow, V., Almeida, N. F., Patil, P. B., Ryan, R. P., et al. (2011). Comparative genomics reveals diversity among xanthomonads infecting tomato and pepper. *BMC Genomics* 12:146. doi: 10.1186/1471-2164-12-146
- Rossier, O., Van den Ackerveken, G., and Bonas, U. (2000). HrpB2 and hrpF from *Xanthomonas* are type III-secreted proteins and essential for pathogenicity and recognition by the host plant. *Mol. Microbiol.* 38, 828–838. doi: 10.1046/j.1365-2958.2000.02173.x
- Ryan, R. P., Vorhölter, F.-J., Potnis, N., Jones, J. B., Van Sluys, M.-A., Bogdanov, A. J., et al. (2011). Pathogenomics of *Xanthomonas*: understanding bacterium–plant interactions. *Nat. Rev. Microbiol.* 9, 344–355. doi: 10.1038/nrmicro2558
- Saddler, G. S., and Bradbury, J. F. (2005). “Order III. Xanthomonadales ord. nov.” in *Bergey’s Manual® of Systematic Bacteriology* (The Proteobacteria), part B (The Gammaproteobacteria), eds D. J. Brenner, N. R. Krieg, J. T. Staley, G. M. Garrity (New York, NY: Springer), 63–122.
- Sallet, E., Gouzy, J., and Schiex, T. (2014). EuGene-PP: a next-generation automated annotation pipeline for prokaryotic genomes. *Bioinformatics* 30, 2659–2661. doi: 10.1093/bioinformatics/btu366
- Salzberg, S. L., Sommer, D. D., Schatz, M. C., Phillippy, A. M., Rabinowicz, P. D., Tsuge, S., et al. (2008). Genome sequence and rapid evolution of the rice pathogen *Xanthomonas oryzae* pv. *oryzae* PXO99A. *BMC Genomics* 9:204. doi: 10.1186/1471-2164-9-204
- Schneider, D., and Lenski, R. E. (2004). Dynamics of insertion sequence elements during experimental evolution of bacteria. *Res. Microbiol.* 155, 319–327. doi: 10.1016/j.resmic.2003.12.008
- Scorticini, M., Marcelletti, S., Ferrante, P., and Firrao, G. (2013). A Genomic Redefinition of *Pseudomonas avellanae* species. *PLoS ONE* 8:e75794. doi: 10.1371/journal.pone.0075794
- Shi, Q., Febres, V. J., Jones, J. B., and Moore, G. A. (2015). Responsiveness of different citrus genotypes to the *Xanthomonas citri* ssp. *citri*-derived pathogen-associated molecular pattern (PAMP) flg22 correlates with resistance to citrus canker. *Mol. Plant Pathol.* 16, 507–520. doi: 10.1111/mpp.12206
- Siguier, P., Gourbeyre, E., and Chandler, M. (2014). Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiol. Rev.* 38, 865–891. doi: 10.1111/1574-6976.12067
- Soto, G. E., and Hultgren, S. J. (1999). Bacterial adhesins: common themes and variations in architecture and assembly MINIREVIEW bacterial adhesins: common themes and variations in architecture and assembly. *J. Bacteriol.* 181, 1059–1071.
- Sun, W., Dunning, F. M., Pfund, C., Weingarten, R., and Bent, A. F. (2006). Within-species flagellin polymorphism in *Xanthomonas campestris* pv. *campestris* and its impact on elicitation of *Arabidopsis* FLAGELLIN SENSING2-dependent defenses. *Plant Cell* 18, 764–779. doi: 10.1105/tpc.105.037648
- Thieme, F., Koebnik, R., Bekel, T., Berger, C., Boch, J., Büttner, D., et al. (2005). Insights into genome plasticity and pathogenicity of the plant pathogenic bacterium *Xanthomonas campestris* pv. *vesicatoria* revealed by the complete genome sequence. *J. Bacteriol.* 187, 7254–7266. doi: 10.1128/JB.187.21.7254-7266.2005
- Tian, F., Yu, C., Li, H., Wu, X., Li, B., Chen, H., et al. (2014). Alternative sigma factor RpoN2 is required for flagellar motility and full virulence of *Xanthomonas oryzae* pv. *oryzae*. *Microbiol. Res.* 170, 177–183. doi: 10.1016/j.micres.2014.07.002
- Touchon, M., and Rocha, E. P. C. (2007). Causes of insertion sequences abundance in prokaryotic genomes. *Mol. Biol. Evol.* 24, 969–981. doi: 10.1093/molbev/msm014
- Triplett, L. R., Verdier, V., Campillo, T., Van Malderghem, C., Cleenwerck, I., Maes, M., et al. (2015). Characterization of a novel clade of *Xanthomonas* isolated from rice leaves in Mali and proposal of *Xanthomonas maliensis* sp. nov. *Antonie Van Leeuwenhoek* 107, 869–881. doi: 10.1007/s10482-015-0379-5
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., et al. (2012). Primer3 - new capabilities and interfaces. *Nucleic Acids Res.* 40, e115. doi: 10.1093/nar/gks596
- Vandroemme, J., Cottyn, B., Baeyen, S., De Vos, P., and Maes, M. (2013b). Draft genome sequence of *Xanthomonas fragariae* reveals reductive evolution

- and distinct virulence-related gene content. *BMC Genomics* 14:829. doi: 10.1186/1471-2164-14-829
- Vandroemme, J., Cottyn, B., Pothier, J. F., Pfleiderer, V., Duffy, B., and Maes, M. (2013a). *Xanthomonas arboricola* pv. *fragariae*: What's in a name? *Plant Pathol.* 62, 1123–1131. doi: 10.1111/ppa.12028
- Vauterin, L., Rademaker, J., and Swings, J. (2000). Synopsis on the Taxonomy of the Genus *Xanthomonas*. *Phytopathology* 90, 677–682. doi: 10.1094/PHYTO.2000.90.7.677
- Vauterin, L., Yang, P., Alvarez, A., Takikawa, Y., Roth, D. A., Vidaver, A. K., et al. (1996). Identification of nonpathogenic *Xanthomonas* strains associated with plants. *Syst. Appl. Microbiol.* 19, 96–105. doi: 10.1016/S0723-2020(96)80016-6
- Vladimirov, N., and Sourjik, V. (2009). Chemotaxis: how bacteria use memory. *Biol. Chem.* 390, 1097–1104. doi: 10.1515/bc.2009.130
- Voloudakis, A. E., Reignier, T. M., and Cooksey, D., a. (2005). Regulation of resistance to copper in *Xanthomonas axonopodis* pv. *vesicatoria*. *Appl. Environ. Microbiol.* 71, 782–789. doi: 10.1128/AEM.71.2.782-789.2005
- Vorhölter, F.-J., Wiggerich, H.-G., Scheidle, H., Sidhu, V. K., Mrozek, K., Küster, H., et al. (2012). Involvement of bacterial TonB-dependent signaling in the generation of an oligogalacturonide damage-associated molecular pattern from plant cell walls exposed to *Xanthomonas campestris* pv. *campestris* pectate lyases. *BMC Microbiol.* 12:239. doi: 10.1186/1471-2180-12-239
- Wang, Y., Wei, X., Bao, H., and Liu, S.-L. (2014). Prediction of bacterial type IV secreted effectors by C-terminal features. *BMC Genomics* 15:50. doi: 10.1186/1471-2164-15-50
- Xu, Z., and Hao, B. (2009). CVTree update: a newly designed phylogenetic study platform using composition vectors and whole genomes. *Nucleic Acids Res.* 37, 174–178. doi: 10.1093/nar/gkp278
- Yokoseki, T., Kutsukake, K., and Ohnishi, K. (1995). Functional analysis of the flagellar genes in the fliD operon of *Salmonella typhimurium*. *Microbiology* 141, 1715–1722. doi: 10.1099/13500872-141-7-1715
- Zerbino, D. R., and Birney, E. (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18, 821–829. doi: 10.1101/gr.074492.107
- Zevenhuizen, L. P. T. M., Dolfing, J., Eshuis, E. J., and Scholten-Koerselman, I. J. (1979). Inhibitory effects of copper on bacteria related to the free ion concentration. *Microb. Ecol.* 5, 139–146. doi: 10.1007/BF02010505
- Zou, L. F., Wang, X. P., Xiang, Y., Zhang, B., Li, Y. R., Xiao, Y. L., et al. (2006). Elucidation of the hrp clusters of *Xanthomonas oryzae* pv. *oryzicola* that control the hypersensitive response in nonhost tobacco and pathogenicity in susceptible host rice. *Appl. Environ. Microbiol.* 72, 6212–6224. doi: 10.1128/AEM.00511-06

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 Cesbron, Briand, Essakhi, Gironde, Boureau, Manceau, Fischer-Le Saux and Jacques. 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.

What makes *Xanthomonas albilineans* unique amongst xanthomonads?

Isabelle Pieretti¹, Alexander Pesic², Daniel Petras², Monique Royer¹, Roderich D. Süssmuth² and Stéphane Cociancich^{1*}

¹ UMR BGPI, Cirad, Montpellier, France, ² Institut für Chemie, Technische Universität Berlin, Berlin, Germany

OPEN ACCESS

Edited by:

Nicolas Denancé,
Institut National de la Recherche
Agronomique, France

Reviewed by:

Joao C. Setubal,
University of São Paulo, Brazil

Julian J. Smith,
Fera Science Ltd., UK

***Correspondence:**

Stéphane Cociancich,
UMR BGPI, Cirad, TA A-54/K,
Campus international de Baillarguet,
F-34398 Montpellier Cedex 5, France
stephane.cociancich@cirad.fr

Specialty section:

This article was submitted to
Plant-Microbe Interaction,
a section of the journal
Frontiers in Plant Science

Received: 24 February 2015

Accepted: 09 April 2015

Published: 24 April 2015

Citation:

Pieretti I, Pesic A, Petras D, Royer M, Süssmuth RD and Cociancich S (2015) What makes *Xanthomonas albilineans* unique amongst xanthomonads? *Front. Plant Sci.* 6:289. doi: 10.3389/fpls.2015.00289

Xanthomonas albilineans causes leaf scald, a lethal disease of sugarcane. Compared to other species of *Xanthomonas*, *X. albilineans* exhibits distinctive pathogenic mechanisms, ecology and taxonomy. Its genome, which has experienced significant erosion, has unique genomic features. It lacks two loci required for pathogenicity in other plant pathogenic species of *Xanthomonas*: the xanthan gum biosynthesis and the Hrp-T3SS (hypersensitive response and pathogenicity-type three secretion system) gene clusters. Instead, *X. albilineans* harbors in its genome an SPI-1 (*Salmonella* pathogenicity island-1) T3SS gene cluster usually found in animal pathogens. *X. albilineans* produces a potent DNA gyrase inhibitor called albicidin, which blocks chloroplast differentiation, resulting in the characteristic white foliar stripe symptoms. The antibacterial activity of albicidin also confers on *X. albilineans* a competitive advantage against rival bacteria during sugarcane colonization. Recent chemical studies have uncovered the unique structure of albicidin and allowed us to partially elucidate its fascinating biosynthesis apparatus, which involves an enigmatic hybrid PKS/NRPS (polyketide synthase/non-ribosomal peptide synthetase) machinery.

Keywords: *Xanthomonas albilineans*, leaf scald disease of sugarcane, genomic features, albicidin, NRPS and PKS genes

Introduction

Xanthomonas albilineans (Ashby) Dowson is known to invade the xylem of sugarcane and to cause leaf scald disease (Rott and Davis, 2000; Birch, 2001). Symptoms of this disease vary from a single, white, narrow, sharply defined stripe to complete wilting and necrosis of infected leaves, leading to plant death. Dissemination of *X. albilineans* occurs mainly mechanically through use of contaminated harvesting tools and by distribution and planting of infected cuttings. However, aerial transmission and potential for epiphytic survival have also been reported for this pathogen (Autrey et al., 1995; Daugrois et al., 2003; Champoiseau et al., 2009).

Xanthomonas albilineans is a representative of the genus *Xanthomonas*, members of which are exclusively Gram-negative plant-associated bacteria that collectively cause dramatic damage to hundreds of plant species of ornamental or agronomical interest. Indeed, both monocotyledonous (e.g., rice, sugarcane, or banana) and dicotyledonous (e.g., citrus, cauliflower, bean, pepper, cabbage, and tomato) plants are targeted worldwide by various *Xanthomonas* species. While sharing numerous phenotypic characteristics, at least 27 species and over 120 pathovars (variants of pathogenicity) of the genus *Xanthomonas* are currently recognized. Each pathovar individually exhibits a very

restricted host range and/or tissue-specificity and this leads to clustering of bacterial strains causing similar symptoms on the same host.

Multilocus sequence analysis (MLSA) with four housekeeping genes resulted in the distribution of *Xanthomonas* species in two clades. The main one contains the majority of species whereas the secondary clade contains *X. albilineans*, *Xanthomonas sacchari*, *Xanthomonas theicola*, *Xanthomonas hyacinthi*, and *Xanthomonas translucens* (Young et al., 2008). Phylogenetic analyses with the *gyrB* sequence indicate that this secondary group also contains several uncharacterized species of *Xanthomonas* isolated mainly on rice, banana or sugarcane (Studholme et al., 2011, 2012). Intriguingly, two multiMLSA studies with 28 genes and 228 genes, respectively, in which *X. albilineans* is the only representative of this secondary clade, resulted in the branching of *Xylella fastidiosa* between *X. albilineans* and the main clade (Rodriguez-R et al., 2012; Naushad and Gupta, 2013). *X. fastidiosa* is a xylem-limited bacterium which is insect-vectored to a variety of diverse hosts, has a reduced genome and lacks the Hrp-T3SS (hypersensitive response and pathogenicity-type III secretion system; Simpson et al., 2000).

Analysis of the *X. albilineans* genome has revealed unusual features compared to other xanthomonads, the most prominent being the absence of the Hrp-T3SS gene cluster and the occurrence of genome erosion. Furthermore, to our knowledge, *X. albilineans* is the only xanthomonad that produces the phytotoxin albicidin. This mini-review aims to summarize the characteristics that, taken together, make *X. albilineans* so unique.

Genome Erosion

The genome of *X. albilineans* strain GPE PC73 has been fully sequenced and annotated. It consists of a 3,768,695-bp circular chromosome with a G+C content of 63%, and three plasmids of 31,555-bp, 27,212-bp and 24,837-bp, respectively (Pieretti et al., 2009). This genome size is much smaller than that of any other xanthomonad sequenced to date (commonly ~5 Mb). Examination of the genome of strain GPE PC73 together with OrthoMCL comparative analyses performed with other sequenced xanthomonads highlights several genomic features that distinguish *X. albilineans* from its near relatives (Pieretti et al., 2009, 2012; Margueretaz et al., 2011; Royer et al., 2013).

Orthologous analyses show that *X. albilineans* and *X. fastidiosa* have experienced a convergent genome reduction during their respective speciation, with a more extensive genome reduction for *X. fastidiosa* (Pieretti et al., 2009). Based on these analyses, *X. albilineans* has lost at least 592 genes that were present in the last common ancestor of the xanthomonads. Interestingly, most of these ancestral genes are conserved in the genome of *X. sacchari* strains NCPPB4393 and LMG 476 and *Xanthomonas* spp. strains NCPPB1131 and NCPPB1132, which are the sequenced strains phylogenetically closest to *X. albilineans* (Studholme et al., 2011, 2012; Pieretti et al., 2015). This indicates that genome erosion is specific to *X. albilineans*. Convergent genome erosion of *X. albilineans* and *X. fastidiosa* could be linked to a similar adaptation to a xylem-invading lifestyle in which interactions with living plant tissues are minimal (Pieretti et al., 2009). More recently, a

study of the somewhat reduced genome of *Xanthomonas fragariae* (4.2 Mb) led to the hypothesis that the convergent genome reduction observed in some xanthomonads could be linked to their endophytic lifestyle and typically to their commitment to a single host (Vandroemme et al., 2013).

Compared to other xanthomonads, a low number of insertion sequences (IS) has been found in the genome of *X. albilineans*. Taken together with a limited recombination of the chromosome and a GC skew pattern containing a low number of distortions, it was postulated that genome erosion of *X. albilineans* was mainly not due to IS and other mechanisms were proposed for this erosion (Pieretti et al., 2009). The low number of IS could be linked to the activity of CRISPR (clustered regularly interspaced short palindromic repeats) systems. Strain GPE PC73 of *X. albilineans* possesses two CRISPR loci. The first one, CRISPR-1, is conserved in *X. oryzae* pv. *oryzae*, *X. axonopodis* pv. *citri*, *X. campestris* pv. *vasculorum*, and *X. campestris* pv. *musacearum*. The second, CRISPR-2, is present in *X. campestris* pv. *raphani* (Pieretti et al., 2012). Interestingly, many spacers of CRISPR-1 and CRISPR-2 of strain GPE PC73 are identical to IS or phage-related DNA sequences present on the chromosome of this strain (Pieretti et al., 2012).

Specific Genes Linked to a Xylem-Invading Lifestyle

Although determinants for host- or tissue-specificity of *X. albilineans* remain unclear, the presence in its genome of genes encoding cell-wall-degrading enzymes (CWDEs) with specific features is probably important for its ability to spread in xylem and for pathogenicity. Indeed, all CWDEs from *X. albilineans* harbor a cellulose-binding domain (CBD) and a long linker region both adapted to the utilization of cell-wall breakdown products as carbon source and to the ability to spread in sugarcane xylem vessels (Pieretti et al., 2012). These enzymes may also be required to disrupt pit membranes in sugarcane, thereby promoting propagation of the bacteria in the plant. Interestingly, *X. fastidiosa* also encodes two CWDEs containing a long linker and a CBD. It has been shown that one of these two CWDEs is involved in the spread of *X. fastidiosa* in the xylem by increasing the pore size of pit membranes. CWDEs are therefore considered as virulence factors (Roper et al., 2007; Chatterjee et al., 2008; Pérez-Donoso et al., 2010). TonB-dependent transporters (TBDTs) may be used by *X. albilineans* to transport cell-wall-degrading products resulting from the activity of CWDEs, and thus may facilitate spread of the organism in the nutrient-poor conditions prevailing in the xylem of sugarcane. In the genome of *X. albilineans*, 35 TBDT genes have been identified, including one specific to this species and two others that are functionally associated to pathogenicity of the bacterium (Rott et al., 2011; Pieretti et al., 2012).

Lack of Hrp-T3SS

Most phytopathogenic bacteria rely on the type III secretion system (T3SS) of the hypersensitive response and pathogenicity family (Hrp1 and Hrp2, respectively). This syringe-like apparatus allows pathogens to deliver, into their host cells, proteins (type

III effectors) that modulate plant physiology and immunity for the benefit of the pathogen. Interestingly, genes encoding the injectisome and associated effectors of the Hrp-T3SS are missing in the genome of *X. albilineans*, as is also the case in the genomes of *X. sacchari* strains NCPPB4393 and LMG 476 and *Xanthomonas* spp. strains NCPPB1131 and NCPPB1132 (Studholme et al., 2011, 2012; Pieretti et al., 2015). Yet, an Hrp system is present in other close neighbor species of *X. albilineans*, such as *X. translucens* pv. *graminis* strain 29, *X. translucens* pv. *translucens* strain DSM18974, and *X. translucens* strain DAR 61454 (Wichmann et al., 2013; Gardiner et al., 2014). Although the Hrp-T3SS is described as a crucial key component in plant–host interactions for most *Xanthomonas* spp., it seems not to be essential in *X. translucens* pv. *graminis* strain 29 for xylem colonization, even though it is involved in symptom development (Ryan et al., 2011; Wichmann et al., 2013). Similarly, despite being devoid of any Hrp T3SS, *X. albilineans* displays pathogenicity and is able to cause serious damage to sugarcane.

Acquisition of a SPI-1 T3SS

The annotated sequence of the genome of *X. albilineans* strain GPE PC73 reveals the presence of a T3SS belonging to the *Salmonella* pathogenicity island-1 (SPI-1) injectisome family. Genes encoding this system are located near the terminus of the replication site of the chromosome and were probably acquired by lateral gene transfer. This secretion system, found mainly in mammals and insects bacterial pathogens or symbionts, exhibits high similarity to that described in *Burkholderia pseudomallei*—a human pathogen causing melioidosis (Stevens et al., 2002). The SPI-1 needle-like assemblies of *X. albilineans* strain GPE PC73 and *B. pseudomallei* strain K96243 are homologous. Both species share all but two genes—*orgA* and *orgB*, encoding putative oxygen-regulated invasion proteins involved in type three secretion that are not conserved in *B. pseudomallei*. The genome composition of the SPI-1 T3SS in *X. albilineans* additionally includes genes encoding translocon components (*xipB*, *xipC*, and *xipD*), injectisome components (*xsaJ* to *xsaS* and *xsaV* to *xsaZ*) and a chaperone (*xicA*). Furthermore, the locus contains 15 additional genes referred to as *xapA*–*xapO*, encoding hypothetical proteins. These genes, which show homology neither to sequences from *B. pseudomallei* nor to sequences available from protein sequence databases, are specific to *X. albilineans* and their products represent good candidates to be considered as effectors for this SPI-1 T3SS (Margueret et al., 2011). Interestingly, this SPI-1 T3SS is conserved in *Xanthomonas axonopodis* pv. *phaseoli* strains CFBP 2534, CFBP 6164 and CFBP 6982, which moreover possess a second T3SS belonging to the Hrp2 family (Alavi et al., 2008; Margueret et al., 2011). Pathogenicity of *X. albilineans* strains seems not to be linked to the presence of the SPI-1 T3SS in their genome; besides, no SPI-1 T3SS locus has been identified in strain PNG130 of *X. albilineans* even though it is able to spread in sugarcane. Functional analyses showed that, *in planta*, multiplication of a SPI-1 T3SS knockout mutant of *X. albilineans* was not impaired when compared to the wild-type, indicating that the SPI-1 T3SS is not required for spread in sugarcane vessels or for development of leaf scald symptoms. The role of

the SPI-1 T3SS of *X. albilineans* remains unclear, although it has been conserved during its evolution in *X. albilineans* without frame-shifting indels or nonsense mutations (Margueret et al., 2011). It remains possible, in conditions other than those tested with our knockout mutant, that the SPI-1 T3SS system may be required for interaction with sugarcane, as in the case of SPI-1 of *Salmonella*, which is involved in interactions with *Arabidopsis thaliana* (Schikora et al., 2011). The SPI-1 T3SS system may also be associated with other aspects of the *X. albilineans* lifestyle, e.g., an involvement in adherence as reported for *Erwinia tasmaniensis* (Kube et al., 2008) or in formation of pellicle or biofilm-like structures (Jennings et al., 2012), which could be related to epiphytic survival on sugarcane leaves. Although no insect vector has been identified for *X. albilineans* to date, we cannot rule out that the SPI-1 T3SS could be involved in insect association or might mediate persistence of the bacterium in an insect vector as was shown for *Pantoea stewartii* (Correa et al., 2012).

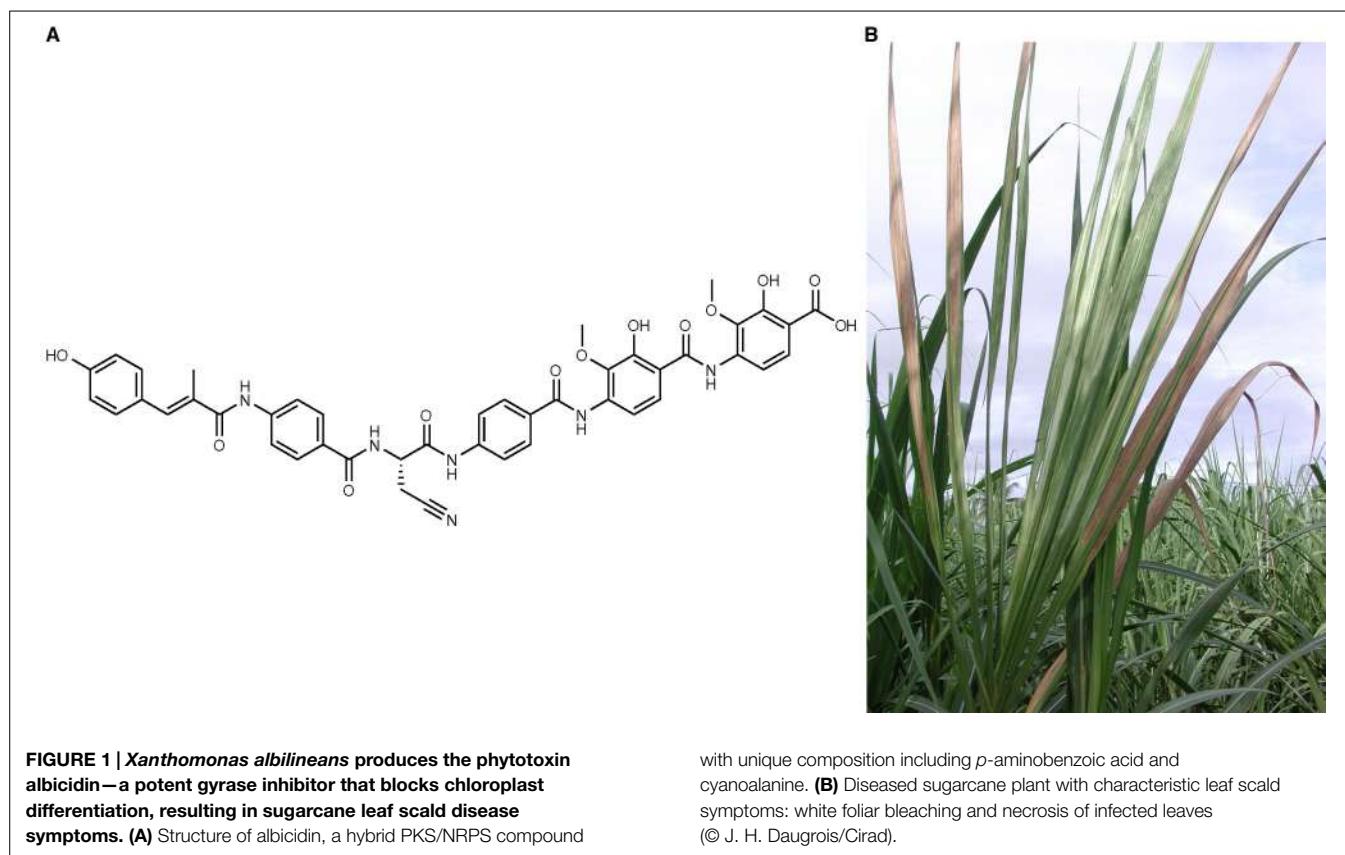
Lack of T6SS and the Xanthan Gum Gene Cluster

Xanthomonas albilineans lacks two other major pathogenicity factors that are common features of most xanthomonads. First, it lacks the gum gene cluster for extracellular polysaccharide (EPS) synthesis. This gene cluster is responsible for biofilm and xanthan gum formation, and is associated with pathogenesis in xanthomonads (Katzen et al., 1998; Kim et al., 2009; Galván et al., 2012). Exceptions are *X. fragariae*, which lacks the *gumN*, *gumo* and *gumP* genes, and *X. albilineans*, which lacks the complete set of gum genes, indicating those are not essential for virulence of both these pathogens (Pieretti et al., 2012; Vandromme et al., 2013).

Xanthomonas albilineans is also devoid of any type VI secretion system (T6SS) described in other xanthomonads, as for example in *Xanthomonas fuscans* pv. *fuscans* strain 4834-R and *Xanthomonas citri* subsp. *citri* strain 306, which each contain a single T6SS (Potnis et al., 2011; Darrasse et al., 2013) or *X. translucens* strain DAR61454, which encodes two distinct T6SS (Gardiner et al., 2014). Structurally, the T6SS looks like an inverted bacteriophage. Functionally, this system is able to interact with both eukaryotic and prokaryotic cells by delivering effectors or toxins into host cells to subvert the signaling process to its own advantage, but also into other bacteria from the same habitat to outcompete them during infection (Filloux, 2013; Russell et al., 2014). Despite its multifunctional roles during host–pathogen interactions, the lack of T6SS in *Xanthomonas campestris* pv. *campestris* strain 8004, *Xanthomonas gardneri* strain 101, and *X. albilineans* seems to have no effect on pathogenesis of these xanthomonads.

Albicidin and Other Non-Ribosomally Synthesized Peptides

A unique feature of *X. albilineans* is the production of albicidin—a phytotoxin causing the white foliar stripe symptoms



characteristic of leaf scald disease of sugarcane (Birch and Patil, 1985). Albicidin is a potent DNA gyrase inhibitor that blocks the differentiation of chloroplasts (Figure 1). It also targets bacterial gyrase by a mechanism different from that of other DNA gyrase inhibitors like coumarins and quinolones (Hashimi et al., 2007). This mode of action accounts for the potent antibacterial activity of albicidin, which inhibits the growth of Gram-positive and Gram-negative pathogenic bacteria at nanomolar concentrations (Birch and Patil, 1985). Albicidin gives a competitive advantage to *X. albilineans* against other bacteria within the xylem vessels of sugarcane (Magnani et al., 2013). Interestingly, two sugarcane-living bacteria harbor an albicidin resistance gene: *Leifsonia xyli* (Monteiro-Vitorello et al., 2004) and *Pantoea dispersa* (Zhang and Birch, 1997).

Albicidin is produced by a hybrid polyketide synthase (PKS)/non-ribosomal peptide synthetase (NRPS) enzyme complex. PKS and NRPS genes are often clustered together with a large set of regulatory, transport or modification (tailoring) genes, as well as genes involved in the biosynthesis of non-proteinogenic amino acids. In addition to a phosphopantetheinyl transferase required for activation of the PKS/NRPS system and a HtpG chaperone, the role of which remains unclear, a locus (*alb* cluster) containing 20 genes is required for albicidin biosynthesis. Among these 20 genes, 3 encode the PKS/NRPS system; 15 others act as transport, regulatory, modification or resistance genes (Royer et al., 2004).

Non-ribosomal peptide synthetases are multimodular megasynthetases used by bacteria and fungi to produce peptides in a

ribosome-independent manner (Stricker et al., 2010). Each module governs the specific incorporation of an amino acid substrate based on signature sequences in the adenylation (A) domains (Stachelhaus and Marahiel, 1995), which are loaded onto peptidyl carrier protein (PCP) domains. Elongation of the peptide is mediated by condensation (C) domains present within each module. PKSs function according to the principles of fatty acid biosynthesis (Weissman and Leadlay, 2005).

For decades, the structure elucidation of albicidin was impeded by its extremely low production yield by *X. albilineans*. A first step to overcome this bottleneck was achieved by transferring the biosynthetic genes into a heterologous host, namely *X. axonopodis* pv. *vesicatoria*, resulting in a significant increase in albicidin production (Vivien et al., 2007). Extensive HPLC purification of albicidin and thorough analysis of the purified compound by means of mass spectrometry and nuclear magnetic resonance spectroscopy then allowed us to unravel its unique structure (Figure 1). Albicidin proved to be a linear pentapeptide composed of cyanoalanine and *p*-amino benzoic acids N-terminally linked to a *p*-coumaric acid derivative (Cocianich et al., 2015). Although over 500 different monomers (amino acid substrates) have been identified to date as being incorporated by NRPS systems, elucidation of the structure of albicidin revealed for the first time the incorporation by NRPSs of cyanoalanine and *p*-amino benzoic acids. Moreover, the incorporation of *p*-amino benzoic acids is the first example of incorporation of a δ -aminoacid by NRPSs, since all NRPSs described to date incorporate only α or β aminoacids. The use of unusual amino acid substrates is linked to unique

features that were identified *in silico* 10 years ago within the albicidin NRPS modules sequence (Royer et al., 2004). The formation and incorporation of cyanoalanine most likely occurs *in situ* through an additional module present in the PKS-NRPS assembly line that was investigated in one of our present studies (Cociancich et al., 2015).

Chemical synthesis of albicidin is now available, allowing both production of high quantities of the compound for further study of its mode of action and activity spectrum, and the synthesis of analogs (Kretz et al., 2015). The uniqueness of its structure and the specific mode of action of this compound make albicidin a strong lead structure for antibiotic development.

Data mining of the genome of *X. albilineans* strain GPE PC73 has led to the identification, in addition to the albicidin biosynthesis locus, of five other NRPS loci (Pieretti et al., 2012; Royer et al., 2013). The first, named Meta-B, encodes megasynthases performing peptidic elongation of a 16-amino acid lipopeptide. This locus also encodes a transcription regulator belonging to the AraC family, a cyclic peptide transporter, and enzymes involved in biosynthesis of the non-proteinogenic amino acids di-amino butyric acid and dihydroxyphenylglycine. Interestingly, the NRPS locus Meta-B has been identified in the genome of strains of three other *Xanthomonas* species, namely *Xanthomonas oryzae* pv. *oryzae* strains BAI3 and X11-5A, *X. translucens* strain DAR61454 and *Xanthomonas* spp. strain XaS3 (Royer et al., 2013). Despite a similar organization of the genes within these loci, the *in silico* prediction of the sequences of the peptides produced indicates that each strain produces a different lipopeptide.

Two other NRPS gene clusters, Meta-A and Meta-C, have been identified in the genome of *X. albilineans* strain GPE PC73. They encode megasynthases that perform the biosynthesis of peptides of 12 and 7 amino acids, respectively. A partial sequence has been predicted for each of these peptides (Royer et al., 2013).

Finally, two short NRPS genes have also been identified on the chromosome of *X. albilineans*: they both encode only one NRPS module. Interestingly, there is an overlap between both these genes and a gene encoding a glycosyltransferase. It has been

hypothesized that these genes encode glycosylated amino acids, to which, however, no precise function could yet be attributed (Royer et al., 2013).

Conclusion

Although most xanthomonads require pathogenicity factors such as *gum* genes, T3SS Hrp and T6SS for survival, growth and spread within host plants, *X. albilineans* lacks these pathogenicity factors, *de facto* reducing its artillery to circumvent sugarcane defense mechanisms and innate immunity. While being “disarmed” could be disadvantageous for a vascular plant pathogen, *X. albilineans* remains able to invade and spread in sugarcane, suggesting that it uses other strategies, such as stealth, i.e., being unobtrusive *in planta*, to minimize inducible host defense responses. On the other hand, the reduced genome of *X. albilineans* has specific features that may be involved in the adaptation of the bacterium to live and spread in sugarcane xylem vessels. For example, specific CWDEs and TBDTs appear to be optimized for life in the nutrient-poor sugarcane xylem environment. The uniqueness of *X. albilineans* resides also in the production of the phytotoxin and antibiotic albicidin. The recently unraveled structure and concomitant development of a chemical synthesis protocol for this compound leads to additional prospects for its use in the antibiotherapy field. According to the specificities deriving from the biological, biochemical, phylogenetic and genomic analyses described in this review, one can truly say that *X. albilineans* is quite unique amongst the genus *Xanthomonas*.

Acknowledgments

Work on albicidin was supported by a grant from the Deutsche Forschungsgemeinschaft (DFG SU239/11-1; SU 18-1), by the Cluster of Excellence “Unifying Concepts in Catalysis (UniCat)” (DFG) and by a grant from the Agence Nationale de la Recherche (ANR-09-BLAN-0413-01). The authors are indebted to Helen Rothnie for English editing.

References

- Alavi, S. M., Sanjari, S., Durand, F., Brin, C., Manceau, C., and Poussier, S. (2008). Assessment of the genetic diversity of *Xanthomonas axonopodis* pv. *phaseoli* and *Xanthomonas fuscans* subsp. *fuscans* as a basis to identify putative pathogenicity genes and a type III secretion system of the SPI-1 family by multiple suppression subtractive hybridizations. *Appl. Environ. Microbiol.* 74, 3295–3301. doi: 10.1128/aem.02507-07
- Autrey, L., Saumtally, S., Dookun, A., Sullivan, S., and Dhayan, S. (1995). Aerial transmission of the leaf scald pathogen, *Xanthomonas albilineans*. *Proc. Int. Soc. Sug. Cane Technol.* 21, 508–526.
- Birch, R. G. (2001). *Xanthomonas albilineans* and the antipathogenesis approach to disease control. *Mol. Plant Pathol.* 2, 1–11. doi: 10.1046/j.1364-3703.2001.00046.x
- Birch, R. G., and Patil, S. S. (1985). Preliminary characterization of an antibiotic produced by *Xanthomonas albilineans* which inhibits DNA synthesis in *Escherichia coli*. *J. Gen. Microbiol.* 131, 1069–1075. doi: 10.1099/00221287-131-5-1069
- Champosieau, P., Rott, P., and Daugrois, J. H. (2009). Epiphytic populations of *Xanthomonas albilineans* and subsequent sugarcane stalk infection are linked to rainfall in Guadeloupe. *Plant Dis.* 93, 339–346. doi: 10.1094/pdis-93-4-0339
- Chatterjee, S., Almeida, R. P. P., and Lindow, S. (2008). Living in two worlds: the plant and insect lifestyles of *Xylella fastidiosa*. *Annu. Rev. Phytopathol.* 46, 243–271. doi: 10.1146/annurev.phyto.45.062806.094342
- Cociancich, S., Pesic, A., Petras, D., Uhlmann, S., Kretz, J., Schubert, V., et al. (2015). The gyrase inhibitor albicidin consists of p-aminobenzoic acids and cyanoalanine. *Nat. Chem. Biol.* 11, 195–197. doi: 10.1038/nchembio.1734
- Correa, V. R., Majerczak, D. R., Ammar, E.-D., Merighi, M., Pratt, R. C., Hogenhout, S. A., et al. (2012). The bacterium *Pantoea stewartii* uses two different type III secretion systems to colonize its plant host and insect vector. *Appl. Environ. Microbiol.* 78, 6327–6336. doi: 10.1128/aem.00892-12
- Darrasse, A., Carrere, S., Barbe, V., Bourreau, T., Arrieta-Ortiz, M., Bonneau, S., et al. (2013). Genome sequence of *Xanthomonas fuscans* subsp. *fuscans* strain 4834-R reveals that flagellar motility is not a general feature of xanthomonads. *BMC Genom.* 14:761. doi: 10.1186/1471-2164-14-761
- Daugrois, J. H., Dumont, V., Champosieau, P., Costet, L., Boisne-Noc, R., and Rott, P. (2003). Aerial contamination of sugarcane in Guadeloupe by two strains of *Xanthomonas albilineans*. *Eur. J. Plant Pathol.* 109, 445–458. doi: 10.1023/a:1024259606468

- Filloux, A. (2013). The rise of the type VI secretion system. *F1000Prime Rep.* 5:52. doi: 10.12703/p5-52
- Galván, E. M., Ielmini, M. V., Patel, Y. N., Bianco, M. I., Franceschini, E. A., Schneider, J. C., et al. (2012). Xanthan chain length is modulated by increasing the availability of the polysaccharide copolymerase protein GumC and the outer membrane polysaccharide export protein GumB. *Glycobiology* 23, 259–272. doi: 10.1093/glycob/cws146
- Gardiner, D. M., Upadhyaya, N. M., Stiller, J., Ellis, J. G., Dodds, P. N., Kazan, K., et al. (2014). Genomic analysis of *Xanthomonas translucens* pathogenic on wheat and barley reveals cross-kingdom gene transfer events and diverse protein delivery systems. *PLoS ONE* 9:e84995. doi: 10.1371/journal.pone.0084995
- Hashimi, S. M., Wall, M. K., Smith, A. B., Maxwell, A., and Birch, R. G. (2007). The phytotoxin albicidin is a novel inhibitor of DNA gyrase. *Antimicrob. Agents Chemother.* 51, 181–187. doi: 10.1128/aac.00918-06
- Jennings, M. E., Quick, L. N., Ubol, N., Shrom, S., Dollahon, N., and Wilson, J. W. (2012). Characterization of *Salmonella* type III secretion hyper-activity which results in biofilm-like cell aggregation. *PLoS ONE* 7:e33080. doi: 10.1371/journal.pone.0033080
- Katzen, F., Ferreiro, D. U., Oddo, C. G., Ielmini, M. V., Becker, A., Pühler, A., et al. (1998). *Xanthomonas campestris* pv. *campestris* gum mutants: effects on xanthan biosynthesis and plant virulence. *J. Bacteriol.* 180, 1607–1617.
- Kim, J.-G., Li, X., Roden, J. A., Taylor, K. W., Aakre, C. D., Su, B., et al. (2009). *Xanthomonas* T3S effector XopN suppresses PAMP-triggered immunity and interacts with a tomato atypical receptor-like kinase and TFT1. *Plant Cell* 21, 1305–1323. doi: 10.1105/tpc.108.063123
- Kretz, J., Kerwat, D., Schubert, V., Grätz, S., Pesic, A., Semsary, S., et al. (2015). Total synthesis of albicidin: a lead structure from *Xanthomonas* *albilineans* for potent antibacterial gyrase inhibitors. *Angew. Chem. Int. Ed.* 54, 1969–1973. doi: 10.1002/anie.201409584
- Kube, M., Migdall, A. M., Müller, I., Kuhl, H., Beck, A., Reinhardt, R., et al. (2008). The genome of *Erwinia tasmaniensis* strain Et1/99, a non-pathogenic bacterium in the genus *Erwinia*. *Environ. Microbiol.* 10, 2211–2222. doi: 10.1111/j.1462-2920.2008.01639.x
- Magnani, G. S., Cruz, L. M., Weber, H., Bespalhok, J. C., Daros, E., Baura, V., et al. (2013). Culture-independent analysis of endophytic bacterial communities associated with Brazilian sugarcane. *Genet. Mol. Res.* 12, 4549–4558. doi: 10.4238/2013.October.15.3
- Marguerettaz, M., Pieretti, I., Gayral, P., Puig, J., Brin, C., Cociancich, S., et al. (2011). Genomic and evolutionary features of the SPI-1 type III secretion system that is present in *Xanthomonas* *albilineans* but is not essential for xylem colonization and symptom development of sugarcane leaf scald. *Mol. Plant Microbe Interact.* 24, 246–259. doi: 10.1094/mpmi-08-10-0188
- Monteiro-Vitorello, C. B., Camargo, L. A., Van Sluys, M. A., Kitajima, J. P., Truffi, D., do Amaral, A. M., et al. (2004). The genome sequence of the gram-positive sugarcane pathogen *Leifsonia xyli* subsp. *xyli*. *Mol. Plant Microbe Interact.* 17, 827–836. doi: 10.1094/mpmi.2004.17.8.827
- Naushad, H. S., and Gupta, R. S. (2013). Phylogenomics and molecular signatures for species from the plant pathogen-containing order Xanthomonadales. *PLoS ONE* 8:e55216. doi: 10.1371/journal.pone.0055216
- Pérez-Donoso, A. G., Sun, Q., Roper, M. C., Greve, L. C., Kirkpatrick, B., and Labavitch, J. M. (2010). Cell wall-degrading enzymes enlarge the pore size of intervesSEL pit membranes in healthy and *Xylella fastidiosa*-infected grapevines. *Plant Physiol.* 152, 1748–1759. doi: 10.1104/pp.109.148791
- Pieretti, I., Bolot, S., Carrère, S., Barbe, V., Cociancich, S., Rott, P., et al. (2015). Draft genome sequence of *Xanthomonas sacchari* LMG 476. *Genome Announc.* 3, e00146-15. doi: 10.1128/genomeA.00146-15
- Pieretti, I., Royer, M., Barbe, V., Carrere, S., Koebnik, R., Cociancich, S., et al. (2009). The complete genome sequence of *Xanthomonas* *albilineans* provides new insights into the reductive genome evolution of the xylem-limited Xanthomonadaceae. *BMC Genom.* 10:616. doi: 10.1186/1471-2164-10-616
- Pieretti, I., Royer, M., Barbe, V., Carrere, S., Koebnik, R., Couloux, A., et al. (2012). Genomic insights into strategies used by *Xanthomonas* *albilineans* with its reduced artillery to spread within sugarcane xylem vessels. *BMC Genom.* 13:658. doi: 10.1186/1471-2164-13-658
- Potnis, N., Krasileva, K., Chow, V., Almeida, N., Patil, P., Ryan, R., et al. (2011). Comparative genomics reveals diversity among xanthomonads infecting tomato and pepper. *BMC Genom.* 12:146. doi: 10.1186/1471-2164-12-146
- Rodríguez-R, L., Grajales, A., Arrieta-Ortiz, M., Salazar, C., Restrepo, S., and Bernal, A. (2012). Genomes-based phylogeny of the genus *Xanthomonas*. *BMC Microbiol.* 12:43. doi: 10.1186/1471-2180-12-43
- Roper, M. C., Greve, L. C., Warren, J. G., Labavitch, J. M., and Kirkpatrick, B. C. (2007). *Xylella fastidiosa* requires polygalacturonase for colonization and pathogenicity in *Vitis vinifera* grapevines. *Mol. Plant Microbe Interact.* 20, 411–419. doi: 10.1094/mpmi-20-4-0411
- Rott, P., and Davis, M. (2000). “Leaf scald,” in *A Guide to Sugarcane Diseases*, eds P. Rott, R. Bailey, J. Comstock, B. Croft, and A. Saumtally (Montpellier: CIRAD-ISSCT press), 339.
- Rott, P., Fleites, L., Marlow, G., Royer, M., and Gabriel, D. W. (2011). Identification of new candidate pathogenicity factors in the xylem-invading pathogen *Xanthomonas* *albilineans* by transposon mutagenesis. *Mol. Plant Microbe Interact.* 24, 594–605. doi: 10.1094/mpmi-07-10-0156
- Royer, M., Costet, L., Vivien, E., Bes, M., Cousin, A., Damais, A., et al. (2004). Albicidin pathotoxin produced by *Xanthomonas* *albilineans* is encoded by three large PKS and NRPS genes present in a gene cluster also containing several putative modifying, regulatory, and resistance genes. *Mol. Plant Microbe Interact.* 17, 414–427. doi: 10.1094/mpmi.2004.17.4.414
- Royer, M., Koebnik, R., Marguerettaz, M., Barbe, V., Robin, G., Brin, C., et al. (2013). Genome mining reveals the genus *Xanthomonas* to be a promising reservoir for new bioactive non-ribosomally synthesized peptides. *BMC Genom.* 14:658. doi: 10.1186/1471-2164-14-658
- Russell, A. B., Peterson, S. B., and Mougous, J. D. (2014). Type VI secretion effectors: poisons with a purpose. *Nat. Rev. Microbiol.* 12, 137–148. doi: 10.1038/nrmicro3185
- Ryan, R. P., Vorhölter, F.-J., Potnis, N., Jones, J. B., Van Sluys, M.-A., Bogdanove, A. J., et al. (2011). Pathogenomics of *Xanthomonas*: understanding bacterium-plant interactions. *Nat. Rev. Microbiol.* 9, 344–355. doi: 10.1038/nrmicro2558
- Schikora, A., Virlogeux-Payant, I., Bueso, E., Garcia, A. V., Nilau, T., Charrier, A., et al. (2011). Conservation of *Salmonella* infection mechanisms in plants and animals. *PLoS ONE* 6:e24112. doi: 10.1371/journal.pone.0024112
- Simpson, A., Reinach, F., and Arruda, P. (2000). The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature* 406, 151–157. doi: 10.1038/35018003
- Stachelhaus, T., and Marahiel, M. A. (1995). Modular structure of genes encoding multifunctional peptide synthetases required for non-ribosomal peptide synthesis. *FEMS Microbiol. Lett.* 125, 3–14. doi: 10.1111/j.1574-6968.1995.tb07328.x
- Stevens, M. P., Wood, M. W., Taylor, L. A., Monaghan, P., Hawes, P., Jones, P. W., et al. (2002). An Inv/Mxi-Spa-like type III protein secretion system in *Burkholderia pseudomallei* modulates intracellular behaviour of the pathogen. *Mol. Microbiol.* 46, 649–659. doi: 10.1046/j.1365-2958.2002.03190.x
- Stricker, M., Tanovic, A., and Marahiel, M. A. (2010). Non-ribosomal peptide synthetases: structures and dynamics. *Curr. Opin. Struct. Biol.* 20, 234–240. doi: 10.1016/j.sbi.2010.01.009
- Studholme, D. J., Wasukira, A., Paszkiewicz, K., Aritua, V., Thwaites, R., Smith, J., et al. (2011). Draft genome sequences of *Xanthomonas* *sacchari* and two banana-associated xanthomonads reveal insights into the *Xanthomonas* Group 1 Clade. *Genes* 2, 1050–1065. doi: 10.3390/genes2010050
- Studholme, D. J., Wasukira, A., Paszkiewicz, K., Aritua, V., Thwaites, R., Smith, J., et al. (2012). Correction: Studholme et al., Draft genome sequences of *Xanthomonas* *sacchari* and two banana-associated xanthomonads reveal insights into the *Xanthomonas* Group 1 Clade. *Genes* 2011, 2, 1050–1065. *Genes* 3, 88–89. doi: 10.3390/genes3010088
- Vandromme, J., Cottyn, B., Baeyen, S., De Vos, P., and Maes, M. (2013). Draft genome sequence of *Xanthomonas fragariae* reveals reductive evolution and distinct virulence-related gene content. *BMC Genom.* 14:829. doi: 10.1186/1471-2164-14-829
- Vivien, E., Pitorre, D., Cociancich, S., Pieretti, I., Gabriel, D. W., Rott, P. C., et al. (2007). Heterologous production of albicidin: a promising approach to overproducing and characterizing this potent inhibitor of DNA gyrase. *Antimicrob. Agents Chemother.* 51, 1549–1552. doi: 10.1128/aac.01450-06
- Weissman, K. J., and Leadlay, P. F. (2005). Combinatorial biosynthesis of reduced polyketides. *Nat. Rev. Micro.* 3, 925–936. doi: 10.1038/nrmicro1287

Wichmann, F., Vorhölter, F.-J., Hersemann, L., Widmer, F., Blom, J., Niehaus, K., et al. (2013). The noncanonical type III secretion system of *Xanthomonas translucens* pv. *graminis* is essential for forage grass infection. *Mol. Plant Pathol.* 14, 576–588. doi: 10.1111/mpp.12030

Young, J. M., Park, D. C., Shearman, H. M., and Fargier, E. (2008). A multilocus sequence analysis of the genus *Xanthomonas*. *Syst. Appl. Microbiol.* 31, 366–377. doi: 10.1016/j.syapm.2008.06.004

Zhang, L., and Birch, R. (1997). The gene for albidin detoxification from *Pantoea dispersa* encodes an esterase and attenuates pathogenicity of *Xanthomonas albilineans*. *Proc. Natl. Acad. Sci. U.S.A.* 94, 9984–9989. doi: 10.1073/pnas.94.18.9984

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 Pieretti, Pesic, Petras, Royer, Süssmuth and Cociancich. 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.



Genome sequencing reveals a new lineage associated with lablab bean and genetic exchange between *Xanthomonas axonopodis* pv. *phaseoli* and *Xanthomonas fuscans* subsp. *fuscans*

Valente Aritua¹, James Harrison², Melanie Sapp³, Robin Buruchara⁴, Julian Smith³ and David J. Studholme^{2*}

¹ International Center for Tropical Agriculture, Kampala, Uganda, ² Biosciences, University of Exeter, Exeter, UK, ³ Fera Science Ltd., York, UK, ⁴ Africa Regional Office, International Center for Tropical Agriculture, Consultative Group for International Agricultural Research (CGIAR), Nairobi, Kenya

OPEN ACCESS

Edited by:

Nicolas Denancé,
Institut National de la Recherche
Agronomique, France

Reviewed by:

Dennis Gross,
Texas A&M University, USA
Marie-Agnès Jacques,
Institut National de la Recherche
Agronomique, France

*Correspondence:

David J. Studholme,
Biosciences, University of Exeter,
Stocker Road, Exeter EX4 4QD, UK
d.j.studholme@exeter.ac.uk

Specialty section:

This article was submitted to
Plant Biotic Interactions,
a section of the journal
Frontiers in Microbiology

Received: 30 May 2015

Accepted: 22 September 2015

Published: 07 October 2015

Citation:

Aritua V, Harrison J, Sapp M,
Buruchara R, Smith J and
Studholme DJ (2015) Genome
sequencing reveals a new lineage
associated with lablab bean and
genetic exchange between
Xanthomonas axonopodis pv. *phaseoli*
and *Xanthomonas fuscans* subsp.
fuscans. *Front. Microbiol.* 6:1080.
doi: 10.3389/fmicb.2015.01080

Common bacterial blight is a devastating seed-borne disease of common beans that also occurs on other legume species including lablab and Lima beans. We sequenced and analyzed the genomes of 26 strains of *Xanthomonas axonopodis* pv. *phaseoli* and *X. fuscans* subsp. *fuscans*, the causative agents of this disease, collected over four decades and six continents. This revealed considerable genetic variation within both taxa, encompassing both single-nucleotide variants and differences in gene content, that could be exploited for tracking pathogen spread. The bacterial strain from Lima bean fell within the previously described Genetic Lineage 1, along with the pathovar type strain (NCPPB 3035). The strains from lablab represent a new, previously unknown genetic lineage closely related to strains of *X. axonopodis* pv. *glycines*. Finally, we identified more than 100 genes that appear to have been recently acquired by *Xanthomonas axonopodis* pv. *phaseoli* from *X. fuscans* subsp. *fuscans*.

Keywords: beans, *Phaseolus vulgaris*, *Phaseolus lunatus*, *Lablab purpureus*, *Dolichos lablab*, *Xanthomonas fuscans*, *Xanthomonas axonopodis*

Introduction

Common bacterial blight (CBB) is a devastating, widespread and seed-borne disease of common beans (*Phaseolus vulgaris*). The bacteria that cause CBB are genetically diverse (Gilbertson et al., 1991; Alavi et al., 2008; Parkinson et al., 2009; Fourie and Herselman, 2011) and include fuscous strains, which produce a brown pigment on tyrosine-containing medium, and non-fuscous strains. Currently, the non-fuscous strains are classified as *Xanthomonas axonopodis* pv. *phaseoli* (*Xap*) while the fuscous strains are classified into a different species as *X. fuscans* subsp. *fuscans* (*Xff*) (Schaad et al., 2005; Bull et al., 2012), though some authors consider the species *X. fuscans* to be a subclade within *X. axonopodis* (Rodriguez-R et al., 2012; Mhedbi-Hajri et al., 2013).

The main host of *Xap* and *Xff* is common bean (*Phaseolus vulgaris*) but they have also been isolated from the closely related Lima bean (*Phaseolus lunatus*) and lablab bean (*Lablab purpureus*,

formerly *Dolichos lablab*) as well as several other legumes, including *Vigna* species (Bradbury, 1986). Lablab bean is a drought-resistant legume that stays green during the dry season and is used to improve soil and to feed livestock (Schaaffhausen, 1963). Lablab bean has been reported to be the main leguminous fodder crop used in Sudan around Khartoum, where it is known as hyacinth bean, bonavist bean or, in Arabic, lubia afin (Schaaffhausen, 1963). It is also grown in Sudan as a pulse legume (Mahdi and Atabani, 1992). Infection by *Xap* has been observed when lablab was sown during the rainy months (Tarr, 1958). It is not currently clear whether a single bacterial population moves frequently between host species or to what extent CBB agents colonizing different plant species represent distinct and genetically isolated populations or distinct taxa. For example, do the strains from Lima bean and lablab bean belong to the same genetic lineages as do strains from common bean?

A key determinant of pathogenicity in *Xanthomonas* species is the Hrp type-three secretion system (T3SS), which functions as a molecular syringe that secretes and translocates a number of bacterial effector proteins into the cytoplasm of the host cell, thereby modifying the host defenses to the advantage of the pathogen (Alfano and Collmer, 1997, 2004; Galán and Collmer, 1999; Grant et al., 2006; Kay and Bonas, 2009). Host ranges of *X. axonopodis* are significantly associated with the bacteria's repertoires of effectors and phylogenetically distinct strains' convergence on a common host plant might be at least partially explained by shared effectors (Hajri et al., 2009). The particular set of effectors expressed by a pathogen has some practical implications; many plant resistance genes trigger host defenses in response to detection of specific pathogen effectors. Effectors may act as virulence factors, enabling the pathogen to overcome host defenses. Therefore, rational deployment of available genetic resistance depends on knowledge of which effectors are likely to be present in a pathogen population. For example, it might be prudent to deploy resistance genes that recognize core effectors that are present in all strains of the pathogen that the plant will encounter rather than against rarely occurring effectors; this was the rationale for a recent study of the genome sequences of 65 strains of *Xanthomonas axonopodis* pv. *manihotis* (Bart et al., 2012).

CBB is currently a serious challenge to bean production in many African countries. In order to make optimal and rational use of limited available resources to contain and manage the impacts of this disease, it is important to understand the spread pathways of the *Xap* and *Xff* pathogens over both long and short geographical distances. Studies of spread rely on molecular markers that can be used to link strains from different times and locations based on their sharing similar genotypes.

According to multi-locus sequence analysis (MLSA), strains from *Phaseolus* species each fell into one of four genetic lineages (GL): GL 1, GL 2, GL 3, and GL *fuscans* (Mhedbi-Hajri et al., 2013) that corresponded to genetic lineages previously determined on the basis of amplified fragment length polymorphism (AFLP) (Alavi et al., 2008). The MLSA-based genetic lineages are consistent with an earlier classification of

X. axonopodis strains into "genetic groups" based on conserved repetitive sequences BOX, enterobacterial repetitive intergenic consensus (ERIC), and repetitive extragenic palindromic (REP) (rep-PCR) (Rademaker et al., 2005), though the MLSA-based classification provides higher resolution. Rademaker's genetic group 9.4 includes GL 1, while genetic group 9.6 includes both GL 2 and GL 3 and GL *fuscans* (Mhedbi-Hajri et al., 2013), implying that GL 2 and GL 3 are more closely related to GL *fuscans* than to GL 1.

Whole-genome sequencing is relatively cheap, easy and quick and readily discovers genetic variation that can be utilized as neutral molecular markers to track specific genotypes (Vinatzer et al., 2014; Goss, 2015). It can also reveal biologically interesting variation and the incidence and distribution of virulence factors (e.g., T3SS effectors) across the pathogen population allowing for rational deployment of genetic resistance in host crop plants, as was recently proposed for cassava and its pathogen *X. axonopodis* pv. *manihotis* (Bart et al., 2012). Other authors have pointed out that deployment of resistance without an awareness of pathogenic variation within the pathogen population could result in costly failure (Taylor et al., 1996; Fourie and Herselman, 2011). At the time of writing (July 2015) sequence assemblies are publicly available for 379 *Xanthomonas* genomes. No genome sequences were currently available for *Xap*, but two *Xff* genome sequences have been published: a finished genome for strain 4834-R (Darrasse et al., 2013b) and a draft assembly for strain 4844 (Indiana et al., 2014). A previous review (Ryan et al., 2011) presented some of the insights into *Xanthomonas* biology revealed by genome sequencing.

In the current study, we aimed to exploit whole-genome sequencing to catalog genetic diversity of CBB pathogens within each of the MLSA-based genetic lineages from common bean and strains from lablab and Lima beans. We also hypothesized that there might be some genetic features that are shared between phylogenetically distant lineages of CBB pathogens that reflect genetic exchange or adaptation to a common host. Therefore, we sequenced and bioinformatically analyzed the genomes of 26 strains deposited in the strain collections as *Xap* or *Xff* spanning six continents and more than four decades.

The specific objectives of this study were:

- To determine the phylogenomic relationships between *Xap* and *Xff* bacterial strains from different host species: common, lablab and Lima beans.
- To identify genetic variation within *Xap* or within *Xff*. These sequence variations could be exploited as molecular markers for use in epidemiological and phylogeographic studies.
- To determine patterns of conservation and variation in the complement of T3SS effectors between and within *Xap* and *Xff*. This knowledge can inform future rational deployment of disease resistance genes in beans.
- To identify genes or alleles that have been recently transmitted between phylogenetically divergent CBB bacteria.
- Identify candidate genetic determinants of the fuscous genotype, i.e., production of the brown pigment on tyrosine-containing medium.

Materials and Methods

Genome Sequencing

Genomic DNA was prepared from overnight liquid cultures of bacteria revived from the NCPPB grown on Yeast extract-Dextrose-Calcium Carbonate solid medium (i.e., agar plates) for 2 days at 28°C. DNA extraction was performed using the QIAamp DNA Mini kit (Qiagen, Hilden, Germany) applying proteinase K incubation for 30 min. We used the Nextera XT kit (Illumina, San Diego, USA) for library preparation following manufacturer's instructions. Purification was carried after fragmentation using AMPure XP beads (Beckman Coulter, High Wycombe, UK) prior to pooling. The 15pM library was then sequenced on an Illumina MiSeq using reagent kit chemistry v3 with 600 cycles.

Bioinformatics

Quality Control on Genomic Sequence Data

The quality of sequence data was checked using FastQC (Andrews)¹. Poor-quality and adaptor-containing reads were filtered and trimmed using FastQ-MCF (Aronesty, 2011).

Alignment of Sequence Reads vs. a Reference Genome Sequence

For alignment of genomic sequence reads against reference genome sequences of *Xff* 4834-R (Darrasse et al., 2013b) and *X. axonopodis* pv. *citri* 306 (Da Silva et al., 2002), we used BWA-MEM (Li, 2014). Resulting alignments were visualized using IGV (Thorvaldsdóttir et al., 2013).

Phylogenetic Analysis and Calling Single-nucleotide Variations

Phylogenetic analysis of the multi-locus sequence data was conducted in MEGA6 (Tamura et al., 2013). Multiple sequence alignments were performed using Muscle (Edgar, 2004). Evolutionary history was inferred using the maximum likelihood method based on the general time reversible model (Nei and Kumar, 2000). Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the maximum composite likelihood (MCL) approach.

For phylogenetic analysis of whole-genome assemblies, we used the Parsnp program from the Harvest suite (Treangen et al., 2014). Phylogenetic trees generated from Parsnp in Newick format were imported into MEGA6 for preparation of the final figures. Parsnp uses FastTree2 to generate approximately maximum likelihood trees (Price et al., 2010). Distributions of single-nucleotide variations, calculated by Parsnp, were visualized using Gingr from the Harvest suite.

To check the reliability of the SNPs called by Parsnp, we further checked them using our previously described method (Mazzaglia et al., 2012; Wasukira et al., 2012; Clarke et al., 2015). For this method, we aligned the sequence reads against the reference genome sequence using BWA-mem version 0.7.5a-r405

(Li, 2013, 2014) with default parameter values and excluding any reads that did not map uniquely to a single site on the reference genome. From the resulting alignments, we generated pileup files using SAMtools version 0.1.19-96b5f2294a (Li et al., 2009). We then parsed the pileup-formatted alignments to examine the polymorphism status of each single-nucleotide site in the entire *Xff* 4834-R reference genome. For each single-nucleotide site we categorized it as either ambiguous or unambiguous. A site was considered to be un- ambiguous only if there was at least 5× coverage by genomic sequence reads from each and every bacterial strain and only if for each and every bacterial strain, at least 95% of the aligned reads were in agreement. Any sites that did not satisfy these criteria were considered to be ambiguous and excluded from further analysis. Over the remaining unambiguous sites, we could be very confident in the genotype for all the sequenced strains.

De Novo Assembly

Prior to assembly, we combined overlapping reads using FLASH (Magoc and Salzberg, 2011). Genomes were assembled using SPAdes version 3.5.0 (Bankevich et al., 2012) with read error correction and with the “- - careful” switch. We assessed the quality of the assemblies and generated summary statistics using Quast (Gurevich et al., 2013) and REAPR (Hunt et al., 2013).

Automated Annotation of Genome Assemblies

Genome assemblies were annotated via the Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP) at the NCBI.

Comparison of Gene Content

To determine the presence or absence of genes in the newly sequenced genomes, we used alignment of genomic sequence reads against a reference pan-genome rather than comparison between genome assemblies. The reference pan-genome consisted of a set of gene sequences, each being a sole representative of a cluster of orthologous genes from all *Xanthomonas* genomes whose sequences were currently available; clustering of orthologous gene sequences was performed using UCLUST (Edgar, 2010). The reason for taking this approach (i.e., alignment of raw reads rather than alignment of assemblies) was to avoid potential errors arising from gaps in the genome assemblies. We aligned sequence reads against the reference genome sequence using BWA-MEM (Li and Durbin, 2009; Li, 2014) and used coverageBed from the BEDtools package (Quinlan and Hall, 2010) to determine the breadth of coverage of each gene in the resulting alignment. These breadths of coverage were visualized as heatmaps using the pheatmap module in R (R Development Core Team, 2013). We also compared genome assemblies using BRIG (Alikhan et al., 2011).

Results

Overview of Sequencing Results

We performed genomic re-sequencing on a collection of 26 *Xap* and *Xff* strains from the strain collections at NCPPB and CIAT as summarized in **Table 1**. For most of the strains, we obtained a depth of coverage of at least 40 x and thus were able

¹ Andrews, S. FastQC A Quality Control tool for High Throughput Sequence Data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> [Accessed May 13, 2015].

TABLE 1 | Sequenced strains.

| Strain | Country and date of isolation | Host | Depth of coverage | Clade | Accession numbers |
|--|-------------------------------|-------------|-------------------|--|---------------------------|
| Xap NCPPB 556 (LMG 829) | Sudan (Shambat) 1957 | Lablab bean | 15 x | "Lablab" | JTJF00000000 SRX1048889 |
| Xap NCPPB 557 (LMG 830) | Sudan (Wad-Medani) 1957 | Lablab bean | 57 x | "Lablab" | JWTE00000000 SRX1048890 |
| Xap NCPPB 2064 (LMG 8015) | Sudan (Wad-Medani) 1965 | Lablab bean | 108 x | "Lablab" | JSEZ00000000 SRX1048891 |
| Xap NCPPB 1713 (LMG 8013,) | Zimbabwe 1962 | Lablab bean | 30 x | "Lablab" | JWTD00000000 SRX1048892 |
| "Xap" NCPPB 3660 | Brazil 1975 | Common bean | 63 x | "Fuscans" | JSEX00000000 SRX1050058 |
| Xff NCPPB 381* (LMG 826, CFBP 6165) | Canada 1957 | Common bean | 40 x | "Fuscans" | JTKK00000000 SRX1050059 |
| Xff CIAT X621 | South Africa (Cedan) 1995 | Common bean | 68 x | "Fuscans" | JXHS00000000 SRX1050082 |
| Xff CIAT XCP631 | Colombia 2004 | Unknown | 33 x | "Fuscans" | JXLW00000000 SRX1050252 |
| Xff NCPPB 1056* (LMG 7457) | Ethiopia 1961 | Common bean | 44 x | "Fuscans" | JSEV00000000 SRX1049856 |
| "Xap" NCPPB 1058** | Ethiopia 1961 | Common bean | 150 x | "Fuscans" | JSEY00000000 SRX1049857 |
| Xff NCPPB 1433* (LMG 8016) | Hungary 1956 | Common bean | 51 x | "Fuscans" | JSBT00000000 SRX1049858 |
| Xff NCPPB 2665* (LMG 841) | Italy 1973 | Common bean | 70 x | "Fuscans" | JSBQ00000000 SRX1049859 |
| Xff NCPPB 1654* (LMG 837) | South Africa 1963 | Common bean | 93 x | "Fuscans" | JSBR00000000 SRX1049860 |
| "Xap" NCPPB 670** (LMG 832) | Uganda 1958 | Common bean | 32 x | "Fuscans" | JRRE00000000 SRX1049872 |
| Xff NCPPB 1402* (LMG 7459) | Uganda 1962 | Common bean | 17 x | "Fuscans" | JSEW00000000 SRX1049873 |
| Xff NCPPB 1158* (LMG 7458) | UK 1961 | Common bean | 40 x | "Fuscans" | JSBS00000000 SRX1049874 |
| Xff NCPPB 1495* (LMG 8017) | UK 1963 | Common bean | 30 x | "Fuscans" | JSEU00000000 SRX1049875 |
| Xap NCPPB 1646 (LMG 8011) | Australia 1964 | Common bean | 48 x | GL1 | JTCT00000000 SRX1050299 |
| Xap NCPPB 301 | Canada pre-1951 | Not known | 50 x | GL1 | JTCU00000000 SRX1050300 |
| Xap CIAT XCP123 | Colombia 1974 | Lima bean | 30 x | GL1 | JXLV00000000 SPX1050292 |
| Xap NCPPB 1420 (LMG 836) | Hungary 1956 | Common bean | 92 x | GL1 | JTCV00000000 SRX1050301 |
| Xap NCPPB 1811 (LMG 8014, CFBP 6164) | Romania 1966 | Common bean | 54 x | GL1 | JWTF00000000 SRX1050302 |
| Xap NCPPB 1680 (LMG 8012) | Tanzania 1964 | Common bean | 66 x | GL1 | JWTG00000000 SRX1050303 |
| Xap NCPPB 3035 (T) (LMG 7455, CFBP 6546) | USA pre-1978 | Common bean | 27 x | GL1 | JSFA00000000 SRX1050305 |
| Xap NCPPB 220 (NCTC 4331) | USA pre-1948 | Not known | 62 x | GL1 | JWTH00000000 SRX1050306 |
| Xap NCPPB 1138 (LMG 834) | Zambia 1961 | Common bean | 45 x | GL1 | JWTI00000000 SRX1050323 |
| "Xap" NCPPB 1128 | Jamaica 1961 | Common bean | 80 x | Unknown <i>Xanthomonas</i> species | LFME00000000.1 SRX1090401 |

All strains had been deposited as Xap, except for those marked with an asterisk (*), which had been deposited as "X. axonopodis pv. phaseoli variant fuscans." Strains marked with two asterisks (**) were deposited as Xap but are reported to produce brown pigment, according to the accession cards that were submitted along with the strains into the NCPPB. Depth of coverage was estimated from alignments of raw sequence reads against the reference genome of *X. axonopodis* pv. *citri* 306 (Da Silva et al., 2002) using BWA-MEM (Li and Durbin, 2009; Li, 2014). GenBank accession numbers are given for the genome assemblies and SRA accession numbers are given for the raw sequence reads. Accession numbers are given for synonymous strains from the Belgian Coordinated Collections of Micro-Organisms (LMG), the Collection Française de Bactéries associées aux Plantes (CFBP) and National Collection of Type Cultures (NCTC).

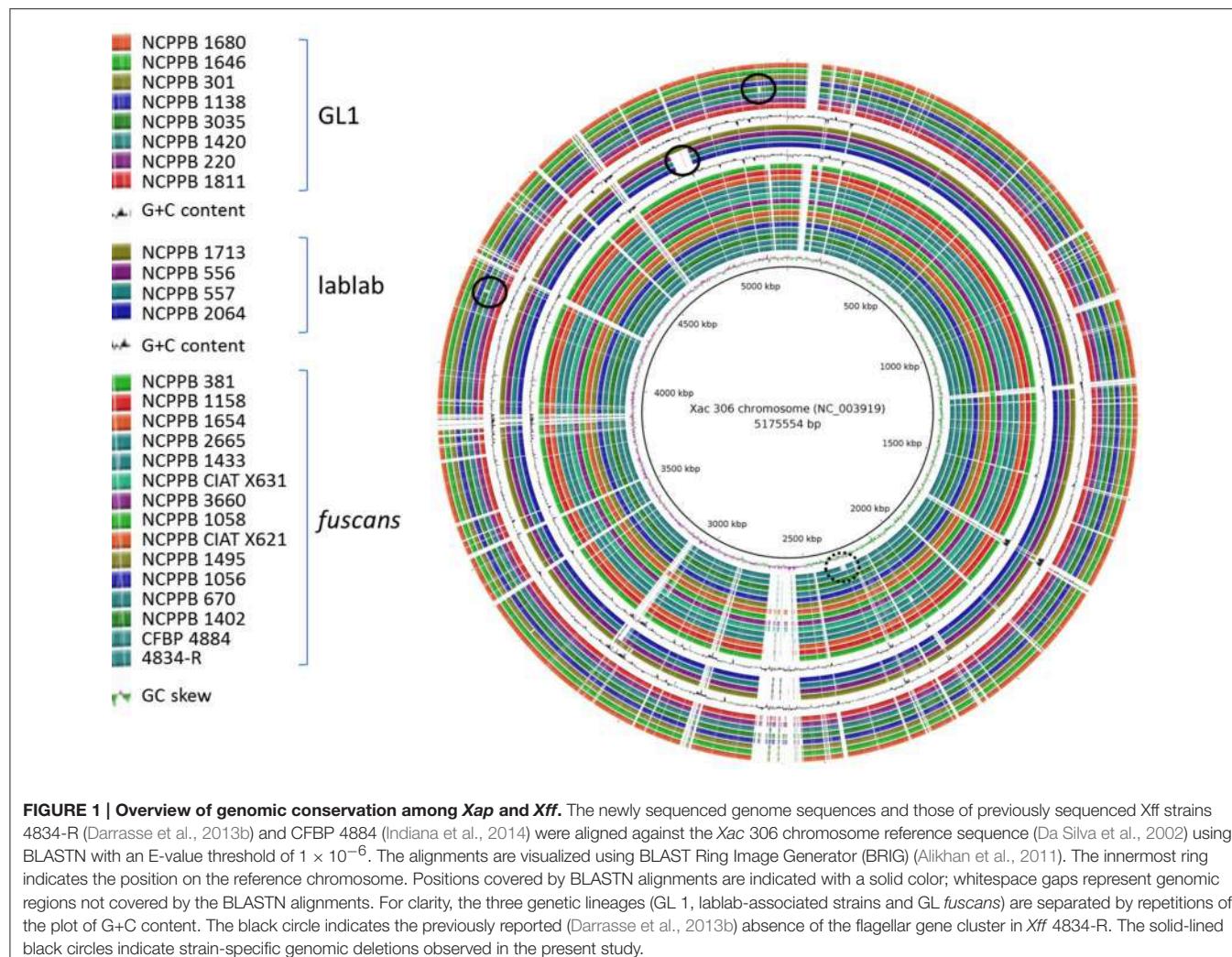
to generate *de novo* genome assemblies. However, for seven of the genomes, there was less than 40 x coverage. We investigated the relationship between coverage depth and assembly quality by assembling subsets of the sequence reads from NCPPB 1058. We found that contig N₅₀ length peaked at around 40 x coverage, with further increases in depth yielding little or no increase in contig lengths.

Figure 1 shows an overview of the *de novo* assemblies of each sequenced Xap and Xff genome aligned against that of the *X. axonopodis* pv. *citri* 306 (Da Silva et al., 2002). See also the Supplementary Figures for genome-wide alignments of the assemblies using Mauve (Darling et al., 2004). Note that the 26 Xap and Xff genomes were assembled *de novo*, using SPAdes (Bankevich et al., 2012), without use of a reference sequence.

The contiguities of the assemblies were comparable to those of previously sequenced *Xanthomonas* genomes. This is illustrated

by the distribution of N₅₀ contig lengths, which ranged from 39.4 to 123.6 kb. The range for a recent study of 65 *X. axonopodis* pv. *manihotis* was 7.4–111.0 kb (Bart et al., 2012). A full summary of assembly statistics, calculated using Quast (Gurevich et al., 2013), is provided in the Supplementary Table S1.

Contiguity of an assembly does not necessarily correlate with accuracy. Therefore, in addition to the Quast analysis of assembly contiguity, we also assessed the accuracies of the assemblies using REAPR (Hunt et al., 2013). This method is based on aligning to the assembly the sequence reads from which it was generated. This allows detection of anomalies in coverage of the assembly by reads and flags two classes of potential errors: fragment coverage distribution (FCD) errors and low fragment coverage errors. We compared the frequencies of these two classes of potential error for each of our genome assemblies and also for each of the 65 previously published *X. axonopodis* pv. *manihotis* assemblies



(Bart et al., 2012); see Supplementary Figure S1. The genome assemblies generated in the present study were of comparable quality to those from the previously published study. However, there is a general trend toward our genome assemblies having more “low fragment coverage” errors and fewer “FCD” errors.

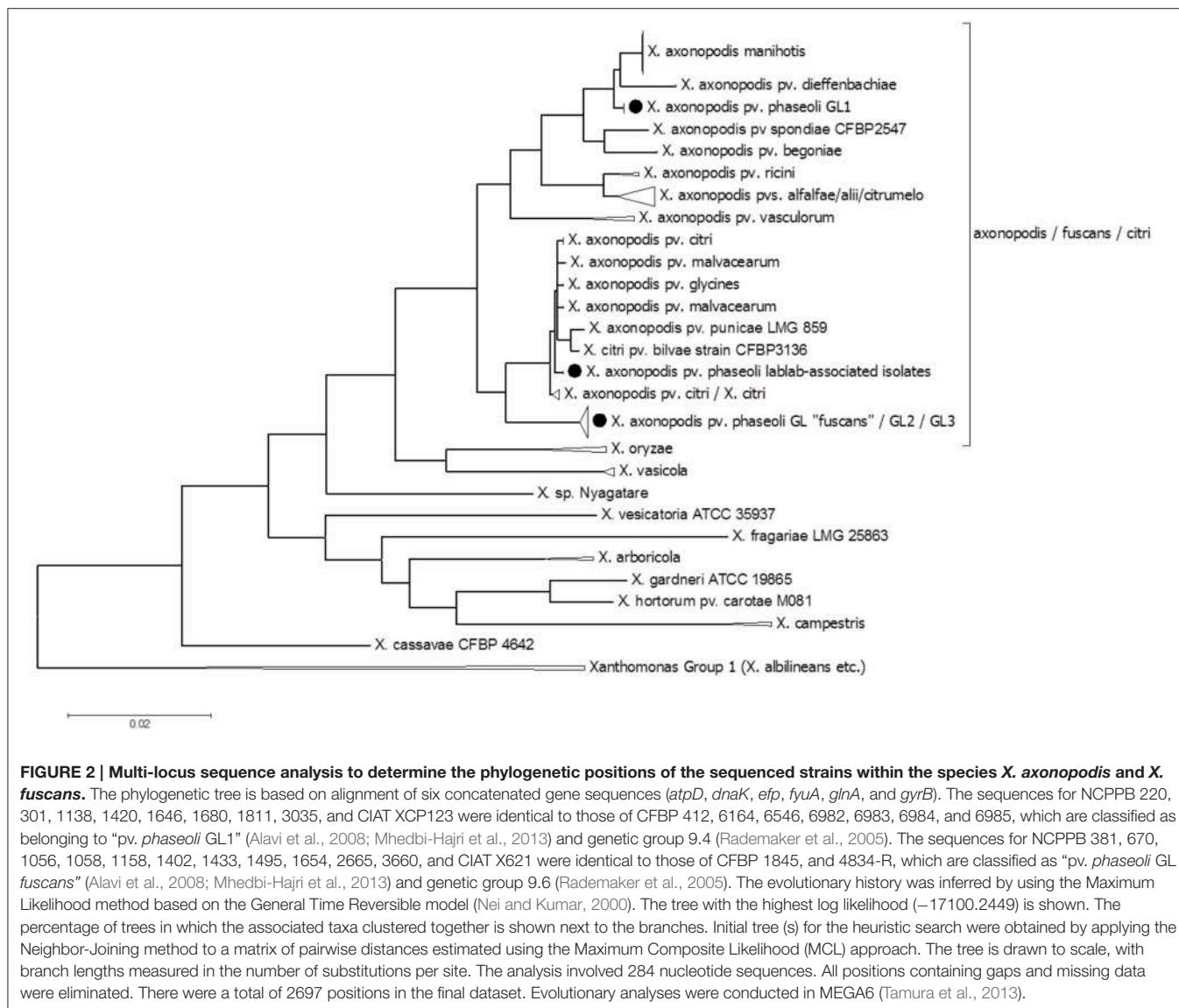
To ascertain the phylogenetic positions of each sequenced strain, we initially used a multi-locus sequence analysis (MLSA) approach, using concatenated sequences from six genes that had been used in previous MLSA studies (Young et al., 2008; Almeida et al., 2010; Hajri et al., 2012; Hamza et al., 2012). This approach had the advantage that we could include in the analysis many *Xap* strains and other xanthomonads whose genomes had not been sequenced but for which MLSA data were available. Nucleotide sequences are available for these six genes from a large number of xanthomonads, either from whole-genome sequence assemblies or from the MLSA studies. We combined the publicly available sequences with homologous sequences extracted from the genomes newly sequenced for this study. The results of the MLSA revealed that the newly sequenced *Xap* and *Xff* genomes each fell into one of three distinct clades: GL 1, GL *fuscans*

and a previously undescribed lineage associated with lablab bean (**Figure 2**).

The newly sequenced strains from lablab bean comprised a third clade, quite distinct from both *Xap* GL1 and from GL *fuscans* and indeed all previously described lineages of bean pathogens. The lablab-associated strains are closely related to members of Rademaker’s genetic group 9.5, along with strains of pathovars *bilvae*, *citri*, *malvacearum*, and *mangiferaeindicae* that are pathogens of diverse plants including Bengal quince, *Citrus* spp., cotton and mango respectively (Bradbury, 1986; Rademaker et al., 2005). Also falling within this MLSA-based clade are strains of *X. axonopodis* pv. *glycines*, causative agent of bacterial pustule in soybean (Jones, 1987).

Genome-wide SNP Analysis Elucidates Phylogeny at Greater Resolution

Based on six-gene MLSA alone, strains could be ascribed to one of the three genetic lineages (GL 1, GL *fuscans*, and GL lablab). However, genome-wide sequence comparisons provided additional resolution and revealed distinct clades



(or sub-lineages) within each lineage. **Figure 3A** illustrates the distribution of single-nucleotide variations across the reference sequence of the chromosome of *Xff* 4834-R for 160 publicly available related genome sequences. **Figure 3B** shows a phylogenetic reconstruction of those 160 genomes based on those variants. Consistent with the MLSA results, the strains sequenced in the present study similarly fell into three clades.

Within the *fuscans* lineage, the genome-wide comparison revealed at least three distinct sub-lineages, depicted in **Figure 4** in red, blue and green respectively. Each of these three lineages includes strains from diverse geographical locations and years. For example, one sub-lineage includes strains from France (1998), Hungary (1956), Italy (1963), South Africa (1963) and the UK (1962). This suggests that this sub-lineage has been circulating in Europe for nearly six decades and has spread between Europe and South Africa at least once, perhaps indirectly via another locality. This pattern is consistent with spread of the pathogen via global trade of seeds.

The genome-wide sequence analysis also reveals that multiple genetic lineages may be present within a single geographical area. For example, NCPPB 1056 and NCPPB 1058 were both isolated in the same country and the same year (Ethiopia, 1961) and fall into two distinct sub-lineages (**Figure 4**).

Similar, intra-lineage variation can be observed for strains within the lablab-associated strains (**Figure 5**) and GL 1 (**Figure 6**). Among lablab-associated strains, those collected in Sudan between 1957 and 1965 cluster together and are distinct from NCPPB 1713, which originates from Zimbabwe in 1962. Within GL 1, there are two multi-strain sub-lineages, which are indicated in blue and green in **Figure 5**. The former sub-lineage spans Australia, Canada, and Tanzania. The latter sub-lineage includes strains from Hungary, Romania, and the USA. Strain NCPPB 1138 (from Zambia, 1961) is distinct from both of these. The single GL 1 strain from Lima bean (CIAT XCP123, Colombia, 1974) is distinct from all of the

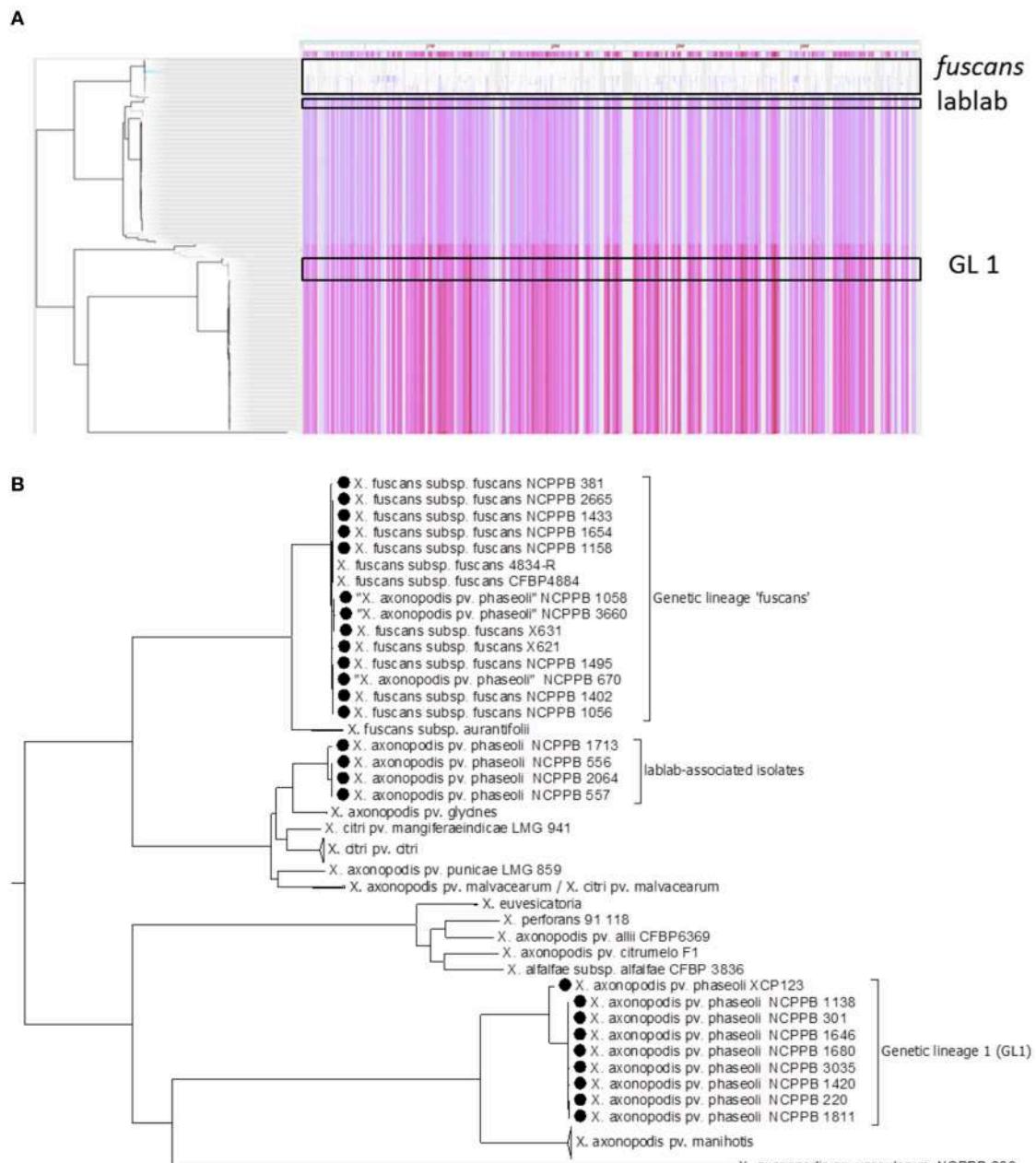
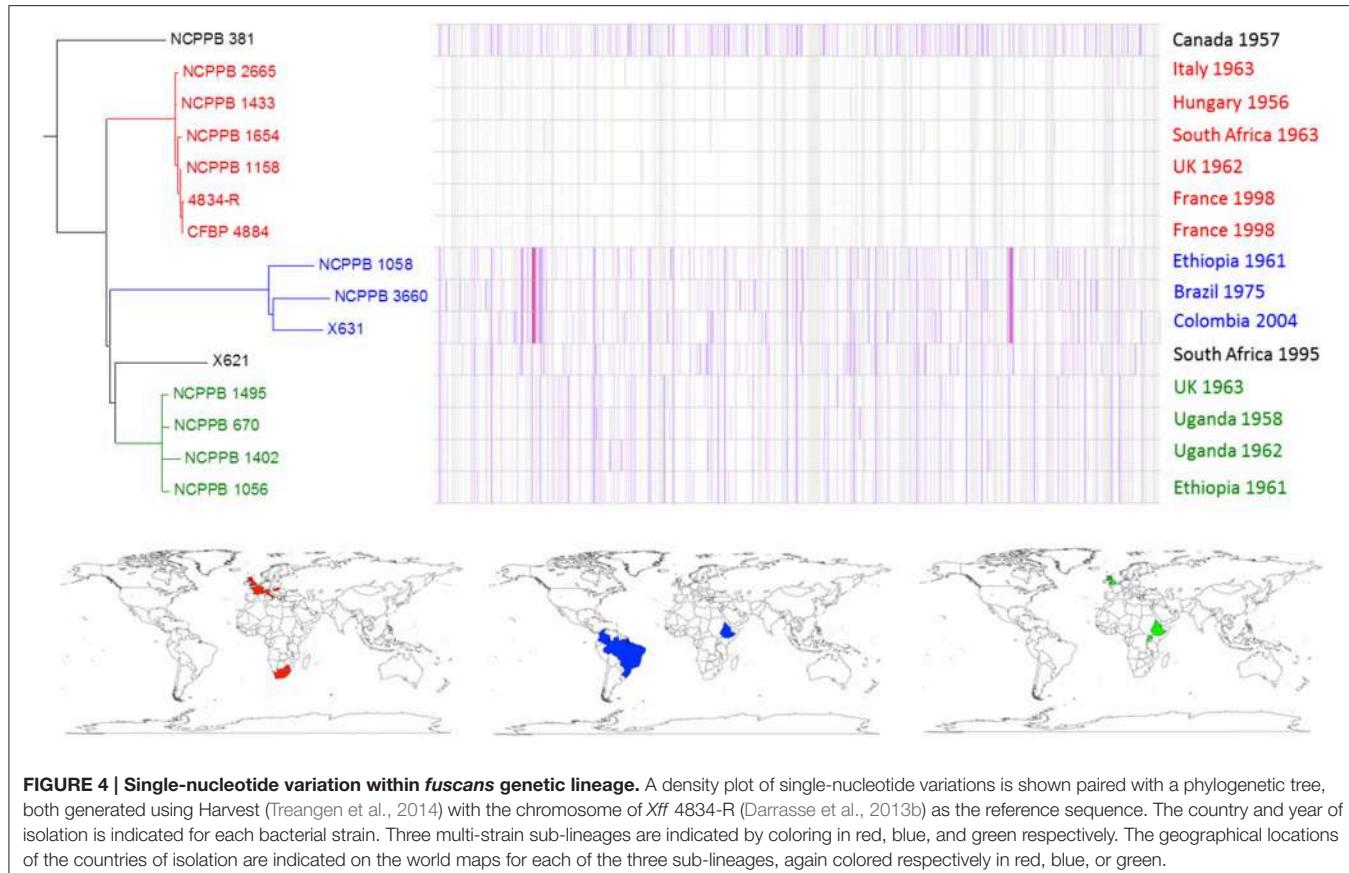


FIGURE 3 | Genome-wide identification of single-nucleotide variations. Panel (A) shows a density plot of single-nucleotide variations paired with a phylogenetic tree, both generated using Harvest (Treangen et al., 2014) with the chromosome of *Xff* 4834-R (Darrasse et al., 2013b) as the reference sequence. Panel (B) shows the same phylogenetic tree as in Panel (A) but with taxon labels and some clades collapsed for clarity. In addition to the 26 newly sequenced genomes, all 134 publicly available genome assemblies from *X. axonopodis*, *X. fuscans*, *X. citri*, and *X. euvesicatoria* were included.

strains from common bean (Figure 5); however, based on MLSA alone, it is indistinguishable from the other GL 1 strains.

Across the 4,981,995-bp chromosome sequence of *Xff* 4834-R, Harvest identified a total of 135,321 SNPs. This number includes all single-nucleotide sites that show variation between any of the 160 genome assemblies included in the analysis. A subset of 61,462 of those SNPs showed polymorphism among the 26 *Xap* and *Xff* genomes sequenced in the present study. The Harvest

SNP calling takes as its input assembled genome sequences. Thus, substitution errors in the assemblies then could appear as false positives. Gaps in the assemblies are unlikely to generate false-positive SNP calls as Harvest only considers the core genome, i.e., those regions of the genome that are present in all of the genome assemblies and discards genomic regions present in only a subset of the assemblies. To assess the reliability of the Harvest SNP calls, we compared the results with a read-based method of SNP calling that we have used previously (Mazzaglia et al., 2012; Wasukira



et al., 2012; Clarke et al., 2015). Read-based methods have the advantage of not being reliant on assembly and they exploit the signal from multiple independent overlapping sequence reads at each site in the genome sequence. However, sequence reads are not available for the majority of *Xanthomonas* genome sequences, since for most studies only the assemblies and not the reads have been deposited in the public repositories. Of the 61,462 SNPs that Harvest called for the *Xap* and *Xff* genomes, our read-based method confirmed 53,811 (87.5%).

It is evident from Figures 3–6 that single-nucleotide variations occur throughout the chromosome. However, the distribution is not uniform and there are several apparent “hotspots” of variation. The most likely explanation for these regions of higher-than-average sequence divergence is horizontal acquisition of genetic material from relatively distantly related strains. Such incongruent patterns of sequence similarity due to horizontal transfer have been reported previously in *Xanthomonas* species (Fargier et al., 2011; Hamza et al., 2012).

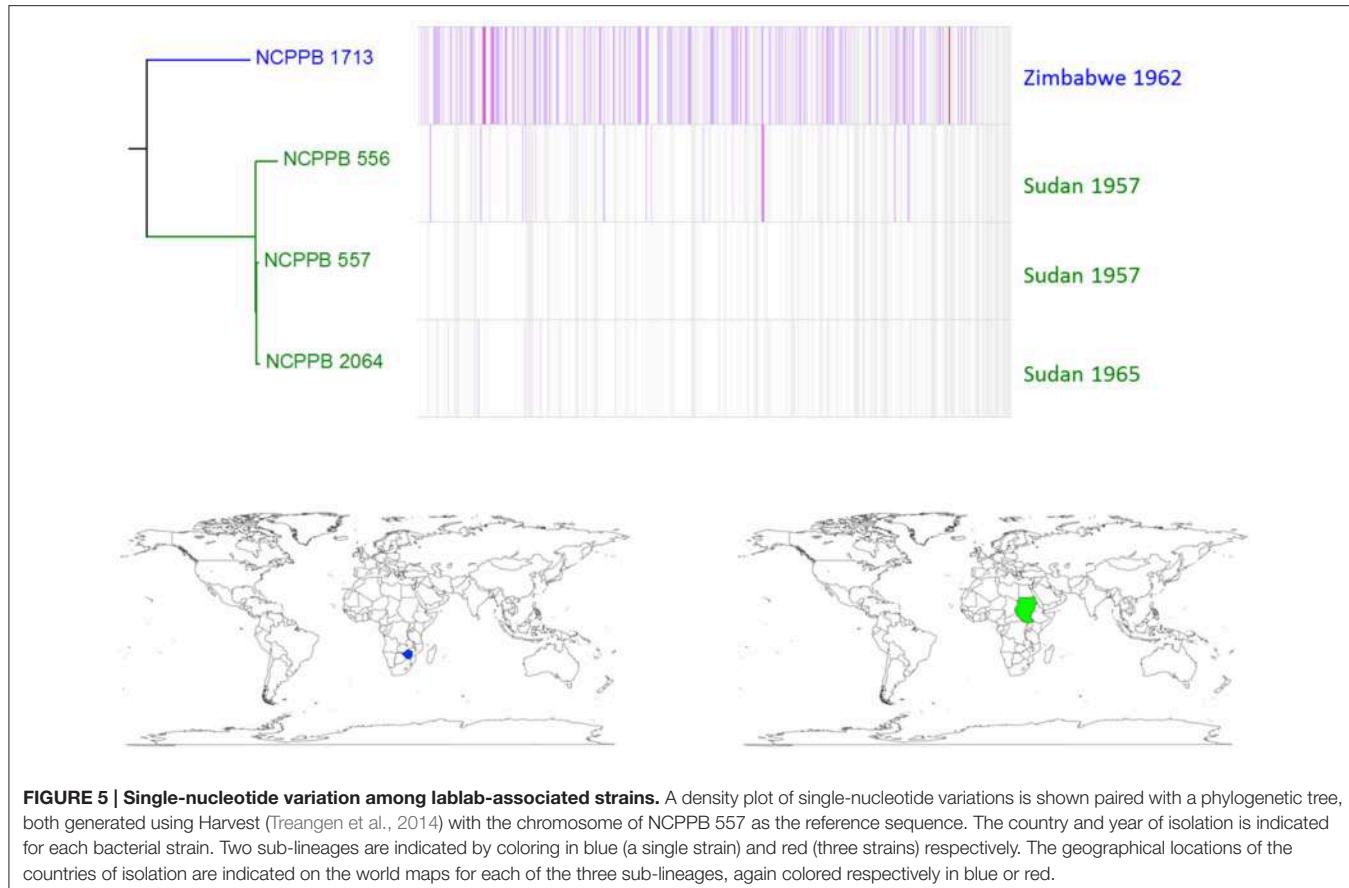
Gene-content Varies between and within Each Clade

Consistent with the indications of horizontal genetic transfer described in the previous section, we observed significant variations in gene presence and absence among strains within each of the three genetic lineages (Figure 7). Within the

fuscans strains, there were 1188 clusters of orthologous genes that were present in at least one strain and absent from at least one other (Figure 7A). Among the lablab-associated strains, 472 orthologous gene clusters showed presence-absence polymorphism (Figure 7B). Among GL 1, the number was 535 (Figure 7A). Clustering of genomes according to gene content is broadly congruent with phylogeny. Supplementary Tables S2–S7 list genes whose presence distinguishes between *Xff*, *Xap* GL 1 and lablab-associated strains. Additionally, the four lablab-associated strains all contain six genes that have no close homologs amongst other sequenced xanthomonads. These are predicted to encode: three hypothetical proteins (KHS05433.1, KKY05378.1, and KHS05434.1), pilus assembly protein PilW (KHS05489.1), an oxidoreductase (KHS05432.1), and an epimerase (KHS05485.1).

Strain-specific Large Chromosomal Deletions

A large chromosomal deletion has been previously reported in *Xff* 4834-R in which a large part of the flagellar gene cluster is absent (Darrasse et al., 2013b). This deletion is visible in Figure 1 at around position 2310 kb in the *Xac* 306 chromosome and indicated by a black circle with broken line. Although, similar deletions were reported in 5% of the strains tested (Darrasse et al., 2013b), this flagellar gene cluster was intact in all of the genomes sequenced in the current study as well as in the previously sequenced *Xff* 4884 (Indiana et al., 2014).



In addition to the strain-specific flagellar deletion, **Figure 1** reveals several other large genomic deletions, examples of which are indicated with black circles. The largest example is a 50-kb region of the *Xac* 306 chromosome sequence that is absent from the three Sudanese lablab-associated strains but present in the Zimbabwe strain. This absence is visible in **Figure 1** at between 4.82 and 4.87 Mb on the reference chromosome sequence and indicated by a black circle. The absence of this region is supported not only by the *de novo* assemblies of NCPPB 556, 557 and 2064, but also by alignment of the raw sequence reads against the *Xac* 306 reference genome, eliminating the possibility that it merely represents an assembly artifact. This region is illustrated in Supplementary Figure S6, includes locus tags XAC4111–XAC4147 and is predicted to encode a type-6 secretion system (Darrasse et al., 2013b).

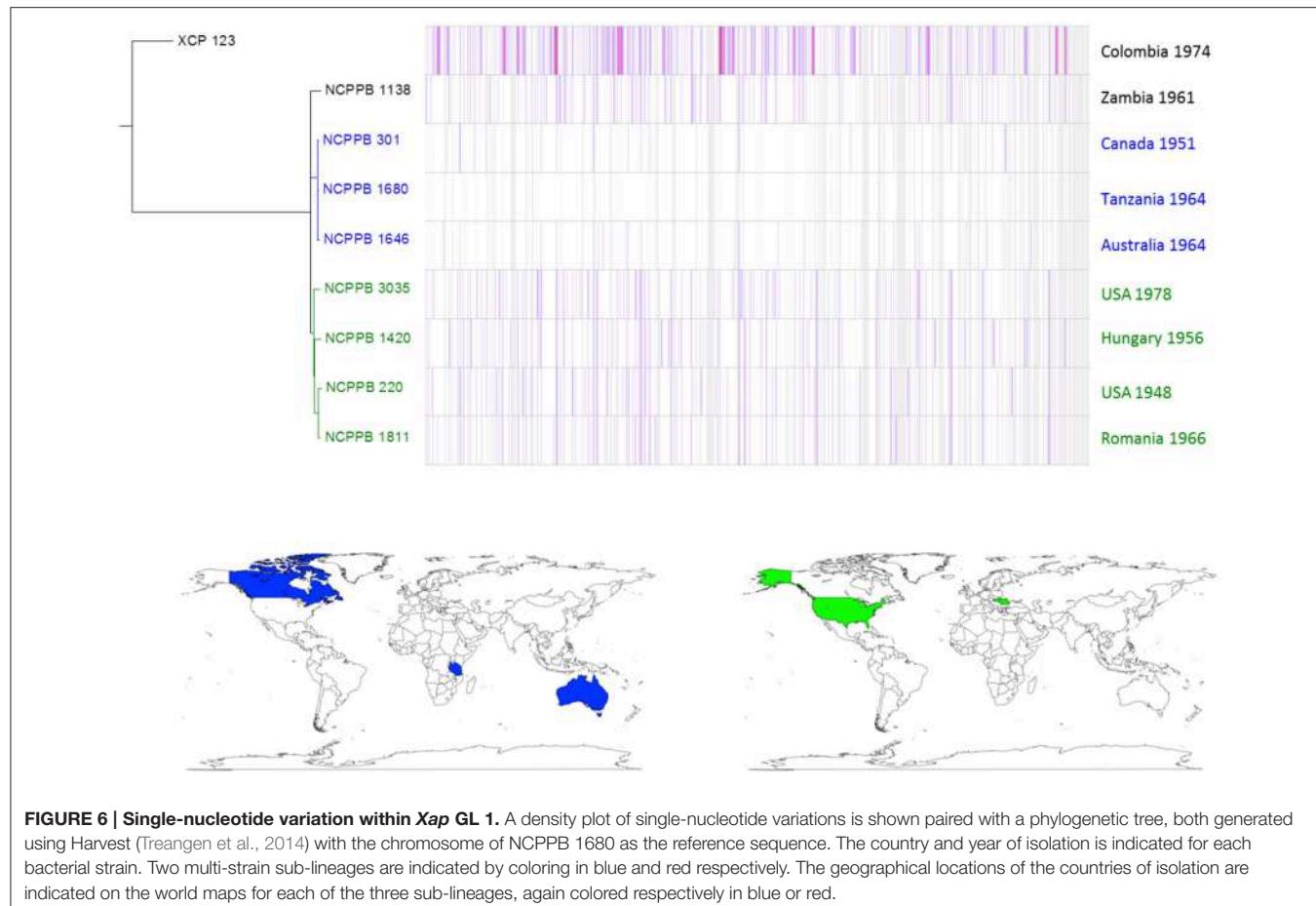
Other examples include a deletion of approximately 9 Kb that is deleted in *Xap* NCPPB 3035, resulting in loss of its ortholog of gene XAC_RS17930 and parts of the two flanking genes XAC_RS17925 and XAC_RS17935 at around position 4.20 Mb on the reference genome (Supplementary Figure S7). A second example of a deletion unique to NCPPB 3035 spans approximately 10 kb at around position 5.10 Mb (Supplementary Figure S8). The deleted region contains genes XAC_RS21755 (predicted plasmid stabilization protein) to XAC_RS21815 (predicted transposase) and likely represents a mobile element.

Strains of *Xap* GL1 Encode a SPI-1-like T3SS

A previously published suppression subtractive hybridizations study comparing bean pathogens and closely related xanthomonads revealed the presence of genes encoding several protein components of a T3SS similar to that of *Salmonella* pathogenicity island 1 (SPI-1) in the genome of *Xap* CFBP 6164 (Alavi et al., 2008). This strain is synonymous with NCPPB 1811 and belongs to lineage GL 1. Subsequently, genome sequencing revealed that *X. albilineans* encodes a SPI-1-like T3SS (Pieretti et al., 2009, 2015) and targeted sequencing confirmed its presence in two further *Xap* GL 1 strains: CFBP 2534 (same as NCPPB 3035) and CFBP 6982 (Marguerettaz et al., 2011). Whole-genome sequencing in the current study indicated that this SPI-1-like T3SS was encoded in the genomes of all GL 1 strains from common bean and Lima bean (**Figure 8**) but was absent from GL fuscans and from the lablab-associated strains. All of the putative structural genes for the T3SS are conserved in *Xap* GL 1 but the *xapABCDEFGHI* genes, hypothesized to encode effectors that are substrates of the T3SS in *X. albilineans* (Marguerettaz et al., 2011), are not conserved in *Xap*.

Repertoires of Hrp T3SS Effectors

Previous genome sequencing of *Xf* 4834-R revealed the presence of genes encoding 30 predicted effectors potentially secreted by the Hrp T3SS (Darrasse et al., 2013b). We searched for orthologs



of these and other *Xanthomonas* T3SS effectors in the newly sequenced *Xap* and *Xff* genomes using TBLASTN (Altschul et al., 1990) to search the genome assemblies against each protein query sequence. The results are summarized in Figure 9. There is a core set of 14 effectors that is encoded in all sequenced strains of *Xap* and *Xff*: XopK, XopZ, XopR, XopV, XopE1, XopN, XopQ, XopAK, XopA, XopL, AvrBs2, and XopX. Four of these are also included in the core set of effectors conserved among 65 strains of *X. axonopodis* pv. *manihotis* (Bart et al., 2012), namely XopE1, XopL, XopN, and XopV. Several others are encoded in most but not all of the newly sequenced genomes, for example: XopC1, XfuTAL2, and XopJ5. Others appear to be limited to just one of the three lineages. For example, XopF2 is limited to lineage *fuscans*, XopC2 is found only in *Xap* GL1 and XopAI is restricted to the lablab-associated strains.

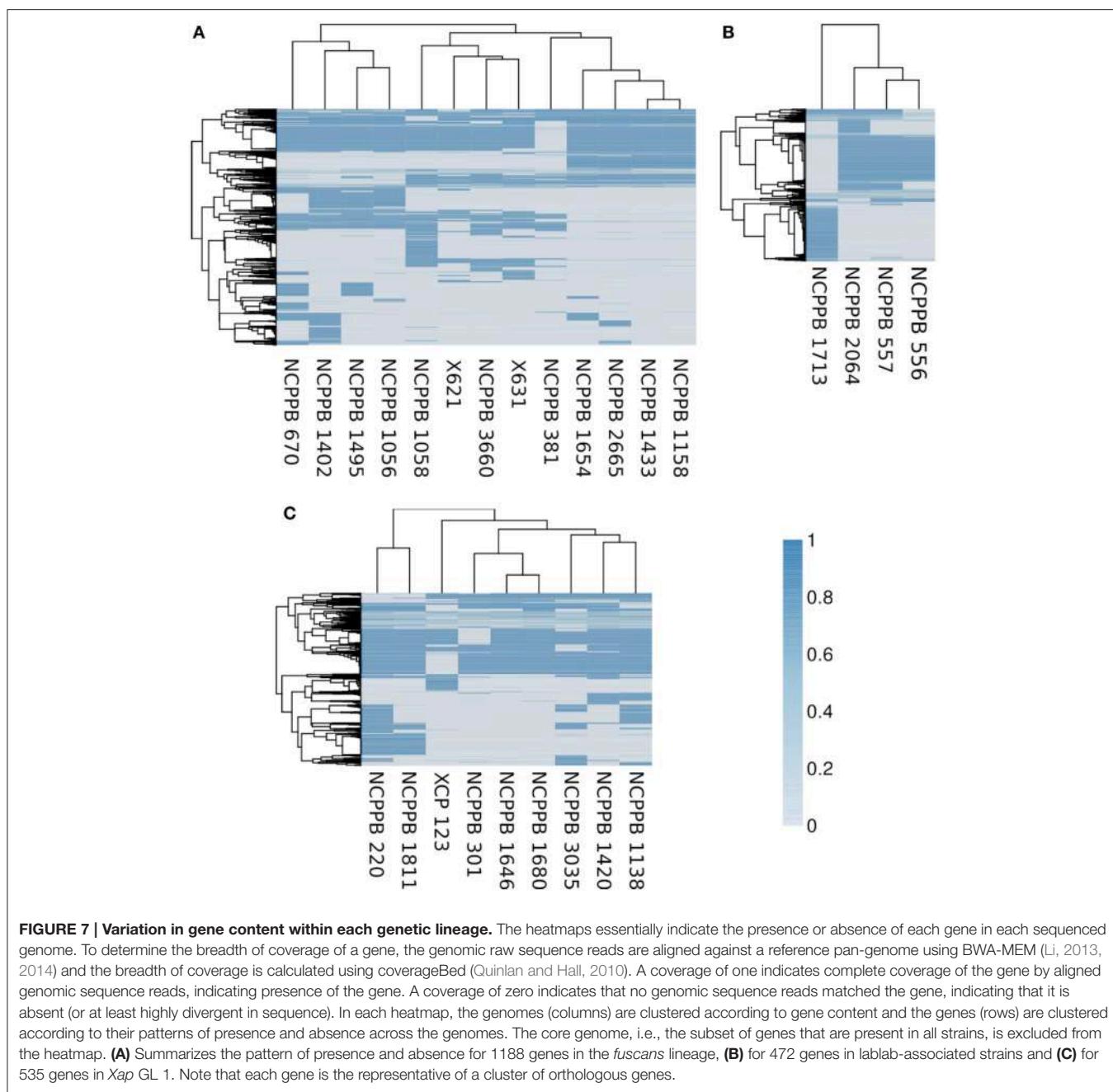
The Molecular basis for Pigmentation

Some bacterial strains from CBB infections produce a brown pigment when grown in tyrosine-containing medium and are therefore described as “fuscous.” The pigment is not believed to be directly associated with virulence (Gilbertson et al., 1991; Fourie, 2002) but fuscous strains tend to be very virulent on bean (Birch et al., 1997; Toth et al., 1998). The brown color arises from oxidized homogentisic acid (2,5 dihydroxyphenyl acetic acid), an intermediate in the tyrosine catabolic pathway

that gets secreted and oxidized in these fuscous strains (Goodwin and Sopher, 1994). Genome sequencing of the fuscous strain *Xff* 4834-R revealed a single-nucleotide deletion in *hmgA*, the gene encoding homogentisate oxygenase (Darrasse et al., 2013b). This enzyme catalyzes a step in the tyrosine degradation pathway that converts tyrosine to fumarate and hence its inactivation likely disrupts tyrosine degradation leading to accumulation of homogentisate and its subsequent oxidation to form the brown pigment. Consistent with this hypothesis, we found that the single-nucleotide deletion was present in all of the sequenced strains belonging to GL *fuscans* resulting in a predicted protein product that is truncated, while the *hmgA* gene was intact in all of the *Xap* GL1 and lablab-associated *Xap* genomes (see Figure 10).

Recent Genetic Exchange between *Xap* GL 1 and GL Fuscans

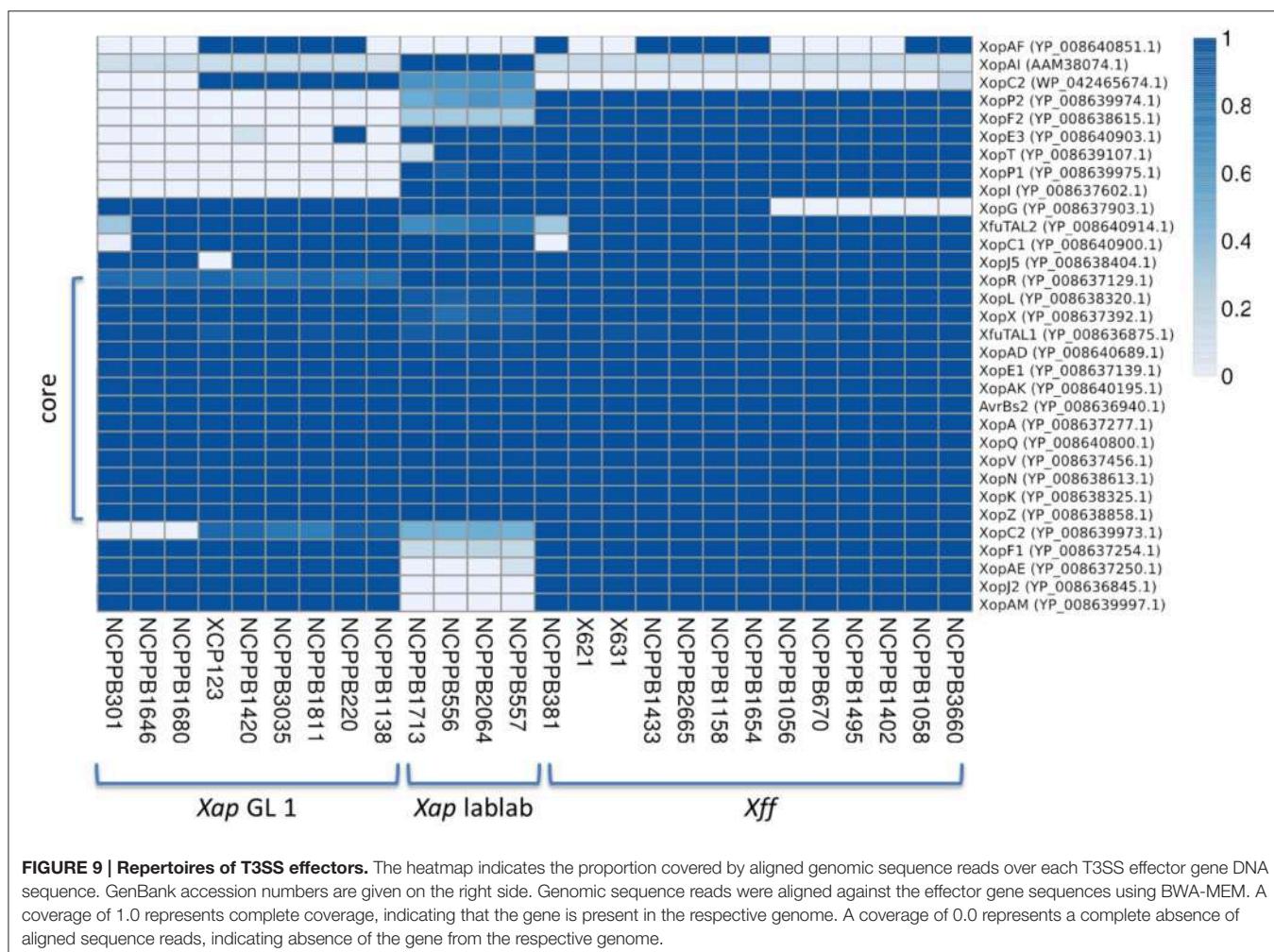
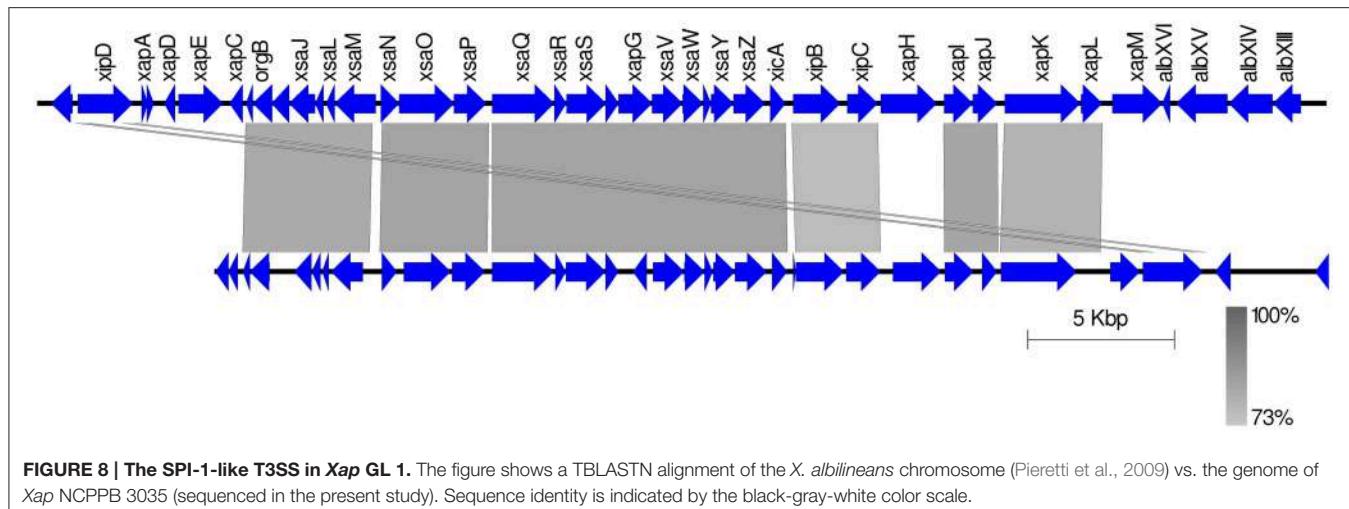
Patterns of single-nucleotide variation (Figure 3A) revealed some regions of the genome where *Xap* GL 1 had many fewer variants with respect to the *Xff* 4834-R reference genome than did the closely related *X. axonopodis* pv. *manihotis*. Closer inspection revealed numerous genes where the *Xap* GL 1 strains shared an identical allele with *Xff*, a pattern that is incongruent with their relatively distant phylogenetic relationship.



To further investigate this phenomenon, we calculated pairwise nucleotide sequence identities for each *Xap* GL 1 gene vs. its closest homolog in other lineages within *X. axonopodis* and *X. fuscans*. The results are summarized in **Figure 11**. Pairwise sequence identities between *Xap* GL 1 and *Xff* (*GL fuscans*) followed a bimodal distribution with peaks at around 96% and at 100%. The peak at 100% was not observed for identities between *Xap* GL 1 and other lineages (*X. axonopodis* pv. *glycines*, *X. axonopodis* pv. *citri*, *X. axonopodis* pv. *manihotis*, *X. fuscans* subsp. *Aurantifolii*, and lablab-associated *Xap*). **Table 2** lists examples of genes with 100% identity between *Xap* GL 1 and *Xff*. Essentially the same set of genes is affected in all of the

Xap GL 1 strains and the alleles are more similar to alleles from pathovars *citri* and *glycines* than to *manihotis*. Therefore, the most parsimonious explanation is that these alleles have been acquired by the ancestors of *Xap* GL 1 from the *fuscans* lineage.

Genome sequencing of *Xap* strains from lablab bean has revealed a previously unknown distinct lineage of *Xap*. This lineage is more closely related to strains of *X. axonopodis* pv. *glycines* than to any of the previously described genetic lineages of *Xap*. The existence of a separate lablab-associated lineage on lablab suggests that there may not be frequent movement of CBB bacteria between this species and common bean. However, conformation of this hypothesis will require genotyping of larger



numbers of strains; with the availability of these genome data it will be straightforward to design PCR-based assays to identify bacterial strains belonging to this newly discovered lineage.

It was previously observed that a *Xap* strain from common bean (NCPPB 302) was less pathogenic on lablab than bacteria isolated from naturally infected lablab (Sabet, 1959). The same

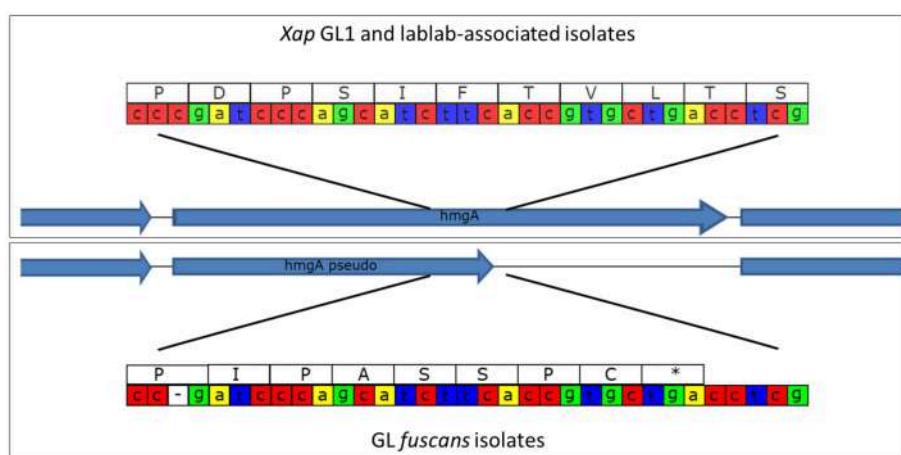


FIGURE 10 | Disruption of the *hmgA* gene in *GL fuscans*. The cartoon illustrates a sequence polymorphism in the *hmgA* gene whereby all of the sequenced *GL fuscans* strains (including previously sequenced 4834-R and CFBP4884) have a single-nucleotide deletion that results in a frame-shift and premature stop codon. All of the *Xap* GL 1 and lablab-associated strains encode a full-length protein product.

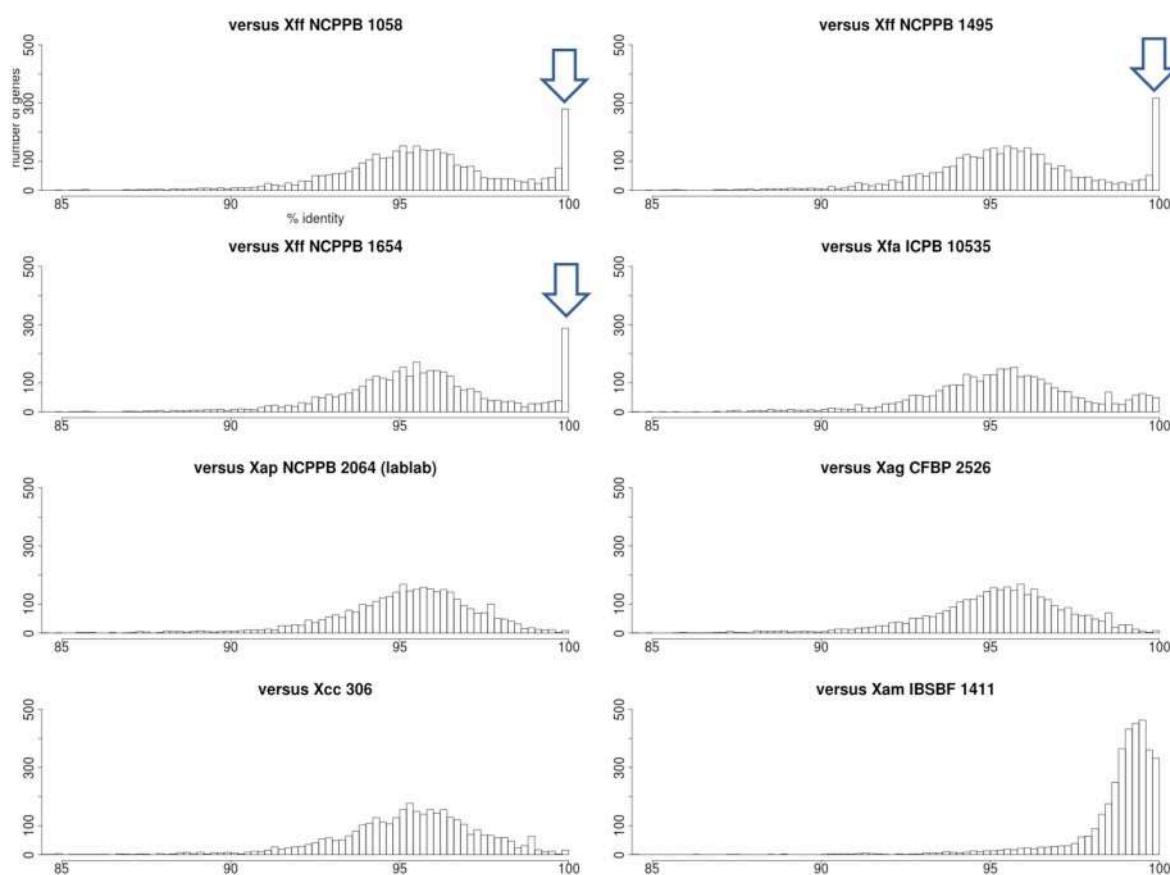


FIGURE 11 | Nucleotide sequence identities of *Xap* GL 1 genes vs. other lineages and pathovars. Each histogram shows the frequency distribution of sequence identity between genes from a *Xap* GL 1 strain (NCPPB 1680) and their closest BLASTN matches in other lineages including *GL fuscans* (NCPPB 1058, 1494 and 1654), lablab-associated *Xap* (NCPPB 2064), *X. fuscans* subsp. *aurantifoliae* (ICPB 10535), *X. axonopodis* pv. *glycines* (CFBP 2526), *X. axonopodis* pv. *citri* (306) and *X. axonopodis* pv. *manihotis* (IBSBF 1411) (Da Silva et al., 2002; Moreira et al., 2010; Bart et al., 2012; Darrasse et al., 2013a). The arrowheads indicate the positions of the peaks at 100% sequence identity between GL 1 and GL fuscans.

TABLE 2 | Genes in *Xap* GL 1 that share 100% nucleotide sequence identity with *Xff* (GL fuscans).

| GenBank accession number and predicted product of <i>Xap</i> NCPPB 1680 | <i>Xag</i> CFBP 2526 (%) | <i>Xam</i> IBSBF 1411 (%) | <i>Xap</i> NCPPB2064 (%) | <i>Xfa</i> ICPB 10535 (%) |
|--|--------------------------|---------------------------|--------------------------|---------------------------|
| KHS20952 glutamine amidotransferase | 97.73 | 99.47 | 97.61 | 98.14 |
| KHS20955 preprotein translocase | 97.28 | 94.42 | 98.88 | 99.20 |
| KHS21044 phospholipase | 96.36 | 93.73 | 96.49 | 99.75 |
| KHS21241 ATPase AAA | 97.92 | 93.84 | 97.92 | 98.67 |
| KHS21242 histidine kinase | 96.95 | 93.91 | 98.54 | 99.58 |
| KHS21578 preprotein translocase subunit SecE | 98.28 | 95.58 | 98.03 | 99.51 |
| KHS21579 transcription antiterminator NusG | 98.03 | 94.97 | 98.03 | 99.82 |
| KHS21580 50S ribosomal protein L11 | 98.36 | 97.20 | 98.13 | 100.00 |
| KHS21581 50S ribosomal protein L1 | 97.71 | 98.14 | 97.56 | 99.28 |
| KHS21586 30S ribosomal protein S12 | 99.20 | 98.66 | 98.93 | 100.00 |
| KHS22189 membrane protein | 94.84 | 93.28 | 0.00 | 99.06 |
| KHS22203 type VI secretion protein | 97.93 | 96.69 | 0.00 | 99.17 |
| KHS22241 UDP-2 3-diacylglicosamine hydrolase | 98.25 | 95.42 | 98.12 | 99.46 |
| KHS22843 cardiolipin synthetase | 94.63 | 94.94 | 94.63 | 99.68 |
| KHS23016 ribose-phosphate pyrophosphokinase | 98.02 | 97.08 | 98.33 | 99.37 |
| KHS23018 peptidyl-tRNA hydrolase | 96.98 | 95.21 | 96.98 | 98.80 |
| KHS23156 heat shock protein GrpE | 99.23 | 95.56 | 99.23 | 99.81 |
| KHS23251 membrane protein | 98.83 | 97.96 | 98.98 | 99.56 |
| KHS23252 membrane protein | 97.67 | 94.25 | 97.67 | 99.46 |
| KHS23254 methionine ABC transporter substrate-binding | 97.78 | 95.64 | 97.78 | 98.75 |
| KHS23255 metal ABC transporter permease | 99.14 | 96.09 | 99.14 | 99.86 |
| KHS23256 methionine ABC transporter ATP-binding protein | 97.12 | 95.13 | 97.02 | 99.21 |
| KHS23257 membrane protein | 97.43 | 97.70 | 97.54 | 97.66 |
| KHS23258 nucleotide-binding protein | 98.56 | 96.91 | 98.14 | 99.38 |
| KHS23269 type III secretion system effector protein XopAK | 94.94 | 88.47 | 94.94 | 97.12 |
| KHS23489 phosphomethylpyrimidine kinase | 98.15 | 95.67 | 98.15 | 99.88 |
| KHS23490 membrane protein | 98.51 | 95.23 | 98.06 | 99.85 |
| KHS23713 S-adenosylmethionine synthetase | 99.01 | 98.02 | 98.92 | 99.83 |
| KHS24197 molybdopterin-guanine dinucleotide biosynthesis | 97.54 | 94.02 | 97.36 | 99.47 |
| KHS24815 50S ribosomal protein L6 | 95.07 | 96.94 | 96.02 | 95.83 |
| KHS24816 30S ribosomal protein S8 | 98.24 | 99.25 | 98.24 | 98.24 |
| KHS24818 50S ribosomal protein L5 | 97.97 | 98.52 | 97.97 | 99.26 |
| KHS24819 50S ribosomal protein L24 | 100.00 | 99.68 | 100.00 | 100.00 |
| KHS24820 50S ribosomal protein L14 | 99.73 | 98.91 | 98.91 | 99.46 |
| KHS24821 30S ribosomal protein S17 | 99.26 | 98.14 | 98.88 | 99.63 |
| KHS24822 50S ribosomal protein L29 | 99.46 | 97.84 | 99.46 | 99.46 |
| KHS25207 histone-like nucleoid-structuring protein | 98.50 | 96.26 | 97.76 | 99.75 |
| KHS25209 membrane protein | 96.86 | 95.48 | 96.86 | 98.82 |
| KHS25211 3-methyladenine DNA glycosylase | 95.97 | 93.71 | 95.65 | 99.19 |
| KHS25388 EF hand domain-containing protein | 98.02 | 93.65 | 98.02 | 98.58 |
| KHS25405 thioredoxin | 99.12 | 97.36 | 98.83 | 100.00 |
| KHS25407 ABC transporter | 99.42 | 97.67 | 99.42 | 100.00 |
| KHS25490 cupin | 96.29 | 93.98 | 96.20 | 98.95 |
| KHS25537 peptide ABC transporter permease | 98.81 | 96.57 | 98.81 | 99.60 |
| KHS26053 RNA-binding protein | 98.35 | 96.69 | 99.17 | 99.59 |
| KHS26541 3-demethylubiquinone-9 3-methyltransferase | 97.47 | 98.61 | 97.47 | 97.61 |
| KHS26908 lytic transglycosylase | 95.53 | 86.24 | 94.99 | 99.11 |
| KHS26912 membrane protein | 97.87 | 96.17 | 97.77 | 99.36 |
| KHS26916 malto-oligosyltrehalose trehalohydrolase | 97.32 | 93.65 | 96.93 | 99.49 |
| KHS27359 S-(hydroxymethyl)glutathione dehydrogenase | 97.92 | 94.94 | 97.83 | 99.82 |

(Continued)

TABLE 2 | Continued

| GenBank accession number and predicted product of <i>Xap</i> NCPPB 1680 | <i>Xag</i> CFBP 2526 (%) | <i>Xam</i> IBSBF 1411 (%) | <i>Xap</i> NCPPB2064 (%) | <i>Xfa</i> ICPB 10535 (%) |
|--|---------------------------------|----------------------------------|---------------------------------|----------------------------------|
| KHS27633 Crp/Fnr family transcriptional regulator | 95.80 | 94.35 | 95.66 | 99.47 |
| KHS27636 ATP phosphoribosyltransferase | 96.28 | 95.30 | 96.50 | 99.67 |
| KHS27638 histidinol-phosphate aminotransferase | 96.70 | 90.87 | 97.07 | 99.45 |
| KHS27639 imidazoleglycerol-phosphate dehydratase | 95.21 | 93.17 | 95.12 | 99.65 |
| KHS27640 imidazole glycerol phosphate synthase | 95.18 | 93.52 | 95.02 | 99.67 |
| KHS28695 GntR family transcriptional regulator | 97.79 | 97.51 | 97.79 | 99.17 |
| KHS28696 vitamin B12 ABC transporter permease | 98.20 | 96.41 | 98.38 | 99.28 |
| KHS28992 NlpC-P60 family protein | 94.19 | 91.30 | 94.48 | 97.88 |
| KHS28993 ErfK/YbiS/YcfS/YnhG family protein | 96.65 | 91.75 | 96.13 | 99.48 |
| KHS29058 membrane protein | 98.59 | 95.07 | 97.54 | 100.00 |
| KHS29941 CDP-diacylglycerol-serine | 99.22 | 97.02 | 98.97 | 99.74 |
| KKY054115 membrane protein | 96.92 | 94.62 | 94.62 | 99.23 |
| KKY05325 membrane protein | 100.00 | 93.75 | 100.00 | 100.00 |

Each of the *Xap* NCPPB 1680 genes shares 100% identity over at least 95% of its length with its ortholog in *Xff* strains NCPPB 1058, 1495, and 1654. For comparison, percentage nucleotide sequence identities are given for each gene vs. *X. axonopodis* pv. *manihotis* (*Xam*), lablab-associated *X. axonopodis* pv. *phaseoli* (*Xap*).

study also reported that the *Xap* strains (Dol1, 2 and 3) were less pathogenic on common bean than was *Xap* NCPPB 302, hinting at the presence of distinct populations of *Xap* differentially adapted to different host species. Furthermore, a subsequent study found that *Xap* strain Dol 3, isolated from lablab in Medani, Sudan, 1965, was pathogenic only on common bean and lablab bean; it was not pathogenic on any of the other leguminous plants that were tested, including several *Vigna* spp., *Rhynchosia memnonia*, mungo bean, pigeon pea, alfalfa, butterfly pea, velvet bean, pea, and white lupin.

To the best of our knowledge, no recent quantitative data are available for the extent and severity of common bacterial blight on lablab. However, in 1959, leaf blight on this crop was reported as widespread and often severe in the Gezira and central Sudan (Sabet, 1959).

The single sequenced bacterial strain from Lima bean clearly fell within *Xap* GL 1, along with strains from common bean, including the pathovar type strain (NCPPB 3035). However, genome-wide phylogenetic reconstruction revealed that the Lima-associated strain was the most early-branching within this lineage and suggests that it has been genetically isolated from the population that is geographically widely dispersed on common bean (Figure 6). Again, the availability of these genomic data will facilitate development of PCR-based assays to rapidly genotype larger panels of strains to elucidate the population genetics.

The newly sequenced genomes confirm and extend previous observations (Alavi et al., 2008; Margueret et al., 2011; Egan et al., 2014), suggesting that a SPI-1-like T3SS is probably universal among *Xap* GL 1 but absent from *Xff* and from the newly discovered lablab-associated lineage. We also confirm that a frame-shift in the *hmgA* gene, resulting in a presumably defective homogentisate 1,2-dioxygenase, is common to all sequenced strains of *Xff* and probably explains the accumulation of brown pigment in fuscous strains (Darrasse et al., 2013b). The *hmgA* gene appeared to be intact in all the GL 1 and

lablab-associated strains consistent with the absence of report of pigment in these.

Previous comparative genomics studies of *Xanthomonas* species have highlighted the presence of rearrangements of fragments of the genome (Qian et al., 2005; Darrasse et al., 2013a). We observed no evidence of such rearrangements among the *Xff*, *Xap* GL1 nor among the lablab-associated *Xap* genomes sequenced in the present study (See Supplementary Figures S3–S5). However, the lack of evidence should not be interpreted as meaning that there are no such rearrangements; draft-quality genome assemblies, such as those generated in the present other related studies (Bart et al., 2012; Indiana et al., 2014; Schwartz et al., 2015), are fragmented into multiple contigs and/or scaffolds and if the breakpoints in the genomic rearrangements coincide with gaps or breakpoints in the assembly, then they would not be detected.

A previous study reported large genomic deletions in about 5% of the examined *Xanthomonas* strains, including *Xff* 4834-R, resulting in loss of flagellar motility (Darrasse et al., 2013b). Although, none of the genomes sequenced in the present study displayed this deletion, there were several other strain-specific multi-kilobase deletions (see Figure 1) suggesting that this is a relatively common phenomenon among xanthomonads.

Discussion

In the present study, we sequenced the genomes of 26 strains of the causative agents of CBB, whose times and places of isolation spanned several decades and several continents. This resource adds to the already published genome sequences of *Xff* 4834-R and CFBP 4884 (Darrasse et al., 2013b; Indiana et al., 2014) with a further 13 sequenced genomes. We also present the first genome sequences for *Xap*, including 9 strains belonging to a previously described lineage known as GL 1. These 9 sequenced GL 1 strains include 8 from common bean and one from Lima bean. We also sequenced a further four strains from lablab bean.

Our data reveal genetic sub-lineages within *Xff* and within *Xap* GL 1, each having a widely dispersed geographical distribution. The availability of these genome sequence data will be a useful source of genetic variation for use in developing molecular markers for distinguishing individual sub-lineages or genotypes and thus aiding the study of routes of pathogen spread (Vinatzer et al., 2014; Goss, 2015). We observed considerable intra-lineage variation with respect to gene content as well as single-nucleotide variations (Figures 4–7).

Whole-genome sequencing revealed the repertoires of predicted T3SS effectors. Our results (Figure 9) were consistent with a previous survey of effector genes (Hajri et al., 2009) except for two apparent discrepancies. First, we find no evidence for presence of *avrRxo1* (*xopAJ*) in the genomes of *Xap* nor *Xff* though Hajri and colleagues found this gene in *Xap* GL 1. Second, genome-wide sequencing sequencing was able to distinguish between *xopF1* and *xopF2*. We find *xopF1* in both *Xff* and *Xap* GL 1 but find *xopF2* only in *Xff*. Hajri reported presence of *xopF2* in both *Xff* and *Xap* GL1; this might be explained by cross-hybridisation of *xopF1* with the *xopF2* probes.

Arguably the most surprising finding to arise from the present study is the observation that *Xff* and *Xap* GL 1 share 100% identical alleles at dozens of loci even though on average most loci share only about 96% identity. This phenomenon is apparent from the bimodal distributions of sequence identities in Figure 11. This phenomenon is apparently restricted to sharing between GL 1 and *Xff*; no such bimodal distribution is seen between GL 1 and the lablab-associated strain nor between GL 1 and *X. fuscans* subsp. *aurantifolii* (which is

closely related to *Xff*). Furthermore, many of the alleles sharing 100% identity between GL 1 and *Xff* show significantly less identity between *Xap* GL 1 and *X. axonopodis* pv. *manihotis*, despite the close phylogenetic relationship between these last two. On the other hand, these shared sequences are more similar to sequences from *X. fuscans* subsp. *aurantifolii* than to sequence from *X. axonopodis* pv. *manihotis*, suggesting that they were acquired by *Xap* GL 1 from *Xff* rather than *vice versa*. Examples are listed in Table 2. It remains to be tested whether these alleles are adaptive for survival on a common ecological niche.

Acknowledgments

We are grateful for assistance from Jayne Hall, Erin Lewis and James Chisholm as well as the curators of NCPPB at Fera and Carlos Jara who kindly provided the CIAT strains and information about their origins. Genome sequencing was supported by the Canadian International Development Agency (Department of Foreign Affairs, Trade, and Development) grant to Pan-Africa Bean Research Alliance. VA was supported on this work by the BBSRC SCPRID Bean Grant BB/J011568/1. JH was supported by a BBSRC PhD studentship. DS was supported by BBSRC Grant BB/L012499/1.

Supplementary Material

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2015.01080>

References

- Alavi, S. M., Sanjari, S., Durand, F., Brin, C., Manceau, C., and Poussier, S. (2008). Assessment of the genetic diversity of *Xanthomonas axonopodis* pv. *phaseoli* and *Xanthomonas fuscans* subsp. *fuscans* as a basis to identify putative pathogenicity genes and a type III secretion system of the SPI-1 family by multiple suppression subtractive h. *Appl. Environ. Microbiol.* 74, 3295–3301. doi: 10.1128/AEM.02507-07
- Alfano, J. R., and Collmer, A. (1997). The type III (Hrp) secretion pathway of plant pathogenic bacteria: trafficking harpins, Avr proteins, and death. *J. Bacteriol.* 179, 5655–5662.
- Alfano, J. R., and Collmer, A. (2004). Type III secretion system effector proteins: double agents in bacterial disease and plant defense. *Annu. Rev. Phytopathol.* 42, 385–414. doi: 10.1146/annurev.phyto.42.040103.110731
- Alfano, J. R., and Collmer, A. (2011). BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12:402. doi: 10.1186/1471-2164-12-402
- Almeida, N. F., Yan, S., Cai, R., Clarke, C. R., Morris, C. E., Schaad, N. W., et al. (2010). PAMDB, a multilocus sequence typing and analysis database and website for plant-associated microbes. *Phytopathology* 100, 208–215. doi: 10.1094/PHYTO-100-3-0208
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Aronesty, E. (2011). ea-utils?: Command-line Tools for Processing Biological Sequencing Data. Available online at: <http://code.google.com/p/ea-utils>
- Mahdi, A. A., and Atabani, I. M. A. (1992). Response of Bradyrhizobium-inoculated soyabean and lablab bean to inoculation with vesicular-arbuscular mycorrhizae. *Exp. Agric.* 28, 399. doi: 10.1017/S001447970002010X
- Bankovich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021
- Bart, R., Cohn, M., Kassen, A., McCallum, E. J., Shybut, M., Petriello, A., et al. (2012). High-throughput genomic sequencing of cassava bacterial blight strains identifies conserved effectors to target for durable resistance. *Proc. Natl. Acad. Sci. U.S.A.* 109, E1972–E1979. doi: 10.1073/pnas.1208003109
- Birch, P. R. J., Hyman, L. J., Taylor, R., Opio, A. F., Bragard, C., and Toth, I. K. (1997). RAPD PCR-based differentiation of *Xanthomonas campestris* pv. *phaseoli* and *Xanthomonas campestris* pv. *phaseoli* var. *fuscans*. *Eur. J. Plant Pathol.* 103, 809–814.
- Bradbury, J. F. (1986). *Guide to Plant Pathogenic Bacteria*. CAB International Available online at: <http://www.cabdirect.org/abstracts/19871324885.html;jsessionid=D5C161DFDC715BFA77A16BC4EE1BF1AE> [Accessed July 19, 2014]
- Bull, C. T., De Boer, S. H., Denny, T. P., Firrao, G., Fischer-Le Saux, M., Saddler, G. S., et al. (2012). List of new names of plant pathogenic bacteria (2008–2010). *J. Plant Pathol.* 94, 21–27. doi: 10.4454/JPP.V96L2.026
- Clarke, C. R., Studholme, D. J., Hayes, B., Runde, B., Weisberg, A., Cai, R., et al. (2015). Genome-enabled phylogeographic investigation of the quarantine pathogen *Ralstonia solanacearum* race 3 biovar 2 and screening for sources of resistance against its core effectors. *Phytopathology* 105, 597–607. doi: 10.1094/PHYTO-12-14-0373-R
- Darling, A. C. E., Mau, B., Blattner, F. R., and Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14, 1394–1403. doi: 10.1101/gr.2289704
- Darrasse, A., Bolot, S., Serres-Giardi, L., Charbit, E., Boureau, T., Fisher-Le Saux, M., et al. (2013a). High-quality draft genome sequences of *Xanthomonas*

- axonopodis* pv. *glycines* Strains CFBP 2526 and CFBP 7119. *Genome Announc.* 1:e01036-13. doi: 10.1128/genomeA.01036-13
- Darrasse, A., Carrère, S., Barbe, V., Boureau, T., Arrieta-Ortiz, M. L., Bonneau, S., et al. (2013b). Genome sequence of *Xanthomonas fuscans* subsp. *fuscans* strain 4834-R reveals that flagellar motility is not a general feature of xanthomonads. *BMC Genomics* 14:761. doi: 10.1186/1471-2164-14-761
- Da Silva, A. C. R., Ferro, J. A., Reinach, F. C., Farah, C. S., Furlan, L. R., Quaggio, R. B., et al. (2002). Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature* 417, 459–463. doi: 10.1038/417459a
- 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
- Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461
- Egan, F., Barret, M., and O'Gara, F. (2014). The SPI-1-like Type III secretion system: more roles than you think. *Front. Plant Sci.* 5:34. doi: 10.3389/fpls.2014.00034
- Fargier, E., Saux, M. F.-L., and Manceau, C. (2011). A multilocus sequence analysis of *Xanthomonas campestris* reveals a complex structure within crucifer-attacking pathovars of this species. *Syst. Appl. Microbiol.* 34, 156–165. doi: 10.1016/j.syapm.2010.09.001
- Fourie, D. (2002). Distribution and severity of bacterial diseases on dry beans (*Phaseolus vulgaris* L.) in South Africa. *J. Phytopathol.* 150, 220–226. doi: 10.1046/j.1439-0434.2002.00745.x
- Fourie, D., and Herselman, L. (2011). Pathogenic and genetic variation in *Xanthomonas axonopodis* pv. *Phaseoli* and its *fuscans* variant in Southern Africa. *African Crop Sci. J.* 19, 393–407. doi: 10.4314/acsj.v19i4
- Galán, J. E., and Collmer, A. (1999). Type III secretion machines: bacterial devices for protein delivery into host cells. *Science* 284, 1322–1328.
- Gilbertson, R. L., Otoya, M. M., Pastor-Corrales, M. A., and Maxwell, D. P. (1991). Genetic diversity in common blight bacteria is revealed by cloned repetitive DNA sequences. *Ann. Rep. Bean Impr. Conf.* 34, 37–38.
- Goodwin, P. H., and Sopher, C. R. (1994). Brown pigmentation of *Xanthomonas campestris* pv. *phaseoli* associated with homogenetic acid. *Can. J. Microbiol.* 40, 28–34. doi: 10.1139/m94-005
- Goss, E. M. (2015). Genome-enabled analysis of plant-pathogen migration. *Annu. Rev. Phytopathol.* 53, 121–135. doi: 10.1146/annurev-phyto-080614-115936
- Grant, S. R., Fisher, E. J., Chang, J. H., Mole, B. M., and Dangl, J. L. (2006). Subterfuge and manipulation: type III effector proteins of phytopathogenic bacteria. *Annu. Rev. Microbiol.* 60, 425–449. doi: 10.1146/annurev.micro.60.080805.142251
- Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086
- Hajri, A., Brin, C., Hunault, G., Lardeux, F., Lemaire, C., Manceau, C., et al. (2009). A “repertoire for repertoire” hypothesis: repertoires of type three effectors are candidate determinants of host specificity in *Xanthomonas*. *PLoS ONE* 4:e6632. doi: 10.1371/journal.pone.0006632
- Hajri, A., Brin, C., Zhao, S., David, P., Feng, J.-X., Koebnik, R., et al. (2012). Multilocus sequence analysis and type III effector repertoire mining provide new insights into the evolutionary history and virulence of *Xanthomonas oryzae*. *Mol. Plant Pathol.* 13, 288–302. doi: 10.1111/j.1364-3703.2011.00745.x
- Hunt, M., Kikuchi, T., Sanders, M., Newbold, C., Berriman, M., and Otto, T. D. (2013). REAPR: a universal tool for genome assembly evaluation. *Genome Biol.* 14:R47. doi: 10.1186/gb-2013-14-5-r47
- Indiana, A., Briand, M., Arlat, M., Gagnevin, L., Koebnik, R., Noël, L. D., et al. (2014). Draft Genome Sequence of the Flagellated *Xanthomonas fuscans* subsp. *fuscans* Strain CFBP 4884. *Genome Announc.* 2, e00966–14. doi: 10.1128/genomeA.00966-14
- Jones, S. B. (1987). Bacterial pustule disease of soybean: microscopy of pustule development in a susceptible cultivar. *Phytopathology* 77:266. doi: 10.1094/Phyto-77-266
- Kay, S., and Bonas, U. (2009). How *Xanthomonas* type III effectors manipulate the host plant. *Curr. Opin. Microbiol.* 12, 37–43. doi: 10.1016/j.mib.2008.12.006
- Li, H. (2013). Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM. 3. Available online at: <http://arxiv.org/abs/1303.3997> [Accessed July 20, 2014].
- Li, H. (2014). Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics* 30, 1–9. doi: 10.1093/bioinformatics/btu356
- 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, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Magoč, T., and Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27, 2957–2963. doi: 10.1093/bioinformatics/btr507
- Margueretaz, M., Pieretti, I., Gayral, P., Puig, J., Brin, C., Cocianich, S., et al. (2011). Genomic and evolutionary features of the SPI-1 type III secretion system that is present in *Xanthomonas* albilineans but is not essential for xylem colonization and symptom development of sugarcane leaf scald. *Mol. Plant Microbe Interact.* 24, 246–259. doi: 10.1094/MPMI-08-10-0188
- Mazzaglia, A., Studholme, D. J., Taratufolo, M. C., Cai, R., Almeida, N. F., Goodman, T., et al. (2012). *Pseudomonas syringae* pv. *actinidiae* (PSA) strains from recent bacterial canker of kiwifruit outbreaks belong to the same genetic lineage. *PLoS ONE* 7:e36518. doi: 10.1371/journal.pone.0036518
- Mhedbi-Hajri, N., Hajri, A., Boureau, T., Darrasse, A., Durand, K., Brin, C., et al. (2013). Evolutionary history of the plant pathogenic bacterium *Xanthomonas axonopodis*. *PLoS ONE* 8:e58474. doi: 10.1371/journal.pone.0058474
- Moreira, L. M., Almeida, N. F., Potnis, N., Digiampietri, L. A., Adi, S. S., Bortolossi, J. C., et al. (2010). Novel insights into the genomic basis of citrus canker based on the genome sequences of two strains of *Xanthomonas fuscans* subsp. *aurantifoliae*. *BMC Genomics* 11:238. doi: 10.1186/1471-2164-11-238
- Nei, M., and Kumar, S. (2000). *Molecular Evolution and Phylogenetics*. Oxford University Press. Available online at: <https://books.google.com/books?hl=en&lr=&id=vtWW9bmVd1IC&pgis=1> [Accessed May 13, 2015].
- Parkinson, N., Cowie, C., Heeney, J., and Stead, D. (2009). Phylogenetic structure of *Xanthomonas* determined by comparison of *gyrB* sequences. *Int. J. Syst. Evol. Microbiol.* 59, 264–274. doi: 10.1099/ijsm.0.65825-0
- Pieretti, I., Pesic, A., Petras, D., Royer, M., Süßmuth, R. D., and Cocianich, S. (2015). What makes *Xanthomonas* albilineans unique amongst xanthomonads? *Front. Plant Sci.* 6:289. doi: 10.3389/fpls.2015.00289
- Pieretti, I., Royer, M., Barbe, V., Carrere, S., Koebnik, R., Cocianich, S., et al. (2009). The complete genome sequence of *Xanthomonas* albilineans provides new insights into the reductive genome evolution of the xylem-limited *Xanthomonadaceae*. *BMC Genomics* 10:616. doi: 10.1186/1471-2164-10-616
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS ONE* 5:e9460. doi: 10.1371/journal.pone.0009490
- Qian, W., Jia, Y., Ren, S. X., He, Y. Q., Feng, J., Lu, L., et al. (2005). Comparative and functional genomic analyses of the pathogenicity of phytopathogen. *Genome Res.* 15, 757–767. doi: 10.1101/gr.3378705
- Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033
- Rademaker, J. L. W., Louws, F. J., Schultz, M. H., Rossbach, U., Vauterin, L., Swings, J., et al. (2005). A comprehensive species to strain taxonomic framework for xanthomonads. *Phytopathology* 95, 1098–1111. doi: 10.1094/PHYTO-95-1098
- R Development Core Team, R. (2013). R: a language and environment for statistical computing. *R Found. Stat. Comput.* 1, 409. doi: 10.1007/978-3-540-74686-7
- Hamza, A. A., Robene-Soustrade, I., Jouen, E., Lefevre, P., Chiroleu, F., Fisher-Le Saux, M., et al. (2012). MultiLocus sequence analysis- and amplified fragment length polymorphism-based characterization of xanthomonads associated with bacterial spot of tomato and pepper and their relatedness to *Xanthomonas* species. *Syst. Appl. Microbiol.* 35, 183–190. doi: 10.1016/j.syapm.2011.12.005
- Rodriguez-R, L. M., Grajales, A., Arrieta-Ortiz, M., Salazar, C., Restrepo, S., and Bernal, A. (2012). Genomes-based phylogeny of the genus *Xanthomonas*. *BMC Microbiol.* 12:43. doi: 10.1186/1471-2180-12-43
- Ryan, R. P., Vorhölter, F.-J., Potnis, N., Jones, J. B., Van Sluys, M.-A., Bogdanove, A. J., et al. (2011). Pathogenomics of *Xanthomonas*: understanding bacterium-plant interactions. *Nat. Rev. Microbiol.* 9, 344–355. doi: 10.1038/nrmicro2558

- Sabet, K. A. (1959). Studies in the bacterial diseases of Sudan crops III. On the occurrence, host range and taxonomy of the bacteria causing leaf blight diseases of certain leguminous plants. *Ann. Appl. Biol.* 47, 318–331.
- Schaad, N. W., Postnikova, E., Lacy, G. H., Sechler, A., Agarkova, I., Stromberg, P. E., et al. (2005). Reclassification of *Xanthomonas campestris* pv. *citri* (ex Hasse 1915) Dye 1978 forms A, B/C/D, and E as *X. smithii* subsp. *citri* (ex Hasse) sp. nov. nom. rev. comb. nov., *X. fuscans* subsp. *aurantifoli* (ex Gabriel 1989) sp. nov. nom. rev. comb. nov., and *X. Syst. Appl. Microbiol.* 28, 494–518. doi: 10.1016/j.syapm.2005.03.017
- Schaaffhausen, R. V. (1963). Dolichos lablab or hyacinth bean: - Its uses for feed, food and soil improvement. *Econ. Bot.* 17, 146–153. doi: 10.1007/BF02985365
- Schwartz, A. R., Potnis, N., Timilsina, S., Wilson, M., Patane, J., Martins, J. J., et al. (2015). Phylogenomics of *Xanthomonas* field strains infecting pepper and tomato reveals diversity in effector repertoires and identifies determinants of host specificity. *Front. Microbiol.* 6:535. doi: 10.3389/fmicb.2015.00535
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- Tarr, S. A. J. (1958). Experiments in the Sudan Gezira on control of wilt of dolichos bean (Dolichos lablab) associated with attack by cockshafer grubs (*Scizonyxa* sp.). *Ann. Appl. Biol.* 46, 630–638. doi: 10.1111/j.1744-7348.1958.tb02246.x
- Taylor, J. D., Teverson, D. M., Allen, D. J., and Pastor Corrales, M. A. (1996). Identification and origin of races of *Pseudomonas syringae* pv. *phaseolicola* from Africa and other bean growing areas. *Plant Pathol.* 45, 469–478. doi: 10.1046/j.1365-3059.1996.d01-147.x
- Thorvaldsdóttir, H., Robinson, J. T., and Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings Bioinforma.* 14, 178–192. doi: 10.1093/bib/bbs017
- Toth, I. K., Hyman, L. J., Taylor, R., and Birch, P. R. J. (1998). PCR-based detection of *Xanthomonas campestris* pv. *phaseoli* var. *fuscans* in plant material and its differentiation from *X. c.* pv. *phaseoli*. *J. Appl. Microbiol.* 85, 327–336. doi: 10.1046/j.1365-2672.1998.00514.x
- Treangen, T. J., Ondov, B. D., Koren, S., and Phillippy, A. M. (2014). The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol.* 15:524. doi: 10.1186/s13059-014-0524-x
- Vinatzer, B. A., Montel, C. L., and Clarke, C. R. (2014). Harnessing population genomics to understand how bacterial pathogens emerge, adapt to crop hosts, and disseminate. *Annu. Rev. Phytopathol.* 52, 19–43. doi: 10.1146/annurev-phyto-102313-045907
- Wasukira, A., Tayebwa, J., Thwaites, R., Paszkiewicz, K., Aritua, V., Kubiriba, J., et al. (2012). Genome-wide sequencing reveals two major sub-lineages in the genetically monomorphic pathogen *xanthomonas campestris* pathovar *musacearum*. *Genes (Basel)* 3, 361–377. doi: 10.3390/genes3030361
- Young, J. M., Park, D.-C., Shearman, H. M., and Fargier, E. (2008). A multilocus sequence analysis of the genus *Xanthomonas*. *Syst. Appl. Microbiol.* 31, 366–377. doi: 10.1016/j.syapm.2008.06.004

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 Aritua, Harrison, Sapp, Buruchara, Smith and Studholme. 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.

OPEN ACCESS

Edited by:

Laurent D. Noël,
Centre National de la Recherche
Scientifique, France

Reviewed by:

Peter Dodds,
Commonwealth Scientific and
Industrial Research Organisation,
Australia

David John Studholme,
University of Exeter, UK

***Correspondence:**

Brian J. Staskawicz,
Department of Plant and Microbial
Biology, University of California,
Berkeley, 241 Koshland Hall, Berkeley,
CA 94705, USA
stask@berkeley.edu

[†]These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Plant-Microbe Interaction,
a section of the journal
Frontiers in Microbiology

Received: 03 April 2015

Accepted: 15 May 2015

Published: 03 June 2015

Citation:

Schwartz AR, Potnis N, Timilsina S,
Wilson M, Patané J, Martins J Jr,
Minsavage GV, Dahlbeck D, Akhunova
A, Almeida N, Vallad GE, Barak JD,
White FF, Miller SA, Ritchie D, Goss E,
Bart RS, Setubal JC, Jones JB and
Staskawicz BJ (2015) Phylogenomics
of *Xanthomonas* field strains infecting
pepper and tomato reveals diversity in
effector repertoires and identifies
determinants of host specificity.
Front. Microbiol. 6:535.
doi: 10.3389/fmicb.2015.00535

Phylogenomics of *Xanthomonas* field strains infecting pepper and tomato reveals diversity in effector repertoires and identifies determinants of host specificity

Allison R. Schwartz^{1†}, Neha Potnis^{2†}, Sujan Timilsina², Mark Wilson³, José Patané⁴, Joaquim Martins Jr.⁴, Gerald V. Minsavage², Douglas Dahlbeck¹, Alina Akhunova⁵, Nalvo Almeida⁶, Gary E. Vallad⁷, Jeri D. Barak⁸, Frank F. White⁵, Sally A. Miller⁹, David Ritchie¹⁰, Erica Goss², Rebecca S. Bart³, João C. Setubal^{4,11}, Jeffrey B. Jones² and Brian J. Staskawicz^{1*}

¹ Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA, USA, ² Department of Plant Pathology, University of Florida, Gainesville, FL, USA, ³ Donald Danforth Plant Science Center, St. Louis, MO, USA,

⁴ Department of Biochemistry, Institute of Chemistry, University of São Paulo, São Paulo, Brazil, ⁵ Department of Plant Pathology, Kansas State University, Manhattan, KS, USA, ⁶ School of Computing, Federal University of Mato Grosso do Sul, Campo Grande, Brazil, ⁷ Gulf Coast Research and Education Center, University of Florida, Wimauma, FL, USA, ⁸ Department of Plant Pathology, University of Wisconsin, Madison, Madison, WI, USA, ⁹ Department of Plant Pathology, Ohio Agricultural Research and Development Center, Wooster, MA, USA, ¹⁰ Department of Plant Pathology, NC State University, Raleigh, NC, USA, ¹¹ Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, VA, USA

Bacterial spot disease of pepper and tomato is caused by four distinct *Xanthomonas* species and is a severely limiting factor on fruit yield in these crops. The genetic diversity and the type III effector repertoires of a large sampling of field strains for this disease have yet to be explored on a genomic scale, limiting our understanding of pathogen evolution in an agricultural setting. Genomes of 67 *Xanthomonas euvesicatoria* (Xe), *Xanthomonas perforans* (Xp), and *Xanthomonas gardneri* (Xg) strains isolated from diseased pepper and tomato fields in the southeastern and midwestern United States were sequenced in order to determine the genetic diversity in field strains. Type III effector repertoires were computationally predicted for each strain, and multiple methods of constructing phylogenies were employed to understand better the genetic relationship of strains in the collection. A division in the Xp population was detected based on core genome phylogeny, supporting a model whereby the host-range expansion of Xp field strains on pepper is due, in part, to a loss of the effector AvrBsT. Xp-host compatibility was further studied with the observation that a double deletion of AvrBsT and XopQ allows a host range expansion for *Nicotiana benthamiana*. Extensive sampling of field strains and an improved understanding of effector content will aid in efforts to design disease resistance strategies targeted against highly conserved core effectors.

Keywords: *Xanthomonas*, type III effector repertoire, phylogenomics, host specificity, bacterial spot disease, AvrBsT, XopQ

Introduction

Species of *Xanthomonas* cause bacterial spot disease on cultivated pepper (*Capsicum annuum*) and tomato (*Solanum lycopersicum*) and are the most devastating to crops grown in warm, humid climates such as in the southeastern and midwestern United States (Obradovic et al., 2008). Once considered a single species, *Xanthomonas vesicatoria* infecting pepper and tomato has been reclassified several times (Stall et al., 1994; Vauterin et al., 1995; Jones and Stall, 2000), but was most recently separated into four distinct species: *X. euvesicatoria* (*Xe*), *X. vesicatoria* (*Xv*), *X. perforans* (*Xp*), and *X. gardneri* (*Xg*) (Jones et al., 2004). While *Xe*, *Xg*, and *Xv* infect both pepper and tomato, *Xp* has only been reported on tomato. Although the four pathogens are present and destructive on a global scale (Jones et al., 1998a; Timilsina et al., 2015), the history and distribution of *Xe*, *Xp*, and *Xg* has changed dramatically in the United States, particularly with the emergence of *Xp* as the dominant tomato pathogen over *Xe* in Florida beginning in the early 1990's (Jones et al., 1998b; Tudor-Nelson et al., 2003; Hert et al., 2005; Stall et al., 2009; Horvath et al., 2012) and *Xg* as a major tomato pathogen in Ohio and Michigan beginning in 2009 (Ma et al., 2011). Outbreaks of *Xv* have not been reported in the United States (Timilsina et al., 2015).

Different phylogenetic analyses found a close evolutionary relationship between *Xe* and *Xp* in comparison to *Xg* and *Xv* (Young et al., 2008; Parkinson et al., 2009; Almeida et al., 2010; Hamza et al., 2010; Midha and Patil, 2014). Comparative genomics of reference strains *Xe*85-10 (Thieme et al., 2005) with *Xp*91-118, *Xg*101, and *Xv*1111 (Potnis et al., 2011) provided the first insights into the shared and unique virulence factors of these pepper and tomato pathogens. A major factor contributing to the virulence and host specificity of these pathogens is the repertoire of effectors secreted into the host plant cell via the type III secretion system (Grant et al., 2006). *Xanthomonads* have evolved effectors with diverse mechanisms to promote virulence, even adopting processes specific to eukaryotes (Kay and Bonas, 2009). The recognition of specific effector proteins by specific cognate resistance (R) proteins leads to defense responses that have been termed Effector Triggered Immunity (ETI), which is accompanied by localized cell death, associated tissue collapse known as the hypersensitive response (HR) at the site of infection, and limited spread of the pathogen (Jones and Dangl, 2006). Several type III effectors are conserved across multiple species and referred to here as core effectors. An additional variable set of effectors may provide specialization to specific hosts and cultivars (Hajri et al., 2009).

The deployment of R proteins in crops that can recognize and respond to core effectors is a potentially durable disease resistance strategy, depending on the evolutionary stability of the targeted cognate effector (Boyd et al., 2013). Because *Xanthomonads* display relatively high genome plasticity, a more comprehensive understanding of the genetic diversity of pepper and tomato pathogens, with specific emphasis on effectors, is necessary for designing informed disease resistance strategies for agricultural areas afflicted by bacterial spot disease (Thieme et al., 2005; Potnis et al., 2011; Timilsina et al., 2015). A

comparative genomic analysis considering many strains from a given geographic region over time will provide a representative view of the effectors present in the regional bacterial population and add insight into the evolutionary trends of effectors, and thus their potential usefulness as targets for R-gene mediated resistance strategies.

To this end we sequenced the genomes of 32 *Xp*, 25 *Xe*, and 10 *Xg* field strains that were collected from diseased peppers and tomatoes in the southeastern and midwestern United States. Here we describe the genetic diversity within and between species using core protein-coding genome phylogeny and whole genome single nucleotide polymorphism (SNP) analysis and present the computationally predicted type III effector repertoires of strains in our collection. The role played by the effectors AvrBsT and XopQ as host specificity determinants for *Xp* infecting pepper and *Nicotiana benthamiana* was also characterized.

Materials and Methods

Xanthomonas Strain Collection

Xe, *Xp*, and *Xg* strains were collected from diseased tomatoes and peppers in the United States (Table 1). *Xp* strains were collected between 1998 and 2013 in Florida and Georgia. *Xg* strains were collected in Ohio and Michigan between 2010 and 2012. *Xe* strains were collected between 1994 and 2012 in Florida, North Carolina, Georgia, and Kentucky.

Genome Sequencing and Effector Predictions

Bacterial genome sequencing and effector prediction were completed as previously described (Bart et al., 2012). Briefly, genomic DNA was isolated with a modified CTAB protocol and prepared for library construction and sequencing on the Illumina platforms. Ten *Xg* libraries were pooled into a single lane of MiSeq (PE250). *Xe* and the *Xp* strains from 2006 were sequenced by multiplexing 48 libraries per lane on an Illumina HiSeq 2000 sequencer (PE100). The *Xp* strains from 2012 were sequenced by multiplexing 20 libraries per lane on an Illumina MiSeq (PE150). Genomic *de novo* assemblies were constructed using CLC Genomics Workbench using a length fraction of 0.9 and a similarity of 1.0. Potential effectors were identified by an in-house Python script utilizing BLAST against a database of known effectors, using a filter of greater than 45% amino acid similarity over 80% of the length of the target sequence (Bart et al., 2012).

Phylogenomic Inference Using Core Protein-coding Genes

All genomes sequenced in this study were annotated using the National Center for Biotechnology Information Prokaryotic Genome Annotation Pipeline (PGAP) (<http://www.ncbi.nlm.nih.gov/books/NBK174280>). Ortholog families were determined using the GET_HOMOLOGUES package (Contreras-Moreira and Vinuesa, 2013), which includes a step of all-against-all BlastP (Altschul et al., 1997) followed by clustering based on OrthoMCL to yield homologous gene clusters (Li et al., 2003). This result was filtered using compare_cluster.pl (a script in the GET_HOMOLOGUES package) with option "-t n," where *n* is the number of genomes, keeping only the gene families that have

TABLE 1 | Summary of *Xanthomonas* field strains sequenced in this paper.

| Species | Strain name | Origin in US | Host isolated | Year isolated | Isolation ID | Collector |
|-------------------------|-------------|--------------------|---------------|---------------|--------------|-----------|
| <i>X. euvesicatoria</i> | Xe073 | North Carolina | P | 1994 | 181 | DR |
| | Xe074 | Raleigh, NC | P | 1994 | 199 | DR |
| | Xe075 | Southwest FL | P | 1995 | 206 | DR |
| | Xe076 | Naples, FL | P | 1995 | 259 | DR |
| | Xe077 | Kentucky | P | 1996 | 315 | DR |
| | Xe078 | Clewiston, FL | P | 1997 | 329 | DR |
| | Xe079 | Jupiter, FL | P | 1998 | 354 | DR |
| | Xe081 | Ft. Pierce, FL | P | 1995 | 376 | DR |
| | Xe082 | Southeast FL | P | 1998 | 455 | DR |
| | Xe083 | Belle Glade, FL | P | 1999 | 490 | DR |
| | Xe085 | Boynton Beach, FL | P | 1999 | 515 | DR |
| | Xe086 | Delray Beach, FL | P | 2000 | 526 | DR |
| | Xe091 | Boca Raton, FL | P | 2003 | 586 | DR |
| | Xe101 | Sampson Co., NC | P | 2008 | 678 | DR |
| | Xe102 | Manatee, FL | P | 2008 | 679 | DR |
| | Xe103 | Pender Co., NC | P | 2009 | 681 | DR |
| | Xe104 | Sampson Co., NC | P | 2010 | 683 | DR |
| | Xe105 | Granville, NC | P | 2010 | 684 | DR |
| | Xe106 | Granville, NC | P | 2010 | 685 | DR |
| | Xe107 | Granville, NC | P | 2011 | 689 | DR |
| | Xe108 | Pender Co., NC | P | 2012 | 695 | DR |
| | Xe109 | Cook Co., GA | P | 2004 | F4-2 | DR |
| | Xe110 | Tift Co. GA | P | 2004 | G4-1 | DR |
| | Xe111 | Colquitt Co., GA | P | 2004 | H3-2 | DR |
| | Xe112 | Brooks Co., GA | P | 2004 | L3-2 | DR |
| <i>X. perforans</i> | Xp4B | Citra, FL | T | 1998 | Xp4B | JJ |
| | Xp2010 | Hendry County, FL | P | 2010 | Xp2010 | JJ |
| | TB6 | Hillsborough, FL | T | 2013 | TB6 | JJ |
| | TB9 | Hillsborough, FL | T | 2013 | TB9 | JJ |
| | TB15 | Hillsborough, FL | T | 2013 | TB15 | JJ |
| | Xp3-15 | Decatur Co., GA | T | 2006 | Xp3-15 | JJ |
| | Xp4-20 | Decatur Co., GA | T | 2006 | Xp4-20 | JJ |
| | Xp5-6 | Decatur Co., GA | T | 2006 | Xp5-6 | JJ |
| | Xp7-12 | Manatee Co., FL | T | 2006 | Xp7-12 | JJ |
| | Xp8-16 | Manatee Co., FL | T | 2006 | Xp8-16 | JJ |
| | Xp9-5 | Manatee Co., FL | T | 2006 | Xp9-5 | JJ |
| | Xp10-13 | Manatee Co., FL | T | 2006 | Xp10-13 | JJ |
| | Xp11-2 | Palm Beach Co., FL | T | 2006 | Xp11-2 | JJ |
| | Xp15-11 | Miami-Dade Co., FL | T | 2006 | Xp15-11 | JJ |
| | Xp17-12 | Collier Co., FL | T | 2006 | Xp17-12 | JJ |
| | Xp18-15 | Collier Co., FL | T | 2006 | Xp18-15 | JJ |
| | GEV839 | Hardee Co., FL | T | 2012 | GEV839 | JJ |
| | GEV872 | Immokalee, FL | T | 2012 | GEV872 | JJ |
| | GEV893 | Collier Co. | T | 2012 | GEV893 | JJ |
| | GEV904 | Hillsborough, FL | T | 2012 | GEV904 | JJ |
| | GEV909 | Collier Co. | T | 2012 | GEV909 | JJ |
| | GEV915 | Hillsborough, FL | T | 2012 | GEV915 | JJ |
| | GEV917 | Hillsborough, FL | T | 2012 | GEV917 | JJ |
| | GEV936 | Lee, FL | T | 2012 | GEV936 | JJ |
| | GEV940 | GCREC, FL | T | 2012 | GEV940 | JJ |

(Continued)

TABLE 1 | Continued

| Species | Strain name | Origin in US | Host isolated | Year isolated | Isolation ID | Collector |
|--------------------|-------------|------------------|---------------|---------------|--------------|-----------|
| | GEV968 | Manatee Co., FL | T | 2012 | GEV968 | JJ |
| | GEV993 | Hendry Co., FL | T | 2012 | GEV993 | JJ |
| | GEV1001 | Quincy, FL | T | 2012 | GEV1001 | JJ |
| | GEV1026 | West Coast, FL | T | 2012 | GEV1026 | JJ |
| | GEV1044 | Collier Co., FL | T | 2012 | GEV1044 | JJ |
| | GEV1054 | Manatee Co., FL | T | 2012 | GEV1054 | JJ |
| | GEV1063 | Collier Co., FL | T | 2012 | GEV1063 | JJ |
| <i>X. gardneri</i> | Xg153 | Gibsonburg, OH | T | 2010 | SM194-10 | SM |
| | Xg156 | Blissfield, MI | T | 2010 | SM177-10 | SM |
| | Xg157 | Blissfield, MI | T | 2010 | SM182-10 | SM |
| | Xg159 | Blissfield, MI | T | 2010 | SM220-10 | SM |
| | Xg160 | Blissfield, MI | T | 2010 | SM234-10 | SM |
| | Xg164 | Ottawa, OH | T | 2011 | SM406-11 | SM |
| | Xg165 | Ottawa, OH | T | 2011 | SM413-11 | SM |
| | Xg173 | Unknown, OH | T | 2011 | SM605-11 | SM |
| | Xg174 | Wayne Co., OH | T | 2012 | SM775-12 | SM |
| | Xg177 | Sandusky Co., OH | T | 2012 | SM795-12 | SM |

Year, host, location, and isolation source are described. DR, David Ritchie; JJ, Jeffrey Jones; SM, Sally Miller.

exactly one representative from each genome considered; the protein-coding genes in these families were considered the “core genome” of these species.

Accuracy checking of each individual gene alignment (using nucleotide sequences) was performed by Guidance (Penn et al., 2010) using the Mafft algorithm (Katoh et al., 2002) anchored by codons with default options, followed by the removal of low-accuracy alignment sites. All edited alignments were concatenated by FASconCAT yielding a nucleotide supermatrix (Kück and Meusemann, 2010). The best partitioning scheme and evolutionary model for each partition were calculated by PartitionFinder (Lanfear et al., 2012), which tests all available models under the Bayesian Information Criterion (BIC) selection procedure (Lanfear et al., 2014). Maximum likelihood (ML) analysis for phylogeny construction was performed using IQTree v.1.1.5 assuming the best partitioning and respective models according to the previous step (Nguyen et al., 2015). A total of 1000 bootstrap pseudoreplicates were performed to assess clade support. Additional taxa included to strengthen the confidence in the phylogenetic relationships are as follows: *Xanthomonas fragariae* (XfrLMG25863, RefSeq PRJNA80793: Vandroemme et al., 2013), *Xanthomonas arboricola* pv. *corylina* (XacNCCB100457, RefSeq PRJNA193452: Ibarra Caballero et al., 2013), *Xanthomonas campestris* pv. *musacearum* (XcmNCPB4384, RefSeq PRJNA73881: Wasukira et al., 2012), *Xanthomonas axonopodis* pv. *citrumelo* F1 (XalfaF1, RefSeq PRJNA73179: Jalan et al., 2011), *Xanthomonas oryzae* pv. *oryzae* (XooKACC10331, RefSeq PRJNA12931: Lee et al., 2005), *Xanthomonas campestris* pv. *campestris* (XccATCC33913, RefSeq PRJNA57887: da Silva et al., 2002), *Xanthomonas euvesicatoria* (also *Xanthomonas campestris* pv. *vesicatoria*, Xe85-10, RefSeq PRJNA58321: Thieme et al., 2005).

Whole Genome SNP Analysis

Illumina reads were trimmed using Trimmomatic version 0.32 (Bolger et al., 2014) and were then mapped to the reference genome *Xanthomonas axonopodis* pv. *citri* strain 306 (Xac306, NC_003919: da Silva et al., 2002) using bowtie2 version 2.1.0 (Langmead and Salzberg, 2012). The Best Practices guidelines of the Broad Institute for variant calling were followed (<https://www.broadinstitute.org/gatk/guide/best-practices>). MarkDuplicates from Picard Tools version 1.118 was used to mark duplicate reads. RealignerTargetCreator and IndelRealigner from GenomeAnalysisToolkit (GATK) version 3.3-0 were used to verify reads were aligned properly (McKenna et al., 2010). HaplotypeCaller from GATK was used to discover variants. SNPs were concatenated as previously described (Bart et al., 2012). A ML phylogenetic tree with bootstrap values was created using RAxML version 8.0 (Stamatakis, 2014).

Effector Allele Analysis

Effectors were compared within each species at the amino acid sequence level for *Xp* and the nucleotide level for *Xe* and *Xg*, and each distinct allele was assigned a number. Neighbor-joining trees were constructed to visualize differences in effector profiles among strains in each species. Simple genetic distances among strains in their effector profiles were calculated for all pairwise comparisons within each species, such that a difference at one effector between two strains equaled a distance of 1.0 and a difference at five effectors equaled a distance of 5.0. *Xp* calculations included an outgroup profile from Xe85-10. Distance was calculated using GenAlEx 6.501. Distance matrices were exported to MEGA format and trees were constructed in MEGA 6.06 (Tamura et al., 2013).

Confirmation of the TAL Effector AvrHah1 in Xg

Xg strains were infiltrated into pepper cv. ECW30R at OD₆₀₀ = 0.3 in order to determine if activation of the *Bs3* resistance gene occurs in response to AvrHah1. Negative and positive controls for AvrHah1 in Xg are strain 1782 and 04T5, respectively (Schornack et al., 2008). Pictures were taken 48 h post-infiltration (hpi). For Southern blot analysis, 5 µg of Xg DNA (extracted as described above) was restriction digested for 2 h with BamHI and run on a 0.7% agarose gel. DNA was transferred overnight to a Hybond-N+ membrane and hybridized overnight with a P³²-labeled probe for the first 705bp of AvrHah1. The size of the predicted BamHI-digested AvrHah1 fragment is 2964bp.

Effector Deletion

Insertion mutants in *Xp* strains (Ω avrBsT) were constructed using site-directed homologous recombination of a partial fragment linked to a gene for antibiotic resistance. Intragenetic partial fragments (approximately 500 bp) of each targeted gene were PCR amplified and cloned into the pCR2.1 TOPO-vector using the TA cloning method (Invitrogen). Positive clones were confirmed by Sanger sequencing. The plasmids were introduced into competent cells of *Xp* recipient strains by electroporation, and transformed cells were selected for kanamycin resistance (kan^R). Single homologous recombination events (due to the integration of the TOPO plasmid containing a portion of the respective gene) disrupted the gene of interest (Sugio et al., 2005). Mutations were confirmed by PCR using a primer flanking the upstream region of the targeted gene and the M13 Forward primer (pCR2.1 TOPO internal primer), followed by Sanger sequencing.

Whole gene knockout strains Xe85-10ΔXopQ, Xg153ΔhrcV, Xp4BΔAvrBsT, and Xp4BΔXopQΔAvrBsT were constructed using the suicide vector pLVC18 containing the contiguous 1kb upstream and 1kb downstream fragments flanking the targeted gene (Lindgren et al., 1986). Double homologous recombination events resulting in markerless deletions were confirmed by PCR or southern blot. Gene deletions were complemented back by conjugation of the stable broad host range plasmid pVSP61 (kan^R) containing the native promoter and the open reading frame of each respective gene.

Inoculation Conditions

Xanthomonas strains were grown on nutrient yeast glycerol agar (NYGA) supplemented, as appropriate, with 100 µg/ml rifampicin (wild type and deletion strains) and 25 µg/ml kanamycin. Strains were incubated at 28°C for 48–72 h. Cells were washed from agar plates with 10mM MgCl₂, and the concentration was adjusted as necessary. For growth assays, leaves were syringe-infiltrated with bacterial suspensions of 10⁵ CFU/mL. For virulence scoring, leaves were syringe infiltrated at 10⁸ CFU/mL and pictures were taken 48 h post-infiltration (hpi) after submerging leaves in water for 10 min to enhance any water-soaked phenotypes. For lesions assays, leaves were syringe infiltrated at 10⁴ CFU/mL and pictures were taken 8–10 days post-infiltration (dpi) after submerging leaves in water for 10 min.

Results

Genome Submission

Draft genome sequences of 32 *Xp*, 25 *Xe* and 10 *Xg* field strains, respectively, from diseased peppers and tomatoes in the United States were obtained by Illumina sequencing (Table 1). Genome assembly statistics for each strain and average *de novo* assembly statistics for *Xe*, *Xp*, and *Xg* are presented in Supplemental Tables 1 and 2, respectively. Draft genome sequences have been deposited in the National Center for Biotechnology (Supplemental Table 1).

Core Genome Phylogenetic Analysis Identifies a Division in the *Xp* Population

The core genome for all three species was identified by sequence similarity, yielding 1152 protein-coding gene families, of which 1017 were considered bona fide orthologs; 135 families were discarded as spurious alignments by the program Guidance. The 1017 families were concatenated, yielding a supermatrix of 916,326 sites. The best partitioning scheme chosen was by codon position in which first, second and third positions are set as separated partitions. The best evolutionary models for each partition were respectively GTR+I+G for the first and second partitions, and TVM+I+G for the third partition.

The Maximum Likelihood (ML) phylogeny based on core genome orthologs displays *Xe*, *Xp*, and *Xg* behaving as separate monophyletic groups (Figure 1). Our results mirrored previous studies, *Xe* and *Xp* being closely related, and *Xg* more distant phylogenetically, with all three species forming monophyletic groups. For *Xp* strains, this analysis showed a division, which we define here as Group 1—further divided into Group 1A and 1B—and Group 2. Group 1A comprises 11 strains (out of 16) from 2012 that form a monophyletic clade (branches in purple). Other strains belonging to Group 1 are defined here as Group 1B (branches in orange), which includes the reference strain Xp91-118, Xp4B (isolated in 1998), and six strains isolated in 2006. Group 1B does not contain any strains isolated in 2012. We define 14 strains as Group 2 (branches in green) which includes five strains from 2006, the single strain from 2010, five strains from 2012, and all three 2013 strains.

Whole Genome SNP Analysis Resolves Genetic Differences among Closely Related Strains

A total of 225,284 SNPs were identified between the *Xe*, *Xg* and *Xp* genomes compared to the reference Xac306, ranging from 22,105 (*Xg164*) to 142,272 (*GEV1063*) (Supplemental Table 3). Average SNPs (\pm standard deviations) between Xac306 and *Xe*, *Xp* and *Xg* field strains are 128,376 \pm 3024, 136,673 \pm 3402, and 30,462 \pm 8015, respectively. Although the majority of *Xp* strains carry more SNPs between Xac306 than *Xe* strains, two *Xp* field strains (TB6 and TB9) show a number of SNPs within the *Xe* range. SNPs were concatenated and used to build a combined species ML tree (Figure 2). We note that differences in sequencing technology used, genome coverage and large deletions or insertions could potentially skew this analysis and therefore conclusions about branch length between the different species should be avoided. The *Xp* Group 1A clade is retained in

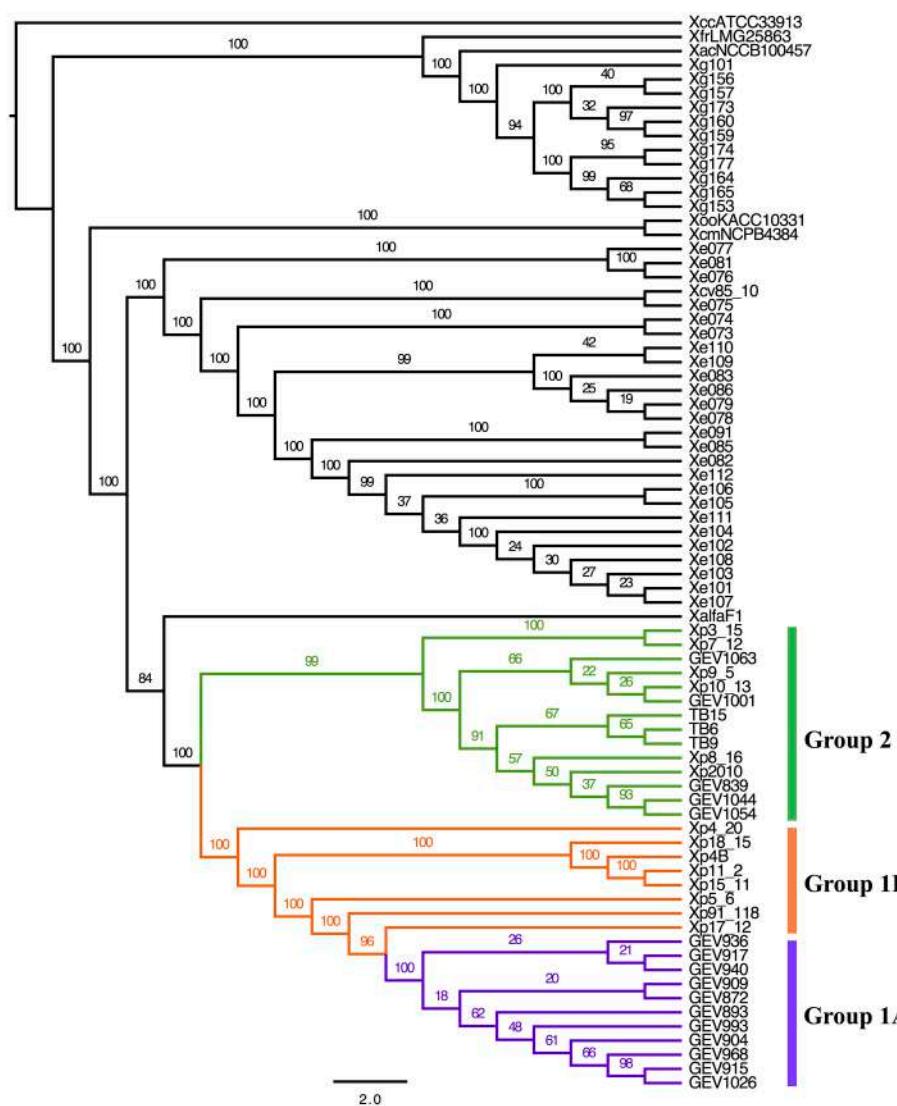


FIGURE 1 | Core genome phylogenetic analysis. Phylogenetic trees obtained by ML (IQTree) analysis, based on partitioned analysis (by codon position) of a total of 916,326 sites (1017 genes families). Branch support

values are shown for each tree, consisting of relative bootstrap proportions. Brackets to indicate *Xp* group designations are colored as follows: Group 1A, purple; Group 1B, orange; Group 2, green.

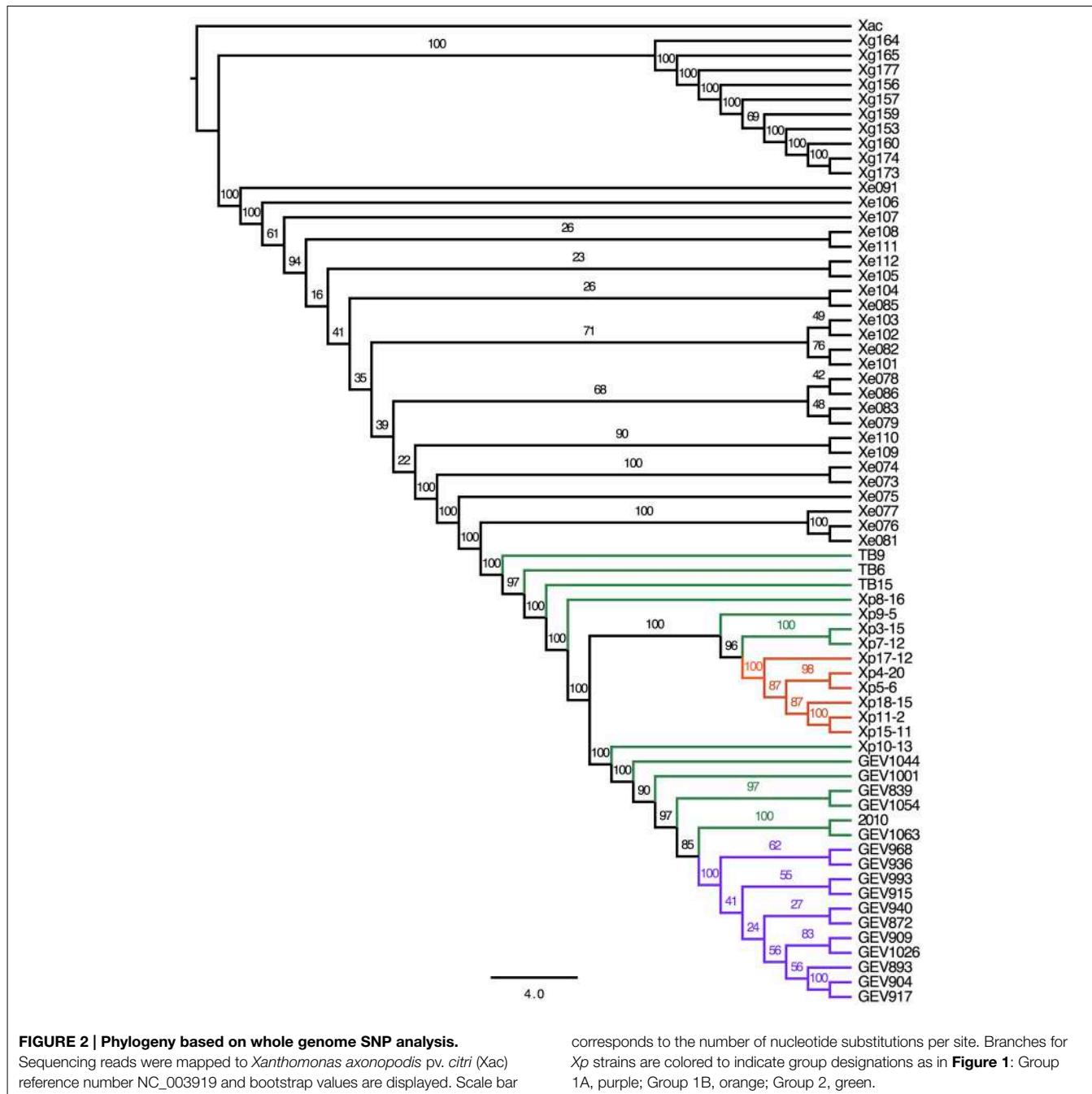
the ML SNP phylogeny (branches marked in purple). However, Group 2 (green branches) is interrupted by Group 1B strains (orange branches).

Effector Predictions For *Xanthomonas* Field Strains Identifies Differences in Effector Content Compared to Reference Genomes

Type III effector repertoires from *Xe*, *Xp*, and *Xg* field strains were compared to the appropriate reference strains *Xe85-10*, *Xp91-118*, and *Xg101* in order to determine if effector repertoires differed between strains with respect to the presence or absence of whole effectors, mutations rendering effectors inactive, or alternate alleles of effectors (Thieme et al., 2005; Potnis et al., 2011).

Xe

Several differences were found in the effector content of *Xe* field strains compared to the reference *Xe85-10* (Supplemental Table 4). Firstly, *Xe85-10* does not have the effector *XopAE*, which is a translational fusion of the *hrp* cluster members *hpaG* and *hpaF* as seen in *Xp91-118* (Potnis et al., 2011). Similar to *Xe85-10*, field strains isolated before 1997 have separate *hpaG* and *hpaF* genes, whereas *Xe* field strains isolated after 1997 possess the predicted *hpaG/hpaF* translational fusion *XopAE*. Secondly, strains collected after 1997 possess a *XopAF*-like effector. The effector has 31% amino acid identity to *XopAF* of *Xp91-118*, 80% amino acid identity to *X. fuscans* *XopAF* (WP_022560489.1) and is identical to an effector of *X. citri* pv. *citri* (WP_015472934.1) except for an in-frame internal 12 amino acid deletion. Similarly,



the *Xe* strains isolated after 1997 possess XopE3, which shares 97% amino acid identity with XopE3 from *X. arboricola* pv. *pruni* (WP_014125894.1). All field strains of *Xe* but one lack XopG, which is carried by the reference strain Xe85-10. A predicted protein-tyrosine phosphatase (abbreviated PTP) was detected in Xe075 that is not present in any other *Xe* strains. Twelve effectors present in all *Xe* strains isolated between 1985 and 2012 have no nucleotide polymorphisms (Table 2). *Xe* field strains in our collection isolated after 1997 did not contain polymorphisms in *xopAA*, *xopF1*, *xopN*, and *xopO*. Except for Xe85-10 and Xe075,

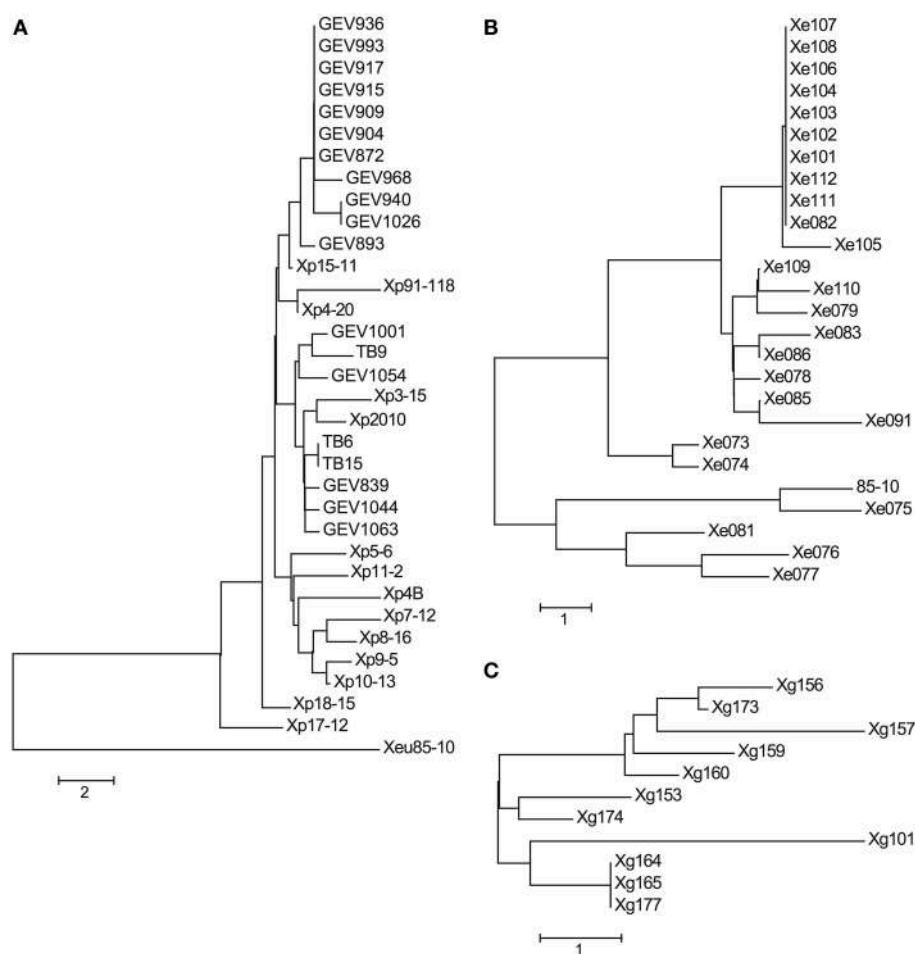
all *Xe* strains have identical sequences for effectors *xopAI*, *xopQ* and *xopV*.

The neighbor-joining tree of the effector alleles displays a grouping of the seven *Xe* strains isolated before 1997, and a clade of 11 strains with nearly identical allele profiles isolated from 2004 to 2012 (Figure 3B). Although Xe111 and Xe112 group with the clade of 11 strains and were isolated in Georgia in 2004, two other Georgia 2004 strains, Xe109 and Xe110, are separated from this clade due to differences in *avrBs2*, *xopE2*, and *xopO*. Interestingly, Xe082 was isolated in 1998 but has an effector allele

TABLE 2 | *Xanthomonas euvesicatoria* nucleotide type III effector alleles.

| Type III Effector | Year | 1985 | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 | 2003 | 2004 | 2008 | 2009 | 2010 | 2011 | 2012 |
|-------------------|------|------|------|------|----------------|----------------|------|------|------|------|------|------|------|------|------|------|
| State | | NC | NC | F | F | K | F | F | F | NC | NC | F | NC | NC | NC | NC |
| Strain | | | | | | | | | | | | | | | | |
| AvrBs2 | 4 | 7 | 7 | 4 | 7 | 3 | 1 | 2 | 6 | 5 | 5 | 1 | 1 | 2 | 2 | 2 |
| XopA | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XopAA | 1 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| XopAD | 1 | 2 | 2 | 1 | 3 | 3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XopAE | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XopAF-like | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XopAI | 1 | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| XopAK | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XopB | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XopC1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XopD | 1 | 1 | 1 | 1 | 2 | 2 | 3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XopE1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XopE2 | 1 | 4 | 3 | 1 | 1 | 3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XopE3 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XopF1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| XopF2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XopG | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| XopI | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XopJ1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XopJ3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XopK | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XopL | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XopN | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| XopO | 1 | 2 | 2 | 2 | 1 ^T | 1 ^T | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| XopP | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XopQ | 1 | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| XopR | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XopV | 1 | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| XopX | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XopZ1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 3 | 1 |
| PTP | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Each distinct nucleotide allele was assigned an arbitrary number. The number 0 indicates the effector is missing from genomic assemblies. Effectors PTP, protein tyrosine phosphatase, Superscripts are as follows: *T*, truncation; *CTG*, containing break in assembly unable to be confirmed via Sanger Sequencing.

**FIGURE 3 | Neighbor-joining trees of effector allele profiles.**

Neighbor-joining trees for *Xp* (A), *Xe* (B), and *Xg* (C) field strains were constructed using nucleotide (*Xe* and *Xg*) and amino acid (*Xp*) pairwise allele

differences between strains. Effector allele designations can be found in **Tables 2–4**. A difference at one effector between two strains equals a distance of 1.0.

profile similar to the 11-member clade made up of strains isolated between 2004 and 2012.

Xp

A shift in pathogen populations from tomato race 3 to tomato race 4 has been observed in Florida (Horvath et al., 2012). All the strains sequenced here (with isolation years spanning from 1998 to 2013 in Florida) are tomato race 4 strains and contain null mutations in the *xopAF/avrXv3* gene of the reference strain Xp91-118 (Supplemental Table 5). All strains possess XopJ4/AvrXv4 with the exception of the pepper strain Xp2010. Another effector, AvrBsT, which has been associated with hypersensitive response (HR) on pepper (Minsavage et al., 1990), has not been previously reported in *Xp*. Xp4B, which was isolated in 1998, has AvrBsT and is non-pathogenic on pepper (Supplemental Figures 2, 3). AvrBsT is also present in nine strains (out of 11) that were isolated in 2006, in all 16 strains collected in 2012, and in one of the three strains collected in 2013. Interestingly, strain Xp17-12 (isolated in 2006) contains two effectors, XopF2 and XopV, that have sequences identical to

the corresponding *Xe*85-10 effector sequence (Table 3). Effectors XopD and XopAD exhibit different alleles in the strains isolated in 2012. All strains have XopE2, which was absent in the reference strain Xp91-118. XopE2 is also present in all *Xe* and *Xg* field and reference strains. A subset of the *Xp* 2006 population have XopE4, which had been reported only in *X. fuscans* pv. *aurantifoliae* (Moreira et al., 2010). However, XopE4 is not present in any strains from 2010, 2012, or 2013. Interestingly, strains belonging to *Xp* Group 2 possess a XopQ identical to the allele from *Xe*85-10. The neighbor-joining tree based on effector alleles shows the conservation of Group 1A, but Group 1B and Group 2 strains were intertwined (Figure 3A).

Xg

The collection of *Xg* field isolates spans 3 years and covers two states (Ohio and Michigan). Effector predictions in *Xg* field strains from this period revealed the presence of four potential effectors that are not present in the reference strain Xg101, which was isolated in the southeastern Europe in 1953 (Supplemental Table 6). *Xg* field strains possess a XopJ1 that is identical to

TABLE 3 | *Xanthomonas perforans* amino acid type III effector alleles.

| Type III effector | Year | 1991 | 1998 | 2006 | | | | 2010 | | | | 2012 | | | | 2013 | | | | 1985 | | | |
|-------------------|------|------|------|------|----|----|---|------|---|----|----|------|----|----|----|------|----|----|---|------|----|----|--|
| State | F | F | G | G | F | F | F | F | F | F | F | F | F | F | F | F | F | F | F | F | F | F | |
| AvrBs2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | |
| XopA | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | |
| XopAP | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 2 | |
| XopAR | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | |
| XopC2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | |
| XopE1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | |
| XopF1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | |
| XopF2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | |
| XopI | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | |
| XopJ4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 3 | |
| XopK | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | |
| XopL | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | |
| XopN | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | |
| XopQ | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | |
| XopR | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | |
| XopV | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 4 | |
| XopX | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | |
| XopAK | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | |
| XopD | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 5 | 4 | 4 | 4 | 4 | 8 | |
| XopAD | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 3 | |
| XopE2 | 0 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | |
| XopE4 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| AvrBsT | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | |
| XopP1 | 1 | 1 | 1 | 0 | 3 | 4 | 0 | 0 | 5 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | |
| XopP2 | 2 | 5 | 17 | 2 | 4 | 8 | 6 | 9 | 0 | 3 | 1 | 18 | 15 | 14 | 1 | 1 | 1 | 1 | 1 | 11 | 12 | 19 | |
| XopP3 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| XopZ1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | |
| XopAE | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | |
| XopAF | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| Group | | 1B | 1B | 2 | 1B | 1B | 2 | 2 | 2 | 1B | 1B | 2 | 2 | 1A | 1A | 1A | 1A | 1A | 2 | 1A | 2 | 2 | |

Each distinct amino acid allele was assigned an arbitrary number. The number 0 indicates the effector is missing from genomic assemblies. AvrBsT and XopP were not used to make the Neighbor-joining effector allele trees for *Xp* in **Figure 3A**. Asterisk (*) indicates newly identified *Xp* pepper pathogens.

the allele in Xe85-10 and a type III effector protein (T3EP) that has 78% amino acid identity to a predicted *Ralstonia* peptidase effector (WP_014619440.1). A predicted effector of the Xg strains shares 65% amino acid similarity to a *X. campestris* pv. *campestris* PTP type III effector (WP_011345706.1). Two copies of XopE2 are present in 7 out of 10 Xg field strains, while the remaining three and the reference strain Xg101 have only one XopE2. Two field strains carry the effector AvrBs7 (Potnis et al., 2012). Because the repetitive nature of TAL effector genes renders them difficult to assemble from short reads, Southern blot analysis was used to identify potential family members (Supplemental Figure 3A). In addition, the ability of each strain to induce a HR on pepper cv. ECW30R, which contains the cognate R gene *Bs3* to the TAL effector AvrHah1 was tested (Supplemental Figure 3B). All field strains of Xg contained a single TAL effector, an apparent AvrHah1, on the basis of band size and activity.

Although the Xg strains were isolated within a 3-year period, only three Xg field strains (Xg164, Xg165, and Xg167) have identical effector allele profiles at the nucleotide level (Figure 3C). Three effectors are highly polymorphic: the *avrBs1*-class effector, of which three alleles were detected, and the two *xopE2* effectors, of which five and three alleles were detected (Table 4). Two alleles of *xopAD* are present at equal frequencies in the Xg field strains, with both alleles present in field strains isolated in the same year in the same state (e.g., Xg165 and Xg173, Ohio 2011) and in the same year in different states (e.g., Xg153 and Xg156, Ohio and Michigan, respectively, 2010).

Common Effectors between Species

Effector predictions of the field strains has identified two new common putative effectors to add to the previously described list of 11 effectors shared between Xe, Xp, and Xg (Potnis et al., 2011). XopE2 was identified in all Xp field strains and, while not in the reference Xp91-118, should, therefore, be considered a commonly shared effector with Xe and Xg. The identification of AvrBsT in the majority of Xp field strains and an identical copy of Xe XopJ1 in Xg field strains indicates the presence of a more broadly defined YopJ-family effector to the commonly shared effector list.

Association of AvrBsT presence or Absence in Host Range Expansion of Xp on Pepper

Xp has previously been considered restricted to tomato as a host. In 2010, we isolated a strain from a greenhouse-grown diseased pepper plant. This strain was confirmed as *X. perforans* based on 16S rRNA sequencing and multilocus sequence analysis (MLSA) (Timilsina et al., 2015), and is designated here as strain Xp2010. Xp2010 does not induce a hypersensitive response (HR) on pepper cv. Early CalWonder (ECW) and is able to create foliar disease lesions (Supplemental Figure 1). Effector predictions for Xp2010 indicated that the absence of AvrBsT, which induces HR on pepper (Kim et al., 2010), may be responsible for its pepper host expansion. We were curious to see if other Xp strains in our collection displayed host expansions to pepper similar to Xp2010 and if this could be explained solely by the absence of AvrBsT. We used PCR to confirm the effector prediction results

for the presence or absence of AvrBsT in the Xp field strains and inoculated pepper cv. ECW with a high inoculum (10^8 CFU/ml) to determine which strains induce HR (Supplemental Table 7). We confirmed that four additional field strains, Xp5-6, Xp17-12, TB9, and TB15 do not possess AvrBsT and also fail to induce HR. Xp17-12, TB9, and TB15 but not Xp5-6 are able to cause disease lesions on pepper cv. ECW when infiltrated at a low inoculum (10^4 CFU/ml) (Supplemental Figure 7), indicating that additional factors restrict the host range of Xp5-6 on pepper.

Three of the newly identified pepper pathogens (Xp2010, TB9, and TB15) belong to Group 2. We observed no HR but differences in pathogenicity and lesion development for the two Group 1B strains that lack AvrBsT (Xp5-6 and Xp17-12). Strain Xp5-6 showed a phenotype similar to Xp91-118ΔavrXv3, which is unable to cause lesions on pepper (Supplemental Figure 1). We hypothesized that Group 2 strains carrying mutations in AvrBsT would exhibit *in planta* growth and virulence similar to that of virulent strains from pepper in our collection. At the same time, strains belonging to Group 1 and carrying mutations in AvrBsT would be non-pathogenic on pepper, similar to Xp91-118ΔavrXv3 (Astua-Monge et al., 2000). To test this hypothesis, AvrBsT insertion mutants were introduced into two Group 2 strains, GEV839 and GEV1001, and two Group 1A strains, GEV872 and GEV909. Indeed, XpGEV839ΩavrBsT and XpGEV1001ΩavrBsT from Group 2 lose ability to elicit HR in pepper and are virulent similar to TB15 (Figure 4). *In planta* population levels for these two mutants were not significantly different from TB15 at Days 4 and 8 post-infiltration, indicating that AvrBsT is the lone factor restricting these two strains on pepper. Also as predicted, insertion mutants of *avrBsT* in Group 1A strains GEV872 and GEV909 lose the ability to induce HR on pepper but do not grow to the same extent as TB15. *In planta* populations of XpGEV872ΩavrBsT and XpGEV909ΩavrBsT were 100-fold higher compared to 91-118ΔavrXv3 but 20–50-fold lower compared to pepper pathogens XpGEV839ΩavrBsT, XpGEV1001ΩavrBsT and TB15, indicating the existence of additional factors restricting the virulence of Group 1A strains on pepper.

Loss of XopQ and AvrBsT Expands the Host Range of Xp to *Nicotiana benthamiana*

Members of both the XopQ and AvrBsT effector families are known to induce a HR in *N. benthamiana* (Wei et al., 2007; Kim et al., 2010). Family members of XopQ occur in Xe, Xp and Xg. It has previously been shown that a *Pseudomonas syringae* pv. *tomato* DC3000 mutant deficient for the XopQ homolog HopQ1-1 causes disease in *N. benthamiana* (Wei et al., 2007). Xe85-10 is not pathogenic on *N. benthamiana*, causing a weak HR (Figure 5, Xe85-10ΔXopQ). A deletion of XopQ in strain Xe85-10 results in a strain that causes water soaking, disease lesions, and grows to a high titer after 6 days on *N. benthamiana* (Figure 5, Xe85-10ΔXopQ). Complementation of the deletion with plasmid pVSP61 carrying the Xe85-10 allele of XopQ restored the original phenotype of low virulence and enhanced the HR phenotype of the complemented strain (Figure 5, Xe85-10ΔXopQ cXopQ).

TABLE 4 | *Xanthomonas gardneri* nucleotide type III effector alleles.

| Type III effector | Year | 1953 | | 2010 | | | | | 2011 | | | 2012 | |
|-------------------|------|-------|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | | State | O | M | M | M | M | O | O | O | O | O | O |
| | | | Strain | Xg101 | Xg153 | Xg156 | Xg157 | Xg159 | Xg160 | Xg164 | Xg165 | Xg173 | Xg174 |
| AvrBs1 class | | | 1 | 2 | 3 | 3 | 3 | 3 | 2 | 2 | 3 | 2 | 2 |
| AvrBs2 | | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| AvrBs7 | | | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| AvrHah1 | | na | na | na | na | na | na | na | na | na | na | na | na |
| AvrXccA1 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XopAD | | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 1 | 1 |
| XopAM | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XopAO | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XopAQ | | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XopAS | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XopB | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XopD | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XopE2_0 | | 1 | 1 | 4 | 5 | 3 | 1 | 5 | 5 | 5 | 4 | 2 | 5 |
| XopE2_1 | | 0 | 1 | 1 | 3 | 2 | 1 | 0 | 0 | 1 | 1 | 0 | 0 |
| XopF1 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XopG | | 1 | FS | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XopJ1 | | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XopK | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XopL | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XopN | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XopQ | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XopR | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XopX | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| ZopZ2 | | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | CTG | 1 | 1 |
| PTP | | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| T3EP | | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

Each distinct nucleotide allele was assigned an arbitrary number. The number 0 indicates the effector is missing from genomic assemblies. PTP, protein tyrosine phosphatase; T3EP, type III effector protein. The TAL effector AvrHah1 could not be assembled (na, not assembled). Superscripts are as follows: CTG, contig break in assembly unable to be confirmed via Sanger Sequencing; FS, a frame shift mutation.

All *Xp* field strains contain XopQ and the majority of *Xp* strains contain AvrBsT. Therefore, *Xp* derivative strains in Xp4B were constructed with single gene deletions of XopQ and AvrBsT, and deletions of both XopQ and AvrBsT. Single knockouts for XopQ and AvrBsT in Xp4B remained incompatible on *N. benthamiana* (HR, low growth, no lesions, **Figure 5**), although Xp4BΔAvrBsT experienced reduced growth compared to Xp4B and Xp4BΔXopQ that was complemented back by the addition of AvrBsT. The double effector deletion mutant Xp4BΔXopQΔAvrBsT gave disease lesions at a low inoculum, showed water soaking at a high inoculum, and grew to levels comparable with Xe85-10ΔXopQ and Xg153 after 6 days on *N. benthamiana* (**Figure 5**). Consistent with the low virulence gain on pepper by Group 1A AvrBsT mutants, Xp4BΔAvrBsT was able to induce weak lesions on pepper cv. ECW (Supplemental Figure 1), but did not grow to comparable population levels of pepper pathogens Xe85-10 or Xg153 (Supplemental Figure 2).

Discussion

The population dynamics of *Xanthomonas*-infecting pepper and tomato has shifted in the United States over the past 25 years. Prior to 1991, *Xe* was the prevalent species and the only species in tomato fields in Florida. *Xp* tomato race 3 was identified first in 1991 and eventually replaced the *Xe* population in tomato fields, a process attributed to the ability of *Xp* race 3 to produce bacteriocins against *Xe* strains (Tudor-Nelson et al., 2003; Hert et al., 2005). Xp4B, a tomato race 4 strain identified in 1998, carries a mutation in the *avrXv3* gene. Field surveys thereafter in 2006 and 2012 recovered a majority of race 4 strains carrying either frameshift mutations or transposon insertions in *avrXv3* (Horvath et al., 2012). The first reports of *Xg* in the United States occurred in Ohio and Michigan during a bacterial spot outbreak on tomato in 2009 (Ma et al., 2011).

Here, we sequenced *Xe*, *Xp*, and *Xg* strains isolated in different years, from different fields/transplant houses

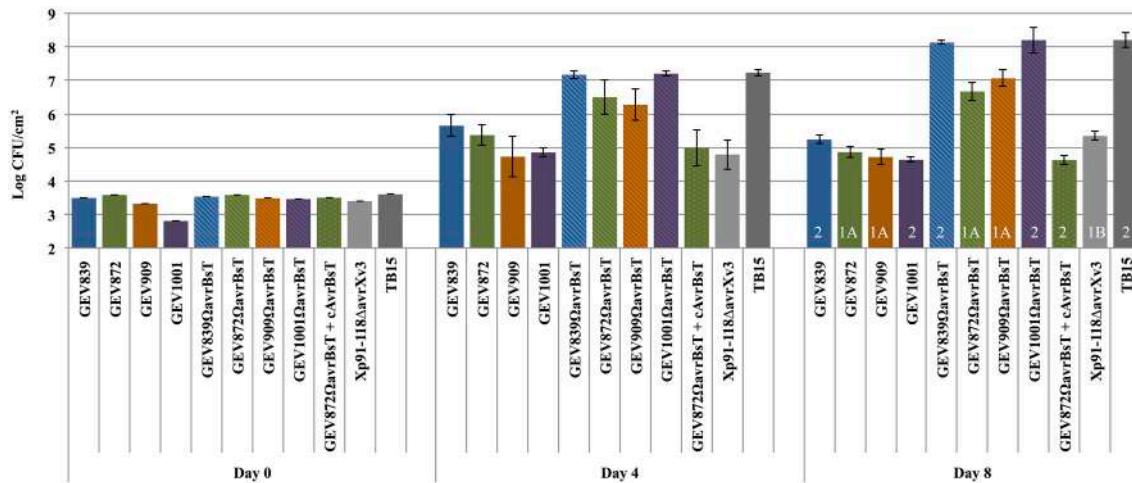


FIGURE 4 | Role of *avrBsT* as host range determinant on pepper cv. Early CalWonder. *In planta* growth of *X. perforans* strains and *avrBsT* insertion mutants was measured at different time points (days 0, 4, and 8)

after infiltration of leaves of pepper cv. Early CalWonder (ECW) using an inoculum concentration of 10^5 CFU/ml. Group designations are marked in white over Day 8 growth.

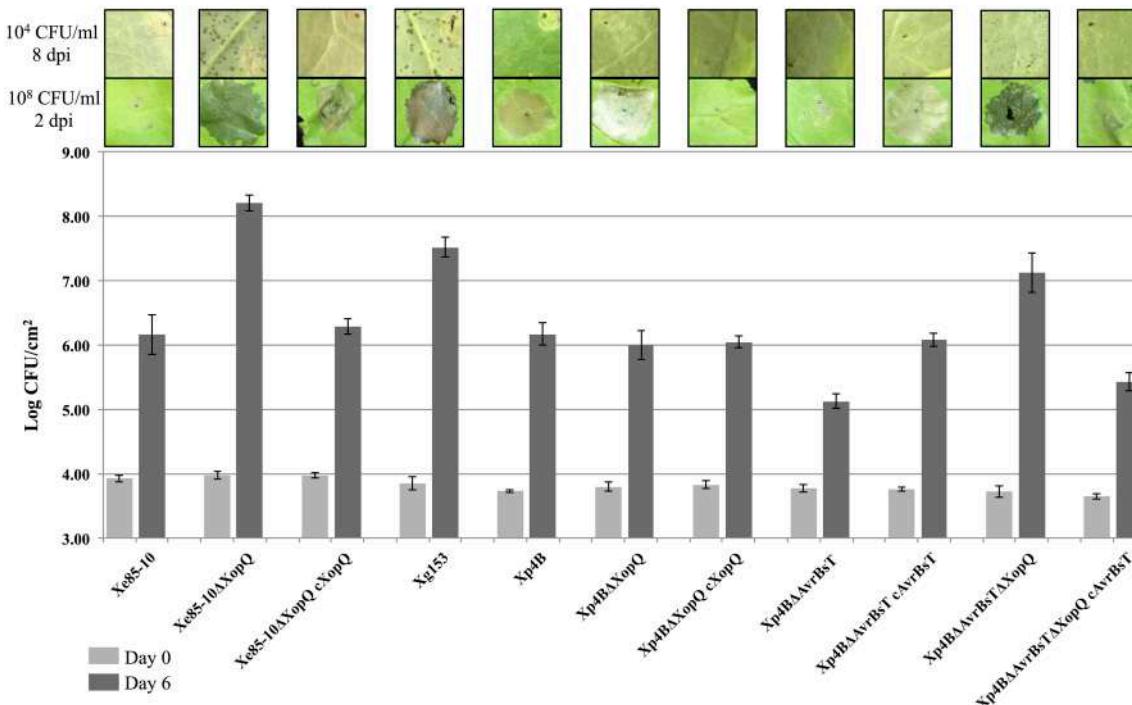


FIGURE 5 | Host expansion of *Xanthomonas* spp. on *Nicotiana benthamiana*. *In planta* growth of *X. perforans* and *AvrBsT* and *KoxQ* deletion mutants measured at Days 0 and 6 with a starting inoculum of 10^5 CFU/mL after infiltration of leaves of *Nicotiana benthamiana*. Infiltrations on *N. benthamiana* were

performed at 10^4 CFU/ml to display lesions and photographed 8 days post-infiltration (8 dpi). High inoculum infiltrated spots were performed at 10^8 CFU/ml to show HR or water soaking and photographed 2 dpi. This experiment was repeated three times with similar results.

throughout southeastern and midwestern United States. We have also sequenced strains collected during the same season from the same field. Following typical population genomic studies, we have taken three components into consideration; location,

time and niche (Monteil et al., 2013). Combining genomic data with metadata such as plant host source, year and location of isolation provides inference of population structure and clues to host adaptation. We have computationally predicted the type III

effector repertoires for each strain, and have used two different methods in order to infer evolutionary relationships of strains based on whole genome data. Phylogeny based on the core genome considers orthologous genes among the set of genomes considered. Phylogeny based on whole genome SNPs included core as well as variable regions of the genome, and thus provides an additional method to describe the genetic diversity within field strains. Phenotypic data, in particular, host range, was then correlated with the whole genome phylogenies.

MLSA studies showed the presence of two distinct groups of *Xp* populations that appeared to be clonal within the lineage (Timilsina et al., 2015). However, these studies were based on 6 genes out of 5000 genes. In our study, core ortholog gene phylogeny also revealed two distinct groups among *Xp* populations (Groups 1 and 2), although we were able to further separate Group 1 into Group 1A and 1B. Group 1A contains 11 strains isolated in 2012, whereas Group 1B contains six strains isolated in 2006 in addition to Xp4B and the reference strain Xp91-118, isolated in 1998 and 1991, respectively. Group 2 comprises five strains from 2006, the single strain from 2010, five strains from 2012, and all three 2013 strains. Additionally, we detected genetic diversity among strains that appeared to be clonal from MLSA in previous work (Timilsina et al., 2015), particularly evident in *Xp* Group 1A. In our study, the *Xp* 2006 population was more diverse than the 2012 population, possibly due to the fact that sampling in 2006 was carried out in a broader geographic range in Florida and Georgia. The diversity within the 2006 population is evident from the core genome and SNP phylogenies.

This study re-emphasizes the role of population genomics for identification of elements involved in host-pathogen arms race. The data revealed the emergence of tomato race 4 strains of *Xp* carrying mutations (either frameshift/transposon insertion) in *avrXv3*. Strain Xp91-118 isolated in 1991 was non-pathogenic on pepper even when mutated in *avrXv3* (Astua-Monge et al., 2000), indicating the existence of other factors that restrict its host range on pepper. The majority of *Xp* strains in our collection, isolated after 1998, have acquired AvrBsT, an avirulence protein responsible for restricting host range on pepper. AvrBsT has been shown to be a virulence factor by suppressing defense responses in tomato (Kim et al., 2010), possibly conferring a competitive advantage to pathogens in tomato fields. Four of the five *Xp* strains isolated after 1998 that do not possess AvrBsT are pathogenic on pepper. Interestingly, mutation in *avrBsT* results in differences in the *in planta* populations in pepper when compared between Group 1A and Group 2. *avrBsT* mutants in Group 2 experience a full virulence gain on pepper, whereas *avrBsT* mutants in Group 1A acquire only a partial growth benefit, indicating that additional factors restrict the host expansion of Group 1A strains onto pepper. Phenotypic characterization, including pepper pathogenicity tests of *avrBsT* mutants, will need to be conducted on other strains in Groups 1 and 2 to support more definitive conclusions.

At the whole genome level, horizontal gene transfer (HGT) of genes that determine phenotypic differences might have occurred frequently enough during evolution to explain the differing degree of pepper pathogenicity between strains belonging to

Group 1B. Two *Xp* strains in Group 1B, Xp17-12, and Xp5-6, do not have AvrBsT and do not induce HR on pepper cv. ECW. However, Xp17-12 is able to induce water-soaked disease lesions on pepper when infiltrated into pepper leaves at a low inoculum (10^4 CFU/mL) whereas Xp5-6 induces only weak lesions. Similar to Xp5-6, an *avrBsT* deletion in the Group 1B strain Xp4B (Xp4B Δ AvrBsT) induces weak disease lesions on pepper and acquires only a partial *in planta* growth increase. Incongruence in degree of pathogenicity and clade could partly be due to the loss or gain of effectors through HGT. Xp17-12 contains effector alleles for XopF2 and XopAD that match those found in Xe85-10 but not those of any other *Xp* strain analyzed here, suggesting the occurrence of HGT events that may have contributed to its ability to infect pepper. Xp5-6 does not share any common effector alleles with Xe85-10. Interestingly, all Group 2 strains contain a XopQ allele identical to XopQ in Xe85-10, while *Xp* strains in Group 1 have a different allele. Previous MLSA analysis also showed evidence for recombination events resulting in haplotypes for two housekeeping genes (*gapA* and *gyrB*) in *Xp* Group 2 strains identical to that found in Xe85-10 (Timilsina et al., 2015). Because mutation in *avrBsT* in the tested Group 2 strains results in complete virulence on pepper, Group 2 strains may have emerged from populations that underwent recombination with an Xe85-10-related strain, acquiring new virulence genes for pepper pathogenicity. Homologous recombination between chromosomal DNA of different *Xanthomonas* species by conjugation *in planta* has been previously observed (Basim et al., 1999), while HGT of virulence-associated genes between different lineages within *X. axonopodis* strains has contributed to host range (Mhedbi-Hajri et al., 2013).

Field strain genomic analysis presents an efficient method for deriving the diversity of type III effector repertoires. Knowledge of the effector load in the population will inform strategies for achieving broad durable resistance strategies based on R gene deployment. Within each species, we identified several differences in the effector repertoires of *Xe*, *Xp*, and *Xg* field strains, including the gain or loss of effector genes, null mutations, and the presence of alternate alleles. We predicted three effector additions to the overall *Xe* field strain repertoire (XopE3, XopAF-like, and XopAE) and one removal (XopG) in comparison to the reference strain Xe85-10. The most polymorphic effector in *Xe* is *avrBs2*, a phenomenon perhaps explained by the selective pressures of the pepper *Bs2* resistance gene deployed in the early 1990's. Several of the previously reported mutations in *avrBs2* are represented here, with no novel polymorphisms detected (Swords et al., 1996; Wichmann et al., 2005). Generally, the effector predictions for *Xe* field strains isolated between 1994 and 2004 show increased effector polymorphisms compared to strains isolated between 2004 and 2012, indicating that the effector repertoires have stabilized over time in our sampling population. *Xp* field strains have evolved their repertoires by losing/gaining effectors (XopE2, XopE4, AvrBsT), through allelic exchange (as seen with XopQ in Group 2 strains) and by frameshift mutations/transposon insertions (in *avrXv3*). Diversity in effector repertoires is seen even in strains collected from the same field during a single growing season.

Strains TB6 and TB15 possess identical type III effector profiles and appear clonal based on core genome phylogeny except for the absence of AvrBsT in TB15. However, this difference has expanded the host range of TB15 to include pepper while TB6 is restricted to tomato. Similar to TB15, TB9 does not possess AvrBsT but has different alleles of XopD and XopE1 compared to TB6 and TB15. We predicted four additions to the *Xg* field strain effector repertoire including a second copy of XopE2 and a XopJ1 identical to *Xe* strains. We also detected allele differences in an AvrBs1-like effector, XopAD, and XopE2. Through this analysis two additional effectors can be added to the previous list of 11 commonly shared effectors between *Xe*, *Xp*, and *Xg* (Potnis et al., 2011): XopE2 and a YopJ-family member (AvrBsT in *Xp*, XopJ1 in *Xg* and *Xe*).

In addition to strain-level variation, allelic diversification in type III effectors was observed at the species level across *Xe*, *Xp*, and *Xg*. Because type III effector repertoires are proposed to be a major factor determining host range (Hajri et al., 2009), it is important to understand the diversity of effectors present in different species that infect common hosts. Although *Xe*, *Xp*, and *Xg* share thirteen core effectors, effector alleles between these three species may be considerably different. For example, the effector AvrBs2 protein sequence shares 99% identity between reference strains Xp91-118 and Xe85-10, but 77% identity to the AvrBs2 in *Xg*101. Similarly, the XopQ alleles of Xp91-118 and Xe85-10 share 99% identity at the amino acid level, but 58% identity to XopQ from *Xg*101. Sampling of a genetically diverse population can be informative to reveal the dominant effector alleles in a specific geographical region, which would be the most appropriate alleles to screen for R protein resistance strategies.

Curiously, we discovered a spectrum of host expansion for *Nicotiana benthamiana* involving the effectors XopQ and AvrBsT. While wild type *Xg* is virulent on *N. benthamiana*, a XopQ deletion in Xe85-10 (*Xe*85-10ΔXopQ) and a double deletion of XopQ and AvrBsT in Xp4B (Xp4BΔXopQΔAvrBsT) results in a *N. benthamiana* host gain. Reducted *in planta* growth of Xp4BΔAvrBsT compared to Xp4B and Xp4BΔXopQ indicates that AvrBsT may play an important virulence role in *N. benthamiana*. Because the XopQ alleles in *Xe* and *Xp* are relatively similar and stable over time in field strains, the potential R protein “R-XopQ” in *N. benthamiana* would be a promising candidate as a resistance tool against *Xe* and *Xp* in pepper and tomato.

The increased speed and dropping cost of DNA sequencing technology combined with the use of genome editing techniques are providing new opportunities for designing resistance strategies against specific pathogens in various crop species. The spread of agricultural pathogens into new niches, either by increasing global movement of food or the emergence of new niches from climate change, makes the continued genomic surveillance of agricultural pathogens a top priority

for food security and resistance strategies. Of particular importance are tracking shifts in dominant species and changes in effector repertoires and alleles. Effector maintenance and stability is a key consideration for the future design of durable resistance strategies using R-gene employment into crops.

Author Contributions

JJ, GV, and BS conceived the project. JJ, FW, BS oversaw genomic sequencing. JJ and GV provided *Xp* strains, DR provided *Xe* strains, SM provided *Xg* strains. GM and AS prepared genomic DNA for sequencing. AS, NP, AA, FW, RB, and JB oversaw genome assemblies. ST helped with *Xp* genome assembly and initial phenotypic characterization of *Xp* strains. NP and AS constructed mutants and tested them phenotypically and for *in planta* growth. GM and DD helped with cloning and constructing mutants. NP, AS, FW, JB, RB, JJ, and BS performed data analyses and interpreted them. NP and AS did effector analyses and EG performed phylogenetic analysis based on effector profiles. Core genome phylogenies were constructed and interpreted by JP, JM, NA, and JS. MW and RB created the whole genome SNP dataset and constructed phylogenies. JM and JP submitted genome sequences to NCBI GenBank. AS and NP wrote final manuscript. All authors approved the final manuscript.

Acknowledgments

AS is funded by the NSF Graduate Research Fellowship Program. NP was supported by grant 2011-670137-30166 from the USDA National Institute of Food and Agriculture (NIFA). The Staskawicz lab is supported by Two Blades Foundation. This research was supported in part with funds from a Specialty Crop Block Grant, award 18015, to G. E. Vallad and J. B. Jones from the Florida Department of Agriculture and Consumer Services and administered by the Florida Specialty Crop Foundation. This work used the KSU Integrated Genomics Facility and the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 Instrumentation Grants S10RR029668 and S10RR027303. FFW was supported by grant IOS-123819 from the Plant Genome Research Program of the National Science Foundation (NSF) and National Research Initiative Competitive Grants Program Grant 2012-67013-19383 from the USDA-NIFA. JS is funded by CAPES grant 3385/2013 and CNPq; JP and JM have CAPES postdoctoral fellowships.

Supplementary Material

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2015.00535/abstract>

References

- Almeida, N. F., Yan, S., Cai, R., Clarke, C. R., Morris, C. E., Schaad, N. W., et al. (2010). PAMDB, a multilocus sequence typing and analysis database and website for plant-associated microbes. *Phytopathology* 100, 208–215. doi: 10.1094/PHYTO-100-3-0208
- Altshul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389
- Astua-Monge, G., Minsavage, G. V., Stall, R. E., Davis, M. J., Bonas, U., and Jones, J. B. (2000). Resistance of tomato and pepper to T3 strains of *Xanthomonas campestris* pv. *vesicatoria* is specified by a plant-inducible avirulence gene. *Mol. Plant Microbe Interact.* 13, 911–921. doi: 10.1094/MPMI.2000.13.9.911
- Bart, R., Cohn, M., Kassen, A., McCallum, E. J., Shybut, M., Petriello, A., et al. (2012). High-throughput genomic sequencing of cassava bacterial blight strains identifies conserved effectors to target for durable resistance. *Proc. Natl. Acad. Sci. U.S.A.* 109, E1972–E1979. doi: 10.1073/pnas.1208003109
- Basim, H., Stall, R. E., Minsavage, G. V., and Jones, J. B. (1999). Chromosomal gene transfer by conjugation in the plant pathogen *Xanthomonas axonopodis* pv. *vesicatoria*. *Phytopathology* 89, 1044–1049. doi: 10.1094/PHYTO.1999.89.11.1044
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Boyd, L. A., Ridout, C., O'Sullivan, D. M., Leach, J. E., and Leung, H. (2013). Plant-pathogen interactions: disease resistance in modern agriculture. *Trends Genet.* 29, 233–240. doi: 10.1016/j.tig.2012.10.011
- Contreras-Moreira, B., and Vinuesa, P. (2013). GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl. Environ. Microbiol.* 79, 7696–7701. doi: 10.1128/AEM.02411-13
- da Silva, A. C. R., Ferro, J. A., Reinach, F. C., Farah, C. S., Furlan, L. R., Quaggio, R. B., et al. (2002). Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature* 417, 459–463. doi: 10.1038/417459a
- Grant, S. R., Fisher, E. J., Chang, J. H., Mole, B. M., and Dangl, J. L. (2006). Suberfuge and manipulation: type III effector proteins of phytopathogenic bacteria. *Annu. Rev. Microbiol.* 60, 425–449. doi: 10.1146/annurev.micro.60.080805.142251
- Hajri, A., Brin, C., Hunault, G., Lardeux, F., Lemaire, C., Manceau, C., et al. (2009). A “repertoire for repertoire” hypothesis: repertoires of type three effectors are candidate determinants of host specificity in *Xanthomonas*. *PLoS ONE* 4:e6632. doi: 10.1371/journal.pone.0006632
- Hamza, A. A., Robène-Soustrade, I., Jouen, E., Gagnevin, L., and Lefevre, P. (2010). Genetic and pathological diversity among *Xanthomonas* strains responsible for bacterial spot on tomato and pepper in the southwest Indian Ocean region. *Plant Dis.* 94, 993–999. doi: 10.1094/PDIS-94-8-0993
- Hert, A. P., Roberts, P. D., Momol, M. T., Minsavage, G. V., Tudor-Nelson, S. M., and Jones, J. B. (2005). Relative importance of bacteriocin-like genes in antagonism of *Xanthomonas perforans* tomato race 3 to *Xanthomonas euvesicatoria* tomato race 1 strains. *Appl. Environ. Microbiol.* 71, 3581–3588. doi: 10.1128/AEM.71.7.3581-3588.2005
- Horvath, D. M., Stall, R. E., Jones, J. B., Pauly, M. H., Vallad, G. E., Dahlbeck, D., et al. (2012). Transgenic resistance confers effective field level control of bacterial spot disease in tomato. *PLoS ONE* 7:e42036. doi: 10.1371/journal.pone.0042036
- Ibarra Caballero, J., Zerillo, M. M., Snelling, J., Boucher, C., and Tisserat, N. (2013). Genome sequence of *Xanthomonas arboricola* pv. *corylina*, isolated from Turkish filbert in Colorado. *Genome Announc.* 1, e00246–e00213. doi: 10.1128/genomeA.00246-13
- Jalan, N., Aritua, V., Kumar, D., Yu, F., Jones, J. B., Graham, J. H., et al. (2011). Comparative genomic analysis of *Xanthomonas axonopodis* pv. *citrumelo* F1, which causes citrus bacterial spot disease, and related strains provides insights into virulence and host specificity. *J. Bacteriol.* 189, 6342–6357. doi: 10.1128/JB.05777-11
- Jones, J. B., Bouzar, H., Somodi, G. C., Stall, R. E., Pernezny, K., El-Morsy, G., et al. (1998b). Evidence for the preemptive nature of tomato race 3 of *Xanthomonas campestris* pv. *vesicatoria* in Florida. *Phytopathology* 88, 33–38. doi: 10.1094/PHYTO.1998.88.1.33
- Jones, J. B., Lacy, G. H., Bouzar, H., Stall, R. E., and Schaad, N. W. (2004). Reclassification of the xanthomonads associated with bacterial spot disease of tomato and pepper. *Syst. Appl. Microbiol.* 27, 755–762. doi: 10.1078/0723202042369884
- Jones, J. B., and Stall, R. E. (2000). Systematic analysis of xanthomonads (*Xanthomonas* spp.) associated with pepper and tomato lesions. *Int. J. Syst. Evol. Microbiol.* 50, 1211–1219. doi: 10.1099/00207713-50-3-1211
- Jones, J. B., Stall, R. E., and Bouzar, H. (1998a). Diversity among xanthomonads pathogenic on pepper and tomato. *Annu. Rev. Phytopathol.* 36, 41–58. doi: 10.1146/annurev.phyto.36.1.41
- Jones, J. D. G., and Dangl, J. L. (2006). The plant immune system. *Nature* 444, 323–329. doi: 10.1038/nature05286
- Katoh, K., Misawa, K., Kuma, K.-I., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436
- Kay, S., and Bonas, U. (2009). How *Xanthomonas* type III effectors manipulate the host plant. *Curr. Opin. Microbiol.* 12, 37–43. doi: 10.1016/j.mib.2008.12.006
- Kim, N. H., Choi, H. W., and Hwang, B. K. (2010). *Xanthomonas campestris* pv. *vesicatoria* effector AvrBsT induces cell death in pepper, but suppresses defense responses in tomato. *Mol. Plant Microbe Interact.* 23, 1069–1082. doi: 10.1094/MPMI-23-8-1069
- Kück, P., and Meusemann, K. (2010). FASconCAT: convenient handling of data matrices. *Mol. Phylogenet. Evol.* 56, 1115–1118. doi: 10.1016/j.ympev.2010.04.024
- Lanfear, R., Calcott, B., Ho, S. Y. W., 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
- Lanfear, R., Calcott, B., Kainer, D., Mayer, C., and Stamatakis, A. (2014). Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Evol. Biol.* 14:82. doi: 10.1186/1471-2148-14-82
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Lee, B. M., Park, Y., Park, D., Kang, H., Kim, J., Song, E., et al. (2005). The genome sequence of *Xanthomonas oryzae* pv. *oryzae* KACC10331, the bacterial blight pathogen of rice. *Nucleic Acids Res.* 33, 577–586. doi: 10.1093/nar/gki206
- Li, L., Stoeckert, C. J. Jr., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189. doi: 10.1101/gr.1224503
- Lindgren, P. B., Peet, R. C., and Panopoulos, N. J. (1986). Gene-cluster of *Pseudomonas syringae* pv. *phaseolicola* controls pathogenicity of bean plants and hypersensitivity on nonhost plants. *J. Bacteriol.* 168, 512–522.
- Ma, X., Lewis Ivey, M. L., and Miller, S. A. (2011). First report of *Xanthomonas gardneri* causing bacterial spot of tomato in Ohio and Michigan. *Plant Dis.* 95, 1584–1584. doi: 10.1094/PDIS-05-11-0448
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110
- Mhedbi-Hajri, N., Hajri, A., Boureau, T., Darrasse, A., Durand, K., Brin, C., et al. (2013). Evolutionary history of the plant pathogenic bacterium *Xanthomonas axonopodis*. *PLoS ONE* 8:e58474. doi: 10.1371/journal.pone.0058474
- Midha, S., and Patil, P. B. (2014). Genomic insights into the evolutionary origin of *Xanthomonas axonopodis* pv. *citri* and its ecological relatives. *Appl. Environ. Microbiol.* 80, 6266–6279. doi: 10.1128/AEM.01654-14
- Minsavage, G. V., Dahlbeck, D., Whalen, M. C., Kearney, B., Bonas, U., Staskawicz, B. J., et al. (1990). Gene-for-gene relationships specifying disease resistance in *Xanthomonas campestris* pv. *vesicatoria*-pepper interactions. *Mol. Plant Microbe Interact.* 3, 41–47. doi: 10.1094/MPMI-3-041
- Monteil, C. L., Cai, R., Liu, H., Llontop, M. E. M., Leman, S., Studholme, D. J., et al. (2013). Nonagricultural reservoirs contribute to emergence and evolution of *Pseudomonas syringae* crop pathogens. *New Phytol.* 199, 800–811. doi: 10.1111/nph.12316
- Moreira, L. M., Almeida, N. F., Potnis, N., Digampietri, L. A., Adi, S. S., Bortolossi, J. C., et al. (2010). Novel insights into the genomic basis of citrus canker based on the genome sequences of two strains of *Xanthomonas fuscans* subsp. *aurantifoliae*. *BMC Genomics* 11:238. doi: 10.1186/1471-2164-11-238

- Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300
- Obradovic, A., Jones, J. B., Balogh, B., and Momol, M. T. (2008). “Integrated management of tomato bacterial spot,” in *Integrated Management of Disease Caused by Fungi, Phytoplasma, and Bacteria, Integrated Management of Plant Pests and Diseases*, eds A. Ciancio and K. G. Mukerji (Dordrecht: Springer Science+Business Media), 211–223.
- Parkinson, N. M., Cowie, C., Heeney, J., and Stead, D. E. (2009). Phylogenetic structure of *Xanthomonas* determined by comparison of *gyrB* sequences. *Int. J. Syst. Evol. Microbiol.* 59, 264–274. doi: 10.1099/ijss.0.65825-0
- Penn, O., Privman, E., Ashkenazy, H., Landan, G., Graur, D., and Pupko, T. (2010). GUIDANCE: a web server for assessing alignment confidence scores. *Nucleic Acids Res.* 38, 23–28. doi: 10.1093/nar/gkq443
- Potnis, N., Krasileva, K., Chow, V., Almeida, N. F., Patil, P. B., Ryan, R. P., et al. (2011). Comparative genomics reveals diversity among xanthomonads infecting tomato and pepper. *BMC Genomics* 12:146. doi: 10.1186/1471-2164-12-146
- Potnis, N., Minsavage, G., Smith, J. K., Hurlbert, J. C., Norman, D., Rodrigues, R., et al. (2012). Avirulence proteins AvrBs7 from *Xanthomonas gardneri* and AvrBs1.1 from *Xanthomonas euvesicatoria* contribute to a novel gene-for-gene interaction in pepper. *Mol. Plant Microbe Interact.* 25, 307–320. doi: 10.1094/MPMI-08-11-0205
- Schornack, S., Minsavage, G. V., Stall, R. E., Jones, J. B., and Lahaye, T. (2008). Characterization of AvrHahl1, a novel AvrBs3-like effector from *Xanthomonas gardneri* with virulence and avirulence activity. *New Phytol.* 179, 546–556. doi: 10.1111/j.1469-8137.2008.02487.x
- Stall, R. E., Beaulieu, C., Egel, D., Hodge, N. C., Leite, R. P., Minsavage, G. V., et al. (1994). Two genetically diverse groups of strains are included in *Xanthomonas campestris* pv. *vesicatoria*. *Int. J. Syst. Bacteriol.* 44, 47–53. doi: 10.1099/00207713-44-1-47
- Stall, R. E., Jones, J. B., and Minsavage, G. V. (2009). Durability of resistance in tomato and pepper to xanthomonads causing bacterial spot. *Annu. Rev. Phytopathol.* 47, 265–284. doi: 10.1146/annurev-phyto-080508-081752
- Stamatakis, A. (2014). RAxML Version 8: a tool for phylogenetic analysis and post-analysis of large phylogenomes. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Sugio, A., Yang, B., and White, F. F. (2005). Characterization of the *hrpF* pathogenicity peninsula of *Xanthomonas oryzae* pv. *oryzae*. *Mol. Plant Microbe Interact.* 18, 546–554. doi: 10.1094/MPMI-18-0546
- Swords, K., Dahlbeck, D., and Staskawicz, B. J. (1996). Spontaneous and induced mutations in a single open reading frame alter both virulence and avirulence in *Xanthomonas campestris* pv. *vesicatoria* *avrBs2*. *J. Bacteriol.* 178, 4661–4669.
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- Thieme, F., Koebnik, R., Bekel, T., Berger, C., Boch, J., Büttner, D., et al. (2005). Insights into genome plasticity and pathogenicity of the plant pathogenic bacterium *Xanthomonas campestris* pv. *vesicatoria* revealed by the complete genome sequence. *J. Bacteriol.* 187, 7254–7266. doi: 10.1128/JB.187.21.7254-7266.2005
- Timilsina, S., Jibrin, M. O., Potnis, N., Minsavage, G. V., Kebede, M., Schwartz, A., et al. (2015). Multilocus sequence analysis of xanthomonads causing bacterial spot of tomato and pepper plants reveals strains generated by recombination among species and recent global spread of *Xanthomonas gardneri*. *Appl. Environ. Microbiol.* 81, 1520–1529. doi: 10.1128/AEM.03000-14
- Tudor-Nelson, S. M., Minsavage, G. V., Stall, R. E., and Jones, J. B. (2003). Bacteriocin-Like Substances from Tomato Race 3 Strains of *Xanthomonas campestris* pv. *vesicatoria*. *Phytopathology* 93, 1415–1421. doi: 10.1094/PHYTO.2003.93.11.1415
- Vandroemme, J., Cottyn, B., Baeten, S., De Vos, P., and Maes, M. (2013). Draft genome sequence of *Xanthomonas fragariae* reveals reductive evolution and distinct virulence-related gene content. *BMC Genomics* 14:829 doi: 10.1186/1471-2164-14-829
- Vauterin, L., Hoste, B., Kersters, K., and Swings, J. (1995). Reclassification of *Xanthomonas*. *Int. J. Syst. Bacteriol.* 45, 472–489. doi: 10.1099/00207713-45-3-472
- Wasukira, A., Tayebwa, J., Thwaites, R., Paszkiewicz, K., Aritua, V., Kubiriba, J., et al. (2012). Genome-wide sequencing reveals two major sub-lineages in the genetically monomorphic pathogen *Xanthomonas campestris* pv. *musacearum*. *Genes* 3, 361–377. doi: 10.3390/genes3030361
- Wei, C.-F., Kvittko, B. H., Shimizu, R., Crabbill, E., Alfano, J. R., Lin, N.-C., et al. (2007). A *Pseudomonas syringae* pv. *tomato* DC3000 mutant lacking the type III effector HopQ1-1 is able to cause disease in the model plant *Nicotiana benthamiana*. *Plant J.* 51, 32–46. doi: 10.1111/j.1365-313X.2007.03126.x
- Wichmann, G., Ritchie, D., Kousik, C. S., and Bergelson, J. (2005). Reduced genetic variation occurs among genes of the highly clonal plant pathogen *Xanthomonas axonopodis* pv. *vesicatoria*, including the effector gene *avrBs2*. *Appl. Environ. Microbiol.* 71, 2418–2423. doi: 10.1128/AEM.71.5.2418-2432.2005
- Young, J. M., Park, D. C., Shearman, H. M., and Fargier, F. (2008). A multilocus sequence analysis of the genus *Xanthomonas*. *Syst. Appl. Microbiol.* 31, 366–377. doi: 10.1016/j.syapm.2008.06.004

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 Schwartz, Potnis, Timilsina, Wilson, Patané, Martins, Minsavage, Dahlbeck, Akhunova, Almeida, Vallad, Barak, White, Miller, Ritchie, Goss, Bart, Setubal, Jones and Staskawicz. 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.



Xanthomonas Whole Genome Sequencing: Phylogenetics, Host Specificity and Beyond

Alice Boulanger and Laurent D. Noël *

Laboratoire des Interactions Plantes-Microorganismes (LIPM), Université de Toulouse, Institut National de la Recherche Agronomique (INRA), Centre National de la Recherche Scientifique (CNRS), Université de Paul Sabatier (UPS), Castanet-Tolosan, France

Keywords: *Xanthomonas*, tomato, pepper, type III effector repertoire, host specificity

A commentary on

Phylogenomics of *Xanthomonas* field strains infecting pepper and tomato reveals diversity in effector repertoires and identifies determinants of host specificity

by Schwartz, A. R., Potnis, N., Timilsina, S., Wilson, M., Patané, J., Martins, J. Jr., et al. (2015). *Front. Microbiol.* 6:535. doi: 10.3389/fmicb.2015.00535

OPEN ACCESS

Edited by:

Choon-Min Ryu,
Korea Research Institute of Bioscience
and Biotechnology, South Korea

Reviewed by:

David John Studholme,
University of Exeter, UK
Tina Britta Jordan,
Eberhard Karls University Tübingen,
Germany

*Correspondence:

Laurent D. Noël
laurent.noel@toulouse.inra.fr

Received: 20 April 2016

Accepted: 30 June 2016

Published: 15 July 2016

Citation:

Boulanger A and Noël LD (2016)
Xanthomonas Whole Genome Sequencing: Phylogenetics, Host Specificity and Beyond.
Front. Microbiol. 7:1100.
doi: 10.3389/fmicb.2016.01100

Crop diseases impact both yield and product quality and result in important health, economic, environmental, and societal issues worldwide (Scholthof, 2003; Strange and Scott, 2005). These diseases are caused by microorganisms like viruses, fungi, and bacteria. Among phytopathogenic bacteria, the genus *Xanthomonas* comprises 20 species which together infect more than 400 plant species, among which several important crops such as rice, cassava, cabbages, citrus, or tomato (Mansfield et al., 2012). In contrast, individual *Xanthomonas* species have a narrow host range usually restricted to a specific plant genus or species. The genetic and molecular bases for host range determination in *Xanthomonas* are essentially unknown at this date (Jacques et al., 2016): It likely involves bacterial factors enhancing the physiological adaptation to the plant environment (e.g., tropism, attachment, nutrition, degradation of plant compounds) or limiting the elicitation of plant immunity (Büttner and Bonas, 2010). Elicitors of plant immunity include both pathogen-associated molecular patterns (PAMP) and type III effectors (T3E) (Jones and Dangl, 2006). PAMP-triggered immunity (PTI) is the first layer of plant's innate immunity and strongly restricts host range for most bacteria. Key to pathogenicity and host range expansion is pathogen's capacity to suppress PTI. This is the prime function of most T3E known to date. T3E proteins are directly injected inside plant cells using a molecular syringe known as the type III secretion system. As a countermeasure, given T3E might specifically elicit potent immune responses known as effector-triggered immunity (ETI) in a set of plant species resulting in a reduced host range. It is thus key to determine pathogen T3E repertoires. Large scale sequencing of collections of phytopathogenic bacteria at the inter- and intra-specific level has revealed a large diversity of T3E repertoires leading to the definition of core vs. accessory T3E (Potnis et al., 2011; Bart et al., 2012; Roux et al., 2015).

The paper by Schwartz et al. (2015) perfectly illustrates the power of comparative genomics to understand pathogenicity and host specificity of this important genus of pathogens. The authors focussed on bacterial spot disease of tomato and pepper caused by three distinct *Xanthomonas* species [*gardneri* (*Xg*), *perforans* (*Xp*), and *euviticatoria* (*Xe*)]. Draft genome sequences were determined for 67 strains sampled in the southeastern and midwestern United States between 1994 and 2013 and subjected to comparative analysis. Importantly, core genome analysis supported the recent reclassification of these strains in three distinct bacterial species (Jones et al., 2004) and evidenced further intraspecific organization of *Xp* strains. Identification, comparison, and analysis of T3E repertoires allowed the definition of surprisingly large core effectomes (24/29 T3E in *Xp*; 26/31 in *Xe*; 19/25

in *Xg*). Thirteen T3E are conserved among all three species. Identification of these core T3E and their sequence conservation should prove instrumental to design durable disease resistance strategies against bacterial spot disease in tomato and pepper. Such large interspecific core T3E repertoire might highlight the need for a future larger worldwide sampling of *Xanthomonas* infecting tomato and pepper. Beyond presence-absence of T3E, the conservation of T3E genes or proteins was also investigated in *Xe*, *Xp*, and *Xg* revealing the evolution and dynamics of these pathogenicity determinants. Overall, a lot more diversity is observed in T3E gene repertoires than in core genome analyses, suggesting a central role of horizontal gene transfer and distinct evolutionary forces applying to T3E genes. In *Xe*, the authors witnessed the evolution of T3E gene profiles over a 20 years period which could have suggested that such effect was driven by plant breeding and the introduction of novel tomato or pepper varieties. However, changes in prevalent bacterial species on tomato observed the last decade in different states (Florida, Ohio, Michigan) is not associated with strong changes in tomato cultivars used in commercial production fields (Potnis et al., 2015). Abiotic conditions could drive diversity and prevalent strains in diseased tomato fields. For instance, *Xg* is generally prevalent in cooler regions as observed for *Pseudomonas syringae* pv. *tomato*, another bacterial spot agent (Potnis et al., 2015). These two pathogens share several T3Es that are not present in other strains of *Xanthomonas* associated with bacterial spot. Horizontal gene transfer between different pathogens found in similar regions could be a strong source of diversity.

The second part of the analysis reports on host range determination in *Xp* by two T3E, AvrBsT and XopQ. Indeed,

the authors demonstrate that the loss of *avrBsT* alone explains *Xp* pathogenicity on pepper while the double loss of *avrBsT* and *xopQ* renders *Xp* pathogenic on *Nicotiana benthamiana*. This result illustrates how plant immunity might restrict host range by ETI activation.

The manuscript by Schwartz and colleagues thus illustrates the potential of large-scale pathogenomics in understanding pathogen evolution and disease emergence and offers perspectives to design durable resistance against bacterial spot in tomato and pepper. Such approaches would be further reinforced by an improved sampling design, an increased number of sequenced genomes, and the use of complete genomes now accessible by SMRT sequencing (Pacific Biosciences). Such bioinformatics analyses should now be extended to pan genomes, plasmids, InDels, SNPs, recombinations, and use the pioneering approach developed by Guttman and colleagues (McCann et al., 2012) which followed evolutionary forces applied to individual *P. syringae* genes resulting in the identification of several novel PAMP elicitors.

AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

AB and LN are supported by the LABEX TULIP (ANR-10-LABX-41 and ANR-11-IDEX-0002-02) and the CROptAL grant (ANR-14-CE19-0002-01).

REFERENCES

- Bart, R., Cohn, M., Kassen, A., McCallum, E. J., Shybut, M., Petriello, A., et al. (2012). High-throughput genomic sequencing of cassava bacterial blight strains identifies conserved effectors to target for durable resistance. *Proc. Natl. Acad. Sci. U.S.A.* 109, E1972–E1979. doi: 10.1073/pnas.1208003109
- Büttner, D., and Bonas, U. (2010). Regulation and secretion of *Xanthomonas* virulence factors. *FEMS Microbiol. Rev.* 34, 107–133. doi: 10.1111/j.1574-6976.2009.00192.x
- Jacques, M.-A., Arlat, M., Boulanger, A., Boureau, T., Carrère, S., Cesbron, S., et al. (2016). Ecology, physiology, and genomics to understand host specificity in *Xanthomonas*: French Network on *Xanthomonads* (FNX). *Annu. Rev. Phytopathol.* 54. doi: 10.1146/annurev-phyto-080615-100147. [Epub ahead of print].
- Jones, J. B., Lacy, G. H., Bouzar, H., Stall, R. E., and Schaad, N. W. (2004). Reclassification of the *Xanthomonads* associated with bacterial spot disease of tomato and pepper. *Syst. Appl. Microbiol.* 27, 755–762. doi: 10.1078/0723202042369884
- Jones, J. D., and Dangl, J. L. (2006). The plant immune system. *Nature* 444, 323–329. doi: 10.1038/nature05286
- Mansfield, J., Genin, S., Magori, S., Citovsky, V., Sriariyanum, M., Ronald, P., et al. (2012). Top 10 plant pathogenic bacteria in molecular plant pathology. *Mol. Plant Pathol.* 13, 614–629. doi: 10.1111/j.1364-3703.2012.00804.x
- McCann, H. C., Nahal, H., Thakur, S., and Guttman, D. S. (2012). Identification of innate immunity elicitors using molecular signatures of natural selection. *Proc. Natl. Acad. Sci. U.S.A.* 109, 4215–4220. doi: 10.1073/pnas.1113893109
- Potnis, N., Krasileva, K., Chow, V., Almeida, N. F., Patil, P. B., Ryan, R. P., et al. (2011). Comparative genomics reveals diversity among *Xanthomonads* infecting tomato and pepper. *BMC Genomics* 12:146. doi: 10.1186/1471-2164-12-146
- Potnis, N., Timilsina, S., Strayer, A., Shanharaj, D., Barak, J. D., Paret, M. L., et al. (2015). Bacterial spot of tomato and pepper: diverse *Xanthomonas* species with a wide variety of virulence factors posing a worldwide challenge. *Mol. Plant Pathol.* 16, 907–920. doi: 10.1111/mpp.12244
- Roux, B., Bolot, S., Guy, E., Denance, N., Lautier, M., Jardinaud, M. F., et al. (2015). Genomics and transcriptomics of *Xanthomonas campestris* species challenge the concept of core type III effectorome. *BMC Genomics* 16:975. doi: 10.1186/s12864-015-2190-0
- Scholthof, K. B. (2003). One foot in the furrow: linkages between agriculture, plant pathology, and public health. *Annu. Rev. Public Health* 24, 153–174. doi: 10.1146/annurev.publhealth.24.090302.155542
- Schwartz, A. R., Potnis, N., Timilsina, S., Wilson, M., Patane, J., Martins, J. Jr., et al. (2015). Phylogenomics of *Xanthomonas* field strains infecting pepper and tomato reveals diversity in effector repertoires and identifies determinants of host specificity. *Front. Microbiol.* 6:535. doi: 10.3389/fmicb.2015.00535
- Strange, R. N., and Scott, P. R. (2005). Plant disease: a threat to global food security. *Annu. Rev. Phytopathol.* 43, 83–116. doi: 10.1146/annurev.phyto.43.113004.133839

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 Boulanger and Noël. 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 genomics of a cannabis pathogen reveals insight into the evolution of pathogenicity in *Xanthomonas*

Jonathan M. Jacobs¹, Céline Pesce^{1,2}, Pierre Lefeuve³ and Ralf Koebnik^{1*}

OPEN ACCESS

Edited by:

Nicolas Denancé,
Institut National de la Recherche
Agronomique, France

Reviewed by:

Matthew James Moscou,
The Sainsbury Laboratory, UK
Sarah Grant,
University of North Carolina, USA

*Correspondence:

Ralf Koebnik,
Institut de Recherche pour le Développement, UMR Interactions – Plantes – Microorganismes – Environnement, Génomique et Transcriptomique des Interactions Plantes-Prokaryotes, 921 avenue Agropolis, 34394 Montpellier, France
koebnik@gmx.de

Specialty section:

This article was submitted to
Plant-Microbe Interaction,
a section of the journal
Frontiers in Plant Science

Received: 15 April 2015

Accepted: 27 May 2015

Published: 16 June 2015

Citation:

Jacobs JM, Pesce C, Lefeuve P and Koebnik R (2015) Comparative genomics of a cannabis pathogen reveals insight into the evolution of pathogenicity in *Xanthomonas*. *Front. Plant Sci.* 6:431.
doi: 10.3389/fpls.2015.00431

¹ Institut de Recherche pour le Développement – Cirad – Université Montpellier, Interactions Plantes Microorganismes Environnement, Montpellier, France, ² Department of Applied Microbiology, Earth and Life Institute, Université Catholique de Louvain, Louvain-la-Neuve, Belgium, ³ Pôle de Protection des Plantes, Cirad, UMR Peuplements Végétaux et Bioagresseurs en Milieu Tropical, Saint-Pierre, Ile de la Réunion, France

Pathogenic bacteria in the genus *Xanthomonas* cause diseases on over 350 plant species, including cannabis (*Cannabis sativa* L.). Because of regulatory limitations, the biology of the *Xanthomonas*-cannabis pathosystem remains largely unexplored. To gain insight into the evolution of *Xanthomonas* strains pathogenic to cannabis, we sequenced the genomes of two geographically distinct *Xanthomonas* strains, NCPPB 3753 and NCPPB 2877, which were previously isolated from symptomatic plant tissue in Japan and Romania. Comparative multilocus sequence analysis of housekeeping genes revealed that they belong to Group 2, which comprises most of the described species of *Xanthomonas*. Interestingly, both strains lack the Hrp Type III secretion system and do not contain any of the known Type III effectors. Yet their genomes notably encode two key Hrp pathogenicity regulators HrpG and HrpX, and *hrpG* and *hrpX* are in the same genetic organization as in the other Group 2 xanthomonads. Promoter prediction of HrpX-regulated genes suggests the induction of an aminopeptidase, a lipase and two polygalacturonases upon plant colonization, similar to other plant-pathogenic xanthomonads. Genome analysis of the distantly related *Xanthomonas maliensis* strain 97M, which was isolated from a rice leaf in Mali, similarly demonstrated the presence of HrpG, HrpX, and a HrpX-regulated polygalacturonase, and the absence of the Hrp Type III secretion system and known Type III effectors. Given the observation that some *Xanthomonas* strains across distinct taxa do not contain *hrpG* and *hrpX*, we speculate a stepwise evolution of pathogenicity, which involves (i) acquisition of key regulatory genes and cell wall-degrading enzymes, followed by (ii) acquisition of the Hrp Type III secretion system, which is ultimately accompanied by (iii) successive acquisition of Type III effectors.

Keywords: comparative genomics, *Xanthomonas*, hemp, cell-wall degrading enzymes, type II secretion system, type III secretion system, *hrp* genes, PIP box

Introduction

Plant pathogenic bacteria in the genus *Xanthomonas* collectively cause major losses worldwide on over 350 plant species, including crops such as banana, tomato, pepper, sugar cane, and many cereals. Over 20 *Xanthomonas* species are divided into two main phylogenetic groups based on 16S rDNA and *gyrB* sequence analysis (Hauben et al., 1997; Parkinson et al., 2007) and subdivided into pathovars loosely corresponding to host specificity. Group 1, also known as the early branching group, comprises highly diverse *Xanthomonas* species including important sugarcane and cereal pathogens (e.g., *Xanthomonas sacchari*, *Xanthomonas albilineans* and *Xanthomonas translucens*, Hauben et al., 1997; Parkinson et al., 2007). Group 2, the largest and best-described group, includes species such as *Xanthomonas oryzae*, *Xanthomonas citri*, *Xanthomonas vasicola*, *Xanthomonas euvesicatoria*, *Xanthomonas axonopodis* and *Xanthomonas campestris* (Hauben et al., 1997; Parkinson et al., 2007). This diverse genus of bacteria infects and associates with many plant hosts, but individual strains typically possess very restricted host ranges limited to a single genus.

Xanthomonas spp. employ a suite of virulence factors to colonize plant tissue, including adhesins, cell wall-degrading enzymes, extracellular polysaccharide and protein secretion systems (Büttner and Bonas, 2010). The Hrp (hypersensitive response and pathogenicity) Type III secretion system (T3SS) is a major virulence trait found in most pathogenic *Xanthomonas* spp. and serves as a molecular syringe to deliver effector proteins into host cells to suppress defenses and modulate plant physiology to promote pathogen growth (White et al., 2009). Plants also evolved resistance proteins that recognize pathogen avirulence effectors and inhibit infection often via a hypersensitivity response (HR), a form of programmed cell death (Bent and Mackey, 2007). A majority of sequenced pathogenic *Xanthomonas* strains have limited host ranges likely due to the plant recognition of Type III (T3)-secreted avirulence effectors (White et al., 2009). In *Xanthomonas* spp., HrpX, an AraC-type regulator, is the transcriptional activator of the genes encoding the T3SS and many of its associated effectors (Koebnik et al., 2006; Tang et al., 2006). HrpG, an OmpR-family and major pathogenicity regulator, positively regulates expression of *hrpX* (Tang et al., 2006). Mutant strains lacking either *hrpX* and *hrpG* are unable to activate expression of the T3SS and thus are non-pathogenic (Wengelnik et al., 1996; Tang et al., 2006; Mole et al., 2007). The importance of the T3SS and many T3-secreted effectors during infection is heavily studied, but the evolutionary history of the acquisition of genes encoding the T3SS, associated T3-secreted effectors and regulators, HrpX and HrpG, remains unclear.

Hemp or cannabis (*Cannabis sativa* L.) is a major, global cash crop with many applications such as seed for human consumption, oil, fiber for clothing or ropes, pulp for paper, plastic and composite material (www.hemp.com). Since 2010 worldwide hemp production has increased, and recent surges of hemp production in the United States, China, Australia, Canada, and many other countries have made hemp a multi-million dollar industry (www.hemp.com, www.faostat.fao.org). A draft

genome is now available for *C. sativa* cv. Purple Kush (van Bakel et al., 2011), potentially providing a base for molecular and evolutionary understanding of this plant species. Hemp plant production is limited by bacteria, fungi, nematodes, and viruses (McPartland et al., 2000), but because of regulatory constraints, little is known about hemp diseases such as bacterial leaf spot of cannabis caused by *Xanthomonas* species.

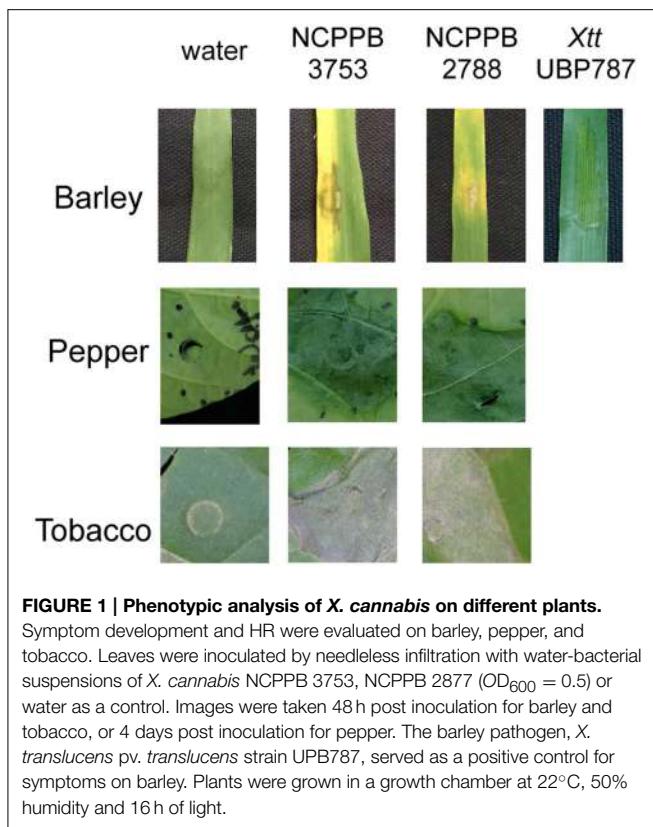
Symptoms associated with *Xanthomonas* bacterial leaf spot include water-soaking lesions followed by necrosis accompanied by a yellow halo (Severin, 1978; Netsu et al., 2014). The host range of these *Xanthomonas* strains appears to be quite large unlike most xanthomonads (Severin, 1978; Netsu et al., 2014). Under laboratory conditions, these bacteria caused symptoms on a wide range of plants including cannabis, tomato, mulberry, geranium and *Ficus erecta* (Severin, 1978; Netsu et al., 2014). These strains further trigger an HR on tobacco, but do not elicit any response after inoculation on common bean (Severin, 1978; Netsu et al., 2014). The factors that contribute to pathogenicity and host range of cannabis-infecting *Xanthomonas* are unknown.

To gain insight into the evolution and pathogenicity of bacterial pathogens of cannabis, we sequenced two geographically distinct *Xanthomonas* strains, NCPPB 3753 and 2877, which were previously isolated from symptomatic hemp leaf tissue from Japan and Romania, respectively (Severin, 1978; Netsu et al., 2014). We tested their ability to infect barley, a previously unreported, compatible monocot host. We determined with comparative whole genome analysis based on average nucleotide identity (ANI) and multilocus sequence analysis (MLSA) the relationship of these cannabis strains to each other and other xanthomonads. We provide evidence that NCPPB 3753 and NCPPB 2877 form a unique species in the genus *Xanthomonas* herein called *Xanthomonas cannabis*. We further describe likely virulence traits encoded by their genomes. Most notably these strains lack a Hrp T3SS but possess the major *hrp* virulence regulators HrpX and HrpG. Based on our comparative genomic analysis in *X. cannabis*, we provide a putative model for acquisition of the T3SS, T3-secreted effectors and the *hrp* regulators in *Xanthomonas* spp.

Results/Methods/Discussion

Phenotypic Evaluation

Two representative strains of *X. cannabis* (also known as *Xanthomonas campestris* pv. *cannabis*), isolated from symptomatic hemp leaves (*C. sativa* L.), were chosen for genome sequencing. Type strain NCPPB 2877 was isolated by I. Sandru at the Lovrin station in the Timiș județ (Romania) in 1974 (Severin, 1978), and strain NCPPB 3753 was isolated by Y. Takiwawa in the Kanuma region of Tochigi Prefecture (Japan) in 1982 (isolate SUPP546; Netsu et al., 2014). *X. cannabis* strains were previously reported to cause disease on many dicot host plants. To determine if *X. cannabis* could infect a monocot host, we inoculated barley (*Hordeum vulgare* L. cv. Morex) leaves with the cannabis strains by infiltration. Overnight cultures grown using PSA (Tsuchiya et al., 1982) or NB medium (Sigma-Aldrich, USA) were pelleted and resuspended in water. Plant leaves



were infiltrated by a needleless syringe with a water-bacterial suspension or water as a control. Leaves developed necrosis around the zone of infiltration followed by leaf yellowing (**Figure 1**). These symptoms closely resembled the leaf spot symptoms on *C. sativa* observed by previous characterizations (Severin, 1978; Netsu et al., 2014). Similar symptoms were observed with lower inoculum ($OD_{600} = 0.05$ and 0.1). Tomato is a compatible host for *X. cannabis* (Severin, 1978), and therefore we decided to test *X. cannabis* virulence of pepper, another solanaceous plant. Pepper leaves were infiltrated with strains NCPPB 2877 and NCPPB 3753 as with barley. Pepper plants displayed water-soaked lesions 48 h post inoculation (**Figure 1**). Both cannabis strains elicited an HR when inoculated on tobacco (**Figure 1**), but the nature of this HR remains to be determined.

Genome Sequencing and Annotation

The genomes of strains NCPPB 2877 and NCPPB 3753 were sequenced using the Illumina Hi-Seq2500 platform (Fasteris SA, Switzerland). The shotgun sequencing yielded 2,921,175 100-bp paired-end reads (730 Mb) for strain NCPPB 2877 and 2,464,521 paired-end reads (616 Mb) for strain NCPPB 3753, with insert sizes ranging from 250 bp to 1.5 kb. Draft genome sequences were assembled using the Edena algorithm v3.131028 (Hernandez et al., 2014), yielding 257 contigs ≥ 200 bp ($N_{50} = 38,306$ bp) with $69 \times$ coverage for strain NCPPB 2877 and 260 contigs ($N_{50} = 35,229$ bp) with $73 \times$ coverage for strain NCPPB 3753. For comparison, draft genome sequences were also assembled using the Velvet algorithm v1.1.04 (Zerbino and Birney, 2008), yielding 564

contigs ≥ 200 bp ($N_{50} = 15,608$ bp) for strain NCPPB 2877 and 469 contigs ($N_{50} = 20,963$ bp) for strain NCPPB 3753. Because of their better quality, Edena-derived contigs were annotated with GeneMarkS + release 2.9 (revision 452131) (Borodovsky and Lomsadze, 2014), as implemented in the NCBI Prokaryotic Genome Annotation Pipeline (http://www.ncbi.nlm.nih.gov/genome/annotation_prok/), which predicted a total of 4095 genes within 4,756,730 bp for strain NCPPB 2877 and 4160 genes within 4,837,471 bp for strain NCPPB 3753. These whole genome shotgun projects have been deposited at DDBJ/EMBL/GenBank under the accession no. JSZE00000000 (NCPPB 2877) and JSZF00000000 (NCPPB 3753). The versions described in this paper are the first versions, JSZE01000000 and JSZF01000000.

Comparison of the Two Genome Sequences

ANI provides a robust method to determine bacterial species definition based on whole genome sequence comparison and is considered the new standard for species definition (Konstantinidis and Tiedje, 2005; Figueras et al., 2014). To determine if NCPPB 2877 and NCPPB 3753 are the same species, the ANI was calculated for both genome sequences using JSpecies (Richter and Rosselló-Móra, 2009). BLAST-based comparison revealed 99.2% ANI for the 92.5% sequences that could be aligned, and MUMmer-based comparison revealed 99.1% ANI for the 96.9% sequences that could be aligned, thus confirming that both strains belong to the same species.

Using our web-based pipeline for prediction of satellites (<http://www.biopred.net/VNTR/>), we then evaluated whether or not both strains belong to a clonal complex. For satellite prediction, the following parameters were chosen (Zhao et al., 2012): algorithm, TRF; region length, 30–1000 bp; unit length, 5–12 bp; and at least 6 tandem repeats with a similarity of at least 80% among the repeats. In total, 45 microsatellites were predicted, 35 of which were found to be present in both genome sequences. For 34 of them, repeat numbers could be derived; while one locus was not informative because it was located at the end of two contigs and thus not completely assembled in NCPPB 3753. To provide further evidence that the calculated repeat numbers were meaningful, the corresponding loci were also analyzed in the Velvet-based genome assemblies. Strikingly, there was not a single discrepancy between the Edena- and Velvet-based data, except for the fact that some satellite loci were not completely assembled by Velvet while they were complete in the Edena assembly. For the complete loci, 28 loci (82%) were different between the two strains with respect to repeat numbers. For the six loci with identical repeat numbers DNA sequence analysis revealed that five of them were identical due to homoplasy, i.e., these loci evolved by convergent evolution to the same number of repeats. Thus, both strains differ by almost all (97%) of their completely assembled microsatellite loci, a finding that indicates that both strains do not belong to a clonal complex.

Taxonomic Position of the Two Cannabis Pathogens

Comparison of 16S rDNA sequences is a method of choice to elucidate the taxonomic positions of bacterial strains, and

was previously used to analyze and delineate 20 species of *Xanthomonas* (Hauben et al., 1997). It was found that the genus *Xanthomonas* exhibited a relatively high level of 16S rDNA sequence identity, with on average 14 single-nucleotide polymorphisms (SNPs) between two different *Xanthomonas* species (Hauben et al., 1997). The 16S rDNA sequences of both cannabis pathogens were found to be identical. When we compared the 16S rDNA sequence of the cannabis pathogens with those of the 20 *Xanthomonas* type strains, the cannabis pathogen grouped with Group 2 strains, which contains the majority of characterized *Xanthomonas* species. Interestingly, GenBank comparison revealed that another recently sequenced strain that was isolated from symptomatic bean plants in Rwanda, Nyagatare, contains the same 16S rDNA sequence (Aritua et al., 2015).

Previously *X. cannabis* strains were also called *X. campestris* pv. *cannabis* based on the similarity of the 16S rDNA sequence to *X. campestris*, but it has been suggested that the name should be changed to *X. cannabis* (Netsu et al., 2014). Since the resolution of the 16S rDNA sequence is very low within Group 2 strains (Hauben et al., 1997) and often only distinguishes a species by one or two SNPs, we performed whole-genome comparisons including one representative strain per species for which genome sequences were available (Figure 2). The pairwise ANI of the two cannabis strains against any of the representative strains was

below 90%, regardless of which algorithm (BLAST or MUMmer) was used, indicating that these two strains belong to an unique and distinct *Xanthomonas* species (Figure 2). We suggest that *X. cannabis* is the appropriate name for this bacterial species based on our ANI analysis and as previously suggested by Netsu et al. (2014).

Guided by the observation that their 16S rDNA sequences were identical to that of the Nyagatare strain, we compared the genomes of *X. cannabis* NCPPB 3753 and NCPPB 2877 and *X. sp.* Nyagatare. JSpecies calculations revealed that the two cannabis strains were 96.3–96.4% identical to the Nyagatare strain, when calculated over the 88.6–91.0% of the genome sequence that could be aligned by the more robust MUMmer algorithm (Richter and Rosselló-Móra, 2009). These values are slightly above the ≈95–96% transition zone, above which strains can be considered to belong to the same taxonomically circumscribed prokaryotic species (Konstantinidis and Tiedje, 2005). Therefore the Nyagatare strain most likely belongs to the *X. cannabis* species. It would be interesting to perform functional studies to determine the similarities and differences between these closely related *X. cannabis* strains because the Nyagatare strain is a reported bean pathogen and *X. cannabis* strains elicit no response when inoculated on bean (Netsu et al., 2014).

Partial sequencing of the *gyrB* and other housekeeping genes for MLSA grouped all *Xanthomonas* species into four major

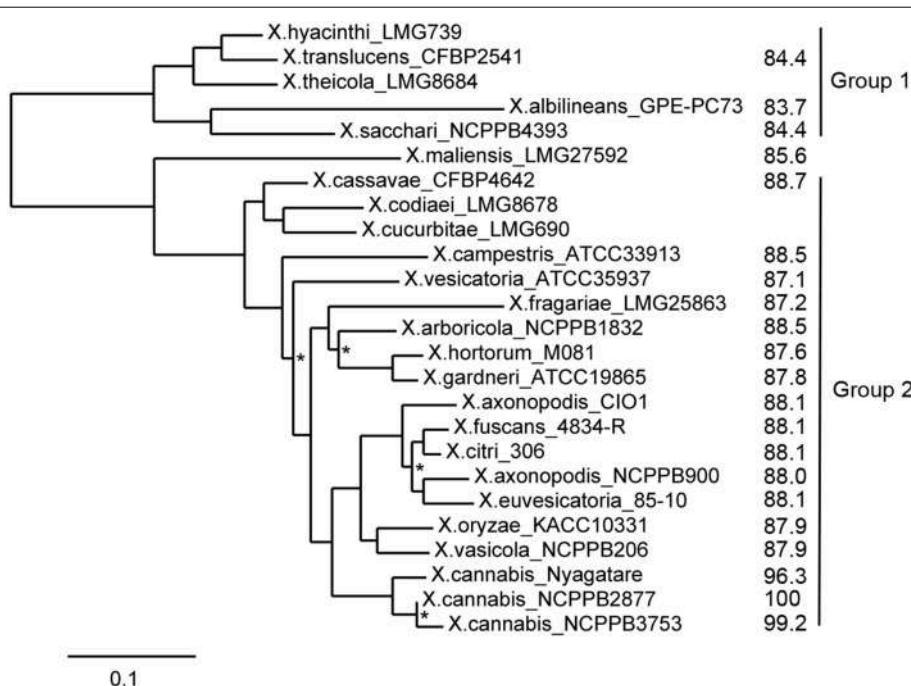


FIGURE 2 | ANI and MLSA-based phylogenetic tree of xanthomonads.

Phylogenetic analysis was performed on the Phylogeny.fr platform (Dereeper et al., 2008). Sequences were aligned with MUSCLE (v3.7) configured for highest accuracy (MUSCLE with default settings). After alignment, ambiguous regions (i.e., containing gaps and/or poorly aligned) were removed with Gblocks (v0.91b) using default parameters. The phylogenetic tree was reconstructed using the maximum likelihood method implemented in the PhyML program (v3.0). The HKY85 substitution model was selected

assuming an estimated proportion of invariant sites and 4 gamma-distributed rate categories to account for rate heterogeneity across sites. The gamma shape parameter was estimated directly from the data. Reliability for internal branch was assessed using the aLRT test (SH-Like). Graphical representation and edition of the phylogenetic tree were performed with TreeDyn (v198.3). All nodes were supported by bootstrap values above 0.9, except for those marked with an asterisk. Calculated ANI values in comparison to NCPPB 2877 and the two major phylogenetic groups are indicated on the right side.

MLSA subgroups (Parkinson et al., 2007; Young et al., 2008). We used MLSA of seven housekeeping genes (*atpD*, *dnaK*, *efp*, *glnA*, *gyrB*, *lepA*, and *rpoD*) and ANI calculations to understand the relationship of the cannabis and Nyagatare strains compared to other xanthomonads. The *X. cannabis* NCPPB 2877 and NCPPB 3753 and Nyagatare strains appear to form a distinct grouping according to MLSA (Figure 2). Based on the comparisons of ANI and of seven concatenated internal portions of the genes, we overall conclude that the cannabis and Nyagatare strains belong to a single species, *X. cannabis* because: (1) they are above the 95–96% ANI threshold for species definition and (2) they appear to belong to a new MLSA clade (Figure 2). This new clade corresponds to the novel species-level clade (slc) 1, as suggested by Parkinson and co-workers, which also contains the pathovars *esculenti* and *zinniae* (Parkinson et al., 2009). Following the pathovar designation, we suggest to name the two cannabis strains *X. cannabis* pv. *cannabis*, and the bean-pathogenic Nyagatare strain *X. cannabis* pv. *phaseoli*.

Comparison of Pathogenicity-related Gene Clusters

Several gene clusters are considered to be important for pathogenicity of xanthomonads and their possible contribution to host- and tissue-specificity has been studied previously (Lu et al., 2008). We therefore analyzed whether and to which extent these gene clusters are conserved in the cannabis strains.

The *rpf* (regulation of pathogenicity factors) gene cluster plays a role in the intercellular signal-response system that links synthesis and perception of the diffusible signal factor (DSF) *cis*-11-methyl-2-dodecanoic acid to the synthesis of extracellular enzymes, extracellular polysaccharide, and biofilm dispersal (Dow, 2008). The genetic organization of the cannabis *rpf* gene cluster resembles that of *X. euvesicatoria* strain 85–10, yet it contains an additional gene, *rpfI*, downstream of gene XCV1913, which is not present in *X. euvesicatoria* strain 85–10. Moreover, the proteins RpfF, responsible for DSF synthesis, and the two-component regulatory system RpfC/RpfG, which is involved in DSF perception and signal transduction (Dow, 2008), are highly conserved in the cannabis pathogens.

The *gum* gene clusters encode proteins that are involved in the exopolysaccharide (EPS) biosynthesis (Becker et al., 1998). The core *gum* gene cluster, consisting of *gumB* to *gumM*, all transcribed in the same direction, is entirely conserved in the cannabis pathogen. As in other xanthomonads, the first gene, *gumB*, is located downstream of a proline-specific tRNA gene. The *gumA* gene which is located upstream of the tRNA gene and downstream *pheS* and *pheT* is most likely not a *bona fide gum* gene. The accessory *gumN* gene, which for instance is disrupted in *X. euvesicatoria* strain 85–10 and *X. oryzae* pv. *oryzicola* strain BLS256, appears to be intact in the cannabis pathogen. *gumO* and *gump*, annotated as a 3-oxoacyl-(acyl carrier protein) synthase and a metal-dependent hydrolase, respectively, are located directly downstream of *gumN*. The contribution of these two genes to EPS biosynthesis is questionable since it is not present in several xanthomonads infecting monocotyledons, such as *X. sacchari*, *X. translucens* and *X. oryzae*, and it is also not present in the recently sequenced *Xanthomonas maliensis* strain

97M, which was isolated from rice leaves in Mali. On the other side, this unique conservation in bacteria colonizing eudicots could suggest a role in host specificity at the level of a division.

Lipopolysaccharide (LPS) is another bacterial polysaccharide, which is firmly attached to the outer membrane, and an aberrant structure of the LPS O-chain has been linked to virulence defects (Mhedbi-Hajri et al., 2011). The gene cluster responsible for LPS biosynthesis is always present between the highly conserved *etfA* and *metB* genes of *Xanthomonas*. A remarkably high degree of variation both in number and in identity of LPS genes has been found in different xanthomonads, even within a single species or pathovar (Patil et al., 2007). In the two cannabis pathogens, 14 genes are predicted to participate in the LPS biosynthesis. The gene content and genetic organization is largely identical to a few other *Xanthomonas* strains belonging to different species: *Xanthomonas gardneri* strain ATCC 19865, *Xanthomonas vesicatoria* strain ATCC 35937, and strains of *X. campestris* (B100, CN14, CN15, CN16, JX, Xca5, and 756C); except that the *X. campestris* strains have a short, additionally predicted gene, *wxcH*, just downstream of the *etfA* gene, which is not present (or frame-shifted) in the other species. Interestingly, the relatively closely related Nyagatare strain shares only the four *etfA*-proximal genes, *wxcK* to *wxcN*, and the two *metB*-proximal genes, *wzm* and *wzt*, with the cannabis strains, while between them three large genes without orthologs in other xanthomonads are predicted. As observed before on a smaller set of *Xanthomonas* genome sequences (Patil et al., 2007; Lu et al., 2008), the cannabis pathogens underscore the fact that the LPS gene cluster varies highly between and within species, suggesting multiple horizontal gene transfers and re-assortments. Yet, it is unclear what evolutionary pressure drives this variation—is it escape from recognition by plant immune receptors, and/or is it escape from recognition by bacteriophages?

Motility is an important feature of bacteria that is governed by flagella-based swimming and/or pilus-based twitching or gliding (Rossez et al., 2015). The flagellum is a complex machinery, the biosynthesis of which depends on dozens of genes, collectively called *fle*, *flg*, *flh*, and *fli* genes, which are organized in large gene clusters. The cannabis strain NCPPB 2877 contains the full gene complement of flagellar biosynthesis. In contrast, strain NCPPB 3753 appears to have lost two large regions in its two genes clusters. One deletion encompasses the genes *flgG* to *flgJ* and the 5' portion of *flgK*. Downstream of *flgF*, remnants of an IS element were present in the genome assembly. The second, even larger deletion goes from the 3' portion of *fliK* until the end of *flhA*, totaling to more than 11 genes. Absence of flagella under standard culture conditions was confirmed for strain NCPPB 3753 by a swimming assay in semisolid medium (Figure 3). Briefly overnight cultures grown in liquid NB medium were fixed to an $OD_{600} = 1.0$ and stabbed into semi-solid NYGA medium (0.3% agar). Motility was defined as observed turbid growth outside the stab zone of inoculation in the semi-solid medium. Tubes inoculated with NCPPB 3753 were turbid after 4 days of growth (Figure 3), which demonstrates that this strain is motile. Growth of NCPPB 2877 was confined to the stab zone of inoculation (Figure 3), which supports our hypothesis that NCPPB 2877 is non-motile because it lacks a flagellum.

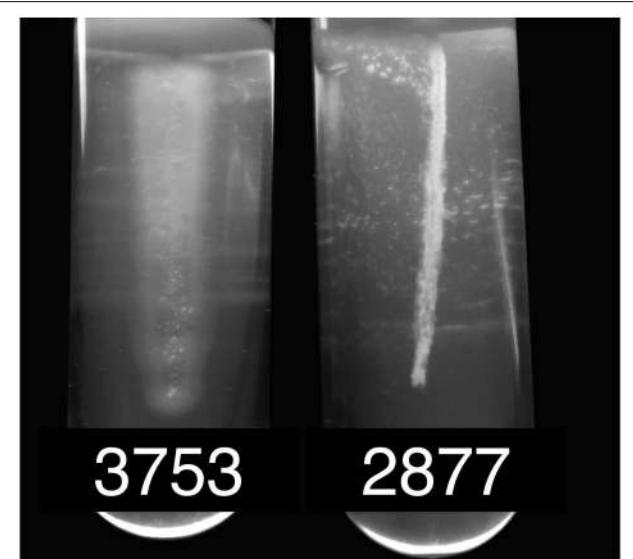


FIGURE 3 | Motility of *X. cannabis* strains. Bacterial motility was determined by stab inoculation in motility medium of either strain NCPPB 3753 or NCPPB 2877. Motility was evaluated qualitatively by turbid growth (motile) compared to localized, fixed growth (non-motile) around the zone of inoculation. Bacteria were grown in semi-solid NYGA agar medium (0.3%) as previously described (Sun et al., 2006).

Loss of flagella in xanthomonads is not without precedent (Darrasse et al., 2013). When Darrasse and co-workers tested 300 *Xanthomonas* strains representing different species and pathovars, five percent of the tested strains turned out to carry a deletion in the flagellar cluster and were non-motile. Co-isolation of flagellate and non-flagellate variants from an outbreak suggested that flagellar motility is not essential for fitness within the plant and that mixed populations could be a strategy to avoid detection by the plant defense system (Darrasse et al., 2013). Indeed, many plants have evolved receptor-like kinases (e.g., FLS2) that detect peptide epitopes of the FliC subunit of the flagellum (e.g., flg22 and flgII-28) (Sun et al., 2006), and it has been reported that otherwise isogenic flagellin-deficient *Xanthomonas* bacteria have a colonization advantage over flagellated bacteria in citrus leaves (Shi et al., 2015). Interestingly, when we analyzed the FliC flagellin sequence of the two cannabis strains, we detected a flg22 variant that deviates from the consensus sequence by as much as nine amino acid residues and is identical to the flg22 sequences from other xanthomonads, such as *X. oryzae* and *X. vasicola* (**Supplemental Figure S1**). In contrast, the related Nyagatare strain has a flg22 sequence that only deviates from the consensus by four residues and is similar to the elicitation-active flagellins that are recognized by FLS2 in *Arabidopsis* (Sun et al., 2006). These findings suggest that the cannabis strains use at least two out of several strategies to escape from recognition by the plant immune surveillance system, either by loss of the flagellum or by allelic variation of flagellin epitope(s) (Rossez et al., 2015).

The cannabis pathogen contains two different Type II protein secretion systems, the Xcs and the Xps system, as found in

several other xanthomonads (Lu et al., 2008). The xcs gene cluster, consisting of 12 genes, *xcsC* to *xcsN*, is located downstream of a TonB-dependent receptor gene and upstream of a GntR-family regulator gene, which is also the case for *X. euvesicatoria* strain 85–10 and strains of *X. campestris*. The related Nyagatare strain contains another four genes between *xcsN* and the GntR-family regulator gene, a feature that is shared with the *X. citri* pv. *citri* strain 306 and the *Xanthomonas arboricola* strain 3004. Interestingly, the xcs gene cluster is also present in the *X. malensis* strain 97M, but absent in the Group 1 strains *X. albilineans* GPE PC-73, *X. sacchari* NCPPB 4393, and *X. translucens* pv. *cerealis* CFBP 2541. The second Type II secretion system, encoded by the xps gene cluster, consists of 11 genes, *xpsE* to *xpsN*, followed by *xpsD*, and is found between the adhesion gene *xadA* and the *pnuC* gene. This position in the bacterial chromosome is largely conserved among *Xanthomonas* strains belonging to different Groups, such as the *X. sacchari* strain NCPPB 4393, the *X. translucens* pv. *cerealis* strain CFBP 2541, the *X. malensis* strain 97M, and strains of Group 2.

Many Gram-negative plant-pathogenic bacteria have evolved another protein secretion system, the T3SS, which plays a pivotal role in the pathogen-host interaction (Büttner, 2012). The Hrp system serves as a molecular syringe that injects T3 effector proteins into the cytosol of the host's cells at the benefit of the bacterium. Yet, many injected effectors can also betray the pathogen to the host plant, triggering a strong defense response, which typically is accompanied by an effector-triggered HR (White et al., 2009). Notably the early-branching Group 1 sugarcane pathogen *X. albilineans* does not have an Hrp T3SS, a finding that could be attributed to the reduced genome size of this pathogen (Pieretti et al., 2009). Since then, new genome sequences became available, including other Group 1 strains belonging to the species *X. sacchari* (Studholme et al., 2011). Some of these strains do not appear to have undergone extensive genome reduction, yet, they do not encode a T3SS. When we analyzed the gene content of the cannabis strains we realized that they do not have any of the *hrc*, *hrp*, or *hpa* genes encoding the T3SS. We were also not able to detect any traces of a T3-effector gene when using the set of described genes as queries (<http://www.xanthomonas.org/t3e.html>). In contrast, the Nyagatare strain has a full gene complement for the Hrp T3SS and also possesses a couple of effector proteins (Aritua et al., 2015). This is an interesting finding for several reasons. First, how do the cannabis pathogens suppress plant defense, a function that became more and more linked to the activity of T3-effectors (e.g., Canonne et al., 2011; Schulze et al., 2012; Sinha et al., 2013; Li et al., 2015; Stork et al., 2015)? Second, which molecular entities are triggering the HR in *Nicotiana tabacum* if not type III effectors? Interestingly, a type II-secreted pectate lyase, XagP, from *X. axonopodis* pv. *glycines* was previously found to be associated with HR induction on tobacco and pepper, but not on cucumber, sesame and tomato, thus giving a prime example that cell wall-degrading enzymes could trigger an HR (Kaewnum et al., 2006). We therefore speculate that similar or other secreted enzymes that disturb the integrity of the plant cell wall could be responsible for HR triggered by the cannabis pathogen independent of a T3SS. Last but not least, these Group

2 genome sequences may serve as a valuable resource to predict new T3 effectors. As a relative of other xanthomonads it will provide a valuable negative training or filtering set for prediction algorithms.

Analysis of the HrpX Regulon

Previous work has shown that the Hrp T3SS and many of its secreted effectors are controlled by a regulatory cascade, consisting of HpaR2/HpaS, HrpG, and HrpX (Büttner and Bonas, 2010; Li et al., 2014). However, HrpG and HrpX regulate pathogenesis beyond the T3SS alone (Guo et al., 2011). We therefore wondered if these components are present in *Xanthomonas* strains that do not possess the Hrp T3SS, and if so, what genes would be controlled by these components. BLAST searches revealed that all four regulatory genes are conserved in the cannabis pathogens (Table 1). *hpaR2* and *hpaS* are also conserved in the *X. maliensis* strain and in strains from Group 1, such as *X. albilineans*, *X. sacchari* and *X. translucens*. Their location is always between the *glmS* and *glmU* gene, with some extra genes for an efflux pump between *glmS* and *hpaR2*. In contrast, *hrpG* and *hrpX* were not found in strains of *X.*

albilineans or *X. sacchari*, but they are present in *X. translucens*. Interestingly, however, *hrpG* and *hrpX* from *X. translucens* cluster together with the rest of the T3SS *hrp* gene cluster while all other xanthomonads have *hrpG* and *hrpX* at a distinct genomic location between *radA* and *hsp70* (Wichmann et al., 2013). We could not determine their genomic organization for *X. maliensis* due to the highly fragmented genome assembly.

Within the Hpa-Hrp regulatory cascade, HrpX is the most downstream component that directly induces the synthesis of pathogenicity factors, such as the Hrp T3SS, T3 effectors and cell wall-degrading enzymes, by binding to a conserved *cis* element, called PIP (plant-induced promoter) box, within the promoter regions of the corresponding genes (Koebnik et al., 2006). For gene activation, both the PIP box as well as a properly spaced –10 promoter motif are required (Furutani et al., 2006). We therefore analyzed the genome sequences of the cannabis pathogens for the presence of the promoter motif TTCGB-N₁₅-TTCGB-N_{30–32}-TYNNNT (B represents C, G, or T; Y represents C or T) (Koebnik et al., 2006). Using this conservative query, four genes were strongly predicted to belong to the HrpX regulon, two polygalacturonases (PehA, synonymous with

TABLE 1 | Presence of type III secretion systems, type III effectors, *hrp/hpa* regulatory genes, and homologs of predicted HrpX regulon members of *X. cannabis* pv. *cannabis* in representative strains of *Xanthomonas*.

| Group | Species | Strain | GenBank accession no. | Hrp T3SS | T3Es | HrpG HrpX | HpaS HpaR2 | PehA | PehD | Amino-peptidase | LPL |
|-------|--|------------|-----------------------|----------|------|-----------|-----------------|------|--------|-----------------|--------|
| 1 | <i>X. albilineans</i> | GPE PC73 | FP565176 | no | no | no | YES | YES | YES | no | YES |
| 1 | <i>X. sacchari</i> | NCPPB 4393 | AGDB01000001 | no | no | no | YES | YES | no | YES | YES |
| 1 | <i>X. translucens</i> pv. <i>cerealis</i> | CFBP 2541 | JWHD01000001 | YES | YES | YES | no ^a | YES | YES | Ψ | YES |
| | <i>X. maliensis</i> | 97M | AQPR01000001 | no | no | YES | YES | no | PIP | PIP[5] | PIP[1] |
| 2 | <i>X. campestris</i> pv. <i>campestris</i> | ATCC 33913 | AE008922 | YES | YES | YES | YES | PIP | PIP[1] | PIP[1] | PIP |
| 2 | <i>X. vesicatoria</i> | ATCC 35937 | AEQV01000004 | YES | YES | YES | YES | PIP | PIP | PIP[1] | PIP |
| 2 | <i>X. vasicola</i> pv. <i>vasculorum</i> | NCPPB 206 | AKBM01000001 | YES | YES | YES | YES | PIP | no | PIP[1] | PIP |
| 2 | <i>X. oryzae</i> pv. <i>oryzae</i> | KACC110331 | AE013598 | YES | YES | YES | no ^b | Ψ | YES | no | PIP |
| 2 | <i>X. fuscans</i> subsp. <i>fuscans</i> | 4834-R | FO681494 | YES | YES | YES | YES | PIP | PIP | PIP[1] | PIP |
| 2 | <i>X. citri</i> pv. <i>citri</i> | 306 | AE008923 | YES | YES | YES | YES | PIP | PIP | PIP[1] | PIP |
| 2 | <i>X. axonopodis</i> pv. <i>manihotis</i> | CIO1 | AKCZ01000001 | YES | YES | YES | YES | PIP | no | PIP[1] | PIP-Ψ |
| 2 | <i>X. axonopodis</i> pv. <i>vasculorum</i> | NCPPB 900 | JPHD01000001 | YES | YES | YES | no ^c | PIP | PIP | PIP[1] | PIP |
| 2 | <i>X. euvesicatoria</i> | 85-10 | AM039952 | YES | YES | YES | YES | PIP | PIP | PIP[1] | PIP |
| 2 | <i>X. cassavae</i> | CFBP 4642 | ATMC01000001 | YES | YES | YES | YES | PIP | PIP | PIP[2] | PIP[3] |
| 2 | <i>X. arboricola</i> pv. <i>celebensis</i> | NCPPB 1832 | JPHC01000001 | YES | YES | YES | YES | PIP | no | PIP[1] | PIP[1] |
| 2 | <i>X. hortorum</i> pv. <i>carotae</i> | M081 | AEEU01000001 | YES | YES | YES | YES | PIP | PIP | PIP[1] | PIP |
| 2 | <i>X. gardneri</i> | ATCC 19865 | AEQX01000001 | YES | YES | YES | YES | PIP | PIP | PIP[1] | PIP |
| 2 | <i>X. fragariae</i> | LMG 25863 | AJRZ01000001 | YES | YES | YES | YES | PIP | PIP | Ψ | Ψ |
| 2 | <i>X. cannabis</i> pv. <i>phaseoli</i> | Nyagatare | JRQI01000001 | YES | YES | YES | YES | PIP | PIP | PIP | PIP |
| 2 | <i>X. cannabis</i> pv. <i>cannabis</i> | NCPPB 2877 | JSZE01000001 | no | no | YES | YES | PIP | PIP | PIP | PIP |
| 2 | <i>X. cannabis</i> pv. <i>cannabis</i> | NCPPB 3753 | JSZF01000001 | no | no | YES | YES | PIP | PIP | PIP | PIP |

^a“YES” indicates presence of a homolog, “no” indicates absence of the protein(s).

^b“Ψ” indicates pseudogenes, i.e., the gene is split into two or more fragments.

^c“PIP” indices the presence of the gene along with a canonical PIP box and a properly spaced –10 promoter motif. Numbers in square brackets indicate the number of single-nucleotide variants with respect to the canonical PIP box and –10 promoter motif.

^a *hpaR2* is eroded while an N-terminally truncated form of *hpaS* is present in *X. translucens* pv. *cerealis* strain CFBP 2541. Remnants are found in two other *X. translucens* strains (ART-Xtg27 and DSM 18974). Interestingly, the *hpaR2/hpaS* locus is intact in *X. translucens* strain DAR61454. This is an example of ongoing gene erosion in one species of *Xanthomonas*.

^b The *hpaR2/hpaS* locus got destroyed by an IS element in *X. oryzae* pv. *oryzae* strain KACC 10331. Yet, *hpaR2/hpaS* is present in *X. oryzae* pv. *oryzicola* strain BLS256. Remnants are found in the African *X. oryzae* pv. *oryzae* strain NA18 and in *X. oryzae* strains from the United States. This finding illustrates ongoing gene erosion in another species of *Xanthomonas*.

^c *hpaR2* is absent while an N-terminally truncated form of *hpaS* is present in *X. axonopodis* pv. *vasculorum* strain NCPPB 900.

PghAxc, and PehD, synonymous with PghBxc and PglA; Hsiao et al., 2008; Wang et al., 2008), one putative aminopeptidase and one putative lysophospholipase (LPL). Indeed, the two polygalacturonases have been previously shown to be regulated by HrpX in *X. campestris* (Wang et al., 2008). For *X. citri*, microarray transcriptome studies have shown previously that three genes with a canonical PIP box and -10 promoter motif (PehA, PehD, LPL; **Table 1**) are positively regulated by HrpX (Guo et al., 2011). Finally, our own RNAseq data show that the LPL is under control of HrpX in an African *X. oryzae* strain (unpublished data). We highly suspected that the four predicted genes with a canonical PIP box and properly spaced -10 promoter motif were indeed under control of HrpX in the cannabis strains.

Genome mining revealed that these four genes are present in most xanthomonads from Groups 1 and 2 (**Table 1**). Yet, in some lineages, one or the other gene apparently got lost. For instance, *X. sacchari*, *X. maliensis*, *X. arboricola*, *X. axonopodis* pv. *manihotis*, *X. oryzae*, and *X. vasicola* lack one or the other polygalacturonase, and the aminopeptidase is absent from *X. albilineans* and *X. oryzae*. In a few cases, these genes appear to have suffered from pseudogenization (**Table 1**) but this needs confirmation by targeted DNA sequencing due to the risk of sequence errors in some of the draft genome sequences.

Since these four genes appear to be under control of HrpX in the cannabis pathogens, we looked for evidence that the same regulation occurs in the other xanthomonads. We therefore compared the upstream regions of these four genes in a representative set of *Xanthomonas* strains. Strikingly, we found PIP boxes and properly spaced -10 promoter motifs for most of the genes in most of the *Xanthomonas* strains (**Figure 4**, **Supplemental Figure S2**), except for the Group 1 strains *X. albilineans*, *X. sacchari*, and *X. translucens*. Multiple sequence alignments of the promoter regions of the three genes show that the PIP boxes are conserved in sequence, context and position (PIP boxes ~ 140 – 150 bp before start codon of *pehA*, ~ 90 bp before start codon of the aminopeptidase gene, and ~ 210 – 250 bp before start codon of *pehD*) indicate that the PIP boxes of the polygalacturonase and aminopeptidase genes evolved early after separation of Group 2 from Group 1. In contrast, the PIP boxes of the lysophospholipase gene do not align with each other and reveal four subgroupings, which are compatible with a MLSA-based phylogeny of Group 2 strains (**Supplemental Figure S2**). This finding could suggest that the PIP box evolved several times independently at early times after separation of the four MLSA subgroups, or that the surrounding sequences evolved too extensively to allow the PIP boxes to be aligned using standard parameters.

To test if HrpX and HrpG could activate expression of the putative targets, we quantified gene expression of *pehA*, one of the four genes with a PIP box, in various mutant backgrounds of *X. cannabis* NCPPB 3753. HrpG* from *X. euvesicatoria* 85–10 is a HrpG variant that mimics the active form of HrpG (Wengelnik et al., 1999). Mutant strains with *hrpG** induce expression of *hrpX* and further genes targeted by HrpX, which includes primarily promoters with PIP boxes (Wengelnik et al., 1999). We therefore transformed *X. cannabis* cells with either pBBR1MCS-5

(empty vector), pBBR1MCS-5::*hrpX* or pBBR1MCS-5::*hrpG**, as previously described (Koebnik et al., 2006). *hrpG** from *X. euvesicatoria* was cloned into pBBR1MCS-5, as described for *hrpX* (Wengelnik et al., 1999; Koebnik et al., 2006). Positive transformants were confirmed by PCR (data not shown). Quantitative PCR (qPCR) was used to determine the relative expression of putative HrpX-target genes in *X. cannabis* NCPPB 3753 pBBR1MCS-5::*hrpG** and NCPPB 3653 pBBR1MCS-5::*hrpX* compared to NCPPB 3753 pBBR1MCS-5 (empty vector) as a control. *atpD* was used as a normalization control. *pehA*, *pehD*, and the gene encoding the putative LPL were dramatically induced in NCPPB 3753 ectopically expressing *hrpG** or *hrpX*, compared to the empty vector control (**Figure 5**). NCPPB 3753 pBBR1MCS-5::*hrpG** specifically upregulated (fold-change [SE]) *pehA* (54.8 ± 17.3), *pehD* (425.8 ± 16.6) and *LPL* (377.1 ± 36.5) compared to the empty vector control. A similar trend was observed with expression of the three genes (26.2 ± 4.1 , 49.4 ± 11.3 , 50.7 ± 11.9 , respectively) in NCPPB 3753 pBBR1MCS-5::*hrpX* compared to the empty vector control. We did not detect differential expression of the putative aminopeptidase in our *hrp* regulatory variant backgrounds under the conditions mentioned above (data not shown). We only tested one set of primers for this gene under one condition, and further experiments need to be performed to determine if this negative result suggests that this gene is not a target of HrpX. We overall conclude that the promoter of *pehA*, *pehD*, and *LPL* are *bona fide* targets of HrpX, validating our bioinformatic analysis. This supports the hypothesis that HrpG and HrpX regulate pathogenesis beyond T3SS alone (Tang et al., 2006). Further investigation of the role of HrpX and HrpG targets remains to be performed, but *X. cannabis* is a potentially interesting model to understand the role of HrpX beyond the T3SS.

One of the polygalacturonase genes, *pehA*, is not only under control of HrpX but was found to be regulated by Clp and RpfF in *X. campestris* (Hsiao et al., 2008). Since Clp and RpfF are conserved over all clades it is tempting to speculate that this layer of regulation has evolved before separation of the clades. However, comparison of the *pehA* promoter regions in a representative set of *Xanthomonas* strains revealed that the predicted Clp-binding site is not conserved (**Supplemental Figure S3**). Notably, multiple sequence alignment instead revealed a conserved sequence motif the position of which is shifted by 8 bp with respect to the predicted Clp-binding site (**Supplemental Figure S3**). In fact, this conserved box is as similar to the canonical Clp box as is the previously predicted Clp-binding site (Dong and Ebright, 1992), and its position is compatible with the mapping of the Clp-binding site by *lacZ* reporter fusions (Hsiao et al., 2008). Probably, the Clp-binding site was slightly mispredicted at that time due to the absence of sufficient sequence data from diverse *Xanthomonas* strains, which help to uncover regulatory DNA elements.

Conclusions

Stimulated by two new genome sequences from cannabis-pathogenic xanthomonads, we explored the world

Polygalacturonase PehA

| | | |
|--------------------|---|-------|
| XAV_NCPPB900 | ACTGGTTCGGTGCCACGGTGGCCGATTTCGCATCGCTGCAGCCAGATGCGCCGGCGCCTT-GTAGAGTC---- | 72 |
| XCC_ATCC33913 | GCTGCTTCGTTTGACGGCAGCCGGTTCCGCG-CTGCCATGCCCGGCTGACGCCCTGCTAGAGTC---- | 80 |
| XE_85-10 | GTGACTTCGCTCCACGGTGTCCGATTTCGCTCGCTGCTA-CAGGCGCCACCACGCTCCGCTAGAGTC---- | 82 |
| XCC_306 | GTGACTTCGCTCCACGGTGTCCGATTTCGCTCGCTGCTA-CAGGCGCCACCACGCTCCGCTAGAGTC---- | 83 |
| XC_CFBP4642 | GTGACTTCGCTCCACGGTGTCCGATTTCGCTCGCTGCTA-CAGGCGCCACCACGCTCCGCTAGAGTC---- | 80 |
| XVV_NCPPB206 | GCTGCTTCGTTCTGCGGCGTCCGTTTCGCTCGCTTGCA-CGCATCGGGCAGTGCTTCGCTAGAGTCGCCCT | 87 |
| XF_LMG25863 | GCC-CTTCGTTCGTACGGCGTCCGGTTTCGCTCGCCCTTTG-CGCATGGGTATCGATCGGTAGAGTC---- | 82 |
| XAC_NCPPB1832 | GCCCCTTCGCATCGGTATCGGGTTTCGCTCGCCCTTG-TGAATAGGCACGCTCGCTAGAGTC---- | 81 |
| XHC_M081 | GCCCCTTCGTTCCGGGGTTTCGCTTGCTTGCG-CATGTGGGTATTGCTCCGCTAGAGTC---- | 81 |
| XG_ATCC19865 | GCGGCTTCGTTCCGGGGGGTTTCGCTTGCTTGCG-CATGTACGGGGTTTCGCTTGCTAGAGTC---- | 81 |
| XAM_CIO1 | CTTGCTTCGTTCGTGGCGCCGGTTTCGCTTGCG-CATGTCACGGGAGTGTCTCCGCTAGAGTC---- | 92 |
| XCC_NCPPB2877/3753 | CTTGCTTCGTTCGTGGCGCCGGTTTCGCTTGCG-CATGCGGGCACAGCGCTTCGCTAGAGTC---- | 80 |
| XCP_Nyagatare | CTTGCTTCGCTCGTGGCGCCGGTTTCGCTTGCG-CATGCGGGCACAGCGCTTCGCTAGAGTC---- | 81 |
| XV_ATCC35937 | GCGGCTTCGCGCCGGCATCGGGTTTCGCCCACGTTCCG-CATGTGGCGCTGCTCCGCTAGAGTC---- | 81 |
| XFF_4834-R | GCGGCTTCGTTCCGGCGCCGGTTTCGCTTGCTTGCTAGAGTC---- | 80 |
| ***** | ***** | ***** |

PehA not present in: XOO_KACC19331

Polygalacturonase PehD

| | | |
|--------------------|---|-------|
| XOO_KACC10331 | --CCACTTGGCGGTGAAATGCTCAGGTCAAGTCAGGTCGCCCTGGACCACGTACAAACGTTGGTGTGCGATCCAGAGTTG--- | 139 |
| XCC_306 | --ACCTTCGGCGCTGCGGTGC----TGAATTCTGTCGCGACCAT--CGAACCTTCGCGAGGGCTCGCCAAAGCTTG---- | 139 |
| XE_85-10 | --ACGTTCCGGCGCTGCGGTGC----TGAATTCTGTCGCGCCAT--CGAACCTTCGCGAGGGCTCGCCAAAGCTTG---- | 139 |
| XFF_4834-R | --ACCTTCGGCGCTGCGGTGC----TGAATTCTGTCGCGCCAT--CGAACCTTCGCGAGGGCTCGCCAAAGCTTG---- | 139 |
| XF_LMG25863 | --CTCTCGCGCGATGCGGCGC----TGGATTCCGCATCCGGACAT--TGCCTCCGGCTAACGCAAATACCCCTTG---- | 154 |
| XG_ATCC19865 | --CGCTTCGGCGATGCGGCGC----TGGATTCCGCATCCGAACA--CGCTCTGGCTAACGCGCACCCCTTG---- | 166 |
| XHC_M081 | --GCCTTCGGCGATGCGGCGC----TGAATTCCGCATCCGACAT--CGCGCGCTGGCATGAAGGGCACGCTTGATC | 153 |
| XCC_ATCC33913 | --CACTTCGGCTGGCGGTGC----CGACTTCGCTGCCGTTTC--CGAGCGCAGACAAACGCGCAGATGGTGA---- | 201 |
| XV_ATCC35937 | --GACTTCGGAGATGCGGCGT----CGAGTTCCGCCACCGACGT--CGGCCGCCAGGATGAGACGCCAGGCTTG---- | 144 |
| XAV_NCPPB900 | --GACTTCGGGAGACGGCGGT----CGAATTCCGCCTCCGACGT--CGGCCGCCACATGAGGCGCACGCTTG---- | 146 |
| XC_CFBP4642 | TTTCCTTCGGCGCGCGCGC----TGGATTCCGCACCCGCCAT--CGACGAACGCGCTCAGGGCGCACGCTTG---- | 148 |
| XCC_NCPPB2877/3753 | --GGCTTCGGAAAGCGGTGC----TGAGTTCCGCCCGACGC--CGGGCAGCGGGCAAGGGCGCACGCTGG---- | 140 |
| XCP_Nyagatare | --GGCTTCGGAAAGCGGTGC----TGAGTTCCGCCCGACGC--CGGGCAGCGGGCAAGGGCGCACGCTGG---- | 140 |
| ***** | ***** | ***** |

PehD not present in: XAC_NCPPB1832, XAM_CIO1, XVV_NCPPB206

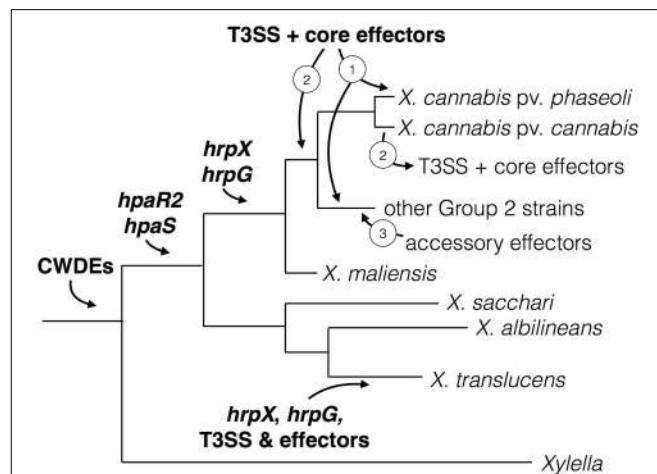
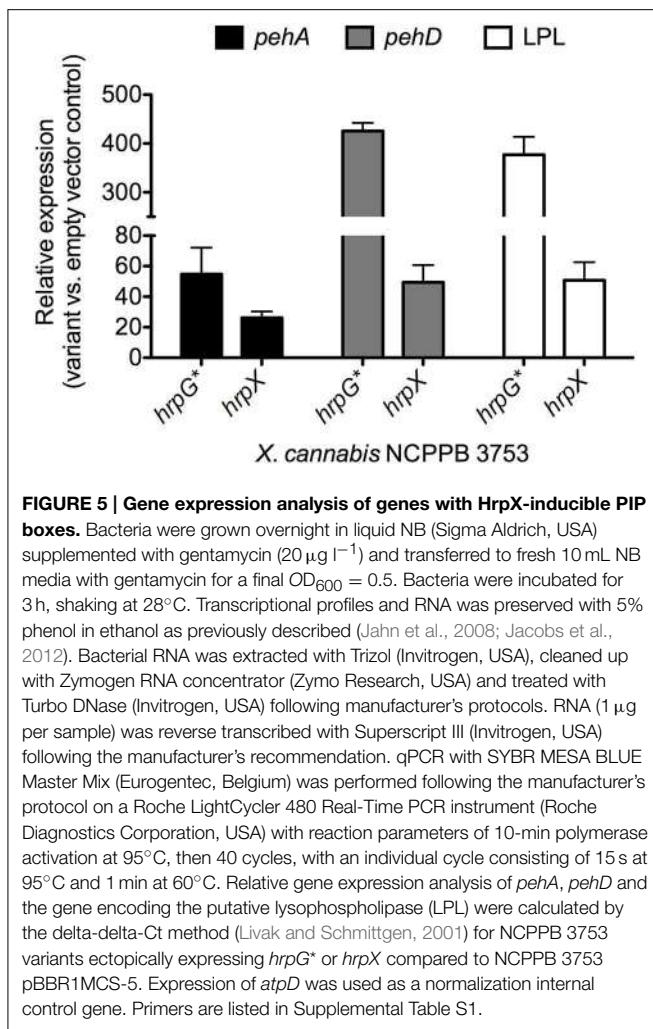
Aminopeptidase

| | | |
|--------------------|--|-------|
| XC_CFBP4642 | ----CTTCGCGC AAAAGGATGACCC TTGCAACGACGGGGACCTTCGCACTGATAAGCT CGGGGTTCCT 22 | |
| XCC_NCPPB2877/3753 | ----CTTCGCGC-CAAGGGATGACCTTTGCAACGGCAGCGGGGACCTTCGCTGATAAGCT CAGGGT-CCCC 20 | |
| XCP_Nyagatare | ----CTTCGCGC-GCAGGGATGACCTTTGCAACGGCAGCGGGGACCTTCGCTGATAAGCT CAGGGT-CCCC 20 | |
| XCC_ATCC33913 | --CCTTCGAC-AGAAGGATGACCTTTGCAACGGCAGACGGGGGACCTTCGCTGATAAGCT CGGAGTCCCC 21 | |
| XVV_NCPPB206 | --ACTTCGCGC-AGAAGGATGACCTTTGCAACGGCAGACGGGGGACTTCGCTGATAAGCT CGGGGT-CCTC 21 | |
| XFF_4834-R | --ACTTCGCGC-AGAAGGATGACCTTTGCAACGGCAGACGGGGGACCTTCGCTGATAAGCT CGGGGT-CCCC 21 | |
| XAM_CIO1 | --ACTTCGCGC-AGAAGGATGACCTTTGCAACGGCAGACGGGGGACCTTCGCTGATAAGCT CGGGGT-CCCC 21 | |
| XAV_NCPPB900 | --ACTTCGCGC-AGAAGGATGACCTTTGCAACGGCAGACGGGGGACCTTCGCTGATAAGCT CGGGGT-CCCC 21 | |
| XCC_306 | --TTCTCGCGC-AAGAACATGACCTTTGCAACGGCAGACGGGGGACCTTCGCTGATAAGCT CGGGGT-CCCC 21 | |
| XE_85-10 | --ACTTCGCGC-AGAAGGATGACCTTTGCAACGGCAGACGGGGGACCTTCGCTGATAAGCT CGGGGT-CCCC 21 | |
| XV_ATCC35937 | CGATCTTCGCGC-ACAGGGATGACCTTTGCAACGGCAGACGGGGGACCTTCGCTGATAAGCT CGGGGTCCCC 21 | |
| XHC_M081 | --ACTTCGCGC-AGGAGGATGACCTTTGCAACGGCAGACGGGGGACCTTCGCTGATAAGCT CGGGGTCCCC 21 | |
| XG_ATCC19865 | --ACTTCGCGC-AGGAGGATGACCTTTGCAACGGCAGACGGGGGACCTTCGCTGATAAGCT CGGGGTCCCC 21 | |
| XAC_NCPPB1832 | --ACTTCGCGC-AGGAGGATGACCTTTGCAACGGCAGACGGGGGACCTTCGCTGATAAGCT CGGGGT-CCCC 21 | |
| ***** | ***** | ***** |

Aminopeptidase not present in: XF_LMG25863, XOO_KACC10331

FIGURE 4 | Promoter sequences of predicted HrpX regulon members and their homologs in strains of *Xanthomonas*. Promoter regions encompassing 350 bp upstream of the translational start codon of a representative set of *Xanthomonas* strains were aligned by MUSCLE. PIP half boxes are shown in blue and the -10 promoter motif is shown in orange. Distance to the translational start codon is indicated on the right side of each sequence block. Deviations from the PIP consensus sequence are highlighted in yellow. The following *Xanthomonas* strains were analyzed: XAC (*X. arboricola* pv. *celebensis*) NCPPB 1832, XAM (*X. axonopodis* pv.

manihotis) CIO1, XAV (*X. axonopodis* pv. *vasculorum*) NCPPB 900, XCC (*X. campestris* pv. *campestris*) ATCC 33913, XCC (*X. cannabis* pv. *cannabis*) NCPPB 2877 and NCPPB 3753, XCP (*X. cannabis* pv. *phaseoli*) Nyagatare, XC (*X. cassavae*) CFBP 4642, XCC (*X. citri* pv. *citri*) 306, XE (*X. euvesicatoria*) 85-10, XF (*X. fragariae*) LMG 25863, XFF (*X. fuscans* subsp. *fuscans*) 4834-R, XG (*X. oryzae*) ATCC 19865, XHC (*X. horstorum* pv. *carotae*) M081, XOO (*X. oryzae* pv. *oryzae*) KACC 19331, XV (*X. vasicola* pv. *vasculorum*) NCPPB 206, and XV (*X. vesicatoria*) ATCC 35937. *Denotes conserved nucleotide.



of pathogenicity determinants in the genus *Xanthomonas*. A plethora of research data as well as our own analyses let us speculate about a stepwise evolution of pathogenicity in *Xanthomonas*. We developed a model (Figure 6) of evolution and acquisition of *Xanthomonas* pathogenicity factors based on a model proposed by Lu et al. (2008). Given the observation that some *Xanthomonas* strains across distinct taxa do not contain *hrpG* and *hrpX*, we speculate a stepwise evolution of pathogenicity (Figure 6), which involves (i) acquisition of key regulatory genes and cell wall-degrading enzymes, followed by (ii) acquisition of the Hrp type III secretion system, which is ultimately accompanied by (iii) successive acquisition of type III effectors. In parallel, as soon as *hrpG* and *hrpX* were acquired, a subset of genes, which can contribute to pathogenicity, evolved PIP boxes in their promoter regions thus ensuring their efficient expression during plant colonization.

Basic pathogenicity factors, such as the Xps type II secretion system and its secreted cell wall-degrading enzymes, as well as the *rpf* gene cluster, were probably already present in the ancestor of *Xanthomonas* and *Xylella*. Perhaps, these components were already expressed in response to environmental conditions, as we

can still observe for the genes that are controlled by RpfF and/or Clp. When the xanthomonads had separated into distinct genetic clades (Group 1 vs. Group 2), *hrpG* and *hrpX* were acquired, and perhaps evolved to be cross-regulated by the *HpaR2/HpaS* two-component regulatory system. It is conceivable that the *X. translucens* lineage acquired *hrpX* and *hrpG* together with the Hrp type III secretion system since the two regulatory genes are physically linked to the *hrp* gene cluster (Wichmann et al., 2013), reminiscent of the situation in *Ralstonia solanacearum*. Notably, the other early-branching xanthomonads, such as *X. albilineans* and *X. sacchari*, did not acquire *hrpX* nor *hrpG*. In contrast, all other xanthomonads have the oppositely transcribed *hrpG* and *hrpX* genes in the same synteny group between *radA* and *hsp70*, suggesting a unique acquisition event.

Interestingly, there are xanthomonads that possess *hrpG* and *hrpX* but not the Hrp Type III secretion system nor any type III effectors, such as the cannabis pathogen, the new *X. maliensis* clade (Triplett et al., 2015), and a recently sequenced strain of *X. arboricola* (Ignatov et al., 2015). This could be taken as evidence

that the Hrp system was either (1) acquired later or (2) got lost in some lineages. We favor the hypothesis that the T3SS and core effectors were gained in the case of the Nyagatare strain and not lost from *X. cannabis* NCPPB 2788 and NCPPB 3753. There seems to be no indication of any remnant features of the T3SS or effectors in the two *X. cannabis* genomes. Moreover, the Nyagatare strain possesses the T3SS and only core effectors, which appear to be located on a pathogenicity island flanked by insertion/transposable elements. More systematic genome sequencing of underexamined genetic lineages will shed light on this evolutionary puzzle.

Acknowledgments

This work benefited from a grant from the National Science Foundation to JJ (DBI - 1306196), a grant from the Fonds pour la formation à la Recherche dans l'Industrie et dans l'Agriculture (093604) to CP, and a grant from the Agence Nationale de la Recherche (ANR-2010-BLAN-1723) to RK. We are grateful to Lionel Moulin and Lucie Poulin, IRD Montpellier, for advice on ANI analyses and for technical assistance, respectively.

Supplementary Material

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2015.00431/abstract>

References

- Aritua, V., Musoni, A., Kato, F., Abang, M. M., Buruchara, R., Sapp, M., et al. (2015). The draft genome sequence of *Xanthomonas* species strain Nyagatare, isolated from diseased bean in Rwanda. *FEMS Microbiol. Lett.* 362. doi: 10.1093/femsle/fnu055
- Becker, A., Katzen, F., Pühler, A., and Ielpi, L. (1998). Xanthan gum biosynthesis and application: a biochemical/genetic perspective. *Appl. Microbiol. Biotechnol.* 50, 145–152.
- Bent, A. F., and Mackey, D. (2007). Elicitors, effectors, and *R* genes: the new paradigm and a lifetime supply of questions. *Annu. Rev. Phytopathol.* 45, 399–436. doi: 10.1146/annurev.phyto.45.062806.094427
- Borodovsky, M., and Lomsadze, A. (2014). Gene identification in prokaryotic genomes, phages, metagenomes, and EST sequences with GeneMarkS suite. *Curr. Protoc. Microbiol.* 32, 7. doi: 10.1002/9780471729259.mc01e07s32
- Büttner, D. (2012). Protein export according to schedule: architecture, assembly, and regulation of type III secretion systems from plant- and animal-pathogenic bacteria. *Microbiol. Mol. Biol. Rev.* 76, 262–310. doi: 10.1128/MMBR.05017-11
- Büttner, D., and Bonas, U. (2010). Regulation and secretion of *Xanthomonas* virulence factors. *FEMS Microbiol. Rev.* 34, 107–133. doi: 10.1111/j.1574-6976.2009.00192.x
- Canonne, J., Marino, D., Jauneau, A., Pouzet, C., Brière, C., Roby, D., et al. (2011). The *Xanthomonas* type III effector XopD targets the *Arabidopsis* transcription factor MYB30 to suppress plant defense. *Plant Cell* 23, 3498–3511. doi: 10.1105/tpc.111.088815
- Darrasse, A., Carrère, S., Barbe, V., Boureau, T., Arrieta-Ortiz, M. L., Bonneau, S., et al. (2013). Genome sequence of *Xanthomonas fuscans* subsp. *fuscans* strain 4834-R reveals that flagellar motility is not a general feature of xanthomonads. *BMC Genomics* 14:761. doi: 10.1186/1471-2164-14-761
- Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., et al. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* 36, W465–W469. doi: 10.1093/nar/gkn180
- Dong, Q., and Ebright, R. H. (1992). DNA binding specificity and sequence of *Xanthomonas campestris* catabolite gene activator protein-like protein. *J. Bacteriol.* 174, 5457–5461.
- Dow, M. (2008). Diversification of the function of cell-to-cell signaling in regulation of virulence within plant pathogenic xanthomonads. *Sci. Signal.* 1:pe23. doi: 10.1126/stke.121pe23
- Felix, G., Duran, J. D., Volko, S., and Boller, T. (1999). Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. *Plant J.* 18, 265–276. doi: 10.1046/j.1365-313X.1999.00265.x
- Figueras, M. J., Beaz-Hidalgo, R., Hossain, M. J., and Liles, M. R. (2014). Taxonomic affiliation of new genomes should be verified using average nucleotide identity and multilocus phylogenetic analysis. *Genome Announc.* 2, e00927-e00914. doi: 10.1128/genomeA.00927-14
- Furutani, A., Nakayama, T., Ochiai, H., Kaku, H., Kubo, Y., and Tsuge, S. (2006). Identification of novel HrpXo regulons preceded by two cis-acting elements, a plant-inducible promoter box and a –10 box-like sequence, from the genome database of *Xanthomonas oryzae* pv. *oryzae*. *FEMS Microbiol. Lett.* 259, 133–141. doi: 10.1111/j.1574-6968.2006.00265.x
- Guo, Y., Figueiredo, F., Jones, J., and Wang, N. (2011). HrpG and HrpX play global roles in coordinating different virulence traits of *Xanthomonas axonopodis* pv. *citri*. *Mol. Plant Microbe Interact.* 24, 649–661. doi: 10.1094/MPMI-09-10-0209
- Hauben, L., Vauterin, L., Swings, J., and Moore, E. R. (1997). Comparison of 16S ribosomal DNA sequences of all *Xanthomonas* species. *Int. J. Syst. Bacteriol.* 47, 328–335. doi: 10.1099/00207713-47-2-328
- Hernandez, D., Tewhey, R., Veyrieras, J. B., Farinelli, L., Østerås, M., François, P., et al. (2014). *In silico* finished 2.8 Mbp *Staphylococcus aureus* genome assembly from 100 bp short and long range paired-end reads. *Bioinformatics* 30, 40–49. doi: 10.1093/bioinformatics/btt590
- Hsiao, Y. M., Zheng, M. H., Hu, R. M., Yang, T. C., and Tseng, Y. H. (2008). Regulation of the *pehA* gene encoding the major polygalacturonase

- of *Xanthomonas campestris* by Clp and RpfF. *Microbiology* 154, 705–713. doi: 10.1099/mic.0.2007/012930-0
- Ignatov, A. N., Kyrova, E. I., Vinogradova, S. V., Kamionskaya, A. M., Schaad, N. W., and Luster, D. G. (2015). Draft genome sequence of *Xanthomonas arboricola* strain 3004, a causal agent of bacterial disease on barley. *Genome Announc.* 3, e01572–e01514. doi: 10.1128/genomeA.01572-14
- Jacobs, J. M., Babujee, L., Meng, F., Milling, A., and Allen, C. (2012). The *in planta* transcriptome of *Ralstonia solanacearum*: conserved physiological and virulence strategies during bacterial wilt of tomato. *MBio* 3, e00114–e00112. doi: 10.1128/mBio.00114-12
- Jahn, C., Charkowski, A., and Willis, D. K. (2008). Evaluation of isolation methods and RNA integrity for bacterial RNA quantitation. *J. Microbiol. Methods* 75, 318–324. doi: 10.1016/j.mimet.2008.07.004
- Kaewnum, S., Prathuangwong, S., and Burr, T. J. (2006). A pectate lyase homolog, *xagP*, in *Xanthomonas axonopodis* pv. *glycines* is associated with hypersensitive response induction on tobacco. *Phytopathology* 96, 1230–1236. doi: 10.1094/PHYTO-96-1230
- Koebnik, R., Krüger, A., Thieme, F., Urban, A., and Bonas, U. (2006). Specific binding of the *Xanthomonas campestris* pv. *vesicatoria* AraC-type transcriptional activator HrpX to plant-inducible promoter boxes. *J. Bacteriol.* 188, 7652–7660. doi: 10.1128/JB.00795-06
- Konstantinidis, K. T., and Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2567–2572. doi: 10.1073/pnas.0409727102
- Li, R. F., Lu, G. T., Li, L., Su, H. Z., Feng, G. F., Chen, Y., et al. (2014). Identification of a putative cognate sensor kinase for the two-component response regulator HrpG, a key regulator controlling the expression of the *hrp* genes in *Xanthomonas campestris* pv. *campestris*. *Environ. Microbiol.* 16, 2053–2071. doi: 10.1111/1462-2920.12207
- Li, S., Wang, Y., Wang, S., Fang, A., Wang, J., Liu, L., et al. (2015). The type III effector AvrBs2 in *Xanthomonas oryzae* pv. *oryzicola* suppresses rice immunity and promotes disease development. *Mol. Plant Microbe Interact.* doi: 10.1094/MPMI-10-14-0314-R. [Epub ahead of print].
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262
- Lu, H., Patil, P., Van Sluys, M. A., White, F. F., Ryan, R. P., Dow, J. M., et al. (2008). Acquisition and evolution of plant pathogenesis-associated gene clusters and candidate determinants of tissue-specificity in *Xanthomonas*. *PLoS ONE* 3:e3828. doi: 10.1371/journal.pone.0003828
- McPartland, J. M., Clarke, R. C., and Watson, D. P. (2000). *Hemp Diseases and Pests: Management and Biological Control: An Advanced Treatise*. Oxfordshire: CABI Publishing.
- Mhedbi-Hajri, N., Jacques, M. A., and Koebnik, R. (2011). Adhesion mechanisms of plant-pathogenic *Xanthomonadaceae*. *Adv. Exp. Med. Biol.* 715, 71–89. doi: 10.1007/978-94-007-0940-9_5
- Mole, B. M., Baltrus, D. A., Dangl, J. L., and Grant, S. R. (2007). Global virulence regulation networks in phytopathogenic bacteria. *Trends Microbiol.* 15, 363–371. doi: 10.1016/j.tim.2007.06.005
- Netsu, O., Kijima, T., and Takikawa, Y. (2014). Bacterial leaf spot of hemp caused by *Xanthomonas campestris* pv. *cannabis* in Japan. *J. Gen. Plant Pathol.* 80, 164–168. doi: 10.1007/s10327-013-0497-8
- Parkinson, N., Aritua, V., Heeney, J., Cowie, C., Bew, J., and Stead, D. (2007). Phylogenetic analysis of *Xanthomonas* species by comparison of partial gyrase B gene sequences. *Int. J. Syst. Evol. Microbiol.* 57, 2881–2887. doi: 10.1099/ijts.0.65220-0
- Parkinson, N., Cowie, C., Heeney, J., and Stead, D. (2009). Phylogenetic structure of *Xanthomonas* determined by comparison of *gyrB* sequences. *Int. J. Syst. Evol. Microbiol.* 59, 264–274. doi: 10.1099/ijts.0.65825-0
- Patil, P. B., Bogdanove, A. J., and Sonti, R. V. (2007). The role of horizontal transfer in the evolution of a highly variable lipopolysaccharide biosynthesis locus in xanthomonads that infect rice, citrus and crucifers. *BMC Evol. Biol.* 7:243. doi: 10.1186/1471-2148-7-243
- Pieretti, I., Royer, M., Barbe, V., Carrere, S., Koebnik, R., Cocianich, S., et al. (2009). The complete genome sequence of *Xanthomonas albilineans* provides new insights into the reductive genome evolution of the xylem-limited *Xanthomonadaceae*. *BMC Genomics* 10:616. doi: 10.1186/1471-2164-10-616
- Richter, M., and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19126–19131. doi: 10.1073/pnas.0906412106
- Rosse, Y., Wolfson, E. B., Holmes, A., Gally, D. L., and Holden, N. J. (2015). Bacterial flagella: twist and stick, or dodge across the kingdoms. *PLoS Pathog.* 11:e1004483. doi: 10.1371/journal.ppat.1004483
- Schulze, S., Kay, S., Büttner, D., Egler, M., Eschen-Lippold, L., Hause, G., et al. (2012). Analysis of new type III effectors from *Xanthomonas* uncovers XopB and XopS as suppressors of plant immunity. *New Phytol.* 195, 894–911. doi: 10.1111/j.1469-8137.2012.04210.x
- Severin, V. (1978). Ein neues Bakterium an Hanf – *Xanthomonas campestris* pathovar *cannabis*. *Arch. Phytopathol. Pflanzenschutz* 14, 7–15.
- Shi, Q., Febres, V. J., Jones, J. B., and Moore, G. A. (2015). Responsiveness of different citrus genotypes to the *Xanthomonas citri* ssp. *citri*-derived pathogen-associated molecular pattern (PAMP) flg22 correlates with resistance to citrus canker. *Mol. Plant Pathol.* 16, 507–520. doi: 10.1111/mpp.12206
- Sinha, D., Gupta, M. K., Patel, H. K., Ranjan, A., and Sonti, R. V. (2013). Cell wall degrading enzyme induced rice innate immune responses are suppressed by the type 3 secretion system effectors XopN, XopQ, XopX and XopZ of *Xanthomonas oryzae* pv. *oryzae*. *PLoS ONE* 8:e75867. doi: 10.1371/journal.pone.0075867
- Stork, W., Kim, J. G., and Mudgett, M. B. (2015). Functional analysis of plant defense suppression and activation by the *Xanthomonas* core type III effector XopX. *Mol. Plant Microbe Interact.* 28, 180–194. doi: 10.1094/MPMI-09-14-0263-R
- Studholme, D. J., Wasukira, A., Paszkiewicz, K., Aritua, V., Thwaites, R., Smith, J., et al. (2011). Draft genome sequences of *Xanthomonas sacchari* and two banana-associated xanthomonads reveal insights into the *Xanthomonas* group 1 clade. *Genes* 2, 1050–1065. doi: 10.3390/genes2041050
- Sun, W., Dunning, F. M., Pfund, C., Weingarten, R., and Bent, A. F. (2006). Within-species flagellin polymorphism in *Xanthomonas campestris* pv. *campestris* and its impact on elicitation of *Arabidopsis FLAGELLIN SENSING2*-dependent defenses. *Plant Cell* 18, 764–779. doi: 10.1105/tpc.105.037648
- Tang, X., Xiao, Y., and Zhou, J. M. (2006). Regulation of the type III secretion system in phytopathogenic bacteria. *Mol. Plant Microbe Interact.* 19, 1159–1166. doi: 10.1094/MPMI-19-1159
- Triplett, L. R., Verdier, V., Campillo, T., Van Malderghem, C., Cleenwerck, I., Maes, M., et al. (2015). Characterization of a novel clade of *Xanthomonas* isolated from rice leaves in Mali and proposal of *Xanthomonas maliensis* sp. nov. *Antonie Van Leeuwenhoek* 107, 869–881. doi: 10.1007/s10482-015-0379-5
- Tsuchiya, K., Mew, T. W., and Wakimoto, S. (1982). Bacteriological and pathological characteristics of wild types and induced mutants of *Xanthomonas campestris* pv. *oryzae*. *Phytopathology* 72, 43–46.
- van Bakel, H., Stout, J. M., Cote, A. G., Tallon, C. M., Sharpe, A. G., Hughes, T. R., et al. (2011). The draft genome and transcriptome of *Cannabis sativa*. *Genome Biol.* 12:R102. doi: 10.1186/gb-2011-12-10-r102
- Wang, L., Rong, W., and He, C. (2008). Two *Xanthomonas* extracellular polygalacturonases, PghAxc and PghBxc, are regulated by type III secretion regulators HrpX and HrpG and are required for virulence. *Mol. Plant Microbe Interact.* 21, 555–563. doi: 10.1094/MPMI-21-5-0555
- Wengelnik, K., Rossier, O., and Bonas, U. (1999). Mutations in the regulatory gene *hrpG* of *Xanthomonas campestris* pv. *vesicatoria* result in constitutive expression of all *hrp* genes. *J. Bacteriol.* 181, 6828–6831.
- Wengelnik, K., Van den Ackerveken, G., and Bonas, U. (1996). HrpG, a key hrp regulatory protein of *Xanthomonas campestris* pv. *vesicatoria* is homologous to two-component response regulators. *Mol. Plant Microbe Interact.* 9, 704–712. doi: 10.1094/MPMI-9-0704
- White, F. F., Potnis, N., Jones, J. B., and Koebnik, R. (2009). The type III effectors of *Xanthomonas*. *Mol. Plant Pathol.* 10, 749–766. doi: 10.1111/j.1364-3703.2009.00590.x
- Wichmann, F., Vorhölter, F., Herseman, L., Widmer, F., Blom, J., Niehaus, K., et al. (2013). The noncanonical type III secretion system of *Xanthomonas translucens* pv. *graminis* is essential for forage grass infection. *Mol. Plant Pathol.* 14, 576–588. doi: 10.1111/mpp.12030

- Young, J. M., Park, D. C., Shearman, H. M., and Fargier, E. (2008). A multilocus sequence analysis of the genus *Xanthomonas*. *Syst. Appl. Microbiol.* 31, 366–377. doi: 10.1016/j.syapm.2008.06.004
- Zerbino, D. R., and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18, 821–829. doi: 10.1101/gr.074492.107
- Zhao, S., Poulin, L., Rodriguez-R, L. M., Forero-Serna, N., Liu, S. H., Wonni, I., et al. (2012). Development of a variable number of tandem repeats typing scheme for the bacterial rice pathogen *Xanthomonas oryzae* pv. *oryzicola*. *Phytopathology* 102, 948–956. doi: 10.1094/PHYTO-04-12-0078-R

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 Jacobs, Pesce, Lefevre and Koebnik. 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.



Interactions of *Xanthomonas* type-III effector proteins with the plant ubiquitin and ubiquitin-like pathways

Suayib Üstün^{1*} and Frederik Börnke^{1,2}

¹ Plant Metabolism Group, Leibniz-Institute of Vegetable and Ornamental Crops, Großbeeren, Germany

² Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany

Edited by:

Nicolas Denancé, Institut National de la Recherche Agronomique, France

Reviewed by:

Mark James Banfield, John Innes Centre, UK

Nemo Peeters, Institut National de la Recherche Agronomique, France

***Correspondence:**

Suayib Üstün, Plant Metabolism Group, Leibniz-Institute of Vegetable and Ornamental Crops, Theodor-Echtermeyer-Weg 1, 14979 Großbeeren, Germany
e-mail: uestuen@igzev.de

In eukaryotes, regulated protein turnover is required during many cellular processes, including defense against pathogens. Ubiquitination and degradation of ubiquitinated proteins via the ubiquitin–proteasome system (UPS) is the main pathway for the turnover of intracellular proteins in eukaryotes. The extensive utilization of the UPS in host cells makes it an ideal pivot for the manipulation of cellular processes by pathogens. Like many other Gram-negative bacteria, *Xanthomonas* species secrete a suite of type-III effector proteins (T3Es) into their host cells to promote virulence. Some of these T3Es exploit the plant UPS to interfere with immunity. This review summarizes T3E examples from the genus *Xanthomonas* with a proven or suggested interaction with the host UPS or UPS-like systems and also discusses the apparent paradox that arises from the presence of T3Es that inhibit the UPS in general while others rely on its activity for their function.

Keywords: *Xanthomonas*, type-III effector, ubiquitin, proteasome, plant defense

INTRODUCTION

The bacterial genus *Xanthomonas* consists of a large group of Gram-negative plant pathogenic bacteria comprising 27 species that infect a wide range of economically important crop plants, such as rice, citrus, banana, cabbage, tomato, pepper, and bean (Ryan et al., 2011). The infection strategies of various *Xanthomonas* species and pathovars are adapted to their different hosts and also exhibit tissue specificity (Ryan et al., 2011). For example, *Xanthomonas campestris* pv. *campestris* and *X. campestris* pv. *musacearum* invade through the vascular system and spread systematically whereas *X. campestris* pv. *vesicatoria* and *X. citri* pv. *citri* colonize the intercellular space (Buttner and Bonas, 2010). The broad host range of the *Xanthomonas* species and the adaptation to different tissues is also reflected in the dynamic nature of the type III effector (T3E) repertoires in a given pathovar or species. To date, ~40 T3Es of the genus *Xanthomonas* have been identified, which are divided into groups based on their sequence identities (White et al., 2009). These T3Es function as virulence and avirulence factors either by suppressing PAMP-triggered immunity (PTI) or through the recognition by host immune receptors (Resistance proteins) and subsequent elicitation of the so called effector-triggered immunity (ETI; Jones and Dangl, 2006). Although, T3Es are assumed to contribute to virulence of *Xanthomonas*, host cellular targets and biochemical activities for many effectors remain unknown.

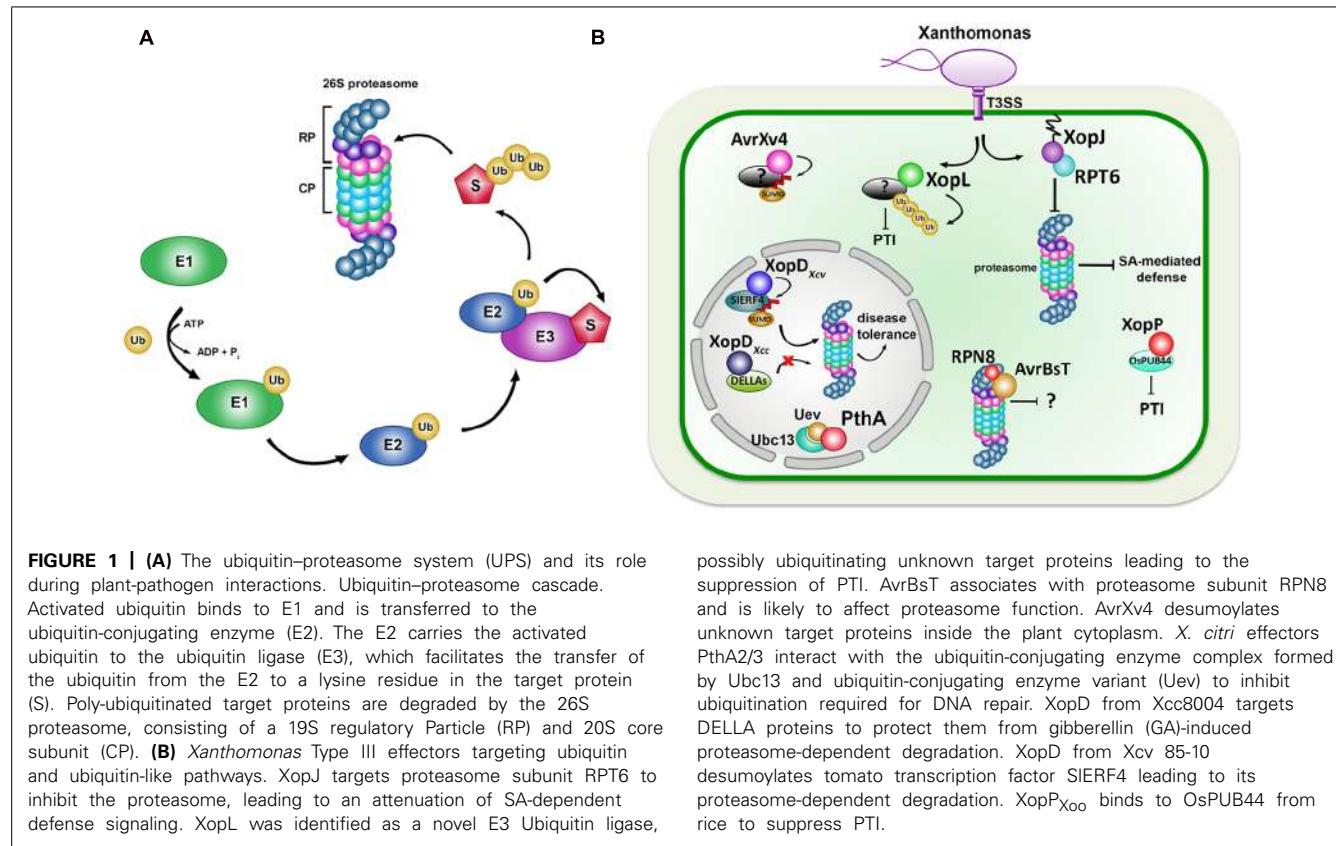
The ubiquitin–proteasome system (UPS) is involved in a broad array of cellular processes, such as signaling, cell cycle, vesicle trafficking, and immunity (Vierstra, 2009). Selective protein degradation by the UPS proceeds from the ligation of one or more ubiquitin proteins to the ε-amino group of a lysine residue within specific target proteins catalyzed by E1, E2, and E3 enzymes (Figure 1A). The ubiquitylated target protein is then recognized by the 26S proteasome for degradation. The 26S proteasome itself

is a 2.5 MDa ATP-dependent protease complex composed of a 20S core protease (CP) and two 19S regulatory particles (RPs), each of which contains a lid and a base subunit (Figure 1A).

Beyond its role in marking target proteins for degradation via the 26S proteasome, ubiquitination can regulate cellular signaling processes. Mono-ubiquitination or multi-ubiquitination is associated with endocytosis, protein sorting, gene expression, and various other cellular pathways (Mukhopadhyay and Riezman, 2007). In addition to ubiquitination, ubiquitin-like modifications, such as SUMO (small ubiquitin-related modifier), play an essential role in various cellular functions. Similar to the ubiquitination pathway, sumoylation requires an E1-E2-E3 enzyme cascade to conjugate SUMO to the target protein. Sumoylation can affect localization, protein-protein interaction, and stability of the modified protein (Vierstra, 2012).

During the past few years, evidence has emerged that ubiquitin- and ubiquitin-like pathways play a major role in immunity and hence are subverted by bacterial pathogens in animal and plant hosts (Boyer and Lemichez, 2004; Perrett et al., 2011; Marino et al., 2012). Several components of the UPS were identified as regulators of plant immunity during PTI and ETI, such as pepper E3 ligase *CaRING1* that is induced upon *Xanthomonas* infection and is required for the activation of cell death (Lee et al., 2011). Moreover, recent studies identified that members of the U-box E3 ligase family are negative regulators of PTI (Trujillo et al., 2008; Stegmann et al., 2012). A direct connection between the UPS and ETI was shown by the fact that the accumulation of certain resistance proteins is controlled by the ubiquitin-mediated degradation via the 26S proteasome (Furlan et al., 2012).

Considering the involvement of the UPS in plant defense mechanisms, co-evolution has selected for T3Es and toxins that can manipulate ubiquitin and ubiquitin-like pathways in order to



interfere with induced defense responses. The best characterized effector proteins or toxins with respect to exploitation of the UPS can be found in *Pseudomonas syringae* pv. *tomato*, a bacterium that causes bacterial speck disease on tomato plants. Some of these effectors mimic E3 ligases, e.g., AvrPtoB, to suppress both PTI and ETI events (Abramovitch et al., 2006; Janjusevic et al., 2006), whereas others, such as HopM1 promote ubiquitination of its target protein to inhibit certain induced defense responses (Nomura et al., 2006). A more direct way to subvert the UPS is achieved by SylA, a secreted toxin from *P. syringae* pv. *syringae*, which directly targets the catalytic subunits of the 26S proteasome to inhibit its activity and to suppress plant immune reactions (Groll et al., 2008; Schellenberg et al., 2010; Misas-Villamil et al., 2013).

In recent years, it has become evident that the UPS has a major role during the interaction of *Xanthomonas* with its plant hosts. Therefore, this mini review summarizes the current knowledge about T3Es of different *Xanthomonas* species with a demonstrated effect on ubiquitin and ubiquitin-like pathways. Possible virulence functions and conflicting actions of T3E proteins promoting or inhibiting the ubiquitin pathway are discussed.

T3Es FROM *Xanthomonas* SPECIES INTERACTING WITH THE HOST UPS

The dual roles of UPS components in defense and development render them to be vulnerable targets for exploitation during infection. Several T3Es from *Xanthomonas* species have been

possibly ubiquitinating unknown target proteins leading to the suppression of PTI. AvrBsT associates with proteasome subunit RPN8 and is likely to affect proteasome function. AvrXv4 desumoylates unknown target proteins inside the plant cytoplasm. *X. citri* effectors PthA2/3 interact with the ubiquitin-conjugating enzyme complex formed by Ubc13 and ubiquitin-conjugating enzyme variant (Uev) to inhibit ubiquitination required for DNA repair. XopD from Xcc8004 targets DELLA proteins to protect them from gibberellin (GA)-induced proteasome-dependent degradation. XopD from Xcv 85-10 desumoylates tomato transcription factor SIERF4 leading to its proteasome-dependent degradation. XopP_{Xoo} binds to OsPUB44 from rice to suppress PTI.

shown or suggested to interact with components of ubiquitin and ubiquitin-like pathways of the host plant in a positive or negative manner (summarized in Table 1; illustrated in Figure 1B).

EFFECTORS INTERACTING WITH UPS COMPONENTS

XopJ is a type III effector of *X. campestris* pv. *vesicatoria* (strain 85-10), although a highly similar sequence is also found in the genome of *X. campestris* pv. *malvacearum*. Apart from that, close homologs are also present in *Pseudomonas* spp., including *Pseudomonas avellanae*, *P. syringae* pv. *actinidiae*, *P. syringae* pv. *lachrymans* and appear to function at least in part in a XopJ-like manner (Üstün et al., 2014). XopJ belongs to the widely distributed YopJ-effector family of cysteine proteases/acetyltransferases (Hotson and Mudgett, 2004; Lewis et al., 2011). Members of this diverse T3E family are present among both plant and animal pathogenic bacteria. Based on structural similarities to cysteine proteases from adenovirus, the archetypal member of this effector family, YopJ from *Yersinia pestis*, was originally assigned to the CE clan of C55 peptidases (Orth et al., 2000). Proteases in this clan share a catalytic triad as a characteristic feature, consisting of the amino acids histidine, glutamic/aspartic acid, and a cysteine. Although recent studies demonstrated that YopJ and other members of this effector protein act as acetyltransferases on their target proteins (Mukherjee et al., 2006; Tasset et al., 2010; Lee et al., 2012; Jiang et al., 2013; Cheong et al., 2014), it has also been shown for YopJ and other members (summarized below) that these T3Es display de-sumoylating and de-ubiquitinating activities, implying that the YopJ effector

Table 1 | *Xanthomonas* effectors interacting with the host UPS.

| Effector | Activity/domain | Target | Role during infection | Reference |
|--------------------------|----------------------------------|--------------------|---|-------------------------|
| AvrXv4 | deSUMOylation | ? | ? | Roden et al. (2004) |
| AvrBsT | Acetyltransferase/SUMO-protease | ACIP1, SnRK1, RPN8 | Suppression of ETI/PTI | Szczesny et al. (2010) |
| XopJ | Cysteine protease? | RPT6 | Inhibition of SA signalling | Üstün et al. (2013) |
| XopD _{Xcv85-10} | deSUMOylation | SLERF4 | Suppression of ethylene responses | Kim et al. (2013) |
| XopD _{Xcc8004} | deSUMOylation? deubiquitination? | DELLAs | Disease tolerance; repression of ROS | Tan et al. (2014) |
| XopL | E3 ubiquitin ligase | ? | Suppression of PTI | Singer et al. (2013) |
| PthA2/3 | TAL | Ubc13/Uev | Interference with DNA repair mechanisms | Domingues et al. (2010) |
| XopI | F-box domain | ? | ? | Schulze et al. (2012) |
| XopP | Unknown | OsPUB44 | Suppression of PTI | Ishikawa et al. (2014) |

family plays a role in manipulation the UPS. Initially, XopJ was identified as an T3E, as its expression is induced dependent on *hrpG* that controls the expression of *hrp* genes, being essential for the pathogenicity of *Xcv* (Noel et al., 2003). Further *in silico* analysis of the amino-terminal part of XopJ revealed a possible myristylation side, being responsible for the plasma membrane localization of XopJ after translocation into the host cytoplasm (Thieme et al., 2007; Bartetzko et al., 2009). Subcellular localization of XopJ is also associated with its function to block the secretory pathway dependent on its catalytic triad and thereby interfering with cell-wall based defense responses (Bartetzko et al., 2009). Further functional analysis revealed that XopJ interacts with the 19S RP subunit RPT6 (RP ATPase 6) of the 26S proteasome. XopJ is able to recruit cytoplasmic RPT6 to the plant plasma membrane leading to the inhibition of the proteasome activity. This effect is dependent on both, its myristylation and its catalytic triad (Üstün et al., 2013). *Xcv* infection of susceptible pepper plants revealed that XopJ is acting as a tolerance factor, attenuating the accumulation of salicylic acid (SA) to delay host tissue necrosis in a proteasome-dependent manner (Üstün et al., 2013). XopJ-mediated inhibition of the proteasome function also interferes with other events during plant immunity, as vesicle trafficking and callose deposition are also affected by the suppression of the proteasome. This also explains the initial observation that XopJ blocks vesicle trafficking during immunity. It is presently not clear how the inhibitory effect of XopJ on the proteasome is related to the suppression of SA-mediated defense responses. Similar to what has been proposed for SylA (Schellenberg et al., 2010), XopJ might be affecting the proteasomal turnover of NPR1, the master regulator of SA signaling, to interfere with SA-dependent immunity. Future studies regarding the protein turnover of putative target proteins of XopJ will shed light on this open question and also reveal other mechanisms implicated in XopJ-triggered immunity suppression.

Given the fact that XopJ so far has only been found in *Xcv* 85-10 and in *X. campestris* pv. *malvacearum*, it is possible that only certain pathovars acquired this effector during evolution to directly target the host cell proteasome as a way of adaptation to different hosts. Alternatively, other *Xanthomonas* pathovars might utilize different effector proteins involving other mechanisms to target components of the UPS. This might be the case for AvrBsT from

Xcv 75-3 that was identified to interact with a UPS component. Szczesny et al. (2010) identified 19S RP subunit RPN8 as a potential interaction partner of AvrBsT in a yeast-2-hybrid assay. Similar to XopJ, AvrBsT is a member of the YopJ-superfamily of cysteine proteases/acetyltransferases, sharing 35% amino acid identity to XopJ. In addition to RPN8, AvrBsT is targeting SNF1-related kinase 1 (SnRK1), an essential regulator of nutrient and stress signaling, to possibly mediate suppression of AvrBs1-triggered hypersensitive response (Szczesny et al., 2010). Intriguingly, SnRK1 is associated with the alpha4/PAD1 subunit of the 20S proteasome to mediate proteasomal binding of a plant SCF ubiquitin ligase (Farras et al., 2001). Taken together, it is possible that AvrBsT is disrupting proteasome-mediated protein turnover similar to XopJ. However, additional experiments are required to assess the role of YopJ-like effector AvrBsT in the manipulation of the UPS machinery. Recently it was demonstrated that AvrBsT displays acetyltransferase activity toward a protein associated with microtubules and immunity (Cheong et al., 2014). Whether SnRK1 or Rpn8 are targets for AvrBsT-mediated acetylation remain to be investigated.

Another example for the exploitation of the UPS by *Xanthomonas*, is the interaction of *X. axonopodis* pv. *citri* type III effectors PthA 2 and 3 with the ubiquitin-conjugation enzyme complex formed by Ubc13 and ubiquitin-conjugation enzyme variant (Uev; Domingues et al., 2010). PthA proteins belong to the AvrBs3/PthA or TAL (transcription activator-like) family that were recently identified to act as transcriptional activators in the plant cell nucleus, where they directly bind to DNA via a central domain of tandem repeats (Boch et al., 2009). Despite the fact that effectors from the TAL family have evolved to target the plant nuclear DNA and modulate host transcription, it could be possible that proteins from this large effector family might associate with other host proteins to regulate host transcription. Both PthA 2 and 3 interact with the heterodimer complex of Ubc 13-Uev, required for ubiquitination of target proteins involved in DNA repair (Domingues et al., 2010). Taken together, this is another example of a *Xanthomonas* T3E possibly hijacking the UPS to modulate host cellular pathways.

Recently, the T3E XopP from *X. oryzae* pv. *oryzae* was shown to target OsPUB44, a rice ubiquitin E3 ligase with a

unique U-box domain, to suppress peptidoglycan (PGN)- and chitin-triggered immunity and resistance to *X. oryzae* (Ishikawa et al., 2014). Although the enzymatic activity of XopP remains unknown, the authors were able to show that XopP inhibits the ubiquitin E3 ligase activity of OsPUB44, leading to its accumulation *in planta* possibly due to a loss of its auto-ubiquitination (Ishikawa et al., 2014). Whether XopP inhibits the E3 ligase activity of OsPUB44 by its biochemical activity or simply by competing for the binding site with an E2 enzyme remains to be shown.

EFFECTORS ENCODING SUMO-PROTEASES

The initial discovery that YopJ-like effectors also share limited structural similarities with the yeast ubiquitin-like Protease 1 [ULP1, also known as small ubiquitin-like modifier (SUMO) protease], led to the assumption that these effectors may act as SUMO proteases. SUMO proteases desumoylate sumo-conjugated target proteins and as sumoylation appears to be connected to pathogen attack and other stress responses, this process might be an attractive target for bacterial invaders to modulate protein functions (Hotson and Mudgett, 2004). The first evidence that *Xanthomonas* effectors mimic SUMO proteases was provided by the functional characterization of XopD (Hotson et al., 2003). In contrast to YopJ-like effectors, XopD shares high similarities with ULPs and hence is classified as a cysteine protease belonging to the C48 family of the CE clan. XopD is localized to subnuclear foci and cleaves plant-specific SUMO precursors interfering with protein sumoylation *in planta* (Hotson et al., 2003). In the nucleus, XopD is able to bind DNA and to repress the transcription of senescence- and defense-related genes leading to the attenuation of SA-dependent senescence in tomato (Kim et al., 2008). Further analysis revealed that XopD targets tomato transcription factor SIERF4 for de-sumoylation to prevent ethylene-mediated defense responses in order to enhance bacterial propagation (Kim et al., 2013). XopD interacts with SIERF4 in the nucleus and catalyzes SUMO1 hydrolysis from lysine 53. This in turn leads to the proteasome dependent destabilization of SIERF4 (Kim et al., 2013). In summary, XopD is an example of a T3E utilizing an ubiquitin-like pathway by acting as a SUMO protease to destabilize its target protein and thereby enhancing the virulence of Xcv during infection of tomato plants.

XopD is also a paradigm for strain specific functions of homolog T3Es, as XopD from *X. campestris* pv. *campestris* (8004) uses a different strategy to modulate plant immunity: XopD_{Xcc8004} targets DELLA protein RGA (repressor of ga1-3) in the nucleus to delay its gibberellin (GA)-mediated degradation via the 26S proteasome (Tan et al., 2014). As a consequence, disease symptom development is suppressed to initiate disease tolerance and promote bacterial survival. Although the authors were not able to show that XopD_{Xcc8004} is de-ubiquitinating or de-sumoylating RGA, the study strongly suggests that XopD_{Xcc8004} somehow modifies RGA to prevent its proteasome-mediated degradation (Tan et al., 2014).

Although members of the YopJ-like effectors share restricted homology to SUMO proteases, *Xanthomonas* YopJ-like effector AvrXv4 was shown to decrease the accumulation of SUMO-modified proteins in plants (Rodén et al., 2004). To date, it remains

unclear whether AvrXv4 possesses SUMO isopeptidase activity and which targets are possibly de-sumoylated by AvrXv4.

EFFECTOR PROTEINS HIJACKING THE UPS BY MIMICKING EUKARYOTIC PROTEINS

Due to the lack of structural or sequence similarities to proteins with known function, enzymatic activities for T3Es of plant pathogenic bacteria have been difficult to predict. However, the determination of the crystal structure of a number of effectors from different bacterial pathogens revealed conserved structural features with components of the host UPS (Perrett et al., 2011). For instance, crystal structure determination of *Xanthomonas* T3E XopL revealed that the protein possesses a novel fold and hence belongs to a new class of E3 ubiquitin ligases (Singer et al., 2013). Structural analysis of XopL revealed similarities to T3E E3 ligases from *Salmonella* or *Shigella*, providing first cues of an E3 ubiquitin ligase activity of XopL. Further biochemical analysis confirmed this observation, as XopL exhibits E3 ubiquitin ligase activity and interacts with specific plant E2 enzymes. The E3 ligase activity of XopL is responsible for cell death induction and also for suppression of plant immunity (Singer et al., 2013).

Alongside E3 ligases, it has been shown that proteins harboring F-box motifs are implicated in protein ubiquitination. The F-box domain is a structural motif that is ~50 amino acids long mediating protein-protein interactions (Perrett et al., 2011). F-box proteins form a heterotetrameric ubiquitin ligase complex (SCF complex), consisting of SKP1 (S-phase-kinase-associated protein 1), Cullin and F-box proteins, mediating ubiquitination of proteins targeted for proteasomal degradation (Sadanandom et al., 2012). The first evidence that F-box proteins play a major role in plant immunity was provided by the identification of the F-box protein CORONATINE INSENSITIVE 1 (COI1), which functions as a receptor for jasmonate (Xie et al., 1998). To date, only one T3E, XopI, from *X. campestris* pv. *vesicatoria* strain 85-10 containing a F-box motif was identified, based on the presence of a PIP (pathogen-inducible promoter) box in its promoter region (Schulze et al., 2012). Type-III dependent secretion and translocation of XopI was shown during the interaction of Xcv with resistant pepper plants. However, plant target(s) of XopI remain to be identified to clarify its role in the manipulation of the UPS.

CONCLUSION

Manipulation of ubiquitin and ubiquitin-like pathways has emerged as an effective virulence strategy for pathogenic bacteria during the past years. Several *Xanthomonas* species and pathovars appear to utilize T3E proteins from widespread families such as the YopJ-like superfamily or XopD-like family to interfere with the UPS. In addition, newly identified T3E with novel structural motifs, such as *Xanthomonas* effector XopL provide further examples. Besides T3Es acting as proteasome inhibitors, others rely on proteasome activity for their function leading to an apparent contradiction. In *X. campestris* pv. *vesicatoria* 85-10, the proteasome inhibitor XopJ and the E3 ligase XopL constitute such a effector pair. This conflicting action of T3E proteins might be resolved if T3Es interfering with the UPS would act spatially separated from each other. Posttranslational myristylation of XopJ is responsible for its plasma membrane localization (Bartetzko et al., 2009). This

feature is essential for the suppression of the proteasome activity, as XopJ interacts with RPT6 at the plasma membrane and only myristoylated XopJ is able to inhibit proteasome activity (Üstün et al., 2013). It is possible that XopL might act as an E3 ligase at a different compartment and thus, action of both T3E are separated spatially. This might be the case for XopJ and XopD, another pair with contradictory functions, as XopD acts in the host nucleus and XopJ at the plant plasma membrane. Another option would be the timing of delivery by the type III secretion system of Xcv, hence avoiding conflicting actions of both effectors.

ACKNOWLEDGMENTS

Work on plant-pathogen interactions in the authors' laboratory is funded by the Deutsche Forschungsgemeinschaft (SFB 796: Reprogramming of host cells by microbial effectors and BO 1916-5/1).

REFERENCES

- Abramovitch, R. B., Janjusevic, R., Stebbins, C. E., and Martin, G. B. (2006). Type III effector AvrPtoB requires intrinsic E3 ubiquitin ligase activity to suppress plant cell death and immunity. *Proc. Natl. Acad. Sci. U.S.A.* 103, 2851–2856. doi: 10.1073/pnas.0507892103
- Bartetzko, V., Sonnewald, S., Vogel, F., Hartner, K., Stadler, R., Hammes, U. Z., et al. (2009). The *Xanthomonas campestris* pv. vesicatoria type III effector protein XopJ inhibits protein secretion: evidence for interference with cell wall-associated defense responses. *Mol. Plant Microbe Interact.* 22, 655–664. doi: 10.1094/MPMI-22-6-0655
- Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., et al. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 326, 1509–1512. doi: 10.1126/science.1178811
- Boyer, L., and Lemichez, E. (2004). Targeting of host-cell ubiquitin and ubiquitin-like pathways by bacterial factors. *Nat. Rev. Microbiol.* 2, 779–788. doi: 10.1038/nrmicro1005
- Buttner, D., and Bonas, U. (2010). Regulation and secretion of *Xanthomonas* virulence factors. *FEMS Microbiol. Rev.* 34, 107–133. doi: 10.1111/j.1574-6976.2009.00192.x
- Cheong, M. S., Kirik, A., Kim, J. G., Frame, K., Kirik, V., and Mudgett, M. B. (2014). AvrBsT acetylates *Arabidopsis* ACIP1, a protein that associates with microtubules and is required for immunity. *PLoS Pathog.* 10:e1003952. doi: 10.1371/journal.ppat.1003952
- Domingues, M. N., De Souza, T. A., Cernadas, R. A., De Oliveira, M. L., Docena, C., Farah, C. S., et al. (2010). The *Xanthomonas citri* effector protein PthA interacts with citrus proteins involved in nuclear transport, protein folding and ubiquitination associated with DNA repair. *Mol. Plant Pathol.* 11, 663–675. doi: 10.1111/j.1364-3703.2010.00636.x
- Farras, R., Ferrando, A., Jasik, J., Kleinow, T., Okresz, L., Tiburcio, A., et al. (2001). SKP1-SnRK protein kinase interactions mediate proteasomal binding of a plant SCF ubiquitin ligase. *EMBO J.* 20, 2742–2756. doi: 10.1093/emboj/20.11.2742
- Furlan, G., Klinkenberg, J., and Trujillo, M. (2012). Regulation of plant immune receptors by ubiquitination. *Front. Plant Sci.* 3:238. doi: 10.3389/fpls.2012.00238
- Groll, M., Schellenberg, B., Bachmann, A. S., Archer, C. R., Huber, R., Powell, T. K., et al. (2008). A plant pathogen virulence factor inhibits the eukaryotic proteasome by a novel mechanism. *Nature* 452, 755–758. doi: 10.1038/nature06782
- Hotson, A., Chosed, R., Shu, H., Orth, K., and Mudgett, M. B. (2003). *Xanthomonas* type III effector XopD targets SUMO-conjugated proteins in planta. *Mol. Microbiol.* 50, 377–389. doi: 10.1046/j.1365-2958.2003.03730.x
- Hotson, A., and Mudgett, M. B. (2004). Cysteine proteases in phytopathogenic bacteria: identification of plant targets and activation of innate immunity. *Curr. Opin. Plant Biol.* 7, 384–390. doi: 10.1016/j.pbi.2004.05.003
- Ishikawa, K., Yamaguchi, K., Sakamoto, K., Yoshimura, S., Inoue, K., Tsuge, S., et al. (2014). Bacterial effector modulation of host E3 ligase activity suppresses PAMP-triggered immunity in rice. *Nat. Commun.* 5:5430. doi: 10.1038/ncomms6430
- Janjusevic, R., Abramovitch, R. B., Martin, G. B., and Stebbins, C. E. (2006). A bacterial inhibitor of host programmed cell death defenses is an E3 ubiquitin ligase. *Science* 311, 222–226. doi: 10.1126/science.1120131
- Jiang, S., Yao, J., Ma, K. W., Zhou, H., Song, J., He, S. Y., et al. (2013). Bacterial effector activates jasmonate signaling by directly targeting JAZ transcriptional repressors. *PLoS Pathog.* 9:e1003715. doi: 10.1371/journal.ppat.1003715
- Jones, J. D., and Dangl, J. L. (2006). The plant immune system. *Nature* 444, 323–329. doi: 10.1038/nature05286
- Kim, J. G., Stork, W., and Mudgett, M. B. (2013). *Xanthomonas* type III effector XopD desumoylates tomato transcription factor SLERF4 to suppress ethylene responses and promote pathogen growth. *Cell Host Microbe* 13, 143–154. doi: 10.1016/j.chom.2013.01.006
- Kim, J. G., Taylor, K. W., Hotson, A., Keegan, M., Schmelz, E. A., and Mudgett, M. B. (2008). XopD SUMO protease affects host transcription, promotes pathogen growth, and delays symptom development in xanthomonas-infected tomato leaves. *Plant Cell* 20, 1915–1929. doi: 10.1105/tpc.108.058529
- Lee, A. H., Hurley, B., Felsensteiner, C., Yea, C., Ckurskumova, W., Bartetzko, V., et al. (2012). A bacterial acetyltransferase destroys plant microtubule networks and blocks secretion. *PLoS Pathog.* 8:e1002523. doi: 10.1371/journal.ppat.1002523
- Lee, D. H., Choi, H. W., and Hwang, B. K. (2011). The pepper E3 ubiquitin ligase RING1 gene, CaRING1, is required for cell death and the salicylic acid-dependent defense response. *Plant Physiol.* 156, 2011–2025. doi: 10.1104/pp.111.177568
- Lewis, J. D., Lee, A., Ma, W., Zhou, H., Guttman, D. S., and Desveaux, D. (2011). The YopJ superfamily in plant-associated bacteria. *Mol. Plant Pathol.* 12, 928–937. doi: 10.1111/j.1364-3703.2011.00719.x
- Marino, D., Peeters, N., and Rivas, S. (2012). Ubiquitination during plant immune response. *Plant Physiol.* 160, 15–27. doi: 10.1104/pp.112.199281
- Misas-Villamil, J. C., Kolodziejek, I., Crabill, E., Kaschani, F., Niessen, S., Shindo, T., et al. (2013). *Pseudomonas syringae* pv. syringae uses proteasome inhibitor syringolin A to colonize from wound infection sites. *PLoS Pathog.* 9:e1003281. doi: 10.1371/journal.ppat.1003281
- Mukherjee, S., Keitany, G., Li, Y., Wang, Y., Ball, H. L., Goldsmith, E. J., et al. (2006). *Yersinia* YopJ acetylates and inhibits kinase activation by blocking phosphorylation. *Science* 312, 1211–1214. doi: 10.1126/science.1126867
- Mukhopadhyay, D., and Riezman, H. (2007). Proteasome-independent functions of ubiquitin in endocytosis and signaling. *Science* 315, 201–205. doi: 10.1126/science.1127085
- Noel, L., Thieme, F., Gabler, J., Buttner, D., and Bonas, U. (2003). XopC and XopJ, two novel type III effector proteins from *Xanthomonas campestris* pv. vesicatoria. *J. Bacteriol.* 185, 7092–7102. doi: 10.1128/JB.185.24.7092-7102.2003
- Nomura, K., Debroy, S., Lee, Y. H., Pumplin, N., Jones, J., and He, S. Y. (2006). A bacterial virulence protein suppresses host innate immunity to cause plant disease. *Science* 313, 220–223. doi: 10.1126/science.1129523
- Orth, K., Xu, Z., Mudgett, M. B., Bao, Z. Q., Palmer, L. E., Bliska, J. B., et al. (2000). Disruption of signaling by *Yersinia* effector YopJ, a ubiquitin-like protein protease. *Science* 290, 1594–1597. doi: 10.1126/science.290.5496.1594
- Perrett, C. A., Lin, D. Y., and Zhou, D. (2011). Interactions of bacterial proteins with host eukaryotic ubiquitin pathways. *Front. Microbiol.* 2:143. doi: 10.3389/fmicb.2011.00143
- Roden, J., Eardley, L., Hotson, A., Cao, Y., and Mudgett, M. B. (2004). Characterization of the *Xanthomonas* AvrXv4 effector, a SUMO protease translocated into plant cells. *Mol. Plant Microbe Interact.* 17, 633–643. doi: 10.1094/MPMI.2004.17.6.633
- Ryan, R. P., Vorholter, F. J., Potnis, N., Jones, J. B., Van Sluys, M. A., Bogdanove, A. J., et al. (2011). Pathogenomics of *Xanthomonas*: understanding bacterium-plant interactions. *Nat. Rev. Microbiol.* 9, 344–355. doi: 10.1038/nrmicro2558
- Sadanandom, A., Bailey, M., Ewan, R., Lee, J., and Nelis, S. (2012). The ubiquitin-proteasome system: central modifier of plant signalling. *New Phytol.* 196, 13–28. doi: 10.1111/j.1469-8137.2012.04266.x
- Schellenberg, B., Ramel, C., and Dudler, R. (2010). *Pseudomonas syringae* virulence factor syringolin A counteracts stomatal immunity by proteasome inhibition. *Mol. Plant Microbe Interact.* 23, 1287–1293. doi: 10.1094/MPMI-04-10-0094
- Schulze, S., Kay, S., Buttner, D., Egler, M., Eschen-Lippold, L., Hause, G., et al. (2012). Analysis of new type III effectors from *Xanthomonas* uncovers XopB and XopS as suppressors of plant immunity. *New Phytol.* 195, 894–911. doi: 10.1111/j.1469-8137.2012.04210.x
- Singer, A. U., Schulze, S., Skarina, T., Xu, X., Cui, H., Eschen-Lippold, L., et al. (2013). A pathogen type III effector with a novel E3 ubiquitin ligase architecture. *PLoS Pathog.* 9:e1003121. doi: 10.1371/journal.ppat.1003121

- Stegmann, M., Anderson, R. G., Ichimura, K., Pecenkova, T., Reuter, P., Zarsky, V., et al. (2012). The ubiquitin ligase PUB22 targets a subunit of the exocyst complex required for PAMP-triggered responses in *Arabidopsis*. *Plant Cell* 24, 4703–4716. doi: 10.1105/tpc.112.104463
- Szczesny, R., Buttner, D., Escobar, L., Schulze, S., Seiferth, A., and Bonas, U. (2010). Suppression of the AvrBs1-specific hypersensitive response by the YopJ effector homolog AvrBsT from *Xanthomonas* depends on a SNF1-related kinase. *New Phytol.* 187, 1058–1074. doi: 10.1111/j.1469-8137.2010.03346.x
- Tan, L., Rong, W., Luo, H., Chen, Y., and He, C. (2014). The *Xanthomonas campestris* effector protein XopDXcc8004 triggers plant disease tolerance by targeting DELLA proteins. *New Phytol.* 204, 595–608. doi: 10.1111/nph.12918
- Tasset, C., Bernoux, M., Jauneau, A., Pouzet, C., Briere, C., Kieffer-Jacquinod, S., et al. (2010). Autoacetylation of the *Ralstonia solanacearum* effector PopP2 targets a lysine residue essential for RRS1-R-mediated immunity in *Arabidopsis*. *PLoS Pathog.* 6:e1001202. doi: 10.1371/journal.ppat.1001202
- Thieme, F., Szczesny, R., Urban, A., Kirchner, O., Hause, G., and Bonas, U. (2007). New type III effectors from *Xanthomonas campestris* pv. vesicatoria trigger plant reactions dependent on a conserved N-myristylation motif. *Mol. Plant Microbe Interact.* 20, 1250–1261. doi: 10.1094/MPMI-20-10-1250
- Trujillo, M., Ichimura, K., Casais, C., and Shirasu, K. (2008). Negative regulation of PAMP-triggered immunity by an E3 ubiquitin ligase triplet in *Arabidopsis*. *Curr. Biol.* 18, 1396–1401. doi: 10.1016/j.cub.2008.07.085
- Üstün, S., Bartetzko, V., and Bornke, F. (2013). The *Xanthomonas campestris* type III effector XopJ targets the host cell proteasome to suppress salicylic-acid mediated plant defence. *PLoS Pathog.* 9:e1003427. doi: 10.1371/journal.ppat.1003427
- Üstün, S., Konig, P., Guttmann, D. S., and Bornke, F. (2014). HopZ4 from *Pseudomonas syringae*, a member of the HopZ type III effector family from the yopI superfamily, inhibits the proteasome in plants. *Mol. Plant Microbe Interact.* 27, 611–623. doi: 10.1094/MPMI-12-13-0363-R
- Vierstra, R. D. (2009). The ubiquitin-26S proteasome system at the nexus of plant biology. *Nat. Rev. Mol. Cell. Biol.* 10, 385–397. doi: 10.1038/nrm2688
- Vierstra, R. D. (2012). The expanding universe of ubiquitin and ubiquitin-like modifiers. *Plant Physiol.* 160, 2–14. doi: 10.1104/pp.112.200667
- White, F. F., Potnis, N., Jones, J. B., and Koebnik, R. (2009). The type III effectors of *Xanthomonas*. *Mol. Plant Pathol.* 10, 749–766. doi: 10.1111/j.1364-3703.2009.00590.x
- Xie, D. X., Feys, B. F., James, S., Nieto-Rostro, M., and Turner, J. G. (1998). COI1: an *Arabidopsis* gene required for jasmonate-regulated defense and fertility. *Science* 280, 1091–1094. doi: 10.1126/science.280.5366.1091

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.

Received: 01 November 2014; accepted: 03 December 2014; published online: 18 December 2014.

Citation: Üstün S and Börnke F (2014) Interactions of *Xanthomonas* type-III effector proteins with the plant ubiquitin and ubiquitin-like pathways. *Front. Plant Sci.* 5:736. doi: 10.3389/fpls.2014.00736

This article was submitted to Plant-Microbe Interaction, a section of the journal *Frontiers in Plant Science*.

Copyright © 2014 Üstün and Börnke. 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.

The *Xanthomonas* effector XopJ triggers a conditional hypersensitive response upon treatment of *N. benthamiana* leaves with salicylic acid

Suayib Üstün¹, Verena Bartetzko² and Frederik Börnke^{1,2,3*}

OPEN ACCESS

Edited by:

Nicolas Denancé,
Institut National de la Recherche
Agronomique, France

Reviewed by:

Susana Rivas,
Laboratoire des Interactions
Plantes-Microorganismes – CNRS,
France
Luis A. J. Mur,
Aberystwyth University, UK
Kee Hoon Sohn,
Massey University, New Zealand

*Correspondence:

Frederik Börnke,
Plant Health, Plant Metabolism
Group, Leibniz-Institute of Vegetable
and Ornamental Crops,
Theodor-Echtermeyer-Weg 1,
14979 Großbeeren, Germany
boernke@igzev.de

Specialty section:

This article was submitted to
Plant Biotic Interactions,
a section of the journal
Frontiers in Plant Science

Received: 18 March 2015

Accepted: 20 July 2015

Published: 03 August 2015

Citation:

Üstün S, Bartetzko V and Börnke F (2015) The *Xanthomonas* effector XopJ triggers a conditional hypersensitive response upon treatment of *N. benthamiana* leaves with salicylic acid. *Front. Plant Sci.* 6:599.
doi: 10.3389/fpls.2015.00599

XopJ is a *Xanthomonas* type III effector protein that promotes bacterial virulence on susceptible pepper plants through the inhibition of the host cell proteasome and a resultant suppression of salicylic acid (SA) – dependent defense responses. We show here that *Nicotiana benthamiana* leaves transiently expressing XopJ display hypersensitive response (HR) –like symptoms when exogenously treated with SA. This apparent avirulence function of XopJ was further dependent on effector myristoylation as well as on an intact catalytic triad, suggesting a requirement of its enzymatic activity for HR-like symptom elicitation. The ability of XopJ to cause a HR-like symptom development upon SA treatment was lost upon silencing of SGT1 and NDR1, respectively, but was independent of EDS1 silencing, suggesting that XopJ is recognized by an R protein of the CC-NBS-LRR class. Furthermore, silencing of NPR1 abolished the elicitation of HR-like symptoms in XopJ expressing leaves after SA application. Measurement of the proteasome activity indicated that proteasome inhibition by XopJ was alleviated in the presence of SA, an effect that was not observed in NPR1 silenced plants. Our results suggest that XopJ – triggered HR-like symptoms are closely related to the virulence function of the effector and that XopJ follows a two-signal model in order to elicit a response in the non-host plant *N. benthamiana*.

Keywords: *Xanthomonas*, type-III effector, XopJ, avirulence, salicylic acid

Introduction

In nature, plants are continuously attacked by a broad range of potential pathogens. However, the majority of plants are resistant to most pathogen species. This form of resistance is known as non-host resistance (NHR) and can be defined as a broad-spectrum plant defense that provides immunity to all members of a plant species against all isolates of a micro-organism that is pathogenic on other plant species (Senthil-Kumar and Mysore, 2013). In order for a pathogen to be successful and cause disease it has to defeat the plant's multilayered immune system. Before it can enter the plant tissue, the pathogen is exposed to a range of preformed physical and chemical barriers, already preventing the entry of many non-adapted pathogens at an early

step. If a pathogen is able to overcome these barriers and comes into contact with the plant cell surface, it will face induced plant defenses. Surface-localized pattern recognition receptors (PRRs) can perceive conserved pathogen molecules (PAMPs, pathogen associated molecular patterns) which in case of bacteria are for instance flagellin, elongation factor Tu (EF-Tu), peptidoglycan (PGN) or lipopolysaccharides (Macho and Zipfel, 2015). This recognition results in the initiation of intra cellular downstream signaling that leads to the production of reactive oxygen species, stimulation of mitogen-activated protein kinase (MAPK) cascades, defense gene induction, and callose deposition at the plant cell wall (Boller and Felix, 2009). These induced defense outputs are in most cases sufficiently effective to eradicate a potential pathogen from infected tissue and are collectively referred to as PAMP-triggered immunity (PTI; Jones and Dangl, 2006). As a response, several Gram-negative pathogenic bacteria use a type-III secretion system (T3SS) to inject a suite of so called type-III effector proteins (T3Es) into their eukaryotic host cell (Galan et al., 2014). These T3Es are targeted to a number of cellular compartments where they influence host cellular processes to provide a beneficial environment for the pathogen to promote pathogen multiplication and disease. In order to counter this, plants have evolved the ability to recognize specific effector proteins through resistance (R) proteins, a class of receptor proteins that typically contain nucleotide-binding domains (NB) and leucine rich repeats (LRRs), (Dodds and Rathjen, 2010). Recognition of T3Es [in that case also referred to as avirulence (Avr) proteins] by NB-LRR proteins can either be directly through physical interaction between both proteins or indirectly through an accessory protein that is part of an NB-LRR protein complex (Dodds and Rathjen, 2010). During indirect recognition, the so called guard hypothesis, it is assumed that the effector interaction is mediated by the effector target protein or a structural mimic thereof (Dangl and Jones, 2001). The activity of the T3E induces structural changes in its target protein that enables recognition by the NB-LRR protein, leading to its activation and finally results in effector-triggered immunity (ETI; Jones and Dangl, 2006). Generally, PTI and ETI give rise to similar responses, although ETI is qualitatively stronger and faster and often involves a rapid form of localized cell death called the hypersensitive response (HR) that is assumed to limit spread of biotrophic pathogens from the site of infection. The NB-LRR repertoire recognition of pathogen effector proteins is highly dynamic and depends on the genotype of a given host cultivar. Thus, ETI primarily protects against specific races of pathogens because it is only triggered when an Avr factor, i.e., a particular T3E, on the pathogen side comes together with a matching R protein on the host side. Although ETI is a major component of host/pathogen race specificity, its role in NHR is not well understood. In some cases, T3Es trigger ETI in non-host plants, suggesting a role for ETI in determining the host range of a pathogen (Staskawicz et al., 1987; Kobayashi et al., 1989; Wei et al., 2007; Wróblewski et al., 2009). The signal transduction and physiological processes leading to HR during ETI are not well understood but it appears that the defense hormone salicylic acid (SA) plays a central role in the induction of such a

resistance response (Glazebrook, 2005). SA depletion in plants by transgenic expression of a bacterial SA hydroxylase encoded by *nahG* suppresses R gene mediated defenses elicited by a range of bacterial, oomycete, and viral pathogens (Delaney et al., 1994; Raidan and Delaney, 2002). While in general SA is active against biotrophic pathogens some necrotrophs have acquired strategies to induce SA signaling during infection in order to promote host cell death and thus virulence. For example, the fungus *Botrytis cinerea* produces an exopolysaccharide, which acts as an elicitor of the SA pathway. In turn, the SA pathway antagonizes the jasmonic acid signaling pathway that would otherwise restrict virulence of this necrotrophic pathogen (El Oirdi et al., 2011).

The Gram-negative phytopathogenic bacterium *Xanthomonas campestris* pv. *vesicatoria* (Xcv) is the causal agent of bacterial spot disease on pepper and tomato plants (Ryan et al., 2011). During infection, it secretes a cocktail of 20–40 T3Es into the plant cell that collectively suppress defense and allow bacterial propagation (Thieme et al., 2007; White et al., 2009). Although these T3Es likely play a role in virulence in susceptible hosts, they can also have Avr function and trigger ETI in certain genotypes of pepper and tomato plants expressing cognate R proteins as well as in non-host plants from other species (Whalen et al., 1993; Bonshtien et al., 2005; Kim et al., 2010; Szczesny et al., 2012). Transient expression of T3Es by infiltration of leaves from *Nicotiana benthamiana* with *Agrobacterium* is widely used to characterize T3E virulence functions in plants (Bartetzko et al., 2009; Gurlebeck et al., 2009; Üstün et al., 2013; Stork et al., 2015). In some cases, expression of T3Es from Xcv in *N. benthamiana* has led to the induction of ETI associated with signs of an HR (Thieme et al., 2007; Schulze et al., 2012; Singer et al., 2013).

The Xcv T3E XopJ is a member of the widespread YopJ-family of effector proteins that is present among plant and animal pathogenic bacteria and whose members are highly diversified in virulence function (Lewis et al., 2011). XopJ and its close homolog HopZ4 from *Pseudomonas syringae* pv. *lachrymans* have been shown to interact with the proteasomal subunit RPT6 in planta to suppress proteasome activity (Üstün et al., 2013, 2014). XopJ-triggered proteasome suppression results in the inhibition of SA-related immune responses to attenuate onset of necrosis and to alter host transcription (Üstün et al., 2013). Transient expression of XopJ in leaves of *N. benthamiana* was instrumental to elucidate its function (Thieme et al., 2007; Bartetzko et al., 2009; Üstün et al., 2013). Using a XopJ-green fluorescent protein (GFP) fusion protein a localization of the effector to the plasma membrane of the host cell mediated by myristylation of the protein could be demonstrated (Thieme et al., 2007; Bartetzko et al., 2009). Furthermore, XopJ's inhibitory effect on protein secretion was shown by transient co-expression of the effector together with a secretable GFP variant (Bartetzko et al., 2009). In some of these experiments XopJ was reported to elicit a cell death reaction in *N. benthamiana* 2–4 days post inoculation, suggesting recognition of the effector in this non-host plant (Thieme et al., 2007). However, this reaction was not observed in other studies (Bartetzko et al., 2009; Üstün et al., 2013). Usually, first signs of an HR become apparent within a few hours after the Avr protein is delivered to the

host cell by the bacterial T3SS (Morel and Dangl, 1997). Although transient overexpression could lead to different kinetics of effector recognition in comparison to T3SS delivery, XopJ triggered cell death in *N. benthamiana* at late time points of expression could also have other reasons than weak recognition by a cognate R protein. For instance, XopJ could interfere with cellular functions requiring proteasome activity leading to a general perturbation of protein homeostasis.

In the present study, we show that XopJ elicits a rapid HR-like response in *N. benthamiana* when leaves transiently expressing the effector are sprayed with SA. Development of HR-like symptoms was closely related to XopJ's virulence function and appears to involve indirect recognition by an R protein. A two-signal model leading to the elicitation of HR-like symptoms by XopJ in *N. benthamiana* is discussed.

Materials and Methods

Plant Material and Growth Conditions

Tobacco plants (*N. benthamiana*) were grown in soil in a greenhouse with daily watering, and subjected to a 16 h light: 8 h dark cycle (25°C: 21°C) at 300 $\mu\text{mol m}^{-2} \text{s}^{-1}$ light and 75% relative humidity.

Transient Expression Assays and SA Treatment

For infiltration of *N. benthamiana* leaves, *A. tumefaciens* C58C1 was infiltrated into the abaxial air space of 4- to 6-week-old plants, using a needleless 2-ml syringe. Agrobacteria were cultivated overnight at 28°C in the presence of appropriate antibiotics. The cultures were harvested by centrifugation, and the pellet was resuspended in sterile water to a final optical density at (OD₆₀₀) of 1.0. SA treatment was performed 24 h after agro-infiltration. Infiltrated leaves were sprayed with 5 mM SA (containing 0,005% v/v Silwet-77) or water (containing 0,005% v/v Silwet-77) and phenotypes were analyzed 24 h later.

Western Blotting

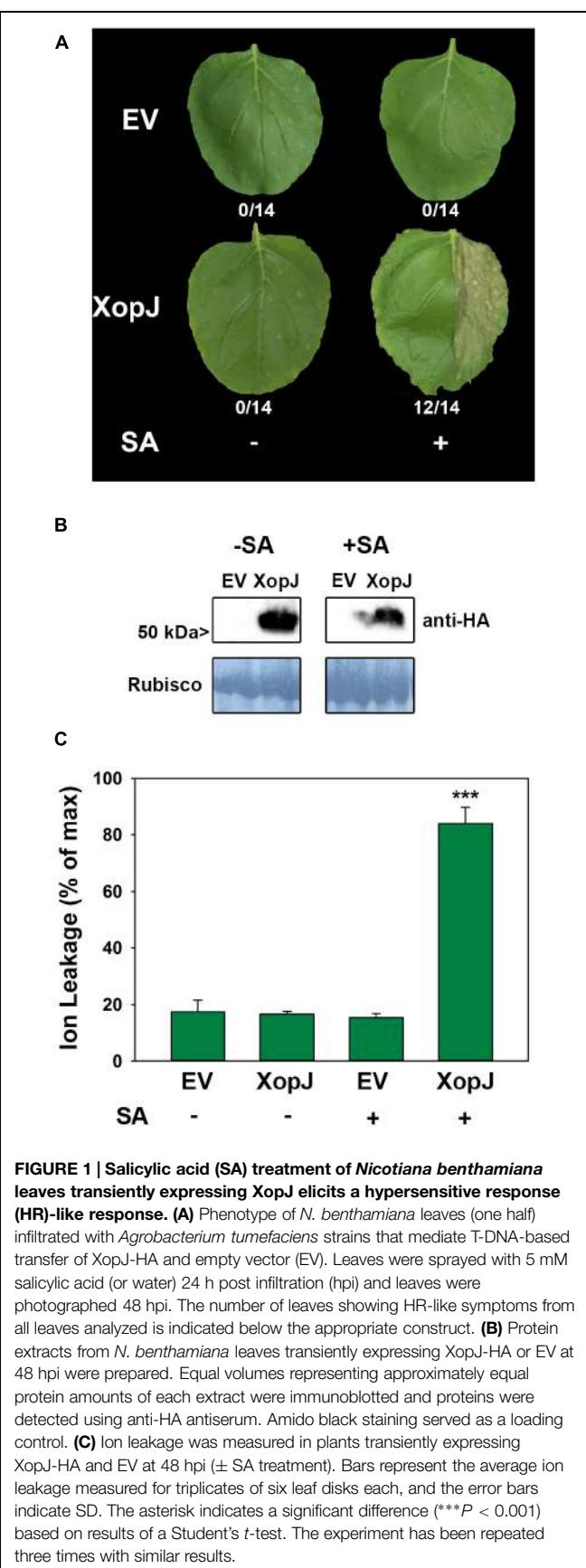
Leaf material was homogenized in sodium-dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) loading buffer (100 mM Tris-HCl, pH 6.8; 9% β -mercapto-ethanol, 40% glycerol, 0.0005% bromophenol blue, 4% SDS) and, after heating for 10 min at 95°C, subjected to gel electrophoresis. Separated proteins were transferred onto nitrocellulose membrane (Porablot, Machery und Nagel, Düren, Germany). Proteins were detected by an anti-HA-Peroxidase high affinity antibody (Roche).

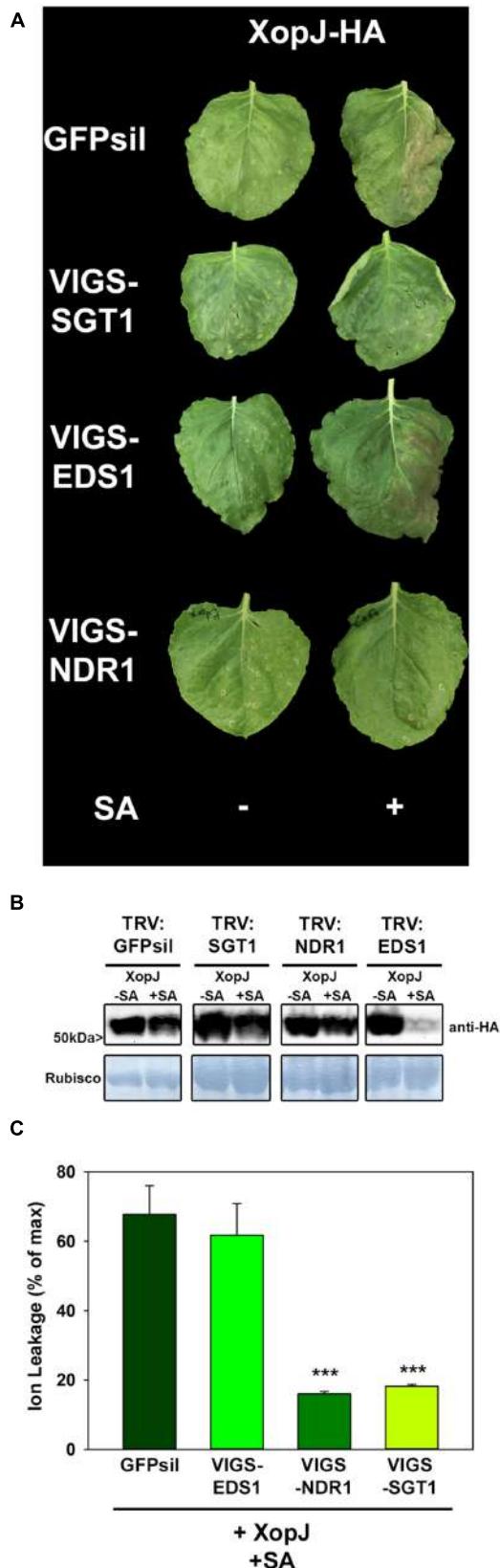
Measurement of Proteasome Activity

Proteasome activity in crude plant extracts was determined spectro-fluorometrically using the fluorogenic substrate succ-LLVY-NH-AMC (Sigma) according to Üstün et al. (2013).

Virus-Induced Gene Silencing of *N. benthamiana*

Virus-induced gene silencing (VIGS) was performed as described previously (Üstün et al., 2012). pTRV2-SGT1, pTRV2-NPR1, and



**FIGURE 2 | Continued**

XopJ-triggered cell-death after SA application requires components of R-protein mediated signaling. **(A)** Phenotypes of *N. benthamiana* SGT1, EDS1, and NDR1 VIGS plants transiently expressing XopJ-HA and EV in comparison to the GFPsil VIGS control leaves expressing XopJ-HA and EV with or without SA treatment. Photographs were taken at 48 hpi. **(B)** Protein extracts from TRV:GFPsil, TRV:SGT1, TRV:NDR1, and TRV:EDS1 transiently expressing XopJ-HA at 48 hpi were prepared. Equal volumes representing approximately equal protein amounts of each extract were blotted onto a nitrocellulose membrane and protein was detected using anti-HA antiserum. Amido black staining served as a loading control. **(C)** Electrolyte conductivity was measured in TRV:GFPsil, TRV:EDS1, and TRV:SGT1 and TRV:NDR1 plants transiently expressing XopJ-HA (sprayed with 5 mM SA 24 hpi) at 48 hpi. Bars represent the average ion leakage measured for triplicates of six leaf disks each, and the error bars indicate SD. Treatments were compared with TRV:GFPsil transiently expressing XopJ-HA and significant differences are indicated by asterisks (**P < 0.001). The experiment was repeated three time with similar results.

pTRV2-EDS1, pYL279-NDR1 (Liu et al., 2002) were obtained from the *Arabidopsis* Biological Resource Center (<http://www.arabidopsis.org>). Briefly, *Agrobacterium* strains with the pTRV1 vector and with pTRV2-GFPsil, pYL279-RPT6, (Üstün et al., 2012, 2013), pTRV2-SGT1, pTRV2-NPR1, and pTRV2-EDS1 (OD₆₀₀ = 1.0) were mixed in a 1:1 ratio, respectively, and the mixture was infiltrated into a lower leaf of a 4-week-old *N. benthamiana* plant using a 1-mL sterile syringe without a needle. Fourteen days of post infiltration silenced plants were used for further transient expression studies.

Ion Leakage Measurements

For electrolyte leakage experiments, triplicates of 1.76 cm² infected leaf material were taken at 48 h post infiltration (hpi). Leaf disks were placed on the bottom of a 15-ml tube. Eight milliliters of deionized water was added to each tube. After 24 h of incubation in a rotary shaker at 4°C, conductivity was determined with a conductometer. To measure the maximum conductivity of the entire sample, conductivity was determined after boiling the samples for 30 min (Üstün et al., 2012).

Results

Treatment of *N. benthamiana* Leaves Transiently Expressing XopJ with SA Rapidly Induces Cell Death

Previous results suggested that during a compatible interaction of Xcv with pepper XopJ exerts its virulence function by inhibiting SA-mediated defense responses (Üstün et al., 2013). However, when Xcv infected pepper leaves were treated with SA they developed necrotic lesions that were comparable to those observed on Xcv Δ XopJ infected leaves at the same time point without SA treatment (Üstün et al., 2013). Thus, Xcv infected tissue remains sensitive to exogenously applied SA even in the presence of XopJ. To further study the role of XopJ in interfering with SA-related processes, we sought to investigate the consequences of SA application on XopJ expressing leaves of the non-host plant *N. benthamiana*. To this end, an HA-tagged

FIGURE 2 | Continued

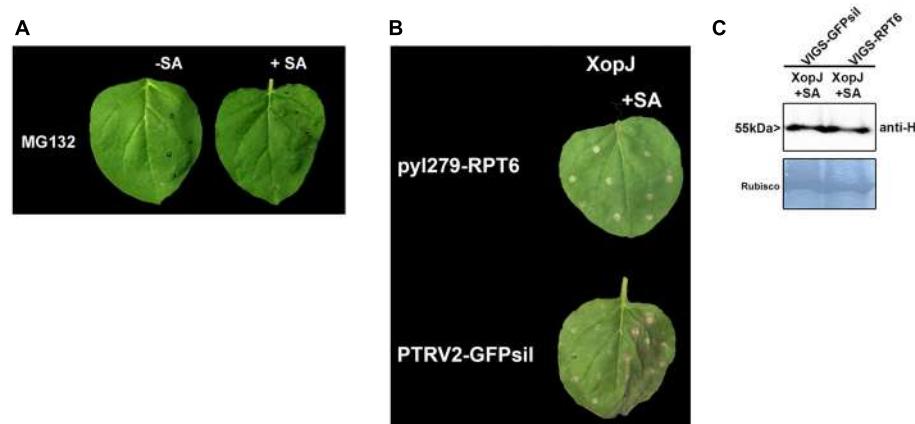


FIGURE 3 | RPT6 is required for XopJ-induced HR-like symptoms.

(A) Phenotypes of *N. benthamiana* leaves infiltrated with 100 μ M MG132 with or without SA. (B) XopJ was transiently expressed in RPT6 silenced leaves or control plants and treated with 5 mM SA 24 hpi. Plant reactions were

photographed 48 h after transient expression of XopJ. (C) Western blot analysis of XopJ-HA protein levels in TRV:RPT6 and TRV:GFPsil control plants using an anti-HA antibody. Amido black staining served as a loading control. The experiment was repeated three times with similar results.

version of XopJ under control of the CaMV35S promoter (XopJ-HA) was transiently expressed in leaves of *N. benthamiana* using *Agrobacterium*-infiltration. Twenty four hours post infiltration (hpi) XopJ-HA infiltrated and empty vector (EV) infiltrated control leaves were sprayed with 5 mM SA. As shown in Figure 1A, XopJ-HA expressing leaves showed tissue collapse and developed necrosis 48 hpi when treated with SA. Untreated leaves showed no signs of tissue damage even in the presence of XopJ, indicating that XopJ alone is not able to trigger necrotic cell death in the time period investigated but requires exogenously applied SA. XopJ protein expression was not affected by the SA-treatment (Figure 1B). The observed phenotype in SA-treated XopJ expressing leaves resembles that of a HR which is usually associated with R-protein-mediated immunity triggered upon recognition of a pathogen-derived Avr protein. An HR is often preceded by an increase in electrolyte leakage in dying cells, and measurement of electrolyte leakage caused by membrane damage is a quantitative measure of HR-associated cell death (Mackey et al., 2003). Ion leakage was strongly increased in SA-treated XopJ expressing leaves as compared to the EV control (Figure 1C). This effect was completely dependent on SA treatment as untreated XopJ expressing tissue did not show signs of cell damage (Figure 1C). Previous results indicated that a myristylation motif at the N-terminus guides XopJ to the plasma membrane and this subcellular localization as well as an intact catalytic triad is required for the effector to function (Üstün et al., 2013). Leaves expressing a XopJ(G2A)-HA protein, which is no longer myristoylated, or a catalytically inactive variant carrying a C to A substitution at position 235 [XopJ(C235A)], developed no visible signs of tissue damage when treated with SA (Supplementary Figure S1). This suggests that development of SA-dependent phenotypes requires XopJ to be fully functional. To confirm that this effect is specific to XopJ, we transiently expressed an unrelated *Xanthomonas* effector, XopS, and treated plants with 5 mM SA. No visible signs of HR-like cell death were visible on leaves either untreated or sprayed with SA

(Supplementary Figure S2), indicating that the induction of tissue collapse after SA application is specific for XopJ.

XopJ-Mediated Cell Death after SA Treatment Requires Signaling Components of R-Protein-Mediated Resistance

The results obtained thus far suggest that tissue collapse and necrosis upon SA treatment of XopJ expressing *N. benthamiana* leaves could involve an HR-like process and thus might be the consequence of R-protein mediated recognition of the effector triggered by SA. Defense signaling by R proteins requires further signaling components such as SGT1 (suppressor of G2 allele of skp1) which, in *N. benthamiana* was found to be required for responses mediated by a diverse range of R proteins against various pathogens (Peart et al., 2002). In order to investigate an involvement of SGT1 in XopJ-mediated cell death after SA-treatment, VIGS with Tobacco rattle virus (TRV), followed by *Agrobacterium*-infiltration and SA-treatment was used. For this purpose, young *N. benthamiana* plants (at the five-leaf stage) were infiltrated with a mixture of *Agrobacterium tumefaciens* strains of pTRV1 (CaMV 35S-driven TRV RNA1) and pTRV2-SGT1 (TRV RNA2 containing the target sequence), or pTRV-GFPsil (serving as a control for infection symptoms). Two weeks after TRV inoculation efficacy of the silencing construct was assessed by RT-PCR (Supplemental Figure S3) and silenced plants and the control were infiltrated with *A. tumefaciens* containing XopJ-HA and sprayed with SA 24 hpi. When SA-treated leaves were inspected after an additional 24 h time period, XopJ infiltrated GFPsil control plants showed clear signs of tissues collapse and necrosis while SA-treatment of XopJ expressing SGT1-silenced plants did not lead to phenotypic alterations (Figure 2A), although immunoblot analysis revealed XopJ protein expression levels to be similar in both plants at 48 hpi (Figure 2B). Measurement of electrolyte leakage showed that SA-treatment of XopJ expressing leaves caused a significant increase in cell membrane disintegration in TRV:GFPsil plants

but not in TRV:SGT1 plants (**Figure 2C**), indicating that SGT1 silencing abrogates tissue damage under these conditions. These results demonstrate that XopJ requires SGT1 to elicit cell death in *N. benthamiana* upon SA-treatment. It is likely, therefore, that XopJ is recognized by a plant R-protein only after treatment of leaves with SA.

R proteins differ in their requirement for signaling components downstream of SGT1. NBS-LRR proteins with amino-terminal Toll and interleukin-1 receptor homology (TIR domain) use EDS1, whereas those with CC domains signal through NDR1 (Aarts et al., 1998). To provide first insights into the nature of a possible R protein involved in the recognition of XopJ after SA-treatment in *N. benthamiana*, VIGS directed against EDS1 and NDR1 was used (Supplementary Figure S3). Silencing of NDR1 resulted in a clear and consistent decrease in HR-like symptom development upon SA-treatment of XopJ infiltrated leaves, while plants with reduced EDS1 expression showed no apparent phenotypical differences when compared with the TRV:GFPsil control (**Figure 2A**). XopJ protein expression was confirmed by immunoblot analyses in EDS1 and NDR1 silenced leaves (**Figure 2B**). In accordance with previous findings (Üstün et al., 2012), we realized that XopJ protein levels are reproducibly lower in SA-treated *EDS1* silenced plants which might be due to an accelerated HR induction in these plants affecting the level of some proteins. Consistent with the observed phenotype, a significant decrease in ion leakage following SA-treatment of XopJ expressing leaves was evident in TRV:NDR1 plants but not in TRV:EDS1 and TRV:GFPsil plants, respectively (**Figure 2C**). This suggests that the R protein mediating the response to XopJ after SA-treatment is a member of the CC-NBS-LRR class.

Elicitation of HR-Like Symptoms is Dependent on RPT6

XopJ acts as a protease to degrade the proteasomal subunit RPT6 in host cells (Üstün and Börnke, 2015). This results in an inhibition of proteasomal activity which finally attenuates SA-dependent defense responses (Üstün et al., 2013). In order to investigate whether inhibition of proteasome activity *per se* is sufficient to elicit HR-like symptoms upon SA treatment or whether this effect requires the action of XopJ on RPT6, *N. benthamiana* leaves were treated with the potent proteasome inhibitor MG132 6 h before SA treatment. In contrast to XopJ expressing tissue, leaves pretreated with MG132 before SA application did not develop any visible signs of HR indicating that a general inhibition of the proteasome is not sufficient to elicit this response in *N. benthamiana* (**Figure 3A**). To assess the requirement of the XopJ virulence target RPT6 for elicitation of SA-dependent HR-like symptoms, *N. benthamiana* leaves transiently expressing XopJ and silenced for *RPT6* expression using VIGS (Supplementary Figure S3) were treated with SA. As shown in **Figure 3B**, the pTRV2-GFPsil control expressing XopJ showed typical signs of cell death while *RPT6* silenced plants did not develop visible symptoms. Similar levels of XopJ protein expression were observed in both types of VIGS plants (**Figure 3C**). These

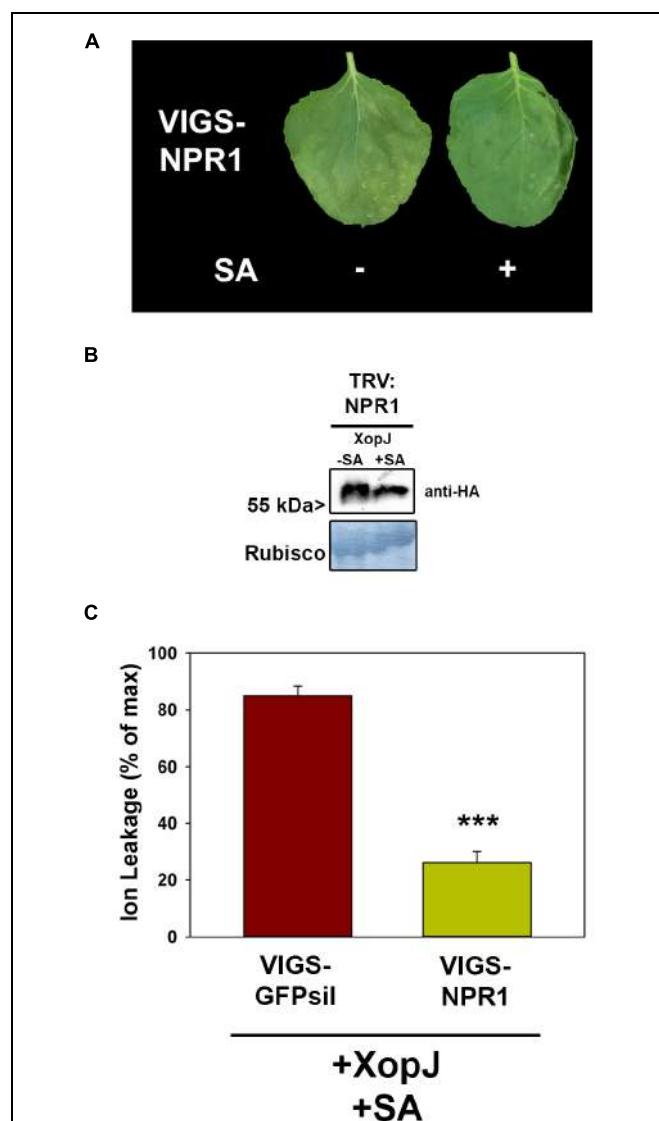


FIGURE 4 | VIGS of *NPR1* prevents XopJ induced HR-like cell death after SA treatment. **(A)** XopJ was transiently expressed in *NPR1* silenced plants and treated with 5 mM SA 24 hpi. Leaf phenotypes were photographed 48 hpi. **(B)** Protein extracts from TRV:NPR1 leaves transiently expressing XopJ-HA treated with or without 5 mM SA. Amido black staining served as a loading control. **(C)** Ion leakage was determined in TRV:NPR1 and TRV:GFPsil plants transiently expressing XopJ following SA treatment. Bars represent the average ion leakage measured for triplicates of six leaf155 disks each, and the error bars indicate SD. Significant differences are indicated by asterisks (**P < 0.001). The experiment was repeated three times with similar results.

finding lend support to the notion that elicitation of SA-dependent HR-like symptoms in XopJ expressing leaves requires interaction of the effector protein with its virulence target RPT6. In order to investigate whether SA treatment would affect the ability of XopJ to degrade RPT6, protein degradation was assessed in leaves co-expressing both proteins after treatment with SA versus the control (Supplementary Figure S4). However, no difference in RPT6 protein amount could be detected between SA treated leaves and the control. Thus,

SA treatment *per se* does not influence RPT6 degradation by XopJ.

The Role of NPR1

The experiments described above suggest a role of SA and SA-signaling in the phenomenon of SA-dependent HR-like symptom elicitation by XopJ. Thus, we sought to investigate the contribution of a SA signaling component to the process. NPR1 (non-expressor of PR1) is a key positive regulator of SA-mediated defense responses notably by activating transcription of a battery of genes in response to rising SA-levels (Pieterse and Van Loon, 2004). Down-regulation by VIGS in *N. benthamiana* plants was used to investigate the involvement of SA-signaling via NPR1 with GFPsil serving as a control. Two weeks after TRV inoculation leaves of silenced plants were infiltrated with *Agrobacteria* harboring XopJ-HA and 24 hpi infiltrated leaves were sprayed with SA. As shown in **Figure 4A**, NPR1 silenced plants did not develop visible signs of HR-like symptoms upon SA treatment of XopJ infiltrated leaves, suggesting a critical role of NPR1 in execution of this response. Western blot analysis showed that XopJ was expressed in VIGS-NPR1 leaves with or without SA-treatment (**Figure 4B**). Measurement of ion leakage confirmed the observed phenotype, as conductivity in NPR1 silenced plants was significantly lower compared to TRV:GFPsil control plants following SA-treatment (**Figure 4C**). Thus, NPR1 appears to be essential for XopJ to trigger HR-like symptoms upon SA-treatment.

XopJ's Inhibitory Effect on the Proteasome is Affected by SA-Treatment and Requires NPR1

XopJ has been shown to dampen proteasome activity during the compatible interaction of Xcv with pepper plants and this leads to a delay in the development of host cell necrosis (Üstün et al., 2013). Further analysis revealed that this effect was dependent on NPR1, as proteasome activity seems to be partially regulated by NPR1 during defense (Üstün et al., 2013). In the light of these observations, we next investigated whether the inhibitory effect on the proteasome is maintained when XopJ expressing plants are treated with SA and whether the associated changes are NPR1 dependent. To circumvent the negative effect of SA-mediated HR-like cell death on the overall proteasome function, we monitored proteasome activity 6 h after spraying *N. benthamiana* leaves expressing either EV or XopJ with SA, as previous results showed that the proteasome is activated upon SA treatment reaching a peak at 6 h after SA-treatment (Üstün et al., 2013). After transient expression of XopJ or EV and subsequent SA-treatment for 6 h, proteasome activity was measured in NPR1 silenced and GFPsil control plants. Spraying plants with SA led to a loss of XopJ's ability to inhibit the proteasome in control plants, whereas in NPR1 silenced plants XopJ was still able to suppress proteasome activity 6 h after SA-treatment (**Figure 5A**). To show that the SA-dependent activation of the proteasome function might be affected in NPR1 silenced plants, proteasome activity was determined in TRV:GFPsil and TRV:NPR1 treated with SA for 1 and 6 h, respectively. In accordance with previous findings, SA significantly elevated proteasome activity in GFPsil control plants but not in *NPR1* silenced plants (**Figure 5B**). These data

suggest that the ability of XopJ to interfere with the proteasome function can be counteracted by the exogenous application of SA.

Discussion

Depending on the genetic context of the plant with which a given bacterial pathovar interacts, T3Es can either act as virulence factors or, upon recognition by cognate R proteins, may function as Avr factors which then trigger a strong defense response typically characterized by an HR (Dodds and Rathjen, 2010). In this study, we investigated the Avr function of T3E XopJ from Xcv in the non-host plant *N. benthamiana* and demonstrated that the effector is able to trigger HR-like symptoms when transiently XopJ expressing tissue is treated with SA. We have previously characterized the virulence activity of XopJ and could show that it acts as a protease to degrade the proteasomal subunit RPT6 which subsequently leads to the inhibition of proteasome activity in host cells. Reduced proteasomal protein turnover interferes with SA-mediated defense responses as well as vesicle trafficking and attenuates host-induced necrosis during infection of susceptible pepper plants (Bartetzko et al., 2009; Üstün et al., 2013; Üstün and Börnke, 2015). Many of the defense responses that XopJ interferes with depend on the central SA-signaling component NPR1 and XopJ-mediated inhibition of the proteasome appears to interfere with proper NPR1 function (Üstün et al., 2013; Üstün and Börnke, 2015).

Like other members of the YopJ-family of effector proteins XopJ possess a catalytic triad that is required for its protease activity and is also essential for its virulence function (Üstün et al., 2013; Üstün and Börnke, 2015). Mutant studies revealed that the Avr activities of the YopJ-family members from plant pathogens depend on the catalytic triad, suggesting that the enzymatic function is required for the recognition by corresponding plant R proteins (Orth et al., 2000; Roden et al., 2004; Bonshtien et al., 2005; Whalen et al., 2008). The fact that the XopJ-induced, SA-dependent HR-like symptom development also requires the catalytic cysteine residue C235 indicates that XopJ is recognized indirectly via its enzymatic activity. This is similar to the *Pseudomonas syringae* T3Es AvrRpt2 and HopAR1 (formerly AvrPphB) which are recognized by their cognate R proteins in *Arabidopsis* via their protease activity (Axtell et al., 2003; Shao et al., 2003; Ade et al., 2007). The resistance protein RPS5 recognizes the proteolytic degradation of the HopAR1 target protein PBS1, while RPS2 is activated upon cleavage of the host target protein RIN4 by AvrRpt2. Silencing of the XopJ host target protein RPT6 prevents development of HR-like symptoms after SA-treatment further supporting the notion that XopJ is recognized indirectly via its proteolytic activity on its host target protein. Furthermore, the predicted myristylation site of XopJ that localizes the protein to the host cell plasma membrane is also required to induce an SA-dependent HR. Mutation of the myristylation site of XopJ has previously been shown to abolish its virulence function, indicating that plasma membrane localization inside the host cell is required for both activities. Thus, the data support a model in which XopJ's ability to elicit HR-like symptoms in *N. benthamiana* is closely linked to its

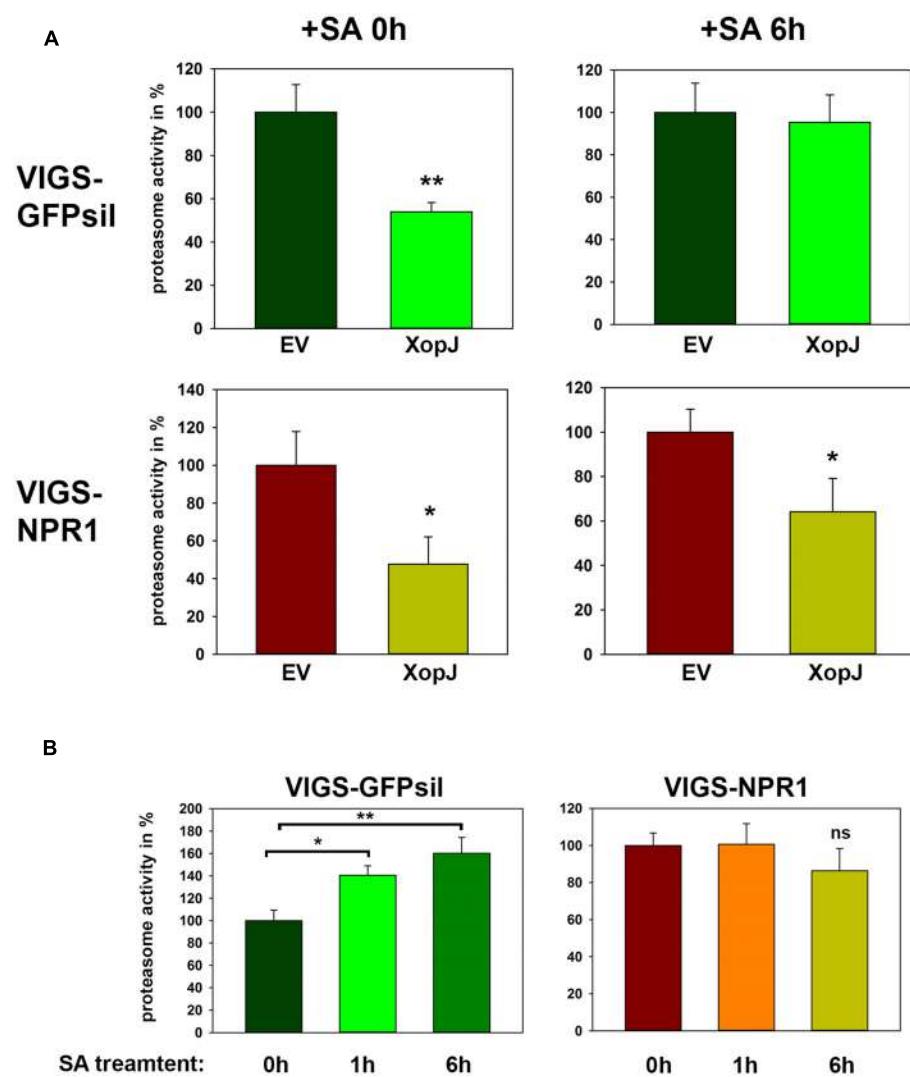
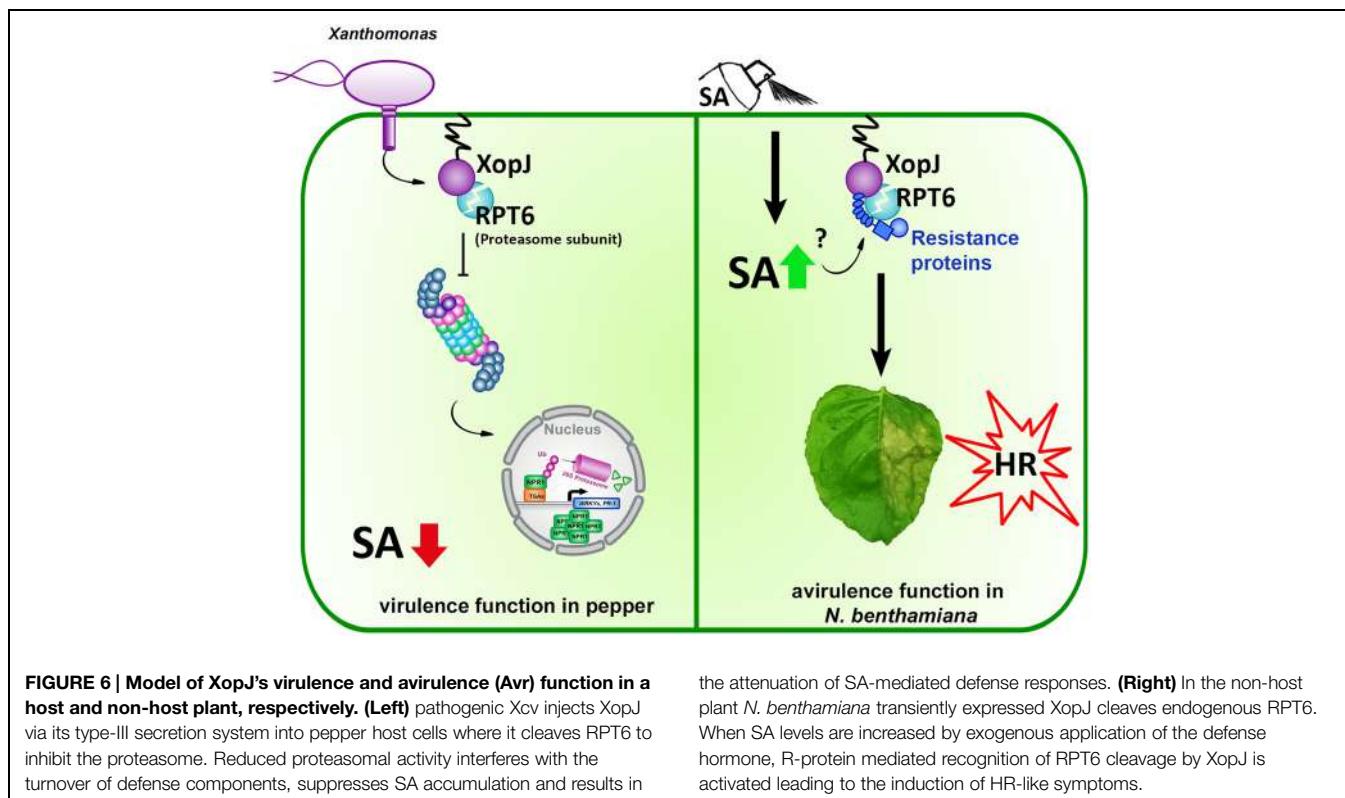


FIGURE 5 | SA treatment interferes with XopJ's ability to suppress proteasome activity in a NPR1-dependent manner. **(A)** XopJ along with an EV control were transiently expressed in leaves of TRV:GFPsil and TRV:NPR1 leaves using Agro-infiltration. After 48 h, leaves were sprayed with 5 mM SA and relative proteasome activity in total protein extracts was determined at 0 and 6 h post SA treatment by monitoring the breakdown of the fluorogenic peptide Suc-LLVY-AMC. The EV control was set to 100%. Data represent the mean SD ($n = 3$). The asterisk

indicates a significant difference ($*P < 0.05$; $**P < 0.01$) based on results of a Student's *t*-test. **(B)** VIGS NPR1 and GFPsil control leaves were sprayed with 5 mM SA and proteasome activity in total leaf extracts was determined at time points indicated in the figure by monitoring the breakdown of the fluorogenic peptide Suc-LLVY-AMC. Data represent the mean SD ($n = 3$). Significant differences are indicated by asterisks ($*P < 0.05$; $**P < 0.01$) and were calculated using Student's *t*-test. (ns = not significant). The experiment has repeated three times with similar results.

virulence target in host plants and in that RPT6 is guarded by a yet unknown R protein whose activation requires two signals (1) degradation of RPT6 and (2) activation of SA-signaling. Previous evidence suggests that RPT2a and RPT2b, two isoforms of another subunit of the proteasomal RP19, interact with the CC-NBS-LRR protein uni-1D from *Arabidopsis* and that this interaction is involved in triggering uni-1D-induced defense signaling (Chung and Tasaka, 2011). Hence, the uni-1D/RPT2 interaction provides an example of a proteasomal subunit that appears to be guarded by an R protein. Therefore, it could be well possible that the same is true for RPT6 in *N. benthamiana*.

XopJ-induced development of HR-like symptoms upon SA-treatment was dependent on SGT1, which is required for resistance mediated by multiple R proteins recognizing a diverse set of pathogens. SGT1 has been shown to control the steady-state level of preactivated R proteins (Peart et al., 2002; Azevedo et al., 2006). Virus-induced silencing of SGT1 in *N. benthamiana* considerably reduced HR-like symptom development in XopJ expressing leaves upon SA-treatment, suggesting the involvement of R-protein-mediated signaling in this process. The resistance protein responsible for the recognition of XopJ in *N. benthamiana* remains to be identified.



Based on the finding that silencing of NDR1 strongly reduces development of XopJ-mediated HR-like symptoms, it might be assumed that the R protein associated with these responses belongs to the CC-NBS-LRR class (Aarts et al., 1998), similar to what has been described for the RPT2/uni-1D couple in *Arabidopsis* (Chung and Tasaka, 2011).

Although a large body of evidence suggests a central role of SA and SA-signaling in the elicitation of cell death during HR (Vlot et al., 2009), it remains unclear why XopJ-triggered HR-like symptoms in *N. benthamiana* depend on the exogenous application of SA to XopJ expressing leaves. Exogenously applied SA appears to trigger the canonical SA-signaling pathway that operates via NPR1 as the central regulator. Plants silenced for NPR1 expression lose the ability to elicit HR-like symptoms upon SA treatment of XopJ expressing leaves. We could show that SA treatment induces proteasome activity in an NPR1 dependent manner and that in the presence of SA XopJ is no longer able to inhibit the proteasome. The reason for this phenomenon is currently unclear but a possible explanation could be that either the induction of proteasome activity by SA is quantitatively stronger than XopJ's ability for its inhibition or that SA can directly interfere with the ability of the effector to degrade RPT6. However, western blot experiments suggest that SA treatment has no effect on the ability of XopJ to degrade RPT6 in transient expression assays. Thus, activation of the proteasome beyond a certain threshold in the presence of a functional XopJ protein could act as a signal for R protein activation. HR-like symptom elicitation by XopJ might follow a two-signal model in which the first signal is the degradation of the host cell protein

RPT6 by XopJ and the second signal is provided by elevated SA levels. A similar model has previously been proposed for the *Pseudomonas syringae* T3Es AvrE and HopM1 (Lindeberg et al., 2012). According to this model AvrE or HopM1 trigger ETI by interfering with a process, e.g., vesicle trafficking, rather than with a specific protein. Plants then reduce spurious cell death responses resulting from vesicle trafficking perturbations by requiring a second signal such as increases in SA to eventually trigger ETI (Lindeberg et al., 2012). For XopJ this would mean that the sole inhibition of proteasomal turnover by removal of RPT6 would not be interpreted as a danger signal by the plant immune system but cell death is only triggered when there is a concomitant rise in SA contents. Alternatively, expression of the R protein required to recognize XopJ action on RPT6 could be dependent on SA as has previously been shown for the R proteins RPW8.1 and RPW8.2 in *Arabidopsis* which show induction on the transcriptional level after exogenous application of SA (Xiao et al., 2003).

Since we used transient expression of XopJ by *Agrobacterium*-infiltration we currently cannot make any statement about the role of XopJ in triggering HR-like symptoms during an incompatible interaction of Xcv with *N. benthamiana*. When inoculated with a high titer into *N. benthamiana* leaves Xcv has been shown to elicit plant cell death (Metz et al., 2005). The T3E responsible for this host response has been identified as XopX. Interestingly, the visual cell death response phenotype was not elicited by *Agrobacterium*-mediated expression of XopX. However, a cell death response could be elicited if the *Agrobacterium*-mediated XopX expression was co-inoculated

with XopX deficient Xcv or with *Xanthomonas campestris* pv. *campestris* that carry a functional T3SS (Metz et al., 2005). XopX has recently been proposed to interfere with PTI responses to promote Xcv virulence (Stork et al., 2015). Thus, XopX triggered cell death responses would also follow a two-signal model in which the second signal is dependent on a functional T3SS (Metz et al., 2005; Stork et al., 2015).

Conclusion

We could show that XopJ's ability to trigger an SA-dependent HR-like host response is tightly linked to its virulence function (Figure 6) and provide another example, in addition to the previously described T3Es XopX, AvrE, and HopM1, for an

effector following a two-signal model to elicit a defense response in plants.

Acknowledgment

This work received funding from the Deutsche Forschungsgemeinschaft (SFB 796: Reprogramming of host cells by microbial effectors TP C5 and BO 1916-5/1) and the IGZ.

Supplementary Material

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2015.00599>

References

- Aarts, N., Metz, M., Holub, E., Staskawicz, B. J., Daniels, M. J., and Parker, J. E. (1998). Different requirements for EDS1 and NDR1 by disease resistance genes define at least two R gene-mediated signaling pathways in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 95, 10306–10311. doi: 10.1073/pnas.95.17.10306
- Ade, J., DeYoung, B. J., Golstein, C., and Innes, R. W. (2007). Indirect activation of a plant nucleotide binding site-leucine-rich repeat protein by a bacterial protease. *Proc. Natl. Acad. Sci. U.S.A.* 104, 2531–2536. doi: 10.1073/pnas.0608779104
- Axtell, M. J., Chisholm, S. T., Dahlbeck, D., and Staskawicz, B. J. (2003). Genetic and molecular evidence that the *Pseudomonas syringae* type III effector protein AvrRpt2 is a cysteine protease. *Mol. Microbiol.* 49, 1537–1546. doi: 10.1046/j.1365-2958.2003.03666.x
- Azevedo, C., Betsuyaku, S., Peart, J., Takahashi, A., Noel, L., Sadanandom, A., et al. (2006). Role of SGT1 in resistance protein accumulation in plant immunity. *EMBO J.* 25, 2007–2016. doi: 10.1038/sj.emboj.7601084
- Bartetzko, V., Sonnewald, S., Vogel, F., Hartner, K., Stadler, R., Hammes, U. Z., et al. (2009). The *Xanthomonas campestris* pv. *vesicatoria* Type III effector protein XopJ inhibits protein secretion: evidence for interference with cell wall-associated defense responses. *Mol. Plant Microbe Interact.* 22, 655–664. doi: 10.1094/MPMI-22-6-0655
- Boller, T., and Felix, G. (2009). A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annu. Rev. Plant Biol.* 60, 379–406. doi: 10.1146/annurev.arplant.57.032905.105346
- Bonshoftien, A., Lev, A., Gibly, A., Debbie, P., Avni, A., and Sessa, G. (2005). Molecular properties of the *Xanthomonas* AvrRxt effector and global transcriptional changes determined by its expression in resistant tomato plants. *Mol. Plant Microbe Interact.* 18, 300–310. doi: 10.1094/MPMI-18-0300
- Chung, K., and Tasaka, M. (2011). RPT2a, a 26S proteasome AAA-ATPase, is directly involved in *Arabidopsis* CC-NBS-LRR protein uni-ID-induced signaling pathways. *Plant Cell Physiol.* 52, 1657–1664. doi: 10.1093/pcp/pcp099
- Dangl, J. L., and Jones, J. D. (2001). Plant pathogens and integrated defence responses to infection. *Nature* 411, 826–833. doi: 10.1038/35081161
- Delaney, T. P., Uknas, S., Vernooij, B., Friedrich, L., Weymann, K., Negrotto, D., et al. (1994). A central role of salicylic acid in plant disease resistance. *Science* 266, 1247–1250. doi: 10.1126/science.266.5188.1247
- Dodds, P. N., and Rathjen, J. P. (2010). Plant immunity: towards an integrated view of plant-pathogen interactions. *Nat. Rev. Genet.* 11, 539–548. doi: 10.1038/nrg2812
- El Oirdi, M., Abd El Rahman, T., Rigano, L., El Hadrami, A., Rodriguez, M. C., Daayf, F., et al. (2011). Botrytis cinerea manipulates the antagonistic effects between immune pathways to promote disease development in tomato. *Plant Cell* 23, 2405–2421. doi: 10.1105/tpc.111.083394
- Galan, J. E., Lara-Tejero, M., Marlovits, T. C., and Wagner, S. (2014). Bacterial type III secretion systems: specialized nanomachines for protein delivery into target cells. *Annu. Rev. Microbiol.* 68, 415–438. doi: 10.1146/annurev-micro-092412-155725
- Glazebrook, J. (2005). Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annu. Rev. Phytopathol.* 43, 205–227. doi: 10.1146/annurev.phyto.43.040204.135923
- Gurlebeck, D., Jahn, S., Gurlebeck, N., Szczesny, R., Szurek, B., Hahn, S., et al. (2009). Visualization of novel virulence activities of the *Xanthomonas* type III effectors AvrBs1, AvrBs3 and AvrBs4. *Mol. Plant Pathol.* 10, 175–188. doi: 10.1111/j.1364-3703.2008.00519.x
- Jones, J. D., and Dangl, J. L. (2006). The plant immune system. *Nature* 444, 323–329. doi: 10.1038/nature05286
- Kim, N. H., Choi, H. W., and Hwang, B. K. (2010). *Xanthomonas campestris* pv. *vesicatoria* effector AvrBsT induces cell death in pepper, but suppresses defense responses in tomato. *Mol. Plant Microbe Interact.* 23, 1069–1082. doi: 10.1094/MPMI-23-8-1069
- Kobayashi, D. Y., Tamaki, S. J., and Keen, N. T. (1989). Cloned avirulence genes from the tomato pathogen *Pseudomonas syringae* pv. *tomato* confer cultivar specificity on soybean. *Proc. Natl. Acad. Sci. U.S.A.* 86, 157–161. doi: 10.1073/pnas.86.1.157
- Lewis, J. D., Lee, A., Ma, W., Zhou, H., Guttman, D. S., and Desveaux, D. (2011). The YopJ superfamily in plant-associated bacteria. *Mol. Plant Pathol.* 12, 928–937. doi: 10.1111/j.1364-3703.2011.00719.x
- Lindeberg, M., Cunnac, S., and Collmer, A. (2012). *Pseudomonas syringae* type III effector repertoires: last words in endless arguments. *Trends Microbiol.* 20, 199–208. doi: 10.1016/j.tim.2012.01.003
- Liu, Y., Schiff, M., Marathe, R., and Dinesh-Kumar, S. P. (2002). Tobacco Rar1, EDS1 and NPR1/NIM1 like genes are required for N-mediated resistance to tobacco mosaic virus. *Plant J.* 30, 415–429. doi: 10.1046/j.1365-313X.2002.01297.x
- Macho, A. P., and Zipfel, C. (2015). Targeting of plant pattern recognition receptor-triggered immunity by bacterial type-III secretion system effectors. *Curr. Opin. Microbiol.* 23, 14–22. doi: 10.1016/j.mib.2014.10.009
- Mackey, D., Belkhadir, Y., Alonso, J. M., Ecker, J. R., and Dangl, J. L. (2003). *Arabidopsis* RIN4 is a target of the type III virulence effector AvrRpt2 and modulates RPS2-mediated resistance. *Cell* 112, 379–389. doi: 10.1016/S0092-8674(03)00040-0
- Metz, M., Dahlbeck, D., Morales, C. Q., Al Sady, B., Clark, E. T., and Staskawicz, B. J. (2005). The conserved *Xanthomonas campestris* pv. *vesicatoria* effector protein XopX is a virulence factor and suppresses host defense in *Nicotiana benthamiana*. *Plant J.* 41, 801–814. doi: 10.1111/j.1365-313X.2005.02338.x
- Morel, J. B., and Dangl, J. L. (1997). The hypersensitive response and the induction of cell death in plants. *Cell Death Differ.* 4, 671–683. doi: 10.1038/sj.cdd.4400309
- Orth, K., Xu, Z., Mudgett, M. B., Bao, Z. Q., Palmer, L. E., Bliska, J. B., et al. (2000). Disruption of signaling by *Yersinia* effector YopJ, a ubiquitin-like protein protease. *Science* 290, 1594–1597. doi: 10.1126/science.290.5496.1594
- Peart, J. R., Lu, R., Sadanandom, A., Malcuit, I., Moffett, P., Brice, D. C., et al. (2002). Ubiquitin ligase-associated protein SGT1 is required for host and

- nonhost disease resistance in plants. *Proc. Natl. Acad. Sci. U.S.A.* 99, 10865–10869. doi: 10.1073/pnas.152330599
- Pieterse, C. M., and Van Loon, L. C. (2004). NPR1: the spider in the web of induced resistance signaling pathways. *Curr. Opin. Plant Biol.* 7, 456–464. doi: 10.1016/j.pbi.2004.05.006
- Rairdan, G. J., and Delaney, T. P. (2002). Role of salicylic acid and NIM1/NPR1 in race-specific resistance in *Arabidopsis*. *Genetics* 161, 803–811.
- Roden, J., Eardley, L., Hotson, A., Cao, Y., and Mudgett, M. B. (2004). Characterization of the *Xanthomonas AvrXv4* effector, a SUMO protease translocated into plant cells. *Mol. Plant Microbe Interact.* 17, 633–643. doi: 10.1094/MPMI.2004.17.6.633
- Ryan, R. P., Vorholter, F. J., Potnis, N., Jones, J. B., Van Sluys, M. A., Bogdanove, A. J., et al. (2011). Pathogenomics of *Xanthomonas*: understanding bacterium-plant interactions. *Nat. Rev. Microbiol.* 9, 344–355. doi: 10.1038/nrmicro2558
- Schulze, S., Kay, S., Büttner, D., Egler, M., Eschen-Lippold, L., Hause, G., et al. (2012). Analysis of new type III effectors from *Xanthomonas* uncovers XopB and XopS as suppressors of plant immunity. *New Phytol.* 195, 894–911. doi: 10.1111/j.1469-8137.2012.04210.x
- Senthil-Kumar, M., and Mysore, K. S. (2013). Nonhost resistance against bacterial pathogens: retrospectives and prospects. *Annu. Rev. Phytopathol.* 51, 407–427. doi: 10.1146/annurev-phyto-082712-102319
- Shao, F., Golstein, C., Ade, J., Stoutemyer, M., Dixon, J. E., and Innes, R. W. (2003). Cleavage of *Arabidopsis PBS1* by a bacterial type III effector. *Science* 301, 1230–1233. doi: 10.1126/science.1085671
- Singer, A. U., Schulze, S., Skarina, T., Xu, X., Cui, H., Eschen-Lippold, L., et al. (2013). A pathogen type III effector with a novel E3 ubiquitin ligase architecture. *PLoS Pathog.* 9:e1003121. doi: 10.1371/journal.ppat.1003121
- Staskawicz, B., Dahlbeck, D., Keen, N., and Napoli, C. (1987). Molecular characterization of cloned avirulence genes from race 0 and race 1 of *Pseudomonas syringae* pv. *glycinea*. *J. Bacteriol.* 169, 5789–5794.
- Stork, W., Kim, J. G., and Mudgett, M. B. (2015). Functional analysis of plant defense suppression and activation by the *Xanthomonas* core type III effector XopX. *Mol. Plant Microbe Interact.* 28, 180–194. doi: 10.1094/MPMI-09-14-0263-R
- Szczesny, R., Büttner, D., Escobar, L., Schulze, S., Seifert, A., and Bonas, U. (2012). Suppression of the AvrBs1-specific hypersensitive response by the YopJ effector homolog AvrBsT from *Xanthomonas* depends on a SNF1-related kinase. *New Phytol.* 187, 1058–1074. doi: 10.1111/j.1469-8137.2010.03346.x
- Thieme, F., Szczesny, R., Urban, A., Kirchner, O., Hause, G., and Bonas, U. (2007). New type III effectors from *Xanthomonas campestris* pv. *vesicatoria* trigger plant reactions dependent on a conserved N-myristylation motif. *Mol. Plant Microbe Interact.* 20, 1250–1261. doi: 10.1094/MPMI-20-10-1250
- Üstün, S., Bartetzko, V., and Börnke, F. (2013). The *Xanthomonas campestris* type III effector XopJ targets the host cell proteasome to suppress salicylic-acid mediated plant defence. *PLoS Pathog.* 9:e1003427. doi: 10.1371/journal.ppat.1003427
- Üstün, S., and Börnke, F. (2015). The *Xanthomonas campestris* type III effector XopJ proteolytically degrades proteasome subunit RPT6. *Plant Physiol.* 168, 107–119. doi: 10.1104/pp.15.00132
- Üstün, S., König, P., Guttman, D. S., and Börnke, F. (2014). HopZ4 from *Pseudomonas syringae*, a member of the HopZ type III effector family from the YopJ superfamily, inhibits the proteasome in plants. *Mol. Plant Microbe Interact.* 27, 611–623. doi: 10.1094/MPMI-12-13-0363-R
- Üstün, S., Müller, P., Palmisano, R., Hensel, M., and Börnke, F. (2012). SseF, a type III effector protein from the mammalian pathogen *Salmonella enterica*, requires resistance-gene-mediated signalling to activate cell death in the model plant *Nicotiana benthamiana*. *New Phytol.* 194, 1046–1060. doi: 10.1111/j.1469-8137.2012.04124.x
- Vlot, A. C., Dempsey, D. A., and Klessig, D. F. (2009). Salicylic acid, a multifaceted hormone to combat disease. *Annu. Rev. Phytopathol.* 47, 177–206. doi: 10.1146/annurev.phyto.050908.135202
- Wei, C. F., Kvirklo, B. H., Shimizu, R., Crabill, E., Alfano, J. R., Lin, N. C., et al. (2007). A *Pseudomonas syringae* pv. *tomato* DC3000 mutant lacking the type III effector HopQ1-1 is able to cause disease in the model plant *Nicotiana benthamiana*. *Plant J.* 51, 32–46. doi: 10.1111/j.1365-313X.2007.03126.x
- Whalen, M., Richter, T., Zakharevich, K., Yoshikawa, M., Al-Azzeh, D., Adefioye, A., et al. (2008). Identification of a host 14-3-3 Protein that interacts with *Xanthomonas* effector AvrXv. *Physiol. Mol. Plant Pathol.* 72, 46–55. doi: 10.1016/j.pmp.2008.05.006
- Whalen, M. C., Wang, J. F., Carland, F. M., Heiskell, M. E., Dahlbeck, D., Minsavage, G. V., et al. (1993). Avirulence gene *avrXv* from *Xanthomonas campestris* pv. *vesicatoria* specifies resistance on tomato line Hawaii 7998. *Mol. Plant Microbe Interact.* 6, 616–627. doi: 10.1094/MPMI-6-616
- White, F. F., Potnis, N., Jones, J. B., and Koebnik, R. (2009). The type III effectors of *Xanthomonas*. *Mol. Plant Pathol.* 10, 749–766. doi: 10.1111/j.1364-3703.2009.00590.x
- Wroblewski, T., Caldwell, K. S., Piskurewicz, U., Cavanaugh, K. A., Xu, H., Kozik, A., et al. (2009). Comparative large-scale analysis of interactions between several crop species and the effector repertoires from multiple pathovars of *Pseudomonas* and *Ralstonia*. *Plant Physiol.* 150, 1733–1749. doi: 10.1104/pp.109.140251
- Xiao, S., Brown, S., Patrick, E., Brearley, C., and Turner, J. G. (2003). Enhanced transcription of the *Arabidopsis* disease resistance genes RPW8.1 and RPW8.2 via a salicylic acid-dependent amplification circuit is required for hypersensitive cell death. *Plant Cell* 15, 33–45. doi: 10.1105/tpc.006940

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 Üstün, Bartetzko and Börnke. 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.

TAL effectors and activation of predicted host targets distinguish Asian from African strains of the rice pathogen *Xanthomonas oryzae* pv. *oryzicola* while strict conservation suggests universal importance of five TAL effectors

OPEN ACCESS

Edited by:

Laurent D. Noël,
Centre National de la Recherche
Scientifique, France

Reviewed by:

Boris Szurek,
Institut de Recherche pour le
Développement, France
Thomas Lahaye,
Ludwig-Maximilians-University
Munich, Germany

***Correspondence:**

Adam J. Bogdanove,
Plant Pathology and Plant-Microbe
Biology Section, School of Integrative
Plant Science, Cornell University, 334
Plant Science Building, Ithaca,
NY 14853, USA
ajb7@cornell.edu

Specialty section:

This article was submitted to
Plant Biotic Interactions,
a section of the journal
Frontiers in Plant Science

Received: 15 April 2015

Accepted: 30 June 2015

Published: 21 July 2015

Citation:

Wilkins KE, Booher NJ, Wang L and
Bogdanove AJ (2015) TAL effectors
and activation of predicted host
targets distinguish Asian from African
strains of the rice pathogen
Xanthomonas oryzae pv. *oryzicola*
while strict conservation suggests
universal importance of five TAL
effectors. *Front. Plant Sci.* 6:536.
doi: 10.3389/fpls.2015.00536

Katherine E. Wilkins^{1,2}, Nicholas J. Booher^{1,2}, Li Wang¹ and Adam J. Bogdanove^{1*}

¹ Plant Pathology and Plant-Microbe Biology Section, School of Integrative Plant Science, Cornell University, Ithaca, NY, USA

² Graduate Field of Computational Biology, Cornell University, Ithaca, NY, USA

Xanthomonas oryzae pv. *oryzicola* (Xoc) causes the increasingly important disease bacterial leaf streak of rice (BLS) in part by type III delivery of repeat-rich transcription activator-like (TAL) effectors to upregulate host susceptibility genes. By pathogen whole genome, single molecule, real-time sequencing and host RNA sequencing, we compared TAL effector content and rice transcriptional responses across 10 geographically diverse Xoc strains. TAL effector content is surprisingly conserved overall, yet distinguishes Asian from African isolates. Five TAL effectors are conserved across all strains. In a prior laboratory assay in rice cv. Nipponbare, only two contributed to virulence in strain BLS256 but the strict conservation indicates all five may be important, in different rice genotypes or in the field. Concatenated and aligned, TAL effector content across strains largely reflects relationships based on housekeeping genes, suggesting predominantly vertical transmission. Rice transcriptional responses did not reflect these relationships, and on average, only 28% of genes upregulated and 22% of genes downregulated by a strain are up- and down- regulated (respectively) by all strains. However, when only known TAL effector targets were considered, the relationships resembled those of the TAL effectors. Toward identifying new targets, we used the TAL effector-DNA recognition code to predict effector binding elements in promoters of genes upregulated by each strain, but found that for every strain, all upregulated genes had at least one. Filtering with a classifier we developed previously decreases the number of predicted binding elements across the genome, suggesting that it may reduce false positives among upregulated genes. Applying this filter and eliminating genes for which upregulation did not strictly correlate with presence of the corresponding TAL effector, we generated testable numbers of candidate targets for four of the five strictly conserved TAL effectors.

Keywords: transcription activator-like (TAL) effector, single molecule real-time (SMRT) sequencing, RNA-Seq, horizontal gene transfer, plant disease resistance, susceptibility (S) genes, population genomics

Introduction

Plant pathogenic *Xanthomonas* spp. inject transcription activator-like (TAL) effectors into host cells, where these proteins activate genes by binding at promoters (Kay et al., 2007; Römer et al., 2010). A TAL effector may activate a susceptibility (*S*) gene that contributes to disease or may activate a resistance (*R*) gene that results in host defense (Boch et al., 2014). In some cases, TAL effectors also or instead activate genes that appear to be inconsequential, “collateral” targets (Li et al., 2012b; Cernadas et al., 2014; Hu et al., 2014). TAL effector targeting is mediated by a modular central repeat region (CRR), with each repeat recognizing one nucleotide according to a degenerate code resulting in specific binding to a contiguous DNA sequence called the effector-binding element (EBE) (Boch et al., 2009; Moscou and Bogdanove, 2009). Nucleotide specificity of each repeat is determined by the hypervariable positions 12 and 13, together known as the repeat variable diresidue (RVD) (Boch et al., 2009; Moscou and Bogdanove, 2009). Crystal structures showed that only the thirteenth residue interacts with the nucleotide, while the twelfth stabilizes the loop that projects the thirteenth residue into the major groove (Deng et al., 2012; Mak et al., 2012). And, while some amino acids at the twelfth position abolish or dramatically reduce binding affinity, RVDs that share the same thirteenth residue typically have similar binding specificities (Boch et al., 2009; Moscou and Bogdanove, 2009; Yang et al., 2014). The thirteenth residue has therefore been referred to as the base-specifying residue (BSR) (De Lange et al., 2014).

Single TAL effectors often determine the outcome of the host-pathogen interaction (Boch et al., 2014). This is clear for TAL effectors that act as avirulence factors by activating a corresponding *R* gene that blocks disease progression (Gu et al., 2005; Römer et al., 2007; Tian et al., 2014), but it is also true for any TAL effector that upregulates an *S* gene that plays a critical role in disease. An example is TAL effector PthXo1 of the bacterial blight of rice pathogen *X. oryzae* pv. *oryzae* (Xoo) strain PXO99^A, which upregulates the *S* gene *OsSWEET11* (also *Os8N3*; Yang et al., 2006), a member of a large family of paralogous sugar transporters. *OsSWEET11* alleles that lack the PthXo1 EBE confer (genetically recessive) resistance (Chu et al., 2006; Yang et al., 2006; Yuan et al., 2009). This resistance can be overcome by strains harboring distinct TAL effectors however (Antony et al., 2010). Indeed, five distinct TAL effectors from different Xoo strains each activate one of five phylogenetically close members of the *SWEET* gene family that function interchangeably as *S* genes (Yang et al., 2006; Antony et al., 2010; Römer et al., 2010; Yu et al., 2011; Streubel et al., 2013; Richter et al., 2014). Similarly, in citrus canker caused by *X. axonopodis* pv. *citri* (divided into three types) and *X. axonopodis* pv. *aurantifoliae* (two types), representative strains of all five types contain one unique TAL effector each that activates the *S* gene *CsLOB1*, which is both necessary and sufficient for the formation of the pustules that are a hallmark of this disease (Hu et al., 2014; Pereira et al., 2014).

In light of their determinative nature, knowledge of TAL effectors and their targets enables innovative strategies for disease control. Removal of EBEs from *OsSWEET* gene promoters by

genome editing resulted in resistance to a diverse collection of Xoo strains (Li et al., 2012a). Every discovery of a novel *S* gene creates new opportunities to engineer host plants with resistance in a similar way. TAL effector-targeted *R* genes can be exploited as well. By engineering an *R* gene promoter to include EBEs for multiple additional TAL effectors, it is possible to trap a broad spectrum of strains and even different pathogens. This strategy applied to the *Xa27* rice bacterial blight *R* gene yielded plants resistant to strains of Xoo lacking AvrXa27 and to each of 10 tested strains of *X. oryzae* pv. *oryzicola* (Xoc), which causes the distinct disease bacterial leaf streak of rice (Hummel et al., 2012). More recently, the strategy was shown to be effective with another bacterial blight *R* gene, *Xa10* (Zeng et al., 2015). Broad effectiveness and durability of these *S* and *R* gene-centered strategies depends, however, on knowledge of TAL effector content within and across populations of the pathogen.

Bacterial blight and bacterial leaf streak of rice constrain production of this staple crop in many parts of the world (Nino-Liu et al., 2006). Four Xoo strains have been fully sequenced, KACC10331, MAFF311018, PXO99^A, and PXO86. These harbor 11, 16, 18, and 18 TAL effectors, respectively. Five bacterial blight *S* genes targeted by eight TAL effectors and three bacterial blight *R* genes each targeted by a unique TAL effector have been identified (reviewed in Boch et al., 2014). By comparison, Xoc is less well-characterized, with only two fully sequenced strains, BLS256 and CFBP7342, which secrete 28 and 24 TAL effectors, respectively (Bogdanove et al., 2011; Booher et al., unpublished), only one identified TAL effector-*S* gene pair (*Tal2g* of BLS256 and the rice sulfate transporter gene *OsSULTR3;6*), and no known TAL effector-activated *R* genes (Cernadas et al., 2014). In a growth chamber assay of TAL effector gene knockout mutants of BLS256, only two showed consistently reduced virulence (including the *tal2g* knockout) (Cernadas et al., 2014), but it is unclear how sensitive this assay is to differences that might be observed in the field. Furthermore, the assay was carried out on only a single rice genotype (cv. Nipponbare) and might have missed other host genotype-specific virulence contributions of TAL effectors. Generating a comprehensive TAL effector gene knockout library is technically demanding and time consuming, and CFBP7342 TAL effector knockouts have not yet been made and tested.

Sequencing of multiple, diverse Xoc strains is a useful alternative approach to identifying uniquely and broadly important TAL effectors, i.e., those that are highly conserved. TAL effectors that perform redundant functions in disease, or no function at all other than perhaps as material for rapid evolution of new TAL effectors via recombination, are likely to be less conserved. In addition to highlighting TAL effectors most likely to function in virulence, TAL effector conservation would inform the design of *R* gene traps, since the inclusion of EBEs bound by highly conserved TAL effectors would mediate broad resistance. In conjunction with transcript profiling of rice inoculated with each of multiple strains, TAL effector sequences are an enabling resource for identifying candidate targets, using the TAL effector DNA binding code for EBE prediction (e.g., Doyle et al., 2012). In particular, TAL effector targets that are activated by many strains even in the absence of TAL effector conservation could represent

S genes to which convergent evolution has resulted in multiple corresponding TAL effectors, or conversely, to which diversifying selection by an *R* gene has resulted in several related but distinct corresponding TAL effectors.

Here, we report comparative analyses of whole genome sequences of 10 Xoc strains from diverse locations (including BLS256, CFBP7342, and eight more sequenced for this study) and of rice transcriptional responses to each strain, with a focus on TAL effectors and their known and candidate targets. Our results allow inference regarding the importance of particular TAL effectors and targets prior to experimentation, and the potential of TAL-effector-centered strategies for durable resistance to bacterial leaf streak.

Materials and Methods

Xoc Strains and Genome Sequences

Xoc strains used in this study are given in **Table 1**. Genome sequences and raw sequence data are available under NCBI BioProject numbers PRJNA280380 and PRJNA283315, respectively. Raw data in.bas.h5/.bax.h5 format are available on request.

Xoc Genome Sequencing and Assembly and TAL Effector Sequence Parsing

Genomes were sequenced using single molecule, real-time (SMRT) technology (Eid et al., 2009) and assembled *de novo* using HGAP (Chin et al., 2013) v3.0 (Pacific Biosciences, Menlo Park, CA) and the PBX toolkit, as described (Booher et al., unpublished). For each strain, 4–7 SMRT cells were used to achieve ~200× coverage. All cells used the XL-C2 chemistry, except for two cells for strain B8-12, which used the P4-C2 chemistry. For RS105 and BXOR1 the large TAL effector gene

cluster from the assembly generated using the PBX toolkit was used in place of the cluster as assembled by HGAP, because HGAP partially collapsed (RS105) or expanded (BXOR1) the cluster. This was detected because very long reads mapping to this region indicated the presence of additional (RS105) or fewer (BXOR1) TAL effector genes in the cluster, matching the PBX toolkit assembly. TAL effector sequences were extracted and parsed using the PBX toolkit (Booher et al., unpublished). The genomic coordinates of all TAL effector genes in each strain and the corresponding RVD sequences are given in tab-delimited text in Supplementary Materials, File S1 (Data Sheet 1).

The existence of any small plasmids that might have been excluded during size selection for the PacBio sequencing was tested by DNA isolation and agarose gel electrophoresis, using *X. campestris* pv. *vesicatoria* strain 85-10 (Thieme et al., 2005) as a positive control, and none were observed (Supplementary Figure S1 in Data Sheet 2). This was carried out as described (Chakrabarty et al., 2010) except that the E.Z.N.A. Plasmid DNA Mini Kit I (Omega Bio-Tek, Norcross, GA) was used for DNA isolation.

Rice Inoculations and RNA Sequencing

Oryza sativa L. ssp. *japonica* cv. Nipponbare plants to be inoculated were grown in LC-1 soil mixture (Sungro, Bellevue, WA) in PGC15 growth chambers (Percival Scientific, Perry, IA) in trays approximately 60 cm below a combination of fluorescent and incandescent bulbs providing approximately 1000 µmoles/m²/s measured at 15 cm, under a cycle of 12 h of light at 28°C and 12 h of dark at 25°C. Plants were inoculated at 2 weeks with bacterial suspensions in 10 mM MgCl₂ at approximately OD₆₀₀ = 0.4 or with a mock inoculum of 10 mM MgCl₂, by infiltration using a needless syringe. For each inoculum, the second and third leaves of each of four plants were infiltrated at 20 contiguous spots per leaf. From each of the eight leaves inoculated with mock inoculum or a single strain, a 12 cm length of the inoculated portion was collected after 48 h and the tissue from all eight leaves pooled together for RNA isolation. For this, RNA was extracted using a modified hot Trizol protocol (Huang et al., 2005) followed by treatment with RNase-free DNase I (Life Technologies, Carlsbad, CA), then purified using the RNeasy MinElute Cleanup kit (Qiagen, Valencia, CA). This experiment was repeated three times for a total of 36 samples. Libraries were prepared with the TruSeq RNA Sample Prep v2 kit (Illumina, San Diego, CA). To perform RNA sequencing (RNA-Seq), samples were indexed and run six per lane on an Illumina Hiseq 2000 following the protocol supplied by the manufacturer for a single-end, 100 cycle run, producing 862.6 million reads across all samples.

RNA-Seq Data Processing, Analysis, and Access

Adapters were trimmed from raw reads using the Trimmomatic v0.22 ILLUMINACLIP function (Bolger et al., 2014) for single end reads, with a maximum of two seed mismatches and a palindrome clip threshold of 30 to trim adapter matches with scores of at least 7. Low quality bases were then trimmed from the ends of reads using BRAT trim v2.0.1 (Harris et al., 2010) to remove bases with Phred scores below 20 (Yu et al., 2012).

TABLE 1 | Xoc strains used in this study.

| Strain | Country of isolation | Year of isolation | Source or references |
|--------------------------|----------------------|-------------------|--|
| BLS256 | Philippines | 1984 | Vera Cruz ^a |
| BLS279 | Philippines | 1984 | Vera Cruz ^a |
| CFBP2286 | Malaysia | 1964 | CIRM-CFBP ^b |
| RS105 | China | 1992 | Fu and Xu, 1997; Song ^c , personal communication |
| B8-12 | China | 2007 | Song ^c |
| L8 | China | Prior to 1995 | Zeng et al., 1995; Yin ^d , personal communication |
| BXOR1 | India | 1996 | Yashitola et al., 2000 |
| CFBP7331 (also MAI10) | Mali | 2003 | Gonzalez et al., 2007 |
| CFBP7341 (also BAI5) | Burkina | 2009 | Wonni et al., 2011 |
| CFBP7342 (also BAI11) | Burkina | 2009 | Wonni et al., 2014 |

^a International Rice Research Institute, Los Baños, the Philippines.

^b International Centre for Microbial Resources – French Collection of Plant-associated Bacteria.

^c Nanjing Agricultural University, Nanjing, China.

^d Temasek Laboratories, Singapore.

Reads that were shorter than 24 base pairs after trimming were dropped. The 852.2 million remaining reads were aligned to the *Oryza sativa* L. ssp. *japonica* (cv. Nipponbare) reference genome (v.7.0) downloaded from Phytozyme 10 using the MSU Release 7.0 annotation downloaded from the Rice Genome Annotation Project (Kawahara et al., 2013). Alignment was completed using Tophat v2.06 (Kim et al., 2013) except for the reads from one BLS279 replicate. That sample caused Tophat v2.06 to crash and so was instead aligned using Tophat v2.05 (Kim et al., 2013). Transcripts were assembled using Cufflinks v2.1.1 and the same reference annotation used for the alignment (Trapnell et al., 2010; Roberts et al., 2011), and the resulting annotation files were combined using the Cufflinks script Cuffmerge v1.0.0. The number of reads aligned to each gene was determined using HTSeq-count (Anders et al., 2015) from the HTSeq framework version 0.5.4p3 with the command line options “-i ID -t gene -s no,” which unambiguously assigned 739.2 million reads (86% of the raw reads) to a gene. Raw reads and read counts are available through the NCBI Gene Expression Omnibus with accession number GSE67588. Differentially expressed genes were identified using QuasiSeq (Lund et al., 2012) with the quartile normalization method (Bullard et al., 2010). The replicate number was used as a cofactor, and genes were filtered if they did not have an average gene expression of at least 1 within every replicate or did not have at least one read in one control sample or in all non-control samples. After these filtering steps, 26,517 genes remained and were tested for differential expression, using the method of Nettleton et al. (2006) to compute *q*-values from *p*-values. All genes with *q*-values less than or equal to 0.05 were taken as being differentially expressed, regardless of fold change.

Phylogenetic Tree Construction and Topology Comparison

The nucleotide sequences of 31 housekeeping genes were extracted from each genome using AMPHORA (Wu and Eisen, 2008), and then aligned at the codon level in MEGA v6.0 (Tamura et al., 2013) with the MUSCLE alignment algorithm (Edgar, 2004) using default parameters and retaining gaps. The model of nucleotide substitution that best fit these sequences was identified with jModelTest v2.1.7 using the Bayesian information criterion (BIC) and separately the Akaike information criterion (AIC) (Guindon and Gascuel, 2003; Darriba et al., 2012). The best model tested using the AIC was the general time reversible model with invariant sites (GTR + I). The MEGA-estimated proportion of invariant sites was zero, so this reduced to a GTR model. Based on the BIC, the best model tested was the Hasegawa-Kishino-Yano model with invariant sites (HKY + I). Each model was used in MEGA v6.0 with 1000 rounds of bootstrapping to create a maximum-likelihood phylogenetic tree. Since the topologies of the two trees were identical, including bootstrap values, arbitrarily the GTR + I tree was retained.

Orthologous pairs of TAL effectors were identified using the reciprocal best BLAST hits method, ranking BLASTP (Altschul et al., 1997) results by bit score and breaking ties by *E*-value (Moreno-Hagelsieb and Latimer, 2008). BLASTP was run using default parameters. The protein sequences of all TAL effectors as well as the protein sequences of all other genes in the genomes,

annotated using RAST (Aziz et al., 2008; Overbeek et al., 2014) and excluding any encoded by reading frames overlapping the TAL effector genes, were included in the search for reciprocal best hits. For every TAL effector, a group containing all orthologs of that TAL effector was created. TAL effectors orthologous to half or fewer of the other TAL effectors in a group were removed from that group, and any TAL effector that was still in two groups after this filtering step was removed from the group in which it had fewer orthologs. This resulted in 39 non-overlapping groups of orthologous TAL effectors. For each strain, corresponding nucleotide sequences were concatenated, and then concatenated sequences were aligned in MEGA v6.0 (Tamura et al., 2013) using the MUSCLE alignment algorithm (Edgar, 2004) using default parameters and retaining gaps. The best model of nucleotide substitution was identified with jModelTest v2.1.7 (Guindon and Gascuel, 2003; Darriba et al., 2012) as above. In this case, the best model, based on either the BIC or the AIC, was the GTR + I model. The MEGA-estimated proportion of invariant sites was zero, so this reduced to a GTR model. A maximum likelihood phylogenetic tree was then created in MEGA using this model with 1000 rounds of bootstrapping. Next, for each ortholog group individually, a phylogenetic tree based on the alignment of the TAL effectors in that group was created in the same way as the tree based on all groups and, for comparison, a phylogenetic tree was created for the strains in each group based on the 31 housekeeping genes in the same way as the housekeeping gene tree for all 10 strains. For each group and for the concatenated alignment of all TAL effector ortholog groups, the one-sided Kishino-Hasegawa test based on pairwise Shimodaira-Hasegawa tests implemented in TREE-PUZZLE (Kishino and Hasegawa, 1989; Shimodaira and Hasegawa, 1999; Goldman et al., 2000; Schmidt et al., 2002) was used to test the null hypothesis that the topology of the maximum likelihood tree created using the TAL effector alignment was not a significantly better fit for those sequence data than the corresponding housekeeping gene tree topology, based on the log likelihood of each topology. Tree topology comparisons were not meaningful for eight ortholog groups containing only two TAL effectors each or for one group containing only six TAL effector sequences that were identical at the nucleotide level.

The tree based on the rice gene expression changes in response to each strain was made using \log_{10} fold change values of the 20,136 genes that had enough reads to be tested for differential expression and were differentially expressed in at least one sample. The tree was built using the function heatmap 0.2 from the R package gplots (Warnes et al., 2015), which uses the standard R function hclust from the stats package (R Core Team, 2013) to perform complete linkage clustering to cluster rows and columns of the heatmap based on the Euclidean distance between them. The tree created using only the known BLS256 TAL effector targets was created in the same way.

Whole Genome Alignments and Recombination Breakpoint Identification

Whole genome alignments were created in MAUVE using the progressiveMauve algorithm with default parameters (Darling et al., 2010). A codon based alignment of TAL effector genes

created using the MUSCLE alignment algorithm (Edgar, 2004) in MEGA v6.0 (Tamura et al., 2013) was searched for recombination breakpoints using GARD (Kosakovsky Pond et al., 2006a,b) on the Datammonkey webserver (Pond and Frost, 2005; Pond et al., 2005; Delport et al., 2010). The GTR model (identified as REV in the GARD menu) was used for this analysis with default settings because this was the best model of nucleotide substitution based on both the BIC and AIC, identified using jModelTest v2.1.7 (Guindon and Gascuel, 2003; Darriba et al., 2012).

TAL EBE Prediction

After TAL effector sequences were extracted from the 10 Xoc genomes, all unique RVD sequences were identified. For every TAL effector containing aberrant repeats of the lengths shown by Richter et al. (2014) to accommodate a single nucleotide deletion at the corresponding position in the DNA, a new TAL effector was artificially added to the list with that repeat missing from the RVD sequence. The original TAL effector was not removed. No TAL effectors with multiple aberrant repeats were observed. TAL effector repeats of atypical lengths other than those characterized by Richter et al. were treated as normal repeats. TAL effectors with fewer than 11 RVDs, which are likely non-functional (Boch et al., 2009), were not included. TAL effector sequences encoded by pseudogenes, characterized by an early stop codon, a frameshift mutation, and/or a large insertion that resulted in the absence of any and all matches to the NLS consensus sequence (K-K/R-X-K/R) (Garcia-Bustos et al., 1991; Van Den Ackerveken et al., 1996) or the lack of a stretch of 35 amino acids with at least 80% sequence identity to the acidic activation domain (Zhu et al., 1998; using BLS256 Tal1c as the reference) were also excluded from binding site predictions.

For the retained TAL effectors, EBEs were predicted using the TALE-NT 2.0 Target Finder tool (Doyle et al., 2012). Based on the RVD binding preferences determined by Yang et al. (2014), we updated our local version of Target Finder to treat HH as NH but left treatment of other previously uncharacterized RVDs unchanged, i.e., as having equal affinity for all nucleotides. Predictions were carried out using the rice promoterome, defined as the 5' UTR (if annotated) plus 1000 base pairs upstream of the transcriptional start site of each transcript; output for a gene includes all unique EBEs predicted in the promoter of any transcript of that gene. Promoters were retrieved from the *Oryza sativa* L. ssp. *japonica* (cv. Nipponbare) reference genome (v.7.0) downloaded from Phytozyme 10 using the MSU Release 7.0 annotation downloaded from the Rice Genome Annotation Project (Kawahara et al., 2013). Predicted EBEs were filtered with a previously developed machine learning classifier. To use the classifier on these large data sets, it was first retrained with transcriptional and translational start site locations from the genome annotation instead of manually curated locations based on EST support as in Cernadas et al. (2014). See Supplementary Table S1 in Data Sheet 2 for updated performance statistics and GitHub, <https://github.com/kwilkins226/TALEffectorClassifier>, for a Weka model file containing the classifier used here and a script to generate classifier input files from Target Finder output. Only EBEs assigned a probability of being true greater than

or equal to 0.5 were included in lists of “passed” binding site predictions.

Results and Discussion

Whole Genome Sequences of and Rice Transcriptional Responses to 10 Diverse Xoc Strains

Based on a Southern blot of *Bam*H-digested genomic DNAs of 34 geographically diverse strains using a TAL effector gene probe (Supplementary Figure S2 in Data Sheet 2), we selected nine strains in addition to BLS256 (**Table 1**) that were broadly representative of the restriction fragment length polymorphism observed, and performed SMRT, whole genome sequencing (NCBI BioProject number PRJNA283315). Only one strain, CFBP2286, carries any plasmid (see also Supplementary Figure S1 in Data Sheet 2).

In addition, we determined rice gene expression changes in response to each strain by RNA-Seq of leaf tissue 48 h after inoculation (NCBI Gene Expression Omnibus accession GSE67588).

TAL Effector Sequences of the 10 Xoc Strains

Across the 10 sequenced strains, 250 TAL effector genes and eight pseudogenes with recognizable TAL effector repeat regions were identified, all chromosomal. One TAL effector gene per strain, including *tal2h* of BLS256, encodes a TAL effector with no activation domain. These each show 95% or greater amino acid sequence identity throughout the C-terminal region (downstream of the CRR) to *Tal2h*. The gene in the African strain CFBP7342 is a pseudogene disrupted by an insertion sequence element, and the genes in the other two African strains, CFBP7331 and CFBP7341, have only six RVDs each, but in all other strains the genes display intact CRRs of 18 or more repeats. Seven of them contain one atypical-length repeat 28 amino acids in length, not matching one of the types characterized by Richter et al. (2014) as being able to accommodate a single base deletion at the corresponding location in the EBE (File S1, Data Sheet 1). Although BLS256 *Tal2h* was originally assumed to be nonfunctional due to its truncated C-terminus and lack of activation domain (Bogdanove et al., 2011), the broad conservation of this TAL effector variant suggests that it may serve an important function, perhaps as a virulence factor (Booher et al., unpublished). The use of engineered TAL effectors missing activation domains as transcriptional repressors in eukaryotes suggests a mechanism by which *Tal2h* could influence host gene expression (Blount et al., 2012; Werner and Gossen, 2014). Notably, the truncated C-terminus retains a predicted nuclear localization signal. Similarly, the atypical repeats of lengths not characterized by Richter et al. could be nonfunctional, but broad conservation suggests otherwise.

The remaining 241 structurally complete TAL effectors collectively represent 98 unique binding specificities, based (solely) on RVD composition and presence or absence of atypical-length repeats (File S1, Data Sheet 1). Six of the 241 have a single atypical-length repeat matching one of the types characterized by Richter et al. (2014) and none has more than

one. Four repeats of 36 amino acids each are also observed in four different TAL effectors. The distribution of RVDs observed in the TAL effector sequences (excluding those encoded in pseudogenes) is similar to that of previously sequenced *Xanthomonas* TAL effectors (Boch and Bonas, 2010), with 75% being HD, NN, NG, or NI (Supplementary Figure S3 in Data Sheet 2). Only 40 of the total 1916 RVDs are of types not assigned any specificity in the TALE-NT 2.0 Target Finder binding site prediction tool (Doyle et al., 2012). Among these are two RVDs, QD and HY, each occurring once, that have not been reported previously in any *Xanthomonas* TAL effector.

TAL Effector Conservation and Known TAL Effector Targets Upregulated by the Xoc Strains

To assess conservation of TAL effector sequences across strains, we grouped TAL effectors by apparent orthology (descent from the same ancestral gene with no duplication events; see Materials

and Methods) (**Figure 1**). To ascertain potential variation in targeting specificity within groups, we then recorded the number of BSR differences among the TAL effectors in each group, using the BLS256 TAL effector in the group as a reference, if present, or the TAL effector that was conserved exactly in the most strains otherwise (**Figure 1**). For orthologs of the 13 BLS256 TAL effectors with one or more known targets (Cernadas et al., 2014), we attempted to infer function and targeting specificity by asking whether the ortholog is predicted to bind the promoter for each corresponding target, and whether each target was upregulated by the corresponding strain in our RNA-Seq experiment. Whenever a BLS256 TAL effector is perfectly conserved at the BSR level in another strain, that strain upregulated the known target(s) (**Figure 1**). For imperfect matches, results were mixed. With one exception discussed below, every BLS256 TAL effector ortholog with no more than six BSR differences has in the promoter of each corresponding

| Strain | From | Tal1a | Tal1b | Tal1c | Os07068970 | Tal2a | Os0249780 | Tal2b | Tal2c | Os03903934 | Tal2d | Os0499194 | Tal2e | Tal2f | Tal2g | Os01952130 | Os069686500 | Tal2h | Tal3a |
|----------|-------------|-------|-------|-------|------------|-------|-----------|-------|-------|------------|-------|-----------|-------|-------|-------|------------|-------------|---------|-------|
| BLS256 | Philippines | 0 | 0 | 0 | 11.61 | 0 | 7.18 | 0 | 0 | 54.23 | 0 | 474.73 | 0 | 0 | 0 | 219.92 | 100.32 | (-AD) 0 | 0 |
| BLS279 | Philippines | X | 0 | 0 | 12.24 | 2 | 1.00 | 1 | 0 | 60.92 | 4 | 28.66 | X | X | 0 | 197.56 | 179.52 | (-AD) 1 | 0 |
| CFBP2286 | Malaysia | X | 0 | 0 | 8.26 | 0 | 3.58 | 1 | 0 | 24.02 | 0 | 211.50 | 0 | 1 | 0 | 94.45 | 77.16 | (-AD) 0 | 0 |
| B8-12 | China | X | 0 | 0 | 9.46 | 0 | 5.34 | 1 | 1 | 110.79 | 4 | 56.25 | 2 | X | 0 | 223.22 | 151.25 | X | 0 |
| L8 | China | X | 0 | 0 | 23.07 | 0 | 9.92 | 1 | 1 | 197.12 | 4 | 60.52 | 2 | X | 0 | 215.34 | 214.25 | (-AD) 1 | 0 |
| RS105 | China | X | 0 | 0 | 9.45 | 0 | 4.60 | 1 | X | 1.00 | X | 1.00 | X | X | 0 | 157.95 | 136.29 | X | 0 |
| BXOR1 | India | X | 0 | X | 1.00 | X | 1.00 | 2 | 3 | 48.32 | 3 | 179.16 | X | X | 0 | 103.06 | 97.75 | X | 0 |
| CFBP7331 | W. Africa | 7 | X | X | 1.00 | X | 1.00 | 6 | 4 | 85.65 | 3 | 120.57 | X | X | 1 | 177.22 | 208.31 | X | 1 |
| CFBP7341 | W. Africa | 7 | X | X | 1.00 | X | 1.00 | 6 | 4 | 53.64 | 3 | 139.41 | X | X | 1 | 163.82 | 175.28 | X | 1 |
| CFBP7342 | W. Africa | X | 1 | X | 1.00 | X | 1.00 | 6 | X | 15.45 | 3 | 292.41 | X | X | 1 | 204.80 | 180.73 | X | 1 |

| Strain | From | Tal1b | Os0234970 | Os05927390 | Os07936430 | Tal3c | Os0247680 | Os0397540 | Tal4a | Os0397840 | Tal4b | Os03932100 | Tal4c | Os06937080 | Tal5a | Os02915290 | Tal5b | Tal6 | Os01931220 | Os06929100 | Os12942970 |
|----------|-------------|-------|-----------|------------|------------|-------|-----------|-----------|-------|-----------|-------|------------|-------|------------|-------|------------|-------|------|------------|------------|------------|
| BLS256 | Philippines | 0 | 243.56 | 217.74 | 86.24 | 0 | 20.81 | 705.86 | 0 | 43.85 | 0 | 43.85 | 0 | 109.91 | 0 | 148.30 | 0 | 0 | 12.86 | 39.21 | 41.08 |
| BLS279 | Philippines | 0 | 219.18 | 237.55 | 162.31 | 0 | 20.66 | 658.77 | 0 | 40.17 | 0 | 40.17 | 0 | 113.09 | 0 | 146.22 | 0 | 0 | 11.96 | 31.10 | 25.61 |
| CFBP2286 | Malaysia | 0 | 77.78 | 115.06 | 64.53 | 0 | 11.36 | 326.54 | X | 1.00 | 1 | 1.00 | 0 | 45.34 | 0 | 103.51 | 0 | 1 | 1.00 | 25.76 | 13.19 |
| B8-12 | China | 0 | 265.41 | 254.94 | 168.33 | 0 | 23.95 | 797.24 | 0 | 42.68 | 0 | 42.68 | 0 | 118.77 | 0 | 132.80 | 0 | 0 | 8.37 | 16.00 | 32.86 |
| L8 | China | 0 | 172.61 | 233.61 | 134.39 | 0 | 27.70 | 625.95 | 0 | 29.78 | 1 | 29.78 | 0 | 117.09 | 0 | 172.70 | 0 | 0 | 22.41 | 86.74 | 20.36 |
| RS105 | China | 0 | 161.23 | 223.77 | 135.07 | 0 | 17.76 | 585.38 | 0 | 35.48 | 0 | 35.48 | 0 | 85.01 | 0 | 117.09 | 0 | 0 | 6.29 | 13.16 | 16.52 |
| BXOR1 | India | 0 | 92.27 | 74.34 | 81.79 | 1 | 12.14 | 252.90 | 1 | 24.55 | 3 | 24.55 | 5 | 42.92 | 2 | 173.96 | 0 | 1 | 2.04 | 15.61 | 8.94 |
| CFBP7331 | W. Africa | 0 | 8.75 | 81.47 | 11.40 | X | 1.00 | 7.04 | 1 | 36.01 | X | 36.01 | X | 26.30 | 16 | 1.00 | X | 6 | 1.00 | 5.14 | 26.58 |
| CFBP7341 | W. Africa | 0 | 11.26 | 74.93 | 13.85 | X | 1.00 | 5.35 | 1 | 37.92 | 4 | 38.67 | X | 1.00 | X | 6 | 1.00 | 1.00 | 24.29 | | |
| CFBP7342 | W. Africa | 1 | 163.06 | 155.67 | 207.48 | X | 17.85 | 56.69 | 2 | 38.25 | X | 38.25 | 5 | 51.19 | X | 1.00 | 0 | 6 | 1.00 | 1.00 | 28.57 |

| Strain | From | Tal7 | Tal8 | Tal9a | Tal9b | Os01951040 | Tal10 | Tal11a | Tal11b | Tal12 | non-BLS256 TAL#1 | non-BLS256 TAL#2 | non-BLS256 TAL#3 | non-BLS256 TAL#4 | non-BLS256 TAL#5 | non-BLS256 TAL#6 | non-BLS256 TAL#7 | non-BLS256 TAL#8 | non-BLS256 TAL#9 | non-BLS256 TAL#10 | unique TAL effectors |
|----------|-------------|------|------|-------|-------|------------|-------|--------|--------|-------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|-------------------|----------------------|
| BLS256 | Philippines | 0 | 0 | 0 | 0 | 30.46 | 0 | 0 | 0 | X | X | X | X | X | X | X | X | X | X | X | 0 |
| BLS279 | Philippines | 0 | 0 | 0 | 3 | 14.66 | 0 | 0 | 0 | 2 | X | X | X | 3 | X | X | X | X | X | X | 0 |
| CFBP2286 | Malaysia | 0 | 0 | 0 | 0 | 23.94 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | X | X | X | X | X | X | 0 |
| B8-12 | China | 0 | 0 | 0 | 3 | 15.47 | 0 | 0 | 0 | 0 | X | X | X | 0 | 0 | X | (-AD) 0 | X | X | X | 0 |
| L8 | China | 0 | 0 | 0 | 3 | 36.39 | 0 | 0 | 0 | 0 | X | X | X | 0 | 0 | X | X | X | X | X | 1 |
| RS105 | China | Ψ | 0 | 0 | 0 | 3 | 16.20 | 0 | 0 | 0 | X | X | X | 0 | 0 | X | (-AD) 0 | X | X | X | 0 |
| BXOR1 | India | 0 | X | 1 | X | 1.00 | X | 2 | 0 | 1 | 0 | X | 0 | X | X | 2 | X | 0 | 0 | X | 3 |
| CFBP7331 | W. Africa | 2 | X | 1 | X | 0.58 | X | 2 | 2 | X | X | 0 | X | X | X | X | X | X | (-AD) 0 | 0 | 0 |
| CFBP7341 | W. Africa | 2 | X | 1 | X | 0.60 | X | 2 | 2 | X | 0 | X | X | 0 | X | Ψ | 0 | 0 | 0 | 0 | 1 |
| CFBP7342 | W. Africa | 3 | X | 1 | X | 1.00 | X | X | 2 | 2 | X | X | 1 | X | 4 | X | 0 | X | 0 | X | 6 |

FIGURE 1 | TAL effector content of diverse Xoc strains in relation to BLS256, and shared upregulation of BLS256 TAL effector targets.

Each column labeled in bold font represents a group of orthologous TAL effectors, identified by finding reciprocal best BLAST hits. Columns are labeled according to the BLS256 ortholog, or if absent, by the designation “non-BLS256 TAL,” numbered arbitrarily. In these columns, the number of BSRs by which each TAL effector in the group (by strain) differs from the reference TAL effector for that group is given. The reference is the BLS256 TAL effector if present and the TAL effector with the most conserved BSR sequence otherwise. Gray fill indicates at least one difference. An “X,” with

black fill, indicates absence of an ortholog. “Ψ” indicates that the TAL effector sequence derives from a pseudogene. “(-AD)” indicates a C-terminus like that of BLS256 Tal2h, missing the activation domain. Columns labeled in regular font with a rice gene name (locus ID) show the fold upregulation of that gene by each strain, with black fill indicating genes that are not significantly upregulated. The genes are grouped by the BLS256 TAL effector that upregulates them, immediately to the right of the column for that TAL effector. Fold change values in red font mark instances in which the gene promoter has no predicted EBE that passes a machine learning filter for the BLS256 TAL effector ortholog present in the strain represented in that row.

BLS256 target gene a predicted EBE that passes a machine learning filter for predicting functionality (see Materials and Methods), and for 99% of these, the target is upregulated. The exception is the BLS256 Tal6 ortholog group. Neither Tal6 nor any of its orthologs has an EBE that passes the filter in the confirmed Tal6 target *Os01g31220*. The reason for this is not clear. Nonetheless, since target activation by BLS256 in all cases was shown to be TAL effector dependent, for any ortholog the presence of a predicted EBE and upregulation of the corresponding target is strong evidence of activation by that ortholog in each case. Where this is the case, the ortholog likely serves the same function as the BLS256 TAL effector, but we cannot exclude the possibility that non-identical orthologs target different genes distinct from the BLS256 TAL effector target(s).

Five TAL effector groups have an ortholog from every strain that is neither pseudogenized nor more than two BSRs diverged from the others. These include the group containing BLS256 Tal2g, a virulence factor that activates the *S* gene *OsSULTR3;6* (Cernadas et al., 2014), and indeed, all 10 strains activate that gene. That all strains use the same TAL effector to upregulate this major *S* gene might reflect a targeting constraint at the promoter that could make modification or deletion of the EBE an effective strategy for durable resistance, one that the pathogen might not readily overcome by targeting a different sequence in the promoter. Another group contains BLS256 Tal11b, a TAL effector of which a knock-out mutant showed reduced virulence on Nipponbare (Cernadas et al., 2014) but for which a target has yet to be identified. The remaining three TAL effector groups include Tal3a, Tal3b, and Tal9a of BLS256, respectively. Although no non-redundant role in virulence for these three TAL effectors was detected based on mutant phenotypes in Nipponbare in a growth chamber (Cernadas et al., 2014), their conservation suggests that they might perform an important function in other rice genotypes or in the field.

While conservation of a TAL effector suggests that the TAL effector is important, conservation of TAL effector target upregulation in the absence of TAL effector conservation suggests that the target gene is important. As an example, Tal3c of BLS256, though conserved in all non-African strains, is absent from the three African strains, yet its two known targets, *Os02g47660* and *Os03g07540* (which are upregulated by all the non-African strains), are upregulated, respectively, by CFBP7342 and by all three African strains (Figure 1). Because the activation of these genes by BLS256 depends on a TAL effector (Cernadas et al., 2014), their upregulation by the African strains seems likely to as well. Consistent with that hypothesis, a TAL effector present only in CFBP7342 has a predicted EBE in the promoter of *Os02g47660* and several predicted EBEs in the promoter of *Os03g07540* that pass the machine learning classifier. The *Os02g47660* EBE does not overlap the Tal3c EBE, but one of the *Os03g07540* EBEs does. Two more TAL effectors, found in all three African strains (one only in the three African strains and the other also in BXOR1), both have multiple, predicted, passing EBEs in *Os03g07540*. One predicted, passing EBE for each of these TAL effectors overlaps the Tal3c EBE by 10 base pairs, and the other overlaps it completely (Supplementary Figure S4 in Data Sheet 2). In addition to pointing to new, strong candidate TAL

effector-target pairs, these observations suggest that *Os03g07540*, the Tal3c target upregulated by all strains though apparently by different TAL effectors, is important in bacterial leaf streak. This is further supported by the fact that *Os03g07540* encodes a member of the bHLH protein family, which includes *UPA20*, the *S* gene target of *X. campestris* pv. *vesicatoria* TAL effector AvrBs3 in pepper (*Capsicum annuum*) (Kay et al., 2007). A *tal3c* mutant strain of BLS256 was not significantly less virulent in a growth chamber using artificial inoculation, but again, it seems likely that such an assay fails to detect differences that would be significant under field conditions.

Relationships of TAL Effectors across Strains in Comparison to Housekeeping Genes

Conservation of TAL effector binding specificities within the Asian and African groups is high. As a point of reference, the four sequenced Xoo strains (Lee et al., 2005; Ochiai et al., 2005; Salzberg et al., 2008; Booher et al., unpublished), all from Asia, in pairwise comparisons share at most 36% of their TAL effectors and on average only 21% (as a percentage of whichever strain has more TAL effectors, and based on perfect identity of BSR sequences). Between every pair of Asian Xoc strains, a minimum of 25% and on average 57% of TAL effectors are conserved. For the African Xoc strains, the minimum is 32% and the average 51%. Across all the Xoc strains, the average is 33%, but there is clear distinction between the Asian and African groups, with only BXOR1 sharing more than 10% of its TAL effectors with any of the African strains. The comparatively high level of TAL effector conservation within the two Xoc groups could be the result of purifying selection or a relative lack of diversifying selection, but could in part reflect broad dissemination of TAL effector genes via horizontal gene transfer. To address this possibility, we generated and compared two phylogenetic trees (Figure 2A), one based on the sequences of 31 housekeeping genes that are recalcitrant to horizontal gene transfer (Jain et al., 1999; Wu and Eisen, 2008) and the other based on the groups of TAL effector orthologs described above. The two trees are nearly identical, but the tree created using the TAL effector orthologs is a significantly better fit for the TAL effector ortholog alignment than the tree created using the housekeeping genes (Kishino-Hasegawa test $p = 0.047$). This result indicates that TAL effectors may have been horizontally transferred among these strains, though probably infrequently given the similar overall topologies of the two trees. Phylogenies based on individual TAL effector ortholog groups support this conclusion, with only six of thirty testable ortholog groups yielding a TAL effector-based topology that is a significantly better fit for the TAL effector alignment than is the topology of a housekeeping gene-based phylogeny (Supplementary Figure S5 in Data Sheet 2).

For each of those six groups, represented by BL256 Tal2b, Tal2c, Tal2g, Tal4b, and Tal12, and by B8-12 Tal5e (non-BLS256 Tal#4 ortholog group in Figure 1 and Supplementary Figure S5 in Data Sheet 2), the genomic context of each TAL effector gene represented in the group, specifically the local spatial relationship to other TAL effector genes, is conserved across the genomes of the strains in the group (Supplementary Figure S5 in Data Sheet 2 and Figure 4). This suggests that either

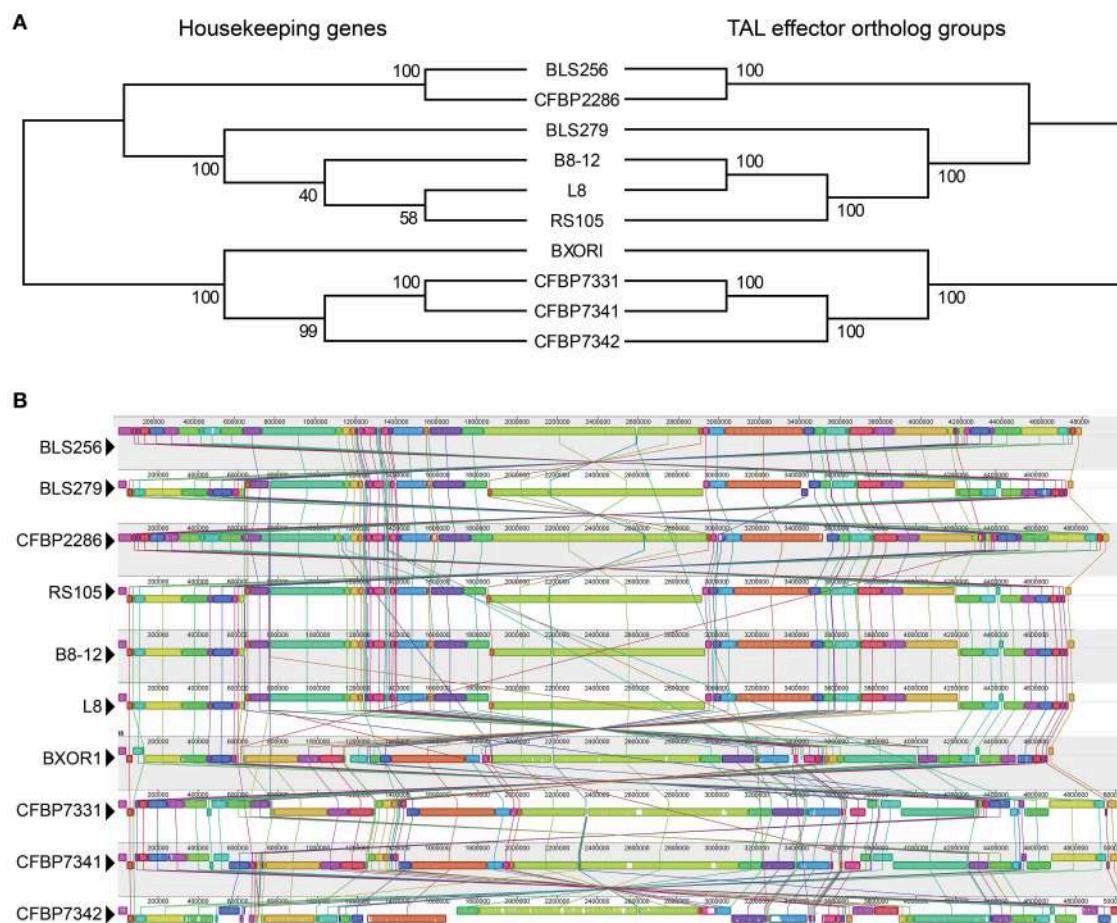


FIGURE 2 | Phylogenetic relationships based on housekeeping genes vs. TAL effectors, and whole genome alignments of the Xoc strains.

(A) Maximum likelihood trees, created using concatenated alignments of the nucleotide sequences of either 31 housekeeping genes or the TAL effectors in each of the 39 ortholog groups shown in Figure 1. Bootstrap values from 1000 replicates are given at the nodes. (B) Alignment of the whole genome

of every strain, generated using progressiveMauve (Darling et al., 2010). Colored blocks represent regions of uninterrupted homology, connected between genomes by matching colored lines. The height of the colored bars within each block indicates relative conservation on average across all genomes. Homologous regions above the centerline are in the same orientation as they are in BLS256 and those below are inverted.

clusters of TAL effector genes were transferred or individual TAL effector genes were transferred to the same locations within those clusters. This second possibility seems unlikely given the overall high sequence similarity across TAL effector genes and often their flanking DNA, which should result in integration of a horizontally transferred TAL effector gene into any cluster randomly. Therefore, if the TAL effector genes in the six groups were horizontally transferred, neighboring TAL effector genes likely were as well. We saw no evidence for this however. For the BLS256 Tal4b and Tal12 groups, TAL effectors encoded by genes on either side yielded tree topologies that fit the TAL effector alignments no better than the housekeeping tree-based topologies did. Similarly, for the groups represented by BLS256 Tal2b, Tal2c, and Tal2g, which are encoded in the same cluster, and for the group represented by B8-12 Tal5e, which is encoded in the orthologous cluster in each of the four genomes in which it occurs, in the same location, TAL effectors encoded by intervening genes in the cluster (Tal2d, Tal2e, and Tal2f in

BLS256) likewise showed no evidence of horizontal transfer (the Tal2f group represents only two strains and was therefore not meaningful). Thus, for these six groups, either the difference between the TAL effector-based topologies and the housekeeping gene-based topologies is not due to horizontal gene transfer, or the lack of difference for the groups representing flanking or intervening genes is in each case a false negative.

Consistent with the high levels of Xoc TAL effector conservation, the Xoc genomes overall are highly structurally conserved. Only 53 recombination breakpoints are required to explain the genome rearrangements observed among the strains (Figure 2B), and no breakpoints detectable by GARD (Kosakovsky Pond et al., 2006b) occur within a TAL effector gene. With minor exceptions that likely arose from exchange between two TAL effector genes by double cross-over, or in some cases local rearrangements associated with insertion sequence elements, all apparent orthologs reside in conserved contexts within their respective genomes (Figure 3).

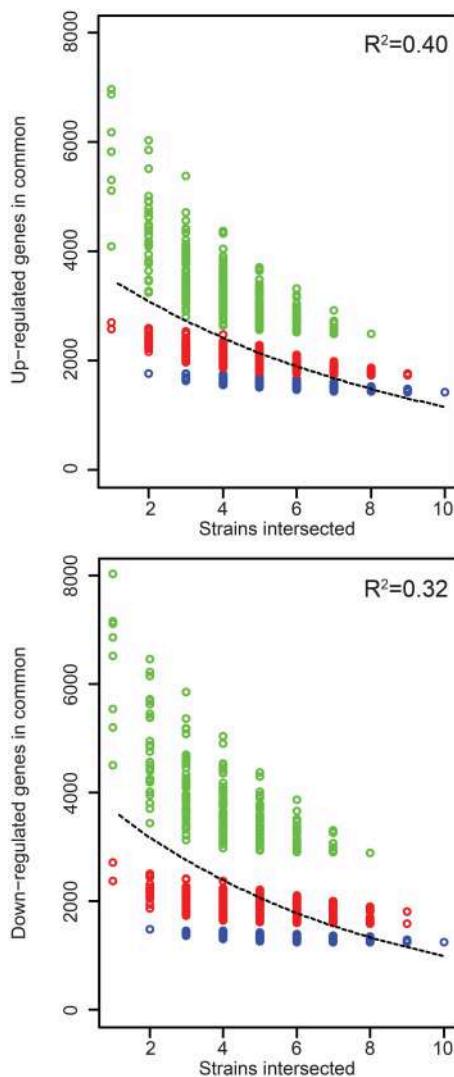


FIGURE 3 | Rice gene expression changes induced in common by the Xoc strains. The number of genes up- or down- regulated in rice (cv. Nipponbare, at 48 h after syringe infiltration) in common by each of n strains across all possible groups of strains is plotted as a function of n . The least-square fit of an exponential decay function is shown as a dashed line, with R^2 values for the fit displayed in the upper-right of each graph. CFBP2286 and BXOR1 induced the fewest gene expression changes (up or down), and this is reflected in stratification of the points in each plot. To make this apparent, groups containing only one of these strains are shown in red and groups containing both in blue.

TAL Effector Content in Relation to Rice Gene Expression Changes

Because even similar TAL effectors may target different sets of host genes and because different TAL effectors may share targets (Yang et al., 2006; Antony et al., 2010; Römer et al., 2010; Yu et al., 2011; Streubel et al., 2013; Richter et al., 2014), differences in TAL effector content between strains by itself may not be predictive of differences in the changes to rice gene expression each strain causes. To examine this, we compared the host gene expression changes in response to all the strains in relation to their TAL

effector content. The RNA-Seq results we obtained show that the Xoc strains on average upregulate 5152 rice genes, but only 1437 (28%) of these are upregulated by every strain (Table 2). The number of genes downregulated by each strain is similar, 5608 genes on average, with 1248 (22%) of these downregulated in common. Every gene expression change uniformly required for the development of bacterial leaf streak of rice by definition is represented among the genes up- or down- regulated by all 10 strains. The relatively small number of these genes up- or downregulated in common might be explained in part by the exclusion of any gene expression change that serves a required function but is substituted for in the case of some strains by a distinct induced change. Upregulation of different SWEET gene family members by different TAL effectors of different Xoo strains in bacterial blight of rice is an example of such essential but interchangeable TAL effector induced changes to host gene expression (Yang et al., 2006; Antony et al., 2010; Römer et al., 2010; Yu et al., 2011; Streubel et al., 2013; Richter et al., 2014). Incidentally, the list of gene expression changes induced in common by every strain almost certainly contains a number of non-essential changes: plotting the number of shared up- and down-regulated genes among all possible combinations of the strains reveals that an exponential decay function fits the data poorly (Figure 4); in other words, the rate at which the number of shared gene expression changes decreases does not decrease as more strains are added, up to the 10 total.

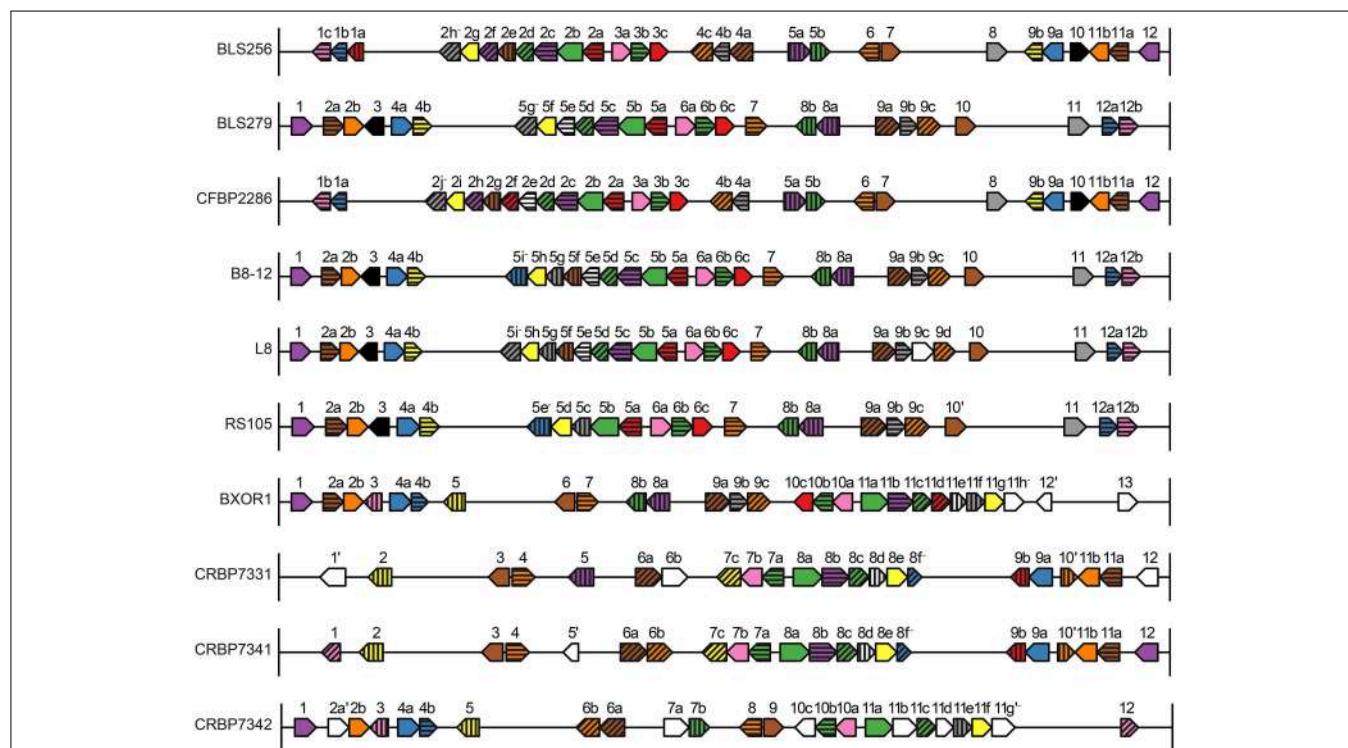
Overall, relationships among the rice transcriptional responses to the 10 Xoc strains, shown as a tree in Figure 5A (see Materials and Methods), do not match those observed for the TAL effectors (Figure 2A). Of particular note are the Asian strains, which cluster distinctly in the TAL effector based tree, but not in the tree based on rice gene expression changes. This greater conservation of induced host gene expression changes than TAL effector content may be due to TAL effector-independent changes, and/or convergent evolution having resulted in diverse TAL effectors sharing targets, as suggested by the targets of BLS256 Tal3c and exemplified by the SWEET genes in bacterial blight, discussed above. Perhaps not surprisingly, however, when only the expression changes of the known targets of BLS256 TAL effectors are examined (Figure 5B), the clustering more closely recapitulates the tree based on the orthologous TAL effector groups (Figure 2A).

Toward Identification of New TAL Effector Targets in Bacterial Leaf Streak of Rice

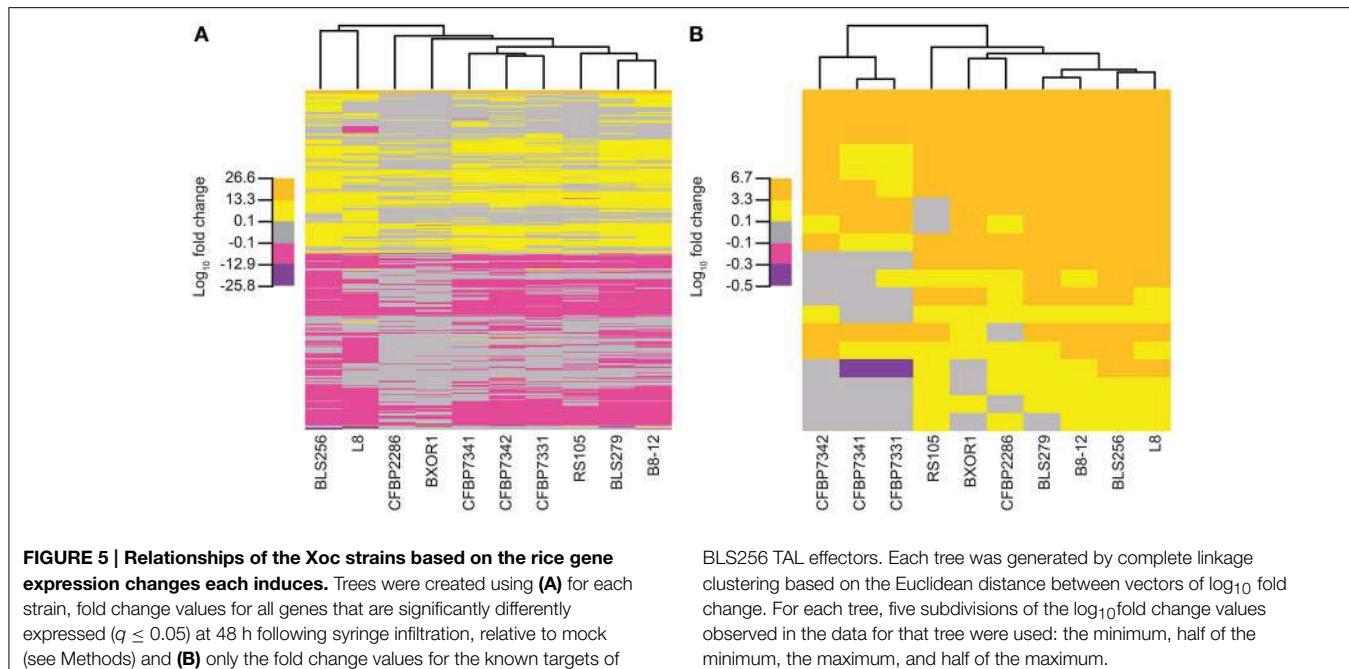
TAL effectors often determine whether an interaction leads to disease or to successful plant defense. In addition to augmenting the promoter of an *R* gene with EBEs for conserved TAL effectors to broaden its effectiveness (Hummel et al., 2012; Zeng et al., 2015), genetic manipulation of host *S* gene promoters to prevent activation, through breeding or by genome editing, can be a successful disease control strategy (Yang et al., 2006; Li et al., 2012b). For broad effectiveness and durability, identification of genes widely targeted by strains within and across pathogen populations, either by conserved TAL effectors or distinct ones in different strains, is an important starting point. Toward this goal, we first sought to identify candidate TAL effector targets

TABLE 2 | Numbers of TAL effectors, induced rice gene expression changes, and target predictions for each Xoc strain.

| Strain or value | TAL effectors ^a | Upregulated genes ^b | Downregulated genes ^c | Predicted EBEs ^d | Passed EBEs ^e | Predicted candidate genes ^f | Passed candidate genes ^g |
|---------------------|----------------------------|--------------------------------|----------------------------------|-----------------------------|--------------------------|--|-------------------------------------|
| B8-12 | 27(1) | 6966 | 7112 | 55,980 | 23,602 | 6966 | 3930 |
| BLS256 | 27(1) | 6881 | 8043 | 55,980 | 24,980 | 6881 | 3976 |
| BLS279 | 25(1) | 6191 | 6872 | 55,980 | 23,539 | 6191 | 3538 |
| BXOR1 | 25(2) | 2592 | 2373 | 55,981 | 30,117 | 2592 | 1771 |
| CFBP2286 | 27(1) | 2708 | 2713 | 55,980 | 25,581 | 2708 | 1609 |
| CFBP7331 | 19(3) | 5117 | 5554 | 55,980 | 32,080 | 5117 | 3623 |
| CFBP7341 | 19(3) | 5306 | 5213 | 55,980 | 31,977 | 5306 | 3715 |
| CFBP7342 | 22(2) | 5830 | 6531 | 55,983 | 45,973 | 5830 | 5065 |
| L8 | 26(3) | 5835 | 7157 | 55,980 | 24,994 | 5835 | 3479 |
| RS105 | 22(2) | 4093 | 4509 | 55,980 | 21,604 | 4093 | 2186 |
| Average | 24 | 5152 | 5608 | 55,980 | 28,445 | 5152 | 3289 |
| Shared ^h | 5 | 1437 | 1248 | 55,979 | 16,425 | 1437 | 648 |

^a The number in parentheses is the number of pseudogenes.^b The number of genes upregulated in rice (cv. Nipponbare) leaves at 48 h after syringe infiltration of the strain relative to mock inoculated leaves ($q < 0.05$).^c The number of genes downregulated in rice (cv. Nipponbare) leaves at 48 h after syringe infiltration of the strain relative to mock inoculated leaves ($q < 0.05$).^d The number of genes with at least one predicted EBE in the promoter.^e The number of genes with at least one EBE in the promoter that passes the machine learning classifier filter.^f The number of upregulated genes with at least one predicted EBE in the promoter.^g The number of upregulated genes with at least one EBE in the promoter that passes the machine learning classifier filter.^h Shared TAL effectors are defined as having no more than two BSR differences.**FIGURE 4 |** Identity, order, and orientation of TAL effector genes across the Xoc strains. Each arrow represents a TAL effector (*tal*) gene and points in the 5' to 3' direction. The genes are to scale relative to one another, as are the intergenic regions, but the genes have been magnified relative to the rest of the genome. Arrows of the same color and pattern

represent orthologs. The same color or pattern alone does not indicate any particular relationship. Genes are numbered and lettered according to Salzberg et al. (2008). An apostrophe indicates a pseudogene, and a minus sign indicates the encoded TAL effector has a BLS256 Tal2h-like C-terminus, lacking the activation domain.



for each of the 10 Xoc strains by predicting whether any of the genes upregulated by a strain contains a possible EBE in its promoter for any of the TAL effectors of that strain. For this, we used the Target Finder tool of TALE-NT 2.0 with a score ratio threshold of 3.0 (Doyle et al., 2012). For any TAL effector with an atypical-length repeat of a type characterized by Richter et al., we ran predictions using the complete RVD sequence as well as that sequence with the RVD of the repeat left out. We treated repeats of other atypical lengths as typical repeats. We did not do predictions for TAL effectors with the truncated Tal2h-like C-terminus, reasoning that if they are functional (1) they might act by targeting outside the promoter (Politz et al., 2013), and (2) their targets might include genes not differentially expressed in comparison to the mock control used in the RNA-Seq experiment, but instead prevented from being otherwise activated by the pathogen. Somewhat surprisingly, every gene upregulated by any strain has a predicted EBE for at least one TAL effector of that strain, meaning that the predictions provide no additional information beyond the RNA-Seq data. In fact, Target Finder predicts an EBE for at least one TAL effector per strain in nearly all of the 55,986 gene promoters in the MSU7 (Kawahara et al., 2013) annotation (Table 2). Consequently, in order to identify a testable list of candidate TAL effector targets from this and like studies of wild type strains with multiple TAL effectors, either more specific (fewer false positive) EBE predictions, or methods to rank and filter candidate genes are essential.

A modified version of the machine learning classifier from Cernadas et al. (2014) (see Methods) appears to perform well as a way to filter candidate genes. If instead of starting with upregulated genes, the whole genome is examined, using only EBEs passing this classifier reduces the number of candidate genes by 36% across all samples, compared to not using the classifier (Table 2). Of the resulting set of genes with passing EBEs, 12% are upregulated, a significant increase from the 9%

of all genes with predicted EBEs that are upregulated ($p < 0.01$). This enrichment suggests that the filtering is beneficial and could be applied to the list of candidates from the upregulated gene set. Having RNA-Seq data for the host interacting individually with each of multiple strains for which the TAL effector content is known enables a second, independent method of filtering (the machine learning classifier does not use the RNA-Seq data as input), eliminating any candidate target that is not upregulated by every strain that contains the corresponding TAL effector. All known BLS256 TAL effector targets pass this filter. Though expected, this result provides some assurance that requiring a candidate gene to be upregulated by all strains containing the relevant TAL effector will not eliminate true TAL effector targets from consideration. The two filters discussed here, the classifier and the shared TAL effector-shared target criterion, can be used together or independently. Supplementary Table S2 (Data Sheet 3) provides a list of candidates for the five TAL effectors shared, with no more than two BSR differences, by all 10 strains, generated using both filters together. These lists and others that could be generated for geographical subsets of these strains represent promising starting points for future experimental analyses to identify additional TAL effector targets, including new BLS S genes.

Conclusion

As discussed in the introduction, new strategies centered on TAL effectors and their targets show great promise for control of plant diseases in which TAL effectors play determinative roles (Boch et al., 2014). One of these, amending an *R* gene promoter with one or more additional EBEs, for breadth and durability requires using EBEs that correspond to conserved TAL effectors or to a minimal set of representative TAL effectors so that it will trap the diversity of strains in a population.

Another, integrating or engineering an allele of a major *S* gene that is immune to activation due to disruption of its EBE(s), requires identifying an *S* gene on which the strains in a population uniformly depend. Characterizing population level TAL effector conservation addresses the first of these requirements, and assessing host genome-wide expression data alongside TAL effector sequences helps address the second by enabling identification of conserved and potentially important candidate TAL effector targets. The analyses we present in this paper of TAL effector sequences from 10 diverse Xoc strains (from eight genome sequences we present here and two we determined previously), and RNA-Seq data we captured for rice responding to each strain, are a significant advance toward these objectives for the increasingly globally important rice disease bacterial leaf streak.

Our RNA-Seq results are consistent with the microarray-based study of BLS256-inoculated rice we published previously (Cernadas et al., 2014): all TAL effector targets identified in that study are upregulated in the RNA-Seq data for all strains that contain a TAL effector with the corresponding RVD sequence. However, because RNA-Seq is more sensitive than hybridization to a microarray, and because, for reasons outlined below, we predicted a greater number of binding sites per TAL effector, on average, than we previously predicted for the BLS256 TAL effectors (Cernadas et al., 2014), here we were able to identify, with filtering, candidate targets for all of the 27 intact BLS256 TAL effectors except Tal2b (which is not, as it was originally described, a pseudogene; Booher et al., unpublished), in contrast to the 19 TAL effectors for which we were able to identify candidate targets using the microarray data, without filtering (Cernadas et al., 2014).

The greater number of EBE predictions here is due, in part, to the presence of more uncharacterized RVDs in this larger collection of TAL effectors. Among the BLS256 TAL effectors, Tal2g contains the only two uncharacterized RVDs. In our previous work, we replaced these with RVDs of known specificity that shared the same BSR (Cernadas et al., 2014). The TAL effectors in the present study include 40 previously uncharacterized RVDs, and based on recent experimental results (Yang et al., 2014; Miller et al., 2015), we chose to replace only some of these with observed RVDs of known specificity. The impact of uncharacterized RVDs on binding site prediction specificity is evident in the filtered candidate gene lists for the five highly conserved groups of orthologous TAL effectors. The filtered candidate gene lists for four of the ortholog groups contain fewer than 11 genes. The TAL effectors in these ortholog groups contain no uncharacterized RVDs. The gene list for the fifth group, which includes Tal2g, has 521 candidate genes that pass both filtering steps. Each TAL effector in that group contains two uncharacterized RVDs. The large number of candidate genes identified for this group almost certainly results from the binding site prediction algorithm treating uncharacterized RVDs as having equal affinity for all four nucleotides.

Another reason for the greater number of candidates in this study is our having considered all potential EBEs in a promoter, instead of just the best scoring one. This allows for the possibility that a poorer scoring EBE that is closer to the transcriptional start site is more likely to activate a gene than a better scoring

EBE that is at a distance, and results in more candidates passing the machine learning filter. We also redefined the promoter as the 5' UTR plus 1000 bp upstream of the transcriptional start site, instead of just the latter, which was used previously. This allows for the possibility that some TAL effectors activate their targets by binding within the 5' UTR of the basal transcript and resetting the transcriptional start site. Finally, in our previous study (Cernadas et al., 2014), we determined the EBE score cutoff for each TAL effector individually using the distribution of the best predicted binding sites in each gene in the rice genome. In the present study, we fixed the score cutoff at three times the best possible score, reflecting the current Target Finder algorithm.

With the greater sensitivity of RNA-seq and the prediction parameters we used, absent a method to more specifically predict EBEs, the filtering is essential to obtain a testable number of candidate targets from RNA-Seq studies using wildtype strains with multiple TAL effectors. Generating individual TAL effector knockout strains for comparison to the wild type is desirable, but not always feasible. Comparison across distinct wildtype strains is a powerful alternative. In the case of BLS256 Tal3c target *Os03g07540*, this approach enabled us to distinguish this gene from among the several targets of that TAL effector as potentially important, because it was the only Tal3c target upregulated by all strains, ostensibly due to being targeted by a TAL effector distinct from Tal3c in some strains.

In light of our overall results, TAL effector-centered strategies for control of BLS appear promising. We observed that within geographic regions, Xoc TAL effectors are highly conserved relative to Xoo TAL effectors. They are not disrupted by breakpoints that lead to large, genomic rearrangements, and they appear rarely or never to have been horizontally transferred within Xoc, consistent with the observation of Ferreira et al. (2015) that Xoc TAL effector genes do not localize with *TnXax1short* inverted repeat sequences to form mobile cassettes, unlike TAL effector genes in other *Xanthomonas* species. Moreover, we identified five Xoc TAL effectors that are shared by all 10 sequenced strains with no more than two BSR differences. As discussed above, one of these groups, the BLS256 Tal2g group, targets the major *S* gene *OsSULTR3;6*, indicating that edited or naturally occurring alleles lacking the corresponding EBE might be broadly effective for disease control. An *R* gene amended with EBEs to capture all five groups also seems likely to be broadly effective and durable.

Finally, the conserved TAL effector groups represent footholds for further dissecting the functions of TAL effectors and their targets. The fact that one of the groups contains BLS256 Tal11b, for which a mutant strain showed reduced virulence but for which no target was identified (Cernadas et al., 2014), supports the conclusion that Tal11b and the orthologs are important but may act in a non-canonical way, such as downregulating a gene by binding to its 5' UTR. The three groups for which no virulence contribution of the BLS256 TAL effector was observed in the growth chamber assay of Nipponbare plants, as discussed, may exemplify TAL effectors with host genotype-specific virulence contributions, or functions not measured by our virulence assay but important under field conditions. These might include contributing to the ability of the pathogen to initiate infection

from a low titer or to disseminate. Broad conservation of the TAL effectors in each of these cases recommends them for further experimentation to test these possibilities.

Author Contributions

LW, KW, and AB conceived and designed the study. LW and KW carried out the experiments. KW, NB, and AB analyzed and interpreted the data. KW and AB prepared the paper, with assistance from LW and NB.

Acknowledgments

This study was supported by the US National Science Foundation (Plant Genome Research Program Award IOS 1238189 to AB and

a Graduate Research Fellowship under Grant No. DGE-1144153 to KW). The authors thank D. Nettleton and C. Du for guidance on the use of QuasiSeq to detect differential gene expression and S. Carpenter for technical assistance. The authors also thank V. Verdier and P. Poitier for the CFBP strains, R. Sonti for BXOR1, C. Song for B8-12 and RS105, Z. Yin for L8, and C.M. Vera-Cruz for BLS256 and BLS279. This work used Red Cloud, servers and storage for cloud computing, which is supported by the Cornell University Center for Advanced Computing.

Supplementary Material

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2015.00536>

References

- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389
- Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. doi: 10.1093/bioinformatics/btu638
- Antony, G., Zhou, J., Huang, S., Li, T., Liu, B., White, F., et al. (2010). Rice *xa13* recessive resistance to bacterial blight is defeated by induction of the disease susceptibility gene *Os11N3*. *Plant Cell* 22, 3864–3876. doi: 10.1105/tpc.110.078964
- Aziz, R. K., Bartels, D., Best, A. A., Dejongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. doi: 10.1186/1471-2164-9-75
- Blootl, B. A., Weenink, T., Vasylechko, S., and Ellis, T. (2012). Rational diversification of a promoter providing fine-tuned expression and orthogonal regulation for synthetic biology. *PLoS ONE* 7:e33279. doi: 10.1371/journal.pone.0033279
- Boch, J., and Bonas, U. (2010). *Xanthomonas AvrBs3* family-type III effectors: discovery and function. *Annu. Rev. Phytopathol.* 48, 419–436. doi: 10.1146/annurev-phyto-080508-081936
- Boch, J., Bonas, U., and Lahaye, T. (2014). TAL effectors – pathogen strategies and plant resistance engineering. *New Phytol.* 204, 823–832. doi: 10.1111/nph.13015
- Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., et al. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 326, 1509–1512. doi: 10.1126/science.1178811
- Bogdanove, A. J., Koebnik, R., Lu, H., Furutani, A., Angiuoli, S. V., Patil, P. B., et al. (2011). Two new complete genome sequences offer insight into host and tissue specificity of plant pathogenic *Xanthomonas* spp. *J. Bacteriol.* 193, 5450–5464. doi: 10.1128/JB.05262-11
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Bullard, J., Purdom, E., Hansen, K., and Dudoit, S. (2010). Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* 11:94. doi: 10.1186/1471-2105-11-94
- Cernadas, R. A., Doyle, E. L., Niño-Liu, D. O., Wilkins, K. E., Bancroft, T., Wang, L., et al. (2014). Code-assisted discovery of TAL effector targets in bacterial leaf streak of rice reveals contrast with bacterial blight and a novel susceptibility gene. *PLoS Pathog.* 10:e1004126. doi: 10.1371/journal.ppat.1003972
- Chakrabarty, P. K., Chavhan, R. L., Ghosh, A., and Gabriel, D. W. (2010). Rapid and efficient protocols for throughput extraction of high quality plasmid DNA from strains of *Xanthomonas axonopodis* pv. *malvacearum* and *Escherichia coli*. *J. Plant Biochem. Biotechnol.* 19, 99–102. doi: 10.1007/BF03323444
- Chin, C. S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* 10, 563–569. doi: 10.1038/nmeth.2474
- Chu, Z., Yuan, M., Yao, J., Ge, X., Yuan, B., Xu, C., et al. (2006). Promoter mutations of an essential gene for pollen development result in disease resistance in rice. *Genes Dev.* 20, 1250–1255. doi: 10.1101/gad.1416306
- Darling, A. E., Mau, B., and Perna, N. T. (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE* 5:e11147. doi: 10.1371/journal.pone.0011147
- Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9, 772–772. doi: 10.1038/nmeth.2109
- De Lange, O., Binder, A., and Lahaye, T. (2014). From dead leaf, to new life: TAL effectors as tools for synthetic biology. *Plant J.* 78, 753–771. doi: 10.1111/tpj.12431
- Delport, W., Poon, A. F. Y., Frost, S. D. W., and Kosakovsky Pond, S. L. (2010). Datamonitor 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* 26, 2455–2457. doi: 10.1093/bioinformatics/btq429
- Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., Zhu, J. K., et al. (2012). Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* 335, 720–723. doi: 10.1126/science.1215670
- Doyle, E. L., Booher, N. J., Standage, D. S., Voytas, D. F., Brendel, V. P., Vandyk, J. K., et al. (2012). TAL effector-nucleotide targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. *Nucleic Acids Res.* 40, W117–W122. doi: 10.1093/nar/gks608
- 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
- Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., et al. (2009). Real-time DNA sequencing from single polymerase molecules. *Science* 323, 133–138. doi: 10.1126/science.1162986
- Ferreira, R. M., De Oliveira, A. C. P., Moreira, L. M., Belasque, J., Gourbeyre, E., Siguier, P., et al. (2015). A TALE of transposition: *Tn3*-like transposons play a major role in the spread of pathogenicity determinants of *Xanthomonas citri* and other xanthomonads. *MBio* 6, e02505–e02514. doi: 10.1128/mBio.02505-14
- Fu, Z., and Xu, Z. (1997). A comparison on the causal organism of rice short leaf streak with that of the bacterial leaf streak of rice and *Leersiae* spp. *Swartz. Acta Phytopathol. Sinica* 27, 225–230.
- Garcia-Bustos, J., Heitman, J., and Hall, M. N. (1991). Nuclear protein localization. *Biochim. Biophys. Acta* 1071, 83–101. doi: 10.1016/0304-4157(91)90013-m
- Goldman, N., Anderson, J. P., and Rodrigo, A. G. (2000). Likelihood-based tests of topologies in phylogenetics. *Syst. Biol.* 49, 652–670. doi: 10.1080/106351500750049752

- Gonzalez, C., Szurek, B., Manceau, C., Mathieu, T., Sere, Y., and Verdier, V. (2007). Molecular and pathotypic characterization of new *Xanthomonas oryzae* strains from West Africa. *Mol. Plant Microbe Interact.* 20, 534–546. doi: 10.1094/MPMI-20-5-0534
- Gu, K., Yang, B., Tian, D., Wu, L., Wang, D., Sreekala, C., et al. (2005). *R* gene expression induced by a type-III effector triggers disease resistance in rice. *Nature* 435, 1122–1125. doi: 10.1038/nature03630
- Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704. doi: 10.1080/10635150390235520
- Harris, E. Y., Ponts, N., Levchuk, A., Roch, K. L., and Lonardi, S. (2010). BRAT: bisulfite-treated reads analysis tool. *Bioinformatics* 26, 572–573. doi: 10.1093/bioinformatics/btp706
- Hu, Y., Zhang, J., Jia, H., Sosso, D., Li, T., Frommer, W. B., et al. (2014). Lateral organ boundaries 1 is a disease susceptibility gene for citrus bacterial canker disease. *Proc. Natl. Acad. Sci. U.S.A.* 111, E521–E529. doi: 10.1073/pnas.1313271111
- Huang, Z., Yeakley, J. M., Garcia, E. W., Holdridge, J. D., Fan, J.-B., and Whitham, S. A. (2005). Salicylic acid-dependent expression of host genes in compatible *Arabidopsis*-virus interactions. *Plant Physiol.* 137, 1147–1159. doi: 10.1104/pp.104.056028
- Hummel, A. W., Doyle, E. L., and Bogdanove, A. J. (2012). Addition of transcription activator-like effector binding sites to a pathogen strain-specific rice bacterial blight resistance gene makes it effective against additional strains and against bacterial leaf streak. *New Phytol.* 195, 883–893. doi: 10.1111/j.1469-8137.2012.04216.x
- Jain, R., Rivera, M. C., and Lake, J. A. (1999). Horizontal gene transfer among genomes: the complexity hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 96, 3801–3806. doi: 10.1073/pnas.96.7.3801
- Kawahara, Y., De La Bastide, M., Hamilton, J., Kanamori, H., Mccombie, W. R., Ouyang, S., et al. (2013). Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* 6:4. doi: 10.1186/1939-8433-6-4
- Kay, S., Hahn, S., Marois, E., Hause, G., and Bonas, U. (2007). A bacterial effector acts as a plant transcription factor and induces a cell size regulator. *Science* 318, 648–651. doi: 10.1126/science.1144956
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14:R36. doi: 10.1186/gb-2013-14-4-r36
- Kishino, H., and Hasegawa, M. (1989). Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *J. Mol. Evol.* 29, 170–179. doi: 10.1007/BF02100115
- Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H., and Frost, S. D. W. (2006a). Automated phylogenetic detection of recombination using a genetic algorithm. *Mol. Biol. Evol.* 23, 1891–1901. doi: 10.1093/molbev/msl051
- Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H., and Frost, S. D. W. (2006b). GARD: a genetic algorithm for recombination detection. *Bioinformatics* 22, 3096–3098. doi: 10.1093/bioinformatics/btl474
- Lee, B.-M., Park, Y.-J., Park, D.-S., Kang, H.-W., Kim, J.-G., Song, E.-S., et al. (2005). The genome sequence of *Xanthomonas oryzae* pathovar oryzae KACC10331, the bacterial blight pathogen of rice. *Nucleic Acids Res.* 33, 577–586. doi: 10.1093/nar/gki206
- Li, C., Wei, J., Lin, Y., and Chen, H. (2012a). Gene silencing using the recessive rice bacterial blight resistance gene *xa13* as a new paradigm in plant breeding. *Plant Cell Rep.* 31, 851–862. doi: 10.1007/s00299-011-1206-8
- Li, T., Liu, B., Spalding, M. H., Weeks, D. P., and Yang, B. (2012b). High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nat. Biotechnol.* 30, 390–392. doi: 10.1038/nbt.2199
- Lund, S. P., Nettleton, D., McCarthy, D. J., and Smyth, G. K. (2012). Detecting differential expression in RNA-sequence data using quasi-likelihood with shrunken dispersion estimates. *Stat. Appl. Genet. Mol. Biol.* 11, 1544–6115. doi: 10.1515/1544-6115.1826
- Mak, A. N., Bradley, P., Bogdanove, A. J., and Stoddard, B. L. (2012). TAL effectors: function, structure, engineering and applications. *Curr. Opin. Struct. Biol.* 23, 93–99. doi: 10.1016/j.sbi.2012.11.001
- Miller, J. C., Zhang, L., Xia, D. F., Campo, J. J., Ankoudinova, I. V., Guschin, D. Y., et al. (2015). Improved specificity of TALE-based genome editing using an expanded RVD repertoire. *Nat. Methods* 12, 465–471. doi: 10.1038/nmeth.3330
- Moreno-Hagelsieb, G., and Latimer, K. (2008). Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics* 24, 319–324. doi: 10.1093/bioinformatics/btm585
- Moscou, M. J., and Bogdanove, A. J. (2009). A simple cipher governs DNA recognition by TAL effectors. *Science* 326, 1501. doi: 10.1126/science.1178817
- Nettleton, D., Hwang, J. T. G., Caldo, R., and Wise, R. (2006). Estimating the number of true null hypotheses from a histogram of *p* values. *J. Agric. Biol. Environ. Stat.* 11, 337–356. doi: 10.1198/108571106X129135
- Nino-Liu, D. O., Ronald, P. C., and Bogdanove, A. J. (2006). *Xanthomonas oryzae* pathovars: model pathogens of a model crop. *Mol. Plant Pathol.* 7, 303–324. doi: 10.1111/j.1364-3703.2006.00344.x
- Ochiai, H., Inoue, Y., Takeya, M., Sasaki, A., and Kaku, H. (2005). Genome sequence of *Xanthomonas oryzae* pv. *oryzae* suggests contribution of large numbers of effector genes and insertion sequences to its race diversity. *Japan Agric. Res. Q.* 39, 275–287. doi: 10.6090/jarq.39.275
- Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., et al. (2014). The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Res.* 42, D206–D214. doi: 10.1093/nar/gkt1226
- Pereira, A., Carazzolle, M., Abe, V., De Oliveira, M., Domingues, M., Silva, J., et al. (2014). Identification of putative TAL effector targets of the citrus canker pathogens shows functional convergence underlying disease development and defense response. *BMC Genomics* 15:157. doi: 10.1186/1471-2164-15-157
- Politz, M. C., Copeland, M. F., and Pfleger, B. F. (2013). Artificial repressors for controlling gene expression in bacteria. *Chem. Commun.* 49, 4325–4327. doi: 10.1039/C2CC37107C
- Pond, S. L. K., and Frost, S. D. W. (2005). Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21, 2531–2533. doi: 10.1093/bioinformatics/bti320
- Pond, S. L. K., Frost, S. D. W., and Muse, S. V. (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21, 676–679. doi: 10.1093/bioinformatics/bti079
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna: The R Foundation for Statistical Computing.
- Richter, A., Streubel, J., Blücher, C., Szurek, B., Reschke, M., Grau, J., et al. (2014). A TAL effector repeat architecture for frameshift binding. *Nat. Commun.* 5:3447. doi: 10.1038/ncomms4447
- Roberts, A., Pimentel, H., Trapnell, C., and Pachter, L. (2011). Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* 27, 2325–2329. doi: 10.1093/bioinformatics/btr355
- Römer, P., Hahn, S., Jordan, T., Strauß, T., Bonas, U., and Lahaye, T. (2007). Plant pathogen recognition mediated by promoter activation of the pepper Bs3 resistance gene. *Science* 318, 645–648. doi: 10.1126/science.1144958
- Römer, P., Recht, S., Strauss, T., Elsaesser, J., Schornack, S., Boch, J., et al. (2010). Promoter elements of rice susceptibility genes are bound and activated by specific TAL effectors from the bacterial blight pathogen, *Xanthomonas oryzae* pv. *oryzae*. *New Phytol.* 187, 1048–1057. doi: 10.1111/j.1469-8137.2010.03217.x
- Salzberg, S. L., Sommer, D. D., Schatz, M. C., Phillippy, A. M., Rabinowicz, P. D., Tsuge, S., et al. (2008). Genome sequence and rapid evolution of the rice pathogen *Xanthomonas oryzae* pv. *oryzae* PXO99^A. *BMC Genomics* 9:204. doi: 10.1186/1471-2164-9-204
- Schmidt, H. A., Strimmer, K., Vingron, M., and Von Haeseler, A. (2002). TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 18, 502–504. doi: 10.1093/bioinformatics/18.3.502
- Shimodaira, H., and Hasegawa, M. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* 16, 1114. doi: 10.1093/oxfordjournals.molbev.a026201
- Streubel, J., Pesce, C., Hutin, M., Koebnik, R., Boch, J., and Szurek, B. (2013). Five phylogenetically close rice *SWEET* genes confer TAL effector-mediated susceptibility to *Xanthomonas oryzae* pv. *oryzae*. *New Phytol.* 200, 808–819. doi: 10.1111/nph.12411
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197

- Thieme, F., Koebnik, R., Bekel, T., Berger, C., Boch, J., Büttner, D., et al. (2005). Insights into genome plasticity and pathogenicity of the plant pathogenic bacterium *Xanthomonas campestris* pv. *vesicatoria* revealed by the complete genome sequence. *J. Bacteriol.* 187, 7254–7266. doi: 10.1128/JB.187.21.7254-7266.2005
- Tian, D., Wang, J., Zeng, X., Gu, K., Qiu, C., Yang, X., et al. (2014). The rice TAL effector-dependent resistance protein XA10 triggers cell death and calcium depletion in the endoplasmic reticulum. *Plant Cell* 26, 497–515. doi: 10.1105/tpc.113.119255
- 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 Ackerveken, G., Marois, E., and Bonas, U. (1996). Recognition of the bacterial avirulence protein AvrBs3 occurs inside the host plant cell. *Cell* 87, 1307–1316. doi: 10.1016/S0092-8674(00)81825-5
- Warnes, G., Bolker, B., Bonebakker, L., Gentleman, R., Liaw, W. H. A., Lumley, T., et al. (2015). *Gplots: Various R Programming Tools for Plotting Data*. Vienna: The R Foundation for Statistical Computing.
- Werner, J., and Gossen, M. (2014). Modes of TAL effector-mediated repression. *Nucleic Acids Res.* 42, 13061–13073. doi: 10.1093/nar/gku1124
- Wonni, I., Cottyn, B., Detemmerman, L., Dao, S., Ouedraogo, L., Sarra, S., et al. (2014). Analysis of *Xanthomonas oryzae* pv. *oryzicola* population in Mali and Burkina Faso reveals a high level of genetic and pathogenic diversity. *Phytopathology* 104, 520–531. doi: 10.1094/PHYTO-07-13-0213-R
- Wonni, I., Ouedraogo, L., and Verdier, V. (2011). First report of bacterial leaf streak caused by *Xanthomonas oryzae* pv. *oryzicola* on rice in Burkina Faso. *Plant Dis.* 95, 72–73. doi: 10.1094/PDIS-08-10-0566
- Wu, M., and Eisen, J. (2008). A simple, fast, and accurate method of phylogenomic inference. *Genome Biol.* 9:R151. doi: 10.1186/gb-2008-9-10-r151
- Yang, B., Sugio, A., and White, F. F. (2006). Os8N3 is a host disease-susceptibility gene for bacterial blight of rice. *Proc. Natl. Acad. Sci. U.S.A.* 103, 10503–10508. doi: 10.1073/pnas.0604088103
- Yang, J., Zhang, Y., Yuan, P., Zhou, Y., Cai, C., Ren, Q., et al. (2014). Complete decoding of TAL effectors for DNA recognition. *Cell Res.* 24, 628–631. doi: 10.1038/cr.2014.19
- Yashitola, J., Reddy, A. P. K., and Sonti, R. V. (2000). A widely distributed lineage of *Xanthomonas oryzae* pv. *oryzae* in India may have come from native wild rice. *Plant Dis.* 84, 465–469. doi: 10.1094/PDIS.2000.84.4.465
- Yu, X., Guda, K., Willis, J., Veigl, M., Wang, Z., Markowitz, S., et al. (2012). How do alignment programs perform on sequencing data with varying qualities and from repetitive regions? *BioData Min.* 5:6. doi: 10.1186/1756-0381-5-6
- Yu, Y., Streubel, J., Balzergue, S., Champion, A., Boch, J., Koebnik, R., et al. (2011). Colonization of rice leaf blades by an African strain of *Xanthomonas oryzae* pv. *oryzae* depends on a new TAL effector that induces the rice nodulin-3 Os11N3 gene. *Mol. Plant Microbe Interact.* 24, 1102–1113. doi: 10.1094/MPMI-11-1-0254
- Yuan, M., Chu, Z., Li, X., Xu, C., and Wang, S. (2009). Pathogen-induced expressional loss of function is the key factor in race-specific bacterial resistance conferred by a recessive *R* gene *xa13* in rice. *Plant Cell Physiol.* 50, 947–955. doi: 10.1093/pcp/pcp046
- Zeng, X., Lai, W., and Xu, D. (1995). Serological specificity of leaf streak pathogen of rice. *J. S. China Agric. Univ.* 16, 65–68.
- Zeng, X., Tian, D., Gu, K., Zhou, Z., Yang, X., Luo, Y., et al. (2015). Genetic engineering of the *Xa10* promoter for broad-spectrum and durable resistance to *Xanthomonas oryzae* pv. *oryzae*. *Plant Biotechnol. J.* doi: 10.1111/pbi.12342. [Epub ahead of print].
- Zhu, W., Yang, B., Chittoor, J. M., Johnson, L. B., and White, F. F. (1998). AvrXa10 contains an acidic transcriptional activation domain in the functionally conserved C-terminus. *Mol. Plant Microbe Interact.* 11, 824–832. doi: 10.1094/MPMI.1998.11.8.824

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 Wilkins, Booher, Wang and Bogdanove. 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.

MorTAL Kombat: the story of defense against TAL effectors through loss-of-susceptibility

Mathilde Hutin^{1†}, Alvaro L. Pérez-Quintero^{1†}, Camilo Lopez^{1,2} and Boris Szurek^{1*}

¹ UMR IPME, Institut de Recherche Pour le Développement, IRD-CIRAD-Université Montpellier 2, Montpellier, France,

² Biology Department, Universidad Nacional de Colombia, Bogota, Colombia

OPEN ACCESS

Edited by:

Laurent D. Noël,
Centre National de la Recherche
Scientifique, France

Reviewed by:

Sebastian Schornack,
University of Cambridge, UK
Adam Bogdanove,
Cornell University, USA

*Correspondence:

Boris Szurek,
UMR IPME, Institut de Recherche
Pour le Développement,
IRD-CIRAD-Université Montpellier 2,
911 Avenue Agropolis BP 64501,
34394 Montpellier Cedex 5, France
boris.szurek@ird.fr

[†]These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Plant Biotic Interactions,
a section of the journal
Frontiers in Plant Science

Received: 16 April 2015

Accepted: 30 June 2015

Published: 14 July 2015

Citation:

Hutin M, Pérez-Quintero AL, Lopez C and Szurek B (2015) MorTAL Kombat:
the story of defense against TAL
effectors through
loss-of-susceptibility.
Front. Plant Sci. 6:535.
doi: 10.3389/fpls.2015.00535

Many plant-pathogenic xanthomonads rely on Transcription Activator-Like (TAL) effectors to colonize their host. This particular family of type III effectors functions as specific plant transcription factors via a programmable DNA-binding domain. Upon binding to the promoters of plant disease susceptibility genes in a sequence-specific manner, the expression of these host genes is induced. However, plants have evolved specific strategies to counter the action of TAL effectors and confer resistance. One mechanism is to avoid the binding of TAL effectors by mutations of their DNA binding sites, resulting in resistance by loss-of-susceptibility. This article reviews our current knowledge of the susceptibility hubs targeted by *Xanthomonas* TAL effectors, possible evolutionary scenarios for plants to combat the pathogen with loss-of-function alleles, and how this knowledge can be used overall to develop new pathogen-informed breeding strategies and improve crop resistance.

Keywords: *Xanthomonas*, plant disease susceptibility S genes, hubs, TAL effectors, agricultural biotechnology, loss-of-function alleles

Introduction

Xanthomonas are a genus of plant pathogenic bacteria that cause devastating diseases on a wide range of hosts, leading to severe impact on yield quantity and quality of important crops such as rice, cassava, cotton, wheat, banana, mango, citrus, and cabbage (Schornack et al., 2013). In order to colonize their host, most of these pathogens rely on a type three secretion system (T3SS) specialized in the injection of virulence factors into the host cell. Also called type three effectors (T3Es), these proteins collectively promote the pathogen's adaptation to specific host tissues, species, and genotypes. This adaptation includes suppression of plant immunity, nutrient acquisition, dispersal, or other virulence-related processes that benefit the pathogen. Upon translocation, T3Es localize to various subcellular compartments such as the plasma membrane or organelles.

The Transcription Activator-Like (TAL) effectors form a particular family of T3Es that act as bona fide plant transcription factors able to reprogram the host transcriptome following nuclear localization. They have a highly conserved structure that is modular and characterized by an N-terminal type three secretion signal, a C-terminal region bearing two to three nuclear localization signals and an acidic transcription activation domain. In between is a central region built of quasi-identical tandem direct repeats of 33–35 amino-acids forming a unique type of DNA-binding domain. Each repeat folds into a super-helical structure that wraps the DNA. Each repeat forms a two-helix bundle in which so-called repeat variable di-residues (RVDs) at positions 12 and 13 reside on the interhelical loop that projects into the DNA major groove. More precisely, RVDs

mediate the binding of TAL effectors to the double-strand DNA in a sequence-specific manner with residue 12 stabilizing the overall interaction while residue 13 specifically contacts the DNA base (reviewed in Mak et al., 2013). Functional differences between TAL effectors are therefore mainly due to the nature of the string of RVDs that determine the sequence of the so-called TAL effector binding element (EBE) in host promoters. TAL effectors can be found in most *Xanthomonas* species and related proteins are also present in *Ralstonia solanacearum* (RipTALs; de Lange et al., 2013) and *Burkholderia rhizoxinica* (Bats; Juillerat et al., 2014; Lange et al., 2014).

Several *Xanthomonas* TAL effectors are known to play an essential role during infection, i.e., the deletion of the genes encoding them leads to reduced disease. Plant genes may contribute to disease susceptibility in different ways, notably by favoring pathogen attraction and attachment to the host, controlling the establishment of a favorable environment for host tissue penetration, colonization, or pathogen dispersal (reviewed by Lapin and Van den Ackerveken, 2013). In the case of plant pathogenic bacteria our knowledge on the diversity and function of disease determinants is still scarce. However, recent breakthroughs in the field of TAL effector biology are significantly improving our understanding of bacterial disease processes, thereby fostering innovative disease control strategies. In this review, we analyze host-pathogen interactions over evolutionary time using the popular video game Mortal Kombat¹, in which combatants strike and counter to survive an epic battle, as a metaphor. We initially focus on the mechanisms underlying susceptibility caused by TAL effectors. Then we explore the strategies that plants have evolved to counter the action of TAL effectors, with emphasis on the strategy of “resistance through loss-of-susceptibility” achieved by avoiding the binding of TAL effectors to host DNA using mutations of their binding sites. Finally, we discuss how, in light of these findings, natural or engineered resistance can be bred to control TAL effector-mediated diseases.

Round 1: Bacteria Attack Using TAL Effectors to Induce Susceptibility Genes

Transcription Activator-Like effectors are found in many plant pathogenic *Xanthomonas* species where they significantly contribute to disease development but their exact function as minor or major virulence factors has been disentangled for only a few of them (Boch and Bonas, 2010). The number of TAL effectors genes contained in the genome of a *Xanthomonas* strain is highly variable between strains. For example, strains of *Xanthomonas oryzae* pv. *oryzae* (*Xoo*) contain from 9 up to more than 15 different TAL effectors. Yet it has been shown that only one or two of them play an essential role in pathogenicity and encode major virulence factors (Yang and White, 2004). For these strains the mutation of the corresponding genes leads to no development of the disease (Yang and White, 2004). The genes targeted by these TAL effectors are called (*S*) susceptibility genes

because their induction is essential to the complete development of the disease. Unlike resistance (*R*) genes that are usually expressed only in the presence of the pathogen, most of the known *S* genes play roles in plant development, and are exploited by pathogens through their overexpression during the infection. The discovery of the TAL effector-DNA binding code combined with transcriptomic data has led to the identification of a *S* gene list for different *Xanthomonas*/host interactions (Table 1). Notably, many of these *S* genes encode either transporters (sugar or sulfate) or transcription factors, and their induction is hypothesized to facilitate bacterial colonization and symptom development.

The SWEET family of sugar transporters represents the best characterized group of *S* genes induced by TAL effectors (Chen, 2014). In rice, these include *OsSWEET11* (*a.k.a Xa13/xa13*) activated by PthXo1 from strain *Xoo* PXO99^A, *OsSWEET13* (*a.k.a Xa25/xa25*) activated by PthXo2 from *Xoo* strains JXO1^A and MAFF311018, and *OsSWEET14* activated by AvrXa7, PthXo3, TalC, and Tal5 from *Xoo* strains PXO86, PXO61, BAI3, and MAI1, respectively (Yang and White, 2004; Chu et al., 2006; Antony et al., 2010; Yu et al., 2011; Streubel et al., 2013; Zhou et al., 2015). They are all members of clade III of the *SWEET* gene family. It was shown that in rice, all five members of clade III can act as a major *S* gene (Li et al., 2013; Streubel et al., 2013). Interestingly, related genes are also induced by *Xanthomonas* TAL effectors in other hosts. For instance *MeSWEET10a* and *CsSWEET1* are, respectively, induced by TAL20 from *X. axonopodis* pv. *manihotis* (*Xam*) in cassava and the PthA series from *X. citri* ssp. *citri* in citrus (Cohn et al., 2014; Hu et al., 2014). *Xam*-mediated induction of *MeSWEET10a* which also belongs to clade III is essential for full development of watersoaking as infiltration of strain *Xam*668ΔTAL20 leads to reduced symptoms (Cohn et al., 2014). Intriguingly the induction of *MeSWEET10a* is necessary for symptom expansion but not for bacterial growth. In contrast, *CsSWEET1* belongs to clade I and no function for citrus canker has been demonstrated to date (Hu et al., 2014). Similarly, AvrBs3 from *X. axonopodis* pv. *vesicatoria* induces *UPA16* from pepper which also belongs to the *SWEET* family but there is no evidence for this gene to play a role for bacterial spot of pepper and tomato (Kay et al., 2009).

At least two members of the *SWEET* family in rice were shown to play a crucial role in plant development. Homozygous lines containing a T-DNA insertion in *OsSWEET14* display delayed growth and seeds smaller than the wild type plant (Antony et al., 2010), and plants silenced for *OsSWEET11* have reduced fertility (Yang et al., 2006). They encode sugar transporters that allow the efflux of carbohydrates from the cell to the intercellular space. Thus, their induction could make carbohydrates easily available for bacteria (Chen et al., 2012), though to date it has not been clearly demonstrated that these carbohydrates help bacterial growth directly. *SWEET* genes may play a role in multiple plant pathogen interactions (Lemoine et al., 2013). *VvSWEET4* in grapevine is up regulated during the interaction with the gray mold fungus *Botrytis cinerea* and a knockout mutant in *Arabidopsis* for the orthologous gene *AtSWEET4* becomes less susceptible to this pathogen (Chong et al., 2014). Also in *Arabidopsis*, differential expression of *SWEET* genes is

¹<http://www.mortalkombat.com/>, Warner Bros. Interactive Entertainment.

TABLE 1 | Plant susceptibility or candidate S genes targeted by *Xanthomonas* Transcription Activator-Like (TAL) effectors.

| Target function | Target name | TAL name | Species | Host | Symptoms | Effect | Reference | Loss S alleles | Reference |
|----------------------|-------------|-------------|---------|---------------|------------------------------|--|--|----------------|--|
| Transcription factor | UPA20 | AvrBs3 | Xav | Pepper | Cell hypertrophy | Increased egress to the leaf surface | Kay et al. (2007) | | |
| | CsLOB1 | PthA series | Xcc | Citrus | Cell expansion | Canker | Hu et al. (2014), Li et al. (2014) | | |
| | PthB | Xca | Citrus | Canker | | | Al-Saadi et al. (2007), Hu et al. (2014) | | |
| | PthC | Xca | Citrus | Canker | | | | | |
| | PthX06 | Xoo | Rice | Water soaking | Increased growth and lesions | | Sugio et al. (2007) | | |
| TFIIA/Y | PthXo7 | Xoo | Rice | Water soaking | Increased growth and lesions | | Hu et al. (2014) | | |
| Sugar Transporter | CsSWEET1 | PthA series | Xcc | Unknown | Unknown | | Chu et al. (2006), Yang et al. (2006) | xat13 | Chu et al. (2006), Yang et al. (2006) |
| | OssWEET11 | PthXo1 | Xoo | Rice | Water soaking | Increased growth and lesions | Chu et al. (2006), Yang et al. (2006) | | |
| | OssWEET13 | PthXo2 | Xoo | Rice | Water soaking | Increased growth and lesions | Zhou et al. (2015) | xat25 | Zhou et al. (2011), Zhou et al. (2015) |
| | OssWEET14 | PthXo3 | Xoo | Rice | Water soaking | Increased growth and lesions | Yang and White (2004) | | Li et al. (2012) |
| | | | | | | | | | |
| | AvrXa7 | Xoo | Rice | Water soaking | Increased growth and lesions | | Antony et al. (2010) | | |
| | Tal5 | Xoo | Rice | Water soaking | Increased growth and lesions | | Streubel et al. (2013) | | |
| | TalC | Xoo | Rice | Water soaking | Increased growth and lesions | | Yu et al. (2011) | | |
| | MesWEET10a | TAL20 | Xam | Cassava | Unknown | | Cohn et al. (2014) | | |
| Sulfate transporter | SULTR3;6 | Tal2g | Xoc | Rice | Water soaking | Increased lesions and egress to the leaf surface | Cernadas et al. (2014) | | |
| miRNA stability | OstHEN1 | PthXo8 | Xoo | Rice | | Not reported | Moscou and Bogdanove (2009) | | |
| | | Tal1C | Xoc | Rice | | None detected | | | |

Xcv, *Xanthomonas axonopodis* pv. *vesicatoria*; Xcc, *X. citri* pv. *citri*; Xca, *X. citri* pv. *aurantifoliae*; Xoo, *X. oryzae* pv. *oryzicola*; Xam, *X. axonopodis* pv. *manihotis*.

observed after infection by *Pseudomonas syringae* pv. tomato, *Golovinomyces cichoracearum*, and *Plamodiophora brassicae* (Siemens et al., 2006; Chen et al., 2010) suggesting that numerous pathogens target the sugar transport mechanisms of the host upon infection.

While *SWEET* genes are essential for the vascular pathogen *Xoo* to cause Bacterial Blight in rice, they are curiously not targeted by the closely related non-vascular pathovar *X. oryzae* pv. *oryzicola* (*Xoc*) that is responsible for bacterial leaf streak of rice. It is then tempting to speculate that induction of *S* genes by TAL effectors may explain tissue specificity to some extent. However, transformation of *Xoc* with *OsSWEET*-inducing TAL effectors from *Xoo* does not allow vascular colonization by *Xoc* (Verdier et al., 2012). Nevertheless, the role of TAL effectors in defining host-ranges and colonization mechanisms is still a field worth exploring. Recently, the systematic mutagenesis of each TAL effector gene of *Xoc* strain BLS256 completed with the functional analysis of their potential targets unmasked a new *S* gene specifically required for leaf streak. *OsSULTR3;6* encodes a sulfate transporter targeted by *Tal2g* and its induction is necessary for lesion expansion and bacterial exudation but not for bacterial growth (Cernadas et al., 2014). How this sulfate transporter contributes to disease has yet to be elucidated. One hypothesis is that its induction modifies the redox potential in a favorable manner for symptom establishment and bacterial expansion. Some orthologous genes have been identified to play a similar role in other pathosystems and in establishment of symbiosis (Cernadas et al., 2014).

Another group of functionally related genes targeted by TAL effectors are transcription factors. The first identified was *UPA20* in pepper. This transcription factor induced by *AvrBs3* from *X. axonopodis* pv. *vesicatoria* belongs to the *bHLH* family and trans-activates the secondary target *UPA7* (Kay et al., 2007). The expression of *UPA7* leads to enlargement of mesophyll cells of pepper leaves that could promote propagation of bacteria (Kay et al., 2007). Another transcription factor acting as a major susceptibility gene is *CsLOB1*, required for pustule formation during citrus bacterial canker. This gene is targeted by five different TAL effectors, including *PthA4*, *PthA^W*, and *PthA^{*}* from *X. citri* ssp. *citri* and *PthB* and *PthC* from *X. citri* ssp. *aurantifolia* (Hu et al., 2014). Interestingly an *Arabidopsis* protein belonging to the LOB family is described as a susceptibility gene for the fungal pathogen *Fusarium oxysporum* (Thatcher et al., 2009). The specific functions in development of *CsLOB1* are not known but it was shown that proteins with LOB domain are responsive to numerous phytohormones, play a role in anthocyanin and nitrogen metabolism and are involved in the regulation of lateral organ development (Majer and Hochholdinger, 2011; Gendron et al., 2012). *OsTFXI* a bZIP transcription factor and *OsTFIIAγ1* a general transcription factor, are targeted, respectively, by *PthXo6* and *PthXo7* from *Xoo* strain PXO99^A (Sugio et al., 2007). *OsTFXI* is particularly interesting as it is induced by all *Xoo* strains assessed to date (White and Yang, 2009). In contrast to the wild type, a transgenic plant overexpressing *OsTFXI* is fully susceptible to a *pthXo6* mutant confirming that it is an *S* gene (Sugio et al., 2007). How *OsTFXI* promotes disease is not yet understood.

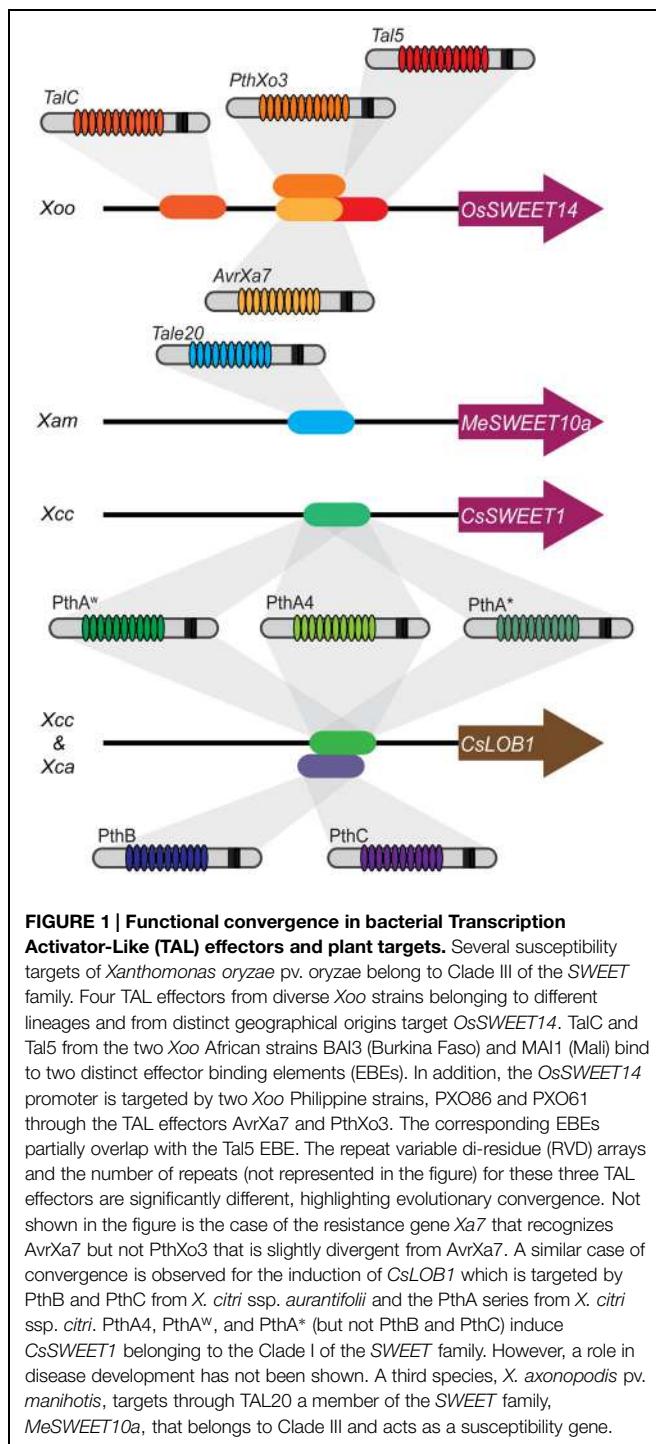
In summary, for *Xoo* species, the *SWEET* transporters are a major target in rice and among them *OsSWEET14* is targeted by four different TALs from genetically and geographically distant *Xoo* strains. The *SWEET* genes are targeted by at least three other *Xanthomonas* species in three other hosts. There is undoubtedly an evolutionary convergence for the induction of this family inter species and intra pathovar. This convergence occurs also for the transcription factors. Five different TAL effectors with different RVD content are able to induce *CsLOB1*, which seems to be the single susceptibility gene induced by both *X. citri* pathovars for the establishment of citrus canker (Hu et al., 2014). Thus, convergence seems to be a good indicator for the discovery of *S* genes (Figure 1). A similar behavior of convergence toward a selected group of targets in the host has also been found for non-TAL effectors from different pathogen groups in *Arabidopsis* (Mukhtar et al., 2011). In addition, computational predictions suggest that targeting of functional “hubs” is indeed a common feature for TAL effectors (Perez-Quintero et al., 2013).

Taking this into account, new potential *S* genes arise, including for example *OsHEN1* in rice. *OsHEN1* encodes a methyl transferase and is the only common target of *Xoo* and *Xoc* via, respectively, *PthXo8* and *Tal1c* (Moscou and Bogdanove, 2009). Although a *tal1c* mutant showed no reduction in virulence (Cernadas et al., 2014), and data for *PthXo8* are yet to be reported, the convergence of different pathovars onto this target may suggest an important role in the respective diseases. *OsHEN1* plays important roles in different physiological processes through the stabilization of small RNAs, introducing the exciting question of how induction of this gene aids in bacterial colonization. On the other hand, there are some TAL effector for which a role in pathogenicity has been shown but the target remains unknown or unconfirmed (Yang et al., 1994; Castiblanco et al., 2012; Munoz-Bodnar et al., 2014).

Round 2: Plant Dodges Bacteria's TAL Effectors Through Mutation in Promoters of S Genes

The evolutionary convergence of TAL effector activity on a restricted set of key host genes of physiological importance (Figure 1) implies that some allelic variants (those that can avoid binding on the effector) could confer broad-spectrum resistance to *Xanthomonas*. To date, EBE variants that abolish TAL effector binding have been found in the promoter of two major susceptibility genes: *OsSWEET11* and *OsSWEET13*, resulting in the recessive resistance alleles known, respectively, as *xa13* and *xa25* (Chu et al., 2006; Yang et al., 2006; Liu et al., 2011; Zhou et al., 2015).

The promoter of the susceptible allele of *OsSWEET11:xa13* but not of *xa13* is induced directly by the TAL effector *PthXo1* from PXO99^A (Yang et al., 2006; Romer et al., 2010). Varieties containing *xa13* are resistant to PXO99^A and the recessive resistance observed is not characterized by the typical hypersensitive response phenotype as it results only on the loss of the induction of a gene essential to disease development (Yuan et al., 2009, 2011). There is a collection of rice varieties



naturally presenting an *xa13* allele. All these alleles differ from the *Xa13* genotype by insertion, deletion, or substitution in the *PthXo1* EBE (Romer et al., 2010). Interestingly a single substitution in the second nucleotide of the EBE is sufficient to avoid induction of *OsSWEET11* by *PthXo1*. In the case of *OsSWEET13*, the susceptible allele *Xa25* but not the race specific recessive resistance gene *xa25* is induced by the *Xoo* strain PXO339 (Liu et al., 2011). Transformation of *Xa25* in a resistant

variety with a *xa25* genotype led to susceptibility to PXO339 (Liu et al., 2011). Recently it was shown that *PthXo2* directly induces *Xa25*. *Xa25* differs from *xa25* by a one nucleotide deletion in the EBE of *PthXo2*. Promoter mutation generated by CRISPR/Cas9 technology confirmed that *Ossweet13* induction is essential for strains that depend on *PthXo2* to lead to disease susceptibility (Zhou et al., 2015).

For *OsSWEET14*, no naturally occurring resistance alleles have been reported yet, but they have been genetically engineered. Li et al. (2012) generated loss of susceptibility alleles altered in the box targeted by *AvrXa7* and *PthXo3* using the TALEN strategy, an approach based on TAL effector binding domains coupled with nucleases (reviewed in Bogdanove and Voytas, 2011; Carroll, 2014). They showed that a deletion of 4 bp is sufficient for loss-of-susceptibility to *AvrXa7* and *PthXo3*. As discussed below Round 4), this genome editing strategy could be extensively used to develop resistant varieties to *Xanthomonas* in various systems.

It is expected that more naturally occurring loss-of-susceptibility alleles can be found by promoter screening of *S* genes in different host germplasm collections (see Round 4). An example of a resource for this type of screening is the recently released SNP data from rice obtained from the sequencing of the genomes of 3000 rice varieties, available in the SNPseek database (Alexandrov et al., 2015). Surprisingly when using this data to look for SNPs in a set of known EBEs in rice (not all from *S* genes) very few SNPs are found within the EBEs (Table 2). It is worth noting, that no indels are included in the SNPseek data and that all the SNPs were obtained from mapping NGS reads onto only one reference genome (Alexandrov et al., 2015). So it is possible that more SNPs in these regions will be found using different tools and datasets.

Nonetheless, this raises the question: **how common are loss-of-susceptibility alleles of *S* genes?**

Gaining resistance through loss-of-susceptibility alleles requires the plant to mutate a region of the promoter to avoid binding of the TAL effector without altering physiologically important regulatory elements in the promoter. This might be particularly hard to achieve given that many TAL effectors seem to bind to crucial elements of the promoter. As a matter of fact, most of the known functional binding sites for TAL effectors are located in a region overlapping with that of core promoter elements of the gene (between 300 bp upstream and 200 bp downstream of the transcription start site). Indeed, TAL effectors seem to bind predominantly to functional motifs, such as the TATA box and the TC box (Grau et al., 2013; Pereira et al., 2014).

Regulatory elements in plant promoters are usually very short; median observed length of 8 bp and notoriously conserved (Reineke et al., 2011). Point mutations (transversions and transitions) in these elements may prevent binding of endogenous plant TFs and probably will not be selected unless they offer a great improvement in fitness (i.e., resistance to pathogenic bacteria). However, point mutations in an EBE could fail to prevent binding since the way TAL effectors bind to DNA allows for enough flexibility to tolerate “mismatches”. For instance some RVDs are able to indiscriminately accommodate different nucleotides, particularly those RVDs with weaker interactions with DNA (Streubel et al., 2012). Also, RVDs in

TABLE 2 | SNPs found in 3000 rice genomes at the EBEs of known TAL effector targets in rice.

| TAL | Xo Strain | Reference | Gene | Effector binding element (EBE) sequence | SNPs EBE | SNPs promoter | SNPs gene |
|--------|-----------|-----------------------------|---------------------------|--|----------|---------------|-----------|
| PtbXo3 | PXO61 | Yang and White (2004) | LOC_Os11g31190 (SWEET14) | TATATAAACCC <u>C</u> TOCAACCAGGT <u>G</u> TAAG | 2 | 20 | 28 |
| AvrXa7 | PXO86 | Antony et al. (2010) | LOC_Os11g31190 (SWEET14) | TATATAAACCC <u>C</u> TOCAACCAGGT <u>G</u> <u>C</u> T | 2 | 20 | 28 |
| Ta13b | BLS256 | Cernadas et al. (2014) | LOC_Os07g36430 | TATAAAGAAGCCT <u>C</u> ACTCG | 1 | 22 | 3 |
| Ta19b | BLS256 | | LOC_Os01g151040 | TAGAAAAC <u>T</u> GCTCTTCCT | 1 | 2 | 22 |
| Ta1C | BA13 | Yu et al. (2011) | LOC_Os11g31190 (SWEET14) | CATGGAATGTCAGCAGCTGGTCAT | 0 | 20 | 28 |
| Ta15 | MA11 | Streubel et al. (2013) | LOC_Os11g31190 (SWEET14) | TAAGCTCATCAAGCCTICA | 0 | 20 | 28 |
| PtbXo6 | PXO99A | Sugio et al. (2007) | LOC_Os09g29820 (OsTFX1) | TATAAAAAGGCCCTCACCAACCCAT | 0 | 5 | 3 |
| PtbXo7 | PXO99A | Yang et al. (2006) | LOC_Os01g73890 (OsTFIAy1) | TATAATTCCCAAATCCCTCTCTC | 0 | 8 | 8 |
| PtbXo1 | PXO99A | Moscou and Bogdanove (2009) | LOC_Os08g42350 (SWEET11) | TGCATCTCCCOCTACTGTACACAC | 0 | 8 | 17 |
| Ta19A | PXO99A | Moscou and Bogdanove (2009) | LOC_Os07g06970 (HEN1) | TCCCTTCCCTAAACCCCACTT | 0 | 4 | 20 |
| Ta1c | BLS256 | Moscou and Bogdanove (2009) | LOC_Os07g06970 (HEN1) | TCCCCCTCGTGCTTCCCT | 0 | 4 | 20 |
| Ta1c | BLS256 | Cernadas et al. (2014) | LOC_Os03g03034 | TCGGGCCCTCTCCCCCCCCACCTGAC | 0 | 5 | 25 |
| Ta1d | BLS256 | | LOC_Os04g49194 | TATTCCTCTGGGTGATC | 0 | 18 | 78 |
| Ta13b | BLS256 | | LOC_Os02g34970 | TATAAGTAGCAGTCGCTCA | 0 | 4 | 9 |
| Ta13b | BLS256 | | LOC_Os05g27590 | TATAAAATAGCCCTCACTCT | 0 | 3 | 0 |
| Ta14a | BLS256 | | LOC_Os03g37840 | TATATATCTCGTTOAGGCCAGGGCCCT | 0 | 12 | 28 |
| Ta13c | BLS256 | | LOC_Os03g07540 | TACACGGTTCCTCCTCACC | 0 | 3 | 9 |
| Ta13c | BLS256 | | LOC_Os02g44760 | TACACATTAGCTTACCAT | 0 | 2 | 11 |
| Ta16 | BLS256 | | LOC_Os09g29100 | TACAAAAGAGAACGCTCCCCC | 0 | 11 | 24 |
| Ta16 | BLS256 | | LOC_Os12g42970 | TAAAAAGCGAAACCCCTCCCTCC | 0 | 16 | 8 |
| Ta15a | BLS256 | | LOC_Os02g15290 | TAATATCCCCCTGCACAGGAAAGCT | 0 | 31 | 2 |
| Ta12g | BLS256 | | LOC_Os01g52130 | TGGCCCGTAGCCCTCTCT | 0 | 0 | 89 |
| Ta12g | BLS256 | | LOC_Os06g46500 | TGGCAAGTGACCTCAGCT | 0 | 9 | 14 |
| Ta14c | BLS256 | | LOC_Os06g37080 | TATAAAAACCTGACAAAGCCTCTCT | 0 | 7 | 46 |
| Ta14b | BLS256 | | LOC_Os09g32100 | TGTATAGTATCCCCCT | 0 | 14 | 14 |

SNPs found elsewhere in the promoter regions (1 kb upstream TSS) and the genes (including intron) are also indicated. Position of the SNPs in the EBE sequences is indicated by bold and underlined nucleotides. SNP data were obtained from the Rice SNPseek database (Alexandrov et al., 2015).

TABLE 2 | SNPs found in 3000 rice genomes at the EBEs of known TAL effector targets in rice.

| TAL | Xo Strain | Reference | Gene | Effector binding element (EBE) sequence | SNPs EBE | SNPs promoter | SNPs gene |
|--------|-----------|-----------------------------|---------------------------|--|----------|---------------|-----------|
| PtbXo3 | PXO61 | Yang and White (2004) | LOC_Os11g31190 (SWET14) | TATATAAACCC <u>C</u> TOCAACCAGGT <u>G</u> TAAG | 2 | 20 | 28 |
| AvrXa7 | PXO86 | Antony et al. (2010) | LOC_Os11g31190 (SWET14) | TATATAAACCC <u>C</u> TOCAACCAGGT <u>G</u> <u>C</u> T | 2 | 20 | 28 |
| Ta13b | BLS256 | Cernadas et al. (2014) | LOC_Os07g36430 | TATAAAGAAGCCT <u>A</u> CTCG | 1 | 22 | 3 |
| Ta19b | BLS256 | | LOC_Os01g1040 | TAGAAAAC <u>T</u> GCTTCCT | 1 | 2 | 22 |
| Ta1C | BA13 | Yu et al. (2011) | LOC_Os11g31190 (SWET14) | CATGGAATGTCAGCAGCTGGTCAT | 0 | 20 | 28 |
| Ta1S | MA11 | Streubel et al. (2013) | LOC_Os11g31190 (SWET14) | TAAGCTCATCAAGCCTICA | 0 | 20 | 28 |
| PtbXo6 | PXO99A | Sugio et al. (2007) | LOC_Os09g29820 (OsTFX1) | TATAAAAGGCCCTCACCAACCCAT | 0 | 5 | 3 |
| PtbXo7 | PXO99A | Yang et al. (2006) | LOC_Os01g73890 (OsTFIAy1) | TATAATTCCCAAATCCCTCTCTC | 0 | 8 | 8 |
| PtbXo1 | PXO99A | Moscou and Bogdanove (2009) | LOC_Os08g42350 (SWET11) | TGCATCTCCCOCTACTGTACACAC | 0 | 8 | 17 |
| Ta19A | PXO99A | Moscou and Bogdanove (2009) | LOC_Os07g06970 (HEN1) | TCCCTTCCCTAAACCCCACTT | 0 | 4 | 20 |
| Ta1C | BLS256 | Moscou and Bogdanove (2009) | LOC_Os07g06970 (HEN1) | TCCCCCTOGCTTGGCTT | 0 | 4 | 20 |
| Ta1C | BLS256 | Cernadas et al. (2014) | LOC_Os03g03034 | TCGGGCCCTCTCCCCCCCCACCTGAC | 0 | 5 | 25 |
| Ta1d | BLS256 | | LOC_Os04g49194 | TATTCCCTCGGCTGATC | 0 | 18 | 78 |
| Ta13b | BLS256 | | LOC_Os02g34970 | TATAAGTAGCAGTCGCTCA | 0 | 4 | 9 |
| Ta13b | BLS256 | | LOC_Os05g27590 | TATAAAATAGCCCTCACTCT | 0 | 3 | 0 |
| Ta14a | BLS256 | | LOC_Os03g37840 | TATATATCTCGTOAGGCCAGGCCCT | 0 | 12 | 28 |
| Ta13c | BLS256 | | LOC_Os03g07540 | TACACGGTCCCCTCACC | 0 | 3 | 9 |
| Ta13c | BLS256 | | LOC_Os02g47660 | TACACATTAGCTTACCAT | 0 | 2 | 11 |
| Ta16 | BLS256 | | LOC_Os09g29100 | TACAAAAGAGAACGCATCCCCC | 0 | 11 | 24 |
| Ta16 | BLS256 | | LOC_Os12g42970 | TAAAAAGCGAAACCCCTCCCTCC | 0 | 16 | 8 |
| Ta15a | BLS256 | | LOC_Os02g15290 | TAATACTCCCCCTGCACAGGAAAGCT | 0 | 31 | 2 |
| Ta12g | BLS256 | | LOC_Os01g2130 | TGGCCCGTAGCCCTCTCT | 0 | 0 | 89 |
| Ta12g | BLS256 | | LOC_Os06g46500 | TGGCAAGTGACCTCAGCT | 0 | 9 | 14 |
| Ta14c | BLS256 | | LOC_Os06g37080 | TATAAAAACCTGACAAAGCCTCTCT | 0 | 7 | 46 |
| Ta14b | BLS256 | | LOC_Os09g32100 | TGTATAGTATCCCCCT | 0 | 14 | 14 |

SNPs found elsewhere in the promoter regions (1 kb upstream TSS) and the genes (including intron) are also indicated. Position of the SNPs in the EBE sequences is indicated by bold and underlined nucleotides. SNP data were obtained from the Rice SNPseek database (Alexandrov et al., 2015).

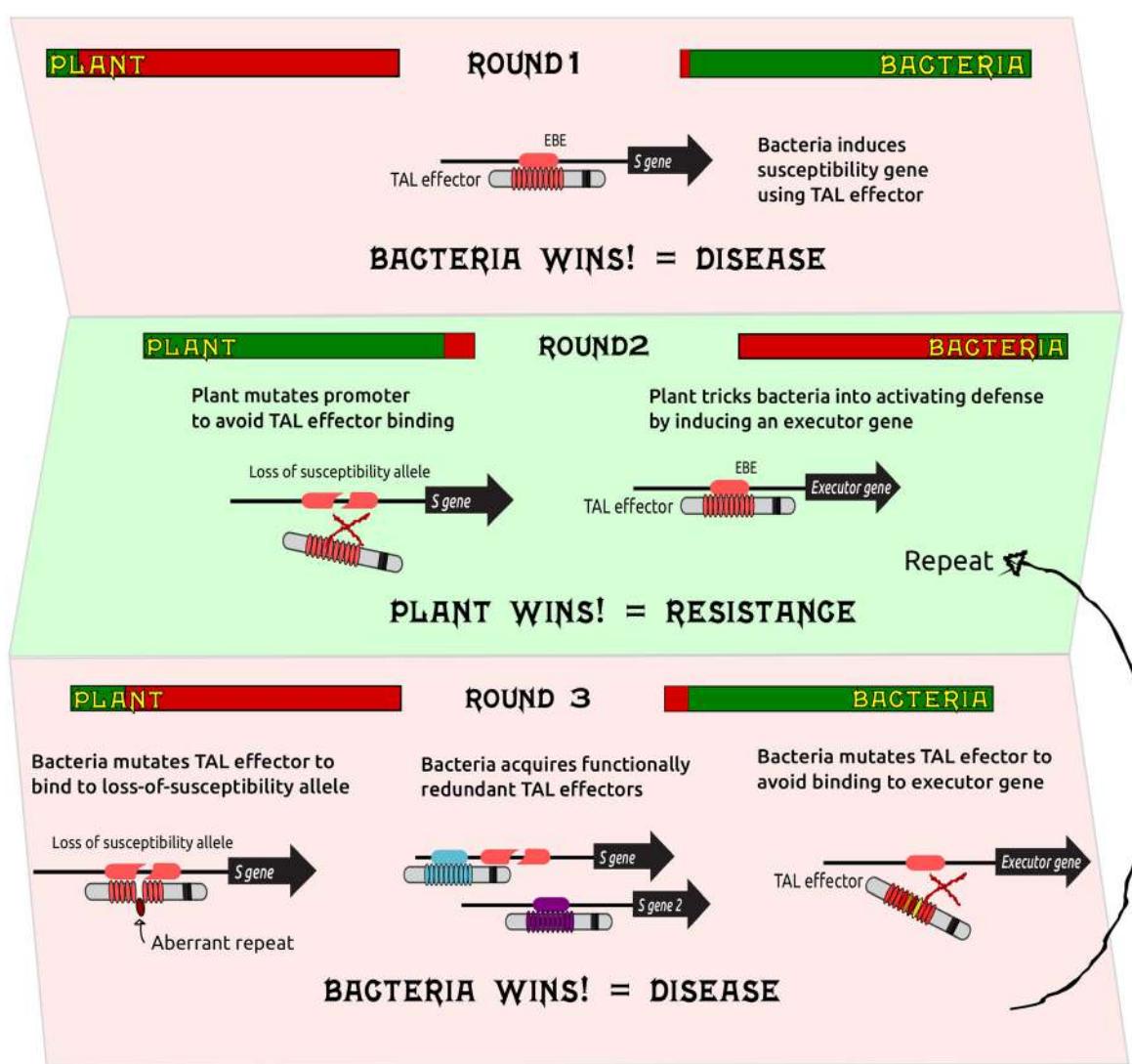


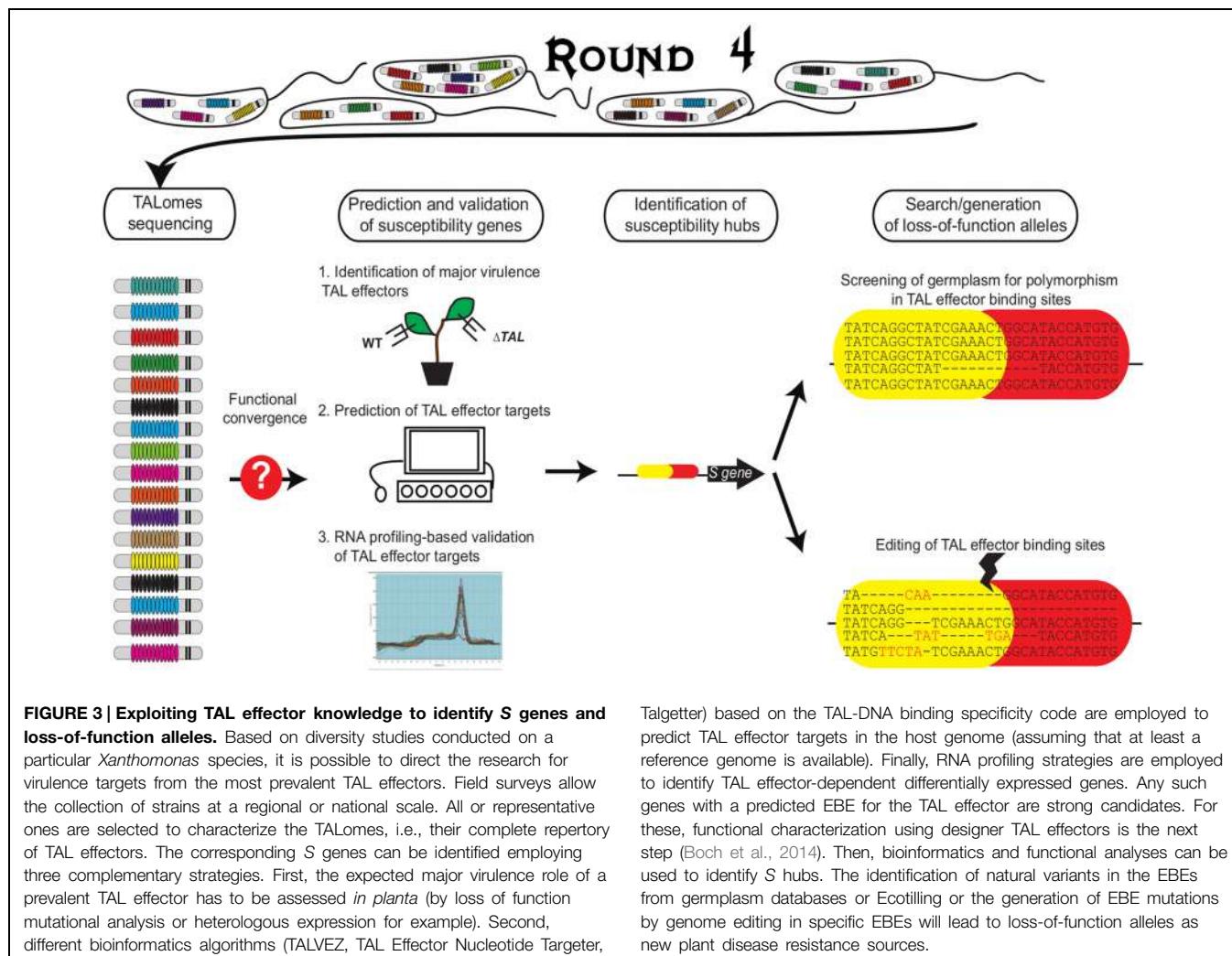
FIGURE 2 | A simplified model for combat involving bacterial TAL effectors and plant susceptibility genes. To cause disease, bacteria use TAL effectors that bind to the promoter regions of susceptibility (*S*) genes. The plant might defend itself against these effectors through mutations in the promoter that prevent binding of the TAL effector(s): so called loss-of-susceptibility alleles. Alternatively, plants can trick bacteria into inducing defenses via an executor gene. Bacteria can in turn counter by mutating TAL effectors to recognize the loss-of-susceptibility alleles for

example aberrant repeats that can accommodate single base deletions in the EBE or by mutating TAL effectors to avoid binding to the promoters of executor genes. Bacteria can also acquire new TAL effectors to redundantly induce *S* genes, the plant can evolve new loss-of-susceptibility alleles and executor genes, and the process becomes cyclical. Bars for plant and bacterium represent “health bars” typically used in video games to indicate the extent to which a combatant is winning or losing. Bold black arrows represent genes.

virus and to some extent *Xanthomonas* is mediated by recessive genes, efforts mainly focused on the identification of dominant *R* genes and the transformation of susceptible varieties using these genes. This type of resistance is often strain-specific and can be easily overcome by the pathogen through mutations in unique and specific avirulence genes (Dangl et al., 2013). In contrast, the knowledge of TAL effector mechanisms offers new exciting opportunities for researchers aiming to develop resistant varieties. A research strategy that exploits TAL effector knowledge to generate broad-spectrum and durable resistance (Figure 3) might include:

The Identification of Major TAL Effectors in Pathogen Populations of Interest

The characterization of a complete or near-complete TAL effector repertoire for a large group of strains of a particular *Xanthomonas* species, collected in geographically diverse regions and through different time periods, allows for the establishment of the more prevalent TAL effectors for that population (Figure 3). The sequencing of TAL effectors is a very complicated task given their repetitive nature. However, through the employment of new, relatively low cost, long-read sequencing technologies such as single molecule, real-time sequencing (Pacific Biosciences),



it is possible to accurately assemble complete TAL effector sequences from whole genome data (A. Bogdanove, personal communication). A more difficult and time-consuming strategy is to clone and sequence TAL effectors individually. A non-tested way could be to generate TAL effector-specific libraries and sequence them. The knowledge gained on “TALome” diversity will allow estimation of the durability of a loss-of-susceptibility allele or an executor gene, which can be anticipated depending on the conservation of a particular TAL effector or a group of functionally equivalent TAL effectors in a given population. Also, a prevalent TAL effector can be taken as candidate virulence factor and prioritized for functional characterization. These TAL effectors can then be tested experimentally to confirm whether or not they have a role in virulence. This knowledge about TALome diversity will be key to design the best breeding resistance strategy. The next step would be:

The Prediction and Validation of Susceptibility Hubs in the Host

Any host gene representing a benefit to bacteria is a potential plant susceptibility gene and can be considered as a candidate

for transformation strategies. However, considering that the aim is to provide broad and durable resistance, those S genes that constitute convergence hubs for various TAL effectors (like the SWEET family) are ideal. These could be identified by combining EBE prediction using the RVD sequences for the major virulence TAL effectors identified and expression data (RNA-seq or microarray). Functional characterization of the S gene would also be ideal.

Once a key S gene targeted by one or more major TAL effectors is identified, a logical step is:

The Screening for Natural Loss-of-Susceptibility Alleles in the Germplasm

For this, researchers can take advantage of the new sequencing technologies and the public genome sequence databases. In addition to the Rice SNPSeek data (Alexandrov et al., 2015), and other genomic databases for rice (Yonemaru et al., 2014; Zhao et al., 2015), similarly large datasets exist for other hosts worth exploring. In *Arabidopsis*, the 1001 project as of September 2014

includes the sequencing of over 1000 lines, representing a valuable resource to find resistance sources for *X. campestris* (Weigel and Mott, 2009). In cassava, an important resource consisting of the partial genome for 600 accessions is available and can be exploited to identify polymorphism in EBEs for targets of TAL effectors from *X. axonopodis* pv. *manihotis*². In tomato the sequence of 84 accessions has been recently reported (Tomato Genome Sequencing Consortium et al., 2014). For important crops with no such databases, alternatives like ecotilling or amplicon sequencing can be considered for identifying natural loss-S alleles (Wang et al., 2012). These alleles could then be introgressed into susceptible lines. An alternative to breeding is:

The Application of Genome Editing Strategies to Achieve Resistance

Once an S gene targeted by a particular TAL effector is identified, genome editing using tools such as TALENs and CRISPR-Cas9 can be employed to modify the EBE to prevent TAL effector binding. As mentioned before, this poses the risk of altering the endogenous regulation of the gene, so ideally several variants should be created and screened for any deleterious effects. The introduction of loss-of-function alleles to commercial, agronomically important varieties using these genome edition strategies can be expected to be subject to less restrictive regulation and greater public acceptance than GMOs (Araki and Ishii, 2015).

Alternatively, an artificial executor gene can be created by engineering the promoter of a classical R gene to have several EBEs corresponding to different TAL effectors. This approach has already been applied in rice and was shown to successfully confer resistance to the desired strains (Hummel et al., 2012; Zeng et al., 2015). A predicted, but not tested, alternative involves silencing of the S gene upon pathogen infection. For example an artificial microRNA can be designed directed specifically against the S host gene. However, to avoid the above-mentioned problem of affecting endogenous function, the expression of this microRNA should be under the control of a pathogen-inducible promoter. In this scenario, the EBE for a particular, prevalent TAL effector

² <http://dev.cassava-genome.cirad.fr/?q=snp>

References

- Alexandrov, N., Tai, S., Wang, W., Mansueto, L., Palis, K., Fuentes, R. R., et al. (2015). SNP-Seek database of SNPs derived from 3000 rice genomes. *Nucleic Acids Res.* 43, D1023–D1027. doi: 10.1093/nar/gk u1039
- Al-Saadi, A., Reddy, J. D., Duan, Y. P., Brunings, A. M., Yuan, Q., and Gabriel, D. W. (2007). All five host-range variants of *Xanthomonas citri* carry one pthA homolog with 17.5 repeats that determines pathogenicity on citrus, but none determine host-range variation. *Mol. Plant Microbe Int.* 20, 934–943. doi: 10.1094/MPMI-20-8-0934
- Antony, G., Zhou, J., Huang, S., Li, T., Liu, B., White, F., et al. (2010). Rice xa13 recessive resistance to bacterial blight is defeated by induction of the disease susceptibility gene Os-11N3. *Plant Cell* 22, 3864–3876. doi: 10.1105/tpc.110.078964

could be employed. The identification of naturally occurring and selected mutations presents a major advantage over *de novo* edited EBE knockouts since it avoids the risk of affecting endogenous cis-regulatory elements and thus plant gene functions. Yet, one should also raise the possibility that viability of such alleles might be genotype-dependent, they need therefore to be tested when moved to any elite recipient variety. In summary, if bacteria have evolved TAL effectors to take advantage of the presence of host genes to induce their expression for their benefit, plant breeders can counter with strategies directed at these TAL effectors to combat these plant pathogens.

Conclusion

The discovery of the TAL effector-DNA binding code is greatly increasing our ability to identify new S proteins and better understand how diseases caused by *Xanthomonas* species occur in several major crops. Because of the highly conserved structure of TAL effectors and their mode of action to promote disease, standardized experimental pipelines can be developed in various *Xanthomonas*-host interactions to search for major S genes, provided that TAL effectors are at play. There is no doubt that in the near future, notably with the help of new sequencing technologies, extensive lists of new S genes will be generated to guide the search for pathogen-adapted loss-of-function alleles useful for marker-assisted breeding of loss-of-susceptibility. Less straightforward will be the challenge of finding such recessive resistance genes that cannot be easily broken, which will require better understanding of TAL effector diversity and evolution in pathogen populations.

Acknowledgments

We are grateful to J. Lang for edits and comments and to L. Lamy for his help in retrieving SNP data. MH and AP-Q are supported by doctoral fellowships awarded by the MESR and the Erasmus Mundus Action 2 PRECIOSA program of the European Community, respectively. This project was supported by a grant from Agence Nationale de la Recherche (ANR-14-CE19-0002) and from Fondation Agropolis (#1403-073).

- Araki, M., and Ishii, T. (2015). Towards social acceptance of plant breeding by genome editing. *Trends Plant Sci.* 20, 145–149. doi: 10.1016/j.tplants.2015.01.010
- Boch, J., and Bonas, U. (2010). *Xanthomonas AvrBs3* family-type III effectors: discovery and function. *Annu. Rev. Phytopathol.* 48, 419–436. doi: 10.1146/annurev-phyto-080508-081936
- Boch, J., Bonas, U., and Lahaye, T. (2014). TAL effectors-pathogen strategies and plant resistance engineering. *New Phytol.* 204, 823–832. doi: 10.1111/nph.13015
- Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., et al. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 326, 1509–1512. doi: 10.1126/science.1178811
- Bogdanove, A. J., and Voytas, D. F. (2011). TAL effectors: customizable proteins for DNA targeting. *Science* 333, 1843–1846. doi: 10.1126/science.1204094

- Carroll, D. (2014). Genome engineering with targetable nucleases. *Annu. Rev. Biochem.* 83, 409–439. doi: 10.1146/annurev-biochem-060713-035418
- Castiblanco, L. F., Gil, J., Rojas, A., Osorio, D., Gutiérrez, S., Muñoz-Bodnar, A., et al. (2012). TALE1 from *Xanthomonas axonopodis* pv. manihotis acts as a transcriptional activator in plant cells and is important for pathogenicity in cassava plants. *Mol. Plant Pathol.* 14, 84–95. doi: 10.1111/j.1364-3703.2012.00830.x
- Cernadas, R. A., Doyle, E. L., Nino-Liu, D. O., Wilkins, K. E., Bancroft, T., Wang, L., et al. (2014). Code-assisted discovery of TAL effector targets in bacterial leaf streak of rice reveals contrast with bacterial blight and a novel susceptibility gene. *PLoS Pathog.* 10:e1003972. doi: 10.1371/journal.ppat.103972
- Chen, L. Q. (2014). SWEET sugar transporters for phloem transport and pathogen nutrition. *New Phytol.* 201, 1150–1155. doi: 10.1111/nph.12445
- Chen, L. Q., Hou, B. H., Lalonde, S., Takanaga, H., Hartung, M. L., Qu, X. Q., et al. (2010). Sugar transporters for intercellular exchange and nutrition of pathogens. *Nature* 468, 527–532. doi: 10.1038/nature09606
- Chen, L. Q., Qu, X. Q., Hou, B. H., Sosso, D., Osorio, S., Fernie, A. R., et al. (2012). Sucrose efflux mediated by SWEET proteins as a key step for phloem transport. *Science* 335, 207–211. doi: 10.1126/science.1213351
- Chong, J., Piron, M. C., Meyer, S., Merdinoglu, D., Bertsch, C., and Mestre, P. (2014). The SWEET family of sugar transporters in grapevine: VvSWEET4 is involved in the interaction with *Botrytis cinerea*. *J. Exp. Bot.* 65, 6589–6601. doi: 10.1093/jxb/eru375
- Chu, Z., Yuan, M., Yao, J., Ge, X., Yuan, B., Xu, C., et al. (2006). Promoter mutations of an essential gene for pollen development result in disease resistance in rice. *Genes Dev.* 20, 1250–1255. doi: 10.1101/gad.1416306
- Cohn, M., Bart, R. S., Shybut, M., Dahlbeck, D., Gomez, M., Morbitzer, R., et al. (2014). *Xanthomonas axonopodis* virulence is promoted by a transcription activator-like effector-mediated induction of a SWEET sugar transporter in cassava. *Mol. Plant Microbe Interact.* 27, 1186–1198. doi: 10.1094/MPMI-06-14-0161-R
- Dangl, J. L., Horvath, D. M., and Staskawicz, B. J. (2013). Pivoting the plant immune system from dissection to deployment. *Science* 341, 746–751. doi: 10.1126/science.1236011
- de Lange, O., Schreiber, T., Schandry, N., Radeck, J., Braun, K. H., Koszinowski, J., et al. (2013). Breaking the DNA-binding code of *Ralstonia solanacearum* TAL effectors provides new possibilities to generate plant resistance genes against bacterial wilt disease. *New Phytol.* 199, 773–786. doi: 10.1111/nph.12324
- Gendron, J. M., Liu, J.-S., Fan, M., Bai, M.-Y., Wenkel, S., Springer, P. S., et al. (2012). Brassinosteroids regulate organ boundary formation in the shoot apical meristem of *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 109, 21152–21157. doi: 10.1073/pnas.1210799110
- Grau, J., Wolf, A., Reschke, M., Bonas, U., Posch, S., and Boch, J. (2013). Computational predictions provide insights into the biology of TAL effector target sites. *PLoS Comput Biol* 9:e1002962. doi: 10.1371/journal.pcbi.1002962
- Hu, Y., Zhang, J., Jia, H., Sosso, D., Li, T., Frommer, W. B., et al. (2014). Lateral organ boundaries 1 is a disease susceptibility gene for citrus bacterial canker disease. *Proc. Natl. Acad. Sci. U.S.A.* 111, E521–E529. doi: 10.1073/pnas.1313271111
- Hummel, A. W., Doyle, E. L., and Bogdanove, A. J. (2012). Addition of transcription activator-like effector binding sites to a pathogen strain-specific rice bacterial blight resistance gene makes it effective against additional strains and against bacterial leaf streak. *New Phytol.* 195, 883–893. doi: 10.1111/j.1469-8137.2012.04216.x
- Juillerat, A., Bertonati, C., Dubois, G., Guyot, V., Thomas, S., Valton, J., et al. (2014). BurrH: a new modular DNA binding protein for genome engineering. *Sci Rep.* 4, 3831. doi: 10.1038/srep03831
- Kay, S., Boch, J., and Bonas, U. (2005). Characterization of AvrBs3-like effectors from a *Brassicaceae* pathogen reveals virulence and avirulence activities and a protein with a novel repeat architecture. *Mol. Plant Microbe Interact.* 18, 838–848. doi: 10.1094/MPMI-18-0838
- Kay, S., Hahn, S., Marois, E., Hause, G., and Bonas, U. (2007). A bacterial effector acts as a plant transcription factor and induces a cell size regulator. *Science* 318, 648–651. doi: 10.1126/science.1144956
- Kay, S., Hahn, S., Marois, E., Wieduwild, R., and Bonas, U. (2009). Detailed analysis of the DNA recognition motifs of the *Xanthomonas* type III effectors AvrBs3 and AvrBs3Deltarep16. *Plant J.* 59, 859–871. doi: 10.1111/j.1365-313X.2009.03922.x
- Lange, O. D., Wolf, C., Dietze, J., Elsaesser, J., Morbitzer, R., and Lahaye, T. (2014). Programmable DNA-binding proteins from *Burkholderia* provide a fresh perspective on the TALE-like repeat domain. *Nucleic Acids Res.* 42, 7436–7449. doi: 10.1093/nar/gku329
- Lapin, D., and Van den Ackerveken, G. (2013). Susceptibility to plant disease: more than a failure of host immunity. *Trends Plant Sci.* 18, 546–554. doi: 10.1016/j.tplants.2013.05.005
- Lemoine, R., La Camera, S., Atanassova, R., Dedaldechamp, F., Allario, T., Pourtau, N., et al. (2013). Source-to-sink transport of sugar and regulation by environmental factors. *Front. Plant Sci.* 4:272. doi: 10.3389/fpls.2013.00272
- Li, T., Huang, S., Zhou, J., and Yang, B. (2013). Designer TAL effectors induce disease susceptibility and resistance to *Xanthomonas oryzae* pv. *oryzae* in rice. *Mol. Plant* 6, 781–789. doi: 10.1093/mp/stt034
- Li, T., Liu, B., Spalding, M. H., Weeks, D. P., and Yang, B. (2012). High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nat. Biotechnol.* 30, 390–392. doi: 10.1038/nbt.2199
- Li, Z., Zou, L., Ye, G., Xiong, L., Ji, Z., Zakria, M., et al. (2014). A potential disease susceptibility gene CsLOB of citrus is targeted by a major virulence effector PthA of *Xanthomonas citri* subsp. *citri*. *Mol. Plant* 7, 912–915. doi: 10.1093/mp/stt176
- Liu, Q., Yuan, M., Zhou, Y., Li, X., Xiao, J., and Wang, S. (2011). A paralog of the MtN3/saliva family recessively confers race-specific resistance to *Xanthomonas oryzae* in rice. *Plant Cell Environ.* 34, 1958–1969. doi: 10.1111/j.1365-3040.2011.02391.x
- Majer, C., and Hochholdinger, F. (2011). Defining the boundaries: structure and function of LOB domain proteins. *Trends Plant Sci.* 16, 47–52. doi: 10.1016/j.tplants.2010.09.009
- Mak, A. N., Bradley, P., Bogdanove, A. J., and Stoddard, B. L. (2013). TAL effectors: function, structure, engineering and applications. *Curr. Opin. Struct. Biol.* 23, 93–99. doi: 10.1016/j.sbi.2012.11.001
- Meckler, J. F., Bhakta, M. S., Kim, M. S., Ovadia, R., Habrian, C. H., Zykovitch, A., et al. (2013). Quantitative analysis of TALE-DNA interactions suggests polarity effects. *Nucleic Acids Res.* 41, 4118–4128. doi: 10.1093/nar/gkt085
- Moscou, M. J., and Bogdanove, A. J. (2009). A simple cipher governs DNA recognition by TAL effectors. *Science* 326, 1501–1501. doi: 10.1126/science.1178817
- Mukhtar, M. S., Carunis, A. R., Dreze, M., Epple, P., Steinbrenner, J., Moore, J., et al. (2011). Independently evolved virulence effectors converge onto hubs in a plant immune system network. *Science* 333, 596–601. doi: 10.1126/science.1203659
- Munoz-Bodnar, A., Perez-Quintero, A. L., Gomez-Cano, F., Gil, J., Michelmore, R., Bernal, A., et al. (2014). RNAseq analysis of cassava reveals similar plant responses upon infection with pathogenic and non-pathogenic strains of *Xanthomonas axonopodis* pv. manihotis. *Plant Cell Rep.* 33, 1901–1912. doi: 10.1007/s00299-014-1667-7
- Pereira, A. L., Carazzolle, M. F., Abe, V. Y., De Oliveira, M. L., Domingues, M. N., Silva, J. C., et al. (2014). Identification of putative TAL effector targets of the citrus canker pathogens shows functional convergence underlying disease development and defense response. *BMC Genomics* 15:157. doi: 10.1186/1471-2164-15-157
- Perez-Quintero, A. L., Rodriguez, R. L., Dereeper, A., Lopez, C., Koebnik, R., Szurek, B., et al. (2013). An improved method for TAL effectors DNA-binding sites prediction reveals functional convergence in TAL repertoires of *Xanthomonas oryzae* strains. *PLoS ONE* 8:e68464. doi: 10.1371/journal.pone.0068464
- Reineke, A. R., Bornberg-Bauer, E., and Gu, J. (2011). Evolutionary divergence and limits of conserved non-coding sequence detection in plant genomes. *Nucleic Acids Res.* 39, 6029–6043. doi: 10.1093/nar/gkr179
- Richter, A., Streubel, J., Blucher, C., Szurek, B., Reschke, M., Grau, J., et al. (2014). A TAL effector repeat architecture for frameshift binding. *Nat. Commun.* 5, 3447. doi: 10.1038/ncomms4447
- Romer, P., Recht, S., Strauss, T., Elsaesser, J., Schornack, S., Boch, J., et al. (2010). Promoter elements of rice susceptibility genes are bound and activated by specific TAL effectors from the bacterial blight pathogen, *Xanthomonas oryzae* pv. *oryzae*. *New Phytol.* 187, 1048–1057. doi: 10.1111/j.1469-8137.2010.03217.x

- Schornack, S., Ballvora, A., Gurlebeck, D., Peart, J., Baulcombe, D., Ganal, M., et al. (2004). The tomato resistance protein Bs4 is a predicted non-nuclear TIR-NB-LRR protein that mediates defense responses to severely truncated derivatives of AvrBs4 and overexpressed AvrBs3. *Plant J.* 37, 46–60. doi: 10.1046/j.1365-313X.2003.01937.x
- Schornack, S., Moscou, M. J., Ward, E. R., and Horvath, D. M. (2013). Engineering plant disease resistance based on TAL effectors. *Annu. Rev. Phytopathol.* 51, 383–406. doi: 10.1146/annurev-phyto-082712-102255
- Siemens, J., Keller, I., Sarx, J., Kunz, S., Schuller, A., Nagel, W., et al. (2006). Transcriptome analysis of *Arabidopsis* clubroots indicate a key role for cytokinins in disease development. *Mol. Plant Microbe Interact.* 19, 480–494. doi: 10.1094/MPMI-19-0480
- Streubel, J., Blucher, C., Landgraf, A., and Boch, J. (2012). TAL effector RVD specificities and efficiencies. *Nat. Biotechnol.* 30, 593–595. doi: 10.1038/nbt.2304
- Streubel, J., Pesce, C., Hutin, M., Koebnik, R., Boch, J., and Szurek, B. (2013). Five phylogenetically close rice SWEET genes confer TAL effector-mediated susceptibility to *Xanthomonas oryzae* pv. *oryzae*. *New Phytol.* 200, 808–819. doi: 10.1111/nph.12411
- Sugio, A., Yang, B., Zhu, T., and White, F. F. (2007). Two type III effector genes of *Xanthomonas oryzae* pv. *oryzae* control the induction of the host genes OsTFIIAgamma1 and OsTFX1 during bacterial blight of rice. *Proc. Natl. Acad. Sci. U.S.A.* 104, 10720–10725. doi: 10.1073/pnas.0701742104
- Thatcher, L. F., Manners, J. M., and Kazan, K. (2009). *Fusarium oxysporum* hijacks COI1-mediated jasmonate signaling to promote disease development in *Arabidopsis*. *Plant J.* 58, 927–939. doi: 10.1111/j.1365-313X.2009.03831.x
- Tomato Genome Sequencing Consortium, Aflitos, S., Schijlen, E., De Jong, H., De Ridder, D., Smit, S., et al. (2014). Exploring genetic variation in the tomato (*Solanum* section *Lycopersicon*) clade by whole-genome sequencing. *Plant J.* 80, 136–148. doi: 10.1111/tpj.12616
- Vedel, V., and Scotti, I. (2011). Promoting the promoter. *Plant Sci.* 180, 182–189. doi: 10.1016/j.plantsci.2010.09.009
- Verdier, V., Triplett, L. R., Hummel, A. W., Corral, R., Cernadas, R. A., Schmidt, C. L., et al. (2012). Transcription activator-like (TAL) effectors targeting OsSWEET genes enhance virulence on diverse rice (*Oryza sativa*) varieties when expressed individually in a TAL effector-deficient strain of *Xanthomonas oryzae*. *New Phytol.* 196, 1197–1207. doi: 10.1111/j.1469-8137.2012.04367.x
- Wang, T. L., Uauy, C., Robson, F., and Til, B. (2012). TILLING in extremis. *Plant Biotechnol. J.* 10, 761–772. doi: 10.1111/j.1467-7652.2012.00708.x
- Weigel, D., and Mott, R. (2009). The 1001 genomes project for *Arabidopsis thaliana*. *Genome Biol.* 10, 107. doi: 10.1186/gb-2009-10-5-107
- White, F. F., and Yang, B. (2009). Host and pathogen factors controlling the rice-*Xanthomonas oryzae* interaction. *Plant Physiol.* 150, 1677–1686. doi: 10.1104/pp.109.139360
- Yang, B., Sugio, A., and White, F. F. (2006). Os8N3 is a host disease-susceptibility gene for bacterial blight of rice. *Proc. Natl. Acad. Sci. U.S.A.* 103, 10503–10508. doi: 10.1073/pnas.0604088103
- Yang, B., and White, F. F. (2004). Diverse members of the AvrBs3/PthA family of type III effectors are major virulence determinants in bacterial blight disease of rice. *Mol. Plant Microbe Interact.* 17, 1192–1200. doi: 10.1094/MPMI.2004.17.11.1192
- Yang, Y., De Feyter, R., and Gabriel, D. W. (1994). Host-specific symptoms and increased release of *Xanthomonas citri* and *X. campestris* pv. *malvacearum* from leaves are determined by the 102-bp tandem repeats of pthA and avrb6, respectively. *Mol. Plant Microbe Interact.* 7, 345–355. doi: 10.1094/MPMI-7-0345
- Yonemaru, J., Ebana, K., and Yano, M. (2014). HapRice, an SNP haplotype database and a web tool for rice. *Plant Cell Physiol.* 55:e9. doi: 10.1093/pcp/pcp188
- Yu, Y. H., Streubel, J., Balzergue, S., Champion, A., Boch, J., Koebnik, R., et al. (2011). Colonization of rice leaf blades by an African Strain of *Xanthomonas oryzae* pv. *oryzae* depends on a new TAL effector that induces the rice nodulin-3 Os11N3 gene. *Mol. Plant Microbe Interact.* 24, 1102–1113. doi: 10.1094/MPMI-11-10-0254
- Yuan, M., Chu, Z., Li, X., Xu, C., and Wang, S. (2009). Pathogen-induced expressional loss of function is the key factor in race-specific bacterial resistance conferred by a recessive R gene xa13 in rice. *Plant Cell Physiol.* 50, 947–955. doi: 10.1093/pcp/pcp046
- Yuan, T., Li, X., Xiao, J., and Wang, S. (2011). Characterization of *Xanthomonas oryzae*-responsive cis-acting element in the promoter of rice race-specific susceptibility gene Xa13. *Mol. Plant* 4, 300–309. doi: 10.1093/mp/ssq076
- Zeng, X., Tian, D., Gu, K., Zhou, Z., Yang, X., Luo, Y., et al. (2015). Genetic engineering of the Xa10 promoter for broad-spectrum and durable resistance to *Xanthomonas oryzae* pv. *oryzae*. *Plant Biotechnol. J.* doi: 10.1111/pbi.12342 [Epub ahead of print].
- Zhao, H., Yao, W., Ouyang, Y., Yang, W., Wang, G., Lian, X., et al. (2015). RiceVarMap: a comprehensive database of rice genomic variations. *Nucleic Acids Res.* 43, D1018–D1022. doi: 10.1093/nar/gku894
- Zhou, J., Peng, Z., Long, J., Sosso, D., Liu, B., Eom, J. S., et al. (2015). Gene targeting by the TAL effector PthXo2 reveals cryptic resistance gene for bacterial blight of rice. *Plant J.* 82, 632–643. doi: 10.1111/tpj.12838

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 Hutin, Pérez-Quintero, Lopez and Szurek. 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.

QueTAL: a suite of tools to classify and compare TAL effectors functionally and phylogenetically

Alvaro L. Pérez-Quintero¹, Léo Lamy¹, Jonathan L. Gordon², Aline Escalon², Sébastien Cunnac¹, Boris Szurek^{1*†} and Lionel Gagnevin^{1*†}

¹ UMR IPME, IRD-CIRAD-Université Montpellier, Montpellier, France, ² UMR PVBMT, CIRAD-Université de la Réunion, Saint-Pierre, France

OPEN ACCESS

Edited by:

Thomas Lahaye,
Ludwig-Maximilians-University
Munich, Germany

Reviewed by:

David John Studholme,
University of Exeter, UK
Tina Britta Jordan,
Eberhard Karls University Tübingen,
Germany

*Correspondence:

Boris Szurek and Lionel Gagnevin,
UMR IPME, IRD-CIRAD-UM, 911, Av.
Agropolis BP 64501, 34394
Montpellier, France
boris.szurek@ird.fr;
lionel.gagnevin@cirad.fr

[†]These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Plant Biotic Interactions,
a section of the journal
Frontiers in Plant Science

Received: 23 April 2015

Accepted: 06 July 2015

Published: 03 August 2015

Citation:

Pérez-Quintero AL, Lamy L, Gordon JL, Escalon A, Cunnac S, Szurek B and Gagnevin L (2015) QueTAL: a suite of tools to classify and compare TAL effectors functionally and phylogenetically. Front. Plant Sci. 6:545.

doi: 10.3389/fpls.2015.00545

Transcription Activator-Like (TAL) effectors from *Xanthomonas* plant pathogenic bacteria can bind to the promoter region of plant genes and induce their expression. DNA-binding specificity is governed by a central domain made of nearly identical repeats, each determining the recognition of one base pair via two amino acid residues (a.k.a. Repeat Variable Di-residue, or RVD). Knowing how TAL effectors differ from each other within and between strains would be useful to infer functional and evolutionary relationships, but their repetitive nature precludes reliable use of traditional alignment methods. The suite QueTAL was therefore developed to offer tailored tools for comparison of TAL effector genes. The program DisTAL considers each repeat as a unit, transforms a TAL effector sequence into a sequence of coded repeats and makes pair-wise alignments between these coded sequences to construct trees. The program FuncTAL is aimed at finding TAL effectors with similar DNA-binding capabilities. It calculates correlations between position weight matrices of potential target DNA sequence predicted from the RVD sequence, and builds trees based on these correlations. The programs accurately represented phylogenetic and functional relationships between TAL effectors using either simulated or literature-curated data. When using the programs on a large set of TAL effector sequences, the DisTAL tree largely reflected the expected species phylogeny. In contrast, FuncTAL showed that TAL effectors with similar binding capabilities can be found between phylogenetically distant taxa. This suite will help users to rapidly analyse any TAL effector genes of interest and compare them to other available TAL genes and should improve our understanding of TAL effectors evolution. It is available at <http://bioinfo-web.mpl.ird.fr/cgi-bin2/quetal/quetal.cgi>.

Keywords: TAL effectors, phylogeny, *Ralstonia*, *Xanthomonas*, functional convergence, EBE

Introduction

Transcription activator-like (TAL) effectors are *Xanthomonas* proteins that are translocated into the plant cell through the type III secretion system and directed to the nucleus where they commandeer the cell metabolism by specifically activating plant genes (Bogdanove et al., 2010). In several pathovars they were demonstrated to be major aggressiveness determinants responsible for symptoms. In some situations they also act as avirulence factors, i.e., triggering the hypersensitive response notably when activating “executor” resistance genes (Boch and Bonas, 2010). Their mode

of action has been detailed and their most outstanding feature is their central repeat domain, which is responsible for their highly specific attachment to DNA in regions known as EBE (effector binding elements). This domain contains 1.5–33.5 repeats of 33–35 amino acids. In each repeat the 12th and 13th amino acids are variable (therefore called “Repeat Variable Di-residue” or RVD) and dictate the specific interaction with a single nucleotide of the target DNA. Hence the successive RVDs in the protein are involved in specific attachment to a sequence of contiguous nucleotides located in the promoter of the gene to be activated. The correspondence between RVD and nucleotide, the “TAL code,” has been deciphered and demonstrated experimentally and may be used to predict targets of TAL effectors in plants (Boch et al., 2009; Moscou and Bogdanove, 2009). Researchers are now confronted by a wide array of potential TAL effector targets that can be experimentally explored to understand the mechanisms of *Xanthomonas* pathogenicity (through susceptibility targets), as well as some mechanisms underlying plant resistance to *Xanthomonas* (through executor resistance genes). Eventually this may help to develop new tools to breed resistant plants, either by escaping susceptibility or by introgressing executor resistance genes (Bogdanove et al., 2010; Boch et al., 2014).

As more TALomes, i.e., repertoires of *TAL* effector genes, are discovered and sequenced, one challenging issue has been to classify and compare them in order to (1) understand phylogenetic relatedness between *TAL* effector genes and decipher their modes of evolution; (2) assess functional similarities between *TAL* effectors and predict cases of functional convergence.

Alignment and distance calculation between *TAL* effector genes at the DNA or protein level are not straightforward due to the high sequence similarity between repeats, which are often identical over the majority of their sequence, with few variable residues not providing enough weight to correctly align orthologous repeats. To avoid this problem several works use alignments of the N-terminal and/or C-terminal regions of *TAL* effectors (Bogdanove et al., 2011; Yu et al., 2011; Pereira et al., 2014). However, sequences for these regions are not always available because sequencing efforts usually concentrate on the central repeat region which is more useful for functional studies. Furthermore, the distal regions are highly similar and may not allow discriminating between genes. In addition, the evolution and diversification of *TAL* effector genes may rely heavily on duplication and recombination, which is facilitated by their frequent localization on mobile insertion cassettes (MICs) (Ferreira et al., 2015) and their repeated structure (Lau et al., 2014). This produces multiple paralogous copies of similar genes sequences differing through insertion, deletion or reshuffling of their repeat units.

Currently, there is no systematic way to predict similar DNA binding capabilities among *TAL* effectors, other than through comparison of outputs from *TAL* effector binding site prediction

software (Noel et al., 2013; Booher and Bogdanove, 2014). This turns to be impractical when dealing with large sets of sequences particularly when different species and pathovars are involved or through visual comparisons of RVD sequences, which in addition to being unworkable leaves out the variable binding inherent in the RVD-DNA code.

In this paper we describe two methods, DisTAL and FuncTAL, to align and classify *TAL* effector gene sequences according to their central repeat region. With DisTAL, we propose a tool that infers phylogenetic relationships between genes by considering each repeat as a unit and calculating distances between arrays of repeats, using an algorithm initially designed to compare microsatellite sequences (ARLEM version 1.0, Abouelhoda et al., 2010). FuncTAL aims to find functionally-related *TAL* effectors by calculating similarities in binding probabilities according to the RVD-DNA code. Together, these programs will help researchers infer evolutionary and functional relationships within and between groups of *TAL* effectors.

Materials and Methods

Datasets

The sequences of 229 *TAL* effectors were obtained from the NCBI protein and nucleotide databases (<http://www.ncbi.nlm.nih.gov/>; accession numbers in **Supplementary Table 1**). This set was used for all analyses unless indicated and is referred to as the public dataset. 496 additional sequences (awaiting publication) were provided by collaborating laboratories, including those reported in Wilkins et al. (2015). These, together with public *TAL* effectors sequences are referred to as the full dataset. The species composition of the full dataset is found in **Supplementary Table 2**.

Program Specifications

FuncTAL and DisTAL are implemented in the Perl and R programming languages, they use the Perl modules Statistics::R, Bio::Perl (Stajich et al., 2002), and the R library APE (Paradis et al., 2004).

DisTAL additionally uses the module Algorithm::NeedlemanWunsch (<http://search.cpan.org/~vbar/Algorithm-NeedlemanWunsch-0.03/lib/Algorithm/NeedlemanWunsch.pm>) to align repeats and the C++ program ARLEM version 1.0 (Abouelhoda et al., 2010) to align sequences of coded repeats. Penalty parameters values for NeedlemanWunsch alignments are gap: 0, mismatch: -1, match: +1, alignment scores are normalized by dividing the score by the maximum length among the analyzed sequences and multiplying by 100 so they can be used by ARLEM. Parameters values for ARLEM are: align = TRUE, -insert = TRUE, ARLEM alignment scores are divided by 100 (so they can be used to build trees). Neighbor-joining trees are generated using the nj function of the package APE with default parameters (Paradis et al., 2004). The input file for DisTAL is a FASTA file containing amino acid sequences of *TAL* effectors. An additional file containing information on the *TAL* effectors can be used to color code the trees generated by the program. The following parameters can be modified: layout of the output tree (default = unrooted), include and compare

Abbreviations: EBE, Effector binding element; Gb, Gigabyte; GHz, Giga hertz; indel, insertion/deletion; LRR, Leucine-rich repeat; PWM, Positional weight matrix; RVD, Repeat variable di-residue; TAL, Transcription activator-like; Xoc, *Xanthomonas oryzae* pv. *oryzicola*; Xoo, *Xanthomonas oryzae* pv. *oryzae*.

input to TAL effectors from the public dataset (default = false), number of similar TAL effectors to output if the public database option is active (default = 5), exclude RVDs from analysis (default = false). Additional parameters can be modified in the standalone version: ARLEM indel penalization (default = 10), ARLEM duplication penalization (default = 10), and Create repeat distance matrix *de novo* (default = false). The outputs generated by DisTAL are:

- a pseudo-FASTA file (*Outputname_CodedRepeats.fa*) with the TAL effectors coded as a string of numbers,
- the set of unique repeats in the input file and their number codes (*Outputname_RepeatsCode.txt*)
- a matrix (*Outputname.mat*) showing the distances between the TAL effectors,
- a tree file (*Outputname.tre*) in Newick format to be used in any tree visualization program,
- a hits file (*Outputname.hits*) if the option to compare to a database was used; this file shows, for each TAL effector, the closest matches in our database.

DisTAL took an average of 0 m 22.3 s to process 200 TAL effector sequences in a computer with a Linux operating system with 15.6 Gb of RAM and an Intel® Core™ i7-4600U CPU @ 2.10 GHz processor. When the option to compare against public TAL effectors is activated, time goes up to 0 m 35.5 s.

FuncTAL uses modified subroutines (readMotifFile, compMotifs, scoreComparison, correlation) from the program compareMotifs.pl from the Homer (Hypergeometric Optimization of Motif EnRichment) suite (Heinz et al., 2010). The program can take as an input a text file with RVD sequences in the format “>RVD_id<tab>HD-NN-HD....” or a FASTA file containing nucleotide or amino acid sequences of TAL effectors. If a FASTA file is entered, the program will first recognize repeats in the TAL effector sequence as described for DisTAL. For each repeat the program next extracts the RVDs, i.e., the 12th and 13th amino acid (e.g., NN-HD). If the 13th amino acid is missing, as is the case for some repeats, the program inserts an asterisk “*.” Neighbor-joining trees are generated using the nj function of the package APE with default parameters (Paradis et al., 2004). The following parameters can be modified: layout of the output tree (default = unrooted), include and compare input to TAL effectors from the public dataset (default = false), and number of similar TAL effectors to output if the public database option is active (default = false). The outputs generated by FuncTAL are:

- a text file (*Outputname.cons*) showing the theoretical most likely binding site for each TAL effector,
- a matrix (*Outputname.mat*) showing the distances between the TAL effectors,
- a tree file (*Outputname.tre*) in Newick format to be used in any tree visualization program,
- a hits file (*Outputname.hits*) if the option to compare to a database was used, this file shows, for each TAL effector, the closest matches in our database according to their binding sites.

FuncTAL took an average of 1 m 22.3 s to process 200 TAL sequences in a computer with a Linux operating system with

15.6 Gb of RAM and an Intel® Core™ i7-4600U CPU @ 2.10 GHz processor. When the option to compare against public TAL effectors is activated, time goes up to 7 m 48.5 s.

The script used for simulated evolution of TAL effectors (Evolve.pl) is also made available at <http://sourceforge.net/projects/quetaleffectors>. This program uses the dist.topo function of the program APE (Paradis et al., 2004) to calculate topological distances between trees using the Penny and Hendy method (Penny and Hendy, 1985). The topological distance is defined as twice the number of internal branches defining different bipartitions of the tips (Penny and Hendy, 1985). The distances were normalized by the number of nodes in a tree. In this way a distance of 0 means identical trees, and the maximum distance of 2 means completely different trees.

The version of DisTAL that uses only sequences of RVDs (DistTAL-OnlyRVDs.pl) is also made available at <http://sourceforge.net/projects/quetaleffectors>. This version extracts the 12th and 13th amino acid from each repeats and then uses the same method for DisTAL, possible alignment scores between RVDs using the Needleman-Wunsch algorithm are 0, 50, and 100, and Indel penalization for ARLEM is 100 (both amino acids are deleted). This version was not extensively tested, thus it is not included in the web version.

ClustalW (Larkin et al., 2007) alignments were made using Clustal 2.1 with default parameters in two steps: clustalw -ALIGN and clustalw -TREE.

Muscle (Edgar, 2004) alignments were made using version 3.8.31 with default parameters. Alignments for t-coffee (Notredame et al., 2000) were made using version 10.00.r1613 (-gapopen = -50, -gapext = 0). And MAFFT (Katoh et al., 2005) alignments used the version 7.123b (E-INS-i -ep 0 -genafpair -maxiterate 1000). Parameters were chosen to allow long gaps in alignments. Trees were generated from these alignments using the dist-align function from the “seqinr” R package (<http://seqinr.r-forge.r-project.org/>) and the nj function from the APE package (Heinz et al., 2010).

Programs Availability

Packages containing the scripts of the FuncTAL and DisTAL programs, as well as additional scripts used in this work are available for download from Sourceforge at <http://sourceforge.net/projects/quetaleffectors>. A web interface and the source code for the suite are also available at <http://bioinfo-web.mpl.ird.fr/cgi-bin2/quetal/quetal.cgi>. The web version was created in Perl cgi-bin with w3c recommendations for CSS level 3 and Html 5.0 <http://www.w3.org/standards/webdesign/htmlcss>.

Results

QueTAL: DisTAL, A Program for the Phylogenetic Classification of TAL Effector Repeat Regions

The overall strategy to compare TAL effectors based on the sequence of their central repeat region consists in considering each repeat as a separate unit, and comparing the TAL effectors according to the nature and order of these units. This strategy is based on the assumption that repeats are the evolutionary units of TAL effector genes and can be deleted and duplicated as a

whole. This is supported by the sequence and structural features of TAL effectors (Deng et al., 2012; Mak et al., 2012), recent models of TAL effectors evolution (Ferreira et al., 2015), as well as by works indicating that TAL effectors' functional specificity can be modified by changing the sequence of repeats (Herbers et al., 1992; Boch et al., 2009; Streubel et al., 2012), and that deletions or duplications occur in nature or and may be responsible for change in virulence and aggressiveness (Vera Cruz et al., 2000).

To phylogenetically classify TAL effectors, we developed the program DisTAL, which classifies the input TAL effectors as a string of coded repeats and then uses the program ARLEM to calculate distances between these strings. The workflow for this program is depicted in **Figure 1**, and described next in detail.

Identification and Coding of Repeats

The program takes as input a set of TAL effectors to be analyzed, and if desired, the input TAL effectors can be compared to a dedicated database of 229 TAL effectors available in public DNA sequence databases (**Supplementary Table 1**). The input file should be a FASTA file containing either nucleotide or amino acid sequences of TAL effectors. If the input is nucleotide sequences these are translated to amino acids (in reading frame +1). It identifies and separates repeats in the input sequences by finding matches to motifs of 7 amino acids found at the start of repeats of known TAL effectors as traditionally defined (Boch and Bonas, 2010) (i.e., LTPDQVV). The program can also identify aberrant repeats (longer or shorter than average) and keep them for analyses. If they exist, the program also identifies and uses missing repeats (identified as strings of X's) which are sometimes included in TAL sequences due to sequencing gaps. It is however not recommended to include sequences with these gaps since these repeats will be assigned the maximum distance to any others.

Each unique repeat type is then assigned a numeric code and the original TAL effector sequences from the input file are transformed into sequences of coded repeats. Additionally the user can decide whether or not to exclude the RVDs from the analyses. If this option is chosen, the sequence analyzed for each repeat will be a concatenation of the 1st to 11th amino acid plus the amino acids from the 14th to the end of the repeat. This reduces the size and complexity of the repeat alphabet and, in theory, avoids biasing effects caused by different selection pressures acting on the RVDs.

Calculating Distances between Unique Repeats

Next, a distance matrix is generated by calculating distances between every pair of unique repeats. For this, a global alignment with sliding ends (no gap penalty) is made for each pair of unique repeats using the Needleman-Wunsch algorithm (Needleman and Wunsch, 1970) as implemented in the Perl package Algorithm::NeedlemanWunsch (<http://search.cpan.org/~vbar/Algorithm-NeedlemanWunsch-0.03/lib/Algorithm/NeedlemanWunsch.pm>). The distances are normalized so that the repeat matrix information can be interpreted as the percentage of amino acids that change between repeats (based on the longest repeat among the two aligned). A distance matrix was generated for a set of 1110 unique TAL effector repeats

found in our full dataset. It is included in the web version and the standalone version of the program to save computational time. If new repeats are found in the input file these are compared to the existing matrix and added to it.

Alternatively, the user can choose to generate this matrix using the Smith-Waterman (Smith and Waterman, 1981) algorithm for pairwise alignments as implemented in the Perl package Bio::Tools::pSW (<http://search.cpan.org/dist/BioPerl/Bio/Tools/pSW.pm>) using different amino acid substitution matrices (PAM30, PAM50, and Blosum62). This strategy is so far only available in the standalone version and it has not been extensively tested, however the results obtained with either matrix are often similar; the average topological distances for 50 trees obtained from 10 randomly selected TAL effector sequences when comparing the trees obtained with the Needleman-Wunsch algorithm to those obtained using Smith-Waterman + PAM30, PAM50, and Blosum62 were 0.36, 0.42, and 0.27, respectively.

Aligning and Calculating Distances between Strings of Coded Repeats

To compare the sequences of coded repeats DisTAL uses the program ARLEM (also referred to as WAMI) (Abouelhoda et al., 2010) which was designed to compare minisatellite maps. A minisatellite map is a sequence of symbols that represents tandem arrays of short repetitive DNA segments such that the set of symbols is in one-to-one correspondence with the set of distinct repeats (Abouelhoda et al., 2010). We propose that, like minisatellites, TAL effector repeats when considered as evolutionary units can undergo three non-mutually exclusive processes: *Unit mutation* (change from one repeat to another), *duplication* (tandem copies of a repeat), and *insertion/deletion* (indels = loss or gain of new repeats) as described in Abouelhoda et al. (2010). In ARLEM each of these events is assigned a cost when aligning the sequences of units (Abouelhoda et al., 2010). In our case, the cost of unit mutation would be defined by the distance matrix generated in the previous steps, that is, the penalization for changing one repeat to another depends on the percentage of amino acids that are different between said repeats. The duplication and indel penalization is 10 for both events by default (a penalization equivalent to changing 10% of the amino acids from one repeat to another). These values were estimated by optimizing the length and score of sample alignments as shown below.

The alignment scores outputted by ARLEM for each pair of TAL effectors are the sums of the penalization values for mismatches, indels and duplications; the scores are then divided by 100. Consequently two TAL effectors with identical repeats will get an alignment score of zero. In contrast, two TAL effectors of the same length (i.e., 15 repeats), aligned with no gaps, and where each pair of aligned repeats differ from each other in 50% of the amino acids will have a score of $50 \times 15/100 = 7.5$.

Creating Trees of TAL Effectors

The scores outputted by ARLEM are organized into a matrix, which is then used to create a neighbor-joining tree using the R package APE (Heinz et al., 2010) in a user-defined format. The tree tip labels can be colored using an additional input file from

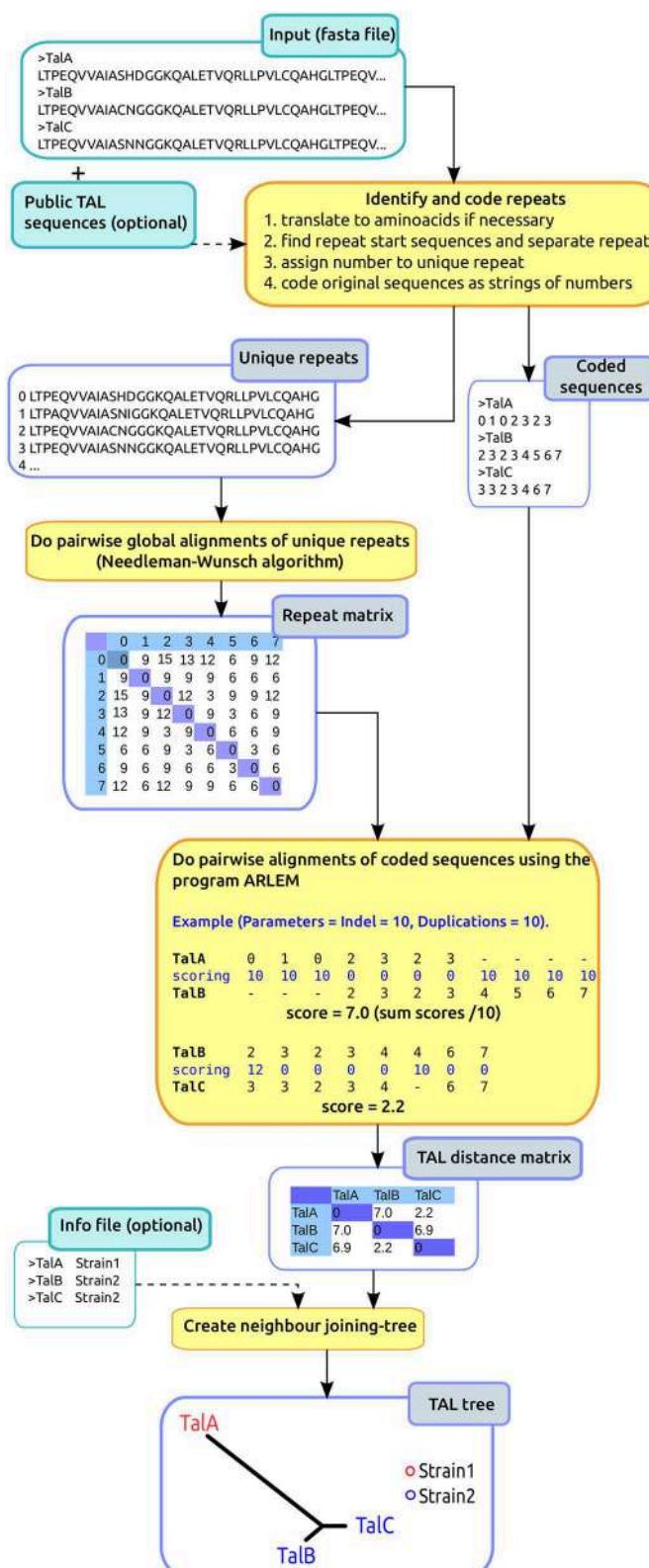


FIGURE 1 | DisTAL workflow. The diagram shows a summarized version of the DisTAL workflow. An example is shown for 3 hypothetical TAL effectors from two strains containing 7–8 repeats each, and 10 unique repeats in total. Processes are shown in the orange squares, inputs (lighter blue) and outputs in blue squares.

the user that contains the TAL effector IDs and categories used to group them (e.g., species or strain).

Optimizing the Values of DisTAL Parameters

To establish adequate penalizations for indels and duplications we optimized the alignments for the TAL effectors PthA1, PthA2, PthA3, and PthA4 from the *X. citri* pv. *citri* strain IAPAR 306 (Da Silva et al., 2002) since these TAL effectors are closely related (Pereira et al., 2014). When aligning the coded sequences for these TAL effectors, too many gaps are introduced if no penalizations for indels or duplications are used, thus resulting in very long alignments where the sequences may not even overlap. In contrast, using penalization values for indels and duplications that are too high results in alignments with fewer gaps but those gaps increase highly the alignment score (Figures 2A–C). As a consequence, the difference in score between gapped and ungapped alignments increases (Figures 2C,D), which could result in biased trees where TAL effectors with the same number of repeats tend to be grouped together.

We used DisTAL to run all pairwise alignments between these four *X. citri* pv. *citri* TAL effectors using different penalization values for indels and duplications, and looked for values that produced short alignments with little variation in the alignment scores for different TAL effector pairs. When keeping a high duplication penalization (100) and changing the indel penalization, the best alignments were found for penalization values between 5 and 10, with 10 producing shorter alignments (Figure 2D). Likewise when keeping a high indel penalization of 100 and changing the duplication penalization, the best alignments were found for penalization values between 5 and 10 (Figure 2E). The same results are found when changing both penalizations simultaneously (Supplementary Figure 1). Similar results were obtained using TAL effectors from the *X. oryzae* pv. *oryzae* (*Xoo*) strain PXO99^A and those from *X. oryzae* pv. *oryzicola* (*Xoc*) strain BLS256 (Supplementary Figure 2). The default value for both penalizations was then decided as 10.

DisTAL Accurately Recreates the Phylogeny of *In silico*-evolved TAL Effectors

To test the ability of DisTAL to decipher the phylogeny of TAL effectors, we designed a script to simulate the evolution of TAL effectors under the assumption of repeats acting as evolutionary units (Figure 3). For this, an initial hypothetical TAL effector is created by randomly selecting 10 repeats out of a set of 344 unique repeats found in *Xoo* TAL effectors in the public dataset. Two copies (descendants) are then generated from this TAL effector and each descendant undergoes 100 evolutionary cycles where in each cycle two different events can occur:

(i) Replacement: a repeat is chosen at random and replaced by another one from the set of 344 repeats. This process is equivalent to mutating a series of amino acids in one repeat, the probability of this occurring in each cycle is designated α .

(ii) Insertion/deletion: a series of X (X = random value from 0 to 3) contiguous repeats are selected in the parent sequence and they have an equal probability of either being deleted, or being inserted into a random position in the TAL effector. The

probability of this event occurring in each cycle is designated β . Note that this event also produces tandem duplications when the repeats are inserted next to their original position.

After 100 cycles the resulting two sequences are duplicated to produce a total of 4 descendants that each undergo the same process again. Finally, eight TAL effector sequences (named A–H) are produced from the initial TAL effector. We expect that a phylogenetic tree of these eight sequences should have this grouping as shown in Figure 3: [((A B)(C D))((E F)(G H))].

Next the resulting eight TAL effectors were fed into DisTAL (under default parameters with duplication and indel penalties equal to 10) and the resulting tree was compared to the expected tree. The topological distance between the trees was calculated using the Penny and Hendy method (Penny and Hendy, 1985), as implemented in the R package APE (Paradis et al., 2004). As shown in Figure 4, this process was repeated 100 times for different combinations of α and β values, from 0 to 0.1 with 0.005 increments (40,000 trees total), to account for different evolution scenarios. DisTAL consistently produced trees that differed little from the expected tree (mean topological distance = 0.09, median = 0). The program worked better when α and β were both higher than 0.02 (at zero all the TAL effectors have the same distance and all the nodes are at the same distance), and slightly better when β was higher than α . The trees obtained with DisTAL were also compared to trees obtained by doing multiple alignments of the repeat regions of the simulated TAL effectors using the programs for multiple alignment ClustalW (Larkin et al., 2007), MAFFT (Katoh et al., 2005), Muscle (Edgar, 2004), and T-coffee (Notredame et al., 2000) and then generating neighbor-joining trees. DisTAL consistently produced trees with closer resemblance to the expected tree than those obtained after alignment with other multiple alignment programs (Figure 4, Supplementary Figure 3).

QueTAL: FuncTAL, A Program for Comparison of TAL Effectors Based on DNA Binding Specificities

TAL effectors act as transcription factors and their binding sites can be predicted according to a code (Boch et al., 2009; Moscou and Bogdanove, 2009; Noel et al., 2013). It is therefore feasible to compare the probable binding sites for TAL effectors using similar strategies as those devised to compare DNA motifs (Heinz et al., 2010). The program FuncTAL was designed to compare DNA binding capabilities for TAL effectors. Briefly, the program translates the RVD sequence of a TAL effector into a position weight matrix (PWM) stating the binding probabilities to nucleotides according to the TAL effector-DNA binding code (Boch et al., 2009; Moscou and Bogdanove, 2009). The PWMs are then compared using the strategy from the program HOMER (Heinz et al., 2010) to compare DNA motifs, which relies on calculating correlations for each position for two PWMs. The workflow for this program is depicted in Figure 5, and is explained in detail below.

Identification of RVDs and Creation of PWMs

The program reads either a tabular file containing RVD sequences or, a FASTA file with nucleotide or amino acid sequences and then extracts RVDs. The sequence of RVDs

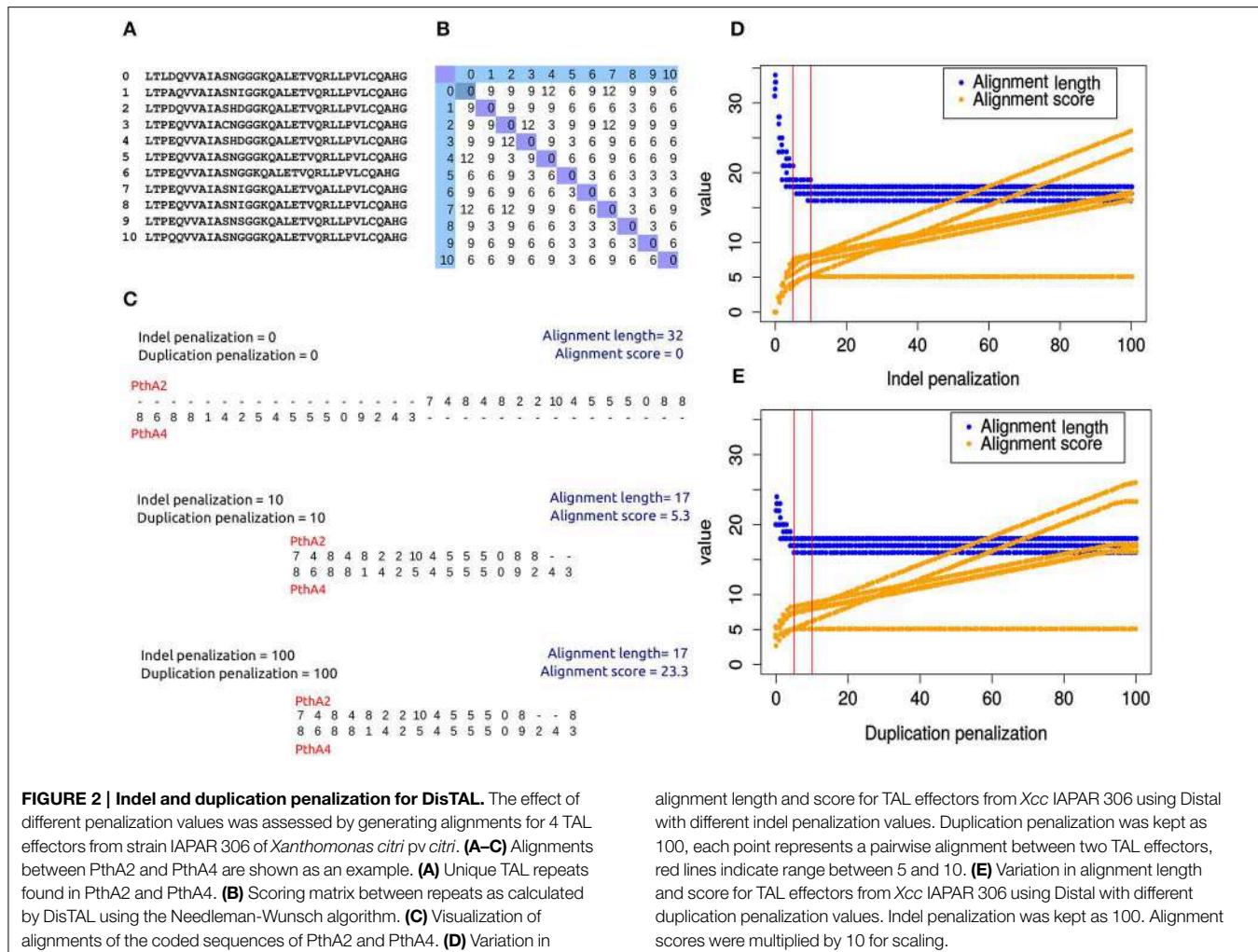


FIGURE 2 | Indel and duplication penalization for DisTAL. The effect of different penalization values was assessed by generating alignments for 4 TAL effectors from strain IAPAR 306 of *Xanthomonas citri* pv. *citri*. **(A–C)** Alignments between PthA2 and PthA4 are shown as an example. **(A)** Unique TAL repeats found in PthA2 and PthA4. **(B)** Scoring matrix between repeats as calculated by DisTAL using the Needleman-Wunsch algorithm. **(C)** Visualization of alignments of the coded sequences of PthA2 and PthA4. **(D)** Variation in

alignment length and score for TAL effectors from Xcc IAPAR 306 using Distal with different indel penalization values. Duplication penalization was kept as 100, each point represents a pairwise alignment between two TAL effectors, red lines indicate range between 5 and 10. **(E)** Variation in alignment length and score for TAL effectors from Xcc IAPAR 306 using Distal with different duplication penalization values. Indel penalization was kept as 100. Alignment scores were multiplied by 10 for scaling.

for each TAL effector is then transformed into a position weight matrix according to a modified version of the RVD-nucleotide association matrix used by the program Talvez (Perez-Quintero et al., 2013) which was shown to perform well for the identification of known binding sites (Perez-Quintero et al., 2013). The matrix was modified to include updated RVD specificities according to the literature (Streubel et al., 2012; De Lange et al., 2013; Deng et al., 2014). The specificities are shown in Supplementary Table 3.

The program also builds and outputs a consensus binding site for each TAL effector by identifying the most probable nucleotide for each position. As with DisTAL, the user can choose whether to compare the input RVD or TAL effector sequences amongst each other or to include a set of RVD sequences available from public databases in the comparison.

Alignment and Scoring of PWMs

To align and score the PWMs the program uses code from the script comparemotifs.pl of the HOMER suite (Heinz et al., 2010), which is designed to compare the binding sites of eukaryotic transcription factors.

The PWM comparisons are made for each pair of TAL effectors in the input. Every possible alignment between two matrices is evaluated by sliding the starting position (offset) of one matrix in respect to another. This process is unidirectional (i.e., the reverse or reverse complement of the binding sites are not compared), the alignments are ungapped, and the unmatched ends at either side of either matrix are filled with equal probabilities of matching any nucleotide.

For each offset, the Pearson correlation coefficient is calculated between two arrays A and B, where A is the ordered binding probabilities for each nucleotide and each position in one PWM, and B is the corresponding probabilities in a second PWM. The best alignment between the two PWMs is chosen by identifying the offset that produced the highest correlation. If x is the highest correlation between two PWMs, $1 - x$ is considered the distance between the matrices. A distance of 0 will correspond to TAL effectors that are identical in length and RVD sequence.

Creating Trees of Binding Probabilities

The distances for each pair of TAL effector PWMs are organized into a matrix, which is then used to create a neighbor-joining

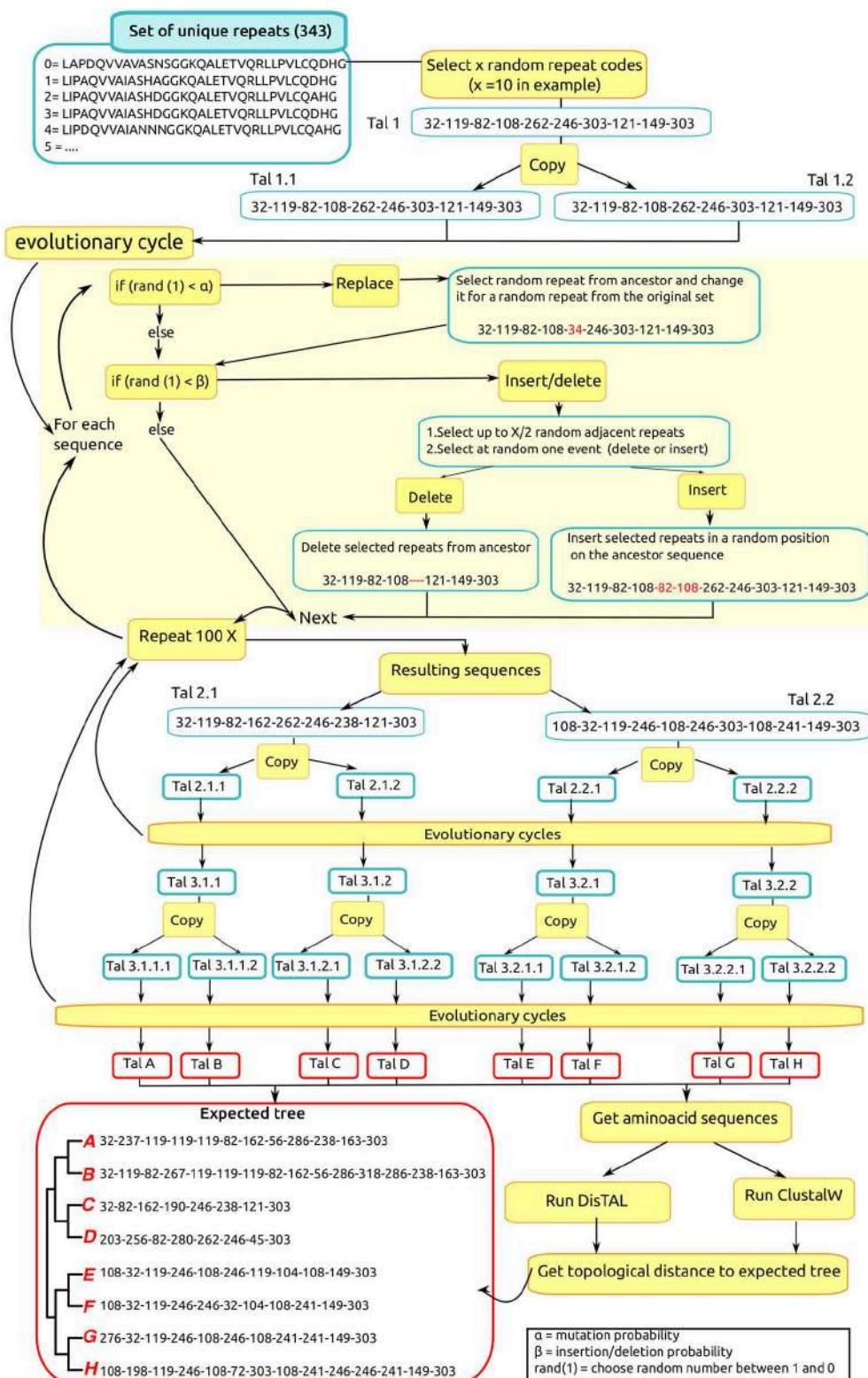
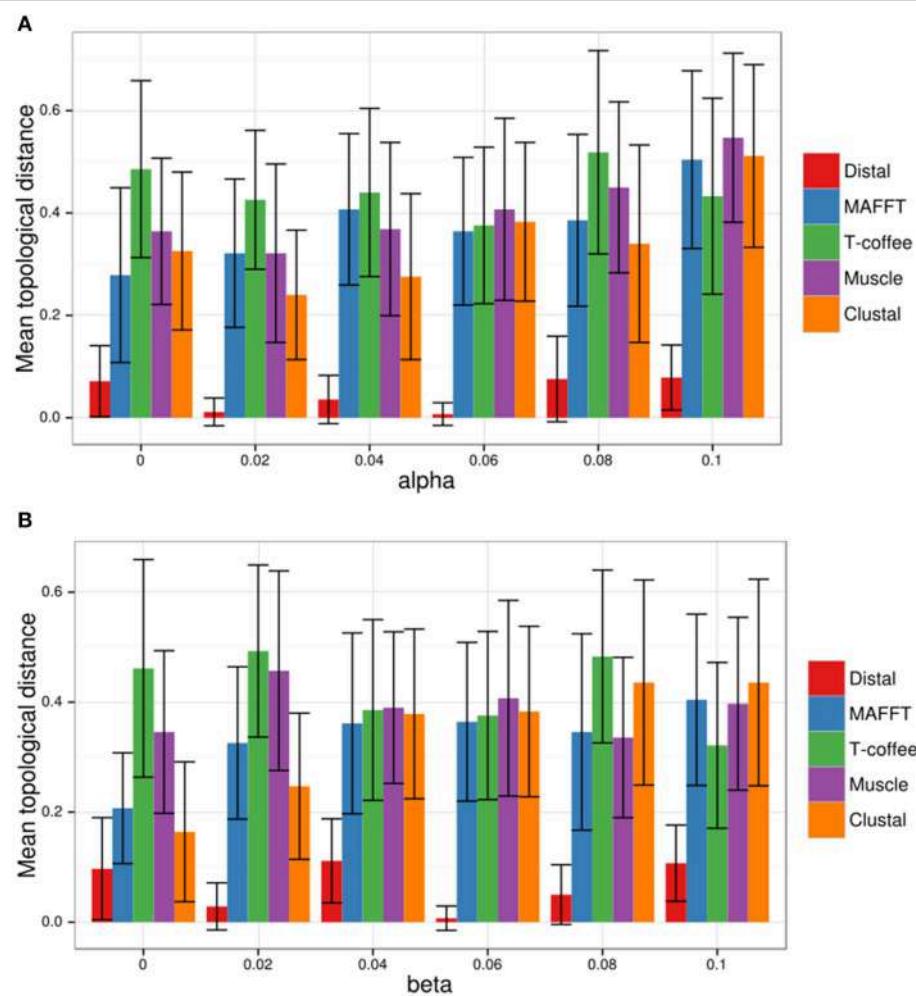


FIGURE 3 | In silico evolution of hypothetical TAL effectors. A hypothetical TAL effector is created and “evolved” as shown in the workflow for an example of a TAL effector with 10 repeats. α , replacement probability; β , indel probability; rand, randomly generated number between 0 and 1; X, TAL

effector length in number of repeats. Intermediate TAL effectors in the process are named with numbers, the first number indicates generation (it increments after each evolutionary cycle) and subsequent numbers indicate descent: when a TAL effector is copied, a number (0.1 or 0.2) is added to the name.

**FIGURE 4 | DisTAL performance with *in silico*-evolved TAL effectors.**

Sets of eight TAL effectors (named A–H) resulting from simulated evolution were fed into DisTAL, ClustalW, MAFFT, Muscle and T-coffee, the resulting trees were compared to the expected tree $[(A\ B)(C\ D)][(E\ F)(G\ H)]$, the scatter plot shows the topological distance. **(A)** Different values of alpha (probability

of repeat replacement) were used to generate the sets of TAL effectors while keeping beta (probability of repeat indel) at a value of 0.06. **(B)** Different values of beta were used to generate the sets of TAL effectors while keeping alpha at a value of 0.06. Each bar represents the average topological distance for 100 sets of TAL effectors, error bars indicate standard deviation.

tree using the R package APE (Paradis et al., 2004) in a user-defined format. As with DisTAL the tree tip labels can be colored using an additional input file from the user that contains the TAL effector IDs and categories used to group them (e.g., species or strain).

FuncTAL Accurately Represents Relations between Functionally Convergent TAL Effectors

To show that FuncTAL can identify TAL effectors with unrelated RVD arrays but similar binding specificities, we decided to take advantage of three cases of experimentally observed functional convergence among TAL effectors.

One of the best-studied cases of functional convergence among TAL effectors is that of the rice *S* susceptibility gene *SWEET14* which is induced by multiple *X. oryzae* pv. *oryzae* TAL effectors targeting at least three different EBEs (**Figure 6A**). AvrXa7 and PthXo3, from strains PXO86 and

PXO61 respectively, target overlapping EBEs in the *SWEET14* promoter (Yang and White, 2004; Chu et al., 2006; Antony et al., 2010). Another TAL effector from *Xoo* strain KACC10331, with similar RVDs but different length than AvrXa7 and PthXo3, is predicted to target the same site (accession number AAW77509.1 or YP_202894.1) (Perez-Quintero et al., 2013). Tal5 from *Xoo* strain MAI1 binds to another EBE in this promoter with minor overlap to that of AvrXa7/PthXo3 (Streubel et al., 2013), and TalC from *Xoo* strain BAI3 binds to an EBE with no overlap to the two other target sites (Yu et al., 2011). From this we expect that when fed to FuncTAL, TAL effectors that target completely overlapping sequences (AvrXa7, PthXo3 and the predicted AAW77509.1) will group together. As an outgroup, we included the TAL effector PthXo1 from *Xoo* strain PXO99^A known to target *SWEET11* which is another *SWEET* member acting as an *S* gene in rice (Yang et al., 2006). Indeed, using FuncTAL on these TAL effectors results in a tree where, AvrXa7, PthXo3, and YP_202894.1 are

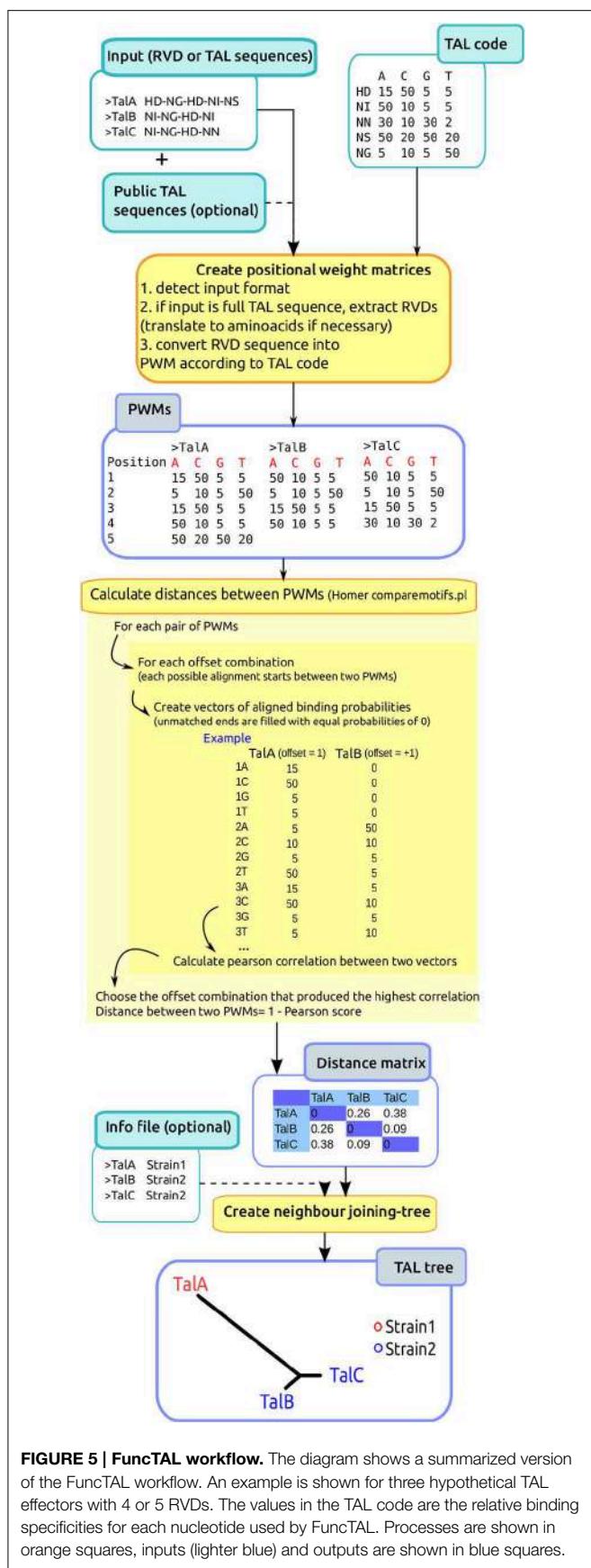


FIGURE 5 | FuncTAL workflow. The diagram shows a summarized version of the FuncTAL workflow. An example is shown for three hypothetical TAL effectors with 4 or 5 RVDs. The values in the TAL code are the relative binding specificities for each nucleotide used by FuncTAL. Processes are shown in orange squares, inputs (lighter blue) and outputs are shown in blue squares.

grouped together (Figure 6D). And although the EBE targeted by this group and that of Tal5 EBE overlap by 3 nucleotides, this is not enough for the program to consider them as functionally similar.

Another example of functional convergence is that of AvrBs3 from *X. euvesicatoria* strain 71–21 and AvrHah1 from *X. gardneri* strain XV444. These TAL effectors both bind to overlapping EBEs in the promoter of the pepper resistance gene *Bs3* (Schornack et al., 2006, 2008; Boch et al., 2009). Additionally, AvrBs3Δrep16 and AvrBs3Δrep109 are two artificial deletion derivatives of AvrBs3 (Herbers et al., 1992). When tested, it was found that AvrBs3Δrep16 lost the ability to bind to the AvrBs3 EBE in the *Bs3* promoter (Boch et al., 2009) (Figure 6B). AvrBs4, a TAL effector that activates the resistance gene *Bs4* was used as an outgroup (Schornack et al., 2004). When using FuncTAL on these TAL effectors, the resulting tree reflected the functional relation shown experimentally (Figure 6E).

Finally, another interesting case is that of a group of TAL effectors from *X. citri* (Figure 6C). PthA4, PthA* and PthA^W originate from *X. citri* pv. *citri* strains IAPAR 306, X0053 and Xc270 respectively. These TAL effectors have somewhat similar RVD sequences, they bind to EBEs situated upstream of the *CsSWEET1* and *CsLOB1* genes, and induce their expression in sweet orange (Hu et al., 2014). Additionally, the *X. citri* pv. *aurantifoliae* TAL effectors PthB and PthC effectively bind to an EBE in the *CsLOB1* promoter which is overlapping to that of PthA (Al-Saadi et al., 2007), however PthB and PthC fail to induce *CsSWEET1* (Hu et al., 2014). From this we expect all these TAL effectors to form a “functionally related” group with two subgroups: one comprising the PthA homologs and the other made of PthB and PthC. We thus fed the RVD sequences for these TAL effectors into FuncTAL. As an outgroup we included PthA3 which is a TAL effector from *Xcc* strain IAPAR 306 that fails to induce either *CsLOB1* or *CsSWEET1*. The resulting tree reflected the expected relations (Figure 6F).

In these analyses the maximum pairwise distances for TAL effectors binding overlapping EBEs were 0.70 in the *X. citri* TAL effectors (between PthC and PthA^W), 0.67 in the *X. oryzae* TAL effectors (between AAW77509.1 and AvrXa7) and 0.44 in the *Bs3*-targeting TAL effectors (between AvrHah1 and AvrBs3Δrep109). Ideally, this data would serve to establish thresholds to group TAL effectors with functional convergence. However, these values might be too variable to make accurate recommendations. More experimental data will be needed to accurately define these thresholds. Meanwhile, FuncTAL distances below 0.5 may be an adequate suggestion to consider TAL effectors as functionally similar.

FuncTAL and DisTAL Show Different Groupings of TAL Effectors

To assess how the results from DisTAL and FuncTAL differ from each other based on different settings we followed an approach based on the comparison of topological distances. For this, a set of n complete TAL amino acid sequences was selected at random from our dataset and five trees were created for that set with the following methods (Figure 7A):

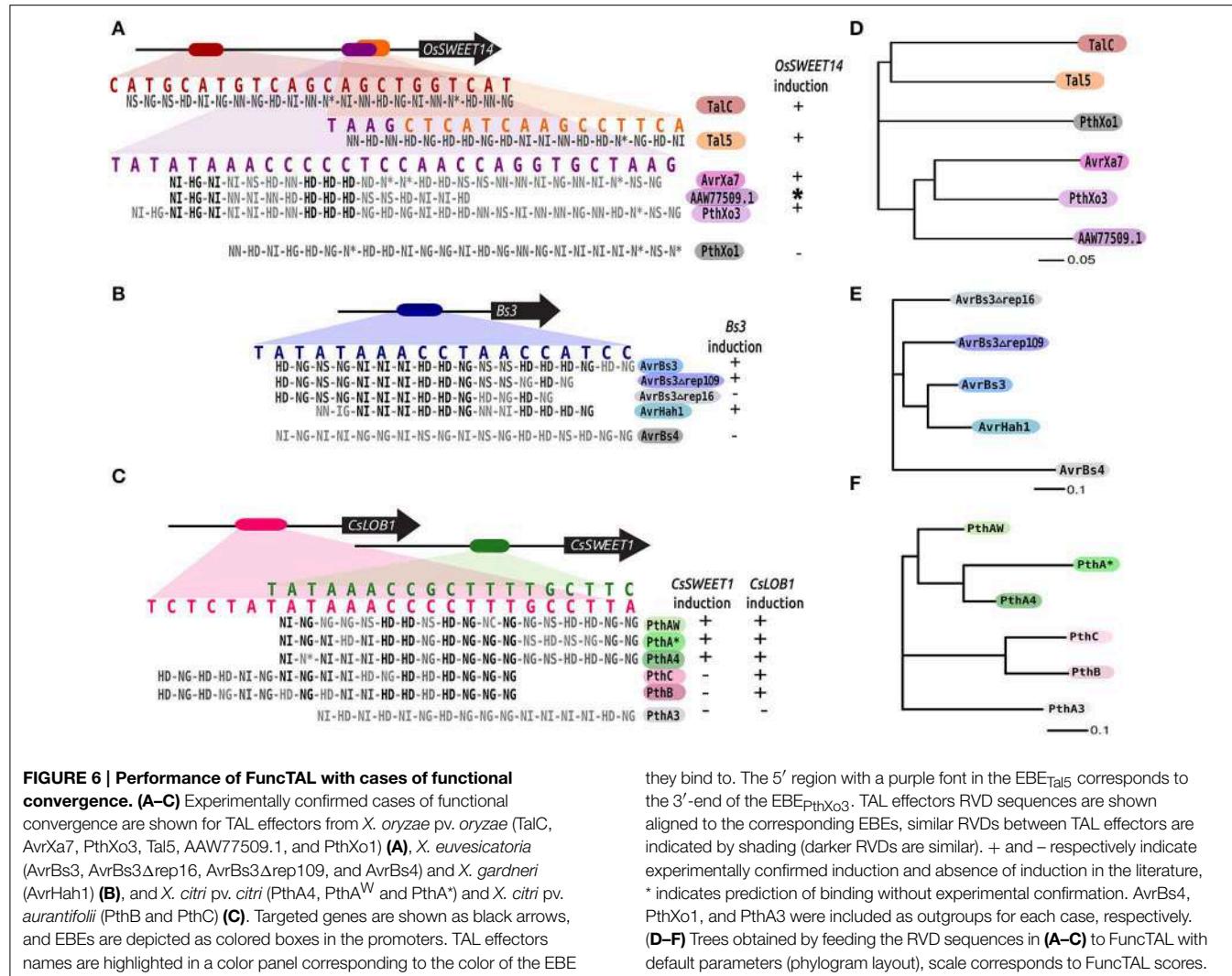


FIGURE 6 | Performance of FuncTAL with cases of functional convergence. (A–C) Experimentally confirmed cases of functional convergence are shown for TAL effectors from *X. oryzae* pv. *oryzae* (*TalC*, *AvrXa7*, *PthXo3*, *Tal5*, *AAW77509.1*, and *PthXo1*) **(A)**, *X. euvesicatoria* (*AvrBs3*, *AvrBs3Δrep16*, *AvrBs3Δrep109*, and *AvrBs4*) and *X. gardneri* (*AvrHah1*) **(B)**, and *X. citri* pv. *citri* (*PthA4*, *PthAW* and *PthA**) and *X. citri* pv. *aurantifoli* (*PthB* and *PthC*) **(C)**. Targeted genes are shown as black arrows, and EBEs are depicted as colored boxes in the promoters. TAL effectors names are highlighted in a color panel corresponding to the color of the EBE

they bind to. The 5' region with a purple font in the EBE_{Tal5} corresponds to the 3'-end of the EBE_{PthXo3}. TAL effectors RVD sequences are shown aligned to the corresponding EBEs, similar RVDs between TAL effectors are indicated by shading (darker RVDs are similar). + and - respectively indicate experimentally confirmed induction and absence of induction in the literature, * indicates prediction of binding without experimental confirmation. *AvrBs4*, *PthXo1*, and *PthA3* were included as outgroups for each case, respectively. **(D–F)** Trees obtained by feeding the RVD sequences in **(A–C)** to FuncTAL with default parameters (phylogram layout), scale corresponds to FuncTAL scores.

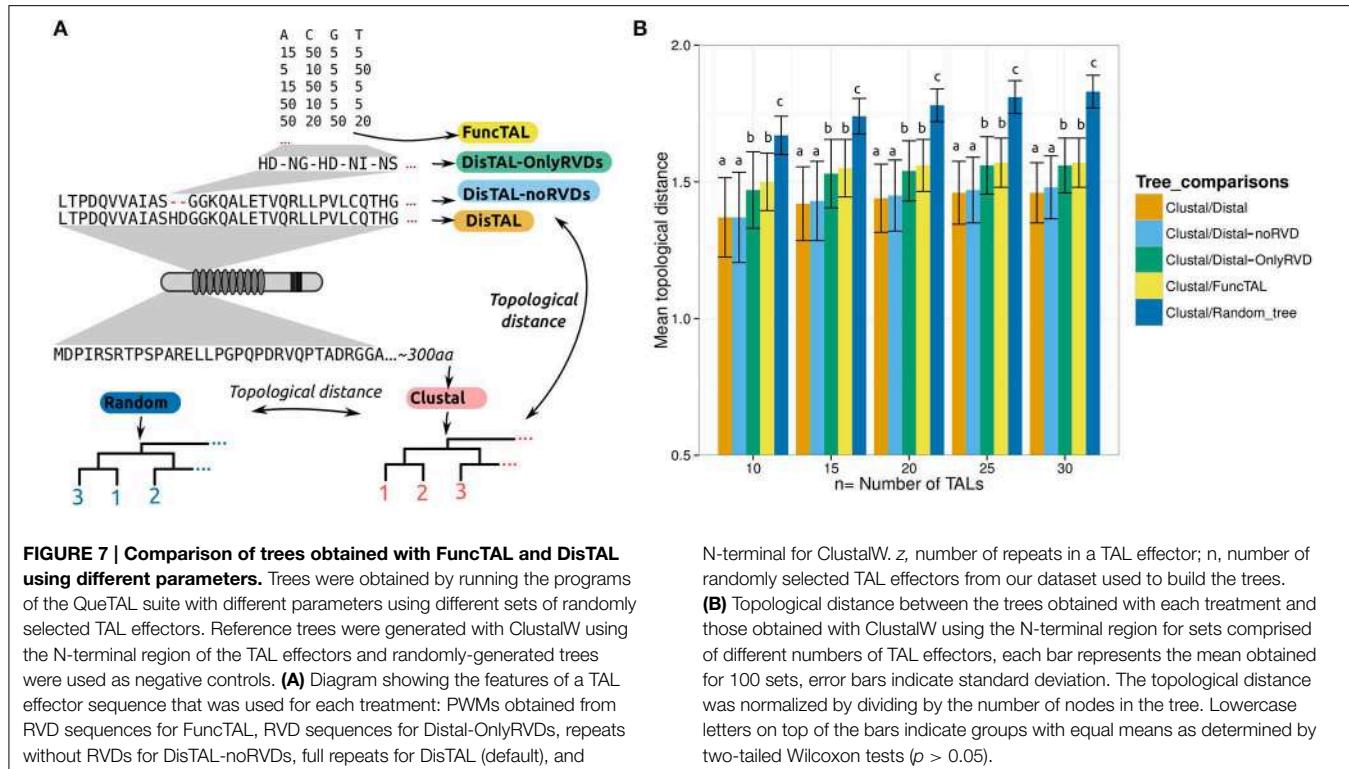
- DisTAL using default parameters (using the full repeat),
- DisTAL excluding RVDs,
- DisTAL using only RVDs,
- FuncTAL using default parameters,
- ClustalW alignment and neighbor-joining phylogenetic tree using only the N-terminal region, to be used as a reference.

The topological distance was calculated as before between each tree and the one obtained with the N-terminal region using ClustalW. As a negative control, the trees were also compared to a random tree [using rtree from the R package APE, (Paradis et al., 2004)]. The process was repeated 100 times for different values of *n*. As a result, the trees obtained with DisTAL using either the full repeats or excluding the RVDs were the most similar to the N-terminal reference (Figure 7B), suggesting that these methods infer a phylogeny similar to that obtained using a more traditional approach. Yet, the average normalized distance between each of the two methods when compared to ClustalW (N-terminal) was higher than 1. This indicates that at least half

the nodes in the trees differed from each other, suggesting that there is different information in the repeat sequences to that in the N-terminal region.

When compared to each other, the trees obtained with DisTAL with or without the RVDs also had a relatively high topological distance (mean = 1.25 when *n* = 20). This difference between the trees may be explained by RVDs being under different selective pressure (related to target sequence specificity) than the rest of the repeat sequence (which is probably under selective pressure for protein conformation). Also, the mean topological distance was higher when comparing the N-terminal trees to those obtained with FuncTAL or with DisTAL using only RVDs (Figure 7B). This indicates that the information contained in RVD sequences is somewhat different from that in the rest of the protein, thus, binding similarities are expected to not necessarily follow the phylogeny due to the selection for them to bind a specific sequence element in the host genome.

Finally, we ran our complete set of 725 TAL effector sequences through DisTAL and FuncTAL (default parameters),



and compared the distribution of taxonomic groups. As seen in **Figure 8**, the tree obtained with DisTAL seems to follow at least partially the expected phylogeny of the groups analyzed. For example, the TAL and RipTAL proteins from respectively *X. citri* pv. *citri* and *Ralstonia solanacearum* form discrete well-defined groups. In contrast, the TAL effectors from the two main pathovars of the species *X. oryzae* appear distributed in many clusters. Additionally, the recently discovered TAL effectors-like proteins from an unknown marine organism identified in metagenomic data (Juillerat et al., 2014) as well as those of *Burkholderia rhizoxinica* (De Lange et al., 2014; Juillerat et al., 2014) appear as separated from the *Xanthomonas* TAL effectors and closer to the *R. solanacearum* RipTAL proteins.

On the other hand, the tree obtained by FuncTAL shows that clusters of “functionally similar” TAL effectors often include sequences coming from different taxa (**Figure 8**). However, TAL effectors from certain clades seem to have very specific clustering, particular examples of this are the *R. solanacearum* RipTAL proteins as well as the TAL effectors from *X. translucens* to some extent, that form clusters in the tree that are distinct from the other clades. This might be due to specific RVD usage in these groups. Indeed RipTAL proteins are predicted to bind to G-C rich DNA regions in contrast to the A-T rich regions predicted for most TAL effectors (De Lange et al., 2013). Naturally occurring targets for these effectors are yet to be confirmed.

Altogether these results show that DisTAL and FuncTAL display different but complementary information that can be used to infer evolutionary relationships between taxons and predict cases of functional convergence between TAL effectors.

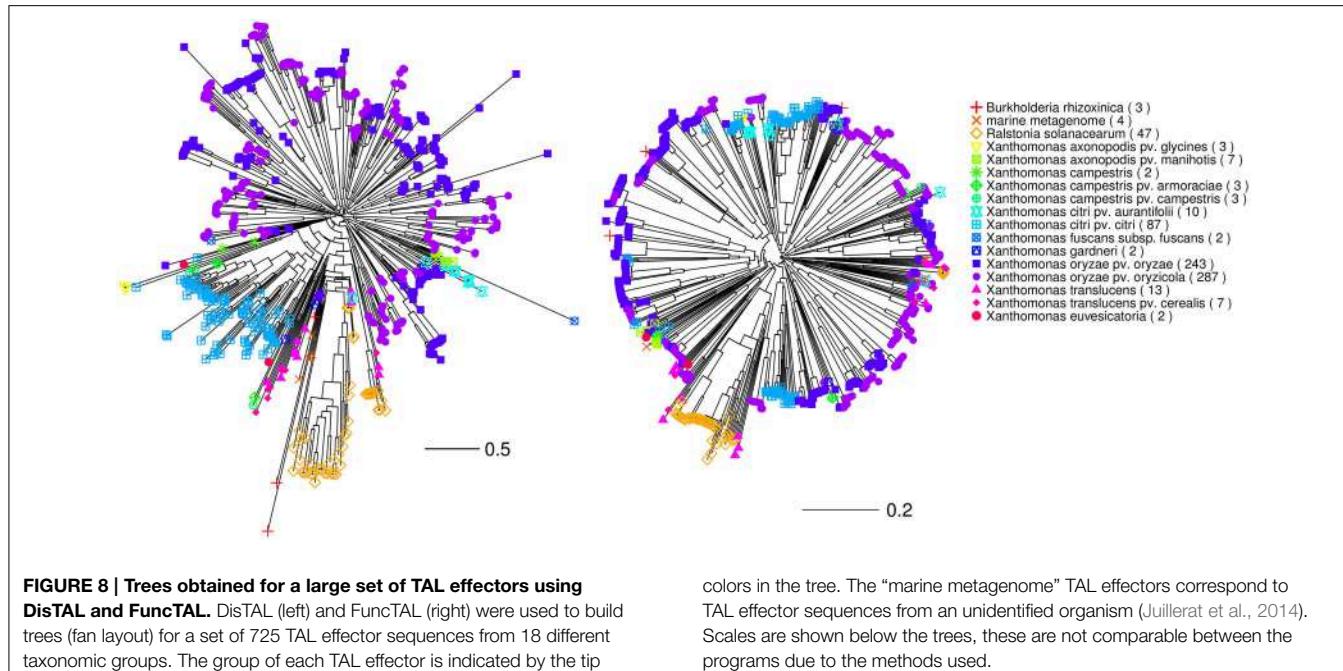
N-terminal for ClustalW. z, number of repeats in a TAL effector; n, number of randomly selected TAL effectors from our dataset used to build the trees.

(B) Topological distance between the trees obtained with each treatment and those obtained with ClustalW using the N-terminal region for sets comprised of different numbers of TAL effectors, each bar represents the mean obtained for 100 sets, error bars indicate standard deviation. The topological distance was normalized by dividing by the number of nodes in the tree. Lowercase letters on top of the bars indicate groups with equal means as determined by two-tailed Wilcoxon tests ($p > 0.05$).

Discussion

In order to understand how TAL effectors or other related proteins differ from each other within and between strains of one or several pathovars, current approaches mainly rely on the evaluation of genetic distances through the alignment of the N-terminal and/or C-terminal regions, thus excluding the central region due its repetitive nature. To fill this gap, the first aim of this work was to adapt existing methods to compare the sequences of TAL effectors repeats and infer evolutionary scenarios. The program DisTAL, used to calculate phylogenetic distances, relies on the hypothesis that one of the major driving forces of evolution of TAL effectors is probably through recombination between repeats or slipped strand mispairing during DNA replication, resulting in duplication, deletion or reorganization of one or several repeats. This hypothesis is supported by the fact that in several strains the TALome appears to be the result of numerous duplications (e.g., in *Xoo* strain PXO99^A (Bogdanove et al., 2011), that deletions occur in nature (Vera Cruz et al., 2000), and that internal recombination events were detected upon experimental evolution assays (Yang et al., 2005) and for TALEN systems *in vitro* (Lau et al., 2014). Since the structure of the genes and the mechanisms of evolution are expected to be similar to that of microsatellites, we chose to adapt an algorithm and program designed to compare coded “maps” representing tandem repeats (Abouelhoda et al., 2010).

DisTAL considers repeats as evolutionary units and finds similarities between arrays of repeats. Using simulated data we showed that the program can accurately infer relationships



between arrays of repeats derived from one ancestor that underwent processes of insertions, deletions and replacement (mutation) of repeats, performing better than a traditional multiple alignment methods. Possible caveats of the method include the fact that duplication breakpoints might not correspond to the way the repetitions have been traditionally defined, though so far there is not enough data to accurately pinpoint where these events occur. A possible workaround this problem that we will try to implement in future versions of DisTAL is to adapt a method that does not restrict tandem repeats by unit boundaries like a graph-based applied to study LRR tandem units in GALA effector proteins from *R. solanacearum* (Szalkowski and Anisimova, 2013).

The DisTAL parameters for penalizations for insertions, deletions and duplications to be used by the ARLEM algorithm were optimized by finding short alignments with low variability in their scores for TALomes of fully sequenced strains. These parameters may not accurately reflect the rate at which these events occur. Studies are needed where the evolution of TAL effectors is followed on natural bacterial populations for which short term evolutionary patterns are known. Alternatively, mutagenesis or artificial evolution experiments on TAL effectors would also be a great resource to understand variation in these proteins, in a similar way as to what has been done to create TAL effector variants with reduced virulence (Yang and White, 2004) or in the way mutational events were studied in viral vectors carrying TAL repeats (Lau et al., 2014). These types of experiments will also help determine the recombination points on these proteins.

So far, DisTAL uses amino acid sequences for all of the comparisons instead of nucleotides because the former are shorter, reducing greatly the computational time to calculate distances. This could represent a loss of information since

synonymous mutations are not taken into account. However, this loss may be minor since, for example, a set of 169 complete and unique nucleotide TAL repeat sequences (from public available databases) corresponds to 168 unique amino acid sequences.

A robust scientific framework to understand and anticipate TAL effector diversity, evolution and dynamics is essential to assess the value of control strategies based on manipulation of their host targets (Boch et al., 2014). DisTAL is the first program to allow classification of TAL effectors in a manner which includes the possibility of repeat rearrangement and duplication as a major determinant of TAL effectors evolution. The program includes pre-processing of any TAL sequence, and alignment of repeat sequences based on the ARLEM program (Abouelhoda et al., 2010). We believe this tool not only is more reliable at comparing and classifying TAL effectors according to their phylogeny but will also offer precious help for future experimental and modeling works on TAL effector evolution.

An important feature of TAL effectors is that their function can, to some extent, be predicted from their RVDs sequence thanks to their modular and specific interaction with DNA (Boch et al., 2009; Moscou and Bogdanove, 2009). Indeed tools already exist to predict candidate EBEs in plant genomes (Doyle et al., 2012; Grau et al., 2013; Pérez-Quintero et al., 2013). As more TAL effectors are discovered, notably through sequencing of entire TALomes (e.g., Wilkins et al., 2015), it is essential to classify them according to what can be hypothesized about their function. The second main output of this study is the design of a tool for comparing TAL effectors through their EBEs, which will facilitate the identification of cases of functional convergences and therefore candidate susceptibility hubs. FuncTAL calculates correlations between potential TAL effector binding sites by translating RVD sequences into PWMs according to the RVD-DNA code. The program successfully

inferred functional relations for known cases of functional convergence among TAL effectors targeting overlapping EBEs. Notably it associated TAL effectors that have very different RVD sequences and for which convergence would normally be difficult to predict (i.e., the association between PthB and PthC to the PthA group).

For now, the program does not take into account binding specificities not encoded by RVDs, such as those for position 0 in the EBE. For *Xanthomonas* TAL effectors, a thymine (T_0) preceding the EBE is required in most cases for binding and activity (Boch et al., 2009) whereas for *Ralstonia* RipTAL effectors, a guanine is required instead (De Lange et al., 2013). Because these requirements are encoded by the degenerated -1 repeat situated upstream of the central repeats, binding is not determined by RVDs but rather by the overall structure of this region (Mak et al., 2012). The specific features in the -1 repeat determining the preference for different nucleotides have yet to be identified. Structure studies suggest that in *Xanthomonas* TAL effectors, binding to T_0 is coordinated by a tryptophane (W232) in the -1 repeat (Mak et al., 2012). However, repeat number and RVD-composition seem to also affect the specificities at position zero (Schreiber and Bonas, 2014). A future version of the program may account for position zero specificity once it is possible to predict it from the TAL effector sequence and calculate binding probabilities from it.

So far the alignments made with FuncTAL are ungapped because TAL effectors bind to DNA in a sequential manner, with one RVD corresponding to one base pair, without gaps. A possible exception to this rule are TAL effectors that contain aberrant or longer than normal repeats, that have been shown to allow flexibility in binding and tolerating short gaps in their corresponding EBE (Richter et al., 2014). However, biological examples for this type of flexibility are rare, the only TAL effectors with aberrant repeats for which binding has been extensively studied are AvrXa7 and PthXo3 (Richter et al., 2014). Once the exact mechanisms of aberrant repeat binding specificities are described they may be included in the program.

It is worth stressing that FuncTAL does not use promoter sequences to infer the relations. This means that cases of functional convergence where the EBEs are not overlapping, such as TalC and AvrXa7 EBEs in rice, will be impossible to predict using these method. Yet, FuncTAL can be of interest to follow the evolution of TALomes in epidemics, notably under selective pressures, e.g., when pathogen populations are constrained by host varieties carrying resistance genes such as recessive loss-of-susceptibility alleles or dominant *R* genes. As for the analysis of the full TAL effectors dataset corresponding to 18 different taxonomic groups, the associations found by FuncTAL are more likely reflective of RVD usage and general binding preference than actual potential for convergent gene induction. We expect that to be the case for the well-defined group found for *R. solanacearum* and *X. translucens*. Indeed, the effectors in this group may have different targets, but their association reflects a preferential targeting for a certain type of sequences (GC rich regions), which may be of biological relevance.

Developing methods to predict true evolutionary and functional convergence is still needed, particularly since TAL

effectors tend to preferentially target specific genes or gene families that are crucial for disease development [Reviewed in Hutin et al. (2015), and elsewhere (Boch et al., 2014)]. Future work will be aimed at predicting these relations by combining expression data, binding site prediction and distances generated with the methods presented here.

Here we obtained trees with DisTAL and FuncTAL from a large set of TAL effector (and related) proteins showing in some cases well-defined groups that often coincide with the species or pathovar phylogeny. In the future, it will be of interest to go more in detail in the analysis of some of these groups and scrutinize which relations arise between particular strains. In particular, it is worth noting that for a few TAL effector genes, *Xoo* and *Xoc* orthologs seem closer than to any of their paralogs. This contrasts with results obtained upon alignment of the N- and C-termini regions of some of these TAL effectors, showing *Xoc* and *Xoo* TALomes to cluster separately (Bogdanove et al., 2011; Yu et al., 2011). Such contrasted results potentially highlight DisTAL's higher accuracy to infer phylogenetic relations, notably because it relies on information coming from the central repeat region. It would also be of interest to evaluate the nature and the function of these "conserved" TAL effectors, knowing that a few rice genes are known to be targeted by both pathovars (Cernadas et al., 2014).

We expect this suite to be a constantly expanding project. Other than the possible improvements mentioned above we expect to be able to add new functionalities and features to the suite including: (1) a way to use TAL effector distances obtained from either FuncTAL or DisTAL to calculate similarities between strains with fully sequenced TALomes, (2) a tool to find over-represented strings of RVD or repeat sequences in TAL effectors that may constitute functional evolutionary units, (3) tools to compare TAL effectors binding sites to plant transcription factor binding sites (aiming to help in genetic engineering strategies where resistance against bacteria is to be achieved by mutating EBEs without altering endogenous regulation of genes).

In conclusion, this work provides a more accurate tool for inferring genetic distances between TAL effector genes through the use of the phylogenetic information encoded by the repeat region. It also offers the possibility to classify groups of TAL effectors with similar DNA-binding specificities, i.e., targeting the same EBEs, thereby highlighting cases of functional convergence on key susceptibility genes. Such information can be precious when dealing with a high number of candidate host targets from which a selection has to be made to choose for the best S genes candidates. Overall in the present context where a relentless flow of TALomes and host genomes are made available through next-generation sequencing methods, we hope the QueTAL suite will be helpful to push forward our understanding of TAL effectors evolution and functional diversity.

Author Contributions

AP, designed the programs and performed the analyses, LL, designed the web platform for the programs, JG and AE devised the strategy for DisTAL, SC, helped in the design of the program and the validation strategies, BS and LG directed the work, BS, LG, and AP wrote the manuscript.

Acknowledgments

We wish to thank Alexis Dereeper for his help on the construction of the web version and to Jonathan Jacobs for fruitful discussions. We are grateful to Adam Bogdanove, Adriana Bernal, Carlos Zarate, Celine Pesce, Daniela Osorio, Katherine Wilkins, Laurent Noel, Li Wang, Nicolas Denancé, Nicholas J. Booher, Niklas Schandry, Orlando De Lange, Ralf Koebnik, Ricardo Oliva, Thomas Lahaye, and Tran Tuan Tu, for contributing with TAL effector sequences for this project. AP is supported by a doctoral fellowships awarded by the Erasmus Mundus Action 2 PRECiosa program of the European Community. This project was supported by a grant from Agence Nationale de la Recherche (ANR-14-CE19-0002) and from Fondation Agropolis (#1403-073). JG received funding from the European Union's Seventh Framework Programme ([FP7/2007-2013]) under grant agreement no. 263958 (RUN-Emerge project).

Supplementary Material

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2015.00545>

References

- Abouelhoda, M., El-Kaliby, M., and Giegerich, R. (2010). WAMI: a web server for the analysis of minisatellite maps. *BMC Evol. Biol.* 10:167. doi: 10.1186/1471-2148-10-167
- Al-Saadi, A., Reddy, J. D., Duan, Y. P., Brunings, A. M., Yuan, Q., and Gabriel, D. W. (2007). All five host-range variants of *Xanthomonas citri* carry one *pthA* homolog with 17.5 repeats that determines pathogenicity on citrus, but none determine host-range variation. *Mol. Plant Microbe Interact.* 20, 934–943. doi: 10.1094/MPMI-20-8-0934
- Antony, G., Zhou, J., Huang, S., Li, T., Liu, B., White, F., et al. (2010). Rice *xal3* recessive resistance to bacterial blight is defeated by induction of the disease susceptibility gene Os-11N3. *Plant Cell* 22, 3864–3876. doi: 10.1105/tpc.110.078964
- Boch, J., and Bonas, U. (2010). *Xanthomonas AvrBs3* family-type III effectors: discovery and function. *Annu. Rev. Phytopathol.* 48, 419–436. doi: 10.1146/annurev-phyto-080508-081936
- Boch, J., Bonas, U., and Lahaye, T. (2014). TAL effectors-pathogen strategies and plant resistance engineering. *New Phytol.* 204, 823–832. doi: 10.1111/nph.13015
- Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., et al. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 326, 1509–1512. doi: 10.1126/science.1178811
- Bogdanove, A. J., Koebnik, R., Lu, H., Furutani, A., Angiuoli, S. V., Patil, P. B., et al. (2011). Two new complete genome sequences offer insight into host and tissue specificity of plant pathogenic *Xanthomonas* spp. *J. Bacteriol.* 193, 5450–5464. doi: 10.1128/JB.05262-11
- Bogdanove, A. J., Schornack, S., and Lahaye, T. (2010). TAL effectors: finding plant genes for disease and defense. *Curr. Opin. Plant Biol.* 13, 394–401. doi: 10.1016/j.pbi.2010.04.010
- Booher, N. J., and Bogdanove, A. J. (2014). Tools for TAL effector design and target prediction. *Methods* 69, 121–127. doi: 10.1016/j.ymeth.2014.06.006
- Cernadas, R. A., Doyle, E. L., Niño-Liu, D. O., Wilkins, K. E., Bancroft, T., Wang, L., et al. (2014). Code-assisted discovery of TAL effector targets in bacterial leaf streak of rice reveals contrast with bacterial blight and a novel susceptibility gene. *PLoS Pathog.* 10:e1003972. doi: 10.1371/journal.ppat.1003972
- Chu, Z., Yuan, M., Yao, J., Ge, X., Yuan, B., Xu, C., et al. (2006). Promoter mutations of an essential gene for pollen development result in disease resistance in rice. *Genes Dev.* 20, 1250–1255. doi: 10.1101/gad.141630
- Da Silva, A. C., Ferro, J. A., Reinach, F. C., Farah, C. S., Furlan, L. R., Quaggio, R. B., et al. (2002). Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature* 417, 459–463. doi: 10.1038/417459a
- De Lange, O., Schreiber, T., Schandry, N., Radeck, J., Braun, K. H., Koszinowski, J., et al. (2013). Breaking the DNA-binding code of *Ralstonia solanacearum* TAL effectors provides new possibilities to generate plant resistance genes against bacterial wilt disease. *New Phytol.* 199, 773–786. doi: 10.1111/nph.12324
- De Lange, O., Wolf, C., Dietze, J., Elsaesser, J., Morbitzer, R., and Lahaye, T. (2014). Programmable DNA-binding proteins from *Brunholderia* provide a fresh perspective on the TALE-like repeat domain. *Nucleic Acids Res.* 42, 7436–7449. doi: 10.1093/nar/gku39
- Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., Zhu, J. K., et al. (2012). Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* 335, 720–723. doi: 10.1126/science.1215670
- Deng, D., Yan, C., Wu, J., Pan, X., and Yan, N. (2014). Revisiting the TALE repeat. *Protein Cell* 5, 297–306. doi: 10.1007/s13238-014-0035-2
- Doyle, E. L., Booher, N. J., Standage, D. S., Voytas, D. F., Brendel, V. P., Vandky, J. K., et al. (2012). TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. *Nucleic Acids Res.* 40, W117–W122. doi: 10.1093/nar/gks608
- 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
- Ferreira, R. M., De Oliveira, A. C., Moreira, L. M., Belasque, J. Jr., Gourbeyre, E., Siguiet, P., et al. (2015). A TALE of transposition: Tn3-like transposons play a major role in the spread of pathogenicity determinants of *Xanthomonas citri* and other xanthomonads. *MBio* 6, e02505–e02514. doi: 10.1128/mBio.02505-14
- Grau, J., Wolf, A., Reschke, M., Bonas, U., Posch, S., and Boch, J. (2013). Computational predictions provide insights into the biology of TAL effector target sites. *PLoS Comput. Biol.* 9:e1002962. doi: 10.1371/journal.pcbi.1002962
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., et al. (2010). Simple combinations of lineage-determining transcription factors

- prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* 38, 576–589. doi: 10.1016/j.molcel.2010.05.004
- Herbers, K., Conrads-Strauch, J., and Bonas, U. (1992). Race-specificity of plant resistance to bacterial spot disease determined by repetitive motifs in a bacterial avirulence protein. *Nature* 356, 172–174. doi: 10.1038/356172a0
- Hu, Y., Zhang, J., Jia, H., Sosso, D., Li, T., Frommer, W. B., et al. (2014). Lateral organ boundaries 1 is a disease susceptibility gene for citrus bacterial canker disease. *Proc. Natl. Acad. Sci. U.S.A.* 111, E521–E529. doi: 10.1073/pnas.1313271111
- Hutin, M., Pérez-Quintero, A. L., Lopez, C. and Szurek, B. (2015). MorTAL Kombat: the story of defense against TAL effectors through loss-of-susceptibility. *Front. Plant Sci.* 6:535. doi: 10.3389/fpls.2015.00535
- Juillerat, A., Bertonati, C., Dubois, G., Guyot, V., Thomas, S., Valton, J., et al. (2014). BurrH: a new modular DNA binding protein for genome engineering. *Sci. Rep.* 4:3831. doi: 10.1038/srep03831
- Katoh, K., Kuma, K., Toh, H., and Miyata, T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* 33, 511–518. doi: 10.1093/nar/gki198
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404
- Lau, C.-H., Zhu, H., Tay, J. C.-K., Li, Z., Tay, F. C., Chen, C., et al. (2014). Genetic rearrangements of variable di-residue (RVD)-containing repeat arrays in a baculoviral TALEN system. *Mol. Ther. Methods Clin. Dev.* 1:14050. doi: 10.1038/mtnm.2014.50
- Mak, A. N., Bradley, P., Cernadas, R. A., Bogdanove, A. J., and Stoddard, B. L. (2012). The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science* 335, 716–719. doi: 10.1126/science.1216211
- Moscou, M. J., and Bogdanove, A. J. (2009). A simple cipher governs DNA recognition by TAL effectors. *Science* 326, 1501–1501. doi: 10.1126/science.1178817
- Needleman, S. B., and Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* 48, 443–453. doi: 10.1016/0022-2836(70)90057-4
- Noel, L. D., Denance, N., and Szurek, B. (2013). Predicting promoters targeted by TAL effectors in plant genomes: from dream to reality. *Front. Plant Sci.* 4:333. doi: 10.3389/fpls.2013.00333
- Notredame, C., Higgins, D. G., and Heringa, J. (2000). T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* 302, 205–217. doi: 10.1006/jmbi.2000.4042
- Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20, 289–290. doi: 10.1093/bioinformatics/btg412
- Penny, D., and Hendy, M. (1985). The use of tree comparison metrics. *Syst. Zool.* 34, 75–82. doi: 10.2307/2413347
- Pereira, A. L., Carazzolle, M. F., Abe, V. Y., De Oliveira, M. L., Domingues, M. N., Silva, J. C., et al. (2014). Identification of putative TAL effector targets of the citrus canker pathogens shows functional convergence underlying disease development and defense response. *BMC Genomics* 15:157. doi: 10.1186/1471-2164-15-157
- Pérez-Quintero, A. L., Rodriguez, R. L., Dereeper, A., Lopez, C., Koebnik, R., Szurek, B., et al. (2013). An improved method for TAL effectors DNA-binding sites prediction reveals functional convergence in TAL repertoires of *Xanthomonas oryzae* strains. *PLoS ONE* 8:e68464. doi: 10.1371/journal.pone.0068464
- Richter, A., Streubel, J., Blucher, C., Szurek, B., Reschke, M., Grau, J., et al. (2014). A TAL effector repeat architecture for frameshift binding. *Nat. Commun.* 5, 3447. doi: 10.1038/ncomms4447
- Schornack, S., Ballvora, A., Gurlebeck, D., Peart, J., Baulcombe, D., Ganap, M., et al. (2004). The tomato resistance protein Bs4 is a predicted non-nuclear TIR-NB-LRR protein that mediates defense responses to severely truncated derivatives of AvrBs4 and overexpressed AvrBs3. *Plant J.* 37, 46–60. doi: 10.1046/j.1365-313X.2003.01937.x
- Schornack, S., Meyer, A., Romer, P., Jordan, T., and Lahaye, T. (2006). Gene-for-gene-mediated recognition of nuclear-targeted AvrBs3-like bacterial effector proteins. *J. Plant Physiol.* 163, 256–272. doi: 10.1016/j.jplph.2005.12.001
- Schornack, S., Minsavage, G. V., Stall, R. E., Jones, J. B., and Lahaye, T. (2008). Characterization of AvrHah1, a novel AvrBs3-like effector from *Xanthomonas gardneri* with virulence and avirulence activity. *New Phytol.* 179, 546–556. doi: 10.1111/j.1469-8137.2008.02487.x
- Schreiber, T., and Bonas, U. (2014). Repeat 1 of TAL effectors affects target specificity for the base at position zero. *Nucleic Acids Res.* 42, 7160–7169. doi: 10.1093/nar/gku341
- Smith, T. F., and Waterman, M. S. (1981). Identification of common molecular subsequences. *J. Mol. Biol.* 147, 195–197. doi: 10.1016/0022-2836(81)90087-5
- Stajich, J. E., Block, D., Boulez, K., Brenner, S. E., Chervitz, S. A., Dagdigian, C., et al. (2002). The Bioperl toolkit: perl modules for the life sciences. *Genome Res.* 12, 1611–1618. doi: 10.1101/gr.361602
- Streubel, J., Blucher, C., Landgraf, A., and Boch, J. (2012). TAL effector RVD specificities and efficiencies. *Nat. Biotechnol.* 30, 593–595. doi: 10.1038/nbt.2304
- Streubel, J., Pesce, C., Hutin, M., Koebnik, R., Boch, J., and Szurek, B. (2013). Five phylogenetically close rice SWEET genes confer TAL effector-mediated susceptibility to *Xanthomonas oryzae* pv. *oryzae*. *New Phytol.* 200, 808–819. doi: 10.1111/nph.12411
- Szalkowski, A. M., and Anisimova, M. (2013). Graph-based modeling of tandem repeats improves global multiple sequence alignment. *Nucleic Acids Res.* 41, e162. doi: 10.1093/nar/gkt628
- Vera Cruz, C. M., Bai, J. F., Ona, I., Leung, H., Nelson, R. J., Mew, T. W., et al. (2000). Predicting durability of a disease resistance gene based on an assessment of the fitness loss and epidemiological consequences of avirulence gene mutation. *Proc. Natl. Acad. Sci. U.S.A.* 97, 13500–13505. doi: 10.1073/pnas.250271997
- Wilkins, K. E., Booher, N. J., Wang, L., and Bogdanove, A. J. (2015). TAL effectors and activation of predicted host targets distinguish Asian from African strains of the rice pathogen *Xanthomonas oryzae* pv. *oryzicola* while strict conservation suggests universal importance of five TAL effectors. *Front. Plant Sci.* 6:536. doi: 10.3389/fpls.2015.00536
- Yang, B., Sugio, A., and White, F. F. (2005). Avoidance of host recognition by alterations in the repetitive and C-terminal regions of AvrXa7, a type III effector of *Xanthomonas oryzae* pv. *oryzae*. *Mol. Plant Microbe Interact.* 18, 142–149. doi: 10.1094/MPMI-18-0142
- Yang, B., Sugio, A., and White, F. F. (2006). Os8N3 is a host disease-susceptibility gene for bacterial blight of rice. *Proc. Natl. Acad. Sci. U.S.A.* 103, 10503–10508. doi: 10.1073/pnas.0604088103
- Yang, B., and White, F. F. (2004). Diverse members of the AvrBs3/PthA family of type III effectors are major virulence determinants in bacterial blight disease of rice. *Mol. Plant Microbe Interact.* 17, 1192–1200. doi: 10.1094/MPMI.2004.17.11.1192
- Yu, Y. H., Streubel, J., Balzergue, S., Champion, A., Boch, J., Koebnik, R., et al. (2011). Colonization of rice leaf blades by an African strain of *Xanthomonas oryzae* pv. *oryzae* depends on a new TAL effector that induces the rice Nodulin-3 Os11N3 gene. *Mol. Plant Microbe Interact.* 24, 1102–1113. doi: 10.1094/MPMI-11-10-0254

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 Pérez-Quintero, Lamy, Gordon, Escalon, Cunnac, Szurek and Gagnevin. 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: MorTALKombat: the story of defense against TAL effectors through loss-of-susceptibility

OPEN ACCESS

Edited by:

Laurent D. Noël,
Centre National de la Recherche
Scientifique, France

Reviewed by:

Sebastian Schornack,
University of Cambridge, UK

***Correspondence:**

Boris Szurek,
UMR IPME, Institut de Recherche
Pour le Développement,
IRD-CIRAD-Université Montpellier 2,
911 Avenue Agropolis BP 64501,
34394 Montpellier Cedex 5, France
boris.szurek@ird.fr

[†]These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Plant Biotic Interactions,
a section of the journal
Frontiers in Plant Science

Received: 31 July 2015

Accepted: 04 August 2015

Published: 21 August 2015

Citation:

Hutin M, Pérez-Quintero AL, Lopez C and Szurek B (2015) Corrigendum:
MorTALKombat: the story of defense
against TAL effectors through
loss-of-susceptibility.
Front. Plant Sci. 6:647.
doi: 10.3389/fpls.2015.00647

Mathilde Hutin^{1†}, **Alvaro L. Pérez-Quintero**^{1†}, **Camilo Lopez**^{1,2} and **Boris Szurek**^{1*}

¹ UMR IPME, Institut de Recherche Pour le Développement, IRD-CIRAD-Université Montpellier 2, Montpellier, France,

² Biology Department, Universidad Nacional de Colombia, Bogota, Colombia

Keywords: *xanthomonas*, plant disease susceptibility S genes, hubs, TAL effectors, agricultural biotechnology, loss-of-function alleles

A corrigendum on

MorTAL Kombat: the story of defense against TAL effectors through loss-of-susceptibility
by Hutin, M., Pérez-Quintero, A. L., Lopez, C., and Szurek, B. (2015). *Front. Plant Sci.* 6:535. doi: 10.3389/fpls.2015.00535

There is an error in the statement about MeSWEET10a function in cassava bacterial blight. The TAL20-dependent activation of *MeSWEET10a* contributes to water soaking symptoms and also to bacterial growth in the plant, in contrast to what is reported in the review. The growth defect seen upon inoculation of Xam668ΔTAL20 is small but it is statistically significant (Cohn et al., 2014). Accordingly, one should also read in Table 1 that TAL20 increases growth and water soaking (column “effect”).

References

Cohn, M., Bart, R. S., Shybut, M., Dahlbeck, D., Gomez, M., Morbitzer, R., et al. (2014). *Xanthomonas axonopodis* virulence is promoted by a transcription activator-like effector-mediated induction of a SWEET sugar transporter in cassava. *Mol. Plant Microbe Interact.* 27, 1186–1198. doi: 10.1094/MPMI-06-14-0161-R

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 Hutin, Pérez-Quintero, Lopez and Szurek. 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.

TAL effectors and the executor *R* genes

Junli Zhang^{1*}, Zhongchao Yin² and Frank White³

¹ Department of Plant Pathology, Kansas State University, Manhattan, KS, USA, ² Temasek Life Sciences Laboratory, National University of Singapore, Singapore, Singapore, ³ Department of Plant Pathology, University of Florida, Gainesville, FL, USA

OPEN ACCESS

Edited by:

Thomas Lahaye,
Ludwig Maximilian University
of Munich, Germany

Reviewed by:

Brian Staskawicz,
University of California, Berkeley, USA
Adam Bogdanove,
Cornell University, USA
Tom Schreiber,
Martin Luther University
of Halle-Wittenberg, Germany

***Correspondence:**

Junli Zhang,
Department of Plant Pathology,
Kansas State University, 4024
Throckmorton Plant Sciences Center,
Manhattan, KS 66506, USA
yuanyuan12543@gmail.com

Specialty section:

This article was submitted to
Plant Biotic Interactions,
a section of the journal
Frontiers in Plant Science

Received: 07 May 2015

Accepted: 02 August 2015

Published: 20 August 2015

Citation:

Zhang J, Yin Z and White F (2015) TAL effectors and the executor *R* genes.
Front. Plant Sci. 6:641.
doi: 10.3389/fpls.2015.00641

Transcription activator-like (TAL) effectors are bacterial type III secretion proteins that function as transcription factors in plants during *Xanthomonas*/plant interactions, conditioning either host susceptibility and/or host resistance. Three types of TAL effector associated resistance (*R*) genes have been characterized—recessive, dominant non-transcriptional, and dominant TAL effector-dependent transcriptional based resistance. Here, we discuss the last type of *R* genes, whose functions are dependent on direct TAL effector binding to discrete effector binding elements in the promoters. Only five of the so-called executor *R* genes have been cloned, and commonalities are not clear. We have placed the protein products in two groups for conceptual purposes. Group 1 consists solely of the protein from pepper, BS3, which is predicted to have catalytic function on the basis of homology to a large conserved protein family. Group 2 consists of BS4C-R, XA27, XA10, and XA23, all of which are relatively short proteins from pepper or rice with multiple potential transmembrane domains. Group 2 members have low sequence similarity to proteins of unknown function in closely related species. Firm predictions await further experimentation on these interesting new members to the *R* gene repertoire, which have potential broad application in new strategies for disease resistance.

Keywords: TAL effectors, *R* gene, *Xanthomonas*

Introduction

Xanthomonas infects monocotyledonous and dicotyledonous plant species, and the pathogenicity of many species depends in part on the effector proteins secreted by a type III secretion (T3S) system (Leyns et al., 1984; Tampakaki et al., 2004). The transcription activator-like (TAL) effector family is a distinct family of type III effectors, which includes members with cognate susceptibility (*S*) and/or resistance (*R*) genes. TAL effectors function as host gene specific transcription factors that can target both *S* and *R* genes, leading to enhanced expression and consequential phenotypic effects (Gu et al., 2005; Yang et al., 2006; Kay et al., 2007; Römer et al., 2007). Susceptibility (*S*) genes are genes with TAL effector-dependent expression and have measurable effects on disease symptoms (Yang et al., 2006; Boch et al., 2014). TAL effector genes are limited to members of the genus *Xanthomonas* and *Ralstonia* (Hopkins et al., 1992; De Feyter et al., 1993; Salanoubat et al., 2002; Heuer et al., 2007). The genes are ubiquitous in some species and have apparent critical functions in a number of diseases (Yang and White, 2004; Cernadas et al., 2014; Cohn et al., 2014; Hu et al., 2014; Schwartz et al., 2015).

Three types of TAL effector associated *R* genes have been reported—recessive, dominant non-transcriptional (classical) and dominant TAL effector-dependent transcriptional based resistance. TAL effector-dependent recessive resistance occurs in rice lines with DNA polymorphisms in *S* gene effector binding elements and will not be discussed in detail here (Hutin et al., 2015). Dominant non-transcriptional based resistance is represented solely by the NBS-LRR resistance gene from tomato,

Bs4, which was identified as the cognate *R* gene to the TAL effector gene *avrBsP/avrBs4* (Bonas et al., 1993; Schornack et al., 2004). However, a transcriptionally functional TAL effector is not required for *Bs4* resistance elicitation as truncated versions of the cognate avirulence gene also trigger resistance. Here, we discuss the third type, namely, TAL effector-dependent *R* genes that are both direct targets of TAL effectors in the host and identified as *R* genes. The genes have been referred to as terminator or, here, executor *E* genes (Bogdanove et al., 2010; Tian et al., 2014). *E* gene expression, like Avr/R gene interactions, is associated with hypersensitive response (HR) on the respective host plants and restricts pathogen growth at the site of infection. Five *E* genes and the cognate TAL effector genes have been cloned, including *Xa27*, *Bs3*, *Bs4C-R*, *Xa10*, and *Xa23* (Gu et al., 2005; Römer et al., 2007; Strauss et al., 2012; Tian et al., 2014; Wang et al., 2015). The TAL effector AvrXa7 may target an as yet uncharacterized *E* gene *Xa7* due to the requirements for the effector nuclear localization signals (NLSs) and the transcription acidic activator domain in *Xa7*-dependent resistance (Hopkins et al., 1992; Yang et al., 2000).

***E* Gene Variation is in the Promoter**

E genes are unique in the panoply of *R* genes in that specificity is not in the *R* gene coding sequence but in the expression of the *R* gene in the presence of the effector (Gu et al., 2005; Römer et al., 2007). The TAL effector, itself, contains two notable regions—the central repetitive region and a C-terminal region with NLS motifs and a potent transcription activation domain (AD). The NLS and AD were shown to be required for *E* gene function in the case of *Bs3*, *Xa10*, *Xa7*, and *Xa27* (Van den Ackerveken et al., 1996; Zhu et al., 1998, 1999; Yang et al., 2000; Szurek et al., 2001). TAL effector specificity is determined by the central repetitive region (Herbers et al., 1992; Yang and Gabriel, 1995; Yang and White, 2004; Yang et al., 2005), and is the structural basis for the TAL effector code, where each repeat specifies the probability of accommodating individual nucleotides (Boch et al., 2009; Moscou and Bogdanove, 2009; Deng et al., 2012; Gao et al., 2012; Mak et al., 2012). The repetitive domain consists of 33–35 amino acid repeats that are polymorphic at amino acid residues 12 and 13, which are referred to as the repeat-variable di-residues (RVDs), each of which can be represented by amino acid residue 13 and corresponds to one DNA base in the effector binding element. Proximal to the N-terminal portion of the repetitive domain are non-canonical repeats that mediate pairing with an initial 5' thymine (Boch et al., 2009; Moscou and Bogdanove, 2009). *E* gene expression occurs upon cognate effector binding to a compatible effector binding element in the respective promoter (**Figure 1A**). The known *E* genes, with the exception of *Xa10*, have dominant and recessive alleles that differ in DNA sequence polymorphisms in the promoter region (**Figure 1B**). AvrBs3, for example, fails to induce *Bs3-E*, an allele of *Bs3* with a 13-bp insertion in the effector binding element in the promoter (**Figure 1B**, iv; Römer et al., 2007, 2009b; Kay et al., 2009). *E* genes, *S* genes, and TAL effector genes, therefore, reflect selective pressures in the evolution of the host and pathogen interaction. In this regard, it is important to note that naturally occurring TAL effectors are not necessarily optimized for the cognate promoters simply in terms of the

binding requirements. Natural TAL effector configurations may reflect adaptive responses to other factors, including the level of target gene expression and frequency of binding sites within a genome.

E Proteins are not Homologs of Classical R Proteins

E proteins are not related on the basis of sequence to any other type of *R* protein. In fact, the proteins, with the exception of the recently reported XA10 and XA23, share no sequence identity with each other. Conceptually, the *E* genes and their products can be divided into two groups. Group 1 consists of proteins that likely have a function in plant development or physiology and whose function has been hijacked by host adaptation to disease. Group 1 consists solely of BS3, which is a member of a conserved family of proteins known as flavin mono-oxygenases (FMO) and, more specifically, a subclass of FMOs known variously as YUCCA or FLOOZY (**Figure 2A**; Römer et al., 2007; Exposito-Rodriguez et al., 2011; Zhao, 2014). Group 2 members, of which there are four, are relatively short proteins that have multiple hydrophobic potential membrane spanning domains (**Figure 2B**). The proteins share no sequence relatedness with proteins of known function and the relatively few related coding sequences occur within close relatives. One related sequence outside the Solanaceae, from grapevine, was reported for *Bs3C-R*. Several of the *E* proteins may have structural similarities. XA27 and XA10 are predicted or have been shown to localize to host cellular membranes and XA10, more specifically, has been shown to localize to the endoplasmic reticulum (ER; Wu et al., 2008; Tian et al., 2014). Prediction software also indicates that BS4C-R may be localized to the ER (Nakai and Horton, 1999; Strauss et al., 2012). It is tempting to speculate that BS3 requires catalytic activity for the *R* gene response and the group 2 proteins function as *R* proteins due to their interaction with host organelles. However, whether the predicted catalytic functions of BS3 are required for the *R* gene response has not been reported, and future analysis of the mechanism-of-action for the respective proteins may indicate some common feature.

***E* Genes in Bacterial Spot Disease on Pepper**

E genes for groups 1 and 2 have been cloned from pepper. The group 1 *Bs3* is recognized by both TAL effectors AvrBs3 and AvrHah from the pathogens *Xanthomonas campestris* pv. *vesicatoria* and *Xanthomonas gardneri*, respectively, both causal organisms of bacterial spot disease of pepper and tomato (Bonas et al., 1998; Schornack et al., 2008). The gene product BS3 is a 342 amino acid protein with a high degree of relatedness with FMOs (Römer et al., 2007; Schornack et al., 2008). FMO proteins are a family of enzymes functioning in all phyla (van Berkel et al., 2006), and play roles in pathogen defense, auxin biosynthesis and metabolism of glucosinolates (Bartsch et al., 2006; Koch et al., 2006; Mishina and Zeier, 2006; Schlaich, 2007). As noted earlier, BS3 falls in a phylogenetic clade consisting of YUCCA and ToFZY members (**Figure 2A**; Römer et al., 2007). The most closely related

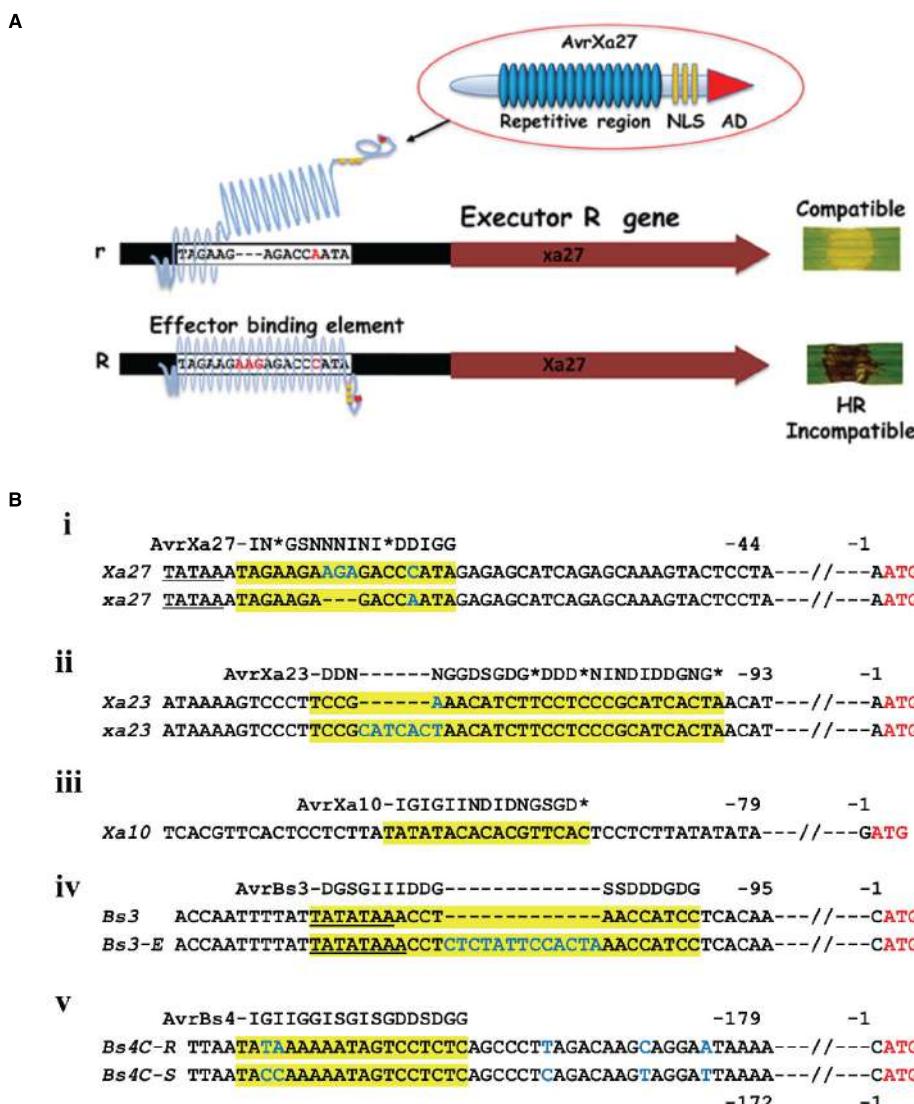
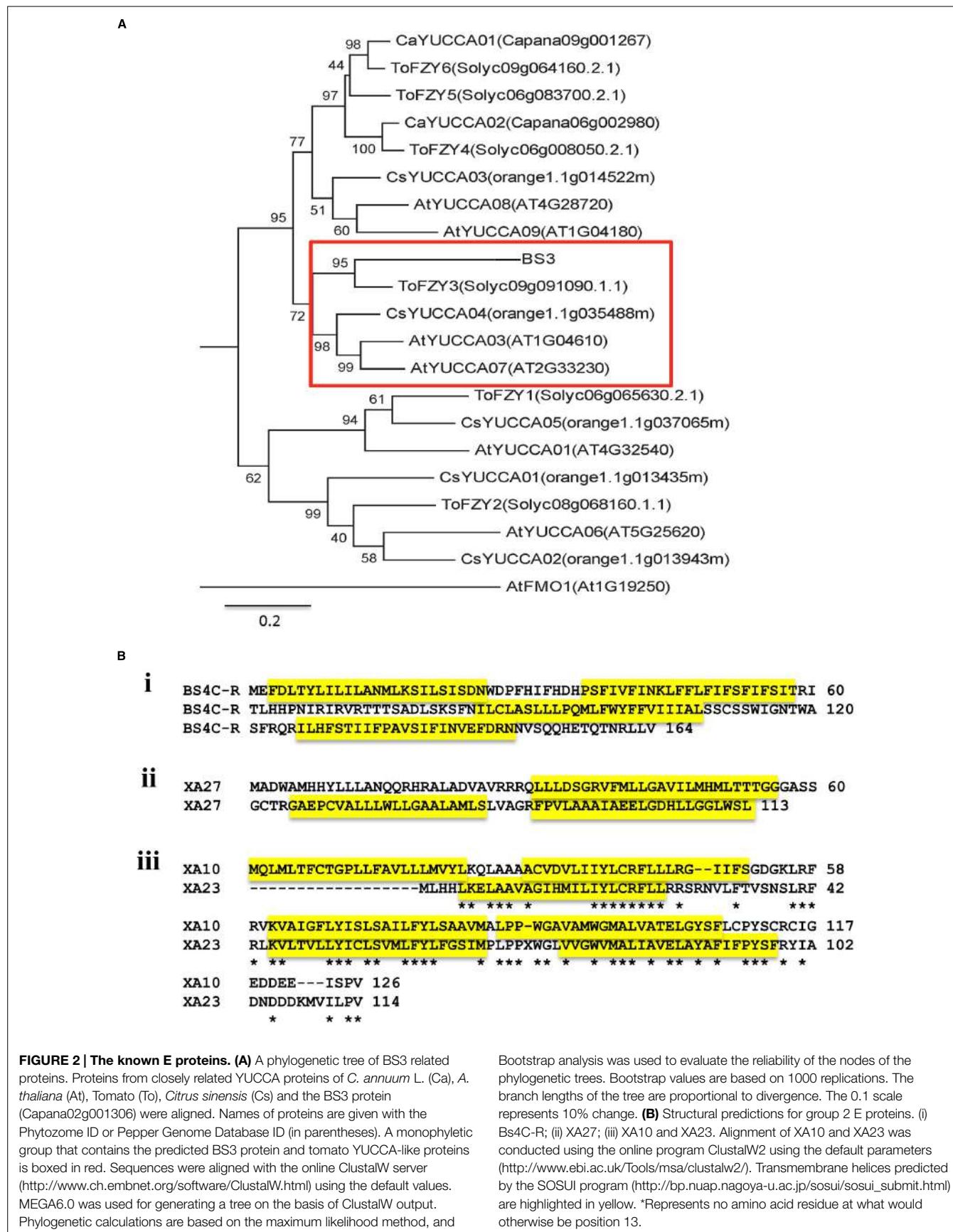


FIGURE 1 | TAL effector and *E* gene interactions. (A) Schematic of the interaction between *AvrXa27* and *Xa27*. Lower case r indicates the ineffective allele that lacks the *AvrXa27* effector binding element. The r allele is missing three nucleotides and different in one nucleotide in the effector binding element of *Xa27* and does not permit binding of *AvrXa7* leading to a compatible interaction (Römer et al., 2010). R indicates the dominant and functional allele of *Xa27*. *Xa27* expression leads to a resistance response and HR on leaves, indicated by the dark discoloring of the inoculation site (here, on a rice leaf). NLS, nuclear localization signal. AD, transcription activation domain of TAL effector. (B) Promoters of *E* genes and polymorphisms in dominant and recessive alleles. (i) Sequence alignment of a part of the promoters of *Xa27* from the rice cultivar IRBB27 (gi|66735941|gb|AY986491.1) and *xa27* from the rice cultivar IR24 (gi|66735943|gb|AY986492.1). (ii) Sequence of a part of the promoter of rice *Xa10* from the rice cultivar IRBB10 (gi|448280729|gb|JX025645.1). (iii) A part of the

promoters of *Xa23* from the rice cultivar CBB23 (gi|721363841|gb|KP123634.1) and *xa23* from the rice cultivar JG30 (gi|721363854|gb|KP123635.1). (iv) A part of the promoters of *Bs3* from *Capsicum annuum* L. cultivar ECW-30R (gi|158851516|gb|EU078684.1) and *Bs3-E* from *C. annuum* L. cultivar ECW (gi|158851512|gb|EU078683.1). (v) A part of the promoters of *Bs4C-R* from *pubescens* cultivar PI 235047 (gi|414148024|gb|JX944826.1) and *Bs4C-S* from *Capsicum pubescens* cultivar PI 585270 (gi|414148026|gb|JX944827.1). The ATG start codon in each case is displayed in red letters. Nucleotides that are identical between the alleles are displayed as black letters. Predicted TATA boxes are underlined. Effector binding elements are highlighted in yellow with blue letters indicating differences between alleles. The TAL effectors are represented by the repeat regions using a single letter represents each RVD (I-NI; G-NG or HG; S-NS; D-HD or ND, *-N*, N-NN). *Represents no amino acid residue at what would otherwise be position 13.

proteins to BS3 have been demonstrated to be involved in auxin biosynthesis and a variety of developmental and physiological responses (Exposito-Rodriguez et al., 2011; Stepanova et al., 2011; Lee et al., 2012; Hentrich et al., 2013; Zhao, 2014). YUCCA/FLOOZY members catalyze a key intermediate in the plant pathway from indole-3-pyruvate (IPA) into indole-3-acetic

acid (IAA) through oxidative decarboxylation reaction (Kim et al., 2011; Stepanova et al., 2011; Dai et al., 2013; Hentrich et al., 2013; Zhao, 2014). A homolog from tomato, *ToFZY*, also functions in auxin biosynthesis (Exposito-Rodriguez et al., 2011). A more distant relative of unknown enzymatic activity, *AtFMO1*, plays a role in systemic acquired resistance (Mishina and Zeier, 2006).



Bs4C-R encodes a member of our group 2 E proteins and is expressed in the presence of the TAL effector AvrBs4 (Strauss et al., 2012). *Bs4C-R* is the only *E* gene isolated on the basis of differential expression between resistant and susceptible cultivars and not the typical gene mapping strategy. A two-nucleotide polymorphism in the region of the effector binding element of a susceptible allele *Bs4C-S* leads to the failure of induction of an AvrBs4-dependent HR (**Figure 1B**, v; Strauss et al., 2012). Both the dominant and recessive alleles encode functionally competent proteins as constitutive expression of either *Bs4C-R* or *Bs4C-S* triggered HR *Nicotiana benthamiana* in leaves (Strauss et al., 2012).

E Genes in Bacterial Blight Disease of Rice

The *E* genes of rice are all included in our group 2 and provide resistance to bacterial blight disease. Bacterial blight of rice is caused by *Xanthomonas oryzae* pv. *oryzae*, and TAL effectors are major avirulence factors for *X. oryzae* pv. *oryzae* when the cognate *E* genes are present in host plants (Mew, 1987; White and Yang, 2009). Three pairs of TAL effectors and cognate *E* genes have been cloned from rice—AvrXa27/Xa27, AvrXa10/Xa10, and AvrXa23/Xa23. No cognate *S* genes or virulence effects for the TAL effectors of AvrXa10, AvrXa23, or AvrXa27 in compatible host cultivars have been reported, despite the presence of AvrXa27 and AvrXa23 in many extant strains of *X. oryzae* pv. *oryzae* (Gu et al., 2004; Wang et al., 2014).

The *Xa27* product is a protein of 113 amino acids without any clear homologs based on sequence similarity in plants other than rice and several related species of the *Oryza* genus (Gu et al., 2005; Bimolata et al., 2013). The resistance conferred by *Xa27* is affected by developmental stage, increasing with the age of the plants and reaching maximum resistance at 5 weeks. Moreover, *Xa27* showed a dosage effect in the cultivar CO39 genetic background (Gu et al., 2004). At least two transmembrane α -helix domains were predicted, depending on the prediction software (Gu et al., 2005). Here, we show three based on the SOSUI program (**Figure 2B**, Hirokawa et al., 1998). Further experimentation has shown that the protein XA27 localizes to cytoplasmic membrane, and some protein appears in the apoplast after plasmolysis (Wu et al., 2008). Localization is dependent on the N-terminal signal anchor-like sequence, which is also essential for resistance to *X. oryzae* pv. *oryzae* (Wu et al., 2008). The protein itself appears to be toxic as gene transfer to compatible rice lines occurs with a reduced efficiency. Nevertheless recombinant lines were recovered, demonstrating that the AvrXa27-dependency of the resistance is indeed linked to the *Xa27* locus (Gu et al., 2005). At the same time, lines were obtained that had elevated expression of *Xa27* and displayed defense reactions, including thickened vascular elements, even in the absence of bacterial inoculation (Gu et al., 2005). The effector binding element is located immediately downstream of the predicted TATA box, and the recessive allele *xa27* in the susceptible rice cultivar IR24 encodes the same protein but has a three-nucleotide deletion and one nucleotide difference in comparison to *Xa27* (**Figure 1B**, i; Römer et al., 2009a). DNA sequence alignment of *Xa27* alleles from 27 lines representing four *Oryza* species revealed that a *Xa27*-related coding sequence was indeed present in all of the

lines. However, only the IRBB27 allele appears to possess the necessary effector binding element for AvrXa27 (Bimolata et al., 2013). A synthetic TAL effector directed at the recessive allele in IR24 induced a resistance reaction, indicating the product of the recessive allele could function similarly to *Xa27*, if expressed (Li et al., 2013).

Xa10 encodes a 126-amino acid protein, containing four potential transmembrane helices (Tian et al., 2014). A consensus effector binding element is present in the promoter region of *Xa10* (**Figure 1B**, iii). *Xa10* differs from *Xa27* and *Bs4C-R* in sequence and by the lack of a nearly identical coding sequence in susceptible plant lines. At the same time, related sequences are found in other lines, including *Xa23* (Wang et al., 2015). Ectopic and weak expression of *Xa10* in rice causes a lesion mimic-like phenotype, while transient expression of *Xa10* in *N. benthamiana* and rice induced HR in plants (Tian et al., 2014). Under the appropriate promoter, *Xa10* also induced programmed cell death (PCD) in mammalian HeLa cells (Tian et al., 2014). In both rice and *N. benthamiana* cells, hydrogen peroxide, swelling and degradation were detected in chloroplasts. Degradation of mitochondria was also observed, supporting the model that XA10 functions as a general inducer of PCD in plant and animal cells (Tian et al., 2014). Further functional characterization revealed that XA10 forms hexamers, localizes on the ER membrane of plant and HeLa cells, and mediates Ca^{2+} depletion, which is consistent with some processes of PCD (Pinton et al., 2008; Williams et al., 2014).

The *E* gene *Xa23* encodes a 113-amino acid protein that shares approximately 50% amino acid sequence identity and 64% nucleotide sequence similarity with XA10 and *Xa10*, respectively (Wang et al., 2015). An identical recessive allele is present on the basis of the coding region, and characterization of the effector binding element of AvrXa23 revealed a 7-bp polymorphism accounts for the failure of *xa23* induction in the recessive rice varieties (**Figure 1B**, ii). The susceptible cultivar JG30, with *xa23*, became resistant to PXO99^A harboring a designed TAL effector specifically targeting the *xa23* promoter region including the 7-bp polymorphism (Wang et al., 2015). Moreover, *Agrobacterium*-mediated transient transformation of *Xa23* indicated that, like *Xa10*, *Xa23* induced an HR in *N. benthamiana*, and also induced an HR in tomato (Wang et al., 2015). Both XA10 and XA23 have a motif of unknown function that is comprised of five acidic amino acid residues (EDDEE and DNDDD, respectively) at the C-termini (Tian et al., 2014). Alteration of the so-called ED motif in XA10 abolished HR activity (Tian et al., 2014).

Prospects for *E* Genes in Disease Control

The question arises whether, as dominant major genes for resistance, the genes are effective in control of the respective diseases. In rice, only *Xa10* has been deployed in field conditions and is effective against a few extant races of the pathogens (Vera-Cruz et al., 2000; Mishra et al., 2013). *Xa27* has been introduced into breeding programs (Luo et al., 2012; Luo and Yin, 2013). *Xa23* and *Xa7* are also in the process of introduction into various breeding programs (Perez et al., 2008; Huang et al., 2012). But how durable is *E* gene mediated resistance? Bacteria can rapidly evolve to avoid *R* gene recognition through avirulence gene loss under

high selection pressure for virulence (Koskineni et al., 2012). More specifically, TAL effectors appear to reflect exquisitely the selective forces of evolution in the form of the repetitive domain. Deletion of repeats in AvrBs3, for example, resulted in the loss of the induction of *Bs3* (Römer et al., 2007). Deployment of an *E* gene that targets a critical TAL effector for virulence has been proposed as an approach to make adaptation less likely as the pathogen would have to maintain virulence in addition to losing *Xa7* recognition. Indeed, *Xa7*, which is triggered by the major TAL effector AvrXa7, was found to be durable in field tests in the Philippines (Vera-Cruz et al., 2000). Field isolated strains that arose showed loss of the ability to induce resistance and were weakly virulent presumably due to associated mutations in *avrXa7* (Vera-Cruz et al., 2000). However, *in vitro* rearrangements and *in vivo* selection for loss of AvrXa7-mediated resistance produced gene variants that maintained strain virulence and avoided *Xa7*-mediated resistance (Yang et al., 2005). Furthermore, a number of extant TAL effectors target the same *S* gene as AvrXa7, namely *OsSWEET14*, without activating *Xa7*-dependent resistance (Antony et al., 2010; Yu et al., 2011; Streubel et al., 2013). Strains can also acquire other major TAL effectors that target alternative paralogs of that *S* gene (Yang et al., 2006; Streubel et al., 2013; Zhou et al., 2015).

In India, field strain surveys have found a diversity of strains, many without AvrXa7 activity, indicating that any benefits in the deployment of *Xa7* would be short-lived (Mishra et al., 2013). Thus, broad application of a single *E* gene like *Xa7* in some environments appears to be limited. At the same time, local conditions, such as in the Philippine tests, may limit the invasion of a particular TAL effector gene into extant pathogen populations and deployment may be both broad and durable (Vera-Cruz et al., 2000). *Xa27* and *Xa23* are interesting from the perspective that the cognate avirulence genes are present in many strains and, therefore, broadly effective (Gu et al., 2004; Wang et al., 2014). In contrast to AvrXa7, no cognate *S* genes have been reported for AvrXa27 or AvrXa23, so we can speculate that these effectors may provide some fitness to the pathogen which has not been detected in laboratory or greenhouse assays so far. AvrBs3 has a phenotypic effect for strains of *Xanthomonas euvesicatoria* harboring the gene; a fitness benefit for the effector has been observed (Marois et al., 2002; Wichmann and Bergelson, 2004), and *Bs3* is effective for many pepper strains of *X. euvesicatoria*. In addition, *Bs3* is also effective against the emerging pepper pathogen *X. gardneri*, which harbors the TAL effector AvrHah1 (Schornack et al., 2008; Schwartz et al., 2015).

***E* Genes and New Strategies for Resistance**

Despite possible shortcomings of endogenous *E* genes, *E* genes hold great potential for breeding broadly and durably resistant crop varieties. Specifically for TAL effector associated diseases, *E* genes can be constructed with so-called super-promoters, consisting of multiple effector binding sites, each recognizing specific corresponding TAL effectors that are expressed in the pathogen populations (Römer et al., 2009a; Hummel et al., 2012; Zeng et al., 2015). *Xa27* was fused to a super promoter including

binding sites for three TAL effectors from *X. oryzae* pv. *oryzae* and three from the bacterial leaf streak pathogen *X. oryzae* pv. *oryzicola*. The plants were resistant to several *X. oryzae* pv. *oryzae* and *X. oryzae* pv. *oryzicola* strains that were originally compatible on wild type homozygous *Xa27* plants (Hummel et al., 2012). Similarly, transgenic rice lines containing *Xa10^{E5}* with binding elements to five TAL effectors proved to be resistant to 27 of the 28 selected *X. oryzae* pv. *oryzae* strains gathered from 11 countries (Zeng et al., 2015). Judicial choices of the effector binding sites for TAL effectors in extant populations may provide resilient barriers to TAL effector associated diseases. However, due to the risk that an added effector binding element might coincidentally contains a *cis* regulatory element which could induce the *E* gene expression in response to particular stimuli and cause cell death without challenge of TAL effectors, such amended promoters should be tested thoroughly before deployment (Hummel et al., 2012). Another approach is to engineer an *E* gene to be under the control of a different type of pathogen inducible promoter. For example, expression of *Xa27* under the control of an disease inducible or defense gene promoter, in this case, the rice PR1 promoter, which is induced by both compatible and incompatible bacteria, conferred broad resistance to *X. oryzae* pv. *oryzae* strains (Gu et al., 2005). This strategy need not be limited to *Xanthomonas* related diseases.

Conclusion

Thirteen *R* genes have been cloned for resistance to *Xanthomonas* diseases—all coming from rice, pepper, or tomato. Four, in addition to *Bs4*, are representatives of the two major classes of *R* genes, the receptor linked kinases (RLK) and nucleotide binding site leucine rich repeat (NBS-LRR) genes which are represented by *Xa21* (RLK, rice), *Xa26* (RLK, rice), *Xa1* (NBS-LRR, rice), and *Bs2* (NBS-LRR, pepper; Yoshimura et al., 1998; Tai et al., 1999; Zhang and Wang, 2013). Three cloned genes are recessive genes from rice and, although not discussed here, can be considered cases of loss of susceptibility (White and Yang, 2009). The five *E* genes and the protein products bear little or no resemblance to the other *R* genes, or, for that matter, other common defense response components. *E* genes, at least phenotypically, trigger host responses, in particular the HR, similarly to some other *R* gene mediated resistances. Whether the *E* proteins intersect other *R* protein mediated resistance pathways in plants remains unknown. Evidence for XA10 indicates that the protein activates PCD, possibly through ER-stress in light of the association with the ER (Williams et al., 2014). Further research into *E* gene functions should enhance their utility for new resistance strategies as well as improve our understanding on plant defense and PCD pathways.

Acknowledgments

The authors thank the members of the Yin laboratory for reviewing the manuscript. JZ and FW are supported by funds from National Science Foundation research award 1238189. ZY is supported by the National Research Foundation and Office of the Prime Minister Competitive Research Programme Singapore award NRF-CRP7-2010-02.

References

- Antony, G., Zhou, J., Huang, S., Li, T., Liu, B., White, F., et al. (2010). Rice *xa13* recessive resistance to bacterial blight is defeated by induction of the disease susceptibility gene *Os11N3*. *Plant Cell* 22, 3864–3876. doi: 10.1105/tpc.110.078964
- Bartsch, M., Gobbato, E., Bednarek, P., Debey, S., Schultze, J. L., Bautor, J., et al. (2006). Salicylic acid-independent ENHANCED DISEASE SUSCEPTIBILITY1 signaling in *Arabidopsis* immunity and cell death is regulated by the monooxygenase FMO1 and the nudix hydrolase NUDT7. *Plant Cell* 18, 1038–1051. doi: 10.1105/tpc.105.039982
- Bimolata, W., Kumar, A., Sundaram, R. M., Laha, G. S., Qureshi, I. A., Reddy, G. A., et al. (2013). Analysis of nucleotide diversity among alleles of the major bacterial blight resistance gene *Xa27* in cultivars of rice (*Oryza sativa*) and its wild relatives. *Planta* 238, 293–305. doi: 10.1007/s00425-013-1891-3
- Boch, J., Bonas, U., and Lahaye, T. (2014). TAL effectors—pathogen strategies and plant resistance engineering. *New Phytol.* 204, 823–832. doi: 10.1111/nph.13015
- Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., et al. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 326, 1509–1512. doi: 10.1126/science.1178811
- Bogdanove, A. J., Schornack, S., and Lahaye, T. (2010). TAL effectors: finding plant genes for disease and defense. *Curr. Opin. Plant Biol.* 13, 394–401. doi: 10.1016/j.pbi.2010.04.010
- Bonas, U., Conradts-Strauch, J., and Balbo, I. (1993). Resistance in tomato to *Xanthomonas campestris* pv *vesicatoria* is determined by alleles of the pepper-specific avirulence gene *avrBs3*. *Mol. Gen. Genet.* 238, 261–269.
- Bonas, U., Stall, R. E., and Staskawicz, B. (1998). Genetic and structural characterization of the avirulence gene *avrBs3* from *Xanthomonas campestris* pv. *vesicatoria*. *Mol. Gen. Genet.* 218, 127–136. doi: 10.1007/BF00330575
- Cernadas, R. A., Doyle, E. L., Nino-Liu, D. O., Wilkins, K. E., Bancroft, T., Wang, L., et al. (2014). Code-assisted discovery of TAL effector targets in bacterial leaf streak of rice reveals contrast with bacterial blight and a novel susceptibility gene. *PLoS Pathog.* 10:e1003972. doi: 10.1371/journal.ppat.1003972
- Cohn, M., Bart, R. S., Shybut, M., Dahlbeck, D., Gomez, M., Morbitzer, R., et al. (2014). *Xanthomonas axonopodis* virulence is promoted by a transcription activator-like effector mediated induction of a SWEET sugar transporter in cassava. *Mol. Plant Microbe Interact.* 27, 1186–1198. doi: 10.1094/MPMI-06-14-0161-R
- DAI, X., Mashiguchi, K., Chen, Q., Kasahara, H., Kamiya, Y., Ojha, S., et al. (2013). The biochemical mechanism of auxin biosynthesis by an *Arabidopsis* YUCCA flavin-containing monooxygenase. *J. Biol. Chem.* 288, 1448–1457. doi: 10.1074/jbc.M112.424077
- De Feyter, R., Yang, Y., and Gabriel, D. W. (1993). Gene-for-genes interactions between cotton *R* genes and *Xanthomonas campestris* pv. *malvacearum* avr genes. *Mol. Plant Microbe Interact.* 6, 225. doi: 10.1094/MPMI-6-225
- Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., Zhu, J., et al. (2012). Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* 335, 720–723. doi: 10.1126/science.1215670
- Exposito-Rodriguez, M., Borges, A. A., Borges-Perez, A., and Perez, J. A. (2011). Gene structure and spatiotemporal expression profile of tomato genes encoding YUCCA-like flavin monooxygenases: the *ToFZY* gene family. *Plant Physiol. Biochem.* 49, 782–791. doi: 10.1016/j.plaphy.2011.02.022
- Gao, H., Wu, X., Chai, J., and Han, Z. (2012). Crystal structure of a TALE protein reveals an extended N-terminal DNA binding region. *Cell Res.* 22, 1716–1720. doi: 10.1038/cr.2012.156
- Gu, K., Tian, D., Yang, F., Wu, L., Sreekala, C., Wang, D., et al. (2004). High-resolution genetic mapping of *Xa27(t)*, a new bacterial blight resistance gene in rice, *Oryza sativa* L. *Theor. Appl. Genet.* 108, 800–807. doi: 10.1007/s00122-003-1491-x
- Gu, K. Y., Yang, B., Tian, D. S., Wu, L. F., Wang, D. J., Sreekala, C., et al. (2005). *R* gene expression induced by a type-III effector triggers disease resistance in rice. *Nature* 435, 1122–1125. doi: 10.1038/nature03630
- Henrich, M., Sanchez-Parral, B., Perez Alonso, M., Carrasco Loba, V., Carrillo, L., Vicente-Carbajosa, J., et al. (2013). YUCCA8 and YUCCA9 overexpression reveals a link between auxin signaling and lignification through the induction of ethylene biosynthesis. *Plant Signal. Behav.* 8, e26363–e26363. doi: 10.4161/psb.26363
- Herbers, K., Conrads-Strauch, J., and Bonas, U. (1992). Race-specificity of plant resistance to bacterial spot disease determined by repetitive motifs in a bacterial avirulence protein. *Nature* 356, 172–174. doi: 10.1038/356172a0
- Heuer, H., Yin, Y., Xue, Q., Smalla, K., and Guo, J. (2007). Repeat domain diversity of *avrBs3*-like genes in *Ralstonia solanacearum* strains and association with host preferences in the field. *Appl. Environ. Microbiol.* 73, 4379. doi: 10.1128/AEM.00367-07
- Hirokawa, T., Boon-Chieng, S., and Mitaku, S. (1998). SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics* 14, 378–379. doi: 10.1093/bioinformatics/14.4.378
- Hopkins, C. M., White, F. F., Choi, S. H., Guo, A., and Leach, J. E. (1992). Identification of a family of avirulence genes from *Xanthomonas oryzae* pv *oryzae*. *Mol. Plant Microbe Interact.* 5, 451–459. doi: 10.1094/MPMI-5-451
- Huang, B., Xu, J. Y., Hou, M. S., Ali, J., and Mou, T. M. (2012). Introgression of bacterial blight resistance genes *Xa7*, *Xa21*, *Xa22* and *Xa23* into hybrid rice restorer lines by molecular marker-assisted selection. *Euphytica* 187, 449–459. doi: 10.1007/s10681-012-0758-1
- Hu, Y., Zhang, J., Jia, H., Sosso, D., Li, T., Frommer, W. B., et al. (2014). Lateral organ boundaries 1 is a disease susceptibility gene for citrus bacterial canker disease. *Proc. Natl. Acad. Sci. U.S.A.* 111, E521–E529. doi: 10.1073/pnas.1313271111
- Hummel, A. W., Doyle, E. L., and Bogdanove, A. J. (2012). Addition of transcription activator-like effector binding sites to a pathogen strain-specific rice bacterial blight resistance gene makes it effective against additional strains and against bacterial leaf streak. *New Phytol.* 195, 883–893. doi: 10.1111/j.1469-8137.2012.04216.x
- Hutin, M., Pérez-Quintero, A., Lopez, C., and Szurek, B. (2015). MorTAL Kombat: the story of defense against TAL effectors through loss-of-susceptibility. *Front. Microbiol.* 6:535. doi: 10.3389/fmicb.2015.00535
- Kay, S., Hahn, S., Marois, E., Hause, G., and Bonas, U. (2007). A bacterial effector acts as a plant transcription factor and induces a cell size regulator. *Science* 318, 648–651. doi: 10.1126/science.1144956
- Kay, S., Hahn, S., Marois, E., Wieduwild, R., and Bonas, U. (2009). Detailed analysis of the DNA recognition motifs of the *Xanthomonas* type III effectors AvrBs3 and AvrBs3 delta rep16. *Plant J.* 59, 859–871. doi: 10.1111/j.1365-313X.2009.03922.x
- Kim, J. I., Murphy, A. S., Baek, D., Lee, S., Yun, D., Bressan, R. A., et al. (2011). YUCCA6 over-expression demonstrates auxin function in delaying leaf senescence in *Arabidopsis thaliana*. *J. Exp. Bot.* 62, 3981–3992. doi: 10.1093/jxb/err094
- Koch, M., Vorwerk, S., Masur, C., Sharifi-Sirchi, G., Olivieri, N., and Schlaich, N. L. (2006). A role for a flavin-containing mono-oxygenase in resistance against microbial pathogens in *Arabidopsis*. *Plant J.* 47, 629–639. doi: 10.1111/j.1365-313X.2006.02813.x
- Koskineni, S., Sun, S., Berg, O. G., and Andersson, D. I. (2012). Selection-driven gene loss in bacteria. *PLoS Genet.* 8:e1002787. doi: 10.1371/journal.pgen.1002787
- Lee, M., Jung, J., Han, D., Seo, P. J., Park, W. J., and Park, C. (2012). Activation of a flavin monooxygenase gene YUCCA7 enhances drought resistance in *Arabidopsis*. *Planta* 235, 923–938. doi: 10.1007/s00425-011-1552-3
- Leyns, F., De Cleene, M., Swings, J., and De Ley, J. (1984). The host range of the genus *Xanthomonas*. *Bot. Rev.* 50, 308–356. doi: 10.1007/BF02862635
- Li, T., Huang, S., Zhou, J., and Yang, B. (2013). Designer TAL effectors induce disease susceptibility and resistance to *Xanthomonas oryzae* pv. *oryzae* in rice. *Mol. Plant* 6, 781–789. doi: 10.1093/mp/sst034
- Luo, Y., Sangha, J. S., Wang, S., Li, Z., Yang, J., and Yin, Z. (2012). Marker-assisted breeding of *Xa4*, *Xa21* and *Xa27* in the restorer lines of hybrid rice for broad-spectrum and enhanced disease resistance to bacterial blight. *Mol. Breed.* 30, 1601–1610. doi: 10.1007/s11032-012-9742-7
- Luo, Y., and Yin, Z. (2013). Marker-assisted breeding of Thai fragrance rice for semi-dwarf phenotype, submergence tolerance and disease resistance to rice blast and bacterial blight. *Mol. Breed.* 32, 709–721. doi: 10.1007/s11032-013-9904-2
- Mak, A. N., Bradley, P., Cernadas, R. A., Bogdanove, A. J., and Stoddard, B. L. (2012). The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science* 335, 716–719. doi: 10.1126/science.1216211
- Marois, E., Van den Ackerveken, G., and Bonas, U. (2002). The *Xanthomonas* type III effector protein AvrBs3 modulates plant gene expression and induces cell hypertrophy in the susceptible host. *Mol. Plant Microbe Interact.* 15, 637–646. doi: 10.1094/MPMI.2002.15.7.637

- Mew, T. W. (1987). Current status and future prospects of research on bacterial blight of rice. *Ann. Rev. Phytopathol.* 25, 559–382. doi: 10.1146/annurev.py.25.090187.002043
- Mishina, T. E., and Zeier, J. (2006). The *Arabidopsis* flavin-dependent monooxygenase FMO1 is an essential component of biologically induced systemic acquired resistance. *Plant Physiol.* 141, 1666–1675. doi: 10.1104/pp.106.081257
- Mishra, D., Vishnupriya, M. R., Anil, M. G., Konda, K., Raj, Y., and Sonti, R. V. (2013). Pathotype and genetic diversity amongst Indian isolates of *Xanthomonas oryzae* pv. *oryzae*. *PLoS ONE* 8:e81996. doi: 10.1371/journal.pone.0081996
- Moscou, M. J., and Bogdanove, A. J. (2009). A simple cipher governs DNA recognition by TAL effectors. *Science* 326, 1501–1501. doi: 10.1126/science.1178817
- Nakai, K., and Horton, P. (1999). PSORT: a program for detecting the sorting signals of proteins and predicting their subcellular localization. *Trends Biochem. Sci.* 24, 34–35. doi: 10.1016/s0968-0004(98)01336-x
- Perez, L. M., Redoña, E. D., Mendioro, M. S., Vera Cruz, C. M., and Leung, H. (2008). Introgression of *Xa4*, *Xa7* and *Xa21* for resistance to bacterial blight in thermosensitive genetic male sterile rice (*Oryza sativa* L.) for the development of two-line hybrids. *Euphytica* 164, 627–636. doi: 10.1007/s10681-008-9653-1
- Pinton, P., Giorgi, C., Siviero, R., Zecchini, E., and Rizzuto, R. (2008). Calcium and apoptosis: ER-mitochondria Ca^{2+} transfer in the control of apoptosis. *Oncogene* 27, 6407–6418. doi: 10.1038/onc.2008.308
- Römer, P., Hahn, S., Jordan, T., Strauss, T., Bonas, U., and Lahaye, T. (2007). Plant pathogen recognition mediated by promoter activation of the pepper *Bs3* resistance gene. *Science* 318, 645–648. doi: 10.1126/science.1144958
- Römer, P., Recht, S., and Lahaye, T. (2009a). A single plant resistance gene promoter engineered to recognize multiple TAL effectors from disparate pathogens. *Proc. Natl. Acad. Sci. U.S.A.* 106, 20526–20531. doi: 10.1073/pnas.0908812106
- Römer, P., Strauss, T., Hahn, S., Scholze, H., Morbitzer, R., Grau, J., et al. (2009b). Recognition of AvrBs3-like proteins is mediated by specific binding to promoters of matching pepper *Bs3* alleles. *Plant Physiol.* 150, 1697–1712. doi: 10.1104/pp.109.139931
- Römer, P., Recht, S., Strauss, T., Elsaesser, J., Schornack, S., Boch, J., et al. (2010). Promoter elements of rice susceptibility genes are bound and activated by specific TAL effectors from the bacterial blight pathogen *Xanthomonas oryzae* pv. *oryzae*. *New Phytol.* 187, 1048–1057. doi: 10.1111/j.1469-8137.2010.03217.x
- Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S., Arlat, M., et al. (2002). Genome sequence of the plant pathogen *Ralstonia solanacearum*. 415, 497–502. doi: 10.1038/415497a
- Schlachter, N. L. (2007). Flavin-containing monooxygenases in plants: looking beyond detox. *Trends Plant Sci.* 12, 412–418. doi: 10.1016/j.tplants.2007.08.009
- Schornack, S., Ballvora, A., Gurlebeck, D., Peart, J., Ganap, M., Baker, B., et al. (2004). The tomato resistance protein Bs4 is a predicted non-nuclear TIR-NB-LRR protein that mediates defense responses to severely truncated derivatives of AvrBs4 and overexpressed AvrBs3. *Plant J.* 37, 46–60. doi: 10.1046/j.1365-313X.2003.01937.x
- Schornack, S., Minsavage, G. V., Stall, R. E., Jones, J. B., and Lahaye, T. (2008). Characterization of AvrHah1, a novel AvrBs3-like effector from *Xanthomonas gardneri* with virulence and avirulence activity. *New Phytol.* 179, 546–556. doi: 10.1111/j.1469-8137.2008.02487.x
- Schwartz, A. R., Neha, P. N., Sujan, T. S., Mark, W. M., José, P. J., Joaquim, M. J., et al. (2015). Phylogenomics of *Xanthomonas* field strains infecting pepper and tomato reveals diversity in effector repertoires and identifies determinants of host specificity. *Front. Microbiol.* 6:535. doi: 10.3389/fmicb.2015.00535
- Stepanova, A. N., Yun, J., Robles, L. M., Novak, O., He, W., Guo, H., et al. (2011). The *Arabidopsis* YUCCA1 flavin monooxygenase functions in the indole-3-pyruvic acid branch of auxin biosynthesis. *Plant Cell* 23, 3961–3973. doi: 10.1104/tpc.111.088047
- Strauss, T., van Poecke, R. M. P., Strauss, A., Römer, P., Minsavage, G. V., Singh, S., et al. (2012). RNA-seq pinpoints a *Xanthomonas* TAL-effector activated resistance gene in a large-crop genome. *Proc. Natl. Acad. Sci. U.S.A.* 109, 19480–19485. doi: 10.1073/pnas.1212415109
- Streubel, J., Pesce, C., Hutin, M., Koebnik, R., Boch, J., and Szurek, B. (2013). Five phylogenetically close rice SWEET genes confer TAL effector-mediated susceptibility to *Xanthomonas oryzae* pv. *oryzae*. *New Phytol.* 200, 808–819. doi: 10.1111/nph.12411
- Szurek, B., Marois, E., Bonas, U., and Van den Ackerveken, G. (2001). Eukaryotic features of the *Xanthomonas* type III effector AvrBs3: protein domains involved in transcriptional activation and the interaction with nuclear import receptors from pepper. *Plant J.* 26, 523–534. doi: 10.1046/j.0960-7412.2001.01046.x
- Tai, T. H., Dahlbeck, D., Clark, E. T., Gajiwala, P., Pasion, R., Whalen, M. C., et al. (1999). Expression of the *Bs2* pepper gene confers resistance to bacterial spot disease in tomato. *Proc. Natl. Acad. Sci. U.S.A.* 96, 14153–14158. doi: 10.1073/pnas.96.24.14153
- Tampakaki, A. P., Fadouoglou, V. E., Gazi, A. D., Panopoulos, N. J., and Kokkinidis, M. (2004). Conserved features of type III secretion. *Cell. Microbiol.* 6, 805–816. doi: 10.1111/j.1462-5822.2004.00432.x
- Tian, D., Wang, J., Zeng, X., Gu, K., Qiu, C., Yang, X., et al. (2014). The rice TAL effector-dependent resistance protein XA10 triggers cell death and calcium depletion in the endoplasmic reticulum. *Plant Cell* 26, 497–515. doi: 10.1105/tpc.113.119255
- van Berkel, W. J. H., Kamerbeek, N. M., and Fraaije, M. W. (2006). Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts. *J. Biotechnol.* 124, 670–689. doi: 10.1016/j.biote.2006.03.044
- Van den Ackerveken, G., Marois, E., and Bonas, U. (1996). Recognition of the bacterial avirulence protein AvrBs3 occurs inside the host plant cell. *Cell* 87, 1307–1316. doi: 10.1016/S0092-8674(00)81825-5
- Vera-Cruz, C. M. V., Bai, J. F., Ona, I., Leung, H., Nelson, R. J., Mew, T. W., et al. (2000). Predicting durability of a disease resistance gene based on an assessment of the fitness loss and epidemiological consequences of avirulence gene mutation. *Proc. Natl. Acad. Sci. U.S.A.* 97, 13500–13505. doi: 10.1073/pnas.250271997
- Wang, C., Qin, T., Yu, H., Zhang, X., Che, J., Gao, Y., et al. (2014). The broad bacterial blight resistance of rice line CBB23 is triggered by a novel transcription activator-like (TAL) effector of *Xanthomonas oryzae* pv. *oryzae*. *Mol. Plant Pathol.* 15, 333–341. doi: 10.1111/mpp.12092
- Wang, C., Zhang, X., Fan, Y., Gao, Y., Zhu, Q., Zheng, C., et al. (2015). XA23 is an executor R protein and confers broad-spectrum disease resistance in rice. *Mol. Plant Pathol.* 8, 290–302. doi: 10.1016/j.molp.2014.10.010
- Wichmann, G., and Bergelson, J. (2004). Effector genes of *Xanthomonas axonopodis* pv. *vesicatoria* promote transmission and enhance other fitness traits in the field. *Genetics* 166, 693–706. doi: 10.1534/genetics.166.2.693
- White, F. F., and Yang, B. (2009). Host and pathogen factors controlling the rice *Xanthomonas oryzae* interaction. *Plant Physiol.* 150, 1677–1686. doi: 10.1104/pp.109.139360
- Williams, B., Verchot, J., and Dickman, M. B. (2014). When supply does not meet demand-ER stress and plant programmed cell death. *Front. Plant Sci.* 5:211. doi: 10.3389/fpls.2014.00211
- Wu, L., Goh, M. L., Sreekala, C., and Yin, Z. (2008). XA27 depends on an amino-terminal signal-anchor-like sequence to localize to the apoplast for resistance to *Xanthomonas oryzae* pv. *oryzae*. *Plant Physiol.* 148, 1497–1509. doi: 10.1104/pp.108.123356
- Yang, Y., and Gabriel, D. W. (1995). Intragenic recombination of a single plant pathogen gene provides a mechanism for the evolution of new host specificities. *J. Bacteriol.* 177, 4963–4968.
- Yang, B., Sugio, A., and White, F. F. (2005). Avoidance of host recognition by alterations in the repetitive and C-terminal regions of AvrXa7, a type III effector of *Xanthomonas oryzae* pv. *oryzae*. *Mol. Plant Microbe Interact.* 18, 142–149. doi: 10.1094/MPMI.18-0142
- Yang, B., Sugio, A., and White, F. F. (2006). Os8N3 is a host disease-susceptibility gene for bacterial blight of rice. *Proc. Natl. Acad. Sci. U.S.A.* 103, 10503–10508. doi: 10.1073/pnas.0604088103
- Yang, B., and White, F. F. (2004). Diverse members of the AvrBs3/PthA family of type III effectors are major virulence determinants in bacterial blight disease of rice. *Mol. Plant Microbe Interact.* 17, 1192–1200. doi: 10.1094/MPMI.2004.17.11.1192
- Yang, B., Zhu, W. G., Johnson, L. B., and White, F. F. (2000). The virulence factor AvrXa7 of *Xanthomonas oryzae* pv. *oryzae* is a type III secretion pathway-dependent nuclear-localized double-stranded DNA-binding protein. *Proc. Natl. Acad. Sci. U.S.A.* 97, 9807–9812. doi: 10.1073/pnas.170286897
- Yoshimura, S., Yamanouchi, U., Katayose, Y., Toki, S., Wang, Z. X., Kono, I., et al. (1998). Expression of Xa1, a bacterial blight-resistance gene in rice, is

- induced by bacterial inoculation. *Proc. Natl. Acad. Sci. U.S.A.* 95, 1663. doi: 10.1073/pnas.95.4.1663
- Yu, Y., Streubel, J., Balzergue, S., Champion, A., Boch, J., Koebnik, R., et al. (2011). Colonization of rice leaf blades by an African strain of *Xanthomonas oryzae* pv. *oryzae* depends on a new TAL effector that induces the rice nodulin-3 *Os11N3* gene. *Molecular Plant Microbe Interact.* 24, 1102–1113. doi: 10.1094/MPMI-11-10-0254
- Zeng, X., Tian, D., Gu, K., Zhou, Z., Yang, X., Luo, Y., et al. (2015). Genetic engineering of the Xa10 promoter for broad-spectrum and durable resistance to *Xanthomonas oryzae* pv. *oryzae*. *Plant Biotechnol. J.* doi: 10.1111/pbi.12342 [Epub ahead of print].
- Zhang, H., and Wang, S. (2013). Rice versus *Xanthomonas oryzae* pv. *oryzae*: a unique pathosystem. *Curr. Opin. Plant Biol.* 16, 188–195. doi: 10.1016/j.pbi.2013.02.008
- Zhao, Y. (2014). Auxin biosynthesis. *Arabidopsis Book* 12, e0173–e0173. doi: 10.1199/tab.0173
- Zhou, J., Peng, Z., Long, J., Soso, D., Liu, B., Eom, J. S., et al. (2015). Gene targeting by the TAL effector PthXo2 reveals cryptic resistance gene for bacterial blight of rice. *Plant J.* 82, 632–643. doi: 10.1111/tpj.12838
- Zhu, W., Yang, B., Wills, N., Johnson, L. B., and White, F. F. (1999). The C terminus of AvrXa10 can be replaced by the transcriptional activation domain of VP16 from the herpes simplex virus. *Plant Cell* 11, 1665–1674. doi: 10.1105/tpc.11.9.1665
- Zhu, W., Yang, B., Chittoor, J. M., Johnson, L. B., and White, F. F. (1998). AvrXa10 contains an acidic transcriptional activation domain in the functionally conserved C terminus. *Mol. Plant Microbe Interact.* 11, 824–832. doi: 10.1094/MPMI.1998.11.8.824

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 Zhang, Yin and White. 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.

Decision tools for bacterial blight resistance gene deployment in rice-based agricultural ecosystems

Gerbert S. Dossa^{1,2}, Adam Sparks³, Casiana Vera Cruz¹ and Ricardo Oliva^{1*}

¹ Plant Breeding, Genetics, and Biotechnology Division, International Rice Research Institute, Metro Manila, Philippines, ² Department of Phytomedicine, Leibniz Universität Hannover, Hannover, Germany, ³ Crop and Environmental Sciences Division, International Rice Research Institute, Metro Manila, Philippines

OPEN ACCESS

Edited by:

Laurent D. Noël,
Centre National de la Recherche
Scientifique, France

Reviewed by:

Tina B. Jordan,
Eberhard Karls University Tübingen,
Germany

Wei Qian,
Chinese Academy of Sciences,
China

*Correspondence:

Ricardo Oliva,
Plant Breeding, Genetics, and
Biotechnology Division, International
Rice Research Institute,
DAPO Box 7777,
Metro Manila 1301, Philippines
r.oliva@irri.org

Specialty section:

This article was submitted to
Plant-Microbe Interaction,
a section of the journal
Frontiers in Plant Science

Received: 26 February 2015

Accepted: 16 April 2015

Published: 05 May 2015

Citation:

Dossa GS, Sparks A, Vera Cruz C and
Oliva R (2015) Decision tools for
bacterial blight resistance gene
deployment in rice-based
agricultural ecosystems.
Front. Plant Sci. 6:305.
doi: 10.3389/fpls.2015.00305

Attempting to achieve long-lasting and stable resistance using uniformly deployed rice varieties is not a sustainable approach. The real situation appears to be much more complex and dynamic, one in which pathogens quickly adapt to resistant varieties. To prevent disease epidemics, deployment should be customized and this decision will require interdisciplinary actions. This perspective article aims to highlight the current progress on disease resistance deployment to control bacterial blight in rice. Although the model system rice-*Xanthomonas oryzae* pv. *oryzae* has distinctive features that underpin the need for a case-by-case analysis, strategies to integrate those elements into a unique decision tool could be easily extended to other crops.

Keywords: customized deployment, forward breeding, TAL effectors, R-genes, genome editing

Why Customize the Deployment of Bacterial Blight Resistance Genes?

Very often, elite rice varieties carrying effective resistance genes are distributed across broad geographic areas to maximize their socioeconomic impact. Eventually, the resistance gene will overlap with virulent pathogen populations that persist in low frequency. Prolonged exposure will increase selected population and leads to an outbreak. In this case, deployment is perhaps the most influential event compromising durability. Preventing disease epidemics requires a deeper understanding of the biological systems and interdisciplinary approaches to interconnect the factors that account for conducive environments, locally effective genes, and pathogen dynamics. A systematic monitoring of the pathogen population, which incorporates current understanding of effector biology, emerges as a key aspect to drive pathogen-informed deployment. However, it is essential that such information is readily transferred to breeding pipelines to guarantee the right variety profiles. We also believe that geographic information systems can be used to couple disease forecasting models with in-field surveys and other on-the-ground work to map epidemics in real time and therefore be integrated into a unique decision tool.

Bacterial blight (BB), caused by *Xanthomonas oryzae* pv. *oryzae* (*Xoo*), is the most important bacterial disease of rice. At least 39 resistance genes (*Xa*) have been identified from wild and cultivated accessions (Khan et al., 2014; Zhang et al., 2014a); among which *Xa4*, *xa5*, *xa13*, and *Xa21* appear to be widely used in breeding programs across Asia (Khan et al., 2014). Disease resistance deployment to control *Xoo* emerges as a perfect case for analysis because (i) the disease is widely distributed across rice-growing regions worldwide, (ii) resistance has a strong race-specific component, and (iii) many *Xa* genes have been incorporated into released varieties. A game-changing feature of the pathosystem is that *Xoo* uses transcription activator-like (TAL) effectors to promote

colonization and ensure nutrient uptake. In contrast to other rice pathosystems, *Xa* genes can be classified into distinct functional categories and only a small number appears to encode for NBS-LRR proteins (Boch et al., 2014). In this paper, we describe the key elements that need to be considered if we are to implement a strategy to customize the deployment of *Xa* genes in rice agroecosystems.

Breeding Fast, Breeding Precisely

Conventional breeding in rice uses pedigree breeding and selection, which employs a forward breeding approach based on traits of interest. The selection of advanced pedigree lines and recombinant inbred lines requires a long process that can take 8–9 years to generate elite lines for varietal release. With the recent advances in technologies, breeding for target traits can be fast-tracked by the application of marker-assisted selection (MAS) focusing on improving tolerance to abiotic stresses—drought, submergence, salinity, and soil problems—for unfavorable environments, along with increased yield, biotic stress resistance, and improved grain quality, traits that are also required for a favorable environment. Recently, the duration for developing improved varieties through forward breeding was decreased to 5–6 years through implementing rapid generation advance and MAS techniques. Further enhancement uses marker-assisted backcrossing (MABC) retaining the good traits of the recipient parents, combined with precision in incorporating specific traits of interest into high-yielding mega-varieties with superior grain quality, thus further reducing the time of producing elite lines to 3–4 years. Recently, a further transformation in breeding has been revolutionizing precision breeding while doubling the rate of genetic gains. The availability of 3,000 sequenced rice genomes provides an unprecedented wealth of data to mine alleles for novel genes (Li et al., 2014). This is coupled with advances in high-throughput SNP genotyping across large breeding populations in various platforms (Fluidigm's Dynamic ArraysTM, Douglas Scientific Array TapeTM, and LGC's automated systems for running KASPTM markers) to accelerate rice improvement (Thomson, 2014). Likewise, genotyping by sequencing (GBS) is currently becoming a choice for low-cost high-density genome-wide scans using multiplexed sequencing.

At IRRI, in-house genotyping services for various breeding programs contribute to fast-tracking the breeding cycle from hybridization to population advancement in 2–3 years to generate elite lines. Structurally, IRRI breeding hubs facilitate multi-environment testing with strategically selected locations in South Asia (India), Southeast Asia (Myanmar and the Philippines), and East and Southern Africa (Burundi), in partnership with other partners from both the public and private sector. Through this transformed breeding process, we envision generating various combinations of effective *Xa* genes in similar or different elite backgrounds for market segmentation. The strategy will likely promote shuffling of resistance mechanisms displayed on the field to prevent rapid pathogen adaptation to single-gene virulence, deployed in a customized manner (Figure 1). Real-time deployment of resistant varieties with various combinations of effective *Xa* genes can be customized through gene rotation or mixture

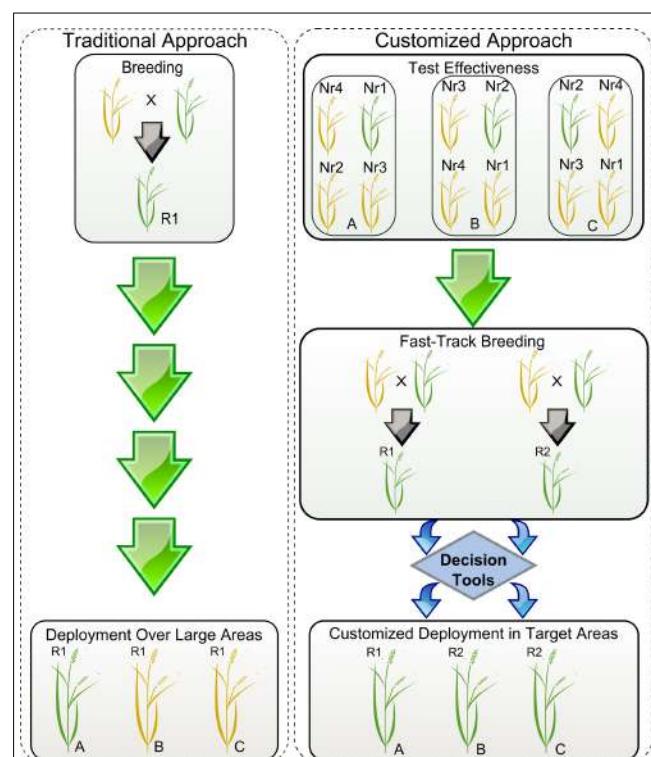


FIGURE 1 | Simplified scheme showing the deployment of resistant varieties using traditional versus customized approaches. R1 and R2 represent resistant elite varieties carrying hypothetical genes 1 and 2. Yellow and green plants represent susceptible and resistant phenotypes, respectively. Locations A, B, and C represent cropping regions that do not share boundaries. Nr1, Nr2, Nr3, and Nr4 are near-isogenic lines (NILs) for each of the available resistance genes 1, 2, 3, and 4. During traditional deployment, variety R1 is bred and released in large areas but is effective only in particular locations. During customized deployment, the effectiveness of the resistance genes and pathogen population structures are monitored using disease hotspots, seasonal collections, and pathogenicity tests done in a confined setting. Using a decision tool, breeding programs can rapidly customize the elite varieties to be deployed in targeted locations based on variety profiles.

in a single genetic background targeted for evolving or dynamic pathogen populations, especially in a BB-endemic rice-growing environment. In a broader context, breeding programs are allowed to prioritize genes that are effective across multiple locations, but also genes that combine different mechanisms and show low turnover rate.

Monitoring of Effective Genes

In order to customize deployment, we need to assess the effectiveness of R-genes in specific target areas and understand the evolutionary potential of the local *Xoo* population. Seasonal monitoring, when timely executed and concerted in a cost-effective manner, could be very useful for breeders to direct breeding efforts (Figure 2). The first aspect of monitoring entails a collection of field strains to understand population genetic structure and local distribution. So far, all markers based on restriction fragment length polymorphism profiling or repeated sequences (Choi et al., 1998; Chen et al., 2012; Mishra et al., 2013) have not succeeded in describing functional groups or simply are not

currently available in high-throughput platforms. However, many large-scale shotgun sequencing projects involving short-reads are underway and would potentially allow us to identify a set of SNP markers for standard use in field genotyping across different regions. Ultimately, the hope is that current advances in long-reads sequencing chemistry will resolve TAL effector sequences within complex samples, thus opening the door for more informative monitoring initiatives. The second aspect may involve the use of near-isogenic lines (NILs) carrying updated sets of *Xa* genes. Traditionally, this material has been used as a tool to identify phenotypic groups under controlled conditions (Ogawa et al., 1988). NILs are also suitable for deployment in disease-endemic areas or “hotspots” to capture low-prevalence genotypes that might escape from seasonal collection or to assess the effectiveness of *Xa* genes at a local scale (Figure 1). Ideally, both aspects could provide regional breeders with real-time decision support for small-scale interventions. Since 2013, IRRI has been deploying NILs in target areas of Asia and Africa and coordinating efforts to collect *Xoo* samples with local partners. Lessons learned from engaging rice research programs suggest that monitoring must be a participatory exercise and information exchange between national partners is more likely to occur under a common platform.

Exploiting Effectors to Drive Deployment

Scientists defined effectors as molecular instruments that facilitate a parasitic life style (Hogenhout et al., 2009). Translational research done in a number of crops has tailored effector biology into a useful tool for disease resistance breeding (Gawehns et al., 2013; Vleeshouwers and Oliver, 2014). In oomycetes and fungal pathosystems, effectors have been used to facilitate cloning R-genes, discovering novel specificities, or avoiding unnecessary breeding effort (Vleeshouwers et al., 2011; Oliver et al., 2012; Rietman et al., 2012). Recent advances in the biochemical function of bacteria TAL effectors suggest that their activity has a major impact on virulence (Kay and Bonas, 2009). TAL effectors activate the expression of specific host susceptibility genes (*S*) in order to create a favorable environment. For instance, increasing sucrose availability or reducing copper-mediated toxicity within the xylem vessels appears to be a clear output of its virulence function during rice-*Xoo* interaction (Chen et al., 2010; Yuan et al., 2010). With the TAL-DNA recognition decoded (Boch et al., 2009; Moscou and Bogdanove, 2009), it is now possible to have a catalog of validated TAL targets in the rice genome (Noël et al., 2013). The increasing evidence that TAL repertoires can determine host specificity in a gene-for-gene fashion (Yang and White, 2004; Gu et al., 2005; Yang et al., 2006; Antony et al., 2010; Tian et al., 2014; Wang et al., 2014) highlights their potential use in translational research. For instance, *Xoo* strains carrying TAL effectors *Avrxa10*, *AvrXa23*, and *AvrXa27* are unable to colonize rice accession containing *Xa10*, *Xa23* and *Xa27*, respectively (Gu et al., 2009; Tian et al., 2014; Wang et al., 2014).

Theoretically, the allelic diversity of TAL effectors in a region can be used to help deployment interventions, but in practice we need to overcome some technical issues: (i) What is the best way to efficiently capture TAL repertoires (TALome)? Probably,

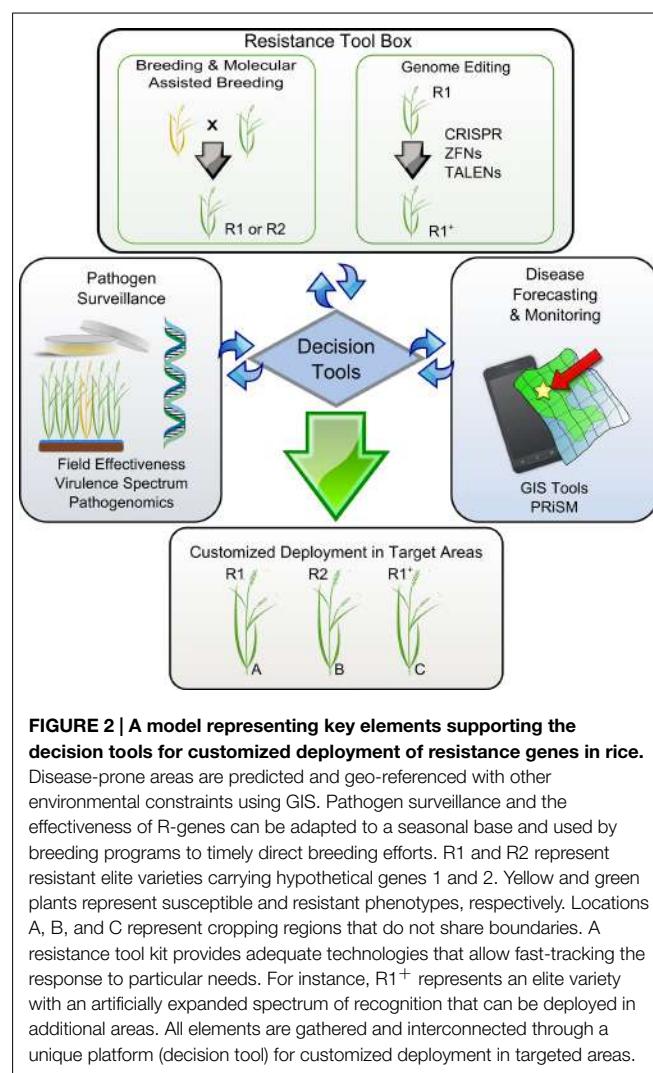


FIGURE 2 | A model representing key elements supporting the decision tools for customized deployment of resistance genes in rice.

Disease-prone areas are predicted and geo-referenced with other environmental constraints using GIS. Pathogen surveillance and the effectiveness of R-genes can be adapted to a seasonal base and used by breeding programs to timely direct breeding efforts. R1 and R2 represent resistant elite varieties carrying hypothetical genes 1 and 2. Yellow and green plants represent susceptible and resistant phenotypes, respectively. Locations A, B, and C represent cropping regions that do not share boundaries. A resistance tool kit provides adequate technologies that allow fast-tracking the response to particular needs. For instance, R1⁺ represents an elite variety with an artificially expanded spectrum of recognition that can be deployed in additional areas. All elements are gathered and interconnected through a unique platform (decision tool) for customized deployment in targeted areas.

a combination of TAL enrichment methods and high-throughput sequencing using long-reads will be enough to map the TALome, although cost-efficiency remains a major limitation. (ii) How do we improve current algorithms to efficiently catalog TAL targets? While many candidate *S*-genes have been validated experimentally, we expect an enhanced accuracy of prediction tools (Doyle et al., 2012; Grau et al., 2013; Pérez-Quintero et al., 2013) as more *Xoo* genomes become available and the algorithms can be trained on natural alleles of candidate targets using enlarged rice data sets (Li et al., 2014). (iii) Is it worthwhile assessing epiallelic variation of TAL effectors among field isolates? So far, no evidence suggests that *Xoo* actively uses this pathway to adapt to the selection pressure imposed by agricultural deployment. For now, it does not seem likely that a diagnostic test will be included in regular surveys because detecting gene expression in field samples can be technically challenging and economically not feasible (Gijzen et al., 2014). In summary, we cannot afford to exclude effectors-based information from a modern rice breeding and deployment program, but addressing these questions is essential for fine tuning the overall strategy. Beside effectors biology, the study of cell to cell signaling pathways or virulence

regulation in response to abiotic factors are some of the emerging areas of research that will need some attention in the near future.

Mapping Disease in Real Time

The technologies currently exist to merge all of these concepts into a working system for targeted deployment. For example, the Philippine Rice Information System (PRISM¹) gathers data from farmers' fields throughout the Philippines. Trained individuals use a standardized survey portfolio based on the IRRI publication "A Survey Portfolio to Characterize Yield-Reducing Factors in Rice" (Savary and Castilla, 2009), collecting data on yield-reducing (biotic stresses) and also yield-limiting (abiotic stresses) factors, current yields, and farmers' agronomic practices using smartphones. The data are submitted via a wireless connection (Wi-Fi or cellular) to a centralized database in near real time. These data allow us to create a profile of a local targeted environment for breeders to reference by linking with other sources of information. The profile could include the most common pathogen population structure (collected in the fields being surveyed), farmers' market preferences, farmers' agronomic practices, and biotic and abiotic stresses in the target area. For example, using this technology, breeders working to develop a variety for an area where flooding is common could see where BB and flooding are most likely to occur together. Then, these data could be coupled with the local *Xoo* population structure and local market preferences to develop a variety that farmers would be likely to adopt and that is tailored to that specific environment's stresses (Figure 2).

Broadening the Spectrum of Recognition, Only if Necessary

Genome editing techniques enable targeting specific DNA sequences and introducing a broad range of precise genetic modifications (Fichtner et al., 2014). Beside the current direction of the regulatory debate, the outcome of this technology is not a GMO product because it does not contain any foreign DNA (Waltz, 2012). While most of the available genome editing tools (ZFNs, TALENs, CRISPR-Cas9) have been successfully tested in rice (Li et al., 2012; Miao et al., 2013; Zhang et al., 2014b; Zhou et al., 2014), current progress on virulence mechanisms promoted by TAL effectors has inspired new ways to immunize crops using such an approach. For instance, resistance can be acquired by disrupting the TAL effector binding site of major S-gene promoters, such as members of the SWEET sucrose-efflux transporter family (Li et al., 2012). Eventually, broad-spectrum resistance can be created if several family members are targeted at the same time, thus limiting the access of the bacteria to alternative nutrient resources. Resistance can also be engineered using multiple decoy TAL-binding sites fused upstream of a single executor R-gene (Römer et al., 2009). Among all the rice R-genes that have been reported (Gu et al., 2005; Tian et al., 2014; Wang et al., 2014), few have

potential applications as executors because they trigger strong localized cell-death and are induced only in the presence of the pathogen. We predict that genome editing tools will be integrated into the next-generation resistance tool kit, but might be considered only when no other alternative is available (Figure 2). For instance, elite varieties with an artificially expanded spectrum of recognition (either *R* or *S*) may become a solution in regions where management practices or pyramiding of existing *Xa* genes are no longer options. Swarna-Sub1 is a high-yielding mega-variety that shows a yield advantage in flood-prone areas of Asia (Ismail et al., 2013) but it is quite susceptible to *Xoo* infection in some of these unfavorable environments. Current attempts to precisely fast-track effective combinations of *Xa* genes into Swarna-Sub1 are under way, but the number of effective *Xa* genes available is limited and alternative strategies to increase the diversity of mechanisms are key for sustainable deployment-based management (Figure 2). Whether we are planning to exploit *R*- or *S*-genes to broaden the spectrum of resistance, it is clear that genome editing tools will be important assets for next-generation resistance breeding.

One Tool to Rule Them All

It is only a matter of time before information and communication technologies (ICTs) lead the research revolution on the agricultural landscape. Currently, very precise information can be retrieved and/or delivered to and from farmers' hands in real time. Platforms that incorporate crop health status under well-characterized environments are coming and will soon become tools for informed interventions. Therefore, it becomes very important that pathologists, breeders, and epidemiologists endeavor to integrate diagnostics, disease models, and breeding efforts into a unique platform for customized deployment (Figure 2). This vision does not exclude other fields of research that also contribute to increased variety adoption and are important in the rice value chain. These are exciting times, like never before, rice scientists have the possibility to adapt their breeding programs and decide which variety will be promoted next season to reduce the chance of future epidemics.

Acknowledgments

Scientists at the International Rice Research Institute (IRRI) are partially funded by the Global Rice Science Partnership (GRiSP), and projects under the Stress-tolerant rice for poor farmers in Africa and South Asia (STRASA) supported by the Bill & Melinda Gates Foundation (BMGF). The Philippine Rice Information System (PRISM) is a collaborative project between Department of Agriculture (DA)-Philippine Rice Research Institute (PhilRice), DA-Bureau of Plant Industry (BPI), DA-Regional Field Offices (RFOs), and IRRI, in support to DA's Food Staples Sufficiency Program, with funding support from the DA-National Rice Program through the DA-Bureau of Agricultural Research. We want to thank Hei Leung and Michael Thomson for reviewing the manuscript.

¹<http://philippinericeinfo.ph>

References

- Antony, G., Zhou, J., Huang, S., Li, T., Liu, B., White, F. E., et al. (2010). Rice xa13 recessive resistance to bacterial blight is defeated by induction of the disease susceptibility gene Os-11N3. *Plant Cell* 22, 3864–3876. doi: 10.1105/tpc.110.078964
- Boch, J., Bonas, U., and Lahaye, T. (2014). TAL effectors-pathogen strategies and plant resistance engineering. *New Phytol.* 204, 823–832. doi: 10.1111/nph.13015
- Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., et al. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 326, 1509–1512. doi: 10.1126/science.1178811
- Chen, L. Q., Hou, B. H., Lalonde, S., Takanaga, H., Hartung, M. L., Qu, X. Q., et al. (2010). Sugar transporters for intercellular exchange and nutrition of pathogens. *Nature* 468, 527–532. doi: 10.1038/nature09606
- Chen, X. L., Yu, L., Gao, L. L., Jiang, T., Li, Q. Y., and Huang, Q. (2012). Elevational variation in diversity of *Xanthomonas oryzae* pv. *oryzae* in South-West China. *J. Phytopathol.* 160, 261–268. doi: 10.1111/j.1439-0434.2012.01892.x
- Choi, S. H., Vera Cruz, C. M., and Leach, J. E. (1998). Distribution of *Xanthomonas oryzae* pv. *oryzae* DNA modification systems in Asia. *Appl. Environ. Microbiol.* 64, 1663–1668.
- Doyle, E. L., Booher, N. J., Standage, D. S., Voytas, D. F., Brendel, V. P., Van Dyk, J. K., et al. (2012). TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. *Nucleic Acids Res.* 40, 117–122. doi: 10.1093/nar/gks608
- Fichtner, F., Castellanos, R. U., and Ülker, B. (2014). Precision genetic modifications: a new era in molecular biology and crop improvement. *Planta* 239, 921–939. doi: 10.1007/s00425-014-2029-y
- Gawehns, F., Cornelissen, B. J. C., and Takken, F. L. W. (2013). The potential of effector-target genes in breeding for plant innate immunity. *Microb. Biotechnol.* 6, 223–229. doi: 10.1111/1751-7915.12023
- Gijzen, M., Ishmael, C., and Shrestha, S. D. (2014). Epigenetic control of effectors in plant pathogens. *Front. Plant Sci.* 5:638. doi: 10.3389/fpls.2014.00638
- Grau, J., Wolf, A., Reschke, M., Bonas, U., Posch, S., and Boch, J. (2013). Computational predictions provide insights into the biology of TAL effector target sites. *PLoS Comput. Biol.* 9:e1002962. doi: 10.1371/journal.pcbi.1002962
- Gu, K. Y., Yang, B., Tian, D. S., Wu, L. F., Wang, D. J., Sreekala, C., et al. (2005). R gene expression induced by a type-III effector triggers disease resistance in rice. *Nature* 435, 1122–1125. doi: 10.1038/nature03630
- Gu, K., Tian, D., Qiu, C., and Yin, Z. (2009). Transcription activator-like type III effector AvrXa27 depends on OsTFIIAgamma5 for the activation of Xa27 transcription in rice that triggers disease resistance to *Xanthomonas oryzae* pv. *oryzae*. *Mol. Plant Pathol.* 10, 829–835. doi: 10.1111/j.1364-3703.2009.00567.x
- Hogenhout, S. A., van der Hoorn, R. A. L., Terauchi, R., and Kamoun, S. (2009). Emerging concepts in effector biology of plant-associated organisms. *Mol. Plant Microbe Interact.* 22, 115–122. doi: 10.1094/MPMI-22-2-0115
- Ismail, A. M., Singh, U. S., Singh, S., Dar, M. H., and Mackill, D. J. (2013). The contribution of submergence tolerant (Sub1) rice variety to food security in flood-prone rainfed areas in Asia. *Field Crops Res.* 152, 83–93. doi: 10.1016/j.fcr.2013.01.007
- Kay, S., and Bonas, U. (2009). How *Xanthomonas* type III effectors manipulate the host plant. *Curr. Opin. Microbiol.* 12, 37–43. doi: 10.1016/j.mib.2008.12.006
- Khan, M. A., Naeem, M., and Iqbal, M. (2014). Breeding approaches for bacterial leaf blight resistance in rice (*Oryza sativa* L.), current status and future directions. *Eur. J. Plant Pathol.* 139, 27–37. doi: 10.1007/s10658-014-0377-x
- Li, J. Y., Wang, J., and Zeigler, R. S. (2014). The 3,000 rice genomes project: new opportunities and challenges for future rice research. *GigaScience* 3:8. doi: 10.1186/2047-217X-3-8
- Li, T., Liu, B., Spalding, M. H., Weeks, D. P., and Yang, B. (2012). High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nat. Biotechnol.* 30, 390–392. doi: 10.1038/nbt.2199
- Miao, J., Guo, D., Zhang, J., Huang, Q., Qin, G., Zhang, X., et al. (2013). Targeted mutagenesis in rice using CRISPR-Cas system. *Cell Res.* 23, 1233–1236. doi: 10.1038/cr.2013.123
- Mishra, D., Vishnupriya, M. R., Anil, M. G., Konda, K., Raj, Y., and Sonti, R. V. (2013). Pathotype and genetic diversity amongst Indian isolates of *Xanthomonas oryzae* pv. *oryzae*. *PLoS ONE* 8:e81996. doi: 10.1371/journal.pone.0081996
- Moscou, M. J., and Bogdanove, A. J. (2009). A simple cipher governs DNA recognition by TAL effectors. *Science* 326, 1501. doi: 10.1126/science.1178817
- Noël, L. D., Denancé, N., and Szurek, B. (2013). Predicting promoters targeted by TAL effectors in plant genomes: from dream to reality. *Front. Plant Sci.* 4:333. doi: 10.3389/fpls.2013.00333
- Ogawa, T., Yamamoto, T., Khush, G. S., Mew, T. W., and Kaku, H. (1988). Near-isogenic lines as international differentials for resistance to bacterial blight of rice. *RGN* 5, 106–107.
- Oliver, R. P., Friesen, T. L., Faris, J. D., and Solomon, P. S. (2012). *Stagonospora nodorum*: from pathology to genomics and host resistance. *Annu. Rev. Phytopathol.* 50, 23–43. doi: 10.1146/annurev-phyto-081211-173019
- Pérez-Quintero, A., Rodriguez, R. L., Dereeper, A., Lopez, C., Koebnik, R., Szurek, B., et al. (2013). An improved method for TAL effectors DNA-binding sites prediction reveals functional convergence in TAL repertoires of *Xanthomonas oryzae* strains. *PLoS ONE* 8:e68464. doi: 10.1371/journal.pone.0068464
- Rietman, H., Bijsterbosch, G., Cano, L. M., Lee, H. R., Vossen, J. H., Jacobsen, E., et al. (2012). Qualitative and quantitative late blight resistance in the potato cultivar Sarpo Mira is determined by the perception of five distinct RXLR effectors. *Mol. Plant Microbe Interact.* 25, 910–919. doi: 10.1094/MPMI-01-12-0010-R
- Römer, P., Recht, S., and Lahaye, T. (2009). A single plant resistance gene promoter engineered to recognize multiple TAL effectors from disparate pathogens. *Proc. Natl. Acad. Sci. U.S.A.* 106, 20526–20531. doi: 10.1073/pnas.0908812106
- Savary, S., and Castilla, N. P. (2009). A survey portfolio for the characterization of rice pest constraints. *IRRI Discuss. Paper* 18, 32.
- Thomson, J. M. (2014). High-throughput SNP genotyping to accelerate crop improvement. *Plant Breed. Biotechnol.* 2, 195–212. doi: 10.9787/PBB.2014.2.3.195
- Tian, D., Wang, J., Zeng, X., Gu, K., Qiu, C., Yang, X., et al. (2014). The rice TAL effector-dependent resistance protein Xa10 triggers cell death and calcium depletion in the endoplasmic reticulum. *Plant Cell* 26, 497–515. doi: 10.1105/tpc.113.119255
- Vleeshouwers, V. G., and Oliver, R. P. (2014). Effectors as tools in disease resistance breeding against biotrophic, hemibiotrophic, and necrotrophic plant pathogens. *Mol. Plant Microbe Interact.* 27, 196–206. doi: 10.1094/MPMI-10-13-0313-IA
- Vleeshouwers, V. G., Raffaele, S., Vossen, J., Champouret, N., Oliva, R., Segretni, M. E., et al. (2011). Understanding and exploiting late blight resistance in the age of effectors. *Annu. Rev. Phytopathol.* 49, 507–531. doi: 10.1146/annurev-phyto-072910-095326
- Waltz, E. (2012). Tiptoeing around transgenics. *Nat. Biotechnol.* 30, 215–217. doi: 10.1038/nbt.2143
- Wang, C., Zhang, X., Fan, Y., Gao, Y., Zhu, Q., Zheng, C., et al. (2014). Xa23 is an executor R protein and confers broad-spectrum disease resistance in rice. *Mol. Plant* 8, 290–302. doi: 10.1016/j.molp.2014.10.010
- Yang, B., Sugio, A., and White, F. F. (2006). Os8N3 is a host disease susceptibility gene for bacterial blight of rice. *Proc. Natl. Acad. Sci. U.S.A.* 103, 10503–10508. doi: 10.1073/pnas.0604088103
- Yang, B., and White, F. F. (2004). Diverse members of the AvrBs3/ PthA family of type III effectors are major virulence determinants in bacterial blight disease of rice. *Mol. Plant Microbe Interact.* 17, 1192–1200. doi: 10.1094/MPMI.2004.17.11.1192
- Yuan, M., Chu, Z., Li, X., Xu, C., and Wang, S. (2010). The bacterial pathogen *Xanthomonas oryzae* overcomes rice defenses by regulating host copper redistribution. *Plant Cell* 22, 3164–3176. doi: 10.1105/tpc.110.078022
- Zhang, F., Zhuo, D. L., Zhang, F., Huang, L. Y., Wang, W. S., Xu, J. L., et al. (2014a). Xa39, a novel dominant gene conferring broad-spectrum resistance to *Xanthomonas oryzae* pv. *oryzae* in rice. *Plant Pathol.* doi: 10.1111/ppa.12283 [Epub ahead of print].
- Zhang, H., Zhang, J., Wei, P., Zhang, B., Gou, F., Feng, Z., et al. (2014b). The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. *Plant Biotechnol. J.* 12, 797–807. doi: 10.1111/pbi.12200
- Zhou, H., Liu, B., Weeks, D. P., Spalding, M. H., and Yang, B. (2014). Large chromosomal deletions and heritable small genetic changes induced by CRISPR/Cas9 in rice. *Nucleic Acids Res.* 42, 10903–10914. doi: 10.1093/nar/gku806
- 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 Dossa, Sparks, Vera Cruz and Oliva. 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.**

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read,
for greatest visibility



COLLABORATIVE PEER-REVIEW

Designed to be rigorous
– yet also collaborative,
fair and constructive



FAST PUBLICATION

Average 85 days from
submission to publication
(across all journals)



COPYRIGHT TO AUTHORS

No limit to article
distribution and re-use



TRANSPARENT

Editors and reviewers
acknowledged by name
on published articles



SUPPORT

By our Swiss-based
editorial team



IMPACT METRICS

Advanced metrics
track your article's impact



GLOBAL SPREAD

5'100'000+ monthly
article views
and downloads



LOOP RESEARCH NETWORK

Our network
increases readership
for your article

Frontiers

EPFL Innovation Park, Building I • 1015 Lausanne • Switzerland
Tel +41 21 510 17 00 • Fax +41 21 510 17 01 • info@frontiersin.org
www.frontiersin.org

Find us on

