

Identifying possibly novel sources of antimicrobial resistance in uncultivated bacterial and archeal metagenome-assembled genomes

This manuscript ([permalink](#)) was automatically generated from [jackiepurdue/paper-2019-amr-uba@c9599f5](#) on July 24, 2019.

Authors

- **Jackie Purdue**

 [0000-0003-3131-0688](#) ·  [jackiepurdue](#) ·  [jackiepurdue](#)

Department of Computer Science, Dalhousie University · Funded by Grant XXXXXXXX

- **Jane Roe**

 [XXXX-XXXX-XXXX-XXXX](#) ·  [janeroe](#)

Department of Something, University of Whatever; Department of Whatever, University of Something

Abstract

Antimicrobial resistance (AMR) is a global threat to human health. Efforts to prevent the spread of AMR rely on surveillance of possible AMR determinants. Under-sampled, and uncultured microbial genomes in the form of metagenome data, are a potential source for AMR gene discovery. The Resistance Gene Identifier (RGI) and the Comprehensive Antibiotic Resistance Database (CARD) [1](#) were used to predict resistomes in a data set of 7903 draft quality metagenome assembled genomes [2](#) (MAGs) from various uncultivated bacterial and archeal sources (UBAs). Genes associated with mobile colistin resistance (MCR), and carbapenem resistance, such as Metallo- β -Lactamases (NDM), *Klebsiella pneumoniae* carbapenemases (KPC), and OXA β -lactamases were found to be represented in the UBAs. [TODO: findings] [TODO: conclusion]

Introduction

An important part of the microbial adaptive process is an organisms ability to adapt to pressure from antibiotic agents in the environment. In the context of pathogenic microbes and pharmaceutical antibiotics, this process is a threat to our treatment of infectious diseases. In both healthcare, and agricultural settings, many antibiotic drugs are becoming less effective due to this antimicrobial resistance. AMR is becoming more prevalent, and is even starting to become an issue with last-line antibiotics such as colistins, and carbapenemases. This problem is arising due to irresponsible practices in healthcare, and in agriculture. Much of this problem is due to indiscriminate application of antibiotics, without regard for current knowledge on the matter. Another portion of this problem is due to a lack of insight into the constantly changing pathogenic resistomes, where bad practices could be carried out in ignorance. In an effort to control the spread of AMR, it is important that we collect as much information as possible about this process. Surveillance and characterization of many AMR genes is currently done through curated databases such as the Comprehensive Antibiotic Resistance Database (CARD). These databases contain plenty of AMR information about resistomes which are commonly studied, but it is possible that a lack of important information exists in the rest of the environment. Observing potential AMR genes for antibiotics such as carbapenemases and colistins, outside of a clinical setting, could lead to valuable new information.

A potential source of discovery of new AMR determinants is an examination of Under-sampled, and uncultured microbial genomes in the form of metagenome data. The Resistance Gene Identifier (RGI) uses the Comprehensive Antibiotic Resistance Database (CARD) [1](#) to predict potential AMR genes from sequence data, would be a useful tool in examining such data. Using canonical reference genes associated with colistin and carbapenem resistance, we can use current knowledge to query these metagenomes, and perhaps glean new knowledge from the phylogenetic relationships.

TODO: introduce MCR, OXA [3](#), KPC, and NDM

Methods

AMR Prediction

A dataset of 7903 draft quality MAGs which were recovered from the Sequence Read Archive by Parks [2](#). TODO: source of UBA sequence files RGI version 5.0 with CARD database version 3.02 [1](#) was run on the contigs of the 7903 MAGs with the inclusion of loose, perfect, and strict hits.

Phylogenetic Analysis

MCR

Obtaining candidatea and reference MCR genes

24 canonical sequence variants as indicated by CARD for the MCR phosphoethanolamine transferase [AMR Gene Family], MCR 3.11, 6.1, 2.2, 3.6, 3.10, 3.5, 1.10, 1.9, 3.9, 3.8, 3.7, 3.2, 3.1, 1.6, 1, 2, 3, 1.2, 1.3, 1.4, 8, 4, 5, 7.1 were obtained from the CARD database.

Output containing only AMR gene prediction data for the MCR phosphoethanolamine transferase gene family was produced by filtering RGI output based on a 'MCR phosphoethanolamine' string search. The filtered data were translated to a blast database. The 24 CARD canonical sequences were used to perform a multiple query BLASTP against the UBA blast database with a e-value threshold of 1e-180 and a query coverage of 90%. Redundant results were filtered from these BLASTP results by retrieving only the longest sequence for each uniquely labeled result.

Additional putative MCR phosphoethanolamine transferase gene family sequences were accumulated by querying the CARD prevalence data version 3.0.4 ("based on sequence data acquired from NCBI on 28-Feb-2019, analyzed using RGI 4.2.2 (DIAMOND homolog detection) and CARD 3.0.1"). The prevalence data was translated to a BLAST database. The 24 CARD canonical sequences were used to perform a multiple query BLASTP against the prevalence blast database with a e-value threshold of 1e-180 and a query coverage of 90%. Redundant results were filtered from these BLAST results by retrieving only the longest sequence for each uniquely labeled result.

The 24 CARD canonical sequences were used to perform a multiple query BLASTP against the NCBI non-redundant database blast database with a e-value threshold of 1e-180 and a query coverage of 90%. Redundant results were filtered from these BLAST results by retrieving only the longest sequence for each uniquely labeled result.

The filtered sequences from NCBI non-redundant data, CARD prevalence data, UBA data were all concatenated to one multi-FASTA format file with the canonical sequences. These the concatenated amino acid sequences were aligned with MAFFT FFT-NS-2. TODO: re-run MAFFT with best parameters and update this file. The aligned sequences were trimmed by trimal using the automated1 option. IQ-TREE was then used to build a bootstrapped tree with -bb 1000 with the LG+R10 model of substitution. all leaves of the trees were labeled with as much taxonomic information as possible for each rank, based on information from metadata.

The result was carried to subsequent analysis.

References

1. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database

Baofeng Jia, Amogelang R. Raphenya, Brian Alcock, Nicholas Waglechner, Peiyao Guo, Kara K. Tsang, Briony A. Lago, Biren M. Dave, Sheldon Pereira, Arjun N. Sharma, ... Andrew G. McArthur

Nucleic Acids Research (2016-10-26) <https://doi.org/f9wbjs>

DOI: [10.1093/nar/gkw1004](https://doi.org/10.1093/nar/gkw1004) · PMID: [27789705](https://pubmed.ncbi.nlm.nih.gov/27789705/) · PMCID: [PMC5210516](https://pubmed.ncbi.nlm.nih.gov/PMC5210516/)

2. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life

Donovan H. Parks, Christian Rinke, Maria Chuvpochina, Pierre-Alain Chaumeil, Ben J. Woodcroft, Paul N. Evans, Philip Hugenholtz, Gene W. Tyson

Nature Microbiology (2017-09-11) <https://doi.org/cczd>

DOI: [10.1038/s41564-017-0012-7](https://doi.org/10.1038/s41564-017-0012-7) · PMID: [28894102](https://pubmed.ncbi.nlm.nih.gov/28894102/)

3. OXA -Lactamases

B. A. Evans, S. G. B. Amyes

Clinical Microbiology Reviews (2014-04-01) <https://doi.org/f5zp8b>

DOI: [10.1128/cmr.00117-13](https://doi.org/10.1128/cmr.00117-13) · PMID: [24696435](https://pubmed.ncbi.nlm.nih.gov/24696435/) · PMCID: [PMC3993105](https://pubmed.ncbi.nlm.nih.gov/PMC3993105/)