Challenges and opportunities for whole-genome sequencing–based surveillance of antibiotic resistance

Prof. Willem van Schaik
Professor in Microbiology and Infection

Institute of Microbiology and Infection
University of Birmingham, United Kingdom
w.vanschaik@bham.ac.uk

Twitter: WvSchaik
Sequence-based pathogen surveillance

use sequence data to infer relatedness of strains
MLST (allelic profile of 5–7 housekeeping genes)

Low resolution

Differences may be apparent, but not clear

Provides a nomenclature to describe clones (e.g. *E. coli* ST131)
Strain A

MLST

WGS

High resolution

Core genome phylogeny
(SNP-based, core genome MLST)

Whole genome phylogeny
(whole genome MLST)

Strain B
Many tools

Which ones to choose?

Or use the tools developed by Torsten Seemann (Univ of Melbourne)

Whole-Genome Sequencing of Bacterial Pathogens: the Future of Nosocomial Outbreak Analysis

Scott Quainoo, Jordy P. M. Coolen, Sacha A. F. T. van Hijum, Martijn A. Huynen, Willem J. G. Melchers, Willem van Schaik, Heiman F. L. Wertheim

Clin Microbiol Rev 30:1015–1063
But there is more in a whole genome sequence...

Antibiotic resistance genes

Presence of mobile genetic elements
Vancomycin-resistant *Enterococcus faecium*

“Human commensal”

“Clinical isolates”

“Animal isolates”

Maximum likelihood tree: concatenated alignment of 1344 core genes in 73 *E. faecium* genomes

Direct mapping of reads: no loss of data

Assembly allows assessment of genomic context

Accommodates multiple bacteria in sample

Fast, validated for S. aureus and M. tuberculosis


- Fastest option
- [a] BWA implemented in SEAR
- [b] Bowtie2 implemented in SRST2
- [c] k-merResistence
- [d] Mykrobe predictor
- [e] BLAST (several tools, including Abricate)
What is an antibiotic resistance gene?

“a gene that confers resistance to an antibiotic in an otherwise susceptible microbial host”

Horizontally acquired resistance genes: e.g. \textit{blaKPC}, \textit{ermB}, \textit{vanA} associated with mobile genetic elements

Mutations in housekeeping genes that confer resistance e.g. target modification, porin mutation

Conserved genes that make a species intrinsically resistant e.g. efflux pumps
Antibiotic Resistance Gene Databases

ARDB: no updates, many intrinsic resistance genes: do not use

CARD: frequently updated, based on ARDB. Contains some intrinsic resistance genes; database for ‘resistance SNPs’ but can give false positives

ResFinder
ARG-ANNOT

Frequently updated. Acquired resistance genes only, small number of intrinsic resistance genes.

ResFams: prediction of resistance genes from metagenomes, many false positives or unvalidated genes
Antibiotic Resistance Gene Databases

Subject Section

ARGs-OAP v2.0 with an Expanded SARG Database and Hidden Markov Models for Enhancement of Antibiotic Resistance Gene Detection

DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomes

MEGARes: an antimicrobial resistance database for high-throughput sequencing

Steven M. Lakin, Chris Dean, Noelle R. Noyes, Adam Dettenwanger, Anne Spencer Ross, Enrique Doster, Pablo Rovira, Zaid Abdo, Kenneth L. Jones, Jaime Ruiz, Keith E. Belk, Paul S. Morley and Christina Boucher
When reporting antibiotic resistance genes....
Sequence-based diagnostics of antibiotic resistance genes?
Sequence-based diagnostics of antibiotic resistance genes?

Mycobacterium tuberculosis

Very slow growth

No horizontal gene transfer

Large collection of strains + resistance phenotypes

WGS accurately predicts susceptibility 93.6% – 99.0% to four key antibiotics

CRyPTIC Consortium and the 100,000 Genomes Project. New Engl J Med. 2018 379:1403
Sequence-based diagnostics of antibiotic resistance genes?

Common causes of bloodstream infections

*Escherichia coli*
*Staphylococcus aureus*
*Enterococcus*

Fast growth

Horizontal gene transfer

Emergence of new resistance determinants (e.g. *mcr*)

Fast WGS (e.g. of bloodstream cultures) for screening while susceptibility testing is ongoing
Prediction of Plasmids from Illumina Data

- 42 completely assembled genomes
- 12 genera
- 148 plasmids (1.5 – 339 kbp)

Sequence repeats in plasmid sequences complicates assembly

Illumina + PacBio sequences available

Prediction of Plasmids from Illumina Data

Comparison of completely automated approaches


or

cBar: detects differences in k-mer composition between plasmids and chromosomes Zhou & Xu Bioinformatics 2010 26:2051


PlasmidSPAdes: plasmids have different coverage (higher copy number) than chromosome Antipov et al. Bioinformatics 2016 32:3380

Prediction of Plasmids from Illumina Data

Number of plasmids correctly predicted (>90% of plasmid sequence)

Results are variable

Good performance PlasmidSPAdes, but all plasmids in single bin

Long reads can cover repeats
Summary

Whole genome sequencing is required for high-resolution typing