PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens

Zankari, Ea; Allesøe, Rosa Lundbye; Joensen, Katrine Grimstrup; Cavaco, Lina; Lund, Ole; Aarestrup, Frank Møller

Published in:
Journal of Antimicrobial Chemotherapy

Link to article, DOI:
10.1093/jac/dkx217

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens

Ea Zankari1,2, Rosa Allesøe2, Katrine G. Joensen3, Lina M. Cavaco1, Ole Lund2 and Frank M. Aarestrup1*

1National Food Institute, Technical University of Denmark, 2800 Kgs Lyngby, Denmark; 2Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, 2800 Kgs Lyngby, Denmark; 3Department of Microbiology and Infectious Diseases, Statens Serum Institut, Copenhagen, Denmark

*Corresponding author. Tel: +45-35886281; E-mail: fmaa@food.dtu.dk

Received 20 March 2017; returned 19 April 2017; revised 2 June 2017; accepted 7 June 2017

Background: Antibiotic resistance is a major health problem, as drugs that were once highly effective no longer cure bacterial infections. WGS has previously been shown to be an alternative method for detecting horizontally acquired antimicrobial resistance genes. However, suitable bioinformatics methods that can provide easily interpretable, accurate and fast results for antimicrobial resistance associated with chromosomal point mutations are still lacking.

Methods: Phenotypic antimicrobial susceptibility tests were performed on 150 isolates covering three different bacterial species: Salmonella enterica, Escherichia coli and Campylobacter jejuni. The web-server ResFinder-2.1 was used to identify acquired antimicrobial resistance genes and two methods, the novel PointFinder (using BLAST) and an in-house method (mapping of raw WGS reads), were used to identify chromosomal point mutations. Results were compared with phenotypic antimicrobial susceptibility testing results.

Results: A total of 685 different phenotypic tests associated with chromosomal resistance to quinolones, polymyxin, rifampicin, macrolides and tetracyclines resulted in 98.4% concordance. Eleven cases of disagreement between tested and predicted susceptibility were observed: two C. jejuni isolates with phenotypic fluoroquinolone resistance and two with phenotypic erythromycin resistance and five colistin-susceptible E. coli isolates with a detected pmrB V161G mutation when assembled with Velvet, but not when using SPAdes or when mapping the reads.

Conclusions: PointFinder proved, with high concordance between phenotypic and predicted antimicrobial susceptibility, to be a user-friendly web tool for detection of chromosomal point mutations associated with antimicrobial resistance.

Introduction

Horizontal gene transfer among bacterial isolates is often considered the main mediator of acquired antimicrobial resistance. However, mutational resistance is another important way to confer resistance.

It has previously been shown that WGS can be an alternative to phenotypic susceptibility testing of bacterial isolates for detection of horizontally acquired resistance.1,2 Databases for mapping to chromosomal mutations have also been developed for Mycobacterium tuberculosis.3 However, at present there is a lack of suitable bioinformatics methods to provide easily interpretable results for antimicrobial resistance associated with chromosomal point mutations for most bacterial species.

In this study, a novel web tool, PointFinder, was developed for detection of chromosomal point mutations associated with antimicrobial resistance, in bacterial WGS data. PointFinder may be run in parallel and become an extension to the already existing web server tool ResFinder,1 which detects horizontally acquired resistance genes in WGS data. The performance was compared with that of an in-house mapping method for detecting point mutations and both results were compared with phenotypic antimicrobial susceptibility tests in order to validate the possibilities of using these methods as alternatives to standard phenotypical testing.
Materials and methods

Chromosomal mutation database

Information regarding mutations in chromosomal genes associated with antimicrobial resistance was collected from published papers (Table 1). The reference sequences were selected from WT *Escherichia coli* strain K-12 (MG1655) for the *E. coli* database, *Salmonella* Typhimurium strain LT2 for the *Salmonella enterica* database and *Campylobacter jejuni* NCTC 11168 for the *C. jejuni* database.

Bacterial isolates and WGS data

In total, 150 isolates covering three species were included in the study: *E. coli* (n = 50) and *Salmonella* (n = 50) isolates from the in-house strain collection at the National Food Institute and *C. jejuni* (n = 50) isolates from the in-house strain collection at Statens Serum Institut. The isolates were selected on the basis of having both WGS data and phenotypes available. The *Salmonella* isolates included strains from 10 different serovars (Tables S1 to S3, available as Supplementary data at JAC Online). All bacterial isolates were sequenced using the Miseq platform (Illumina) to obtain paired-end sequences and assembled de novo using Velvet (reference software). Bacterial strains were screened for phenotypic resistance using MIC determinations interpreted according to EUCAST (www.eucast.org). Only the susceptibility tests relevant for antimicrobial resistance associated with chromosomal point mutations for each species were analysed (Table 2). As resistance to some of the antimicrobial agents can be caused by either acquired genes or chromosomal point mutations, ResFinder-2.1 (www.genomicepidemiology.org)31 was used to detect known acquired resistance genes in the WGS data, using a threshold of 98% identity (%ID) and 60% length (minimum percentage length of the resistance gene to be covered). All isolates with disagreement between the phenotypic and predicted susceptibility were re-tested.

PointFinder

PointFinder consists of two databases: a chromosomal gene database, with all reference sequences in fasta format; and a chromosomal mutation database containing information on codon positions and substitutions. PointFinder uses BLASTn for identifying the best match for each gene in the chromosomal gene database, and only hits with an identity of >80% are further analysed. The program goes through each alignment comparing each position for the query (sequence found in input sequence) with the corresponding position in the subject (database sequence). All mismatches are saved and compared with the chromosomal mutation database. It is

<table>
<thead>
<tr>
<th>Species</th>
<th>Gene</th>
<th>Chromosomal mutations</th>
<th>Resistance</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pmrA</td>
<td>S39I, R81S</td>
<td>colistin</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>pmrB</td>
<td>V161G</td>
<td>colistin</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>folP</td>
<td>P63R, P64L, P64S, P64A, P64H</td>
<td>sulphonamides</td>
<td>9</td>
</tr>
<tr>
<td>23S</td>
<td>A2059G</td>
<td>macrolide</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>16S rrsB</td>
<td>A964G, G1053A, C1054T, A1055G, G1058C</td>
<td>tetracycline</td>
<td>13,16</td>
<td></td>
</tr>
<tr>
<td>16S rrsB</td>
<td>T1406A, A1408G</td>
<td>gentamicin</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>16S rrsR</td>
<td>C1192T</td>
<td>spectinomycin</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td></td>
<td>parC</td>
<td>T66I, G78D, S80R, S80I, E84K, E84G</td>
<td>quinolone</td>
<td>4,21</td>
</tr>
<tr>
<td>16S rrsD</td>
<td>C1065T, C1192T</td>
<td>spectinomycin</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>23S</td>
<td>A2074G, A2074T, A2074C, A2075G</td>
<td>macrolide</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td></td>
<td>cmeR</td>
<td>A86G</td>
<td>macrolide</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>rpsL</td>
<td>A103C</td>
<td>spectinomycin</td>
<td>30</td>
</tr>
</tbody>
</table>

Materials and methods

Chromosomal mutation database

Information regarding mutations in chromosomal genes associated with antimicrobial resistance was collected from published papers (Table 1). The reference sequences were selected from WT *Escherichia coli* strain K-12 (MG1655) for the *E. coli* database, *Salmonella* Typhimurium strain LT2 for the *Salmonella enterica* database and *Campylobacter jejuni* NCTC 11168 for the *C. jejuni* database.

Bacterial isolates and WGS data

In total, 150 isolates covering three species were included in the study: *E. coli* (n = 50) and *Salmonella* (n = 50) isolates from the in-house strain collection at the National Food Institute and *C. jejuni* (n = 50) isolates from the in-house strain collection at Statens Serum Institut. The isolates were selected on the basis of having both WGS data and phenotypes available. The *Salmonella* isolates included strains from 10 different serovars (Tables S1 to S3, available as Supplementary data at JAC Online). All bacterial isolates were sequenced using the Miseq platform (Illumina) to obtain paired-end sequences and assembled de novo using Velvet (reference software). Bacterial strains were screened for phenotypic resistance using MIC determinations interpreted according to EUCAST (www.eucast.org). Only the susceptibility tests relevant for antimicrobial resistance associated with chromosomal point mutations for each species were analysed (Table 2). As resistance to some of the antimicrobial agents can be caused by either acquired genes or chromosomal point mutations, ResFinder-2.1 (www.genomicepidemiology.org)31 was used to detect known acquired resistance genes in the WGS data, using a threshold of 98% identity (%ID) and 60% length (minimum percentage length of the resistance gene to be covered). All isolates with disagreement between the phenotypic and predicted susceptibility were re-tested.

PointFinder

PointFinder consists of two databases: a chromosomal gene database, with all reference sequences in fasta format; and a chromosomal mutation database containing information on codon positions and substitutions. PointFinder uses BLASTn for identifying the best match for each gene in the chromosomal gene database, and only hits with an identity of >80% are further analysed. The program goes through each alignment comparing each position for the query (sequence found in input sequence) with the corresponding position in the subject (database sequence). All mismatches are saved and compared with the chromosomal mutation database. It is
The fastq files corresponding to the paired-end reads were mapped against the chromosomal gene sequence database using the assimpler.py script described in Joensen et al. In brief, 17mers from the reads were mapped to the reference sequence and extended to ungapped alignments that were considered significant if they had a score of at least 50, using a match score of 1 and a mismatch score of −3. A base was called if \( Z = (X - Y)/(X + Y) \) was > 3.29, where \( X \) is the number of observations of the most common nucleotide and \( Y \) is the number of other nucleotides at that position. Furthermore, nucleotide calls were considered significant only when the most common nucleotide was at least 10 times more abundant than other nucleotides at the position. All mismatches in positions from the chromosomal database were outputted except silent mutations, which were discarded. In cases with disagreement between PointFinder and mapping, the isolates were re-assembled de novo using SPAdes and re-analysed by PointFinder.

**Results**

**MIC and predicted antimicrobial resistance**

The 150 isolates were each tested against four to six different antimicrobial agents (Table 2), leading to a total of 684 susceptibility test results (Tables S1–S3). These results were compared with the results from PointFinder, mapping and ResFinder. Resistance to colistin, sulphonamides, tetracycline, erythromycin and spectinomycin can be caused by both chromosomal point mutations and acquired resistance genes; therefore results from both PointFinder and ResFinder were used to explain resistance.

For all Salmonella isolates, complete agreement between tested and predicted susceptibility was observed (Tables S1–S3). Disagreements in *E. coli* and *C. jejuni* were observed in five and four cases, respectively (Table 3).

The point mutation pmrB V161G was found by PointFinder in five *E. coli* isolates (E30–E34), but all tested phenotypically susceptible to colistin (MIC ≤ 1 mg/L). In *C. jejuni*, two isolates (C23 and C39) tested phenotypically resistant to ciprofloxacin (MIC 8) and nalidixic acid (MIC > 64 mg/L), while two (C8 and C24) tested erythromycin resistant (MIC > 128 mg/L), but neither mutations nor acquired genes were found that could explain the resistance.

**PointFinder versus mapping**

Mapping and PointFinder found the same mutations in all isolates except the five pmrB V161G mutations found by PointFinder in *E. coli* strains (E30–E34). The five isolates were re-assembled de novo using SPAdes and run through PointFinder, and this time no mutations were found in pmrB in any of the isolates. The codon change detected in the Velvet assembly of the five isolates was GTG→GGG, and when looking further into the sequences, the mapping showed that 28%–37% (Table 4) of the reads mapping to pmrB contained GGG instead of GTG.

**Discussion**

This study showed a high agreement between phenotypic susceptibility tests and WGS-predicted resistance, with only 11 (1.6%) mismatches. However, since the number of isolates included in the evaluation was very limited and selected, this has to be further verified in future studies. The six disagreements observed in *C. jejuni* all involved predicted susceptibility, whereas the isolates were phenotypically resistant, which may be due to unknown novel genes or mutations, as neither ResFinder nor PointFinder can detect novel resistance mechanisms.

We found that the BLAST-based method was dependent on the assembly method, which can cause either false-positive or -negative results. As the mapping method does not depend on the assembly this method gives a more precise result, which is consistent with a recent study by Clausen et al. Exploring the sequences mapping to pmrB, we found that ~1/3 of the isolate sequences for
Table 4. Mapped sequences to pmrB position 161 (amino acid position)

<table>
<thead>
<tr>
<th>Isolate ID</th>
<th>Total mapped sequences</th>
<th>No. of sequences mapping to GTG (%)</th>
<th>No. of sequences mapping to GGG (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>E30</td>
<td>69</td>
<td>50 (72)</td>
<td>19 (28)</td>
</tr>
<tr>
<td>E31</td>
<td>70</td>
<td>44 (63)</td>
<td>26 (37)</td>
</tr>
<tr>
<td>E32</td>
<td>80</td>
<td>54 (68)</td>
<td>26 (33)</td>
</tr>
<tr>
<td>E33</td>
<td>70</td>
<td>47 (67)</td>
<td>23 (33)</td>
</tr>
<tr>
<td>E34</td>
<td>89</td>
<td>64 (72)</td>
<td>25 (28)</td>
</tr>
</tbody>
</table>

each isolate contained the V161G mutation, indicating that there may be more than one copy of the pmrB gene present or that the sequence consisted of more than a single isolate. For many professionals working with WGS data, assemblies are still the preferred format. Due to their smaller size, assemblies are easier to share, upload and manage. Therefore, users working with assembly-based methods should consider that the data quality and method of assembly might influence the output.

Both ARG-ANNOT and CARD have tried to incorporate chromosomal point mutations in their databases. ARG-ANNOT has a database with partial sequences for chromosomal mutational regions of genes associated with mutational resistance, as well as information about position and mutation in the corresponding gene. ARG-ANNOT does not automatically detect these mutations, so the user has to manually browse through the alignment to detect potential mutations. CARD’s resistance gene identifier (RGI) protein variant formation about position and mutation in the corresponding gene. Unfortunately, neither ARG-ANNOT nor CARD takes the bacterial species into account. This means that both methods also output possible mutations/sequences related to mutations in quinolone-resistant Escherichia coli and Klebsiella pneumoniae isolates using whole genomic sequence data. J Antimicrob Chemother 2013; 68:2234–44.


Transparency declarations

None to declare.

Supplementary data

Tables S1 to S3 are available as Supplementary data at JAC Online.

References


