J Antimicrob Chemother 2012; **67**: 2640–2644
doi:10.1093/jac/dks261 Advance Access publication 10 July 2012

Journal of Antimicrobial Chemotherapy

Identification of acquired antimicrobial resistance genes

Ea Zankari[1,2]*, Henrik Hasman[1], Salvatore Cosentino[2], Martin Vestergaard[1], Simon Rasmussen[2], Ole Lund[2], Frank M. Aarestrup[1] and Mette Voldby Larsen[2]

[1]*National Food Institute, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark;* [2]*Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark*

*Corresponding author. Tel: +45-35887183; E-mail: east@food.dtu.dk

Received 13 March 2012; returned 26 April 2012; revised 8 June 2012; accepted 13 June 2012

Objectives: Identification of antimicrobial resistance genes is important for understanding the underlying mechanisms and the epidemiology of antimicrobial resistance. As the costs of whole-genome sequencing (WGS) continue to decline, it becomes increasingly available in routine diagnostic laboratories and is anticipated to substitute traditional methods for resistance gene identification. Thus, the current challenge is to extract the relevant information from the large amount of generated data.

Methods: We developed a web-based method, ResFinder that uses BLAST for identification of acquired antimicrobial resistance genes in whole-genome data. As input, the method can use both pre-assembled, complete or partial genomes, and short sequence reads from four different sequencing platforms. The method was evaluated on 1862 GenBank files containing 1411 different resistance genes, as well as on 23 *de-novo*-sequenced isolates.

Results: When testing the 1862 GenBank files, the method identified the resistance genes with an ID=100% (100% identity) to the genes in ResFinder. Agreement between *in silico* predictions and phenotypic testing was found when the method was further tested on 23 isolates of five different bacterial species, with available phenotypes. Furthermore, ResFinder was evaluated on WGS chromosomes and plasmids of 30 isolates. Seven of these isolates were annotated to have antimicrobial resistance, and in all cases, annotations were compatible with the ResFinder results.

Conclusions: A web server providing a convenient way of identifying acquired antimicrobial resistance genes in completely sequenced isolates was created. ResFinder can be accessed at www.genomicepidemiology.org. ResFinder will continuously be updated as new resistance genes are identified.

Keywords: antibiotic resistance, genotype, ResFinder, resistance gene identification

Introduction

The introduction of antimicrobial agents for treatment of infectious diseases is one of the most important achievements of the 20th century. However, soon after their introduction, isolates with acquired resistance emerged and this pattern has followed the introduction of each new antimicrobial agent.

A large number of different genes can be responsible for antimicrobial resistance. Identification of these genes is important to understand resistance epidemiology, for verification of non-susceptible phenotypes and for identification of resistant strains, when genes are weakly expressed *in vitro*. Detection of resistance genes has typically been performed using PCR[1] or microarrays.[2] However, in several cases, it is necessary to perform supplementary sequencing of the amplified PCR products.[3] As a result, it is expensive and time-consuming to perform a complete identification of resistance genes present in a strain collection.

The cost of DNA sequencing has steadily gone down, by roughly 10-fold every five years. As a consequence, DNA sequencing is becoming increasingly accessible for routine use and was recently utilized for complete characterization of antimicrobial resistance and virulence gene content during the safety evaluation of 28 strains intended for use in human nutrition.[4] The challenge is, however, to extract the relevant information from the large amount of data that is generated by these techniques.

The Center for Genomic Epidemiology (www.genomicepidemiology.org) aims at providing the bioinformatic and scientific foundation for processing and handling whole-genome sequencing (WGS) information in a standardized way useful for outbreak investigation, source tracking, diagnostics and epidemiological surveillance. The services are publically available through web servers specifically designed to be user-friendly—and also for investigators with limited bioinformatics experience.

We here present ResFinder, a web server that uses WGS data for identifying acquired antimicrobial resistance genes in bacteria.

Methods

Databases

Data on acquired resistance genes was collected from databases (http://faculty.washington.edu/marilynr/, http://ardb.cbcb.umd.edu/ and http://www.lahey.org/Studies/) and published papers including reviews.[5,6] All sequences were collected from the NCBI nucleotide database (http://www.ncbi.nlm.nih.gov/nuccore/) and used to build the ResFinder database. To our knowledge, we have created the largest collection of acquired antimicrobial resistance genes (see Table S1, available as Supplementary data at *JAC* Online).

Identifying resistance genes in completely sequenced bacteria

Draft assembly of short sequence reads was done as previously described.[7] All genes from the ResFinder database were BLASTed against the assembled genome, and the best-matching genes were given as output. For a gene to be reported, it has to cover at least 2/5 of the length of the resistance gene in the database. The best-matching genes were identified as previously[7]. It is possible to select a % identity (ID) threshold (the percentage of nucleotides that are identical between the best-matching resistance gene in the database and the corresponding sequence in the genome). The default ID is 100%.

Evaluation of method

Verification of the databases was made by testing ResFinder with the 1862 GenBank files from which the genes were collected, to verify that the method would find all genes with ID=100%.

Short sequence reads from 23 isolates of five different species, *Escherichia coli*, *Klebsiella pneumoniae*, *Salmonella enterica*, *Staphylococcus aureus* and *Vibrio cholerae*, were also submitted to ResFinder. All 23 isolates had been sequenced on the Illumina platform using paired-end reads. A ResFinder threshold of ID=98.00% was selected, as previous tests of ResFinder had shown that a threshold lower than this gives too much noise (e.g. fragments of genes). Phenotypic antimicrobial susceptibility testing was determined as MIC determinations, as previously described.[8]

With '(chromosome and plasmid)(multi-drug or antimicrobial or antibiotic)(resistant or resistance) pathogen' as search criteria, one isolate from each species with completely sequenced and assembled, and chromosome and plasmid data were collected from the NCBI Genomes database (http://www.ncbi.nlm.nih.gov/genome). This resulted in 30 isolates, from 30 different species, containing 85 chromosome/plasmid sequences. All sequences were run through all databases in ResFinder with a selected threshold of ID=98.00%.

Results

Using ResFinder

Short sequence reads can be assembled to draft genomes by the server. It is also possible to input a complete or partial, pre-assembled genome. ResFinder gives the option to run the input against one or several antimicrobial classes simultaneously, and it uses BLAST to identify the acquired resistance genes. It is possible to search for genes with specified similarity from 80%–100% identity, and the best-matching genes are given as output. An example of the output format is shown and explained at www.cbs.dtu.dk/services/ResFinder/output.php.

Evaluation of method

In all cases, ResFinder identified the acquired resistance genes in the 1862 GenBank files from which the databases were created, with an ID=100%.

Table 1 shows antimicrobial genes found by ResFinder, the predicted resistance profile and the phenotypic antimicrobial susceptibility test results for five bacterial isolates covering five different species. Tests for all 23 bacterial isolates covering the five different species can be seen in Table S2 (available as Supplementary data at *JAC* Online). Almost complete agreement between *in silico* predictions and phenotypic testing was found. The exceptions were two *S. aureus* isolates that contained the *mecA* gene but were phenotypically susceptible to penicillins, and two *S. aureus* isolates, one resistant to spectinomycin and the other to tiamulin, neither of which was found to contain genes matching these phenotypes. The *catB3* gene was found in all four *K. pneumoniae* isolates with an ID=100%, but not in full length, consistent with all four testing phenotypically susceptible to chloramphenicol. One *V. cholera* isolate contained part of *floR* and tested phenotypically susceptible to florfenicol.

Acquired antimicrobial resistance genes were found in 10 of the 30 strains from the NCBI genomes database (Table 2). For all except two isolates this coincided with the ResFinder results. *K. pneumoniae* KCTC 2242 was annotated to contain bla_{TEM}, whereas ResFinder detected bla_{SHV}. *Nocardia farcinica* IFM 10152 was annotated to contain a β-lactam gene as well as *aph(2″)*, *aph(3′)* and *aph(6)*, but ResFinder detected only the bla_{FAR-1} gene. These genes were further examined with BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi), which demonstrated that the genes detected by ResFinder were correct.

Discussion

Since their original development by Alexander Fleming, phenotypic disc diffusion and MIC determinations have been the gold standard for antimicrobial susceptibility testing. These methods have the great advantage of determining the 'true' *in vitro* relationship between the antimicrobial agent and the strain tested, and will detect any new emerging resistance mechanisms.

Genotypic testing of suspected resistant isolates is often performed to verify phenotypic observations and for epidemiological purposes. The most widely used approach has been to perform PCR to detect the presence of selected genes. In many cases only a single or a few genes mediating resistance are tested, and such studies will often miss the simultaneous presence of multiple genes encoding the same resistance.

WGS has the great benefit that it potentially provides complete information, and thus new experiments do not have to be performed to search for the presence of novel genes—the analysis can simply be rerun. One major obstacle is the lack of available bioinformatics tools allowing simple and standardized analysis of the large amounts of data generated by WGS.

We have developed, implemented and evaluated ResFinder, a method to detect the presence of 1862 different resistance genes from 12 different antimicrobial classes in WGS data (www.genomicepidemiology.org). The current version only

Table 1. ResFinder results for isolates of five different species compared with antimicrobial susceptibility data

Species	Isolate	ResFinder profile	Predicted phenotype	Detected phenotype
E. coli	Ødemsyge-186	*tet(A)*	TET	TET
K. pneumoniae	Kleb-6-1-264y	*aac(3)-IIa*[a]	GEN	GEN
		strA, *strB*	STR	STR
		$bla_{CTX\text{-}M\text{-}15}$	XNL, CTX, AMP	XNL, CTX, AMP
		$bla_{TEM\text{-}1}$	AMP	AMP
		$bla_{OXA\text{-}30}$	AMP, AMC	AMP, AMC
		$bla_{SHV\text{-}28}$	XNL, CTX, AMP	XNL, CTX, AMP
		aac(6′)Ib-cr	CIP	CIP[f]
		catB3[b]	CHL	—
		sul2	SMX	SMX
		tet(A)	TET	TET
		dfrA14[a]	TMP	TMP
		—	—	NAL[f]
S. enterica	Styph-0210H31581	*aac(6′)-Iaa*	[c]	—
		aadA2	SPT, STR	SPT, STR
		$bla_{CARB\text{-}2}$	AMP	AMP
		floR[a]	FFN, CHL	FFN, CHL
		sul1[b]	SMX	SMX
		tet(G)	TET	TET
		—	—	CIP[f], NAL[f]
S. aureus	2007-70-91-4	*aac(3)-Ik*[a]	[d]	—
		mecA	FOX	—
		blaZ	PEN	PEN
		tet(K), *tet(38)*[a], *tet(M)*[a]	TET	TET
		dfrG	TMP	TMP
		fusA[a]	FUS	—
V. cholerae	Vchole-002	*strA*, *strB*	STR	STR
		catB9	CHL[e]	—
		sul2	SMX	SMX
		dfrA1, *dfrA31*	TMP	TMP
		—	—	CIP[f], NAL[f], CST[f]

AMC, amoxicillin/clavulanate (2:1); AMP, ampicillin; CHL, chloramphenicol; CST, colistin; CTX, cefotaxime; FOX, cefoxitin; FFN, florfenicol; FUS, fusidic acid; GEN, gentamicin; PEN, penicillin; SMX, sulfamethoxazole; SPT, spectinomycin; STR, streptomycin; TET, tetracycline; TMP, trimethoprim; XNL, ceftiofur.

[a]The gene is found with an ID < 100%.

[b]The found gene is shorter than the resistance gene.

[c]Resistance to antimicrobials that were not included in the phenotypic antimicrobial susceptibility tests.

[d]Phenotype not known.

[e]Phenotypically silent in native position (19).

[f]Antimicrobial drug associated with chromosomal mutations.

covers horizontally acquired resistance genes and not resistance mediated by mutations, e.g. in housekeeping genes. ResFinder can also be used to ignore known acquired resistance genes in a search for new resistance genes.

ResFinder successfully identified all the genes from which the database was built, and correctly identified all genes present in 30 isolates of whole-genome data collected from the NCBI genomes database (http://www.ncbi.nlm.nih.gov/genome). Furthermore, phenotypic antimicrobial susceptibility tests of 23 isolates from five different species were compared with the results from ResFinder. With a few exceptions, complete agreement between predicted and observed phenotypes was found. All the *V. cholerae* isolates contained the *catB9* gene, which has previously been shown to be phenotypically silent in its native position,[9] consistent with all isolates testing phenotypically susceptible. The five *S. aureus* isolates examined in this study were from a collection of methicillin-resistant *S. aureus* (MRSA).[10] Phenotypic detection of *mecA*-harbouring isolates can be difficult, indicating the superiority of WGS compared with phenotypic testing. Two of the *S. aureus* isolates, 9B and PR11_08, showed phenotypic resistance to spectinomycin and tiamulin, respectively, but without containing any matching resistance genes. Interestingly, we found two extended-spectrum β-lactamase (ESBL)-related genes ($bla_{CTX\text{-}M\text{-}15}$ and $bla_{SHV\text{-}28}$) in

Table 2. ResFinder results for completely sequenced and assembled chromosome and plasmid data from 30 different species

Strain	Annotated resistance	Chromosome	Plasmid
Edwardsiella tarda EIB202	*tet(A)*, *tet(R)*, *strA*, *strB*, *sul2*	no genes found	*strA* 100% *strB* 100% *catA3* 99.84% *sul2* 100%; *tet(A)* 99.92%
Enterobacter cloacae subsp. *cloacae* ATCC 13047	—	*sul2* 100%	no genes found
Enterococcus faecalis	*tet(M)*	*tet(M)* 100%	no genes found
Fusobacterium nucleatum subsp. *polymorphum* ATCC 10953	—	*tet(K)* 100%	no genes found
Klebsiella pneumoniae KCTC 2242	β-lactam (bla_{TEM})	bla_{SHV-99} 99.88%	no genes found
Nocardia farcinica IFM 10152	β-lactam, *aph(2″)*, *aph(3′)*, *aph(6)*	bla_{FAR-1} 98.77%	no genes found
Ochrobactrum anthropi ATCC 49188	—	bla_{OCH-3} 99.91%	no genes found
Ralstonia pickettii 12J	—	bla_{OXA-60} 100%	no genes found
Staphylococcus aureus subsp. *aureus* COL	methicillin resistant	*aac(3)-Ik* 98.87% *mecA* 100% *tet(38)* 99.85%	*tet(K)* 100%
Streptococcus suis BM407	*tet(M)*, *tet(O)*, *tet(L)*, chloramphenicol acetyltransferase	*erm(B)* 99.86% *tet(M)* 99.83% *tet(O)* 99.64% *tet(L)* 100%	no genes found
Xanthomonas campestris pv. *vesicatoria* str. 85-10	streptomycin resistance	*strA* 100% *strB* 100%	no genes found

All sequences were run through all databases in ResFinder with a selected threshold of ID = 98.00%. The following strains had no annotated resistance, and no resistance genes were detected by ResFinder: *Bacillus anthracis* str. 'Ames Ancestor', *Bacillus cereus* 03BB102, *Bacillus thuringiensis* BMB171, *Burkholderia glumae* BGR1, *Burkholderia multivorans* ATCC 17616, *Clavibacter michiganensis* subsp. *michiganensis* NCPPB 382, *Coxiella burnetii* CbuK_Q154, *Cronobacter turicensis* z3032, *Erwinia amylovora* CFBP143, *Erwinia pyrifoliae* DSM 12163, *Helicobacter pylori* B8, *Legionella longbeachae* NSW150, *Listeria monocytogenes* 08-5578, *Pantoea ananatis* AJ13355, *Ralstonia solanacearum* GMI100, *Vibrio harveyi* ATCC BAA-1116, *Vibrio vulnificus* YJ016, *Yersinia enterocolitica* subsp. *enterocolitica* 8081 and *Yersinia pseudotuberculosis* PB1+.

all four *K. pneumonia* isolates. If we had used PCR to detect genes, we would probably not have found more than one, as it is common to cease looking for more genes after a matching gene is found. ResFinder can therefore potentially give more information than the existing method.

ResFinder is a further step in our development of bioinformatics tools for analyzing WGS data; the tools are specifically designed to be easy to use—and for investigators with limited bioinformatics experience. An online tool allowing identification of multilocus sequence types is already available.[7] Additional tools under development include those for the identification of virulence genes and species, and identification and phylogenetic analysis based on single-nucleotide polymorphism and pan-genome analysis.

ResFinder will continuously be updated to include additional and novel emerging resistance genes as they are identified.

Acknowledgements

We are grateful to Inge M. Hansen and John Damm Sørensen for excellent technical assistance.

Funding

This study was supported by the Center for Genomic Epidemiology (www.genomicepidemiology.org) grant 09-067103/DSF from the Danish Council for Strategic Research and by the European Union Reference Laboratory for Antimicrobial Resistance.

Transparency declarations

None to declare.

Supplementary data

Tables S1 and S2 are available as Supplementary data at *JAC* Online (http://jac.oxfordjournals.org/).

References

1 Aarestrup FM, Agerso Y, Gerner-Smidt P *et al*. Comparison of antimicrobial resistance phenotypes and resistance genes in *Enterococcus faecalis* and *Enterococcus faecium* from humans in the community, broilers, and pigs in Denmark. *Diagn Microbiol Infect Dis* 2000; **37**: 127–37.

2 Batchelor M, Hopkins KL, Liebana E *et al*. Development of a miniaturised microarray-based assay for the rapid identification of antimicrobial resistance genes in Gram-negative bacteria. *Int J Antimicrob Agents* 2008; **31**: 440–51.

3 Hasman H, Mevius D, Veldman K *et al*. β-Lactamases among extended-spectrum β-lactamase (ESBL)-resistant *Salmonella* from poultry, poultry products and human patients in The Netherlands. *J Antimicrob Chemother* 2005; **56**: 115–21.

4 Bennedsen M, Stuer-Lauridsen B, Danielsen M *et al*. Screening for antimicrobial resistance genes and virulence factors via genome sequencing. *Appl Environ Microbiol* 2011; **77**: 2785–7.

5 Shaw KJ, Rather PN, Hare RS *et al*. Molecular genetics of aminoglycoside resistance genes and familial relationships of the aminoglycoside-modifying enzymes. *Microbiol Rev* 1993; **57**: 138–63.

6 van Hoek AH, Mevius D, Guerra B *et al*. Acquired antibiotic resistance genes: an overview. *Front Microbiol* 2011; **2**: 203.

7 Larsen MV, Consentino S, Rasmussen S *et al*. Multilocus sequence typing of total-genome-sequenced bacteria. *J Clin Microbiol* 2012; **50**: 1355–61.

8 Hendriksen RS, Seyfarth AM, Jensen AB *et al*. Results of use of WHO Global Salm-Surv external quality assurance system for antimicrobial susceptibility testing of *Salmonella* isolates from 2000 to 2007. *J Clin Microbiol* 2009; **47**: 79–85.

9 Rowe-Magnus DA, Guerout AM, Mazel D. Bacterial resistance evolution by recruitment of super-integron gene cassettes. *Mol Microbiol* 2002; **43**: 1657–69.

10 Price LB, Stegger M, Hasman H *et al*. *Staphylococcus aureus* CC398: host adaptation and emergence of methicillin resistance in livestock. *MBio* 2012; **3**: e00305–11.