Published online 2 October 2008 *Nucleic Acids Research, 2009, Vol. 37, Database issue* ***D443–D447***
doi:10.1093/nar/gkn656

ARDB—Antibiotic Resistance Genes Database

Bo Liu[1] and Mihai Pop[1,2,*]

[1]Center for Bioinformatics and Computational Biology and [2]Department of Computer Science, University of Maryland, College Park, MD 20742, USA

Received August 14, 2008; Revised September 15, 2008; Accepted September 16, 2008

ABSTRACT

The treatment of infections is increasingly compromised by the ability of bacteria to develop resistance to antibiotics through mutations or through the acquisition of resistance genes. Antibiotic resistance genes also have the potential to be used for bio-terror purposes through genetically modified organisms. In order to facilitate the identification and characterization of these genes, we have created a manually curated database—the Antibiotic Resistance Genes Database (ARDB)—unifying most of the publicly available information on antibiotic resistance. Each gene and resistance type is annotated with rich information, including resistance profile, mechanism of action, ontology, COG and CDD annotations, as well as external links to sequence and protein databases. Our database also supports sequence similarity searches and implements an initial version of a tool for characterizing common mutations that confer antibiotic resistance. The information we provide can be used as compendium of antibiotic resistance factors as well as to identify the resistance genes of newly sequenced genes, genomes, or metagenomes. Currently, ARDB contains resistance information for 13 293 genes, 377 types, 257 antibiotics, 632 genomes, 933 species and 124 genera. ARDB is available at http://ardb.cbcb.umd.edu/.

INTRODUCTION

The discovery of penicillin in 1928 by Alexander Fleming has revolutionized the treatment of bacterial infections. The large-scale use of antibiotics, however, has also led to an increase in the number of microbes that can resist treatment. Drug resistant bacteria are an increasing threat to public health, as highlighted by a recent estimate that in the US methicillin-resistant *Staphylococcus aureus* (MRSA) may contribute to more deaths than HIV (1). Methicillin-resistant strains of *S. aureus* were initially documented in the 1960s (2) and have been associated with higher mortality rates (3,4) than their drug-sensitive counterparts. Similar challenges are posed by the emergence of multidrug- and extensively-drug resistant tuberculosis (MDR-TB and XDR-TB, respectively) (5,6). Antibiotic resistance can result from large genomic changes, such as the acquisition of entire plasmids or mobile elements encoding resistance factors. Recent studies are, however, revealing the important role small mutations play in the evolution of resistance. For example, only 35 point-mutations distinguish a vancomycin-resistant strain of *S. aureus* from its sensitive counterpart, and these mutations evolved in just 3 months within an infected patient (7). Furthermore, antibiotic resistance genes have the potential to be used for bioterrorism purposes through genetically modified organisms. These factors emphasize the urgent need for a better understanding of the mechanisms through which bacteria develop resistance, as well as for the development of new techniques for the rapid identification of resistance factors. The database presented in this article provides a first component of an informatics infrastructure aimed at enabling such studies.

Several mechanisms have been characterized through which bacteria become resistant to antibiotics (8): (i) the production of enzymes that digest/metabolize the antibiotic; (ii) efflux pumps that eliminate the drug from the cell; (iii) modifications to the cellular target of the antibiotic that prevent binding; (iv) activation of an alternate pathway that bypasses drug action; and (v) particularly for gram-negative bacteria, down-regulation or elimination of transmembrane porins through which drugs enter the cell (9). The annotation information commonly associated with genes deposited in public databases is insufficiently detailed for representing this variety of resistance mechanisms and the additional meta-information relevant in this context. Specifically, each resistance gene is associated with a resistance profile (set of antibiotics or classes of antibiotics targeted by the gene), yet this information is usually not available. Second, resistance often requires the cooperation of multiple genes, usually within a same operon [e.g. vancomycin resistance VanA operon requires seven genes (10)], while most annotation information is targeted at individual genes. Finally, resistance frequently results from

*To whom correspondence should be addressed. Tel: +1 301 405 7245; Fax: +1 301 314 1341; Email: mpop@umiacs.umd.edu

modifications to, or the disruption of an individual gene (e.g. modifications of the drug target), information incompatible with standard annotation procedures. Consequently, specialized resources are necessary for annotating and cataloging information related to antibiotic resistance.

Several recent efforts have been made to partially unify this information, such as Antibiotic Resistance Genes Online (ARGO) (11), MvirDB (12) and a compendium of TEM β-lactamase genes at the Lahey Clinic (http://www.lahey.org/Studies/). All, however, have limited functionality. ARGO only contains part of β-lactamase, vancomycin and tetracycline resistance genes. In addition, it does not include rich annotation information such as resistance profile, mechanism of action, operon information or gene sequence. Furthermore, many of the links between ARGO and GenBank target incorrect records (e.g. links to a genome instead of the relevant gene record). MvirDB is a broad repository of virulence-associated genes, including toxins, virulence factors and antibiotic resistance. The latter information is simply a replicate of the ARGO database. The Lahey Clinic website is a comprehensive collection of TEM type β-lactamases, which attempts to standardize the nomenclature for these genes. In addition to these specialized resources, antibiotic resistance information can be extracted in a restricted manner from GenBank (13) and SwissProt (14), databases that lack many important types of information relevant in this domain.

To address the limitations of currently available public resources, and to facilitate the identification and characterization of antibiotic resistance genes, we have created a manually curated database [Antibiotic Resistance Genes Database (ARDB)] unifying most of the publicly available genes and related information. Our motivations in creating ARDB are (i) to provide a centralized compendium of information on antibiotic resistance; (ii) to facilitate the consistent annotation of resistance information in newly sequenced organisms; and (iii) to facilitate the identification and characterization of new genes. We believe this resource will be found useful by a broad range of scientists, including microbiologists, clinicians and the biodefense research community.

DATABASE CONTENTS AND CONSTRUCTION

The diversity of antibiotic resistance genes, types and mechanisms, combined with the fact that related information, such as resistance profile, is mostly 'paper-bound' made the construction of ARDB both difficult and time-consuming. To compile, confirm and validate this collection of data, several textbooks and several hundred journal articles were searched and summarized.

The majority of protein and nucleic acid sequences of known antibiotic resistance genes were retrieved from the NCBI nucleotide and protein databases and additional sequences were retrieved from the Swiss-Prot database. Genes were grouped into resistance types based on their protein sequence similarity using the following approach. First, the sequence of an experimentally confirmed representative was identified for every type of resistance, based on literature searches and meta-information provided by the NCBI protein database. These representative resistance genes were then used to 'fish out' additional homologues using similarity searches against the NCBI nr database. The similarity cutoff was set at 80% unless a different value was recommended in the literature for a specific resistance type. Using this approach we identified 13 254 protein sequences putatively involved in antibiotic resistance. We filtered this set by removing vector sequences, synthetic constructs and redundant genes, resulting in a non-redundant set of 6206 proteins. This set was further refined by removing incomplete sequences, thereby yielding a core set of 4554 antibiotic resistance proteins. Each sequence was associated with corresponding CDD, COG, ontology and source organism information. Furthermore, the genes were grouped into resistance types, corresponding to clusters of genes with similar resistance profiles, operon membership and mechanism of action. In addition, basic information about known antibiotics was extracted from KEGG DRUG (15), PubChem, PubMed MeSH database and the Chemical Entities of Biological Interest (ChEBI) ontology. Although ARDB is mainly targeted at antibiotic resistance genes, 12 additional drug targets have also been included into ARDB with relevant information [16S rRNA (16), 23S rRNA, gyrA (17), gyrB, parC, parE, rpoB, katG, pncA, embB, folP, dfr], whose modification has been shown to confer resistance.

The data flow for the curation process is highlighted in Figure 1. ARDB is implemented as a MySQL relational database, and the corresponding schema is available on our website. Access to this database is provided through a CGI-based web interface.

ONTOLOGY INFORMATION

No comprehensive ontology is currently available for annotating antibiotic resistance information. To facilitate the computational analysis of antibiotic resistance information we have created a set of ontology terms aimed at characterizing both the resistance profile conferred by a specific gene and its specific mechanism of action. Specifically, for every antibiotic X, we have created a set of 'X resistance' terms. Furthermore, we classify several mechanisms of action, including drug target modification, replacement or protection, drug enzymatic destruction and drug transport. Drug transport is further subclassified into ATP-binding cassette (ABC) drug efflux, major facilitator superfamily (MFS) drug efflux, small multidrug resistance (SMR) drug efflux and resistance-nodulation-cell division (RND) drug efflux, following the terminology used in (18). These terms are defined within an Antibiotic Resistance (AR) ontology and are associated with each record present in our database. We are currently working with the broader ontology community to further refine this information and integrate it within existing ontology development efforts.

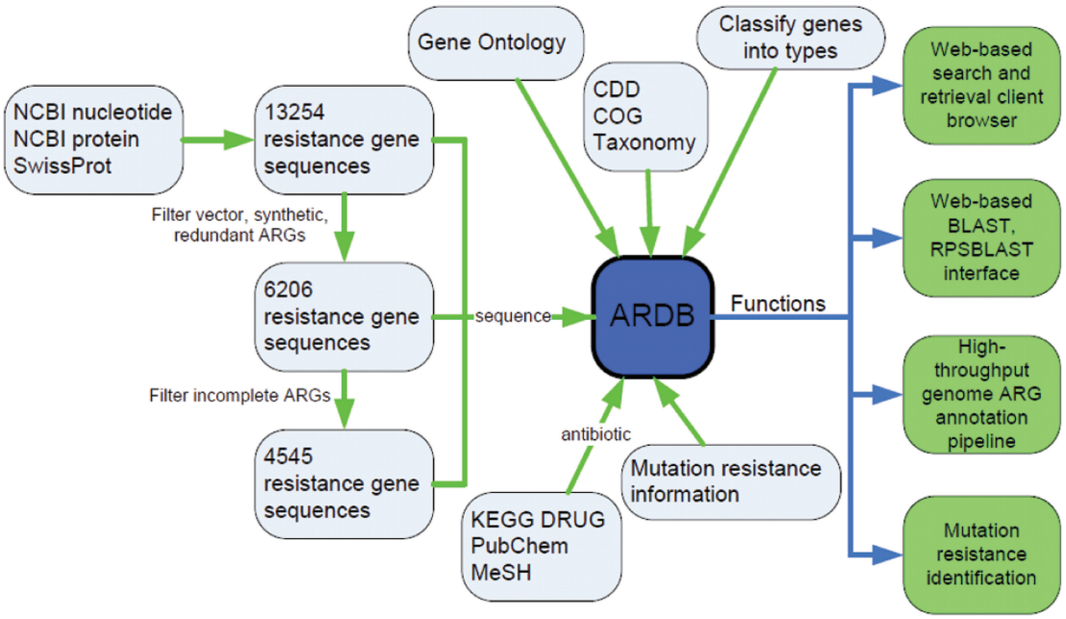

Figure 1. ARDB curation data flow.

DATA ACCESS AND DATA MINING

Users can access our database through a web interface at http://ardb.cbcb.umd.edu. This interface provides several modes of interaction as highlighted below.

Keyword searches

Simple keyword search is available at the top of each page of ARDB website (Figure 2a), providing a quick means for searching a specific object in our database (gene, type, antibiotic, genome and genus) (Figure 2b and f). Users can search all of the data, or narrow down the search to a specific type of information. For example, users interested in the molecular mechanisms of resistance to tetracycline can search for the keyword 'tetracycline' within the 'Resistance Type' database. An advanced search function is also available, allowing users to select from among the available keywords associated with each database field.

Similarity searches

BLAST. To help identify and annotate antibiotic resistance genes, a BLAST interface is also provided. One or more sequences can be provided to this interface in a multi-FASTA file, corresponding to a set of gene sequences. Furthermore, both nucleotide and amino-acid sequences are accepted by our system. The results can be visualized as standard BLAST output, however additional displays are provided that are specific to antibiotic resistance information. Our 'ARDB annotation format' groups individual BLAST hits according to resistance type as inferred from the level of similarity to the genes within the database associated with a specific type of resistance (Figure 2c). A second view allows users to download a tab-delimited spreadsheet summary of the antibiotic resistance genes identified within the uploaded file.

RPSBLAST. In addition to BLAST we also provide an RPSBLAST (19) interface relying on Position Specific Scoring Matrix (PSSM) created from sequences associated with each resistance type, using an approach similar to the NCBI Conserved Domain Database (20). The output of this interface is similar to that provided by the BLAST interface mentioned above.

Polymorphism detection. Additionally, a mutation-specific search function is provided to identify polymorphisms previously characterized to confer resistance (Figure 2d). For example, a G-C mutation at position 1058 of the *Escherichia coli* 16S rRNA has been shown to confer resistance to tetracycline (21). This information is extracted from the detailed BLAST alignment between the query sequence and a reference sequence in our database. Currently this function is available for 12 genes (16S rRNA, 23S rRNA, gyrA, gyrB, parC, parE, rpoB, katG, pncA, embB, folP, dfr), and we expect to extend it as more information becomes available in the literature.

Pre-annotated information

The antibiotic resistance profiles of 632 complete bacterial genomes have already been annotated and deposited in ARDB allowing quick search. This information can be conveniently extracted through keyword searches against the 'genome' database, or through the 'Genome Resistance Profiles Comparison' link from the front page. The latter approach allows users to summarize and compare the resistance profiles of multiple organisms present in our database.

Browse

A 'browse' function is available that allows the users to visualize several classes of antibiotic resistance genes, grouped by their resistance profile. This functionality

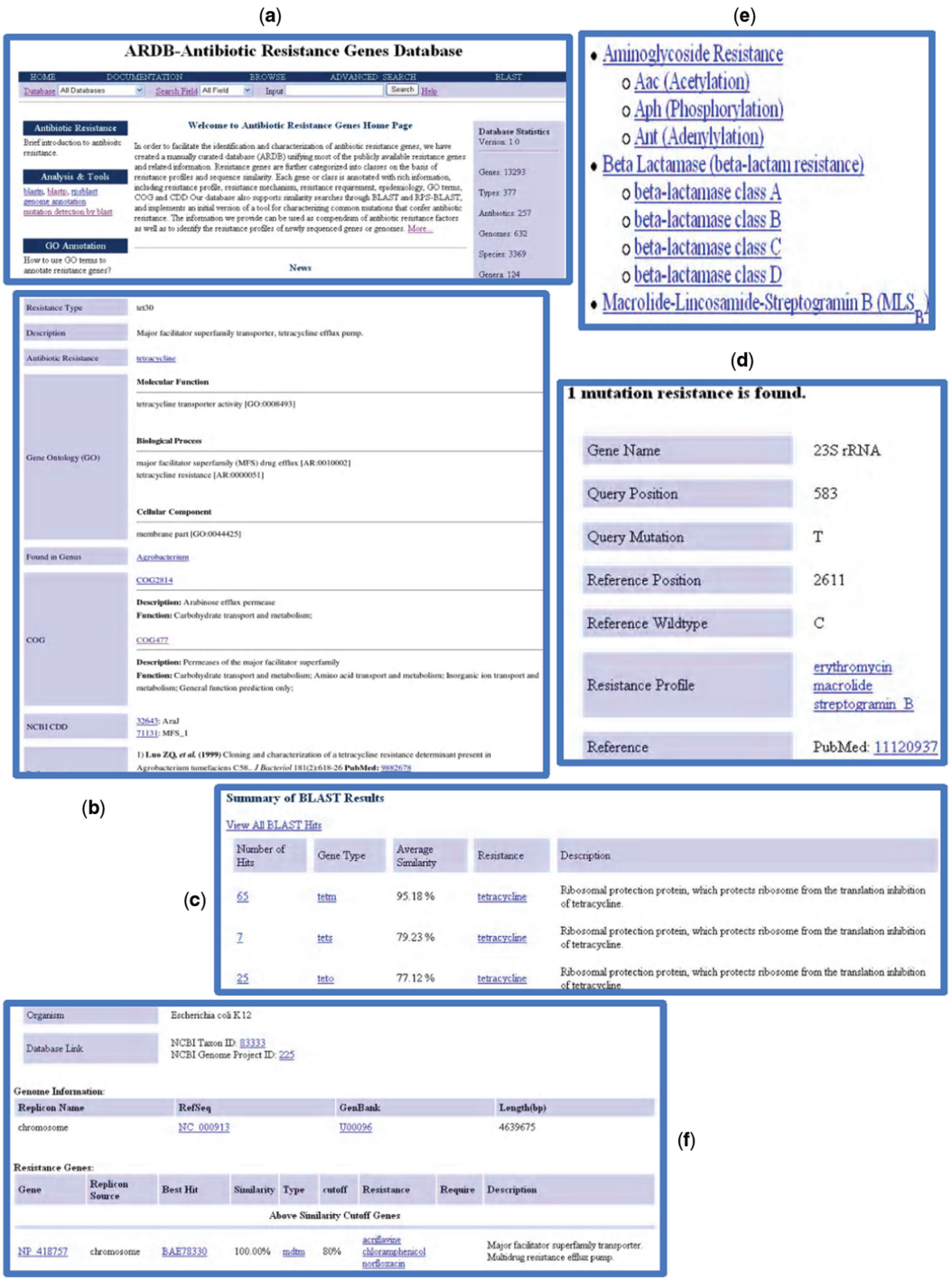

Figure 2. Sample web pages from ARDB. (**a**) Front page, (**b**) resistance type, (**c**) blast result, (**d**) mutation annotation, (**e**) browse and (**f**) genome information.

is currently available for aminoglycoside, β-lactam, macrolide–lincosamide–streptogramin B, multidrug transporter, tetracycline and vancomycin resistance (Figure 2e).

Submission

In order to facilitate community-driven refinement of our database we provide an interface through which users can submit information about novel resistance genes. This interface captures several types of information not commonly available in other databases [Minimum Inhibitory Concentration (MIC), resistance type, ontology, citation information, etc.]. Furthermore we provide a simple file format and upload functionality to facilitate the submission of information for multiple genes. The information received will be vetted and inserted into the database. We are also planning to develop an interface that allows community-deposited information to be

directly added to the database as 'provisional' records, pending additional manual curation.

CONCLUSION AND DISCUSSION

The database described in this article, ARDB, unifies most of the publicly available antibiotic resistance genes and provides a reliable annotation service to researchers investigating the molecular basis for resistance in bacteria. Because of the large diversity and the rapid identification of new resistance genes, the current version of ARDB is just a first catalog of currently available information, and will continue to be updated over the coming months and years. We plan to coordinate our development efforts with researchers actively involved in antibiotic resistance research as well as with the developers of biological ontologies and of databases storing related information (such as virulence factors or toxins). As part of these efforts we aim to refine the structure of our database, better determine the types of information stored and identify additional requirements for the user interface. Future efforts will also target the development of new approaches for cataloguing and characterizing polymorphisms correlated with resistance, as well as for annotating changes to cellular regulatory networks that underlie the mechanisms of drug tolerance.

ACKNOWLEDGEMENTS

We would like to thank Kim Bishop-Lilly and Tim Read for providing initial feedback on our database and for their insightful comments and advice.

FUNDING

Uniformed Services University of the Health Sciences, administered by the Henry Jackson Foundation (HU001-06-1-0015 to M.P.). Funding for open access charge: Uniformed Services University of the Health Sciences, administered by the Henry Jackson Foundation (HU001-06-1-0015).

Conflict of interest statement. None declared.

REFERENCES

1. Bancroft,E.A. (2007) Antimicrobial resistance: it's not just for hospitals. *JAMA*, **298**, 1803–1804.
2. Barber,M. (1961) Methicillin-resistant *staphylococci*. *J. Clin. Pathol.*, **14**, 385–393.
3. Selvey,L.A., Whitby,M. and Johnson,B. (2000) Nosocomial methicillin-resistant *Staphylococcus aureus* bacteremia: is it any worse than nosocomial methicillin-sensitive *Staphylococcus aureus* bacteremia? *Infect. Control. Hosp. Epidemiol.*, **21**, 645–648.
4. Delaney,J.A., Schneider-Lindner,V., Brassard,P. and Suissa,S. (2008) Mortality after infection with methicillin-resistant *Staphylococcus aureus* (MRSA) diagnosed in the community. *BMC Med.*, **6**, 2.
5. Sekiguchi,J., Miyoshi-Akiyama,T., Augustynowicz-Kopec,E., Zwolska,Z., Kirikae,F., Toyota,E., Kobayashi,I., Morita,K., Kudo,K., Kato,S. *et al.* (2007) Detection of multidrug resistance in *Mycobacterium tuberculosis*. *J. Clin. Microbiol.*, **45**, 179–192.
6. Gandhi,N.R., Moll,A., Sturm,A.W., Pawinski,R., Govender,T., Lalloo,U., Zeller,K., Andrews,J. and Friedland,G. (2006) Extensively drug-resistant tuberculosis as a cause of death in patients co-infected with tuberculosis and HIV in a rural area of South Africa. *Lancet*, **368**, 1575–1580.
7. Mwangi,M.M., Wu,S.W., Zhou,Y., Sieradzki,K., de Lencastre,H., Richardson,P., Bruce,D., Rubin,E., Myers,E., Siggia,E.D. *et al.* (2007) Tracking the in vivo evolution of multidrug resistance in *Staphylococcus aureus* by whole-genome sequencing. *Proc. Natl Acad. Sci USA*, **104**, 9451–9456.
8. Alekshun,M.N. and Levy,S.B. (2007) Molecular mechanisms of antibacterial multidrug resistance. *Cell*, **128**, 1037–1050.
9. Vila,J., Marti,S. and Sanchez-Cespedes,J. (2007) Porins, efflux pumps and multidrug resistance in *Acinetobacter baumannii*. *J. Antimicrob. Chemother.*, **59**, 1210–1215.
10. Courvalin,P. (2006) Vancomycin resistance in gram-positive cocci. *Clin. Infect. Dis.*, **42 (**Suppl. 1), S25–S34.
11. Scaria,J., Chandramouli,U. and Verma,S.K. (2005) Antibiotic Resistance Genes Online (ARGO): a Database on vancomycin and beta-lactam resistance genes. *Bioinformation*, **1**, 5–7.
12. Zhou,C.E., Smith,J., Lam,M., Zemla,A., Dyer,M.D. and Slezak,T. (2007) MvirDB—a microbial database of protein toxins, virulence factors and antibiotic resistance genes for bio-defence applications. *Nucleic Acids Res.*, **35**, D391–D394.
13. Benson,D.A., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J. and Wheeler,D.L. (2008) GenBank. *Nucleic Acids Res.*, **36**, D25–D30.
14. UniProt Consortium. (2008) The universal protein resource (UniProt). *Nucleic Acids Res.*, **36**, D190–D195.
15. Kanehisa,M., Araki,M., Goto,S., Hattori,M., Hirakawa,M., Itoh,M., Katayama,T., Kawashima,S., Okuda,S., Tokimatsu,T. *et al.* (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res.*, **36**, D480–D484.
16. Bilgin,N., Richter,A.A., Ehrenberg,M., Dahlberg,A.E. and Kurland,C.G. (1990) Ribosomal RNA and protein mutants resistant to spectinomycin. *EMBO J*, **9**, 735–739.
17. Ruiz,J., Moreno,A., Jimenez de Anta,M.T. and Vila,J. (2005) A double mutation in the gyrA gene is necessary to produce high levels of resistance to moxifloxacin in *Campylobacter spp.* clinical isolates. *Int. J. Antimicrob. Agents*, **25**, 542–545.
18. Higgins,C.F. (2007) Multiple molecular mechanisms for multidrug resistance transporters. *Nature*, **446**, 749–757.
19. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
20. Marchler-Bauer,A., Anderson,J.B., Cherukuri,P.F., DeWeese-Scott,C., Geer,L.Y., Gwadz,M., He,S., Hurwitz,D.I., Jackson,J.D., Ke,Z. *et al.* (2005) CDD: a Conserved Domain Database for protein classification. *Nucleic Acids Res.*, **33**, D192–D196.
21. Ross,J.I., Eady,E.A., Cove,J.H. and Cunliffe,W.J. (1998) 16S rRNA mutation associated with tetracycline resistance in a gram-positive bacterium. *Antimicrob. Agents Chemother.*, **42**, 1702–1705.