D566–D573 *Nucleic Acids Research, 2017, Vol. 45, Database issue* *Published online 26 October 2016*
doi: 10.1093/nar/gkw1004

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

Baofeng Jia[1], Amogelang R. Raphenya[1], Brian Alcock[1], Nicholas Waglechner[1], Peiyao Guo[1], Kara K. Tsang[1], Briony A. Lago[1], Biren M. Dave[1], Sheldon Pereira[1], Arjun N. Sharma[1], Sachin Doshi[1], Mélanie Courtot[2], Raymond Lo[2], Laura E. Williams[3], Jonathan G. Frye[3], Tariq Elsayegh[4], Daim Sardar[1], Erin L. Westman[1], Andrew C. Pawlowski[1], Timothy A. Johnson[1], Fiona S.L. Brinkman[2], Gerard D. Wright[1] and Andrew G. McArthur[1,*]

[1]M.G. DeGroote Institute for Infectious Disease Research, Department of Biochemistry and Biomedical Sciences, DeGroote School of Medicine, McMaster University, Hamilton, Ontario L8S 4K1, Canada, [2]Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia, V5A 1S6, Canada, [3]Bacterial Epidemiology and Antimicrobial Resistance Research Unit, USDA-ARS U.S. National Poultry Research Center, U.S. Department of Agriculture, Athens, GA 30605, USA and [4]School of Medicine, Royal College of Surgeons in Ireland, Dublin 2, Republic of Ireland

Received August 15, 2016; Revised September 30, 2016; Editorial Decision October 14, 2016; Accepted October 17, 2016

ABSTRACT

The Comprehensive Antibiotic Resistance Database (CARD; http://arpcard.mcmaster.ca) is a manually curated resource containing high quality reference data on the molecular basis of antimicrobial resistance (AMR), with an emphasis on the genes, proteins and mutations involved in AMR. CARD is ontologically structured, model centric, and spans the breadth of AMR drug classes and resistance mechanisms, including intrinsic, mutation-driven and acquired resistance. It is built upon the Antibiotic Resistance Ontology (ARO), a custom built, interconnected and hierarchical controlled vocabulary allowing advanced data sharing and organization. Its design allows the development of novel genome analysis tools, such as the Resistance Gene Identifier (RGI) for resistome prediction from raw genome sequence. Recent improvements include extensive curation of additional reference sequences and mutations, development of a unique Model Ontology and accompanying AMR detection models to power sequence analysis, new visualization tools, and expansion of the RGI for detection of emergent AMR threats. CARD curation is updated monthly based on an interplay of manual literature curation, computational text mining, and genome analysis.

INTRODUCTION

Antimicrobial resistance (AMR) has been observed since the first antibiotics were discovered, yet antimicrobial abuse and misuse has resulted in increasing levels of clinical resistance (1). As a consequence, AMR has become a global health crisis, exacerbated by a faltering drug discovery pipeline (2). Based on data from common clinical pathogens (e.g. the ESKAPE pathogens (3)), it is estimated that without appropriate action the death toll from highly or totally resistant infections could increase up to 10 million lives each year by 2050, at a cumulative cost to the global economic output of 100 trillion USD (1). A key component to combatting this crisis is collection of high-quality AMR surveillance data and, while this has traditionally been anchored in measurement of resistance phenotype (i.e. the 'antibiogram'), genome and metagenome sequencing is increasingly being used for surveillance of AMR genotypes, revealing important information on the relative roles of intrinsic resistance, mutation and the mobilome, in addition to the underlying mechanisms of resistance (3). With the advent of next-generation sequencing, enormous amounts of biological and clinical data are being generated throughout the biosciences (4) and AMR research is no exception. However, the exponential increase of biological data demands

[*]To whom correspondence should be addressed. Tel: +1 905 525 9140 (Ext. 21663); Fax: +1 905 528 5330; Email: mcarthua@mcmaster.ca
Present addresses:
Mélanie Courtot, European Bioinformatics Institute (EMBL-EBI), European Molecular Biology Laboratory, Wellcome Trust Genome Campus, Hinxton CB10 1SD, UK.
Peiyao Guo, David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada.
Laura E. Williams, Department of Biology, Providence College, Providence, RI 02918, USA.

improvements in current methods of data management, analysis and accessibility. AMR research and surveillance must increasingly gather data across biological scales, from molecules to populations, with parallel development of new data analysis and sharing paradigms. Construction of this 'Big Data' infrastructure will lead to data-driven predictions that complement traditional research methodologies. However, multidisciplinary understanding is difficult to fulfill due to the increasing volume of literature that makes it difficult to manually find relevant papers and extract valuable information. Biocuration, the act of making biological information accessible to both humans and computers via databases and data sharing tools, is thus an increasingly important part of biomedical research (5) and CARD seeks to fulfill this need for AMR research and surveillance.

There are currently a number of knowledge resources in the field of antimicrobial resistance, such as the Comprehensive Antibiotic Resistance Database (CARD) (6), Antibiotic Resistance Genes Online (ARGO) (7), Antibiotic Resistance Genes Database (ARDB) (8), Antimicrobial Peptide Database (APD3) (9), Collection of Anti-Microbial Peptides (CAMP) (10), Database of Antimicrobial Activity and Structure of Peptides (DBAASP) (11), Antibiotic Resistance Gene-ANNOTation (ARG-ANNOT) (12), BacMet (13) and ResFinder (14). The main underlying differences between these databases are the scope of resistance mechanisms they cover and their reflection of the latest knowledge on the molecular basis of AMR. Unlike a static sequence of a eukaryotic genome, upon which genome annotation and experimental data are curated, antimicrobial pressure favoring new AMR mutations, transfer of AMR genes among pathogens, and emergence of new AMR genes from environmental sources and the proto-resistome (15) create an ever-changing molecular landscape underlying AMR. CARD is thus an actively curated database of molecular sequence reference data for prediction of AMR genotype from genomic data and is focused on comprehensive biocuration of the molecular sequences underlying AMR, including intrinsic resistance, dedicated resistance genes, and acquisition of resistance via mutation of antimicrobial targets and associated elements. ARDB focussed on all antibiotic resistance genes but has not been updated since 2009, ARGO is concerned with resistance to two classes of antibiotics, BacMet specifically collects data on biocide and metal resistance genes, and APD3 concentrates on antimicrobial peptides. Overall, CARD (6) and ARG-ANNOT (12) are currently the two most extensive AMR sequence databases, with ResFinder (14) similarly so for the acquired AMR subset. At the core of CARD is the novel Antibiotic Resistance Ontology (ARO), a controlled vocabulary for describing antimicrobial molecules and their targets, resistance mechanisms, genes and mutations, and their relationships. In addition to the highly developed ARO, CARD includes the Resistance Gene Identifier (RGI) software, which predicts antibiotic resistance genes from genome sequence data, including un-annotated genome sequence assembly contigs. The RGI integrates the ARO, bioinformatics models and molecular reference sequence data to broadly analyze antibiotic resistance at the genome level. With continued biocuration and algorithm development, CARD provides a bioinformatics resource for AMR surveillance in healthcare, agricultural and environmental settings. We here describe an expansion of the CARD reflecting a large increase in curated reference sequences, expanded curation of AMR conferring mutations, and re-organization of the data schema and tools to incorporate explicit AMR gene detection models, including new ontologies and expanded online tools.

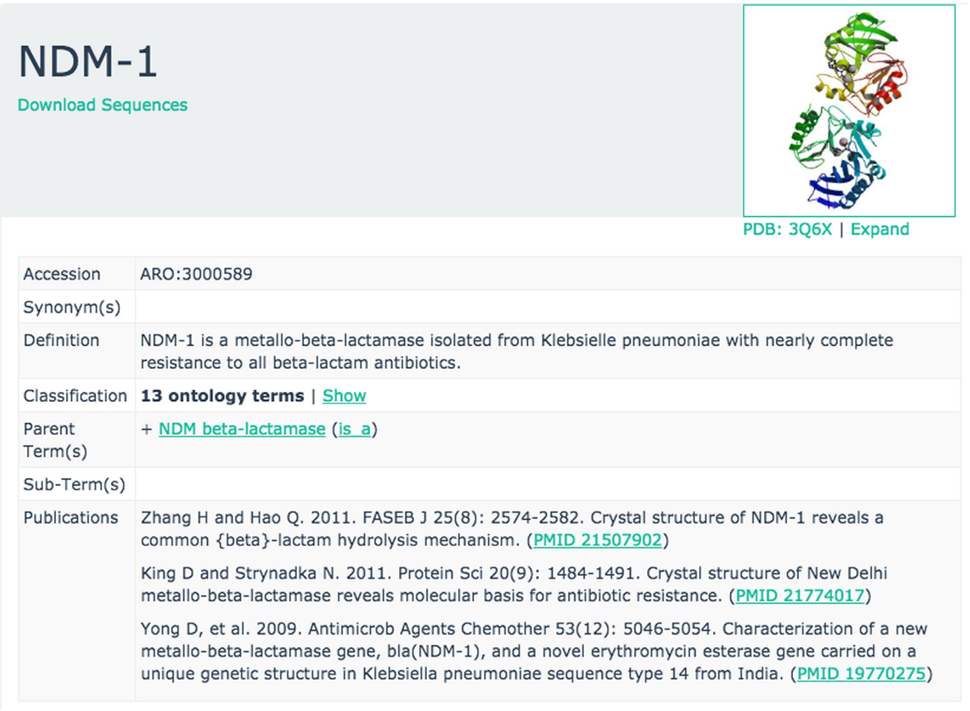
EXPANSION OF CARD

Growth of the Antibiotic Resistance Ontology (ARO)

At its core CARD is an ontology-driven database, which utilizes four central ontologies: the ARO (Antibiotic Resistance Ontology), MO (Model Ontology), RO (a customized subset of the Relationship Ontology) and NCBITaxon (a customized subset of the NCBI Taxonomy Ontology). The ARO contains terms related to antimicrobial resistance genes and mechanisms, as well as antibiotics and their targets; the MO describes AMR gene detection models and parameters for prediction of resistome from genome sequences; the RO defines the relationship types used to connect ontology terms; and finally, the NCBI Taxonomy ontology classifies the bacterial species and strains represented in CARD (mirroring GenBank's taxonomy identifiers). The ARO is organized into six branches giving details on antimicrobial compounds, resistance genes and mutations, drug targets and resistance mechanisms (Table 1), with ontology terms inter-related using a suite of relationship types (Table 2). Each ARO term describing a resistance gene in CARD has ontological relationships to three ARO branches: Determinant of Antibiotic Resistance (ARO:3000000), Antibiotic Molecule (ARO:1000003) and Mechanism of Antibiotic Resistance (ARO:1000002). Strict curation rules define parent–child relationships within the ARO, as each term must be correctly placed within the overall ontological structure. For example, a newly reported resistance gene may belong to a particular family of AMR enzymes, but confers resistance to a different drug. In such cases, the gene family requires multiple ontological branches or an adjustment of relationships to avoid invalid or conflicting resistance information. In addition to expanding the conceptual framework of the ARO to cover the breadth of AMR mechanisms, curators also associate ARO terms with the relevant publications, chemical structures (via curation of PubChem information), and protein structures (via curation of Protein Data Bank information (16)) (Figure 1). Since 2013 (6), the number of ARO terms has grown dramatically: as of August 2016, the ARO contains 3567 ontology terms covering the breadth of AMR mechanisms, supported by 2136 publications.

A new model ontology

A major improvement in CARD since its initial publication is development of AMR detection models and an accompanying new Model Ontology (MO). Detection models are associated with ARO terms such that each ontological concept is paired with information for genome analysis. The MO defines the reference sequences to be used for AMR determinant detection, search parameters, and (where needed) the parameters needed for detection of AMR conferring



NDM-1

Download Sequences

PDB: 3Q6X | Expand

Accession	ARO:3000589
Synonym(s)	
Definition	NDM-1 is a metallo-beta-lactamase isolated from Klebsielle pneumoniae with nearly complete resistance to all beta-lactam antibiotics.
Classification	**13 ontology terms** | Show
Parent Term(s)	+ NDM beta-lactamase (is_a)
Sub-Term(s)	
Publications	Zhang H and Hao Q. 2011. FASEB J 25(8): 2574-2582. Crystal structure of NDM-1 reveals a common {beta}-lactam hydrolysis mechanism. (PMID 21507902) King D and Strynadka N. 2011. Protein Sci 20(9): 1484-1491. Crystal structure of New Delhi metallo-beta-lactamase reveals molecular basis for antibiotic resistance. (PMID 21774017) Yong D, et al. 2009. Antimicrob Agents Chemother 53(12): 5046-5054. Characterization of a new metallo-beta-lactamase gene, bla(NDM-1), and a novel erythromycin esterase gene carried on a unique genetic structure in Klebsiella pneumoniae sequence type 14 from India. (PMID 19770275)

Figure 1. Example of an ARO term (NDM-1 β-lactamase). Each ARO term in CARD incorporates a definition, its parent and sub-terms within the ARO, relevant peer-reviewed publications, and links to protein or chemical structure information (if available).

Table 1. Major branches of the Antibiotic Resistance Ontology (ARO)

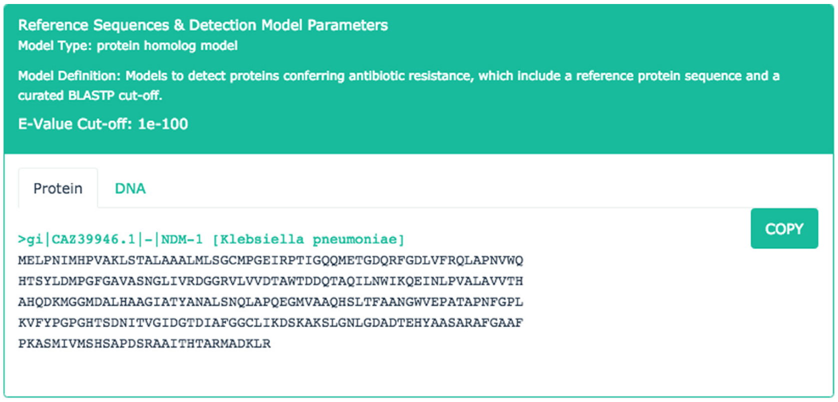
ARO branch	ARO accession	Description
Determinant of Antibiotic Resistance	ARO:3000000	Ontology terms describing genes, mutation, or genomic elements conferring antimicrobial resistance.
Mechanism of Antibiotic Resistance	ARO:1000002	Ontology terms describing mechanisms of antimicrobial resistance.
Antibiotic Target	ARO:3000708	Ontology terms describing targets of antimicrobial compounds.
Antibiotic Molecule	ARO:1000003	Ontology terms describing the chemical diversity of antimicrobial compounds.
Inhibitor of Antibiotic Resistance	ARO:0000076	Ontology terms describing molecules that inhibit antimicrobial resistance.

All major branches are part of 'Process or Component of Antibiotic Biology or Chemistry' (ARO:1000001).

Table 2. Relationship types used in the Antibiotic Resistance Ontology (ARO)

Relationship type	Description
is_a	An axiomatic relationship ontology term in which the subject is placed into a higher order classification.
part_of	A relationship ontology term in which the subject is but part of the object.
derives_from	A relationship ontology term in which the subject has its origins from the object.
regulates	A relationship ontology term in which the subject regulates expression of the object.
confers_resistance_to	A relationship ontology term in which the subject confers antimicrobial resistance to the object.
confers_resistance_to_drug	A relationship ontology term in which the subject (usually a gene) confers clinically relevant antimicrobial resistance to a specific antibiotic.
targeted_by	A relationship ontology term in which the subject is targeted by the object (usually a class of antibiotics).
targeted_by_drug	A relationship ontology term in which the subject is targeted by a specific antibiotic.

mutations. Two detection model types are used for the majority of resistance genes: Protein Homolog and Protein Variant models. Protein Homolog models detect AMR protein sequences based on their similarity to curated AMR reference sequence, using curated BLASTP cut-offs (Figure 2). CARD cut-offs are currently based on BLAST Expect values (E), which are a function of both BLAST database size and query length, but use of the more informative bit score (S') is under development. Protein Variant models perform a similar search, but secondarily screen query sequences for curated sets of AMR-conferring mutations (Figure 3) to differentiate between wild type, antimicrobial sensitive alleles and AMR alleles. To perform these screens, CARD includes extensive catalogues of clinically relevant single nucleotide substitutions, as well as nonsense mutations, insertions, deletions and frameshifts. Resistance vari-



Reference Sequences & Detection Model Parameters

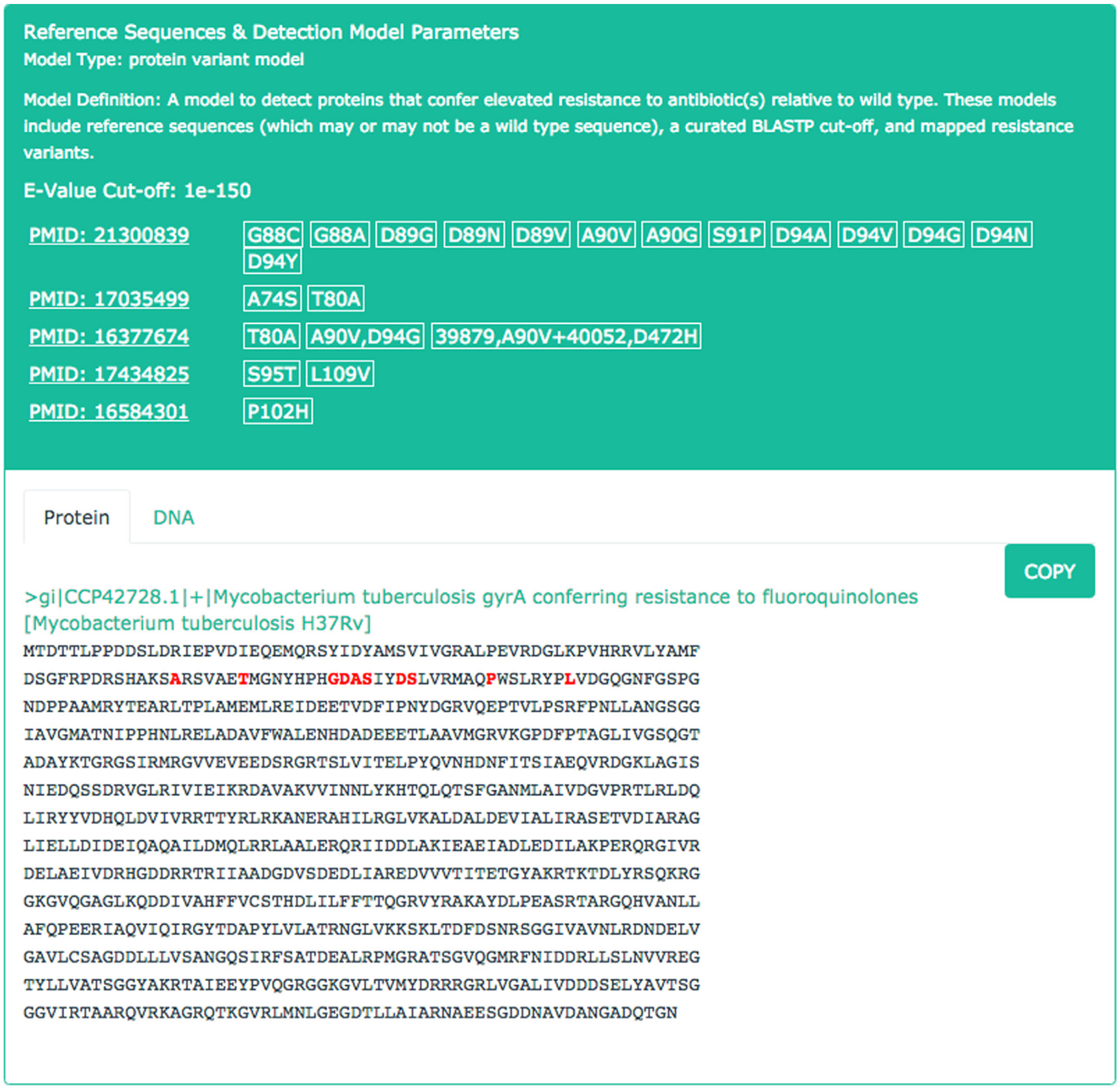
Model Type: protein homolog model

Model Definition: Models to detect proteins conferring antibiotic resistance, which include a reference protein sequence and a curated BLASTP cut-off.

E-Value Cut-off: 1e-100

```
>gi|CAZ39946.1|-|NDM-1 [Klebsiella pneumoniae]
MELPNIMHPVAKLSTALAAALMLSGCMPGEIRPTIGQQMETGDQRFGDLVFRQLAPNVWQ
HTSYLDMPGFGAVASNGLIVRDGGRVLVVDTAWTDDQTAQILNWIKQEINLPVALAVVTH
AHQDKMGGMDALHAAGIATYANALSNQLAPQEGMVAAQHSLTFAANGWVEPATAPNFGPL
KVFYPGPGHTSDNITVGIDGTDIAFGGCLIKDSKAKSLGNLGDADTEHYAASARAFGAAF
PKASMIVMSHSAPDSRAAITHTARMADKLR
```

Figure 2. The Protein Homolog detection model for NDM-1 β-lactamase, involving a curated reference sequence and BLASTP cut-off to limit hits to functional homologs.



Reference Sequences & Detection Model Parameters

Model Type: protein variant model

Model Definition: A model to detect proteins that confer elevated resistance to antibiotic(s) relative to wild type. These models include reference sequences (which may or may not be a wild type sequence), a curated BLASTP cut-off, and mapped resistance variants.

E-Value Cut-off: 1e-150

PMID: 21300839	G88C G88A D89G D89N D89V A90V A90G S91P D94A D94V D94G D94N D94Y
PMID: 17035499	A74S T80A
PMID: 16377674	T80A A90V,D94G 39879,A90V+40052,D472H
PMID: 17434825	S95T L109V
PMID: 16584301	P102H

```
>gi|CCP42728.1|+|Mycobacterium tuberculosis gyrA conferring resistance to fluoroquinolones
[Mycobacterium tuberculosis H37Rv]
MTDTTLPPDDSLDRIEPVDIEQEMQRSYIDYAMSVIVGRALPEVRDGLKPVHRRVLYAMF
DSGFRPDRSHAKSARSVAETMGNYHPHGDASIYDSLVRMAQPWSLRYPLVDGQGNFGSPG
NDPPAAMRYTEARLTPLAMEMLREIDEETVDFIPNYDGRVQEPTVLPSRFPNLLANGSGG
IAVGMATNIPPHNLRELADAVFWALENHDADEEETLAAVMGRVKGPDFPTAGLIVGSQGT
ADAYKTGRGSIRMRGVVEVEEDSRGRTSLVITELPYQVNHDNFITSIAEQVRDGKLAGIS
NIEDQSSDRVGLRIVIEIKRDAVAKVVINNLYKHTQLQTSFGANMLAIVDGVPRTLRLDQ
LIRYYVDHQLDVIVRRTTYRLRKANERAHILRGLVKALDALDEVIALIRASETVDIARAG
LIELLDIDEIQAQAILDMQLRRLAALERQRIIDDLAKIEAEIADLEDILAKPERQRGIVR
DELAEIVDRHGDDRRTRIIAADGDVSDEDLIAREDVVVTITETGYAKRTKTDLYRSQKRG
GKGVQGAGLKQDDIVAHFFVCSTHDLILFFTTQGRVYRAKAYDLPEASRTARGQHVANLL
AFQPEERIAQVIQIRGYTDAPYLVLATRNGLVKKSKLTDFDSNRSGGIVAVNLRDNDELV
GAVLCSAGDDLLLVSANGQSIRFSATDEALRPMGRATSGVQGMRFNIDDRLLSLNVVREG
TYLLVATSGGYAKRTAIEEYPVQGRGGKGVLTVMYDRRRGRLVGALIVDDDSELYAVTSG
GGVIRTAARQVRKAGRQTKGVRLMNLGEGDTLLAIARNAEESGDDNAVDANGADQTGN
```

Figure 3. The Protein Variant detection model for *Mycobacterium tuberculosis* gyrA conferring resistance to fluoroquinolones, involving a curated reference sequence, BLASTP cut-off, and a catalogue of positional single nucleotide polymorphisms (SNPs) that confer resistance to fluoroquinolones. Locations of resistance SNPs are highlighted red in the reference sequence.

ants are manually curated and each variant is associated with a publication in order to streamline data validation and interpretation. Unlike Protein Homolog models, Protein Variant models are differentiated by bacterial species within the ARO, as mutations are often pathogen-specific.

As of August 2016, CARD contains 2260 AMR detection models. Of these, 2102 are Protein Homolog models and 92 are Protein Variant models. Currently only these two model types are supported by CARD's Resistance Gene Identifier (RGI) software, but curators are already designing new model types for use by future software improvements: 47 rRNA mutation models, 11 gene order models and 2 protein absence models. Overall, CARD's detection models include 2441 model reference sequences, 853 single nucleotide substitutions, plus a growing number of indels, frameshift, and nonsense mutations associated with antimicrobial resistance. These models and parameters have been developed from over 2000 publications.

Improved curation

To be included in CARD, an AMR gene or gene variant must confer resistance to a known antimicrobial compound either *in vivo* or *in vitro*, evidence of which must be described in a peer-reviewed scientific publication, and its sequence must be available in NCBI's GenBank repository. As such, only published, experimentally verified AMR genes and mutations, with subsequent submission of sequence to GenBank, are curated into the CARD. CARD is maintained and updated through manual biocuration of the scientific literature, with experts in the field ensuring accuracy of the data and validating ontological concepts. Biocuration is performed regularly with strict guidelines to maintain the integrity of the ontology and consistency of the information provided. These efforts typically involve either adding new ontology terms and detection models or updating pre-existing data to reflect recent advancements in the field. Biocuration includes manual tracking of the published AMR literature as well as drug class or mechanism targeted literature reviews. Yet, the AMR literature is fast moving, so CARD now additionally uses custom 'CARD*Shark' text-mining algorithms to prioritize the scientific literature for biocuration on a monthly basis. CARD*Shark constructs word-association scoring matrices based on the title and abstract of literature associated with the ARO, which are then used to priority score and assign to drug class publications newly available in PubMed. High scoring publications are manually reviewed by the CARD biocuration team, including review of the papers cited within, with all steps in the curation process tracked using an internal Gitlab repository. AMR detection models are further assessed by routine analysis of pathogen genome, plasmid, and whole-genome shotgun assembly sequences available in GenBank to assess false positive and negative rates.

Expanded tools

Outside of online tools to browse AMR ontology terms and detection models, two bioinformatic tools are available for use within CARD. The first is a standard BLAST, allowing searches of the CARD reference sequences using the family of BLAST algorithms (17). The second tool is the RGI, which uses CARD's curated AMR detection models to predict complete resistome from genome sequences, including simultaneous detection of both dedicated resistance genes, such as β-lactamases, and mutations conferring AMR in otherwise sensitive drug targets. The CARD website also includes a new RGI visualization tool that organizes resistome predictions by detected genes, ARO functional categories, known versus emergent threat, and quality of prediction (Figure 4). The RGI provides a preliminary annotation of DNA or protein sequences, based upon the data available in CARD. The RGI currently supports two detection model types (Protein Homolog and Protein Variant) and analyzes sequences under three paradigms—Perfect, Strict, and Loose (a.k.a. Discovery). The Perfect algorithm is most often applied to clinical surveillance as it detects perfect matches to the curated reference sequences and mutations in CARD. In contrast, the Strict algorithm detects previously unknown variants of known AMR genes, including secondary screen for key mutations, using detection models with curated similarity cut-offs to ensure the detected variant is likely a functional AMR gene. The Loose algorithm works outside of the detection model cut-offs to provide detection of new, emergent threats and more distant homologs of AMR genes, but will also catalog homologous sequences and spurious partial hits that may not have a role in AMR. Combined with phenotypic screening, the Loose algorithm allows researchers to hone in on new AMR genes. It is important to note that, as CARD curation evolves, the results of the RGI evolve. A detailed publication describing the RGI algorithms is in preparation.

Schema and information technology

The original version of CARD (6) used the Generic Model Organism Database's Chado data schema (18) for data storage and curation, but this has now been replaced with the custom Broad Street relational schema, named for the 1854 Broad Street cholera outbreak and pioneering epidemiological efforts of Dr John Snow (19). Derived from Chado, Broad Street was introduced to be simple, lightweight, and to focus on ontologies and detection models. It contains five modules: controlled vocabulary, detection model, publication, external reference, and administrative. The schema and data are managed using PostgreSQL 9.4 and the CARD website and curator tools designed using the Laravel 5.2 PHP framework, PHP 7, Apache and PostgreSQL 9.4, with additional statistics generated using PHP and Python. The website, software, data and resolution of all data curation issues are all version controlled using GitLab (http://gitlab.com/).

Updates and availability

CARD is divided into two branches, Development and Production, with the Production branch available to the public at the CARD website and updated monthly from the constantly curated Development branch. Before every monthly update of the Production branch, quality control scripts are used to validate the data in the Development branch, including external identifiers, citations, AMR gene detection model data and parameters, and a suite of rules underly-

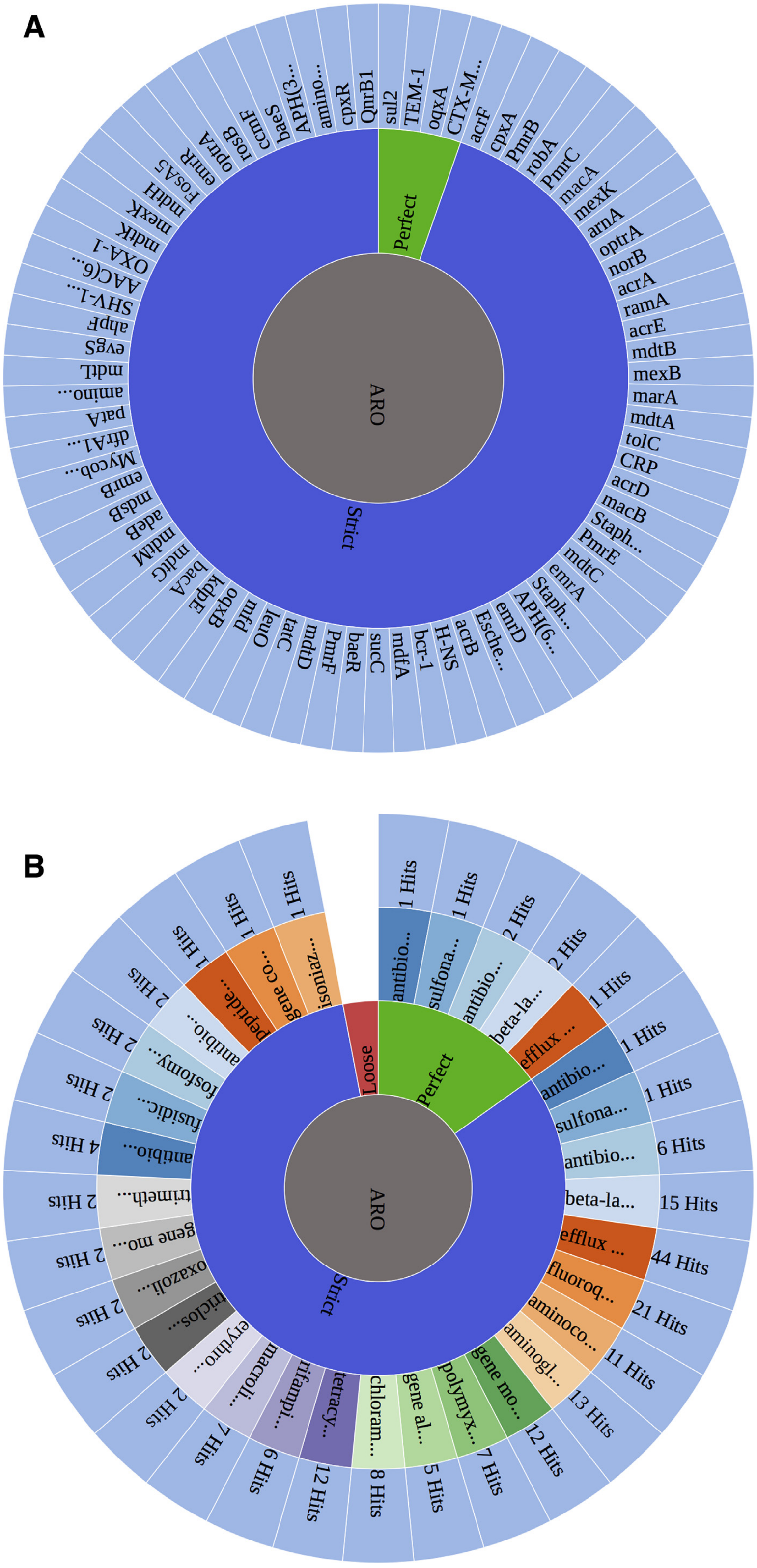


Figure 4. Resistance Gene Identifier visualizations available at the CARD website, based on analysis of a recent clinical MDR *Klebsiella pneumoniae* isolate. (**A**) Individual AMR genes detected, based on 'Perfect' (green) matches or 'Strict' (blue) hits to CARD reference amino acid sequences, including secondary screening for AMR-conferring mutations where appropriate. Strict hits are defined as being within the similarity cut-offs of the individual AMR detection models and represent likely (but not tested) homologs of AMR genes. (**B**) The same results organized by Antibiotic Resistance Ontology functional categories, reflecting both drug classes and resistance mechanisms. Hits with weak similarity (i.e. 'Loose') are not shown. Images provided by the CARD web interface—full details are available by mouse hover as well as by clicking on terms.

ing the structure of the ARO. The public CARD web interface can be found online at http://arpcard.mcmaster.ca, providing tools for browsing and searching the Antibiotic Resistance Ontology knowledgebase, definition of AMR detection model types within the Model Ontology, and sequences, mutation information and parameters for individual AMR detection models. This includes tools for tracking changes in CARD curation between releases. The CARD website also contains an online version of the Resistance Gene Identifier (RGI), including visualization tools for online or offline RGI results. The data and ontologies curated into the CARD are available for download at the CARD website in a variety of file formats, e.g. OBO, tab-delimited, FASTA, and JSON. The website also includes a downloadable, stand-alone version of the RGI, plus CARD AMR detection model data and the command-line version of the RGI are additionally available within the Bioconda project, https://bioconda.github.io, for use within the Conda open source package management system. CARD users are encouraged to join the CARD-L mailing list or follow the CARD twitter feed to receive information on monthly updates and software releases, see http://arpcard.mcmaster.ca/about. The CARD curators and developers are available for contact at card@mcmaster.ca.

CONCLUSIONS

CARD aims to address the molecular basis of AMR by creating an ontology- and model-based framework for curation and detection of known resistance genes, variants existing along the environment—agriculture—clinical axis, and newly emergent threats. Since its initial release (6), CARD has undergone dramatic changes to improve both depth of data and functionality. These changes mainly stem from the development of the new Broad Street schema, coupled with an improved online user interface. Significant improvements in the quality of curated resistance information in the form of detection models, reference sequences, and mutations have allowed for the integration and expanded use of analysis software (including RGI and BLAST). Through the new website, users are now able to access a large amount of data from the ARO, MO and RGI in various downloadable formats, with transparent version control and tracking information. Overall, the increased flexibility provided by the Broad Street schema and incorporation of detection models through the MO has enabled CARD to become a more functional resource for both data curation and screening, allowing it to keep pace with the rapidly evolving AMR crisis.

ACKNOWLEDGEMENTS

The CARD developers wish to thank their user community and software/website beta-testers for helpful feedback, with particular thanks to the trainees of the M.G. DeGroote Institute for Infectious Disease Research at McMaster University, Drs Daniel Haft, Bill Klimke, and Michael Feldgarden of the National Center for Biotechnology Information (NCBI), Dr Gary Van Domselaar, Franklin Bristow, and Aaron Petkau of the Public Health Agency of Canada (PHAC) and members of the Integrated Rapid Infectious Disease Analysis (IRIDA) and Genomic Epidemiology Ontology (GenEpiO) projects. Dr Torsten Seemann and Simon Gladman (University of Melbourne) and Dr Björn Grüning (University of Freiburg), provided assistance with Bioconda deployment.

FUNDING

Canadian Institutes of Health Research; Natural Sciences and Engineering Research Council; Canada Research Chair (to G.D.W.'s laboratory); A.G.M. holds a Cisco Research Chair in Bioinformatics, supported by Cisco Systems Canada, Inc., and his AMR research is funded by the Canadian Foundation for Innovation, Genome Canada, Canadian Institutes of Health Research; Natural Sciences and Engineering Research Council (Canada); McMaster University Faculty of Science Interdisciplinary Research Fund; M.G. DeGroote Institute for Infectious Disease Research Summer Student Fellowship (to B.A.L.); Genome Canada/Genome BC (to F.S.L.B., R.L., M.C.); McMaster Service Lab and Repository (MSLR) computing cluster, funded in part by a grant from the Canadian Foundation for Innovation [34531 to A.G.M.]; USDA Project Numbers 6040-32000-006-00 and 6040-32000-009-00 (to J.G.F.); Postdoctoral Research Associate Fellowship awarded by the USDA, Agricultural Research Service Headquarters (to L.E.W.). Funding for open access charge: internal funds.

Conflict of interest statement. None declared.

REFERENCES

1. ONeill,J. (2014) Antimicrobial resistance: tackling a crisis for the health and wealth of nations. Review on Antimicrobial Resistance, London, United Kingdom.
2. Brown,E.D. and Wright,G.D. (2016) Antibacterial drug discovery in the resistance era. *Nature*, **529**, 336–343.
3. McArthur,A.G. and Wright,G.D. (2015) Bioinformatics of antimicrobial resistance in the age of molecular epidemiology. *Curr. Opin. Microbiol.*, **27**, 45–50.
4. Scholz,M.B., Lo,C.-C. and Chain,P.S.G. (2012) Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis. *Curr. Opin. Biotechnol.*, **23**, 9–15.
5. Burge,S., Attwood,T.K., Bateman,A., Berardini,T.Z., Cherry,M., O'Donovan,C., Xenarios,L. and Gaudet,P. (2012) Biocurators and Biocuration: surveying the 21st century challenges. *Database*, **2012**, bar059.
6. McArthur,A.G., Waglechner,N., Nizam,F., Yan,A., Azad,M.A., Baylay,A.J., Bhullar,K., Canova,M.J., de Pascale,G., Ejim,L. *et al.* (2013) The comprehensive antibiotic resistance database. *Antimicrob. Agents Chemother.*, **57**, 3348–3357.
7. 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.
8. Liu,B. and Pop,M. (2009) ARDB–Antibiotic Resistance Genes Database. *Nucleic Acids Res.*, **37**, D443–D447.
9. Wang,G., Li,X. and Wang,Z. (2016) APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Res.*, **44**, D1087–D1093.
10. Waghu,F.H., Barai,R.S., Gurung,P. and Idicula-Thomas,S. (2015) CAMPR3: a database on sequences, structures and signatures or antimicrobial peptides. *Nucleic Acids Res.*, **44**, D1094–D1097.
11. Pirtskhalava,M., Gabrielian,A., Cruz,P., Griggs,H.L., Squires,R.B., Hurt,D.E., Grigolava,M., Chubinidze,M., Gogoladze,G., Vishnepolsky,B. *et al.* (2016) DBAASP v.2: an enhanced database of structure and antimicrobial/cytotoxic activity of natural and synthetic peptides. *Nucleic Acids Res.*, **44**, D1104–D1112.
12. Gupta,S.K., Padmanabhan,B.R., Diene,S.M., Lopez-Rojas,R., Kempf,M., Landraud,L. and Rolain,J.-M. (2014) ARG-ANNOT, a

new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob. Agents Chemother.*, **58**, 212–220.

13. Pal,C., Bengtsson-Palme,J., Rensing,C., Kristiansson,E. and Larsson,D.G.J. (2014) BacMet: antibacterial biocide and metal resistance genes database. *Nucleic Acids Res.*, **42**, D737–D743.
14. Zankari,E., Hasman,H., Cosentino,S., Vestergaard,M., Rasmussen,S., Lund,O., Aarestrup,F.M. and Larsen,M.V. (2012) Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.*, **67**, 2640–2644.
15. Wright,G.D. (2010) The antibiotic resistome. *Expert Opin. Drug Discov.*, **5**, 779–788.
16. Rose,P.W., Prlić,A., Bi,C., Bluhm,W.F., Christie,C.H., Dutta,S., Green,R.K., Goodsell,D.S., Westbrook,J.D., Woo,J. *et al.* (2014) The RCSB protein data bank: views of structural biology for basic and applied research and education. *Nucleic Acids Res.*, **43**, D345–D356.
17. Camacho,C., Coulouris,G., Avagyan,V., Ma,N., Papadopoulos,J., Bealer,K. and Madden,T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421.
18. Mungall,C.J., Emmert,D.B. and FlyBase Consortium (2007) A chado case study: an ontology-based modular schema for representing genome-associated biological information. *Bioinformatics*, **23**, i337–i346.
19. Newsom,S.W.B. (2006) Pioneers in infection control: John Snow, Henry Whitehead, the Broad Street pump, and the beginnings of geographical epidemiology. *J. Hosp. Infect.*, **64**, 210–216.