

An Introduction to CARD & Biocuration

Brian Alcock

May 25, 2022

A Bit About Me...



- Born and raised in Newfoundland
- McMaster M.Sc. (Biology) - 2013-2015
- McArthur Lab - December 2015

A Bit About Me...



- Born and raised in Newfoundland
- McMaster M.Sc. (Biology) - 2013-2015
- McArthur Lab - December 2015
- *No background in CompSci prior to my MSc!*

What We Do



- **The Comprehensive Antibiotic Resistance Database**
- Antimicrobial surveillance and stewardship
- Biocuration and software development
- Data standardization and harmonization

What We Do



Brian: Curation, Ontologies, AMR detection models

Amos: Software Engineering, RGI, AMR prediction

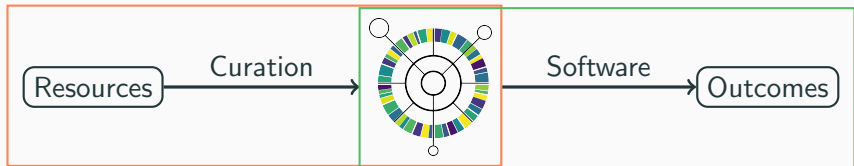
What We Do



Brian: Curation, Ontologies, AMR detection models

Amos: Software Engineering, RGI, AMR prediction

What We Do



Brian: Curation, Ontologies, AMR detection models

Amos: Software Engineering, RGI, AMR prediction

What is (Bio-)curation?

Curation is the collection and organization of items (data, specimens, collectibles, etc.) to make them more readily accessible and widely available. I like to think of curation as part science and part art.

What is (Bio-)curation?

Curation is the collection and organization of items (data, specimens, collectibles, etc.) to make them more readily accessible and widely available. I like to think of curation as part science and part art.

- **Science:** What is included and why
- **Art:** How you represent it to the world

What is (Bio-)curation?

Curation is the collection and organization of items (data, specimens, collectibles, etc.) to make them more readily accessible and widely available. I like to think of curation as part science and part art.

- **Science:** What is included and why
- **Art:** How you represent it to the world

Everyday examples: museums; radio stations; Twitter; TikTok

What Gets Included?

CARD's Golden Rule: to be included in CARD an AMR determinant must be described in a peer-reviewed scientific publication, with its DNA sequence available in GenBank, including clear experimental evidence of elevated minimum inhibitory concentration (MIC) over controls. (CARD 2020)

(Over time we've added some exceptions but let's not worry about those right now...)

How CARD is Structured

- CARD is built around *ontologies* — think of these like a Word Bank where each term is connected — mainly the *Antibiotic Resistance Ontology* (ARO)

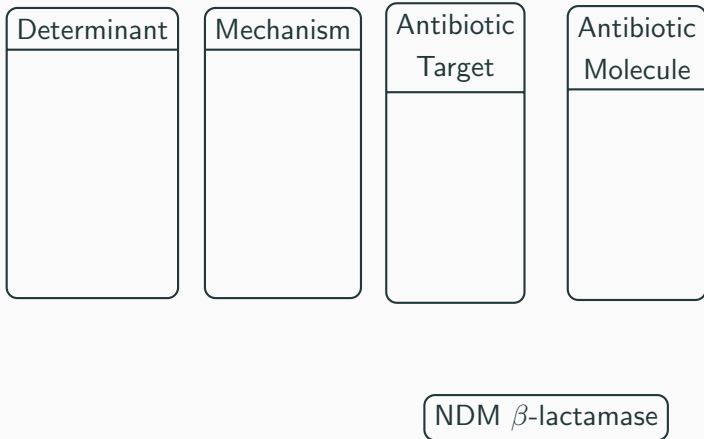
How CARD is Structured

- CARD is built around *ontologies* — think of these like a Word Bank where each term is connected — mainly the *Antibiotic Resistance Ontology* (ARO)

NDM β -lactamase

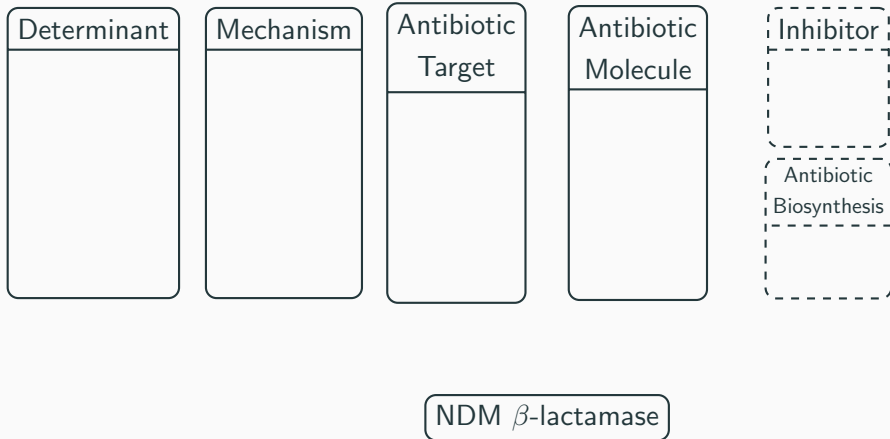
How CARD is Structured

- CARD is built around *ontologies* — think of these like a Word Bank where each term is connected — mainly the *Antibiotic Resistance Ontology* (ARO)



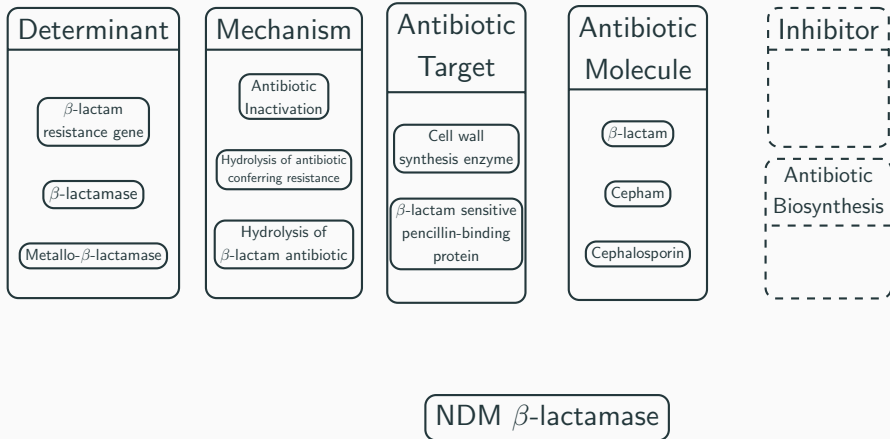
How CARD is Structured

- CARD is built around *ontologies* — think of these like a Word Bank where each term is connected — mainly the *Antibiotic Resistance Ontology* (ARO)



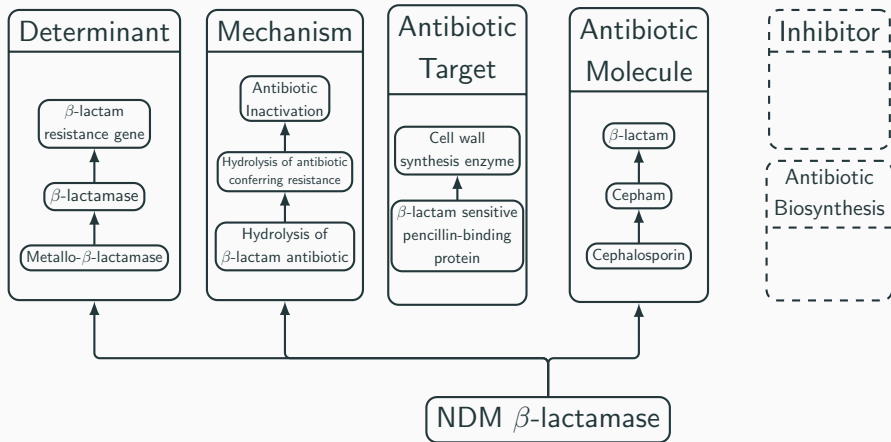
How CARD is Structured

- CARD is built around *ontologies* — think of these like a Word Bank where each term is connected — mainly the *Antibiotic Resistance Ontology* (ARO)



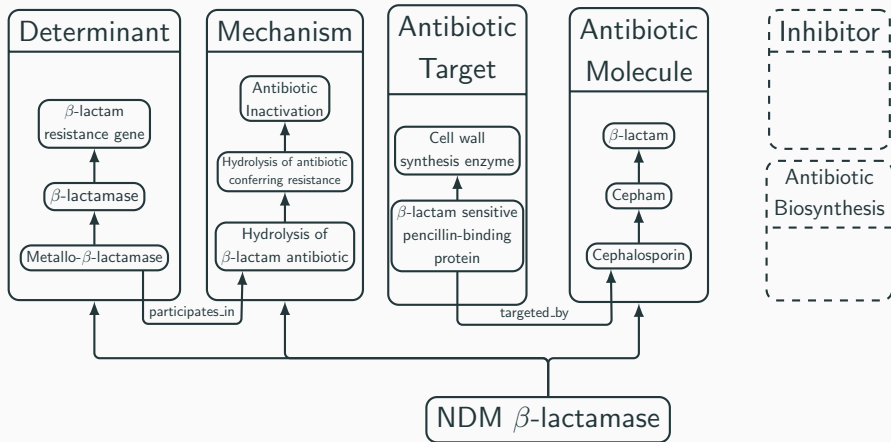
How CARD is Structured

- CARD is built around *ontologies* — think of these like a Word Bank where each term is connected — mainly the *Antibiotic Resistance Ontology* (ARO)



How CARD is Structured

- CARD is built around *ontologies* — think of these like a Word Bank where each term is connected — mainly the *Antibiotic Resistance Ontology* (ARO)



NDM on CARD

NDM beta-lactamase [AMR Gene Family]

[Download Sequences](#)

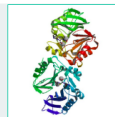
Accession	ARO:3000057
Definition	NDM beta-lactamases or New Delhi metallo-beta-lactamases are class B beta-lactamases that confer resistance to a broad range of antibiotics including carbapenems, cephalosporins and penicillins.
Drug Class	carbapenem , cephalosporin , cephamycin , penam
Resistance Mechanism	antibiotic inactivation
Classification	13 ontology terms Hide <ul style="list-style-type: none">+ process or component of antibiotic biology or chemistry+ antibiotic molecule+ mechanism of antibiotic resistance+ antibiotic inactivation [Resistance Mechanism]+ beta-lactam antibiotic+ determinant of antibiotic resistance+ determinant of beta-lactam resistance+ hydrolysis of antibiotic conferring resistance+ antibiotic inactivation enzyme+ hydrolysis of beta-lactam antibiotic by metallo-beta-lactamase+ beta-lactamase+ cephem+ class B (metallo-) beta-lactamase
Parent Term(s)	5 ontology terms Hide <ul style="list-style-type: none">+ confers_resistance_to carbapenem [Drug Class]+ confers_resistance_to cephalosporin [Drug Class]+ confers_resistance_to cephamycin [Drug Class]+ confers_resistance_to penam [Drug Class]+ subclass B1 (metallo-) beta-lactamase
Sub-Term(s)	17 ontology terms Hide <ul style="list-style-type: none">+ NDM-1+ NDM-2

Adding a New Resistance Gene

One incredibly useful aspect of ontologies is we do not need to record redundant information for new genes:

NDM-1

[Download Sequences](#)



Accession	ARO:3000589
Definition	NDM-1 is a metallo-beta-lactamase isolated from <i>Klebsiella pneumoniae</i> with nearly complete resistance to all beta-lactam antibiotics.
AMR Gene Family	NDM beta-lactamase
Drug Class	penam , carbapenem , cephamycin , cephalosporin
Resistance Mechanism	antibiotic inactivation
Resistomes	<i>Acinetobacter baumannii</i> ^{P1-1595} , <i>Acinetobacter nosocomialis</i> ²⁵⁹⁵ , <i>Citrobacter freundii</i> ^{P1-1595} , <i>Enterobacter asburiae</i> ²⁵⁹⁵ , <i>Enterobacter cloacae</i> ^{P1-1595} , <i>Enterobacter hormaechei</i> ^{P1-1595} , <i>Enterobacter kobei</i> ²⁵⁹⁵ , <i>Escherichia coli</i> ^{P1-1595} , <i>Klebsiella oxytoca</i> ²⁵⁹⁵ , <i>Morganella morganii</i> ^{P1-1595} , <i>Proteus mirabilis</i> ^{P1-1595} , <i>Providencia rettgeri</i> ^{P1-1595} , <i>Providencia stuartii</i> ^{P1-1595} , <i>Serratia marcescens</i> ^{P1-1595} , <i>Shigella sonnei</i> ²⁵⁹⁵ , <i>Vibrio cholerae</i> ²⁵⁹⁵ , <i>Vibrio parahaemolyticus</i> ²⁵⁹⁵
Classification	23 ontology terms Show
Parent Term(s)	5 ontology terms Hide + <i>confers_resistance_to_drug</i> ertapenem [Antibiotic] + <i>confers_resistance_to_drug</i> meropenem [Antibiotic] + NDM beta-lactamase [AMR Gene Family] + <i>confers_resistance_to_drug</i> imipenem [Antibiotic] + <i>confers_resistance_to_drug</i> amoxicillin-clavulanic acid
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)

AMR Detection Models

We append “resistance genes” with genetic sequence, mutation and bit-score information to create an AMR Detection Model — a bioinformatic tool to detect resistance genes from genomes:

Detection Models

protein homolog model

Model Type: protein homolog model

Model Definition: The protein homolog model is an AMR detection model. Protein homolog models detect a protein sequence based on its similarity to a curated reference sequence. A protein homolog model has only one parameter: a curated BLASTP bitscore cutoff for determining the strength of a match. Protein homolog model matches to reference sequences are categorized on three criteria: perfect, strict and loose. A perfect match is 100% identical to the reference sequence along its entire length; a strict match is not identical but the bitscore of the matched sequence is greater than the curated BLASTP bitscore cutoff. Loose matches are other sequences with a match bitscore less than the curated BLASTP bitscore.

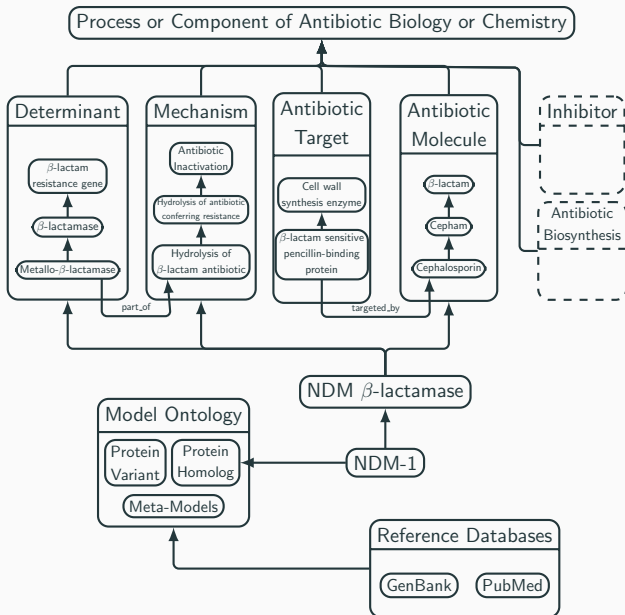
Bit-score Cut-off (blastP): 500

Protein ☒ DNA

>gb|CAZ39946.1|-|NDM-1 [Klebsiella pneumoniae]
MELPFNIMHPVAKLSTALAAALMLSGCMFGEIRPTIGQQMETGQRFGLVFRQLAPNVWQHTSYLDMPGFGAVASNLIVRDGGRVLVVD
TAWTDQQTALILNWIQENLNPVALAVVTHAHQDKMGHDALHAAGIATYANALSNQLAPQEGHVAAQHSLLTFAANGWVEFATAPNFGPL
KVFPYGPHTSDNITVGIDGTDIAFGGCLIKDSKAKSLGNLGDADTEHYAASARAFGAAPFKASMIVMSHSAFDSKRAITHTARMADKLK

COPY

Putting it Together



Core takeaways from this brief introduction:

1. CARD is a **platform and bioresource** for amalgamating AMR data from peer-reviewed sources

Takeaways

Core takeaways from this brief introduction:

1. CARD is a **platform and bioresource** for amalgamating AMR data from peer-reviewed sources
2. ARO is critical for providing a **controlled vocabulary** - universally understood AMR language

Takeaways

Core takeaways from this brief introduction:

1. CARD is a **platform and bioresource** for amalgamating AMR data from peer-reviewed sources
2. ARO is critical for providing a **controlled vocabulary** - universally understood AMR language
3. Biocuration is both **the collection and presentation** of these data for improved user experience

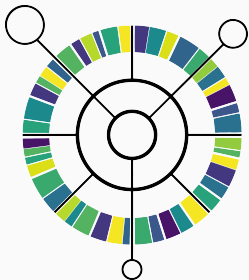
Takeaways

Core takeaways from this brief introduction:

1. CARD is a **platform and bioresource** for amalgamating AMR data from peer-reviewed sources
2. ARO is critical for providing a **controlled vocabulary** - universally understood AMR language
3. Biocuration is both **the collection and presentation** of these data for improved user experience
4. Our software uses these data structures and **AMR detection models** for computational AMR prediction

Thanks for listening!

(Questions are welcome!)



Michael G. DeGroote
INSTITUTE FOR INFECTIOUS DISEASE RESEARCH