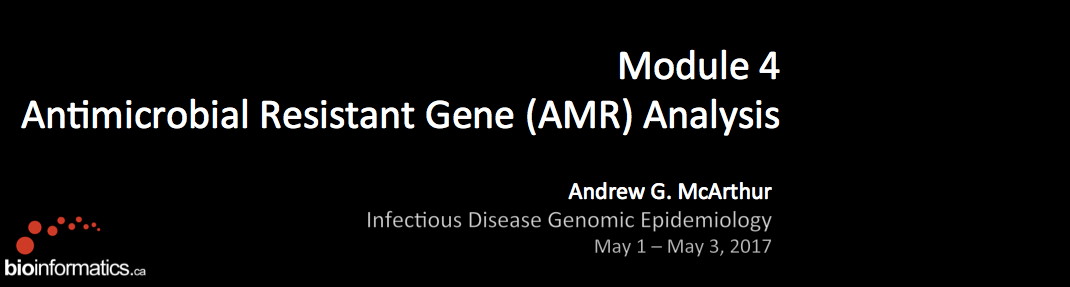
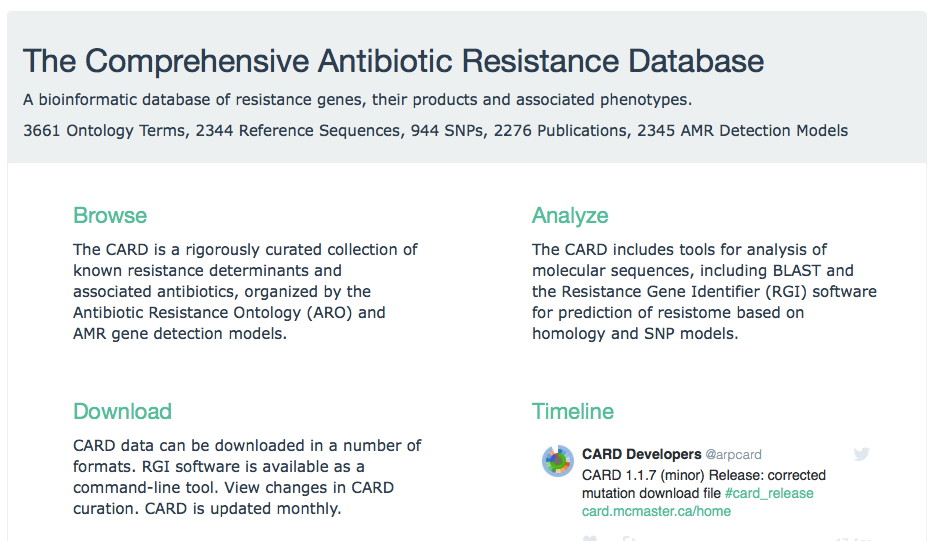
****

**Introduction**

This module gives an introduction to prediction of antimicrobial resistome and phenotype based on comparison of genomic or metagenomic DNA sequencing data to reference sequence information. While there is a large diversity of reference databases and software, the tutorial is focused on the Comprehensive Antibiotic Resistance Database (CARD) for genomic analysis and MEGARs for metagenomic analysis to illustrate the core principles of AMR prediction.

**From Analysis to Interpretation**

The relationship between AMR genotype and AMR phenotype is complicated and no tools for complete prediction of phenotype from genotype exist. Instead, analyses focus on prediction or catalog of the AMR resistome – the collection of AMR genes and mutants in the sequenced sample. While BLAST and other sequence similarity tools can be used to catalog the resistance determinants in a sample via comparison to a reference sequence database, interpretation and phenotypic prediction are often the largest challenge. To start the tutorial, we will use the Comprehensive Antibiotic Resistance Database (http://card.mcmaster.ca) to examine the diversity of resistance mechanisms, how they influence bioinformatics analysis approaches, and how CARD’s Antibiotic Resistance Ontology (ARO) can provide an organizing principle for interpretation of bioinformatics results.



CARD’s website provides the ability to:

1. Browse the Antibiotic Resistance Ontology (ARO) knowledgebase.
2. Browse the underlying AMR detection models, reference sequences, and SNP matrices.
3. Download the ARO, reference sequence data, and indices in a number of formats for custom analyses.
4. Performed integrated genome analysis using the Resistance Gene Identifier (RGI).

In this part of the tutorial, your instructor will walk you through the following use of CARD:

1. Examine the mechanisms of resistance as described by the Antibiotic Resistance Ontology.
2. Examine the NDM-1 beta-lactamase protein, it’s mechanism of action, conferred antibiotic resistance, and it’s detection model. [BLASTP of NDM-1 against CARD]
3. Examine the AAC(6')-If aminoglycoside acetyltransferase, it’s mechanism of action, conferred antibiotic resistance, and it’s detection model. [BLASTP of AAC(6')-If against CARD]
4. Examine the recently reported colistin resistance MCR-1 protein, it’s mechanism of action, conferred antibiotic resistance, and it’s detection model. [BLASTP of MCR-1 against CARD]
5. Examine the fluoroquinolone resistant gyrB, it’s mechanism of action, conferred antibiotic resistance, and it’s detection model. [Why would BLASTP be inappropriate for this resistance determinant?]
6. Examine the glycopeptide resistance gene cluster VanA, it’s mechanism of action, conferred antibiotic resistance, and it’s detection model(s). [Why would BLASTP be inappropriate for this resistance determinant?]
7. Examine the MexAB-OprM efflux complex, it’s mechanism of action, conferred antibiotic resistance, and it’s detection model(s). [Why would BLASTP be inappropriate for this resistance determinant?]

**Genome Analysis**

As illustrated by the exercise above, the diversity of antimicrobial resistance mechanisms requires a diversity of detection algorithms and a diversity of detection limits. CARD’s Resistance Gene Identifier (RGI) currently integrates two CARD detection models: Protein Homolog Model and Protein Variant Model. Unlike naïve analyses, CARD detection models use curated cut-offs, currently based on BLAST expectation values, but migrating in Summer 2017 to BLAST bitscore cut-offs. Many other available tools are based on BLASTN or BLASTP without defined cut-offs and avoid resistance by mutation entirely.

In this part of the tutorial, your instructor will walk you through the following use of CARD:

1. Resistome prediction for the multidrug resistant *Acinetobacter baumannii MDR*-TJ, complete genome [NC\_017847]
2. Resistome prediction for the plasmid isolated from Escherichia coli strain MRSN388634 plasmid [KX276657]
3. Explain the difference in triclosan resistance between two clinical strains of *Pseudomonas aeruginosa* that appear clonal based on identical MLST [Pseudomonas1.fasta, Pseudomonas2.fasta]

Preview: Some resistance mechanisms involve a number of determinants, particularly for glycopeptide resistance and efflux. To accurately predict phenotype based on DNA sequencing, higher-order meta-models are required. CARD will be introducing glycopeptide resistance cluster meta-models in Summer 2017, view the supplied HTML output to preview meta-model level analysis provided by CARD’s draft Glycopeptide Resistance Predictor (GRP): http://agmfilehosting.ca/grp

**AMR Sequence Diversity “In the Wild”**

CARD and other AMR reference databases include sequences and mutations from the published literature with clear experimental evidence of elevated minimum inhibitory concentration (MIC). This has implications for molecular surveillance as sequences in clinical, agricultural, or environmental strains may differ in sequence from characterized & curated reference sequences:

1. To what extent to differences between reference and isolate reflect divergence in function / phenotype? How different is too different? Often variation is a continuum between AMR proteins and homologous sequences interacting with non-antimicrobial small molecules or targets. CARD’s cut-offs are designed to avoid false positives, but detected variants may still require expression and MIC testing. **What is your surveillance goal – how much variation do you want to tolerate?**
2. Algorithms for metagenomic read mapping may require high nucleotide similarity between reference and isolate, biasing analyses towards characterized AMR sequences but possibly missing functional variants. **Do Burrow-Wheeler Transform or other read mapping metagenomic analyses to detect AMR determinants suffer from Type II error?**

The CARD team has used their Wild\*CARD algorithms to assess sequence diversity for the ESKAPE pathogens based on the genome, plasmid, and WGS sequences available in GenBank. Examine the two spreadsheets to see the relative abundance of PERFECT and STRICT RGI hits in the available sequences in GenBank. Hint: Compare the total number of determinants predicted for each pathogen under each criteria, and then also compare the most prevalent determinants for each pathogen under each criteria. **What are you conclusions for genomic or metagenomic surveillance of AMR? What criteria are you going to use in your work?**

**Analysis of Metagenomic Data**

The most common tools for metagenomic data annotation are based on high-stringency read mapping, such as the Burrows-Wheeler Transform. Available methods almost exclusively focus on acquired resistance genes, not those involving resistance via mutation. MEGARs is representative of this methodology, but also includes a carefully designed classification system for interpretation of these results (acyclical classifiers instead of ontologies).

In this part of the tutorial, your instructor will walk you through the MEGARs tutorial data, using the McArthur laboratory’s Galaxy server:

Server: http://galaxylab.mcmaster.ca

User: cr\_galax@mcmaster.ca

Password: Fall2015!

Steps in the analysis:

1. Uploading data
2. Trimmomatic read clean-up
3. Host genome filtering of reads
4. Burrows-Wheeler Transform alignment of reads to AMR reference data
5. Classification of results by gene, gene families, drug class, and mechanism
6. Rarefaction analysis