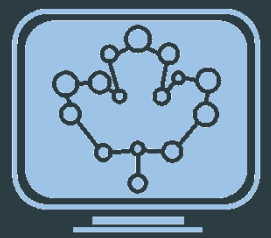
Bioinformatics for Antimicrobial Resistance Research

Rodrigo Ortega Polo
Agriculture and Agri-Food Canada
Lethbridge Research and Development Centre

One Health Summer Institute

June 16, 2021



Canada

Goals of the workshop

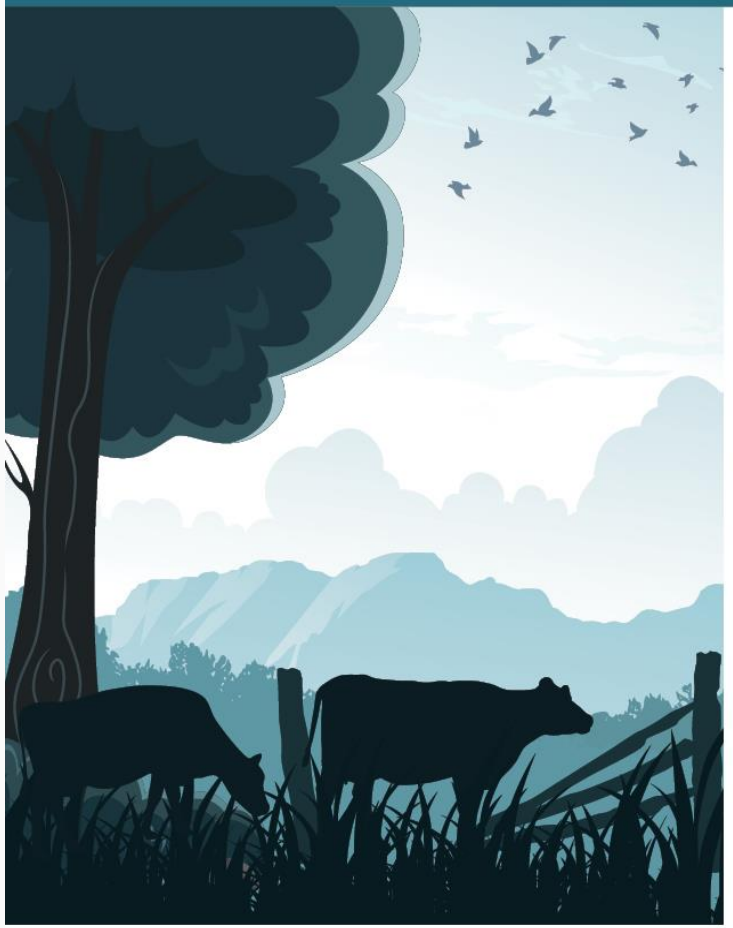
- Overview of methods for bacterial genomics and molecular epidemiology
- Introduction to databases and tools for the analysis of antimicrobial resistance genes

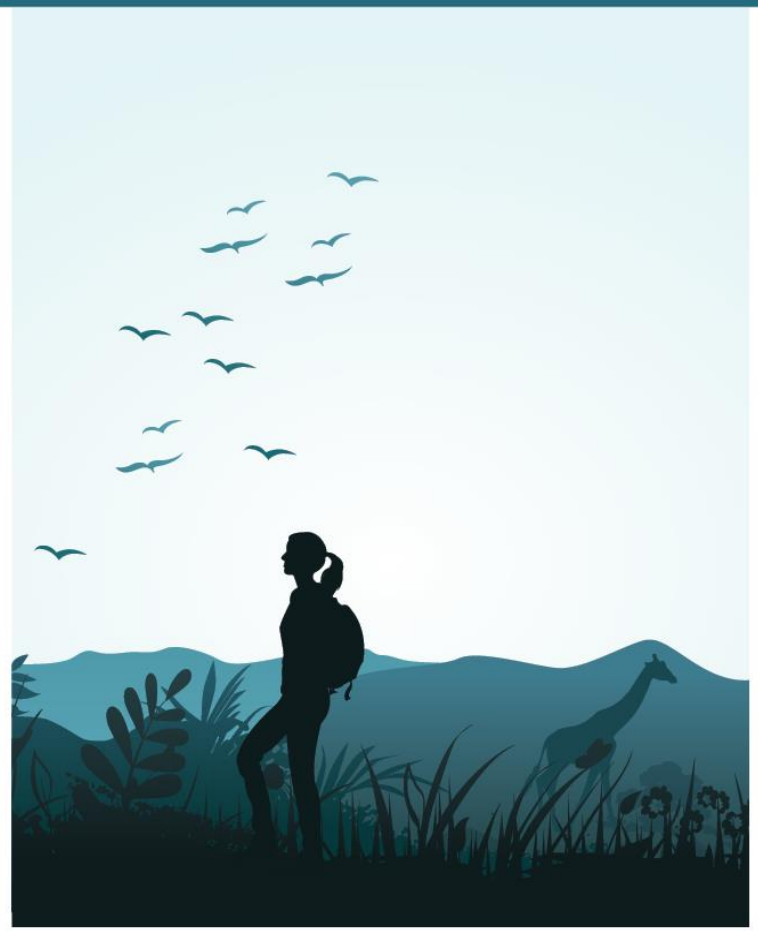
Introduction

- Head of the Bioinformatics Research Support Unit in a research and development centre of Agriculture and Agri-Food Canada (federal government)
- Located in Lethbridge, Alberta
- Distributed team of staff, students and collaborators (AB, BC, ON)
- Our unit is part of a national network of agriculture bioinformatics units
- I am a biologist who **applies** computational tools to address **research questions** in **Agriculture**
- Background:
  - Canadian Food Inspection Agency: research on foot-and-mouth disease virus
  - University of Lethbridge: analysis of One Health continuum metagenomics data
- We can do our own **experiments** with computational infrastructure
- Bioinformatics analysis is **driven by the research questions** and **study design**

ONE HEALTH

U.S. Department of Health and Human Services
Centers for Disease Control and Prevention





CONNECTING HUMAN, ANIMAL, AND ENVIRONMENTAL HEALTH

CS295032A

One Health is the idea that the health of people is connected to the health of animals and our shared environment.

When we protect **one**,
we help protect **all**.

www.cdc.gov/onehealth

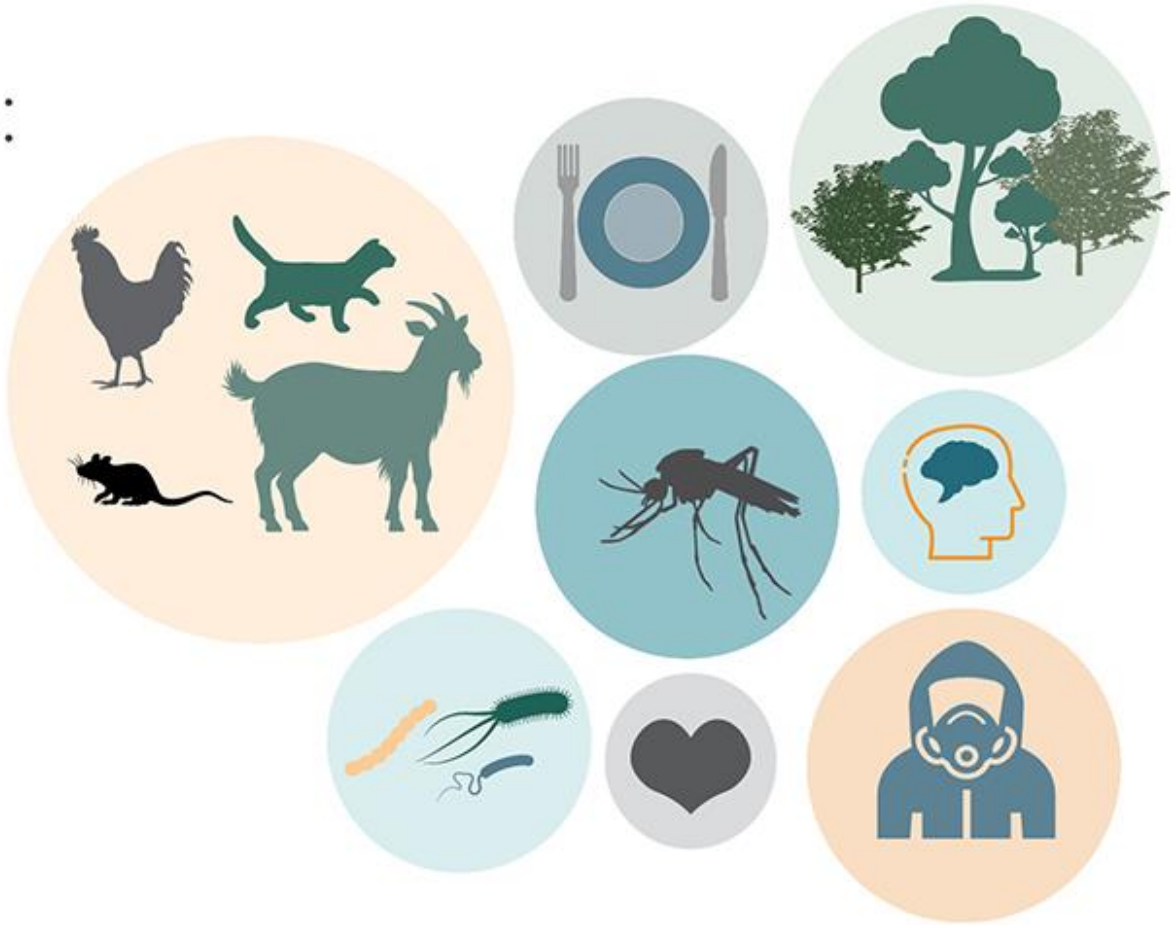
CDC

CS311436-A

Did You Know?

One Health issues include:

- Zoonotic diseases
- Antibiotic resistance
- Food safety and security
- Vector-borne diseases
- Environmental health
- Chronic diseases
- Mental health
- Occupational health



…And more!

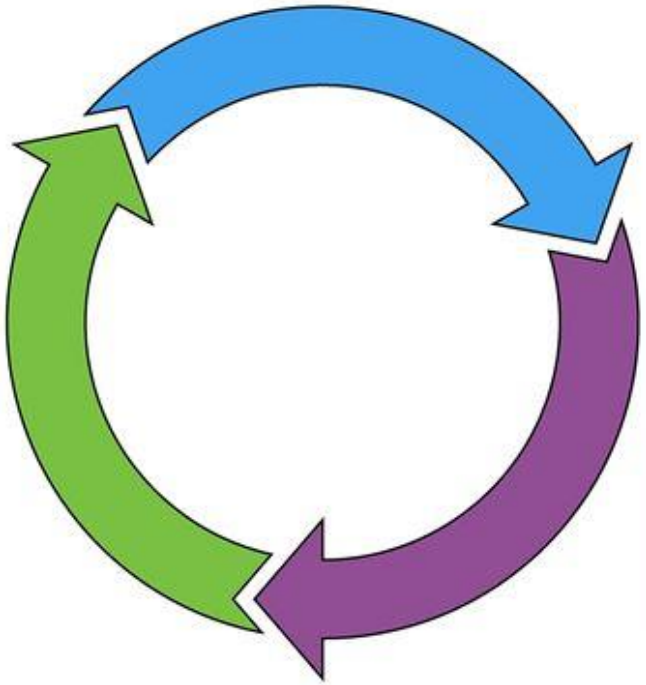
www.cdc.gov/onehealth

What do you study or want to study?

- What aspect of One Health do you work on?
- Are you focusing on a single pathogen? Or microbiomes / communities?
- Are you interested in presence/absence of pathogen? Or do you have other questions: AMR genes, virulence factors, genotyping?

Bacterial Typing
HTS for *in silico* typing and feature extraction to identify bacteria and provide brief characterization of known targets (i.e. ARG, virulence attributes)

Results from pathogenomic studies can reveal new genetic factors to develop/update current typing schemes and reference databases for additional feature extractions.



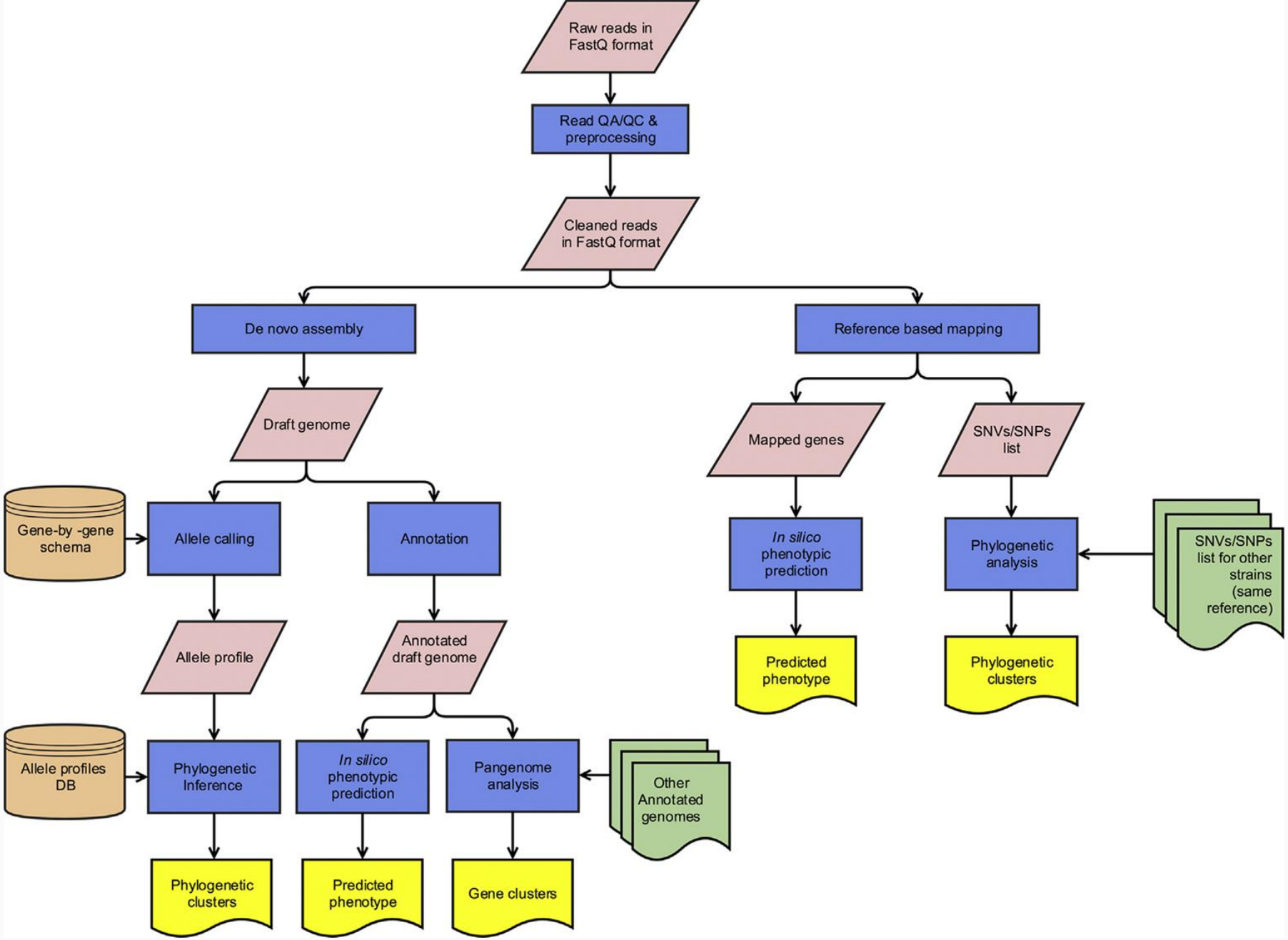
Bacterial typing informs isolate selection for comparison and provides genetic summary of known attributes relevant for outbreak investigations

Pathogenomics
Utilizes HTS data in exploratory research to address larger questions. Expand dataset using global repositories and employ additional methods to progress research field (proteomics, transcriptomics, gene knockout experiments etc.)

Molecular Epidemiology
Applies HTS data to resolve outbreak investigations and molecular trends in the population (i.e. surveillance of ARG in circulating bacterial infections of a specific geographical region).

Molecular Epidemiology provides new comparative characteristics that may define novel genome(s) or populations requiring further research to understand.

- Tarah Lynch, Aaron Petkau, Natalie Knox, Morag Graham, Gary Van Domselaar. Clinical Microbiology Reviews Sep 2016, 29 (4) 881-913; **DOI:** 10.1128/CMR.00001-16



- Commonly used HTS analysis and workflows (Carrico et al., 2018)

Focusing on a single species

- Bacterial whole genome sequence analysis
- High resolution analysis of populations
- Molecular epidemiology
- Hundreds (or thousands) of isolates

SCIENTIFIC REPORTS

Article | Open Access | Published: 03 March 2020

Surveillance of *Enterococcus spp*. reveals distinct species and antimicrobial resistance diversity across a One-Health continuum

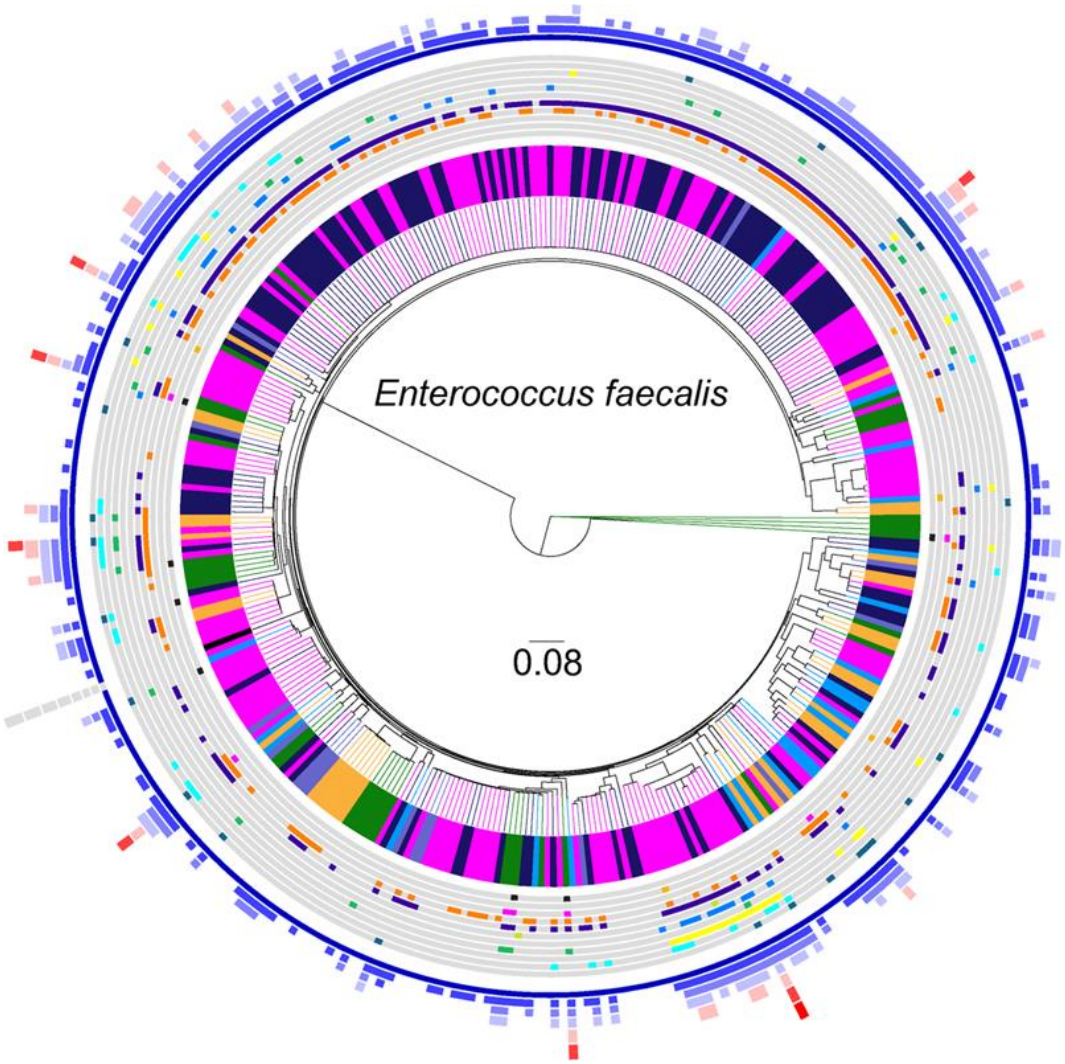
Rahat Zaheer, Shaun R. Cook, Ruth Barbieri, Noriko Goji, Andrew Cameron, Aaron Petkau, Rodrigo Ortega Polo, Lisa Tymensen, Courtney Stamm, Jiming Song, Sherry Hannon, Tineke Jones, Deirdre Church, Calvin W. Booker, Kingsley Amoako, Gary Van Domselaar, Ron R. Read & Tim A. McAllister

Scientific Reports **10**, Article number: 3937 (2020) | Cite this article

2742 Accesses | **2** Citations | **36** Altmetric | Metrics

Enterococcus spp. study

- **Goal:** to compare hundreds of isolates from clinical and agricultural settings
- *De novo* assemblies of whole genome sequences
- Genome annotation
- Antimicrobial resistance gene searches
- Analysis of 500+ isolates was orchestrated using a Snakemake workflow in the computer cluster of the Public Health Agency of Canada using the SLURM scheduler



Inner ring

Isolate source

- Clinical (Non-sterile & Sterile)
- Urban wastewater
- Natural water sources
- Feedlot catchbasin
- Bovine feces
- Beef processing

Middle ring

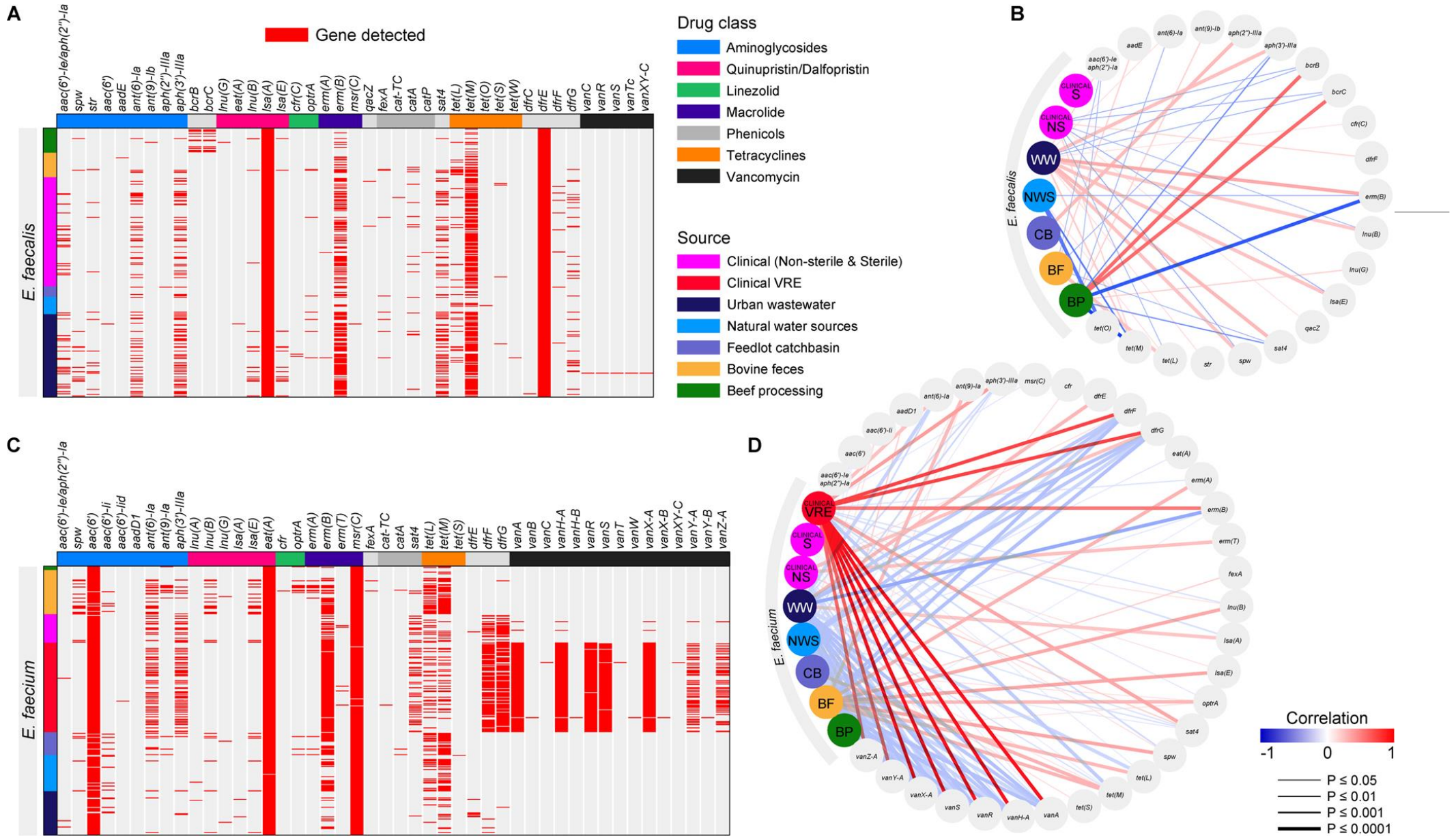
Phenotypic drug resistance

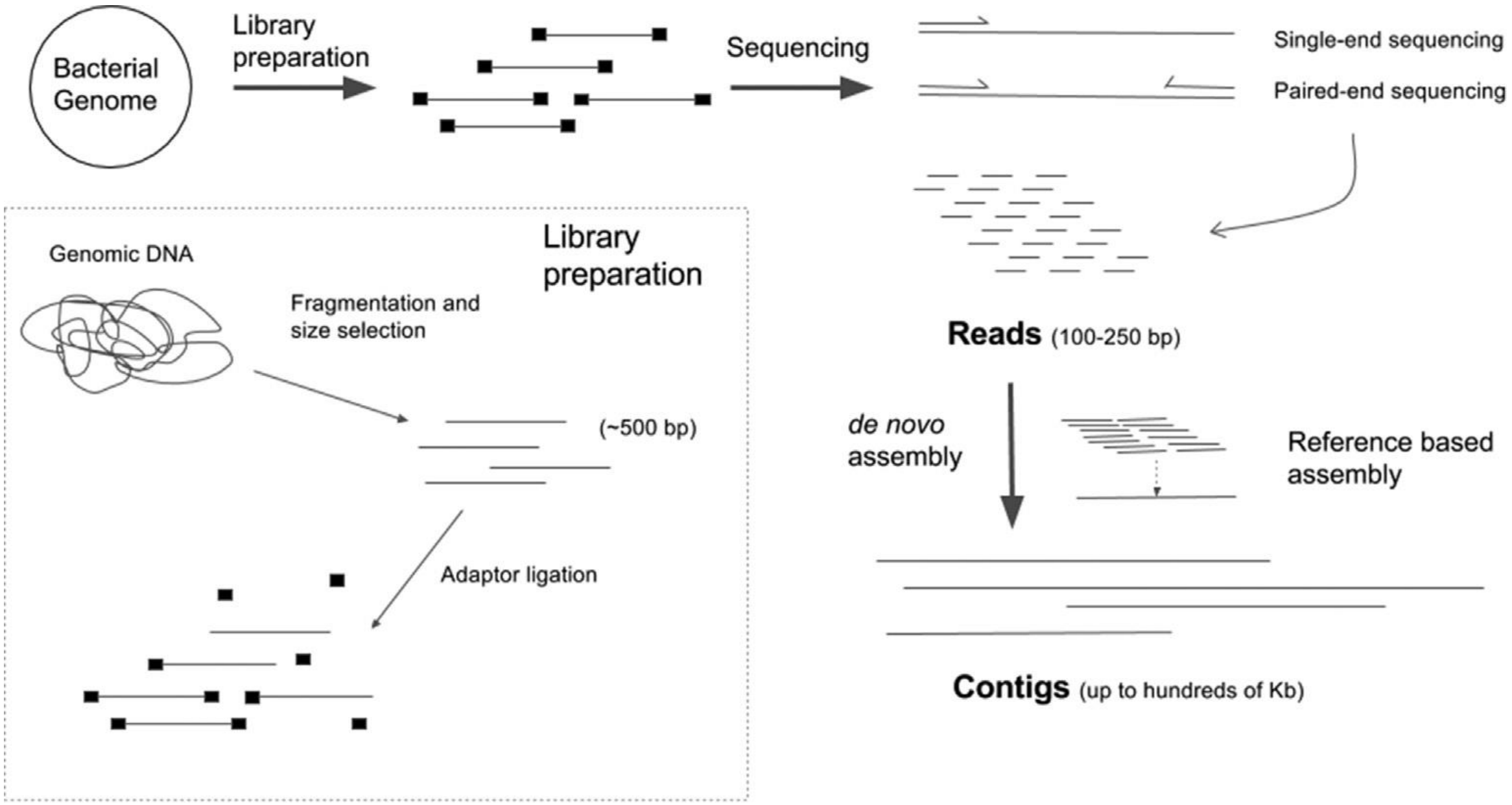
- Tigecycline
- Streptomycin
- Fluoroquinolones
- Linezolid
- Gentamicin
- Nitrofurantoin
- Erythromycin
- Tetracyclines
- β-lactam (penams)
- Teicoplanin
- Vancomycin

Outer ring

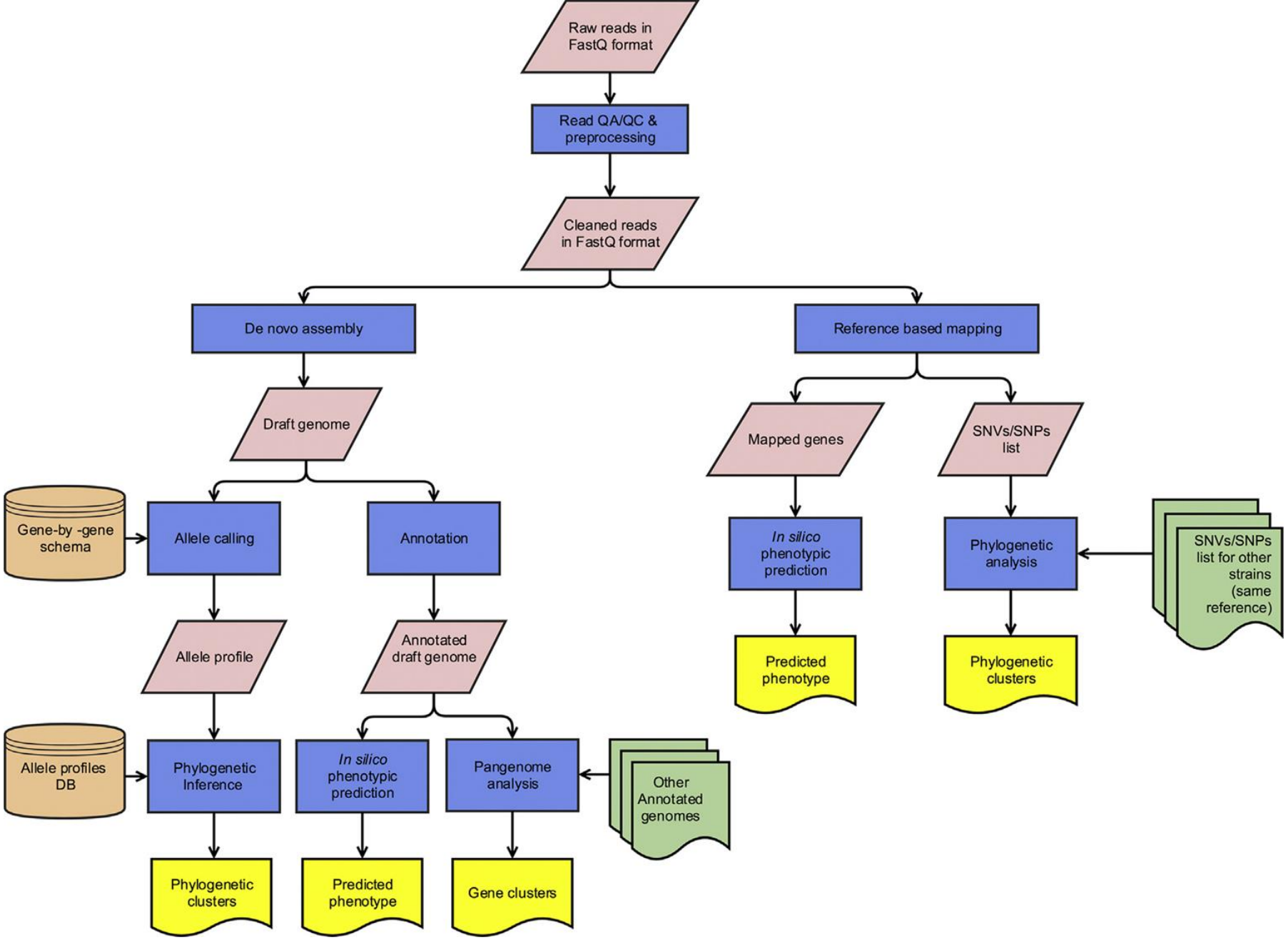
MDR

- Susceptible
- 1 drug
- 2 drugs
- 3 drugs
- 4 drugs
- 5 drugs
- 6 drugs
- 7 drugs

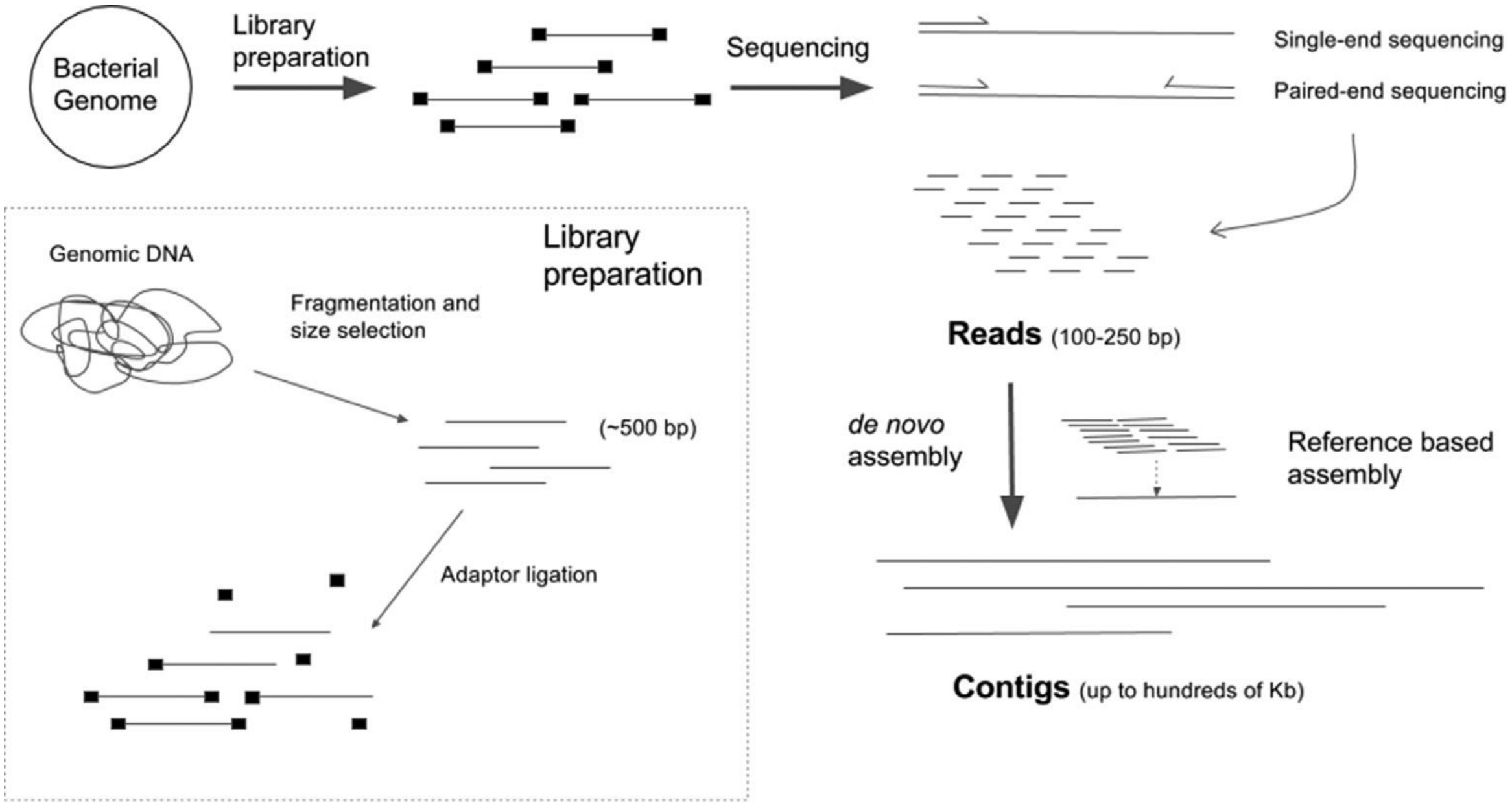

A
Gene detected
aac(6')-Ie/aph(2'')-Ia
spw
str
aac(6')
aadE
ant(6)-Ia
ant(9)-Ib
aph(2'')-IIIa
aph(3')-IIIa
bcrB
bcrC
lnu(G)
eat(A)
lnu(B)
lsa(A)
lsa(E)
cfr(C)
optrA
erm(A)
erm(B)
msr(C)
qacZ
fexA
cat-TC
catA
catP
sat4
tet(L)
tet(M)
tet(O)
tet(S)
tet(W)
dfrC
dfrE
dfrF
dfrG
vanC
vanR
vanS
vanTc
vanXY-C
E. faecalis
Drug class
Aminoglycosides
Quinupristin/Dalfopristin
Linezolid
Macrolide
Phenicols
Tetracyclines
Vancomycin
Source
Clinical (Non-sterile & Sterile)
Clinical VRE
Urban wastewater
Natural water sources
Feedlot catchbasin
Bovine feces
Beef processing
B
E. faecalis
CLINICAL S
CLINICAL NS
WW
NWS
CB
BF
BP
C
E. faecium
aac(6')-Ie/aph(2'')-Ia
spw
aac(6')
aac(6')-Ii
aac(6')-Iid
aadD1
ant(6)-Ia
ant(9)-Ia
aph(3')-IIIa
lnu(A)
lnu(B)
lnu(G)
lsa(A)
lsa(E)
eat(A)
cfr
optrA
erm(A)
erm(B)
erm(T)
msr(C)
fexA
cat-TC
catA
sat4
tet(L)
tet(M)
tet(S)
dfrE
dfrF
dfrG
vanA
vanB
vanC
vanH-A
vanH-B
vanR
vanS
vanT
vanW
vanX-A
vanX-B
vanXY-C
vanY-A
vanY-B
vanZ-A
D
E. faecium
CLINICAL VRE
CLINICAL S
CLINICAL NS
WW
NWS
CB
BF
BP
Correlation
-1
0
1
P ≤ 0.05
P ≤ 0.01
P ≤ 0.001
P ≤ 0.0001



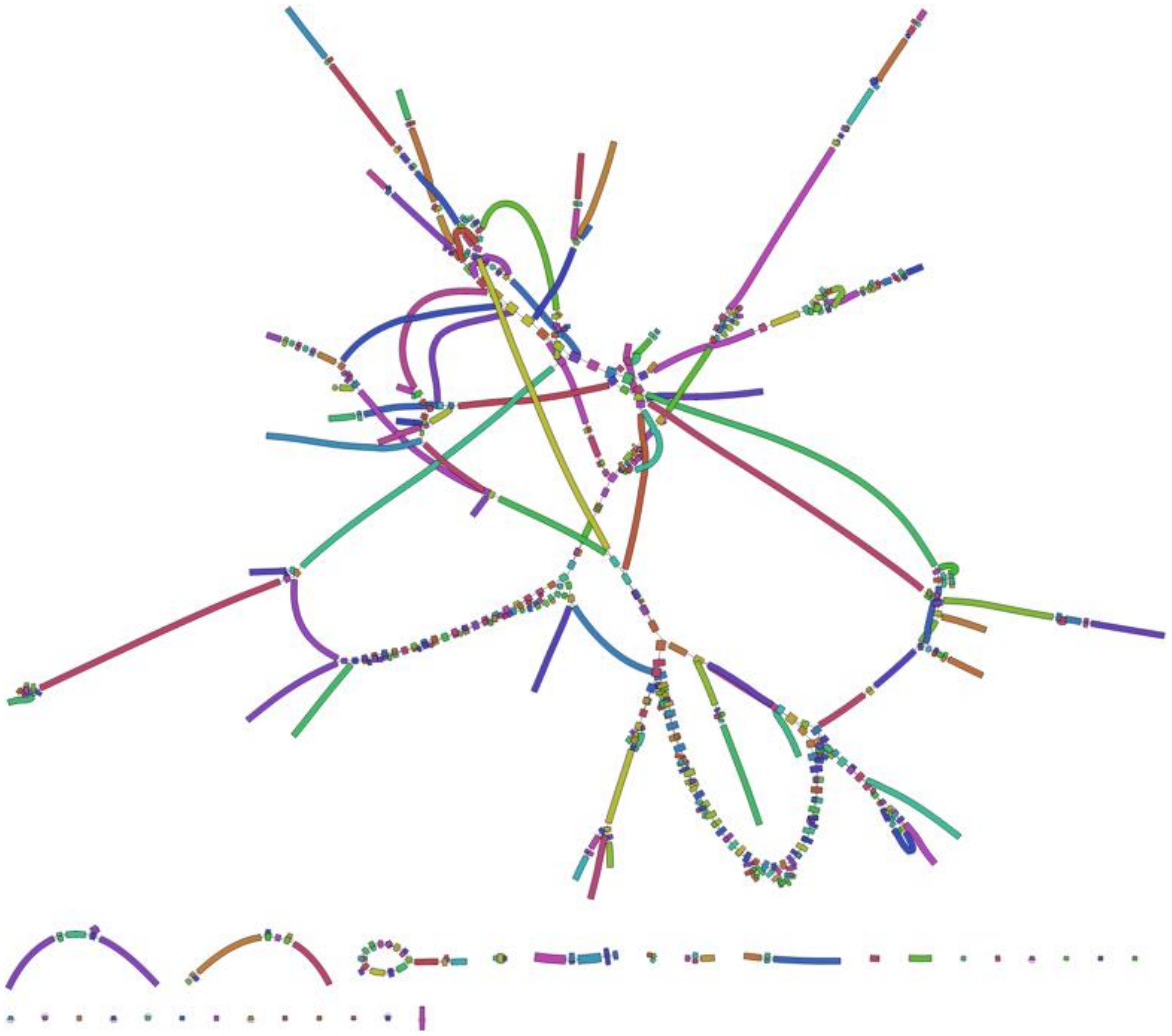
Sequencing concepts from bacterial genomes to contigs (Carrico et al., 2018)



Commonly used HTS analysis and workflows (Carrico et al., 2018)



Sequencing concepts from bacterial genomes to contigs (Carrico et al., 2018)

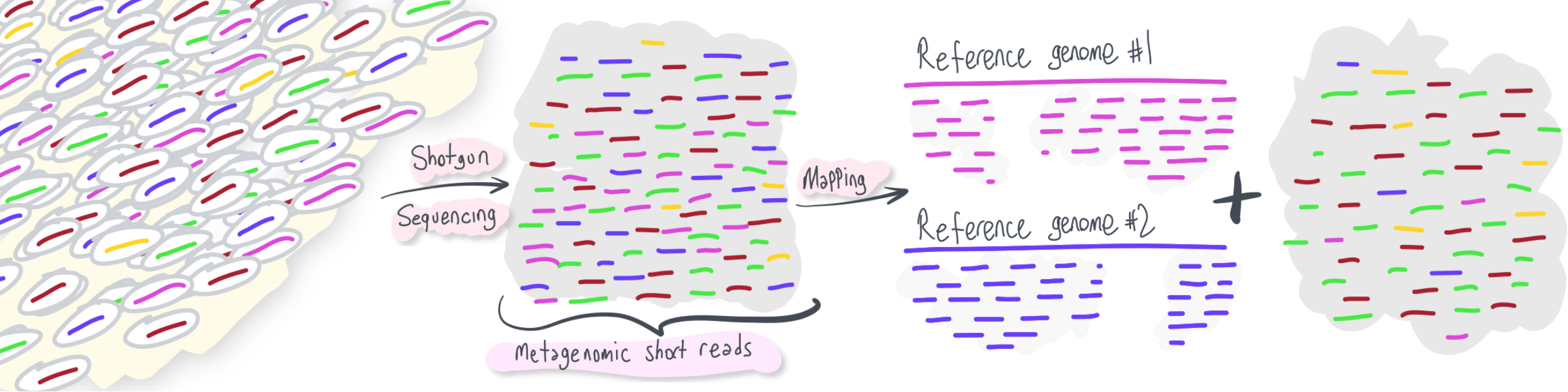


https://github.com/rrwick/Bandage/wiki/Effect-of-kmer-size

Visualization of draft assembly of *Salmonella* genome with Bandage software

Contigs compose an assembly graph

We can use this as a way to assess assembly quality



Functional profiling by read mapping

- Mapping reads with alignment software (e.g. BWA or Bowtie2)
- Reference gene database (e.g. MEGARes database for antimicrobial resistance genes)
- Count number of reads and gene fraction covered

Figure by Murat Eren (Univ. of Chicago) https://merenlab.org/momics/

Data analysis

How do we go from:

- 32 microbiome taxonomic assignment reports
- 32 rarefaction analyses
- 32 resistome classification datasets

to figures and analyses?

Working with the R and Python programming languages

R packages

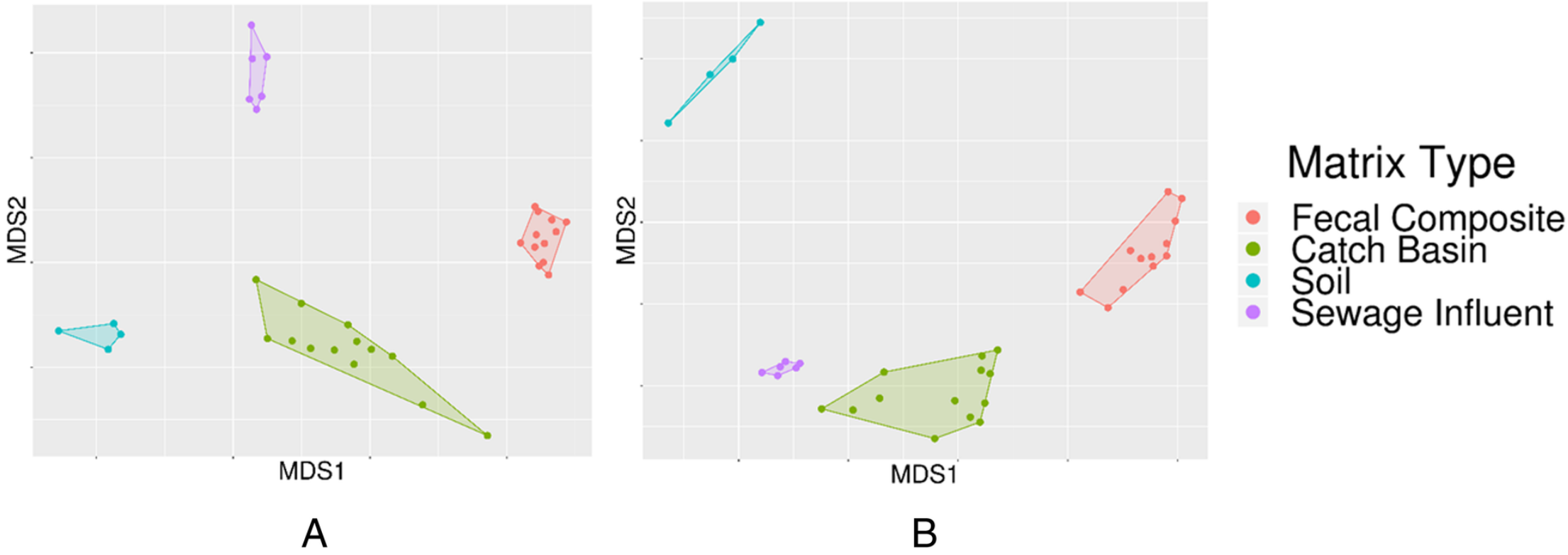
tidyverse - data analysis paradigm for R; encompasses many packages like *dplyr* and *ggplot2*

metagenomeSeq - analysis of metagenomic data

vegan - community ecology

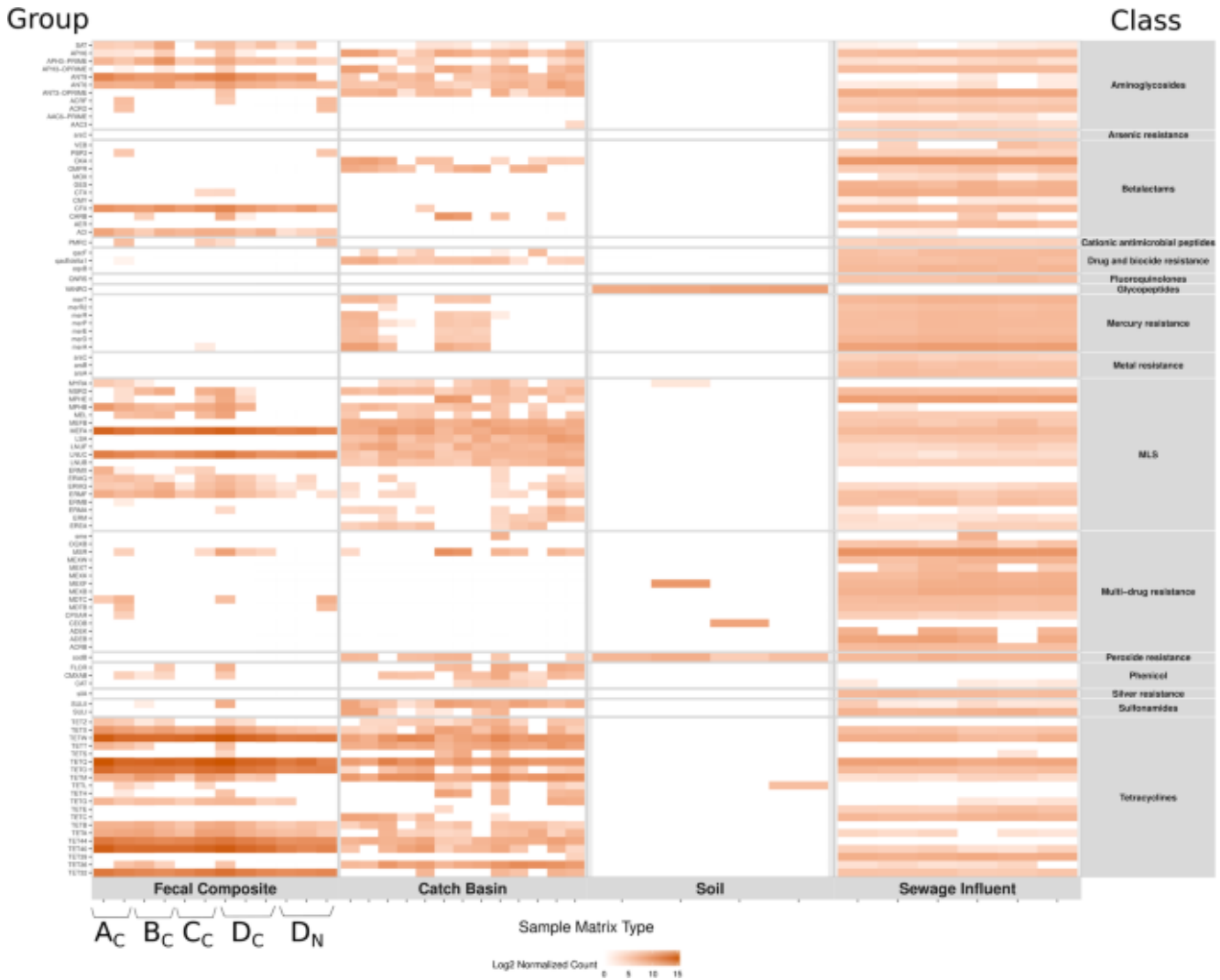
Analyzing metagenomic data at different scales

From the general…



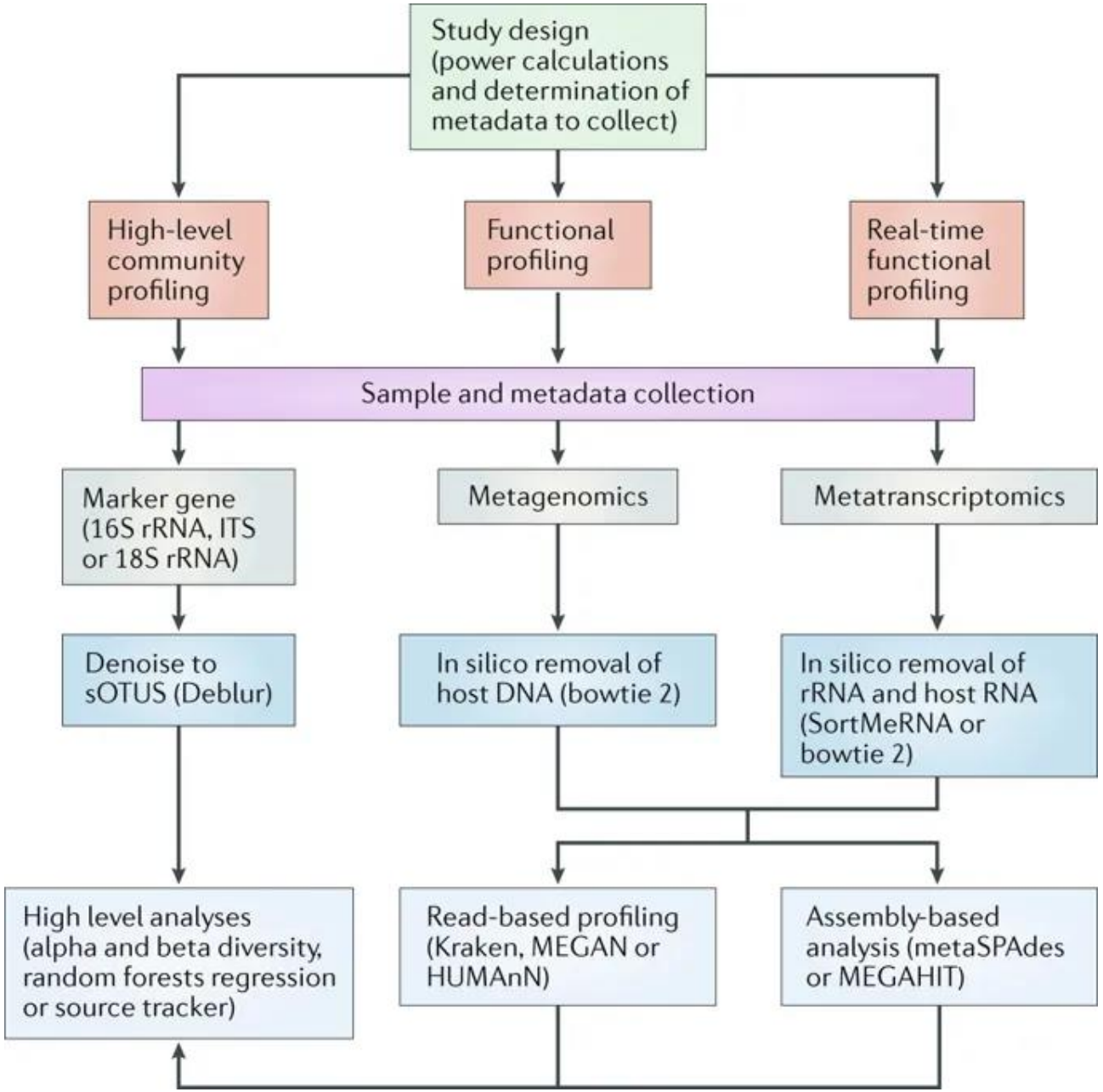
Zaheer, R., Lakin, S.M., Polo, R.O. *et al.* Comparative diversity of microbiomes and resistomes in beef feedlots, downstream environments and urban sewage influent. *BMC Microbiol* **19,** 197 (2019). https://doi.org/10.1186/s12866-019-1548-x

Analyzing metagenomic data at different scales



…to the particular

Zaheer, R., Lakin, S.M., Polo, R.O. *et al.* Comparative diversity of microbiomes and resistomes in beef feedlots, downstream environments and urban sewage influent. *BMC Microbiol* **19,** 197 (2019). https://doi.org/10.1186/s12866-019-1548-x



These approaches address different questions:

- Marker gene (amplicon)
  - **Who is there?**
- Metagenomics (shotgun sequencing)
  - **Who is there?**
  - **What genes do they have?**
- Metatranscriptomics
  - **Who is there?**
  - **What are they doing?**

Knight, R., Vrbanac, A., Taylor, B.C. *et al*. Best practices for analysing microbiomes. *Nat Rev Microbiol* 16, 410-422 (2018). https://doi.org/10.1038/s41579-018-0029-9

Thank you!

Agriculture and Agri-Food Canada

Agriculture et Agroalimentaire Canada

AAFC Lethbridge Bioinformatics Team

Mathew Richards

Lilia Mesina

Matvey Ryabov

Tara Newman

Lance Lansing

Shefali Vishwakarma

Catrione Lee

Morgan Cunningham

Gale Chen

@ropolo

rodrigo.ortegapolo@canada.ca

https://orcid.org/0000-0003-3411-923X

https://profils-profiles.science.gc.ca/en/profile/rodrigo-ortega-polo