

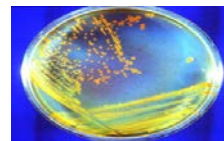
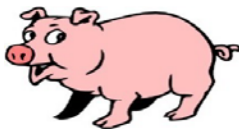
# Update on NCBI Pathogen Detection and Antimicrobial Resistance Activities

William Klimke -GenomeTrakr, Sept 26-28, Crystal City, VA

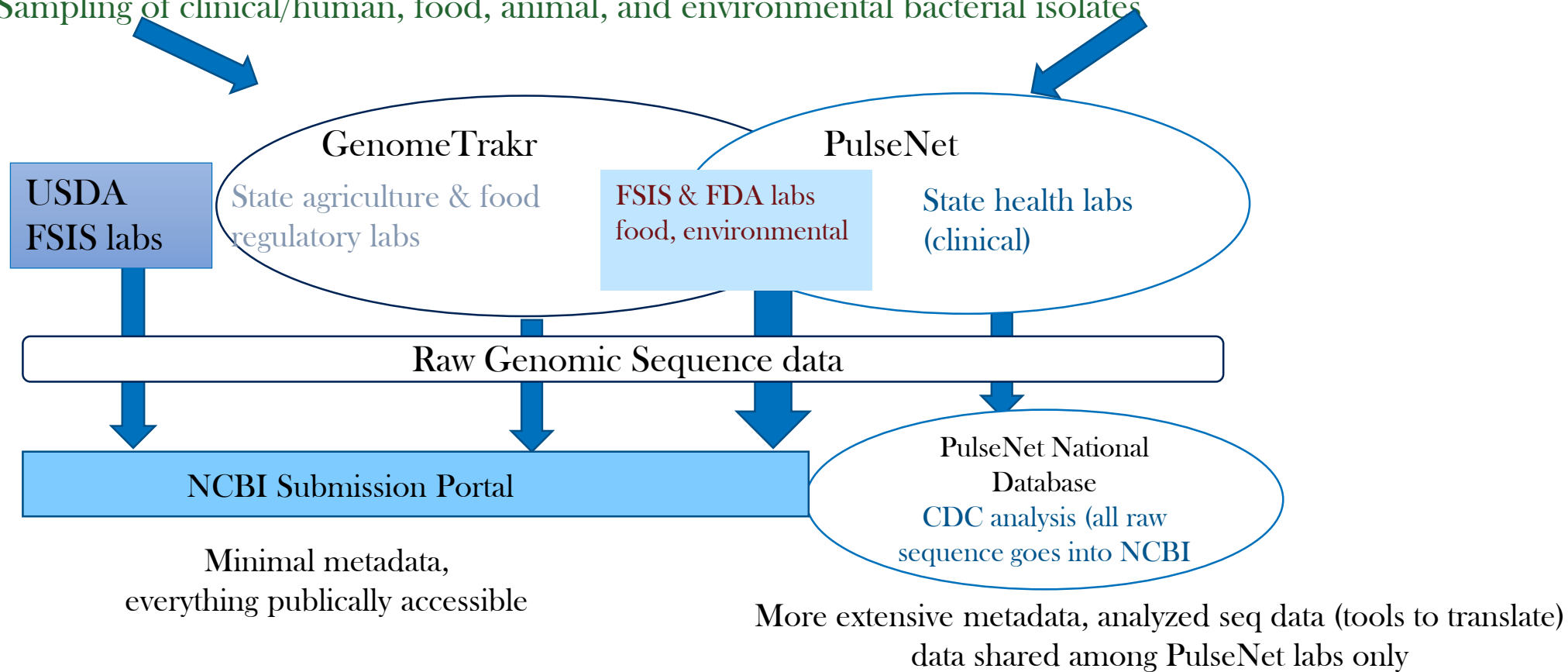


U.S. National Library of Medicine  
National Center for Biotechnology Information

# Shared Network pathways and data streams for outbreak detection and investigations



Sampling of clinical/human, food, animal, and environmental bacterial isolates

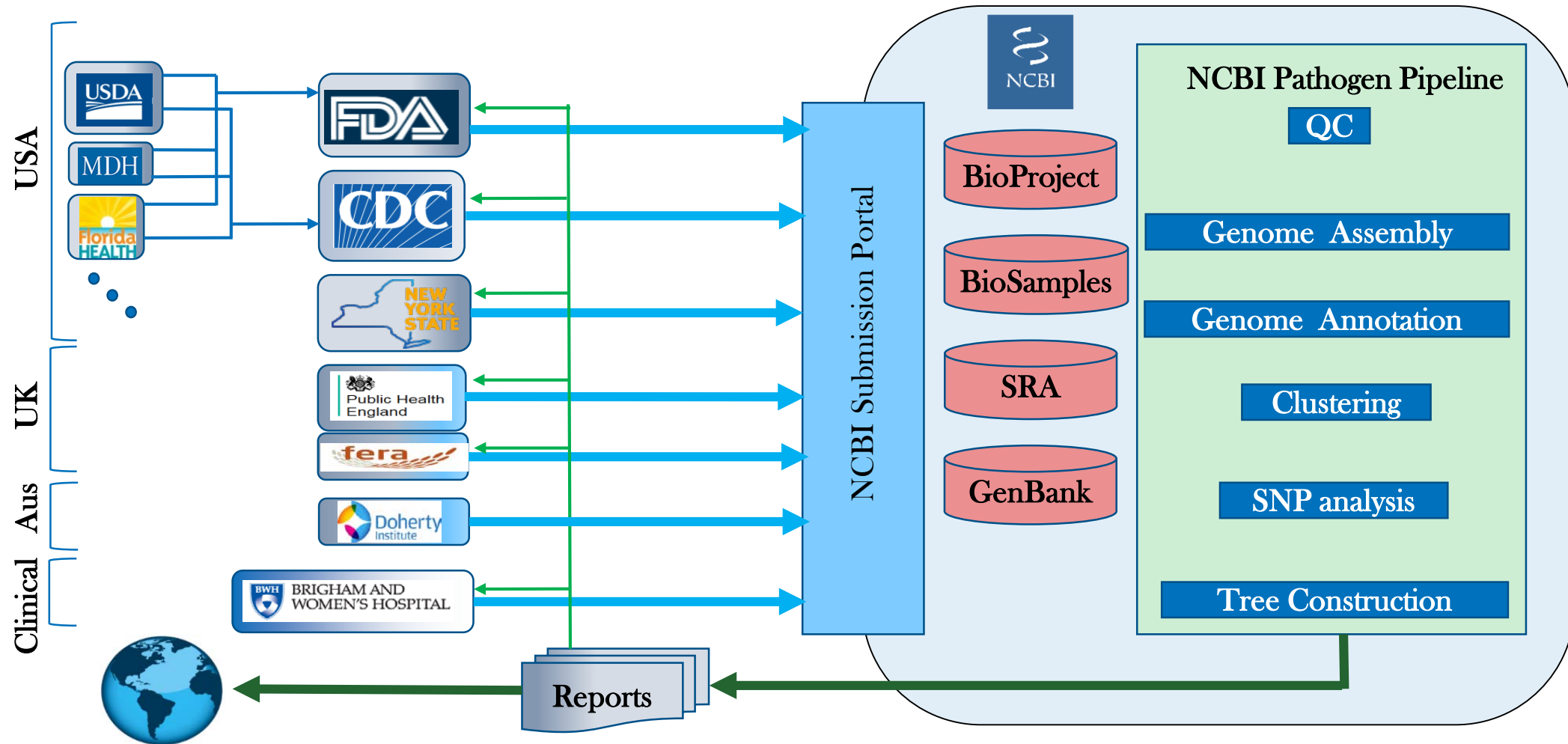


# Analysis goals

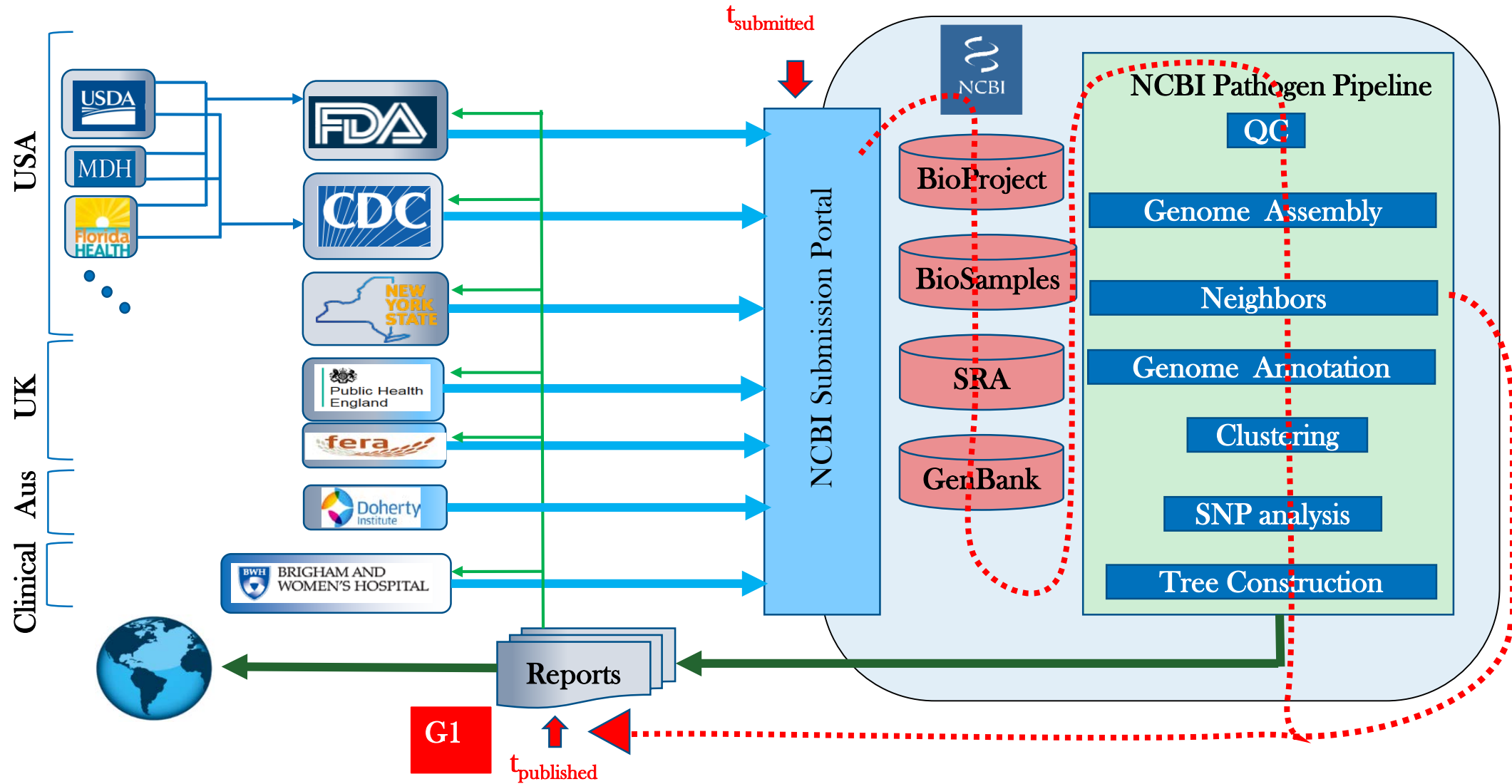
1. Are these isolates clonally related?
2. What is the anti-microbial resistance gene repertoire of this isolate



# Pathogen Detection Pipeline



# Goal: Improve Turnaround Time from Submission to Report



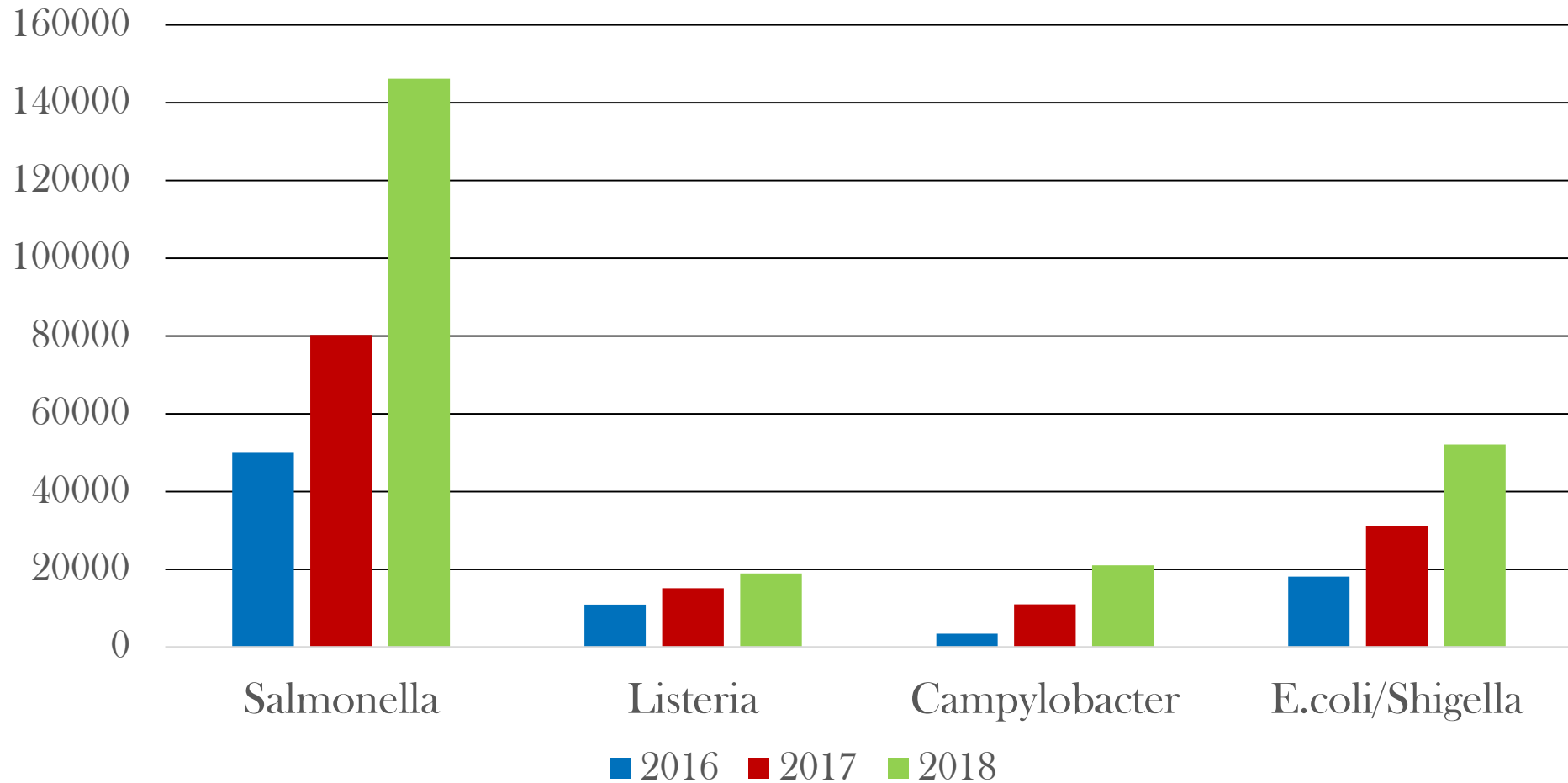
# Pipeline Changes

- SKESA assembler released  
<https://github.com/ncbi/SKESA>  
*SKESA publication accepted (Genome Biology)*
- SKESA running on all new submissions
- SKESA reassembly for all four foodborne pathogens

# Pipeline Changes

- wgMLST schemes developed for all foodborne pathogens
- wgMLST now running for all four foodborne pathogens
  - used for rapid reports (within one hour of submission report a table of nearest neighbors)
  - used for clustering to generate clusters for input to SNP pipeline

# Growth: Total Isolates in Pathogen Detection





## wgMLST/SNP Pipeline Comparison (hours)

| Organism        | No. of Isolates | Avg. Time to Publish (hours) From Submission |                           |
|-----------------|-----------------|--|---------------------------|
|                 |                 | Old SNP Pipeline                             | New wgMLST + SNP Pipeline |
| Campylobacter   | 21,214          | 13   | 5                         |
| E.coli/Shigella | 52,801          | 23   | 9                         |
| Listeria        | 19,146          | 14   | 5                         |
| Salmonella      | 147,150         | 92   | <b>52</b>                 |



# Pathogen Browser

- Will be covered by Arjun Prasad - Friday
  - navigation panel
  - subtree creation
  - sharing of highlighted isolates
  - tree labeling
- Automated alerts
  - first phase - stored searches send automatic notifications
  - second phase - 'watched' isolate mode

These are the last two major features needed to aid outbreak detection

# Automated Alerts: First phase (Stored Searches)

Health > Pathogen Detection > Isolates Browser > Saved searches

taxgroup\_name:"Klebsiella pneumoniae" AND new:1 ✕   Saved searches

Klebsiella pneumoniae ▼ ✕

Matched clusters

| # | Organism groups       | SNP cluster     | Matched isolates | Matched clinical isolates | Matched enviro |
|---|-----------------------|-----------------|------------------|---------------------------|----------------|
| 1 | Klebsiella pneumoniae | PDS000015678_34 | 1                | 1                         | 0              |

Page 1 of 1 | 3 | View 1 - 1 of 1

Name:  ✕

Search:

### Saved searches

|   | Name                       | Query  | Edit                                | Delete                           |
|---|----------------------------|--|-------------------------------------|----------------------------------|
| 1 | new mcr genes 2            | (AMR_genotypes:blaKPC* OR AMR_genotypes:mcr*) AND new:1                                  | <input type="button" value="Edit"/> | <input type="button" value="✕"/> |
| 2 | new usa salmonella         | new:1 AND taxgroup_name:"Salmonella enterica" AND geo_loc_name:USA* AND mindiff:[0 TO 4] | <input type="button" value="Edit"/> | <input type="button" value="✕"/> |
| 3 | new salmonella less than 4 | new:1 AND taxgroup_name:"Salmonella enterica" AND mindiff:[0 to 3]                       | <input type="button" value="Edit"/> | <input type="button" value="✕"/> |
| 4 | newest salmonella          | taxgroup_name:"Salmonella enterica" AND new:1  | <input type="button" value="Edit"/> | <input type="button" value="✕"/> |

Requires MyNCBI login

# Automated Alerts: Email Notifications

Dear Colleagues:

New Salmonella enterica results are available.

2018-09-17T18:37:20Z

1. PDS000032654.6 hits:14 cluster\_size:82 mindiff:0 clinical:11 environmental:3  
<https://www.ncbi.nlm.nih.gov/Structure/tree/#/tree/Salmonella/PDG000000002.1189/PD>
2. PDS000032578.8 hits:5 cluster\_size:258 mindiff:0 clinical:4 environmental:1  
<https://www.ncbi.nlm.nih.gov/Structure/tree/#/tree/Salmonella/PDG000000002.1189/PD>
3. PDS000027927.5 hits:3 cluster\_size:27 mindiff:0 clinical:1 environmental:2  
<https://www.ncbi.nlm.nih.gov/Structure/tree/#/tree/Salmonella/PDG000000002.1189/PD>
4. PDS000018424.16 hits:2 cluster\_size:128 mindiff:0 clinical:2 environmental:0



# Exceptions Report

- enumerates the reason an isolate has failed for those organisms undergoing wgMLST clustering
- text file on FTP
- [ftp://ftp.ncbi.nlm.nih.gov/pathogen/Results/Listeria/latest\\_snps/Exceptions/](ftp://ftp.ncbi.nlm.nih.gov/pathogen/Results/Listeria/latest_snps/Exceptions/)

| exception type              | exception                 | consequence   | lower limit | upper limit | actual value | biosample_acc |
|-----------------------------|---------------------------|---------------|-------------|-------------|--------------|---------------|
| Assembly validation failure | Low contig N50            | Not published | 10000       | NULL        | 2736         | SAMN05173249  |
| Assembly validation failure | High contig L50           | Not published | NULL        | 200         | 330          | SAMN05173249  |
| Readset validation failure  | Insufficient coverage     | Not published | 20          | NULL        | 18           | SAMN10079593  |
| wgMLST validation failure   | Too few wgMLST loci found | Not clustered | 2000        | NULL        | 1723         | SAMN03761820  |

- will have strain and SRA\_center added as columns

# NCBI's Role in Combatting Antibiotic Resistance

## NATIONAL ACTION PLAN FOR COMBATING ANTIBIOTIC-RESISTANT BACTERIA

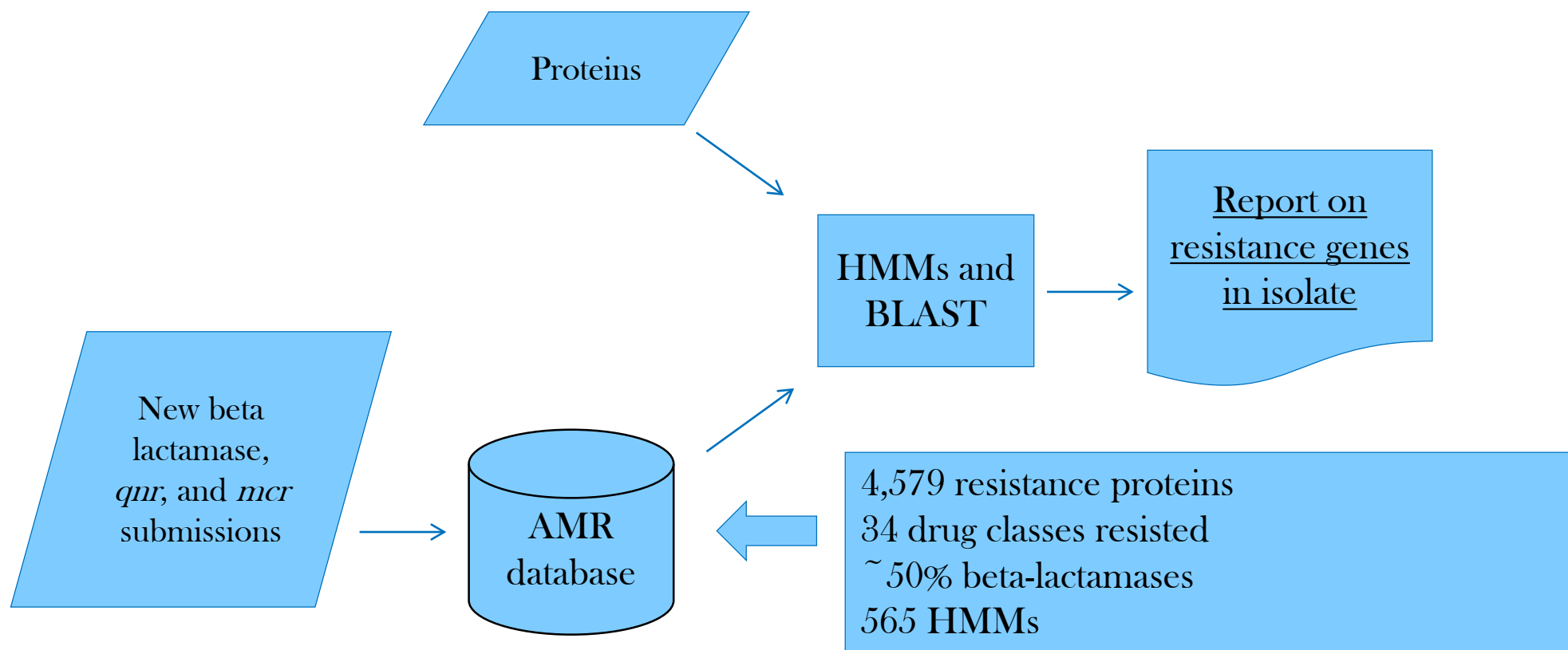
MARCH 2015



1. Build AMR reference database (reference proteins and hidden Markov models (HMMs) and protein family hierarchy)
2. Build AMRFinder tool to identify AMR proteins using reference database
3. Use AMRFinder to identify AMR proteins in all pathogen isolates integrated into NCBI Pathogen Browser
4. Capture antibiotic susceptibility test data (AST)
5. Integrate AST into NCBI Pathogen Browser

[https://ftp.ncbi.nlm.nih.gov/pub/factsheets/Factsheet\\_AMR\\_Project.pdf](https://ftp.ncbi.nlm.nih.gov/pub/factsheets/Factsheet_AMR_Project.pdf)

# From Annotation to Resistance Genes



# AMRFinder Alpha Release

[Health](#) > [Pathogen Detection](#) > [Antimicrobial Resistance](#) > AMRFinder

## AMRFinder **ALPHA**

NCBI has developed AMRFinder, a tool that identifies AMR genes using either protein annotations or nucleotide sequence. AMRFinder is used in the [Pathogen Detection pipeline](#), and these data are displayed in [NCBI's Isolate Browser](#). AMRFinder can identify acquired antimicrobial resistance genes in either protein datasets or nucleotide data, including genomic data. AMRFinder relies on NCBI's curated AMR gene database and curated collection of Hidden Markov Models. For more information on how AMRFinder operates, please see the available [README file](#).

To enable researchers to incorporate these tools into their own analytical pipelines, NCBI has made the following resources publicly available:

1. A downloadable version of [AMRFinder using CWL and Docker](#). See the [README](#) for installation instructions.
2. NCBI's Bacterial Antimicrobial Resistance Reference Gene Database, which contains over 4,000 curated AMR protein sequences. There is also a tab-delimited table that lists accessions, gene names, and product names for all AMR genes in the [reference database](#).
3. NCBI's [curated HMM collection](#) of AMR protein sequences.

Note that this version of AMRFinder is an alpha version. If you have any questions about or experience problems running AMRFinder, please contact [pd-help@ncbi.nlm.nih.gov](mailto:pd-help@ncbi.nlm.nih.gov)

<https://www.ncbi.nlm.nih.gov/pathogens/antimicrobial-resistance/AMRFinder/>





# AMRFinder and Antimicrobial Resistance Resources

- scientific publication describing AMRFinder and genotype/phenotype correlations on ~6K NARMS isolates
- organism-specific point mutations (29 genes) covering resistance to quinolones, macrolides, etc.
- AMR resource landing pages
- AMR Gene Browser - for every AMR gene in every isolate
  - starting with some high priority organisms (Klebsiella)
  - requires annotated genome to be available
- stress response and virulence genes

# NCBI Pathogen Detection Roadmap

Sept. - Oct/2018

Nov - Dec/2018

2019

Pathogen  
Browser

automated alerts

'watched isolates'

browser publication

Pathogen  
Detection  
Pipeline

improve time to report

improve time to report

wgMLST publication

begin GenBank  
submission

wgMLST for other  
pathogens

pipeline publication

AMRFinder  
and Resources

point mutations

publication on AMRFinder

combined protein/  
translated searches

stress response and  
virulence genes

AMR Gene Browser

# Thank you.

Richa Agarwala  
Azat Badretdin  
Slava Brover  
Joshua Cherry  
Jinna Choi  
Vyacheslav Chetvernin  
Robert Cohen  
Michael DiCuccio  
Boris Fedorov  
Mike Feldgarden  
Lewis Geer  
Dan Haft  
Lianyi Han  
Avi Kimchi  
Michel Kimelman  
William Klimke  
Alex Kotliarov  
Valerii Lashmanov  
Aleksandr Morgulis  
Eyal Moses  
Chris O'Sullivan  
Arjun Prasad

Edward Rice  
Kirill Rotmistrovskyy  
Alejandro A. Schaffer  
Nadya Serova  
Stephen Sherry  
Sergey Shirayev  
Martin Shumway  
Oleg Shutov  
Alexandre Souvorov  
Tatiana Tatusova  
Igor Tolstoy  
Chunlin Xiao  
Leonid Zaslavsky  
Alexander Zasyarkin  
Lukas Wagner  
Hlavina Wratko  
Eugene Yaschenko  
  
David Lipman  
James Ostell  
Kim Pruitt

This research was supported by the  
Intramural Research Program of the  
NIH, National Library of Medicine.

[pd-help@ncbi.nlm.nih.gov](mailto:pd-help@ncbi.nlm.nih.gov)

CDC  
FDA/CFSAN  
GenFS  
USDA-FSIS  
PHE/FERA  
NIHGRI  
NIAID  
WRAIR  
Broad  
Wadsworth/MDH  
Vendors: PacBio, Illumina, Roche