

Screening biological products for antimicrobial resistant genes using the HIVE AMR pipeline



Sean Smith, Ph.D.
Office of Biostatistics and Pharmacovigilance
Center for Biologics Evaluation and Research



Introduction



Definition

Bacterial antimicrobial resistance (AMR) occurs when changes in bacteria increase resistance of the bacteria to drug treatments

How Bacteria and Fungi Fight Back Against Antibiotics

Antibiotics fight germs (bacteria and fungi). But germs fight back and find new ways to survive. Their defense strategies are called **resistance mechanisms**. Only germs, not people, become resistant to antibiotics.


Germ develops new cell processes that avoid using the antibiotic's target.

Germ changes or destroys the antibiotic with enzymes, proteins that break down the drug.

Germ restricts access by changing the entryways or limiting the number of entryways.

Germ changes the antibiotic's target so the drug can no longer fit and do its job.

Germ gets rid of antibiotics using pumps.



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

How Antibiotic Resistance Moves Directly Germ to Germ

Any antibiotic use can lead to antibiotic resistance. Antibiotics kill germs like bacteria and fungi, but the resistant survivors remain.

Resistance traits can be inherited generation to generation. They can also pass directly from germ to germ by way of **mobile genetic elements**.

Mobile Genetic Elements


- Plasmids**
Circles of DNA that can move between cells.
- Transposons**
Small pieces of DNA that can go into and change the overall DNA of a cell. These can move from chromosomes (which carry all the genes essential for germ survival) to plasmids and back.
- Phages**
Viruses that attack germs and can carry DNA from germ to germ.

How Mobile Genetic Elements Work

Transduction
Resistance genes can be transferred from one germ to another via phages.

Conjugation
Resistance genes can be transferred between germs when they connect.

Transformation
Resistance genes released from nearby live or dead germs can be picked up directly by another germ.

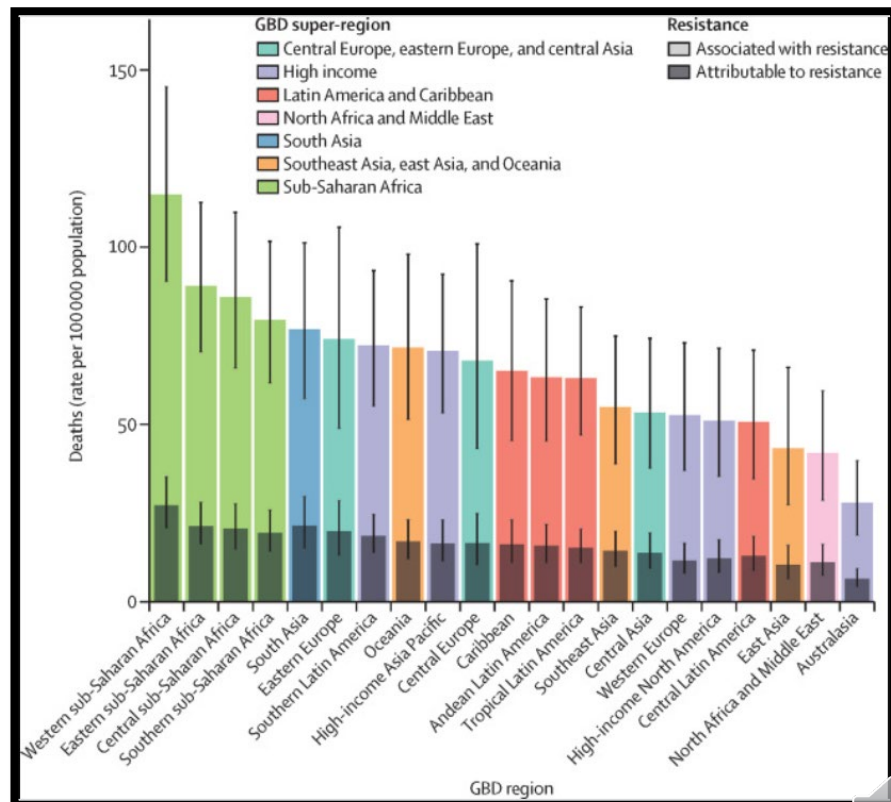


U.S. Department of Health and Human Services
Centers for Disease Control and Prevention

Introduction, cont.

• World-wide impact

- AMR estimates for year 2019 →
 - 4.95 million AMR-associated deaths¹
 - 1.27 million AMR-attributable deaths¹
- Predicted estimates for 2050
 - 10 million AMR-associated deaths annually (more than currently die from cancer)²



¹Murray, 2021, Lancet. ²O'Neill, Review on Antimicrobial Resistance.

Objectives



- **FDA Review**

- Biologic therapeutics (bacteria, phages) reviewed at FDA
 - Screen for safety issues, antibiotic resistance genes

- **Application of technology / scientific computing**



- Developed **antimicrobial resistance pipeline** through the **High-performance Integrated Virtual Environment (HIVE)**
- Accepts high-throughput sequencing (HTS) reads as input and provides a **detailed report** on **potential antibiotic resistant (ABR)** genes present in the biologic product

AMR Pipeline



Designed, optimized for different high-throughput sequencing technologies:

Short reads (dark blue)

Illumina, Ion Torrent
Implemented in HIVE

Long reads

PacBio, Oxford Nanopore

Hybrid

Both

Second Generation
Massively Parallel Sequencing

A T C G

- Sequencing by synthesis
- Amplified templates are generated during sequencing, reducing the requirements for starting material
- High accuracy
- Short read lengths

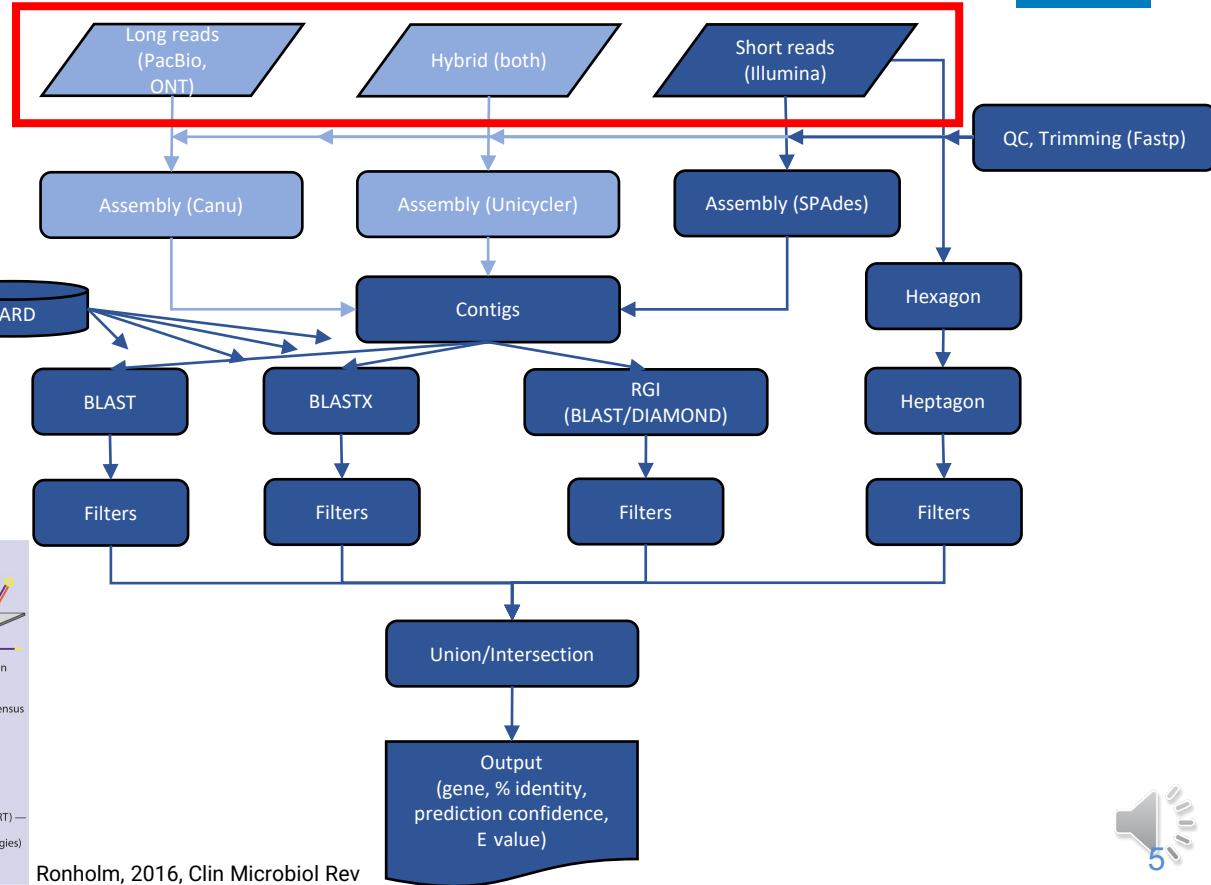
e.g., MiSeq (Illumina), Ion Torrent (Thermo Fisher Scientific)

Third Generation
Single-molecule Sequencing

Mask hairpin sequences
Generate consensus sequence

- Single-molecule templates
- Low accuracy
- Long read lengths

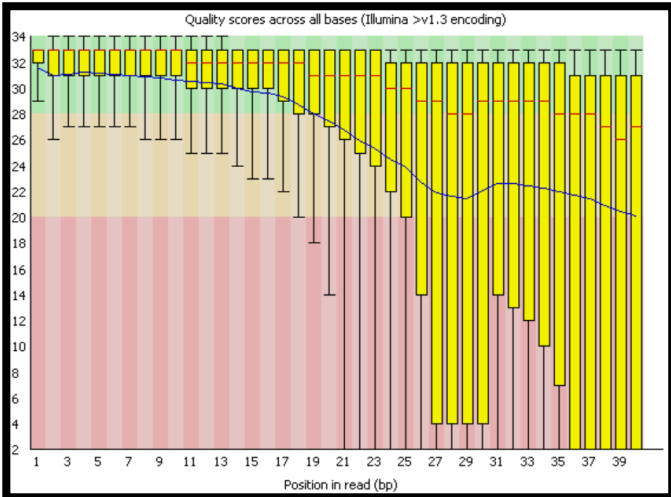
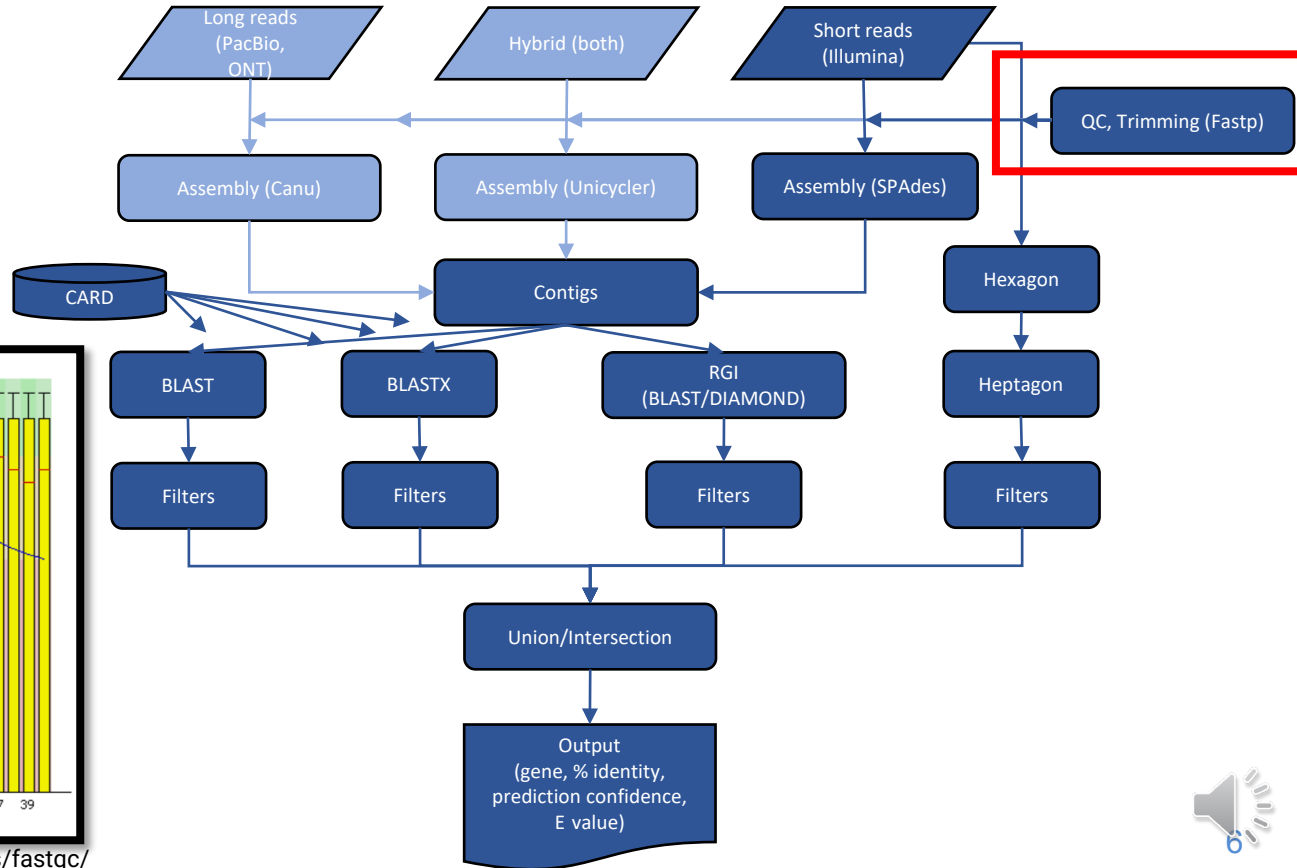
e.g., Single-Molecule Real-Time (SMRT) — Sequencing (Pacific Biosciences), MinION (Oxford Nanopore Technologies)



AMR Pipeline



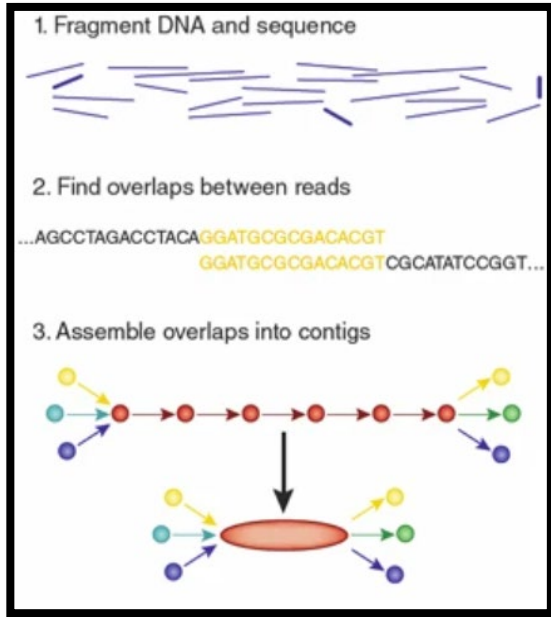
Quality control and trimming of reads



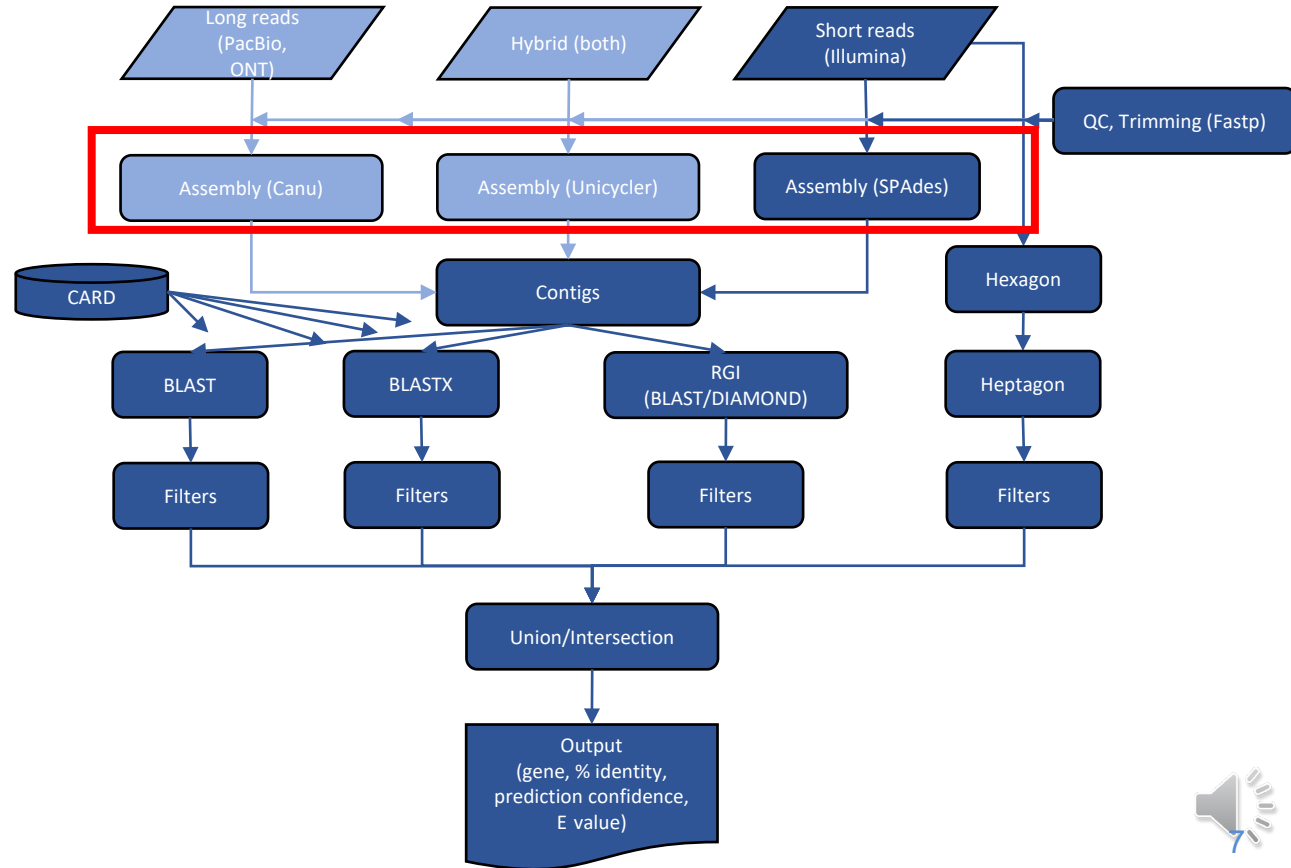
AMR Pipeline



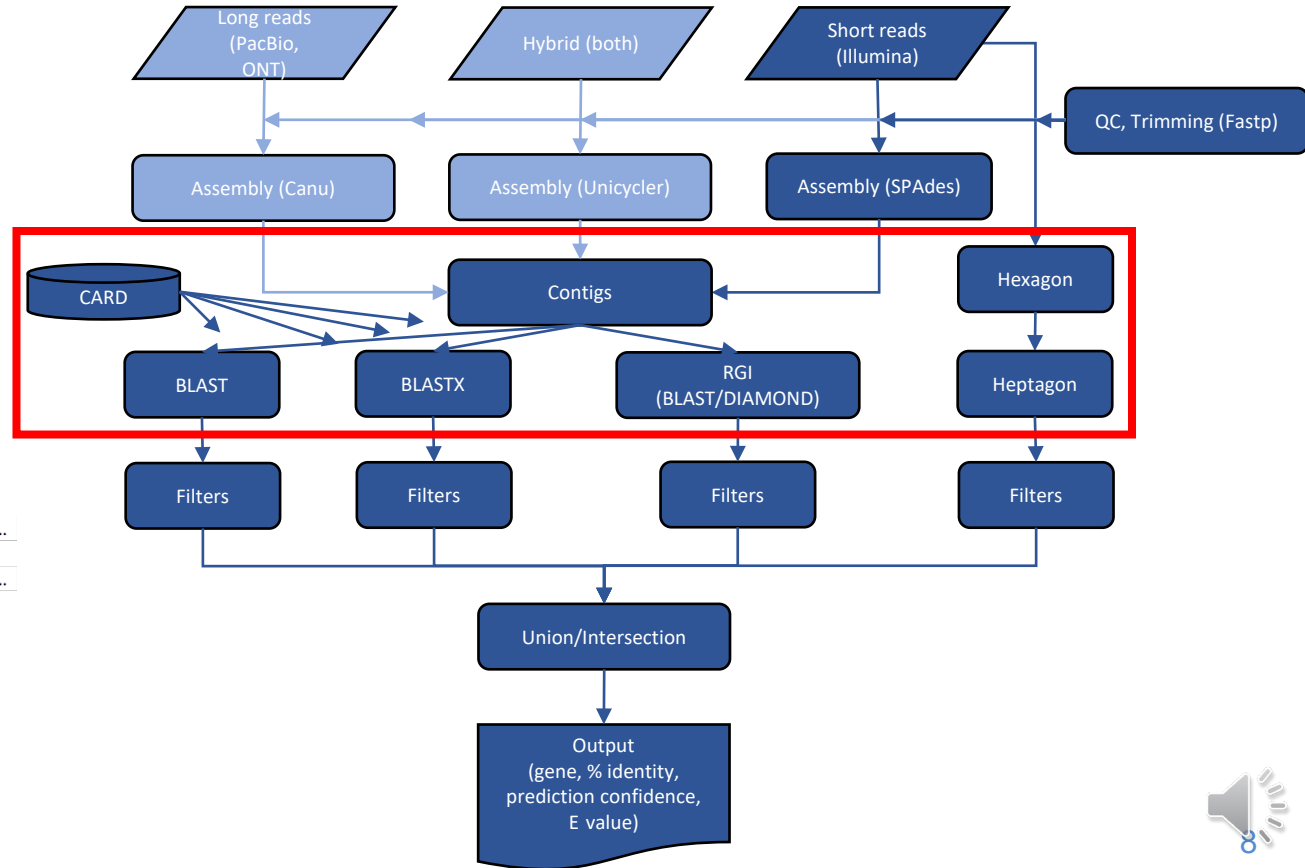
Assemble reads into contigs



Baker, 2012, Nature Methods



AMR Pipeline



Align contigs to CARD database

Contig

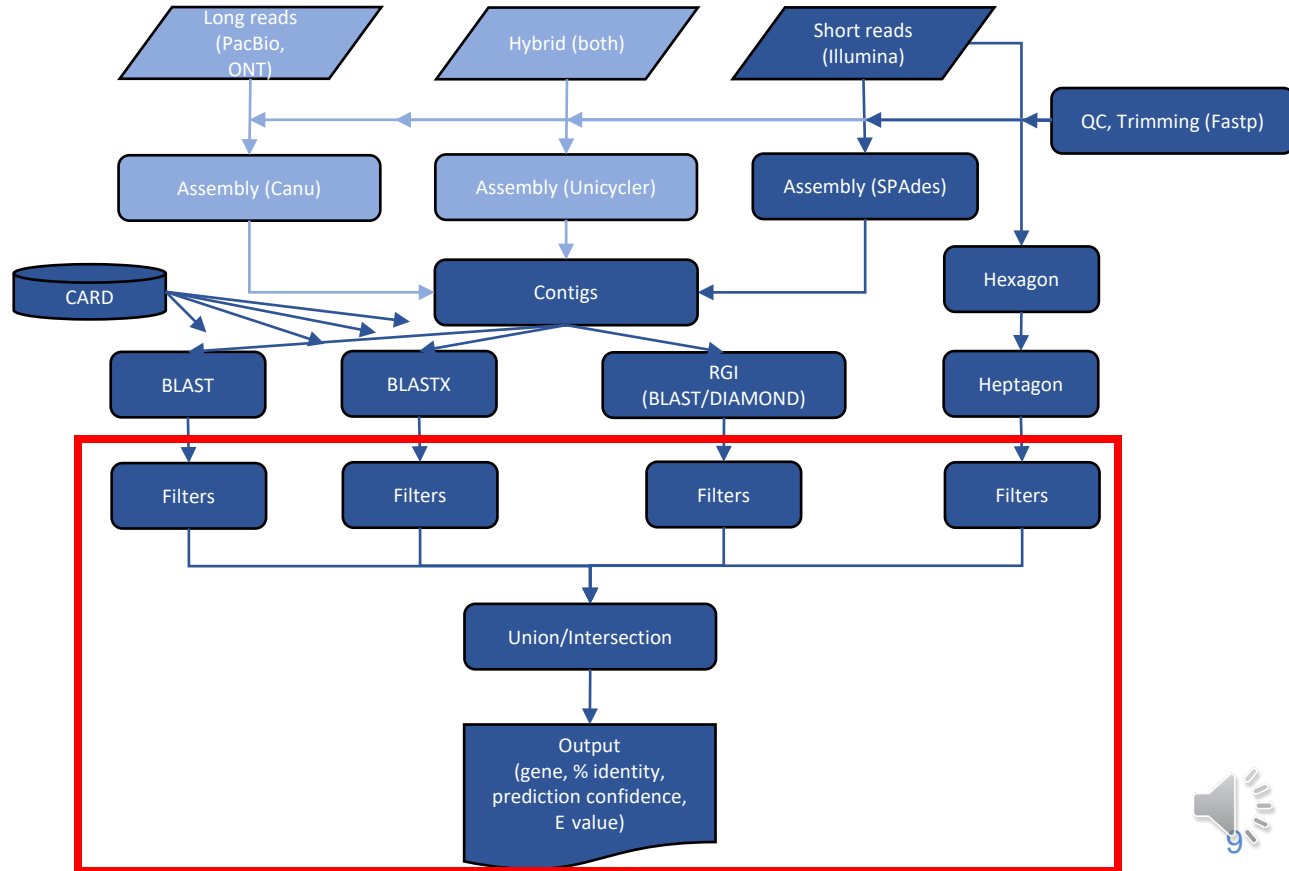
A T G T C A C A T ...

AMR Gene

T C C A T T A T G T C T C A T ...



AMR Pipeline



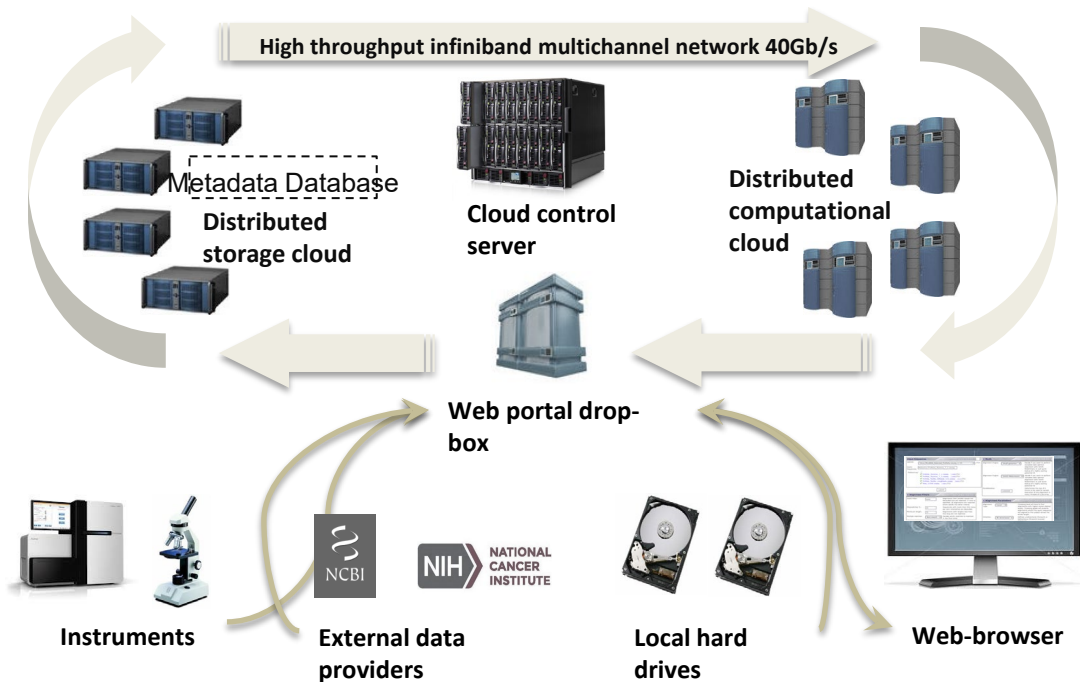
Python script (developed internally) to filter/merge/format results for output to user



HIVE Platform



- A **cloud-based environment** that comprises both a storage library of data and a powerful computing capacity.
- Can **consume, digest, analyze, manage, and share** all this data.



HIVE Interface

The screenshot shows the configuration page for the "AMR Short Reads Pipeline". The left sidebar contains a navigation menu with sections: Parameters (General, System, Advanced), Progress, and Results. The main content area is divided into tabs: General, System, and Advanced. The "General" tab is active, showing the following configuration options:

- Name: [text input]
- Pipeline Input Parameters: [Warning icon]
- Input Reads:
 - Read 1: [dropdown]
 - Read 2: [dropdown]
- Comprehensive Antibiotic Resistance Database (CARD):
 - CARD Reference Genome json: [dropdown]
 - CARD Reference Genome nucleotide fasta: [dropdown]
 - CARD Reference Genome protein fasta: [dropdown]
 - ARO Index: [dropdown]
- SPAdes:
 - Careful:
 - IonTorrent:
- RGI BLAST:
 - Include Loose:
- RGI DIAMOND:
 - Include Loose:
- AMR Merge:
 - Minimum Nucleotide Length:
 - Minimum Nucleotide Percent Identity:
 - Minimum Number of Tools:
 - Minimum Protein Percent Identity:

A blue "SUBMIT" button is located at the bottom right of the configuration area.

The screenshot shows the "Service Help" page for the "AMR Short Pipeline". It features a blue question mark icon and the title "AMR Short Pipeline". Below the title is a subtitle: "Process short single-ended and paired reads for anti-microbial resistant genes".

The page contains the following text:

AMR Short Pipeline

This pipeline consists of 14 chained processes. The most notable programs within this pipeline are SPAdes, BLAST, BLASTX, RGI BLAST, RGI DIAMOND, and Hexagon. Output from these 6 programs are then processed by a Python script to return the desired outcome.

The main inputs for this pipeline are the read(s) and the CARD database, in both json and fasta formats.

For the database parameters, please go to the Official CARD Database and download the "CARD Data" of the desired version. Unzip the files, and upload the "nucleotide_fasta_protein_homolog_model.fasta"; "protein_fasta_protein_homolog_model.fasta" and "card.json", respectively.

Only short-reads sequencers for which Spades is recommended are Illumina and IonTorrent. Not others such as Helicos are recommended.

Output

The screenshot shows the "AMR Merge" results page. The top navigation bar includes: MAIN, HOME, HOME NEW, DATASETS, PORTAL, LINKS, PROJECTS, SERVICES, CONTACT, HELP, LOGOUT, REPORT. The main content area is titled "AMR Merge" and shows "Step 14" for "enterococcus_faecalis_shortread".

The "Available Files to Download" section displays a table with the following data:

download	archive	name	size
		[All files]	23.42 KB
		amr-merge-output.txt	14.01 KB
		amr-merge-output.txt.compressed.txt	9.42 KB

The left sidebar contains a navigation menu with sections: Parameters, Progress (Main Progress), Results (Available Files to Download), and What's Next?

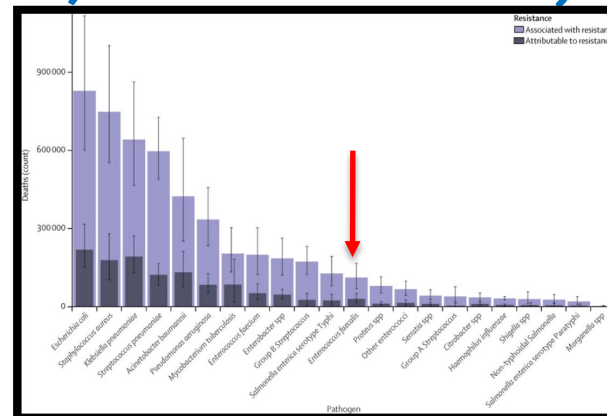


Output (*Staphylococcus aureus*)



ARO	Gene	Protein Product	Protein Accession	Drug Class	Resistance Mechanism	Tool	Length	Pident	Bitscore	Number of Tools Identifying Potential AMR Gene
3000839	arlS	conserved hypothetical protein	ABD30512.1	acridine dye;disinfecting agents and intercalating dyes;fluoroquinolone antibiotic	antibiotic efflux	Blastx	451	100	812	5
3000026	mepA	MepA	AAU95768.1	glycylcycline;tetracycline antibiotic	antibiotic efflux	Blastx	451	100	885	5
3004661	fosB	fosfomycin resistance protein FosB	EHS19134.1	fosfomycin	antibiotic inactivation	Hexagon	420	97.38		5
3005097	blaZ	beta lactamase protein BlaZ	CBZ41939	penam	antibiotic inactivation	Blast	405	99.01		5
3000617	mecA	MecA	AGC51118.1	carbapenem;cephalosporin;cephamycin;monobactam;penam	antibiotic target replacement	Hexagon	2007	99.85		5
3000250	ErmC	macrolide-lincosamide-streptogramin B-resistance protein	AAA98296.1	lincosamide antibiotic;macrolide antibiotic;streptogramin antibiotic	antibiotic target alteration	Hexagon	735	99.86		5
3004572	LmrS	drug resistance transporter, EmrB/QacA subfamily	AAW38464.1	aminoglycoside antibiotic;diaminopyrimidine antibiotic;macrolide antibiotic;oxazolidinone antibiotic;phenicol antibiotic	antibiotic efflux	Blast	1443	99.86		5
3004667	norA	norA	BAA14147.1	fluoroquinolone antibiotic	antibiotic efflux	Hexagon	1167	99.91		5

Output (*Enterococcus faecalis*, *Blautia obeum*)



Murray, 2021, Lancet

Enterococcus faecalis

ARO	Gene	protein Product	Protein Accession	Drug Class	Resistance Mechanism	Tool	Length	Pident	Bitscore	Number of Tools Identifying Potential AMR Gene
3002875	dfrE		EOD99669.1	diaminopyrimidine antibiotic	antibiotic target replacement	Blastx	164	98.171	335	5
3003948	efrA		WP_104671188.1	fluoroquinolone antibiotic;macrolide antibiotic;rifamycin antibiotic	antibiotic efflux	RGI-Blast		99.83	1168.3	5
3003551	emeA	multidrug efflux pump	BAC11911.1	acridine dye;disinfecting agents and intercalating dyes	antibiotic efflux	Blast	736	99.46		5
3000300	IsaA	Lsa	AAO43110.1	lincosamide antibiotic;macrolide antibiotic;oxazolidinone antibiotic;phenicol antibiotic;pleuromutilin antibiotic;streptogramin antibiotic;tetracycline antibiotic	antibiotic target protection	Blastx	498	98.193	954	5
3000186	tetM	Tet(M)	BAB82500.1	tetracycline antibiotic	antibiotic target protection	Hexagon	1920	96.77		5

Blautia obeum

ARO	Gene	protein Product	Protein Accession	Drug Class	Resistance Mechanism	Tool	Length	Pident	Bitscore	Number of Tools Identifying Potential AMR Gene
	none									



Conclusion



- Antibiotic resistance is a serious global health concern
- **AMR pipeline** developed and optimized, hosted on the **HIVE** platform, and available to registered users
- Pipeline identified antibiotic resistant genes in agreement with current knowledge of the bacterial species analyzed
- Pipeline is ready to be utilized by FDA reviewers
 - Applies technology and scientific computing to provide **fast, efficient, consistent,** and **unbiased** results to validate that products are free from antibacterial resistant genes
 - User friendly, bioinformatics knowledge not needed

Availability



- sciHIVE
 - scihive.fda.gov
- regHIVE

AMR Short Reads Pipeline

General System Advanced

> Name []

> Pipeline Input Parameters

> Input Reads

> Read 1 []

> Read 2 []

> Comprehensives Antibiotic Resistance Database (CARD)

> CARD Reference Genome json []

> CARD Reference Genome nucleotide fasta []

> CARD Reference Genome protein fasta []

> ARO Index []

> SRRs

> Caret

> IonTorrent

> RGI BLAST

> Include Loose

> RGI DIAMOND

> Include Loose

> AMR Merge

> Minimum Nucleotide Length []

> Minimum Nucleotide Percent Identity []

> Minimum Number of Ticks []

> Minimum Protein Percent Identity []

SUBMIT

Service Help

? **AMR Short Pipeline**

Process short single-ended and paired reads for anti-microbial resistant genes

AMR Short Pipeline

This pipeline consists of 14 chained processes. The most notable programs within this pipeline are SRRs, BLAST, BLASTX, RGI, BLAST, RGI, DIAMOND, and Hmmer.

Output from these 6 programs are then processed by a Python script to return the desired outcome.

The main inputs for this pipeline are the read(s) and the CARD database, in both json and fasta formats.

For the database parameters, please go to the Official CARD Database and download the "CARD Data" of the desired version. Unzip the files, and upload the "nucleotide_fasta_protein_homolog_model.fasta", "protein_fasta_protein_homolog_model.fasta" and "card.json", respectively.

Only short-reads sequencers for which Spades is recommended are Illumina and IonTorrent. Not others such as Helicos are recommended.

Acknowledgements

Paul Carlson Group (FDA)

Deepika Prasad

HIVE (FDA)



Grace Alterovitz

Arya Eskandarian

Tigran Ghazanchyan

Anton Golikov

Hannah Howell

Konstantinos Karagiannis

Luis Santana-Quintero

Sean Smith

Thank you!





U.S. FOOD & DRUG
ADMINISTRATION