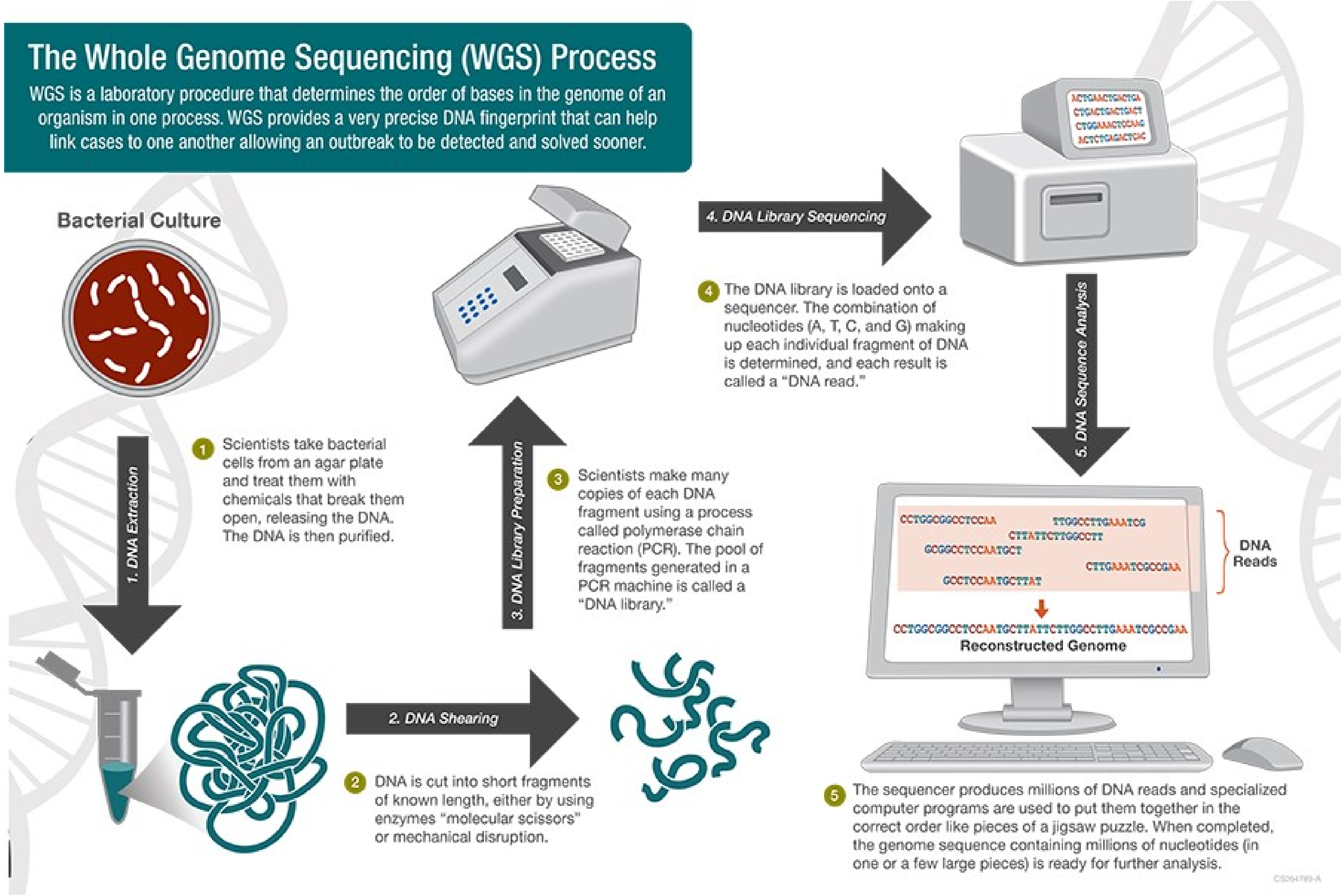
Ana Rita Rebelo
Technical University of Denmark

Review on WGS-based methods for AMR prediction

WGS-BASED ANALYSIS OF BACTERIA - OVERVIEW



The Whole Genome Sequencing (WGS) Process

WGS is a laboratory procedure that determines the order of bases in the genome of an organism in one process. WGS provides a very precise DNA fingerprint that can help link cases to one another allowing an outbreak to be detected and solved sooner.

① Scientists take bacterial cells from an agar plate and treat them with chemicals that break them open, releasing the DNA. The DNA is then purified.

② DNA is cut into short fragments of known length, either by using enzymes "molecular scissors" or mechanical disruption.

③ Scientists make many copies of each DNA fragment using a process called polymerase chain reaction (PCR). The pool of fragments generated in a PCR machine is called a "DNA library."

④ The DNA library is loaded onto a sequencer. The combination of nucleotides (A, T, C, and G) making up each individual fragment of DNA is determined, and each result is called a "DNA read."

⑤ The sequencer produces millions of DNA reads and specialized computer programs are used to put them together in the correct order like pieces of a jigsaw puzzle. When completed, the genome sequence containing millions of nucleotides (in one or a few large pieces) is ready for further analysis.

https://www.cdc.gov/pulsenet/php/infographics/index.html

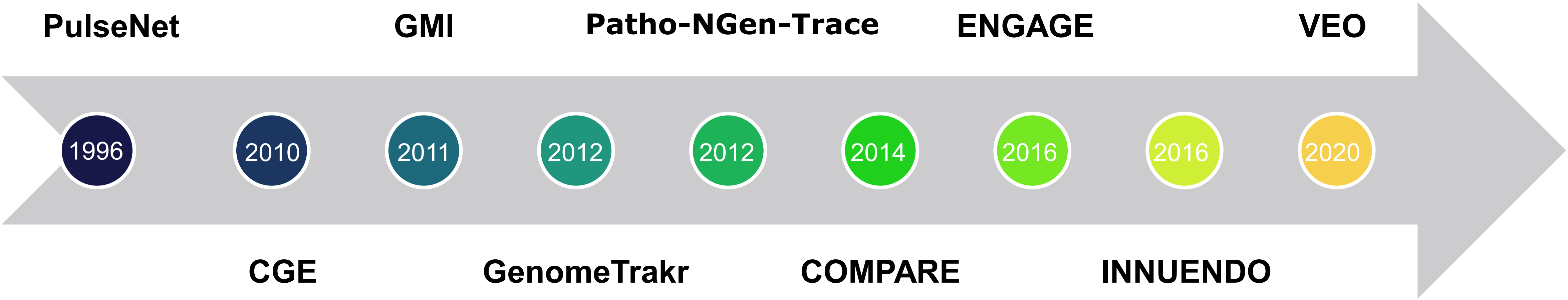
CURRENT EUROPEAN UNION AND INTERNATIONAL GUIDANCE

(ADAPTED FROM EURGEN-REFLABCAP)

ECDC	2016: “Expert opinion on whole genome sequencing for public health surveillance” 2016: “ECDC roadmap for integration of molecular and genomic typing into European-level surveillance and epidemic preparedness” 2019: “ECDC strategic framework for the integration of molecular and genomic typing into European surveillance and multi-country outbreak investigations”
EFSA	2019: “Whole genome sequencing and metagenomics for outbreak investigation, source attribution and risk assessment of food-borne microorganisms”
ISO	TBA: ISO/DIS 23418 standard “Microbiology of the food chain – Whole genome sequencing for typing and genomic characterization of foodborne bacteria – General requirements and guidance”
WHO	2018: “Landscape paper on whole genome sequencing for foodborne disease surveillance” 2020: “Global Antimicrobial Resistance and Use Surveillance System (GLASS) document on whole-genome sequencing for surveillance of antimicrobial resistance”

BIOINFORMATICS DEVELOPMENT INITIATIVES

(ADAPTED FROM EURGEN-REFLABCAP)



PulseNet **GMI** **Patho-NGen-Trace** **ENGAGE** **VEO**

1996 2010 2011 2012 2012 2014 2016 2016 2020

CGE **GenomeTrakr** **COMPARE** **INNUENDO**

CGE: Centre for Genomic Epidemiology

GMI: Global Microbial Identifier

VEO: Versatile Emerging Infectious Disease Observatory

WGS-BASED ANALYSIS OF BACTERIA - REQUIREMENTS

- Expertise on DNA extraction methods
- Expertise on library preparation methods

Not too technically demanding
Ideally a dedicated room

- Access to sequencing platform
- Access and expertise on bioinformatics tools

Main challenges: cost, implementation

- Data management infrastructure

Main challenges: cost, compatibility

WGS-BASED ANALYSIS OF BACTERIA - QUALITY CONTROL

Many different:

- DNA extraction kits
- Sequencing platforms
- Bioinformatics approaches
- Bioinformatics tools

Is harmonization feasible?

Well defined set of QC parameters

- For the raw data
  - *E.g. nr. and length of raw reads, depth of coverage*
- For the assembled genomes
  - *E.g. N50, nr. of contigs, genome size*
- For the performance of the tools
  - *E.g. accurately detect PMs and ARGs in sets of benchmarking data*

BIOINFORMATICS TOOLS - OVERVIEW

Purpose

Quality control
Assembly
Taxonomic analysis
Phylogeny
Serotyping/Subtyping
Detection of AMR determinants
Detection of other determinants

Accessibility

Web-based
Command line
Local vs. server

Maintenance

Benchmarked
Curated

Data

Raw data as input
Assemblies as input
Integration of metadata

Cost

Open access
Subscription

BIOINFORMATICS TOOLS FOR PREDICTION OF AMR

Tool	Target species	Reference database	Output	Comments
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ARIBA	All	CARD, ResFinder, ARG-ANNOT, MEGARes, NCBI Bacterial AMR Reference Gene Database, PubMLST, or others defined by user	Reference sequences and respective % of identity, % of coverage	Also phylogeny, VFs and plasmids, depending on provided databases (such as plasmidfinder, VFDB, VirulenceFinder)
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ResFinder	All	Own	ARGs and respective % of identity, % of coverage, position in genome, predicted phenotype	NA
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
RGI	All	CARD	Reference sequences and respective % of identity, % of coverage, other options	Integrated in the Galaxy server; allows proteome analysis
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
SSTAR	All	Own (resGANNOT created by merging ResFinder and ARG-ANNOT)	Reference sequences and respective % of coverage, depth	Can be used with other reference databases
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
CARD	All	Own	ARGs and point mutations, respective prevalence and predicted phenotype	Highly focused on ontology and standardization. VFs and mobile genetic elements currently being added

BIOINFORMATICS PIPELINES FOR PREDICTION OF AMR

Tool	Target species	Reference database	Output	
[redacted]	[redacted]	[redacted]	[redacted]	
Enterobase	Limited	Tools’ databases	Genome assembly and annotation, serotyping, MLST, cgMLST, rMLST, phylogenetic analysis	
[redacted]	[redacted]	[redacted]	[redacted]	
NCBI Pathogen Detection	Limited	Own	ARGs, VFs, SNP-based phylogenetic analysis	
[redacted]	[redacted]	[redacted]	[redacted]	
Ridom SeqSphere+	All	Own, tools’ databases; Includes NCBI AMRFinder and VFDB	Assemblies, QC, taxonomy, ARGs, VFs, MLST, cgMLST, phylogenetic analysis	€
[redacted]	[redacted]	[redacted]	[redacted]	€

LIMITATIONS OF WGS-BASED PREDICTION OF AMR

Limitations:	**Future perspectives:**
• Can’t predict phenotypes	Genotype-phenotype correlation studies
• Can’t detect new mechanisms of resistance	Sentinel surveillance Machine learning
• Can’t account for over-expression mechanisms	Transcriptomics / proteomics

CONCLUSIONS

There’s no one-size-fits-all approach
There’s no perfect bioinformatics tool

Essential requirements:

QC
Benchmarking datasets
→ Ensuring accuracy of your approach

Knowing which database supports each tool
Understanding how the tool works
→ Knowing what data to provide
Knowing the limitations of your results

THANK YOU ON BEHALF OF THE FWD AMR-REFLABCAP

ANRIRE@FOOD.DTU.DK