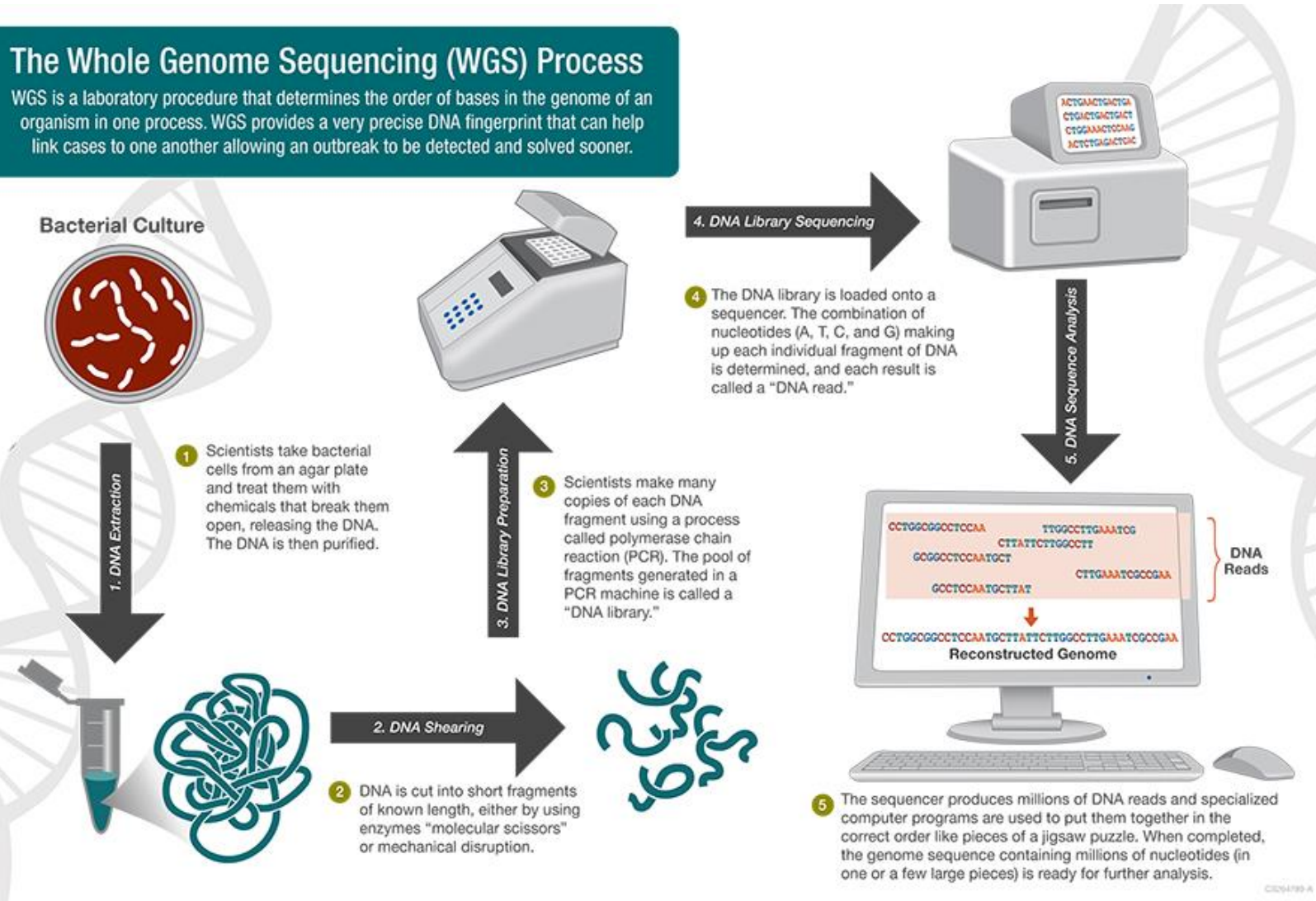


Ana Rita Rebelo
Technical University of Denmark

Review on WGS-based methods for AMR prediction

WGS-BASED ANALYSIS OF BACTERIA - OVERVIEW



CURRENT EUROPEAN UNION AND INTERNATIONAL GUIDANCE

(ADAPTED FROM EURGEN-REFLABCAP)



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"



2019: "Whole genome sequencing and metagenomics for outbreak investigation, source attribution and risk assessment of food-borne microorganisms"



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"

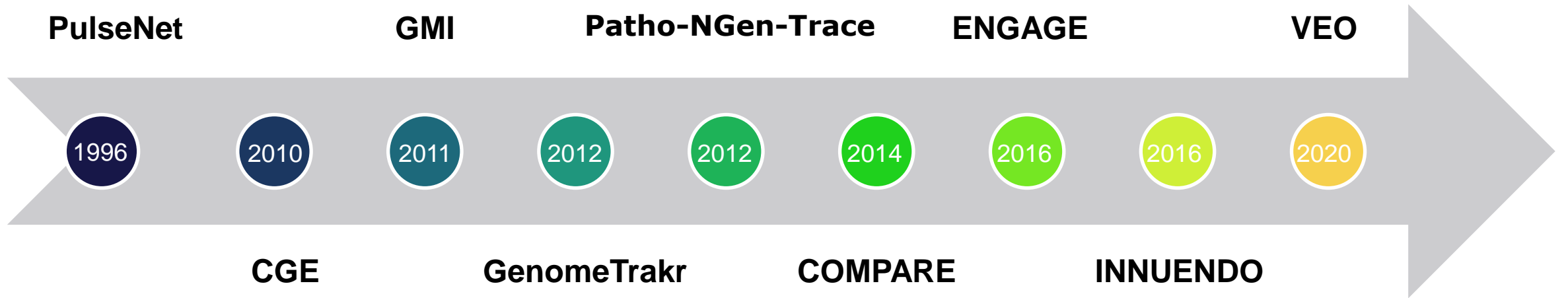


2018: "Landscape paper on whole genome sequencing for foodborne disease surveillance"

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)



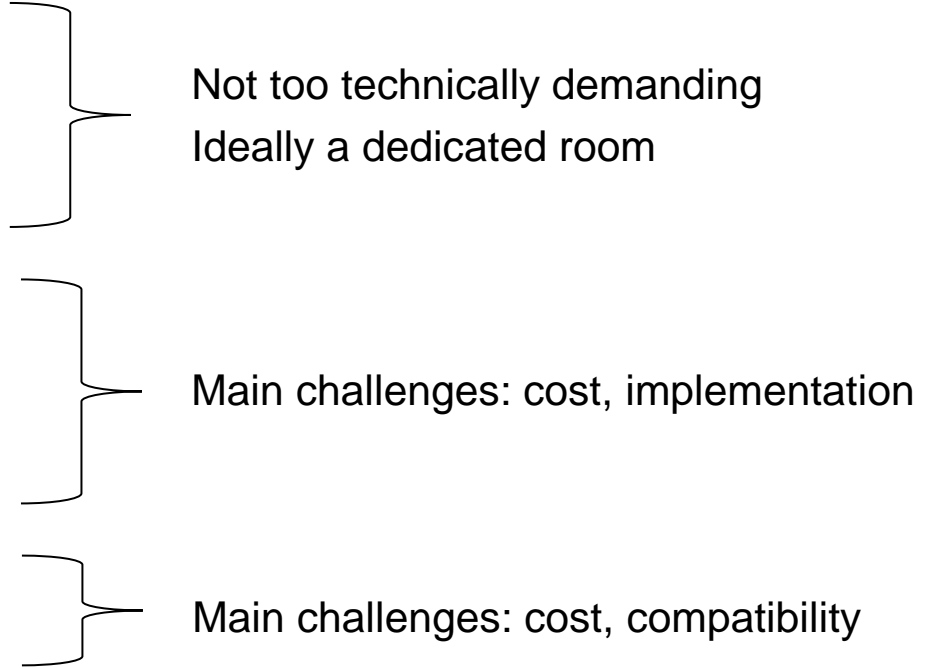
CGE: Centre for Genomic Epidemiology

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
- Access to sequencing platform
- Access and expertise on bioinformatics tools
- Data management infrastructure



Not too technically demanding
Ideally a dedicated room

Main challenges: cost, implementation

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
SRST2	All	CARD, PubMLST, or others defined by user	Reference sequences and respective % of coverage, depth	Also taxonomy, phylogeny, VFs, plasmids, other, depending on provided databases
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)
KmerResistance	All	Own	Reference genomes, ARGs and respective % of identity, % of coverage	Also taxonomy
ResFinder	All	Own	ARGs and respective % of identity, % of coverage, position in genome, predicted phenotype	NA
PointFinder	Limited	Own	Mutated gene, protein translation, predicted phenotype	Included in ResFinder but can be used locally by itself. Currently under development for Klebsiella spp.
RGI	All	CARD	Reference sequences and respective % of identity, % of coverage, other options	Integrated in the Galaxy server; allows proteome analysis
AMRFinder; AMRFinderPlus	All (Limited PMs)	NCBI RefSeq	Reference sequences and respective % of identity, % of coverage	Included in NCBI Pathogen Detection
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
ABRicate	All	CARD, ResFinder, ARG-ANNOT, MEGARES, NCBI AMRFinderPlus, or others defined by user	Reference sequences and respective % of identity, % of coverage	Also VFs and plasmids, depending on provided databases Also VFs and plasmids, depending on provided databases (such as plasmidfinder and VFDB)
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
Pathogenwatch	Limited	Own, tools' databases	Taxonomy, MLST, cgMLST and clustering. Other functionalities for Klebsiella spp. derived from Kleborate
Enterobase	Limited	Tools' databases	Genome assembly and annotation, serotyping, MLST, cgMLST, rMLST, phylogenetic analysis
BIGSdb	All	PubMLST BIGSdb	Annotation, taxonomy, ARGs, plasmids, MLST, rMLST, cgMLST, phylogenetic and spatio-phylogenetic analysis, comparative genomics
NCBI Pathogen Detection	Limited	Own	ARGs, VFs, SNP-based phylogenetic analysis
PATRIC	All	Own, but also includes others such as CARD, NDARO and VFDB	Assemblies, QC, annotation, taxonomy, ARGs, phenotype prediction, VFs, mobile elements, phylogenetic analysis, variation analysis, genome alignments, comparative genomics, other options
Ridom SeqSphere+	All	Own, tools' databases; Includes NCBI AMRFinder and VFDB	Assemblies, QC, taxonomy, ARGs, VFs, MLST, cgMLST, phylogenetic analysis
Bionumerics	All (wgMLST schemes available for limited species)	Own, tools' databases, others provided by user	Assemblies, QC, annotation, taxonomy, ARGs, plasmids, MLST, rMLST, cgMLST, wgMLST, phylogenetic analysis, comparative genomics, other options for E. coli (ARGs, PMs, VFs, plasmids, serotypes)



LIMITATIONS OF WGS-BASED PREDICTION OF AMR

Limitations:

- Can't predict phenotypes
- Can't detect new mechanisms of resistance
- Can't account for over-expression mechanisms



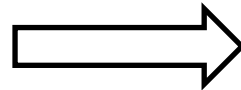
Future perspectives:

- Genotype-phenotype correlation studies
- Sentinel surveillance
Machine learning
- Transcriptomics / proteomics

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

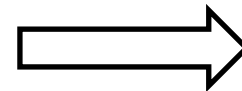
Essential requirements:

QC
Benchmarking datasets



Ensuring accuracy of your approach

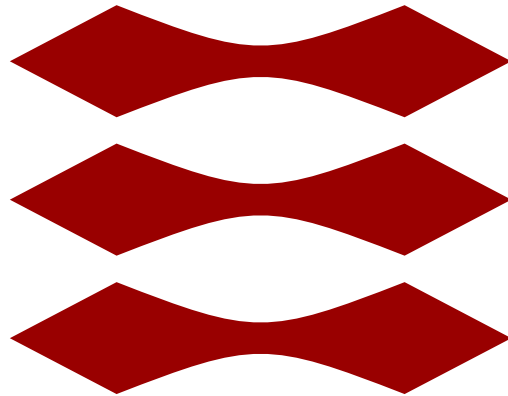
Knowing which database supports each tool
Understanding how the tool works



Knowing what data to provide
Knowing the limitations of your results

THANK YOU ON BEHALF OF THE FWD AMR-REFLABCAP

DTU



ANRIRE@FOOD.DTU.DK