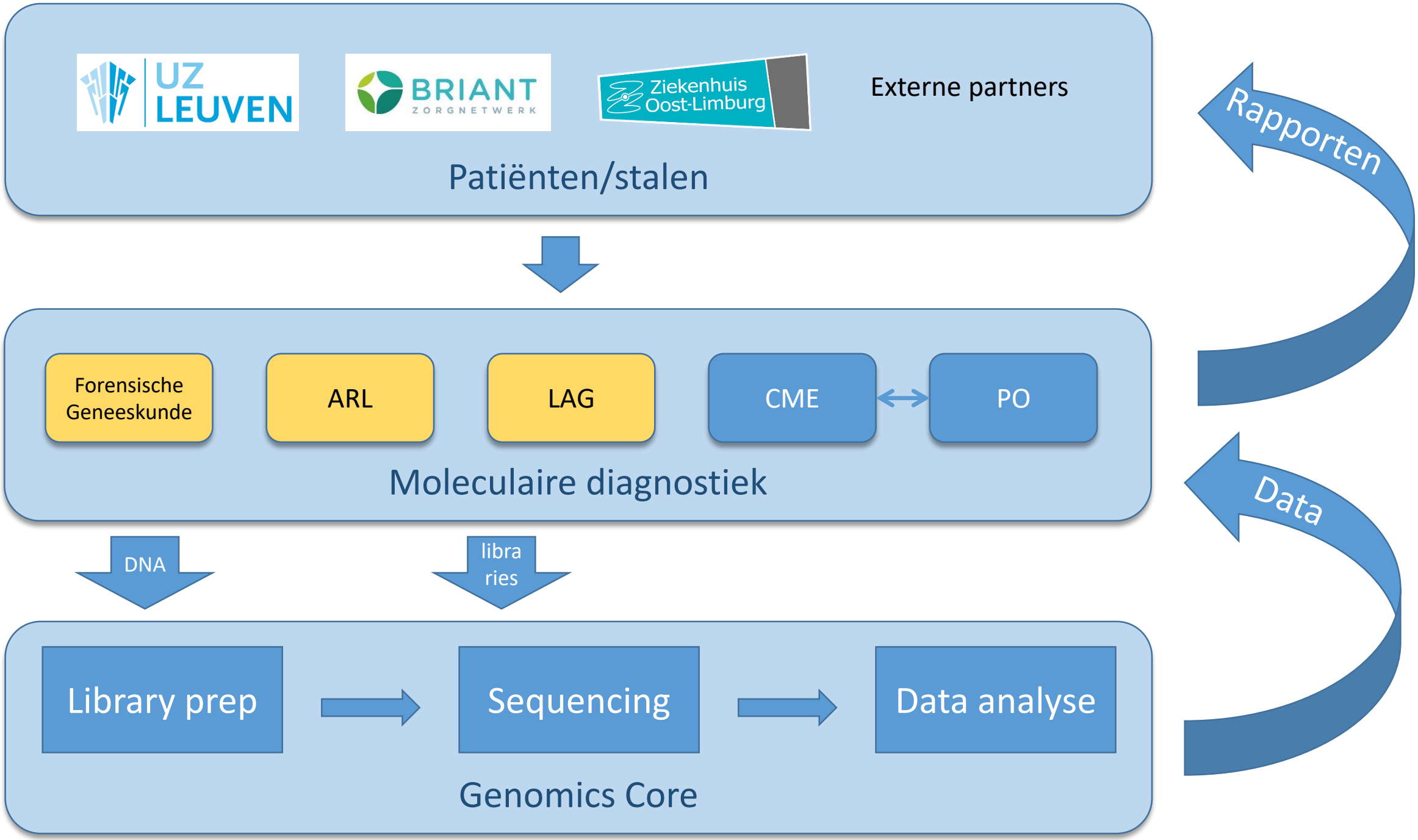
NGS is microbiology

Wouter Bossuyt

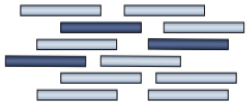
Genomics Core, Leuven

What’s the use of technology if it is not usable?


UZ LEUVEN
BRIANT ZORGNETWERK
Ziekenhuis Oost-Limburg
Externe partners
Patiënten/stalen
Rapporten
Forensische Geneeskunde
ARL
LAG
CME
PO
Moleculaire diagnostiek
Data
DNA
libra ries
Library prep
Sequencing
Data analyse
Genomics Core

NGS workflow

Genomic DNA



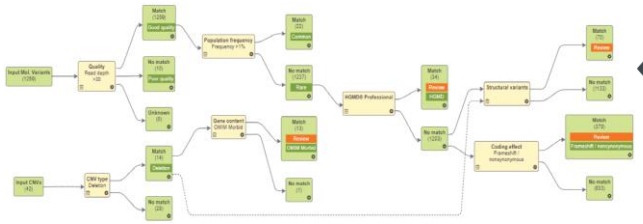
Fragments

Library

DNA sequencing

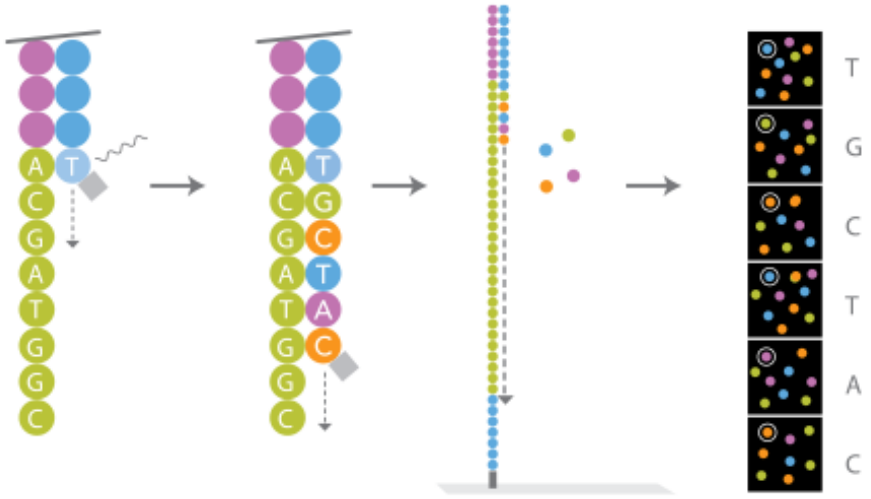
AGGTCGTTACGTACGCTAC
GACCTACATCAGTACATAG
GCATGACAAAGCTAGGTGT

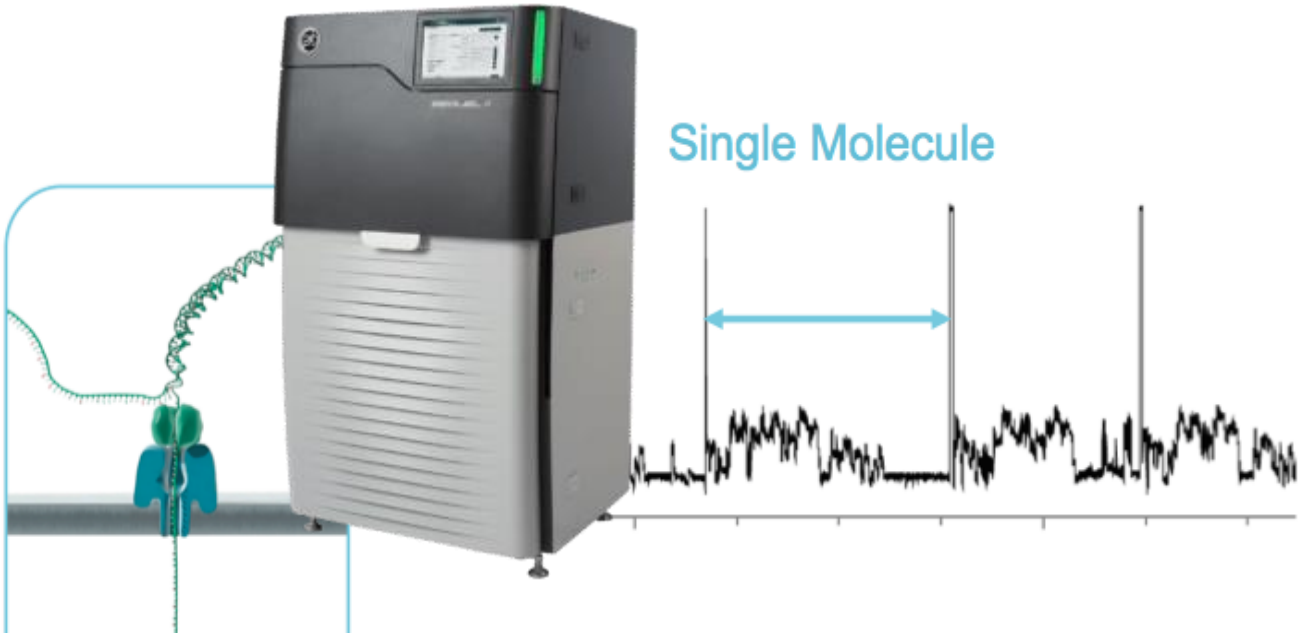
Mapping, alignment, variant calling



Variant filtering & interpretation

Two main technologies





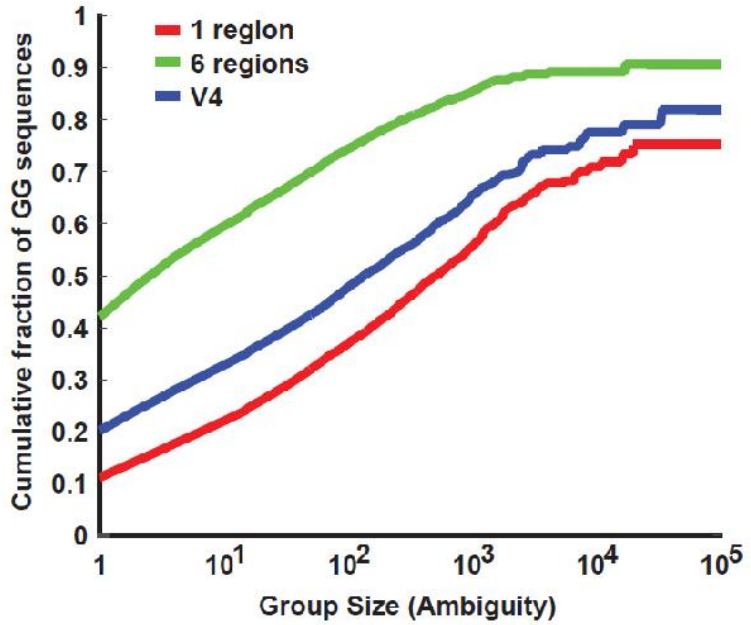
applications

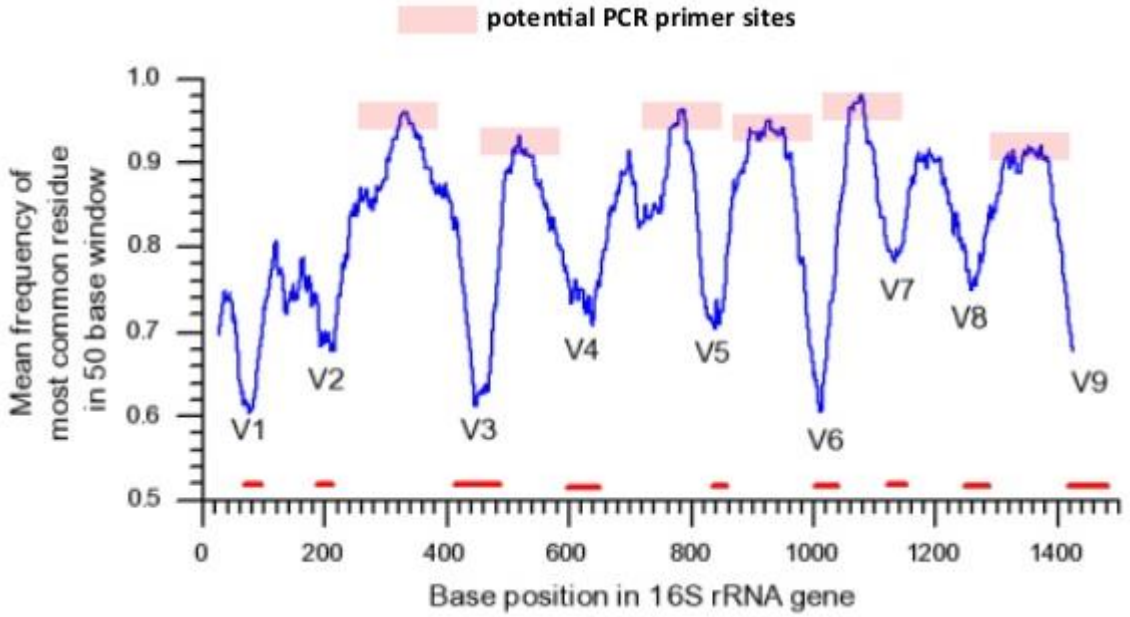
- Targeted amplicon sequencing
  - 16s
  - NGS MLST
- Bacterial WGS
- Metagenomics
- outbreak monitoring

- Virus discovery and detection

Targeted approach: 16S RNA typing

- 16S typing
  - Fast analysis of microbial species in sample





Targeted approach: NGS MLST

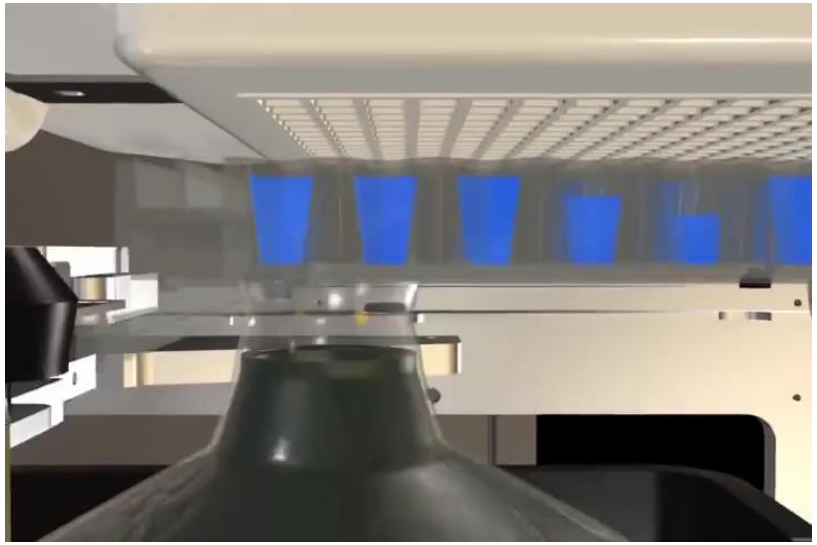
- Multilocus serotyping using NGS
- Multiplex PCR for 7 housekeeping genes → multiplex samples → sequence
  - Gene list dependent on taxonomy of interest
  - Sequence on long read sequencer
  - Depth of coverage important
- Limited info on resistance/virulence

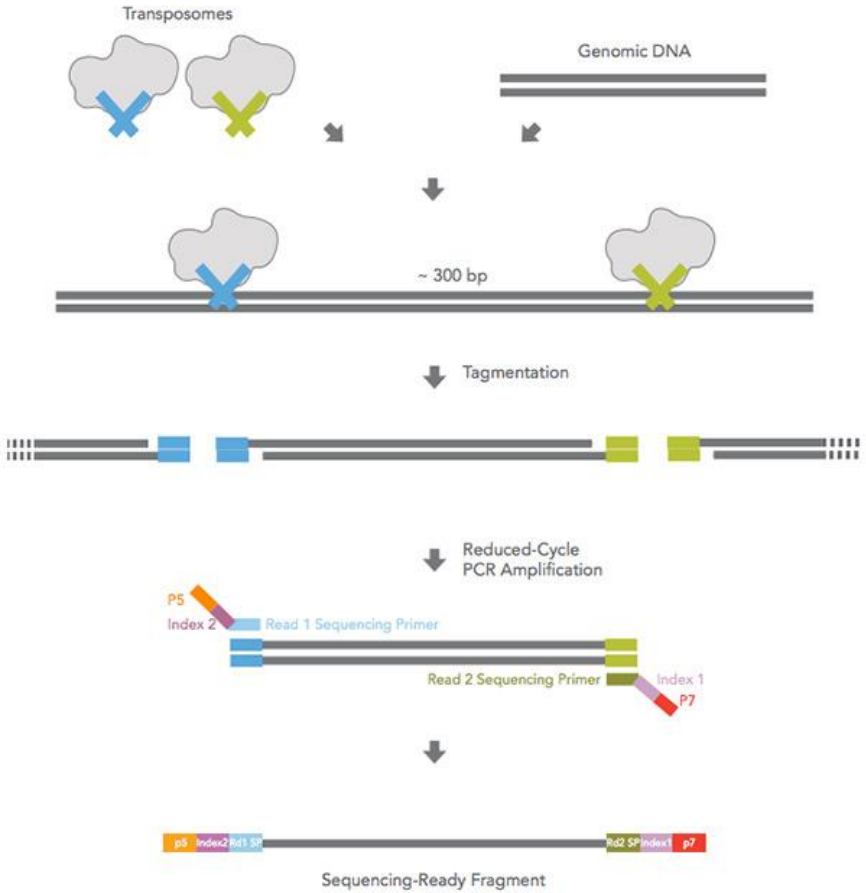
Bacterial WGS

- Full sequencing of bacterial genomes
- Culture bacteria and isolate DNA
- Complete information
  - Typing
  - Pathogenecity monitoring during outbreak
  - Resistance monitoring
    - Identify known resistance mutations
    - Culture-based resistance monitoring in parallel to build more knowledge
  - *de novo* assembly to identify structural variants
  - Evolutionary tracking

Automated library prep for bacterial WGS

- Nextera XT on Echo Labcyte
  - Proven method
  - Scaled to very low volumes
  - In 384-well plates



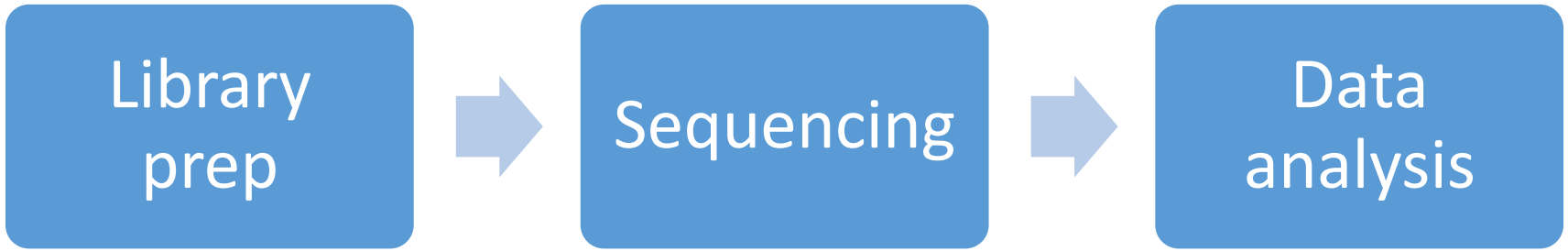


Limited adoption of bacterial WGS

- Whole genome sequencing of bacterial genomes:
  - gives a lot of information
  - can combine many other tests in one test
- No large scale adoption
- What is holding people back?
  - Cost
  - Scalability
  - TAT
  - Knowhow

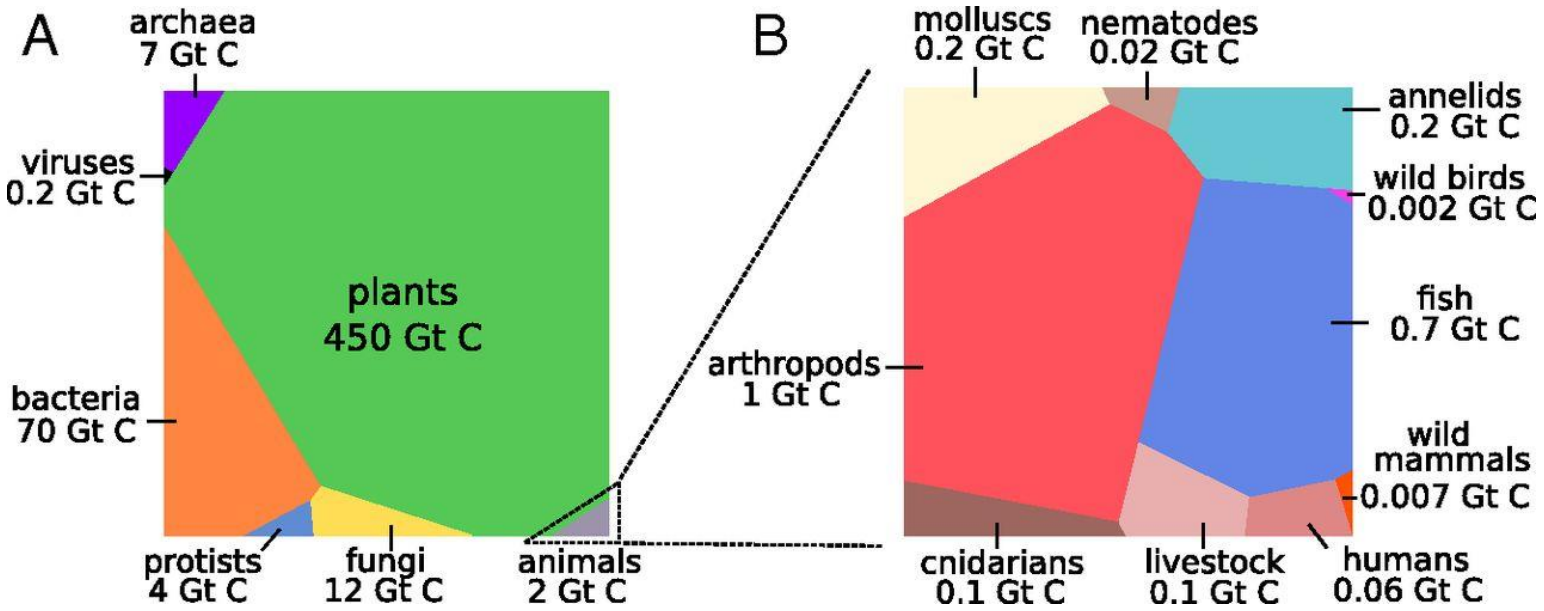
Optimization of bacterial WGS workflow

- Possibility of automated sample submission
- LIMS integration
- Reservation and scheduling of necessary equipment
- Workflow optimization



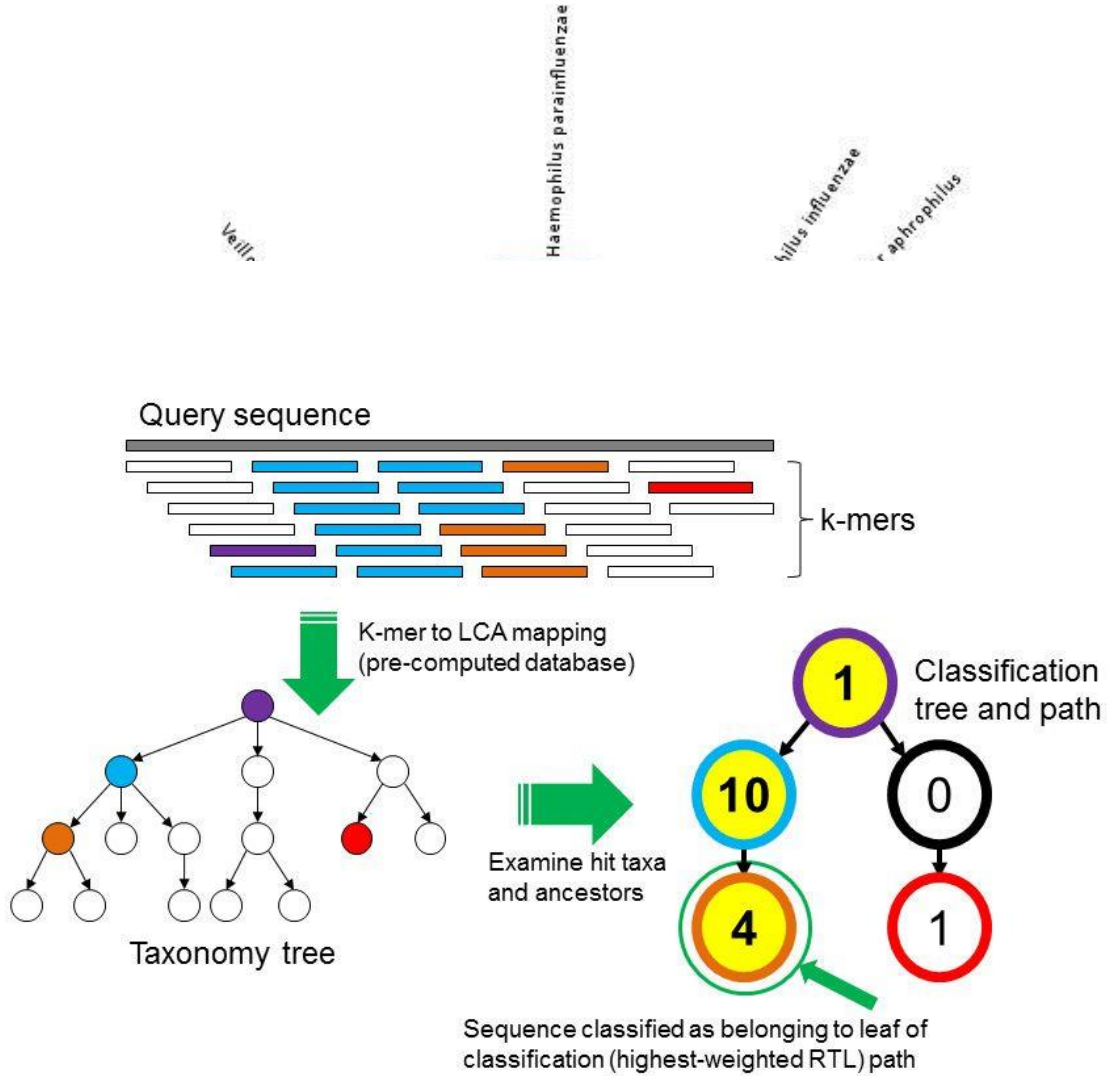
Metagenomics

- Get DNA from a ROI → sequence
- Identification of different species in an 'environmental' sample



Metagenomics

- Kraken as a tool for species identification
  - Identification of species
  - Quantification of species
- Possibility to select only reads from species of interest for further analysis

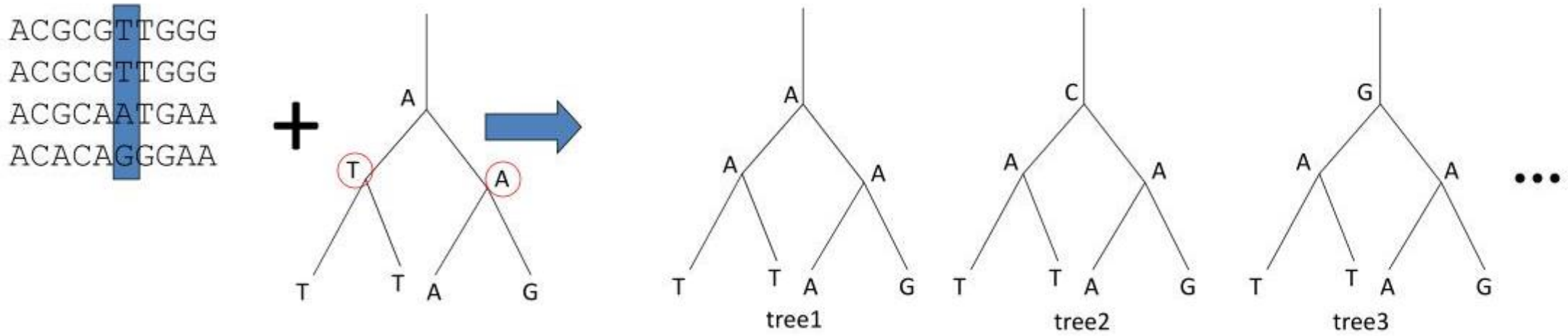


Outbreak monitoring

Metagenomics approach

1. Identify different bacterial species
   - Continue with likely species one at the time
2. Identify SNPs in sequence of closely related species
3. Use maximum likelihood approach to assemble phylogenetic tree

Maximum likelihood



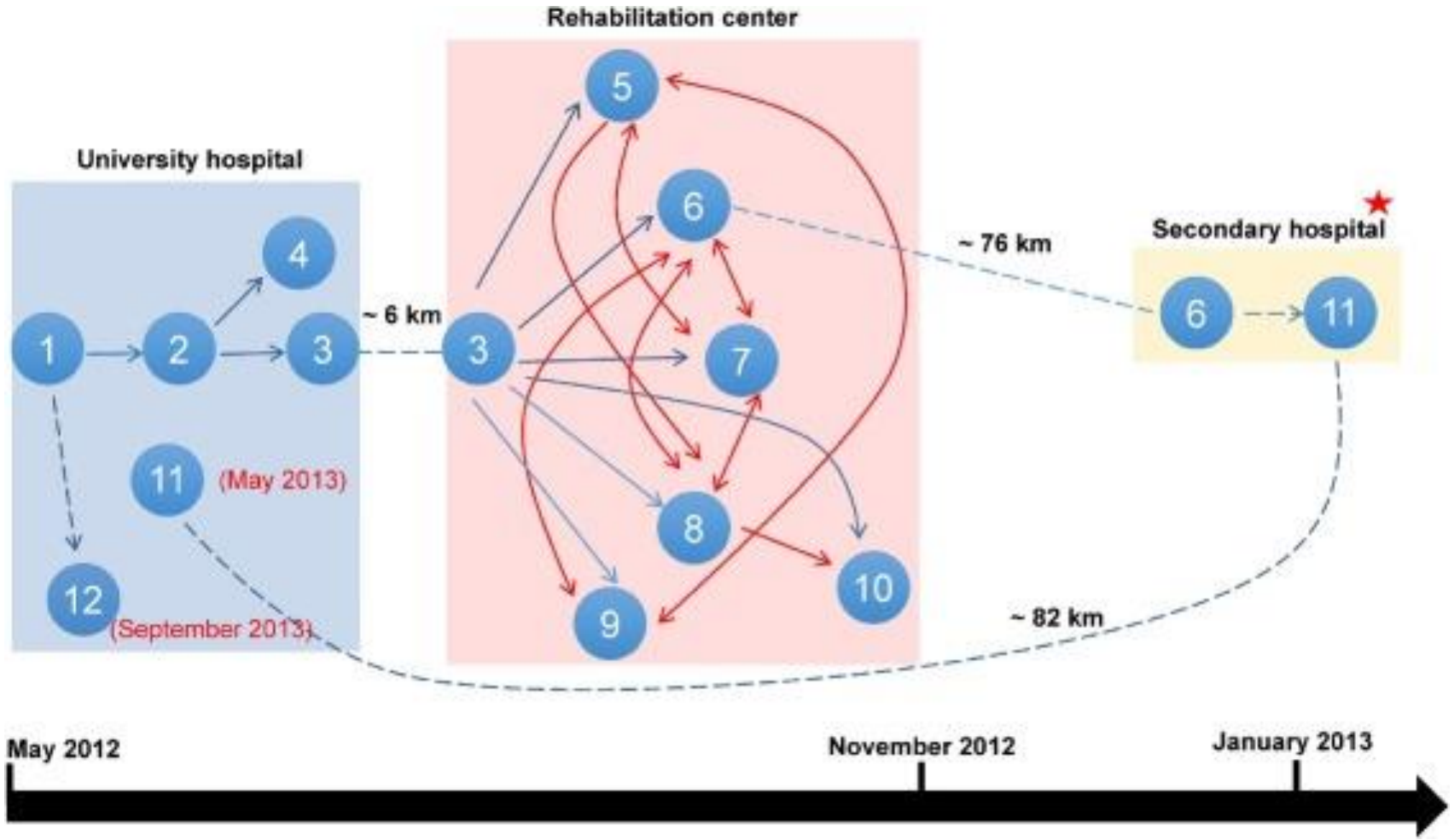
1. Probability is sum of probability of each branch
2. Calculate probability for each <u>possible</u> tree
3. Choose phylogenetic tree with maximum likelihood

Outbreak monitoring

Metagenomics approach

1. Identify different bacterial species
   - Continue with likely species one at the time
2. Identify SNPs in sequence of closely related species
3. Use maximum likelihood approach to assemble phylogenetic tree
4. Align phylogenetic relations ships with epidemiological evidence

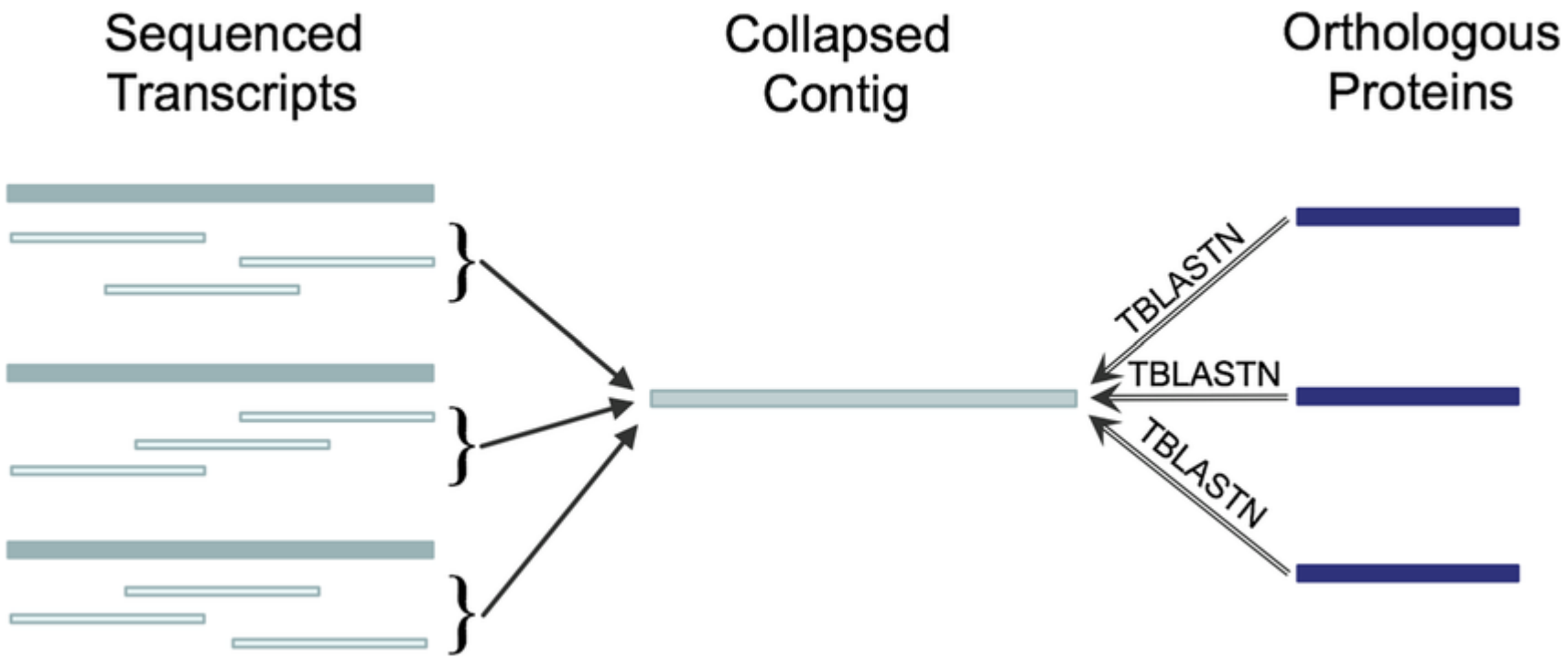
Route of transmission

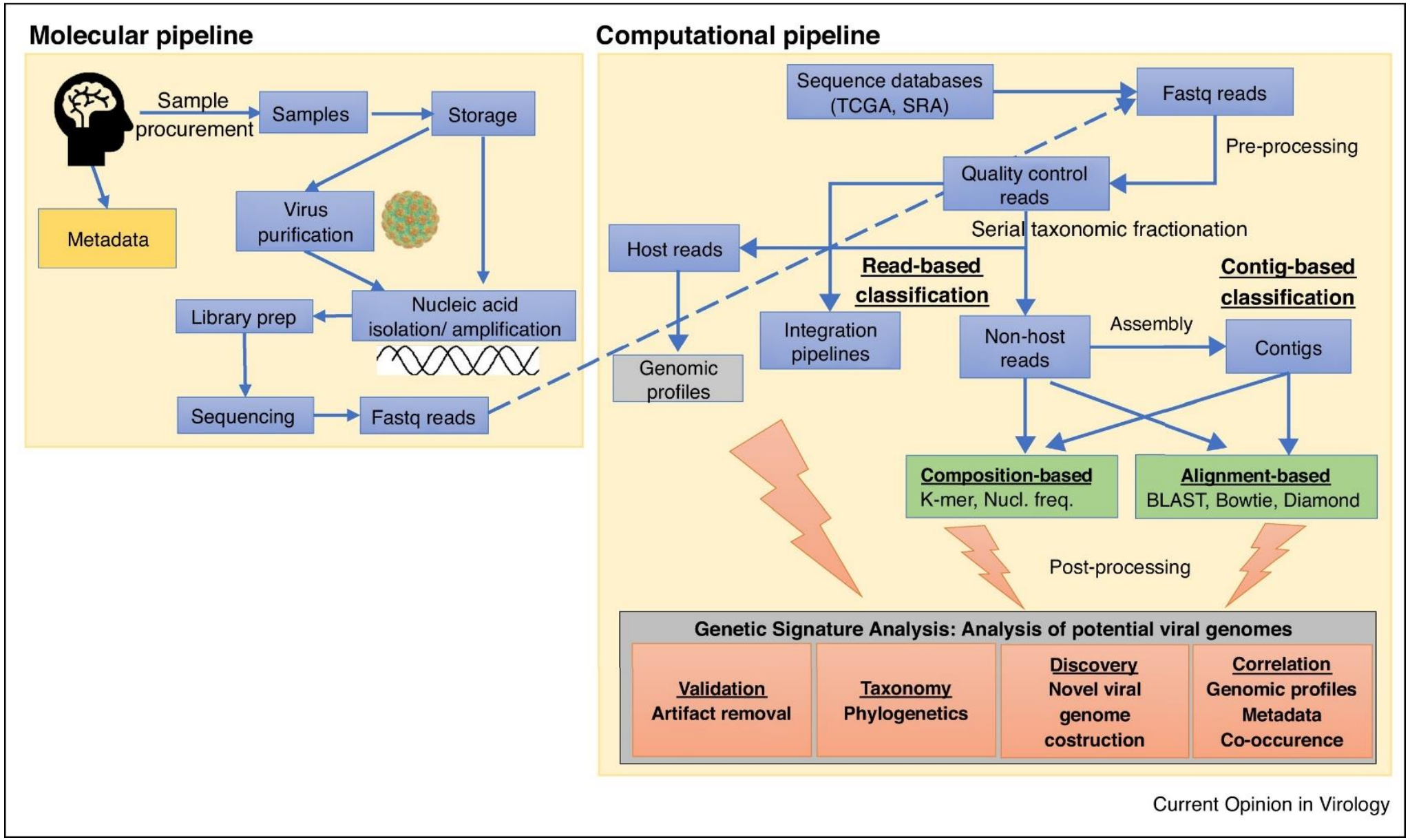


Viral metagenomics

- Huge diversity!
- We can’t yet guess the number of virus species
- Viral genomes are frequently less conserved
- Identification needs other approach
    1. Substract host reads first: except integration sites!
    2. Assembly of viral contigs, starting from viral gene database
    3. Tblastn to compare with viral database: less sensitive to variation

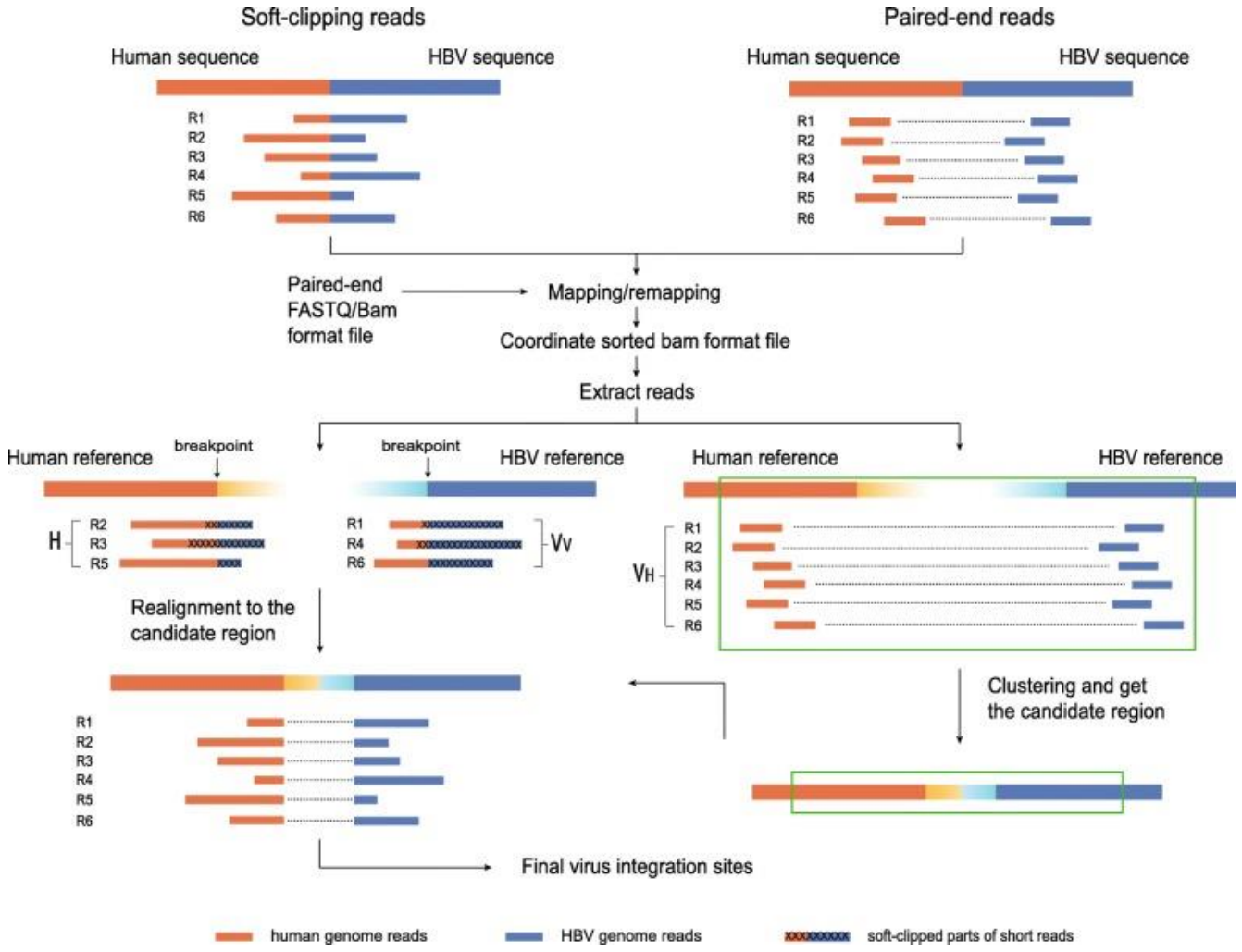
Reducing search space in variable genomes




Molecular pipeline
Computational pipeline
Sample procurement
Samples
Storage
Metadata
Virus purification
Nucleic acid isolation/ amplification
Library prep
Sequencing
Fastq reads
Sequence databases (TCGA, SRA)
Fastq reads
Pre-processing
Quality control reads
Serial taxonomic fractionation
Host reads
Read-based classification
Contig-based classification
Integration pipelines
Non-host reads
Assembly
Contigs
Genomic profiles
Composition-based
K-mer, Nucl. freq.
Alignment-based
BLAST, Bowtie, Diamond
Post-processing
Genetic Signature Analysis: Analysis of potential viral genomes
Validation
Artifact removal
Taxonomy
Phylogenetics
Discovery
Novel viral genome costruction
Correlation
Genomic profiles
Metadata
Co-occurence
Current Opinion in Virology

Viral integration site detection

- Sequence tissues with expected viral integration
- Use split-read info and repetitive softclip-mapping cycles
- Determine insertion site



Contamination checks

- Purified/cultured samples are sometimes contaminated
- Metagenomics tools added to existing NGS pipelines:

What other DNA do we have in here?

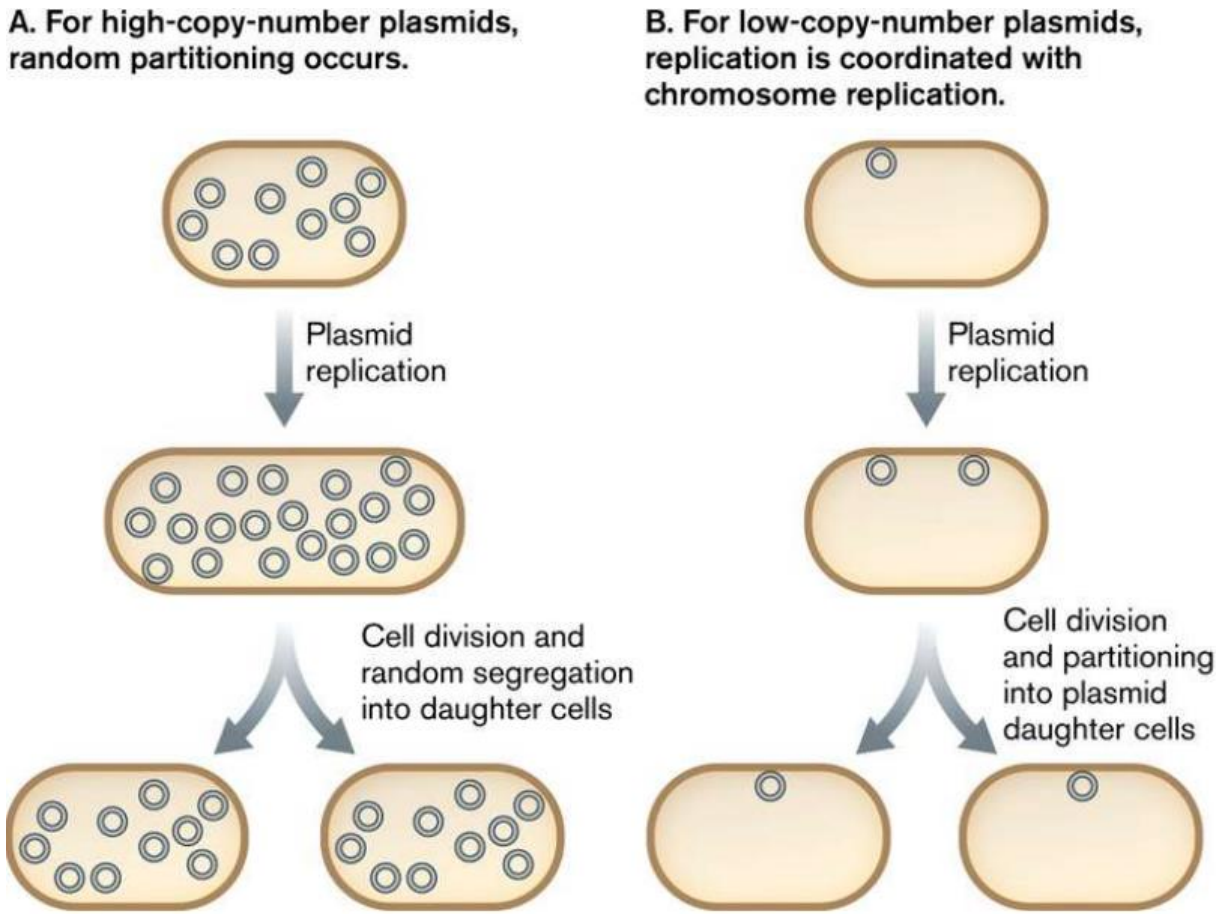
- Quality control for entire workflow
- Early warning system for contamination
- Potential reduction in contamination cultures

Making it difficult

- Repetitive elements
  - Difficult to sequence
  - Low mappability
  - Problematic for *de novo* assembly

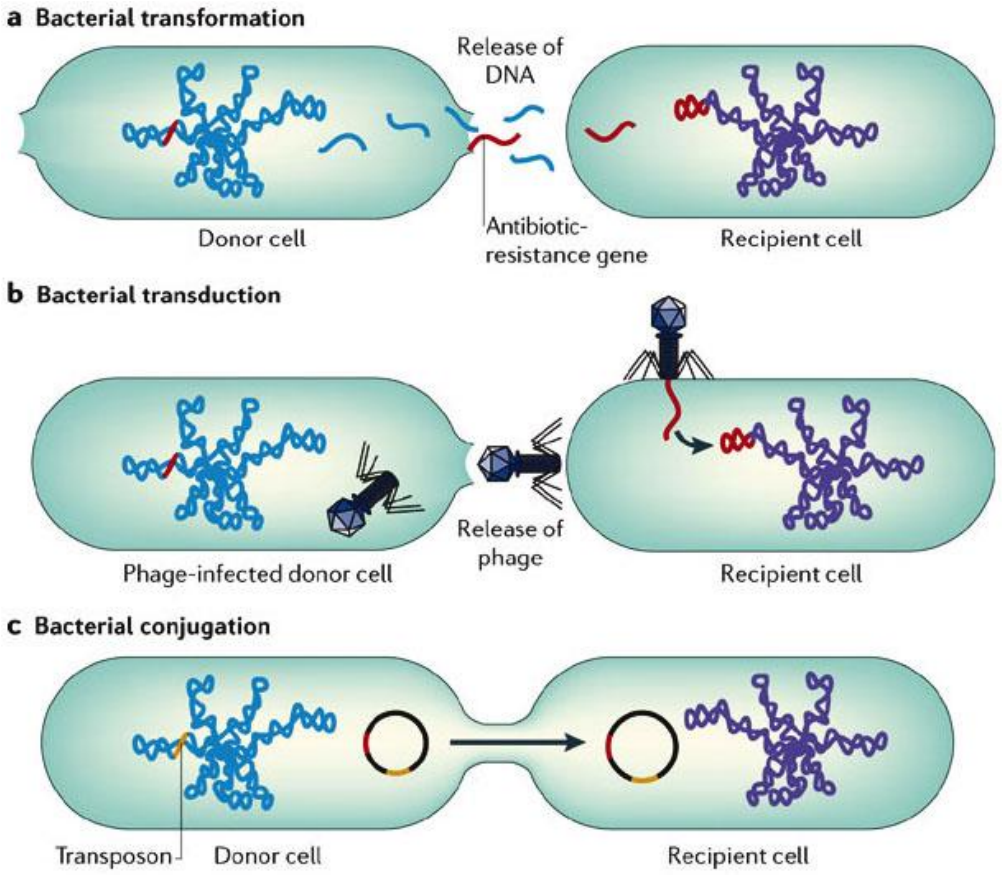
Making it difficult

- Repetitive elements
- Plasmids
  - Abundance variable
  - Presence variable



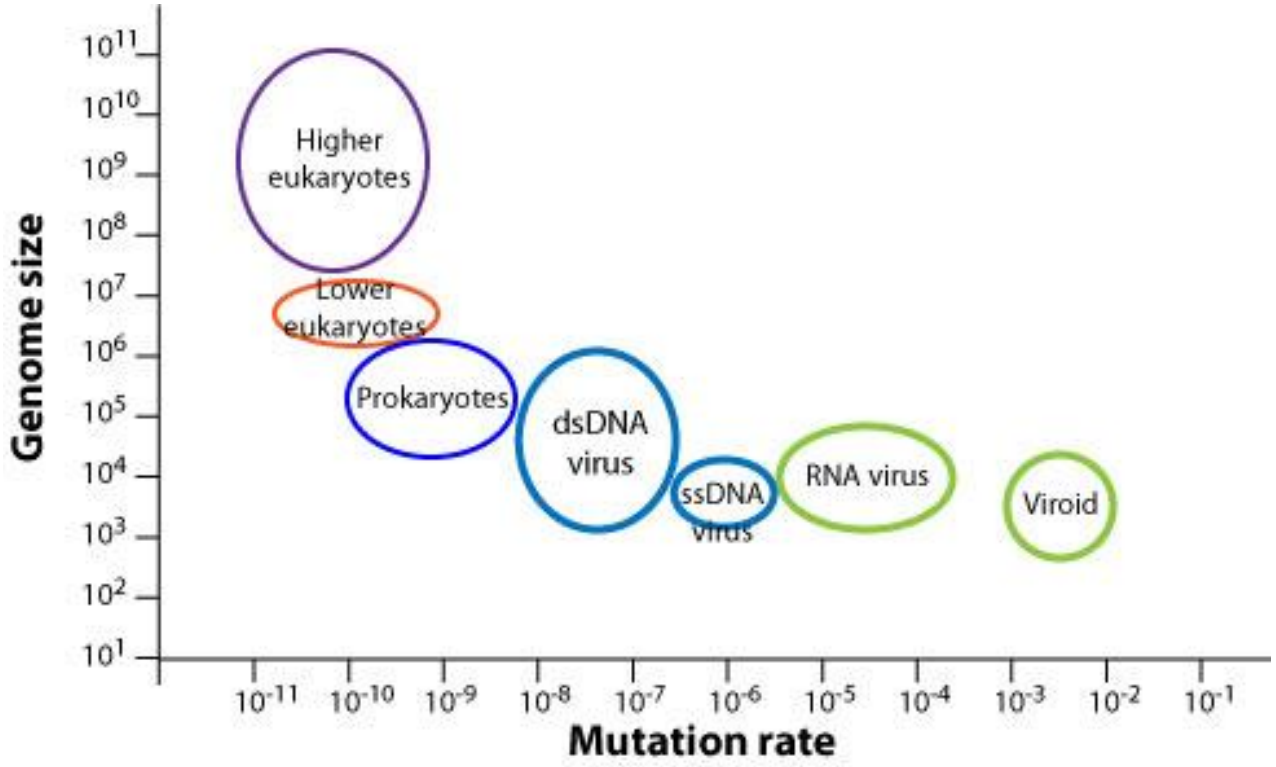
Making it difficult

- Repetitive elements
- Plasmids
- Horizontal gene transfer
  - Transformation/conjucation/transduction
  - Species identification more difficult
  - Phylogeny gives conflicting results
  - Route of transmission difficult to trace

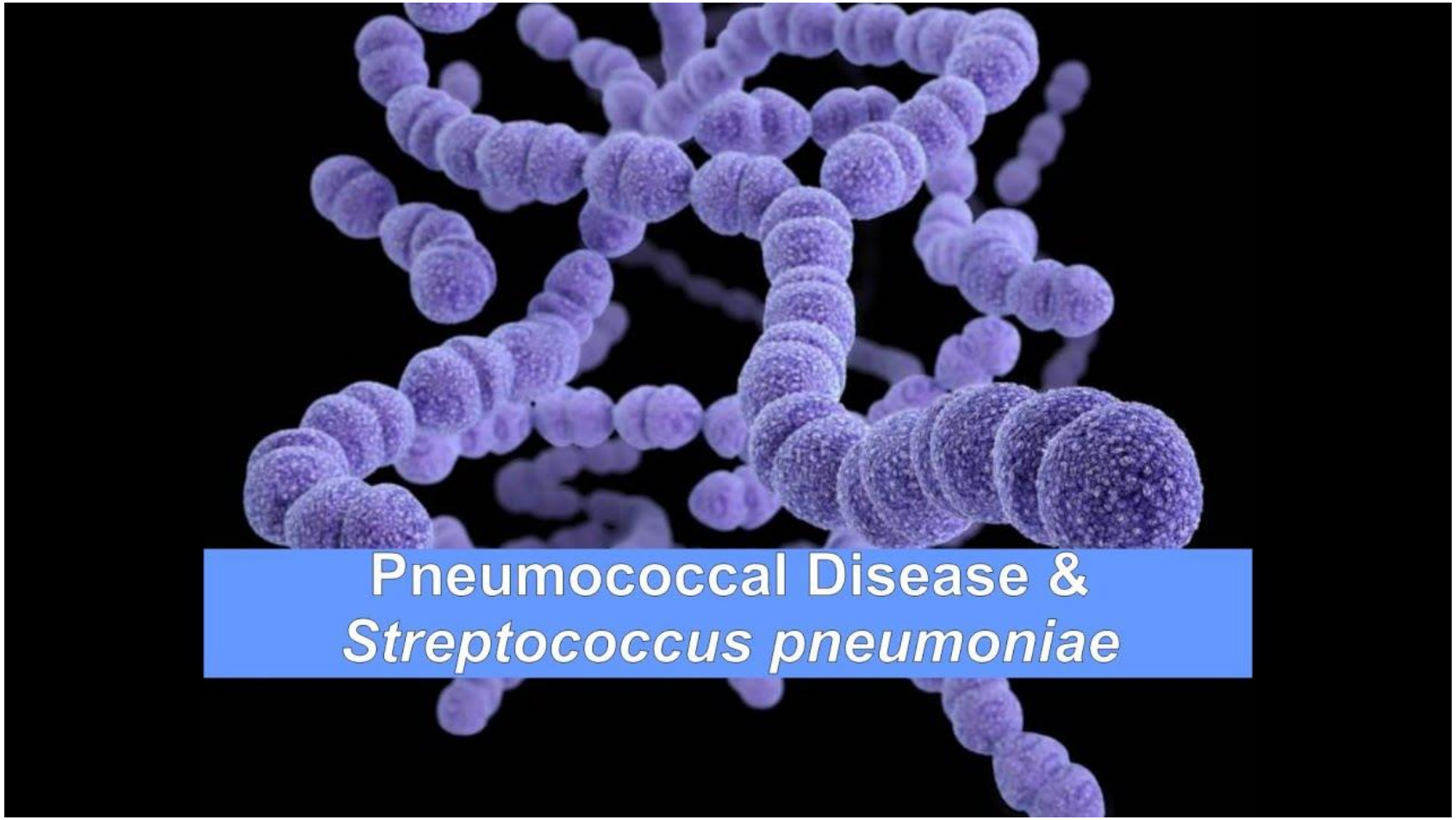


Making it difficult

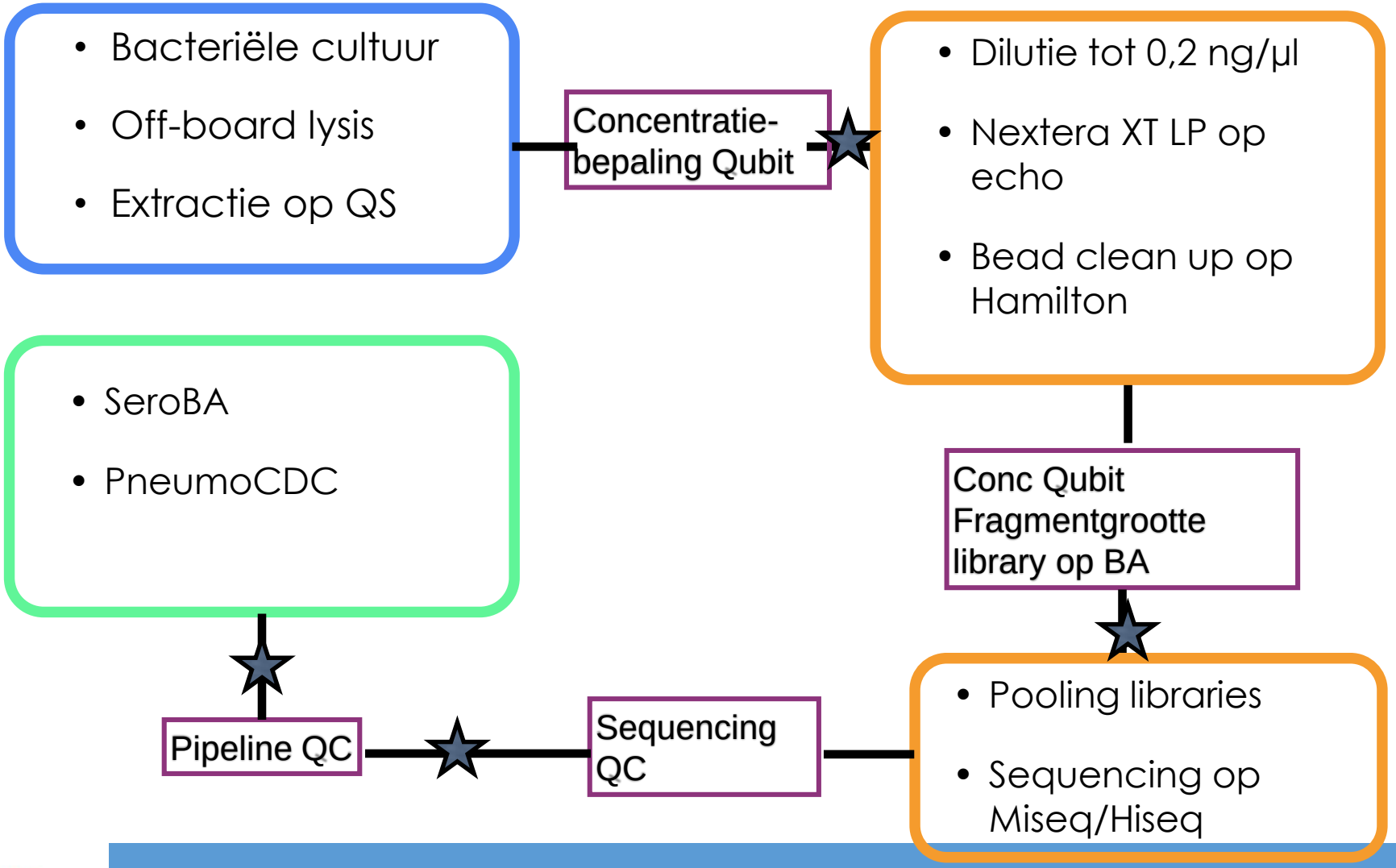
- Repetitive elements
- Plasmids
- Horizontal gene transfer
- High recombination/mutation rates



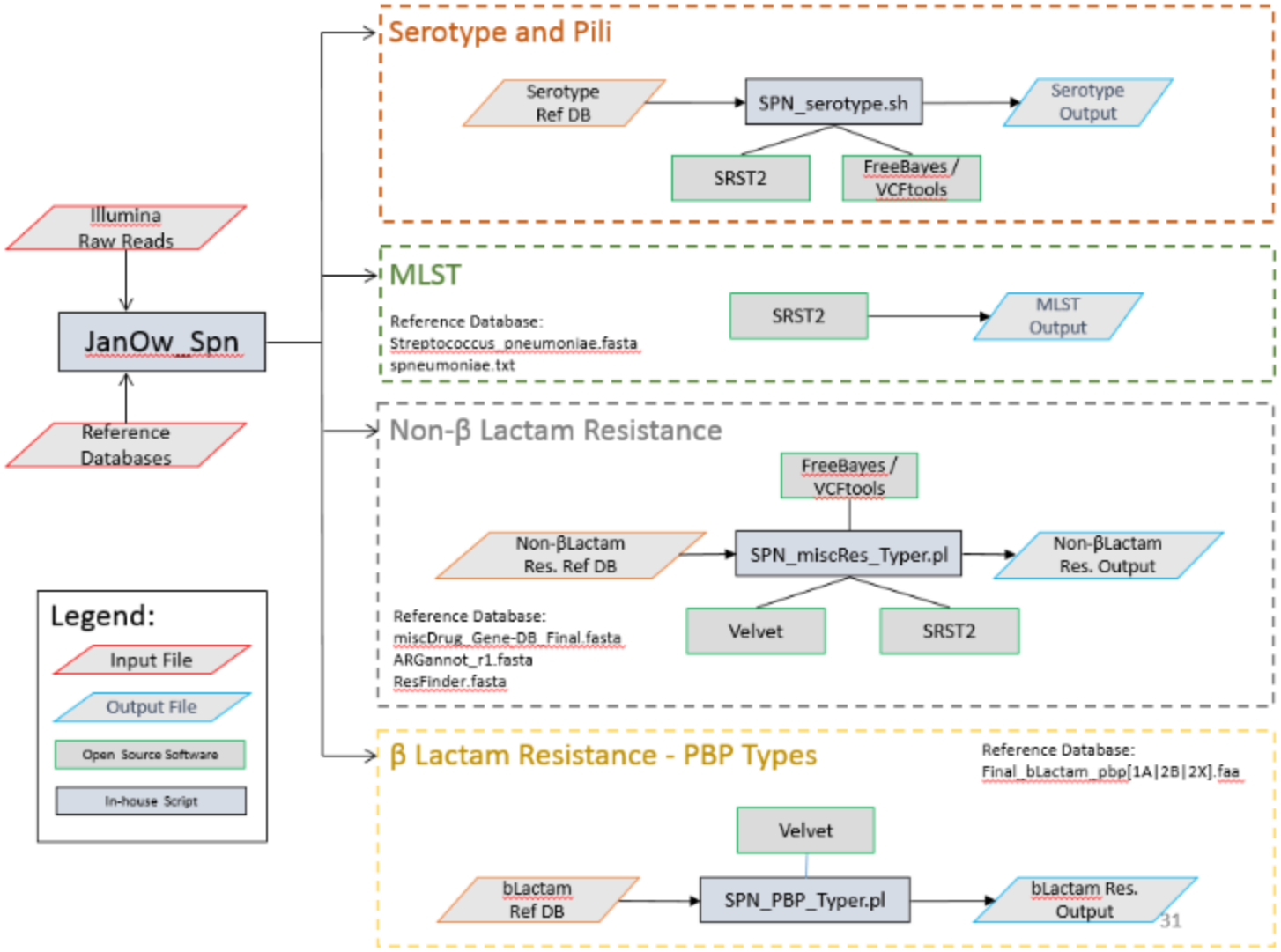
Practical example



SPN Workflow



CDC pipeline - Sero/Pili, MLST, AB resistance

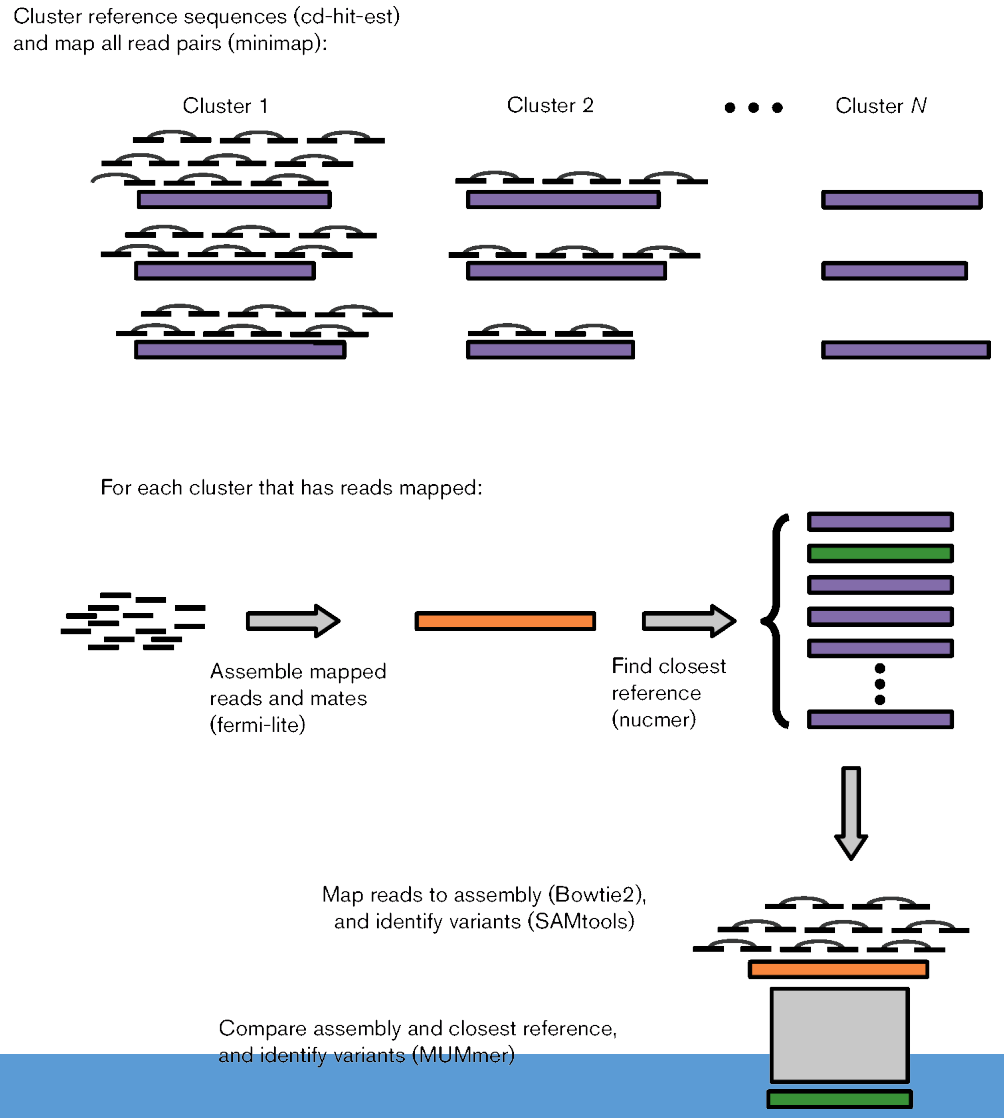


Metcalf BJ et al, Clin Microb Infec. 2016; 22(1)

SeroBA - Serotyping

ARIBA
Antimicrobial Resistance
Identification By Assembly

mapping/alignment and targeted
local assembly approach



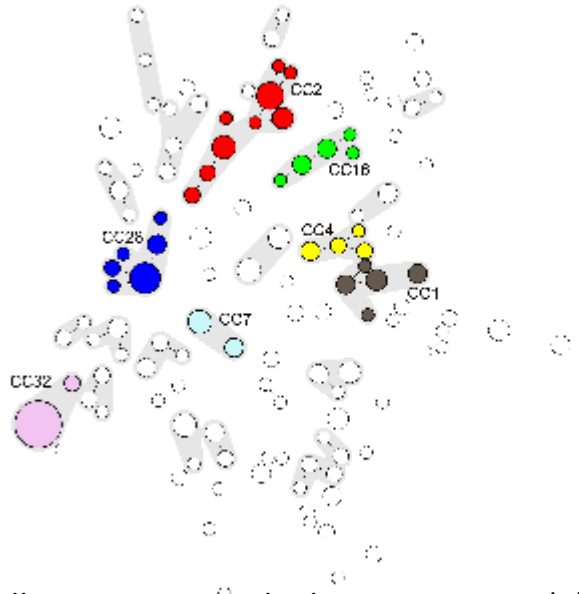
Hunt M et al. Microb Genom. 2017; 3(10)

MLST typing

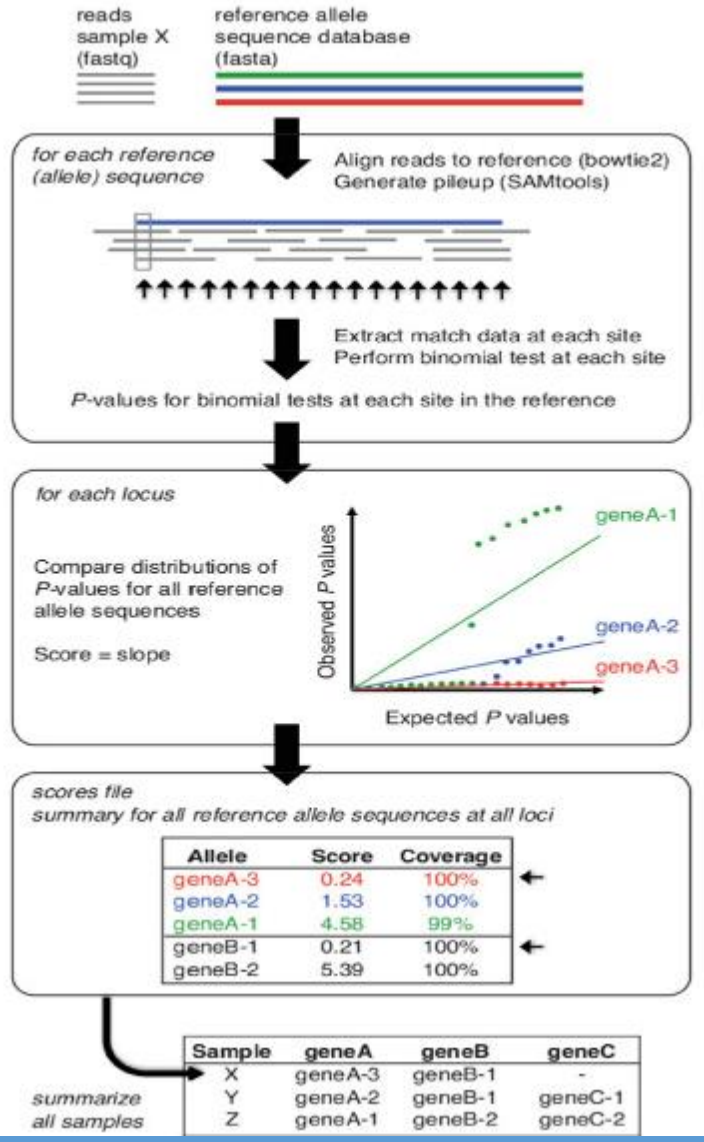
Isolate characterization by multilocus sequence typing

Based on 7 housekeeping genes

SRST2 Short Read Sequence Typing for Bacterial Pathogens



Elberse K E M et al, Plos One 2011; 6(5)



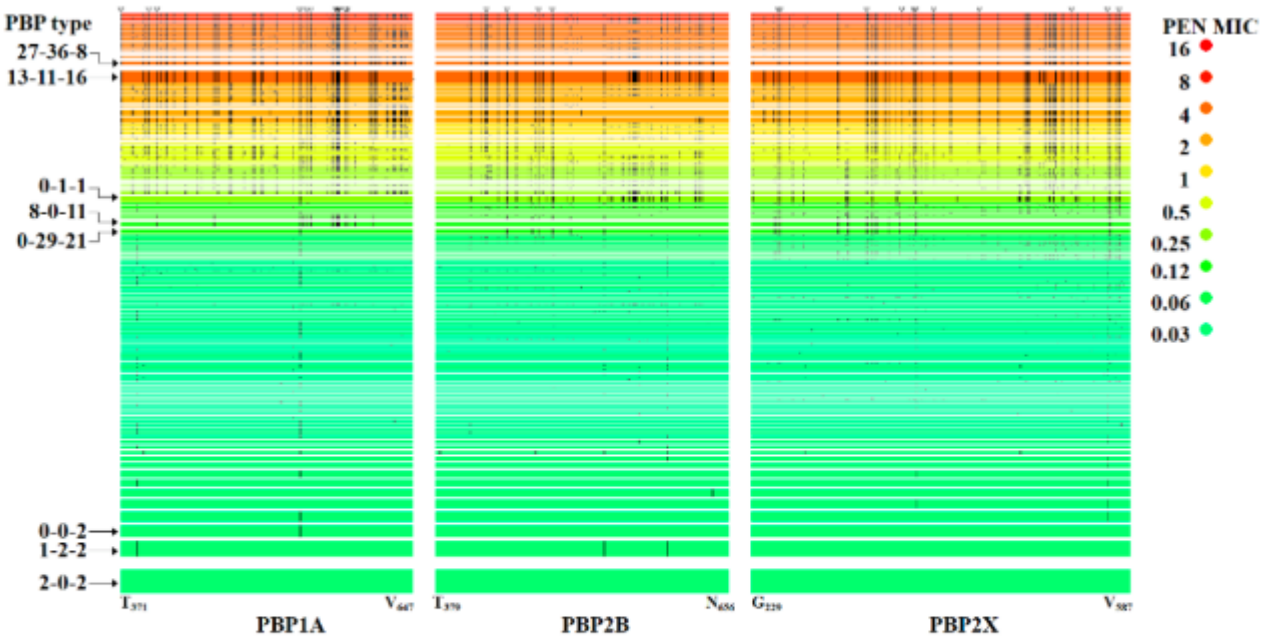
Inouye M et al, Genome Med 2014; 6(11)

Antimicrobial resistance – β-Lactam AB

Mosaic genes – de novo assembly
Random forest model - PBP genes to MIC

	S	I	R
WGS_PEN_SIR_Meningitis	**≤0.06**	-	≥0.12
WGS_PEN_SIR_Nonmeningitis	**≤2**	4	≥8
WGS_AMO_SIR	**≤2**	4	≥8
WGS_MER_SIR	**≤0.25**	0.5	≥1
WGS_TAX_SIR_Meningitis	**≤0.5**	1	≥2
WGS_TAX_SIR_Nonmeningitis	**≤1**	2	≥4
WGS_CFT_SIR_Meningitis	**≤0.5**	1	≥2
WGS_CFT_SIR_Nonmeningitis	**≤1**	2	≥4
WGS_CFX_SIR	**≤0.5**	1	≥2

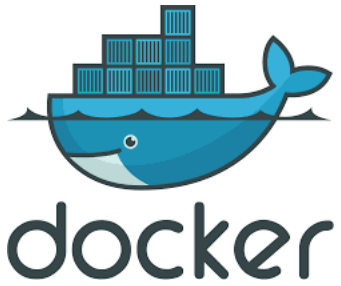
PEN: penicillin; AMO: amoxicillin; MER: meropenem
TAX: cefotaxime; CFT: ceftriaxone; CFX: cefuroxime



Li, Y. *et al.*, MBio. 2016 Jun 14;7(3). pii: e00756-16.
Li, Y. *et al.*, BMC Genomics. 2017 Aug 15;18(1):621.
Metcalf, BJ. *et al.* Clin Microbiol Infect. 2016 Jan;22(1):60.e9-60.e29.

Automated cloud computing

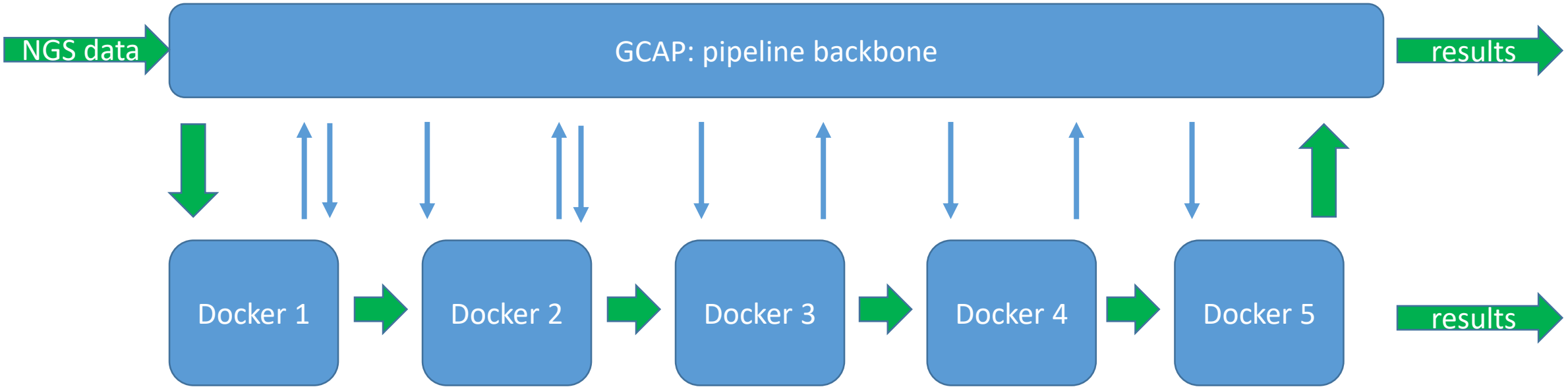
- Bring pipeline to the cloud using Docker



Docker is an open platform for developing, shipping, and running applications.

- Automated transfer of data upon sequencing
- Initiates different docker modules in the cloud
- Data and results stored and shareable through the cloud

Cloud computing approach

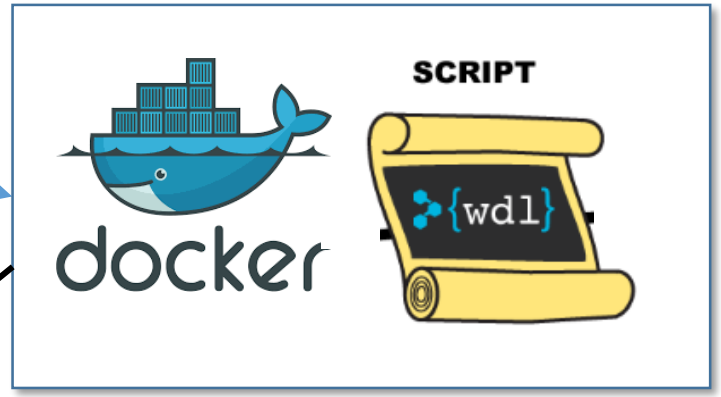


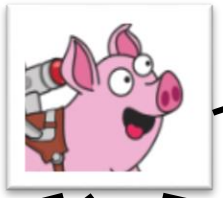
Case: *Streptococcus pneumoniae*

NGS goal: diagnosis, typing, susceptibility testing

Paper+
GitHub distribution

GENOMICS CORELEUVEN



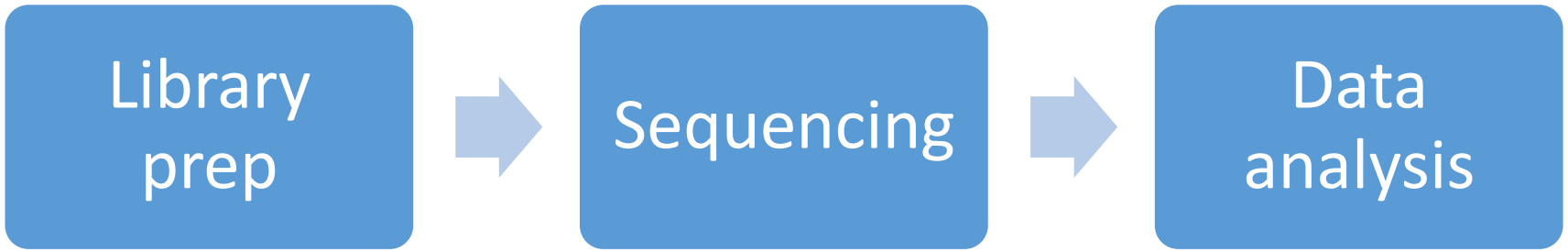


Local
cluster

Labs worldwide

Timeline and cost reduction

- Library prep cost: 80% cost reduction by Echo use for library prep
- Sample number is scalable



Day 1: QC
Day 2: library prep

Day 3: sequencing

Day 4: data transfer
and automatic cloud analysis

Benefits of NGS in microbiology

- Less need to grow bacteria
- Reduction in diagnostic time possible
- Get more and wider information
  - More detailed information
    - Species identification
    - Typing
    - Resistance gene
  - Ability to trace route of infection