

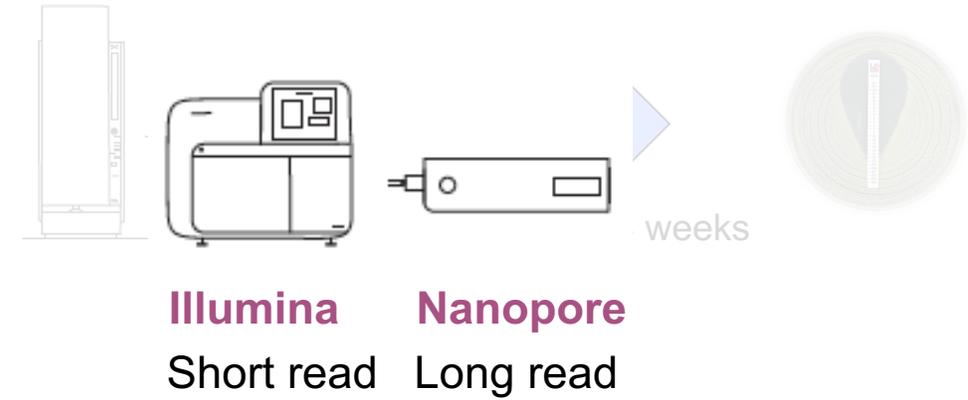
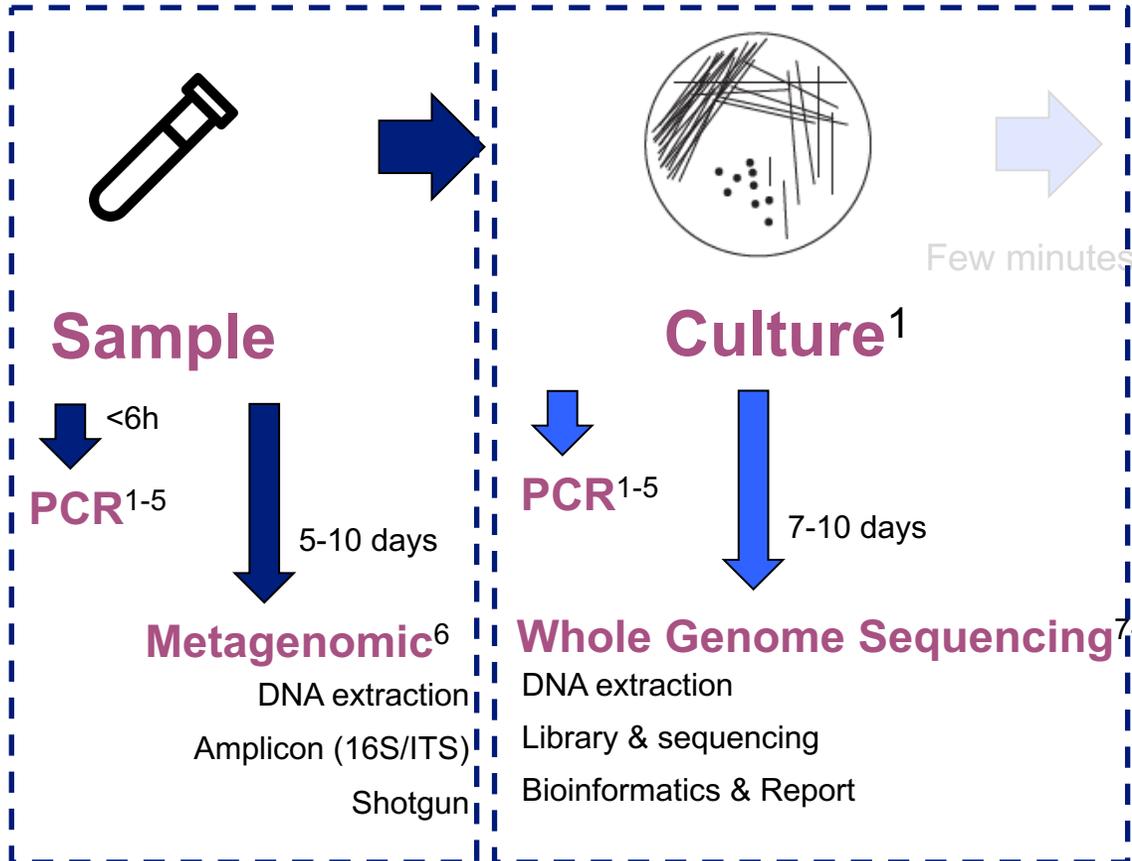


Bio 296: Microbial Bioinformatics

Routine diagnostics and NGS

Tim Roloff

Genotypic axis of diagnostic



→ **Two sequencing techniques**

¹ Hinic V, Amrein I, ... et Egli A J Micro Met 2017; ² Dierig A, Frei R, Egli A, Ped Infect Dis J 2015 ³ Egli A et al. Transpl Infect Dis 2015; ⁴ Osthoff M, ... et Egli, A. Clin Microbiol Infect 2017; ⁵ Weis C, ... Egli A, Borgwardt K, Bioinformatics 2020; ⁶ Egli A, Schmid H, et al. Clin Microbiol Infect 2017; ⁷ Hinic V, Reist J, Egli A J Microbiol Met 2018

Samples from various customers

Customers sending us samples for NGS include

Hospitals (Outbreak investigations, resistance mechanisms)

Reference labs (typing)

Authorities (surveillance)

Food companies (typing, outbreak investigations)

Pharma companies (strain characterization)

Standardized workflow in the lab

Standardization is very important for diagnostic workflows

Results can have a big impact for customer/patient

Workflow for typing and detection of resistance gene accredited according to ISO/IEC 17025

- Standard operating procedures (SOPs) for every step
- Internal quality controls (PhiX, QC strains)
- External quality controls (ring trials)
 - Samples sent to different laboratories and results compared
- Audits

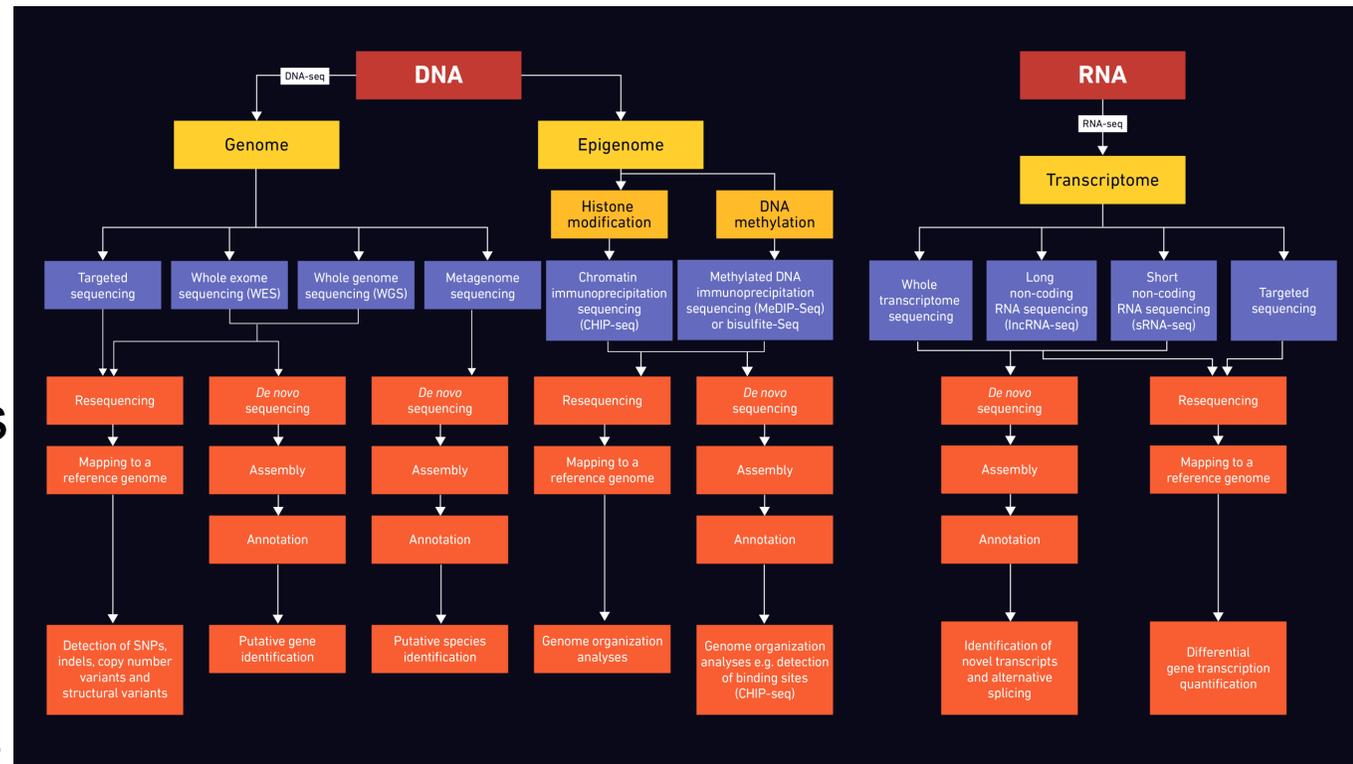


NGS in medical microbiology routine diagnostics

Analyze the genome of a pathogen or bacterial community to study

- Identity (species)
- Resistance genes and virulence factors
- Relationship between isolates
- Replace outdated diagnostics

NGS offers many more possibilities but that are not (yet) used in diagnostics



WGS data analysis for isolates: IMMense

Modular pipeline starting from bcl, fastq or fasta files

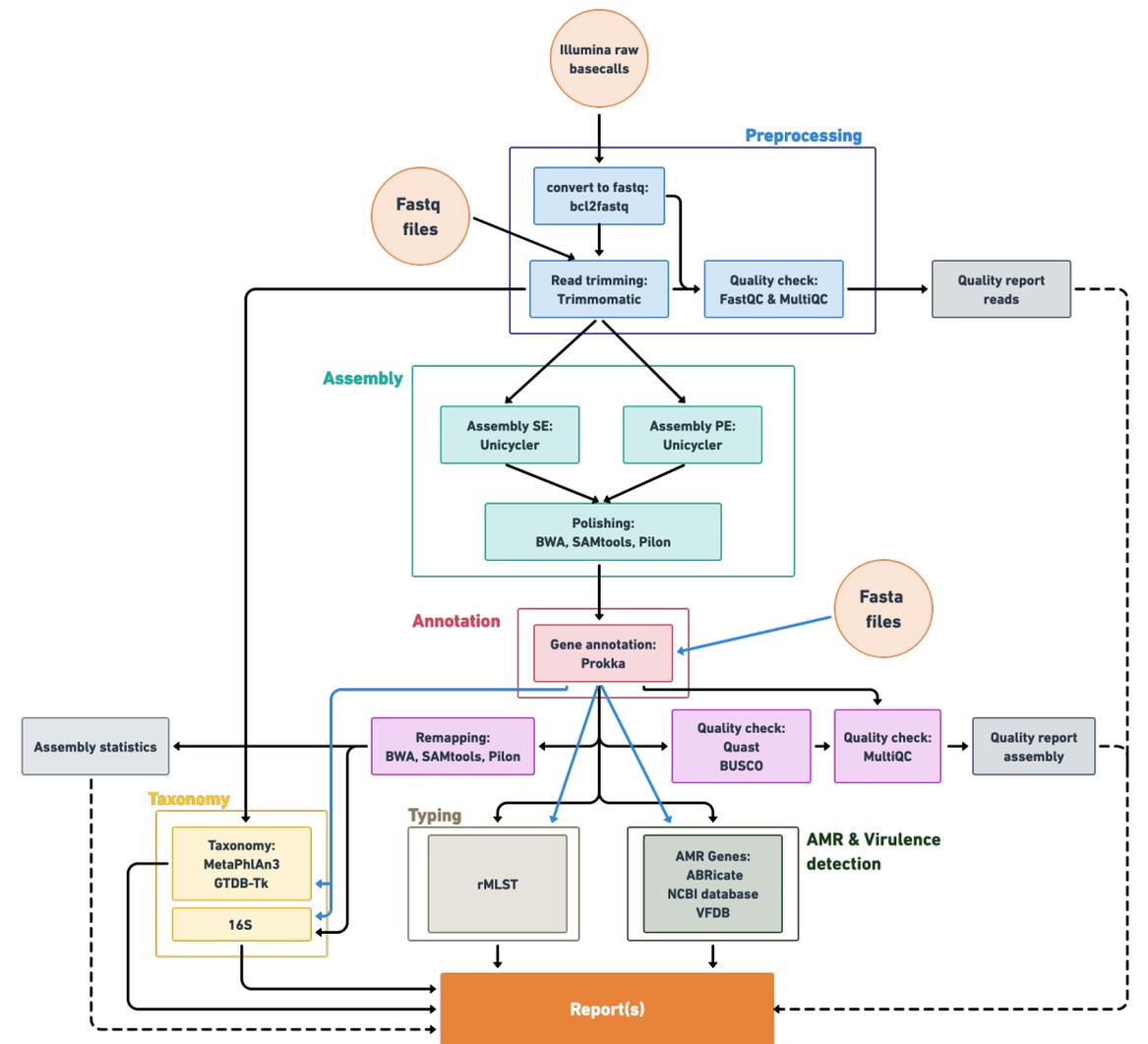
- QC, pre-processing, genome assembly, assembly QC, taxonomy, resistance and virulence gene detection
- NextFlow for **reproducibility** and **portability**
- Installation on IMM server for patient data security
- Automatically triggered after each sequencing run

Next steps:

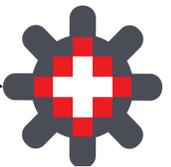
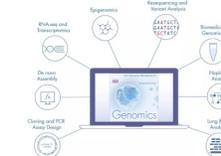
Automated customer report

Direct communication with other tools

In this course we will be using components of this workflow

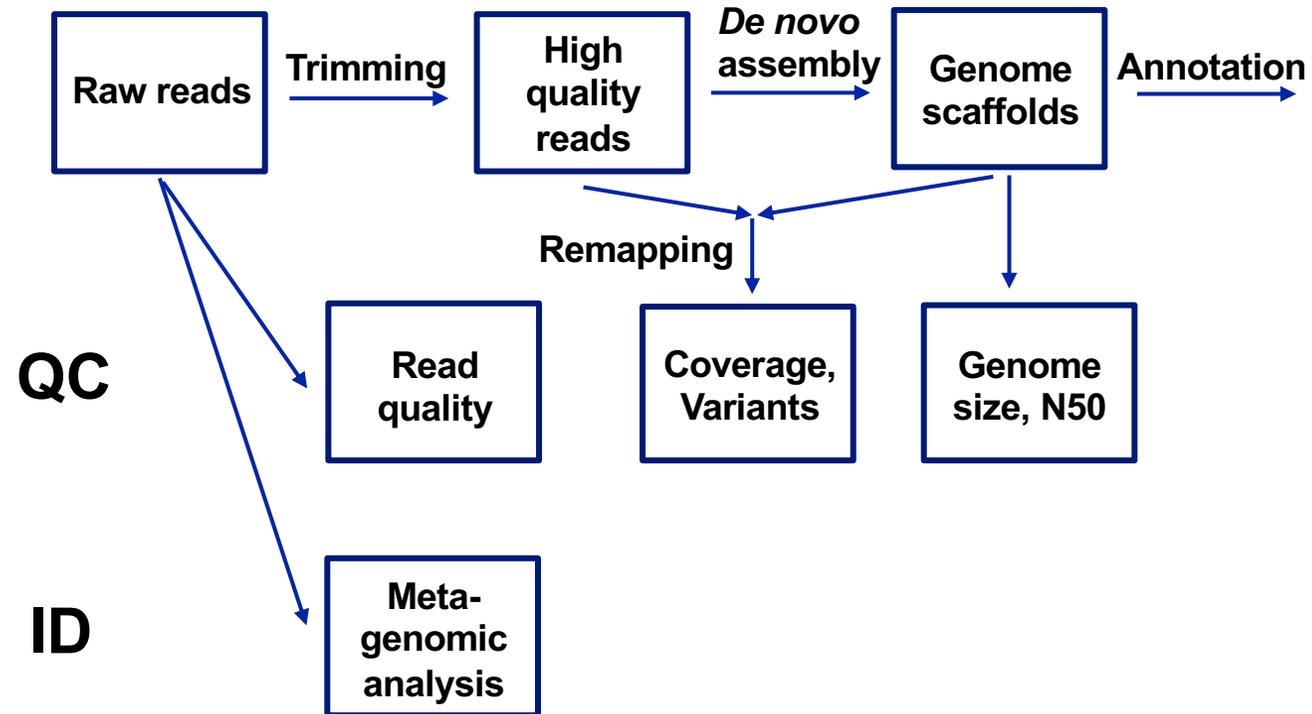


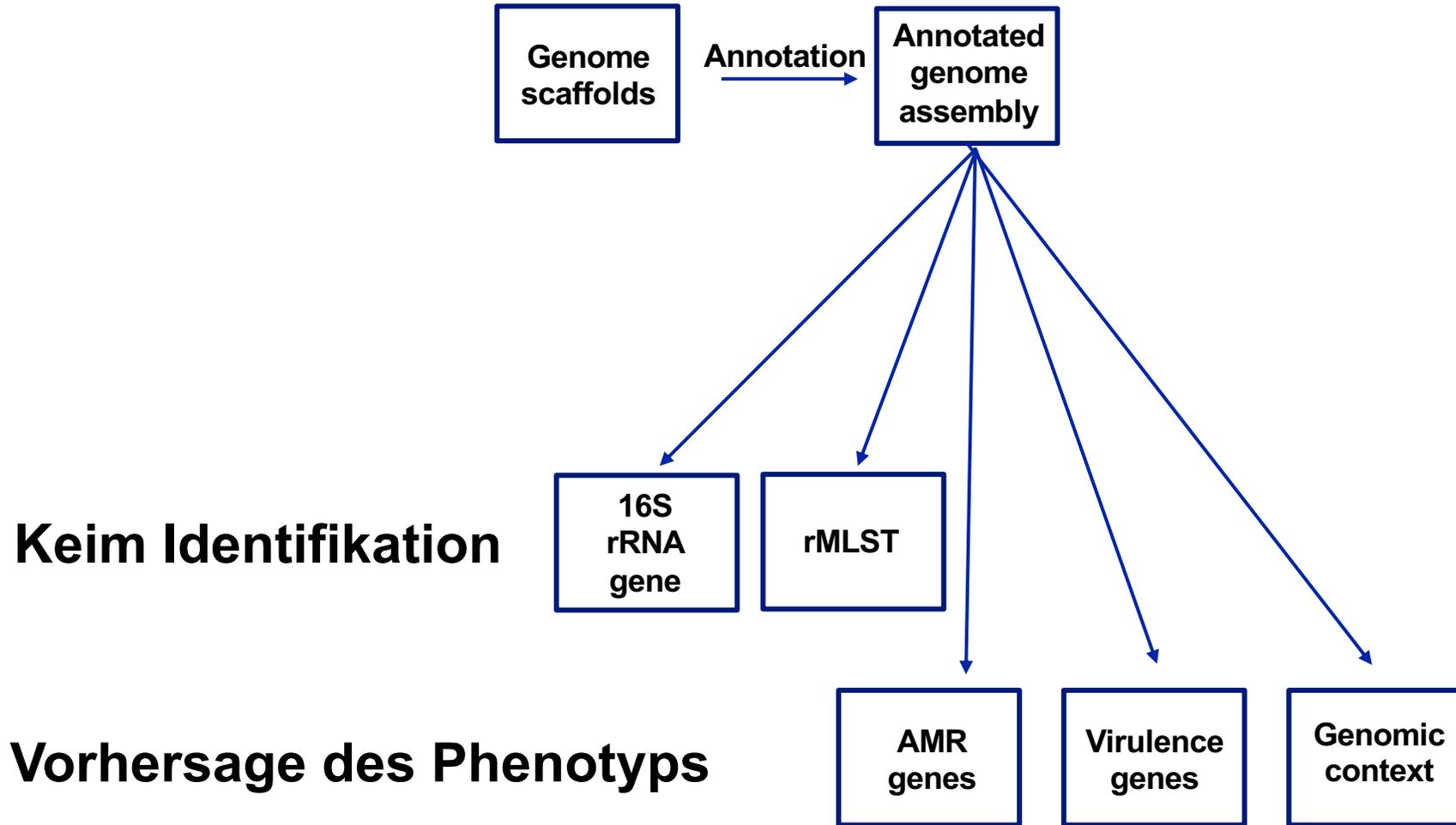
ridom
BIOINFORMATICS



SPSP

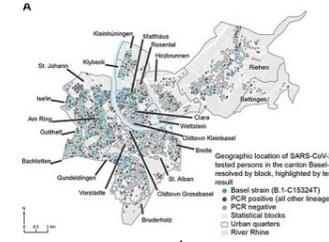
Swiss Pathogen Surveillance Platform





Data integration

- FAIR data (findable, accessible, interoperable and reusable)
- LIMS system (DiData) as **hub to integrate sample related data**
 - Fastq + QC
 - Assemblies, annotation + QC (21000 assemblies)
 - Patient data from LIS
 - MALDI-TOF MS ID and spectra
 - Anti-microbial resistance profiles e.g. MICs, SIR
 - Epidemiological metadata e.g., geography
- **Structured data** allows for data integration and machine learning
 - Phenotypic resistance vs genotype
 - Interpretation of Maldi profiles
 - Biomarkers for clinical phenotypes



MALDI Biotyper® (Bruker)



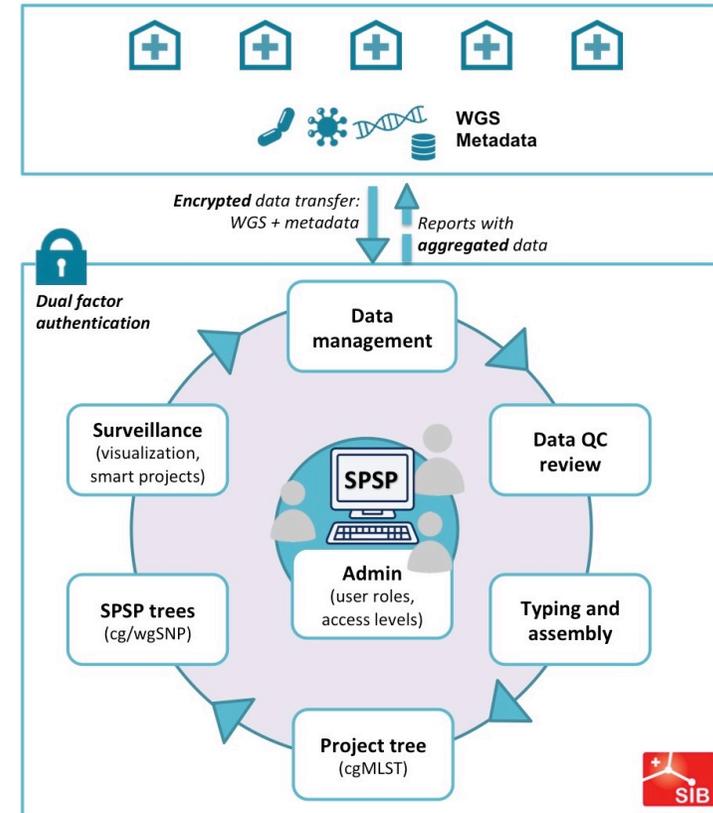
DiData



Data sharing

Swiss Pathogen Surveillance Platform

- Started with MRSA
- Heavily used for COVID-19 data sharing
- All 5 university hospitals connected
- Surveillance beyond cantonal borders



<http://spsp.ch>

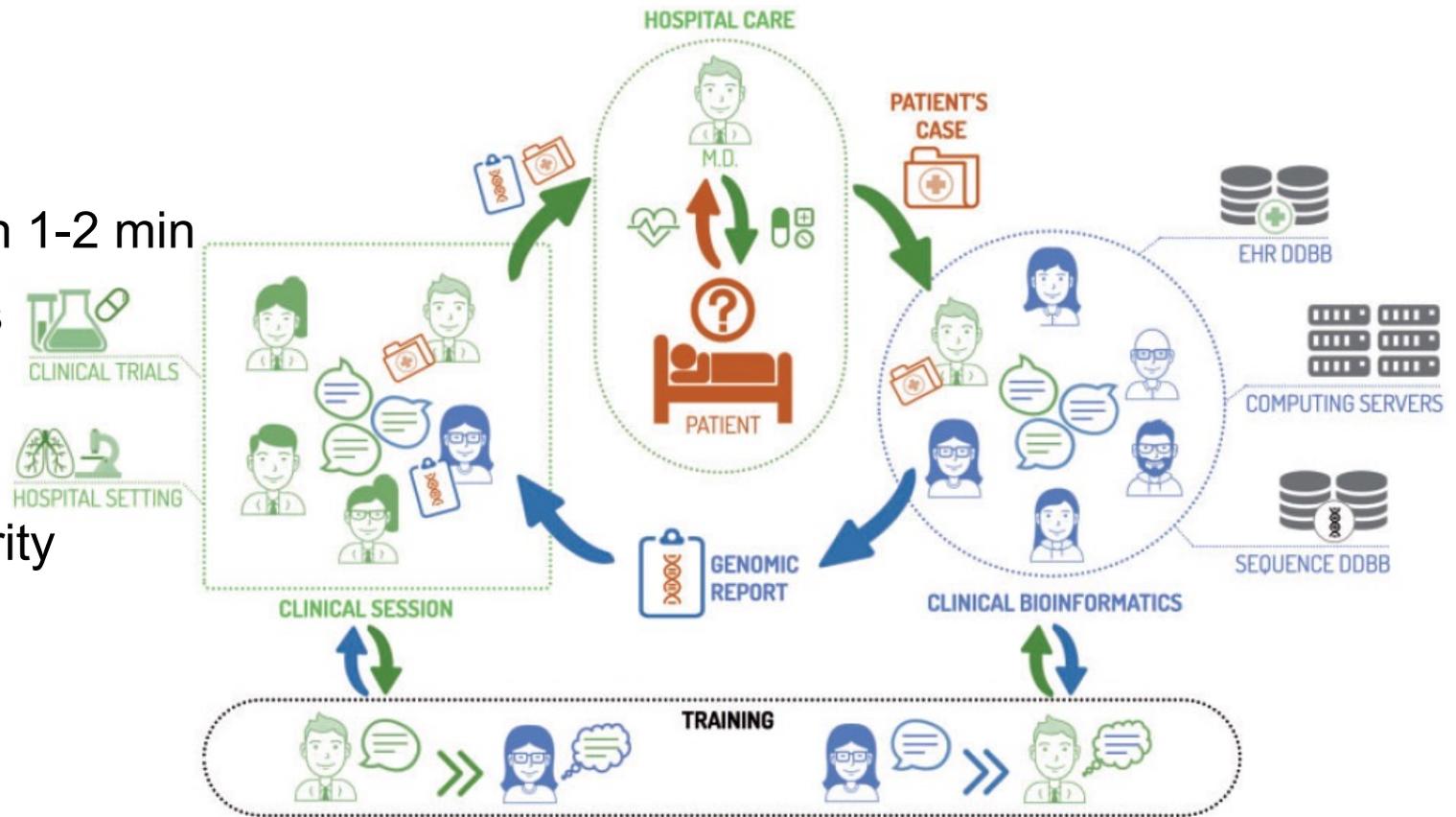
Reporting to interdisciplinary teams

Results reported to various stakeholders

Different levels of complexity

- Clinician needs key message in 1-2 min
- Statistician wants lots of details

Survey across labs and hospitals in Europe to define report granularity for each customer group



Gómez-López *Brief Bioinf* 2017

Carrico *Clin Microbiol Infect* 2018

Standardized report



Universität
Zürich^{uzh}

Institut für
Medizinische Mikrobiologie

Universität Zürich
Institut für Medizinische Mikrobiologie
Gloriastrasse 30/32
CH-8000 Zürich
Telefon +41 44 634 27 00
Telefax +41 44 634 49 06
www.imm.uzh.ch

Prof. Dr. med. Dr. phil. Adrian Egli, FAMH
Direktor, Institut für Medizinische Mikrobiologie
+41 44 634 28 60
aegli@imm.uzh.ch

Einsender Adresse
Dr. XXX FAMH
Institut für XXX
Spital XXX
Postfach XXX
XXXX

Zürich, 29. November 2022

Molekulare Typisierung mittels Genom-Sequenzierung (WGS)

Zusammenfassung

Eine Zusammenfassung der Daten bis September 2022 wird vorgelegt.

Analyse: H. Seth-Smith
Bioinformatik

Freigabe der Resultate
Prof. Dr. Dr. A. Egli,
Direktor, Institut für Medizinische Mikrobiologie

Summary / Interpretation

Seite 14



Universität
Zürich^{uzh}

Institut für
Medizinische Mikrobiologie

Qualitätssicherung und Durchführung / Methoden und Qualitätssicherung

Gesamtenom-Sequenzierung (WGS) wurde mit Hilfe eines MiSeq Illumina Sequenziergerätes durchgeführt. Die erhaltenen Sequenzdaten wurden mittels einer bioinformatischen Pipeline bzgl. statistischer Qualitätskriterien kontrolliert (Seth-Smith et al., 2019, *Front Public Health*, 3591/publ/2019.00241). Die Resultate wurden mit Ridom SeqSphere (v8.3.4) analysiert durch definierte Schemen (cgMLST.org).

core genome MLST (cgMLST) Methodik vergleicht alle gemeinsamen Gene (Kerngenom) der analysierten Isolate. Dabei wird untersucht, wie viele unterschiedliche Allele die einzelnen Isolate im Kerngenom zueinander aufweisen. Je kleiner diese Anzahl ist, desto näher sind die Isolate miteinander verwandt. Insgesamt werden mehr als 1'400 Gene in den Vergleich eingeschlossen. Durch entsteht eine sehr hohe Auflösung für die Typisierung einzelner Bakterien.

Sequenzierungen und Analysen aller Samples haben die internen Qualitätsstandards erreicht (Deckung > 30x Durchschnitt; Genomgröße; Spezies bestimmt).

Personelle Auskunft

Dr. phil. Helena Seth-Smith
Bioinformatikerin
44 634 2624
hsmith@imm.uzh.ch

Dr. rer. nat. Tim-C. Roloff Handschin
Bioinformatiker, technischer Leiter NGS Facility
+41 44 634 0257
trollof@imm.uzh.ch

Methods



Phylogeny

Sample table

WGS in pathogen identification

- New species *Mycobacterium basiliense* identified and characterised
 - ANI to closest species $\approx 81\%$ ($<95\%$)
 - Digital DDH to closest species $\approx 23\%$ ($<70\%$)

- Case study: travel returner with fever
 - *Borrelia persica* identified as pathogen by shotgun metagenomics on blood sample
 - 684 of 7.8M reads mapped to *B. persica* (database assembly AYOT)

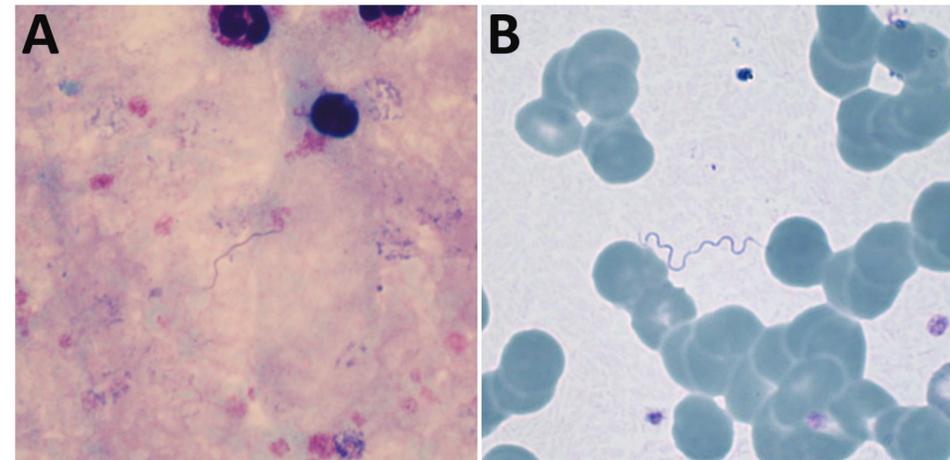
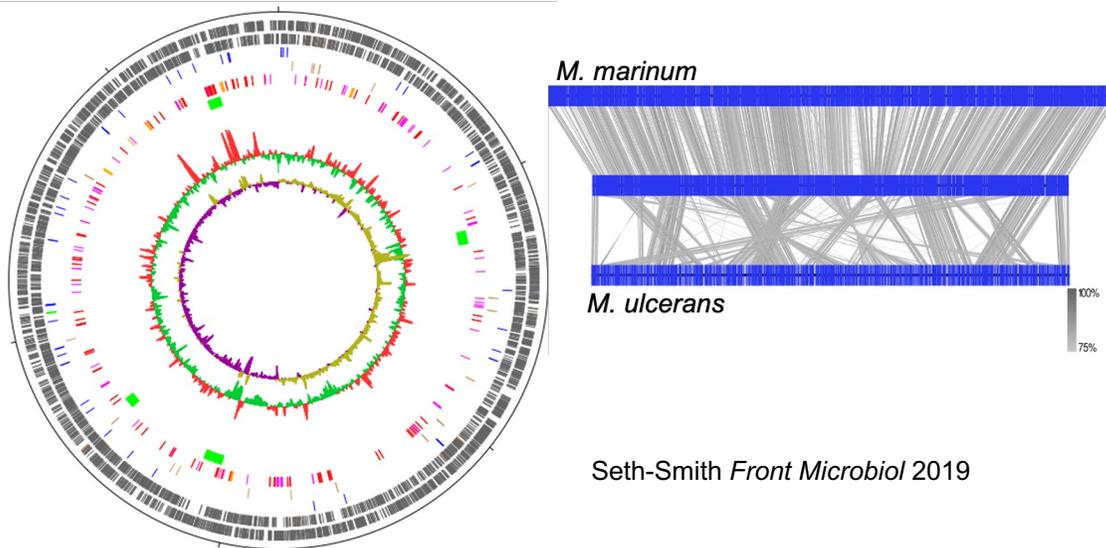


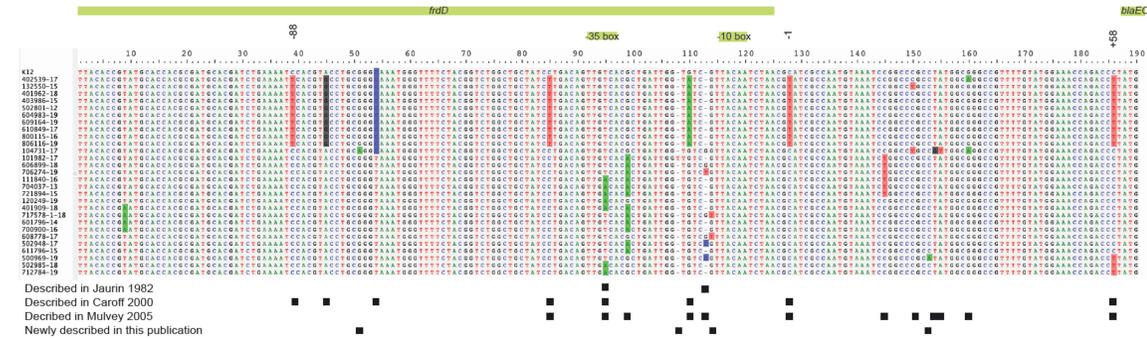
Figure. Giemsa-stained thick (A) and thin (B) blood films, demonstrating extracellular spirochetes. Original magnifications $\times 1,000$.

Muigg /Seth-Smith
Emerg Infect Dis 2020

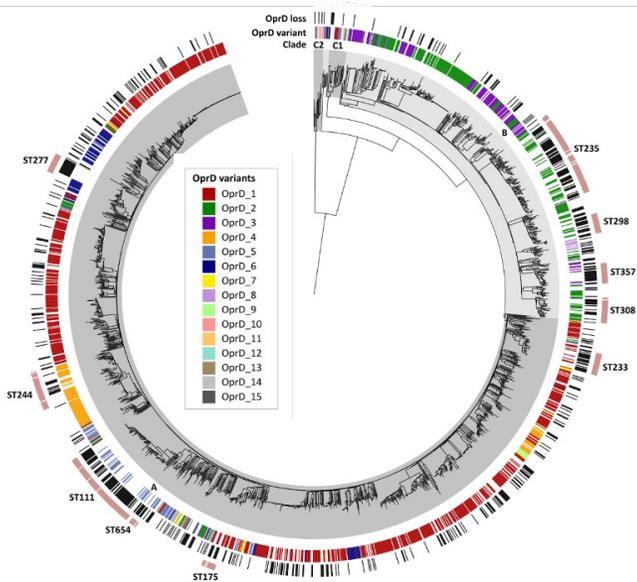
WGS in AMR prediction

- PorinPredict software developed:
 - Predicts OprD porin loss of function in *Pseudomonas aeruginosa* (meropenemR)
 - Phenotype-genotype comparisons in 1,078 strains, 79 from USB

- Novel *bla_{EC}* promoter mutations found in AmpC phenotype *E. coli* Hinic *J Antimicrob Chemother* in review



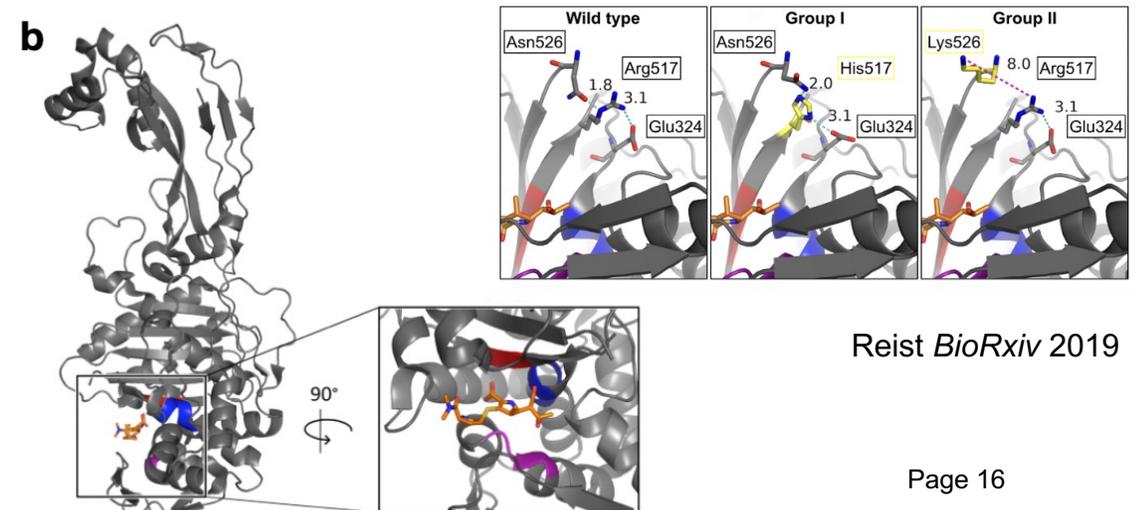
Distribution of porin loss in 2088 genomes



Frameshifts / terminations / promoter mutations

Biggel *Microbiol Spectrum* in press

- Phenotypic impact of amino acid substitutions in *Haemophilus influenzae pbp3*

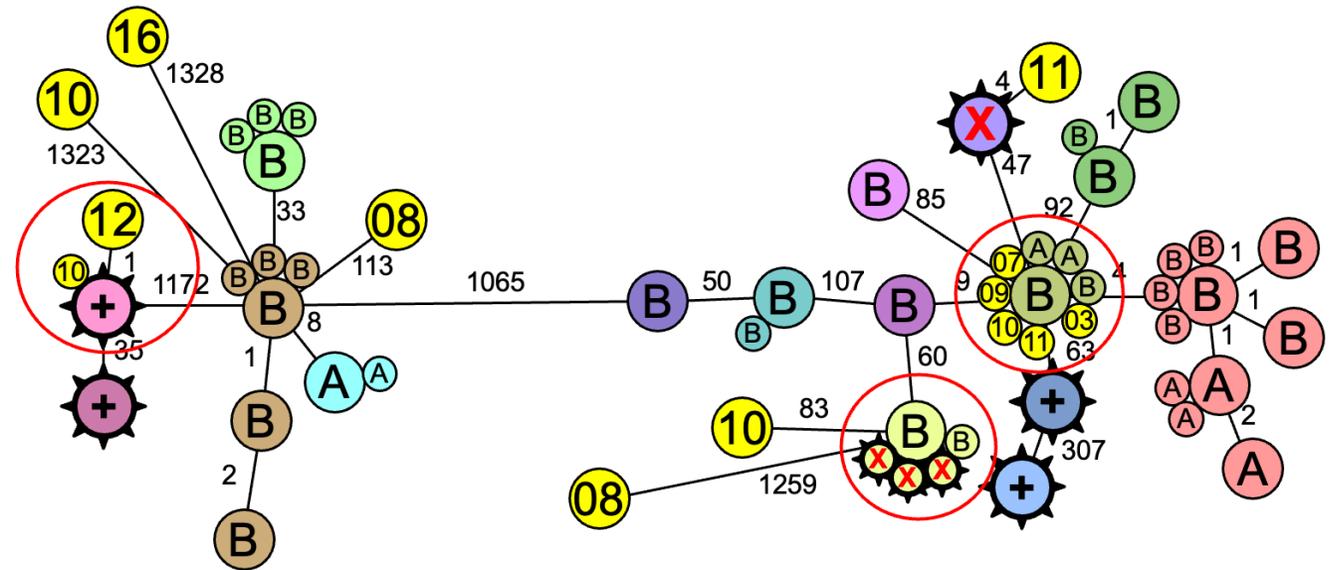
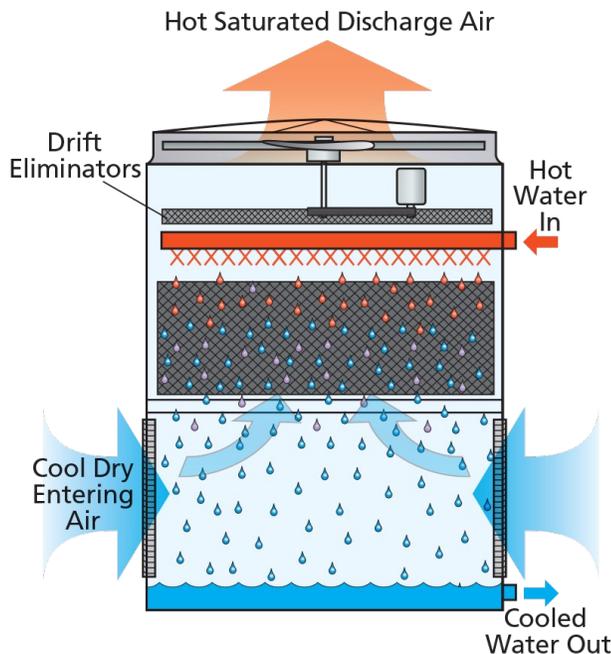


Reist *BioRxiv* 2019

WGS to solve outbreaks: *Legionella pneumophila*

Legionella pneumophila infections from same urban district

- *Legionella pneumophila* detected in cooling towers
- Cooling towers (A and B) have diverse strains, but also share similar strains
- Three 2017 outbreak isolates (X) identical to cooling tower B isolates
- Clinical isolates from previous years are closely related to other cooling tower strains



cgMLST Sequence types (Cluster types)

277 278 228 177 281 283 276 279 117 174 199 280 282 284 37 90

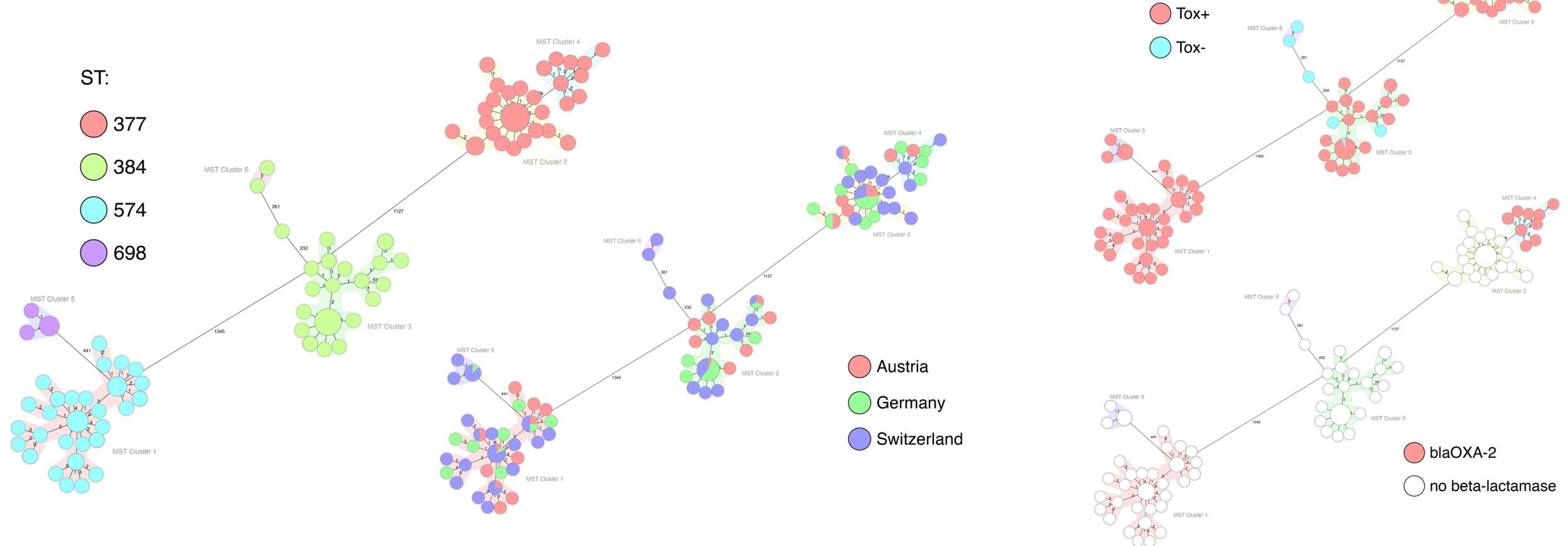
⚙ Isolates 2017 X Isolates 4058 00 Isolates last decade

Wüthrich D et al. *Eurosurveill* 2018

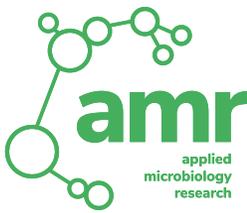
WGS in outbreaks: *Corynebacterium diphtheriae* 2022

Increase in diphtheriae cases in refugee camps across Europe

- Determination of MLST, clusters, toxin presence, ARG from the genome assembly
- Visualised in Ridom Seqsphere+, superimposed on Minimum Spanning Tree



NGS in routine diagnostics - summary



- Used for identification, outbreak investigation, surveillance and resistance / virulence prediction
- Highest resolution for outbreak investigations and surveillance
- Interdisciplinary team needed (BMAs, bioinformaticians, clinicians/specialists)
- ISO accredited workflow
- Proper communication of results is crucial
- Rather expensive and slow diagnostic tool (work in progress)
- Proper data storage needed to make most out of the data, e.g. with machine learning (FAIR principles)

Time for a break and some labwork

