Microbes and Infection 26 (2024) 105410

Contents lists available at ScienceDirect

Microbes and Infection

journal homepage: www.elsevier.com/locate/micinf

Meeting report

Conference report: the first bacterial genome sequencing pan-European network conference

Zoja Germuskova [a, 1], Elisa Sosa [a, 1], Amaya Campillay Lagos [b, 1], Hege Vangstein Aamot [c, d], Mathew A. Beale [e], Claire Bertelli [f], Jonas Björkmann [g], Natacha Couto [h], Lena Feige [i], Gilbert Greub [f], Erika Tång Hallbäck [j], Emma B. Hodcroft [k, l], Dag Harmsen [m], Laurent Jacob [n], Keith A. Jolley [o], Andre Kahles [p], Alison E. Mather [q, r], Richard A. Neher [l, s], Aitana Neves [l], Stefan Niemann [t], Oliver Nolte [a], Sharon J. Peacock [u], Mohammad Razavi [v], Tim Roloff [a], Jacques Schrenzel [w], Per Sikora [j, x], Martin Sundqvist [b], Paula Mölling [b, **, 2], Adrian Egli [a, *, 2]

[a] Institute of Medical Microbiology, University of Zurich, Zurich, Switzerland
[b] Department of Laboratory Medicine, Clinical Microbiology, Faculty of Medicine and Health, Örebro University, Örebro, Sweden
[c] Department of Microbiology and Infection Control, Akershus University Hospital, Lørenskog, Norway
[d] Department of Nursing, Health, and Laboratory Science, Østfold University College, Fredrikstad, Norway
[e] Parasites and Microbes Programme, Wellcome Sanger Institute, Hinxton, United Kingdom
[f] Institute of Microbiology, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland
[g] Center for Molecular Diagnostics, Department of Clinical Genetics, Pathology and Molecular Diagnostics, Office for Medical Services, Region Skåne, Lund, Sweden
[h] Centre for Genomic Pathogen Surveillance, Pandemic Sciences Institute, University of Oxford, United Kingdom
[i] Federal State Agency for Consumer and Health Protection Rhineland-Palatinate, Germany
[j] Department of Clinical Microbiology, Sahlgrenska University Hospital, Gothenburg, Sweden
[k] Swiss Tropical and Public Health Institute, University of Basel, Allschwil, Switzerland
[l] Swiss Institute of Bioinformatics, Geneva, Switzerland
[m] Department of Periodontology and Operative Dentistry, University Hospital Münster, Münster, Germany
[n] Sorbonne Université, France
[o] Department of Biology, University of Oxford, United Kingdom
[p] Institute for Machine Learning, Department of Computer Science, ETH Zurich, Switzerland
[q] Quadram Institute Bioscience, Norwich, United Kingdom
[r] University of East Anglia, Norwich, United Kingdom
[s] Biozentrum, University of Basel, Basel, Switzerland
[t] Forschungszentrum Borstel, Leibniz Lungenzentrum, Germany
[u] University of Cambridge, United Kingdom
[v] Division of Clinical Microbiology, Department of Laboratory Medicine, Karolinska Institutet, Stockholm, Sweden
[w] University Hospital Geneva, Geneva, Switzerland
[x] Bioinformatics Data Center, Core Facilities, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

ARTICLE INFO

Article history:
Received 9 August 2024
Accepted 20 August 2024
Available online 30 August 2024

Keywords:
Bacterial genomes
DNA sequencing

* Corresponding author. Institute of Medical Microbiology, University of Zurich, Gloriastrasse 28/30, 8006 Zurich, Switzerland.
** Corresponding author. Department of Laboratory Medicine, Clinical Microbiology, Örebro University Hospital, Örebro, Sweden.
E-mail addresses: paula.molling@regionorebrolan.se (P. Mölling), aegli@imm.uzh.ch (A. Egli).
[1] Shared first authorship.
[2] Share last authorship.

https://doi.org/10.1016/j.micinf.2024.105410

Standardization
Community network
Data sharing
Database

1. Introduction

Genome sequencing of microorganisms has become increasingly important for outbreak investigation and pathogen surveillance in recent years, proving to be invaluable in these areas due to its high resolution. However, its applications in clinical settings are still limited [1]. To integrate genome sequencing, in particular whole genome sequencing (WGS), into patient care, we need to overcome several challenges. These include technical limitations due to the complexity and reproducibility of the large amount of sequencing data produced, which calls for advanced IT solutions; the lack of standardized and automated wet-lab protocols; high costs per sample when processing only a few samples in clinical diagnostics; and the lack of unified reporting of results [2]. Achieving this integration demands focused dialogues, a willingness to evolve and innovate, and interdisciplinary solutions. A united community can standardize efforts, facilitate resource and data sharing, and amplify collective advancements.

The first conference of the "Bacterial Genome Sequencing pan-European Network" aimed to serve as a unifying platform for field leaders, collaboratively driving the future of genomic medicine in clinical microbiology. We aimed to generate scientific exchange and form a community. The conference was co-organized by Adrian Egli (Switzerland) and Paula Mölling (Sweden) and hosted 26 speakers (Table 1). The conference venue in Engelberg, Switzerland, ensured active interactions and effective networking among attendees (Fig. 1). The conference featured a 4-day program with a series of presentations, panel discussions, interactive workshops, and networking activities. Sessions covered topics on data for public health, setting standards, practical use cases, and fostering WGS as a technology in clinical diagnostics. The topics are summarised in this paper.

2. Data for public health

2.1. Swiss pathogen surveillance platform (SPSP)

Adrian Egli detailed the functionality and evolution of the SPSP (www.spsp.ch), particularly highlighting its development during the SARS-CoV-2 pandemic [3–5]. SPSP is now expanding to other pathogens, including features like antibiotics resistance prediction. He addressed the challenge of sharing research and diagnostic data among laboratories and making data available to the authorities and public while keeping confidential data secure. The core principle of SPSP is FAIR data (findable, accessible, interoperable, and reusable). The legal, ethical and governance framework of public health-related genomic data was discussed, as well as the sustainability of funding models combining government and academic sources, and the challenge of balancing public health functions with academic research and patient care.

2.2. Genomic medicine Sweden (GMS): national genomic platform for data sharing

Paula Mölling outlined the objectives of GMS (https://genomicmedicine.se/en/), which aims to implement precision medicine with broad gene sequencing in healthcare for enhanced diagnostics and equitable patient care using joint national technology solutions [6,7]. Ongoing GMS projects within the field of infectious diseases were presented, including a pilot study where the quality and accuracy of whole genome sequencing (WGS) data generated at nine different laboratories were examined and analysed by common pipelines, working towards implementing a national monitoring approach for methicillin-resistant *Staphylococcus aureus* (MRSA) in Sweden. In addition, insights were shared from a broad national long-read 16S sequencing project with Oxford Nanopore Technologies (ONT) conducted at 20 different laboratories in Sweden, including a validation of 16S ONT sequencing against short-read sequencing. The need for data sharing between regions was highlighted to improve surveillance and demonstrated the impact of sharing epidemiological data. Challenges of convincing stakeholders about the benefits of sequencing, particularly in terms of cost and efficiency compared to traditional methods were discussed.

2.3. miGenomeSurv: connecting across countries

Dag Harmsen emphasized the importance of international collaborations in genomic surveillance and presented miGenomeSurv (https://www.medizin.uni-muenster.de/migenomesurv/home.html) as a platform facilitating data sharing and connectivity across countries, where users retain full control of their data. He reviewed various national and international initiatives like NCBI Pathogen Detection, the UK Health Security Agency's centralized approach, and CDC PulseNet's decentralized model with central analysis. He further advocated for collaboration between different national systems and stressed the role of organizations like the European Center for Diseases Control and Surveillance (ECDC) and the World Health Organization (WHO) in creating an international data sharing framework.

2.4. Surveillance in Europe, in Elixir's perspective

Aitana Neves introduced the Swiss Institute of Bioinformatics (SIB; https://www.sib.swiss/) and its role within ELIXIR (https://elixir-europe.org/), an intergovernmental organization with 25 nodes in Europe dedicated to promoting data sharing standards and tools, with a focus on making data more accessible and discoverable. ELIXIR aims to facilitate the sharing of raw, consensus, and contextual data (e.g. size of the dataset) across various domains, including human, animal, and transcriptomic data. In general, there is a need for sustainable infrastructure for data sharing and pathogen surveillance, however there are challenges with national constraints, resource competition and the need for government support.

2.5. PubMLST: public data repositories

Keith Jolley presented on PubMLST (https://pubmlst.org/), a public database repository with over 20 years of history in indexing multi-locus sequence typing (MLST) alleles and bacterial genomes. PubMLST supports sequence typing schemes for more than 130 microorganisms, indexing combinations of alleles to assign sequence types. He discussed the significance of PubMLST in providing public access to microbial sequence data and its role in research and surveillance efforts. He also addressed questions

Table 1
List of participants and their affiliations. Every attendee also contributed to the program.

Name	Country	Affiliation	Link
Hege Vangstein Aamot	Norway	Akershus University Hospital & Østfold University College	www.ahus.no www.hiof.no
Mathew Beale	United Kingdom	Wellcome Sanger Institute	www.sanger.ac.uk
Claire Bertelli	Switzerland	Lausanne University Hospital & University of Lausanne	https://www.chuv.ch/fr/microbiologie/imu-home
Jonas Björkmann	Sweden	Lund University Hospital	Division of Clinical Genetics | Lund University
Natacha Couto	United Kingdom	University of Oxford	https://www.pathogensurveillance.net/team/
Adrian Egli	Switzerland	University of Zurich	https://www.imm.uzh.ch/de.html
Lena Feige	Germany	Federal State Agency for Consumer and Health Protection Rhineland-Palatinate	https://lua.rlp.de/
Zoja Germuskova	Switzerland	University of Zurich	https://www.imm.uzh.ch/de.html
Gilbert Greub	Switzerland	Lausanne University Hospital & University of Lausanne	https://www.chuv.ch/fr/microbiologie/imu-home
Erika Tång Hallbäck	Sweden	University of Gothenburg	https://www.gu.se/en
Dag Harmsen	Germany	University Hospital Münster, Münster, Germany	https://www.ukm.de/university-hospital-muenster
Emma Hodcroft	Switzerland	Swiss Tropical and Public Health Institute & Swiss Institute of Bioinformatics	https://www.swisstph.ch/en https://www.sib.swiss/
Laurent Jacob	France	Sorbonne Université	https://www.sorbonne-universite.fr/
Keith Jolley	United Kingdom	University of Oxford	https://www.ox.ac.uk/
Andre Kahles	Switzerland	ETH Zurich	https://bmi.inf.ethz.ch/
Amaya Campillay Lagos	Sweden	Örebro University Hospital	https://www.oru.se/english/about-us/organisation-and-governance/faculties/faculty-of-medicine-and-health/
Alison Mather	United Kingdom	Quadram Institute & University of East Anglia	https://quadram.ac.uk/ https://www.uea.ac.uk/
Paula Mölling	Sweden	Örebro University Hospital	https://www.oru.se/english/about-us/organisation-and-governance/faculties/faculty-of-medicine-and-health/
Richard Neher	Switzerland	University of Basel & Swiss Institute of Bioinformatic	https://www.biozentrum.unibas.ch/ https://www.sib.swiss/
Aitana Neves	Switzerland	Swiss Institute of Bioinformatics	https://www.sib.swiss/
Stefan Niemann	Germany	Forschungszentrum Borstel, Leibniz Lungenzentrum	https://fz-borstel.de/index.php/de/
Oliver Nolte	Switzerland	University of Zurich	https://www.imm.uzh.ch/de.html
Sharon Peacock	United Kingdom	University of Cambridge	https://www.cam.ac.uk/
Mohammad Razavi	Sweden	Karolinska Institutet	https://ki.se/en/labmed/divisions
Tim Roloff	Switzerland	University of Zurich	https://www.imm.uzh.ch/de.html
Jacques Schrenzel	Switzerland	University Hospital Geneva	https://www.hug.ch/en
Per Sikora	Sweden	University of Gothenburg &	https://www.gu.se/en
Elisa Sosa	Switzerland	University of Zurich	https://www.imm.uzh.ch/de.html
Martin Sundqvist	Sweden	Örebro University Hospital	https://www.oru.se/english/about-us/organisation-and-governance/faculties/faculty-of-medicine-and-health/

about managing quality assurance with 300 curators and the relevance of allelic approaches with the increasing use of long-read sequencing.

2.6. Nextstrain

Richard Neher presented on Nextstrain (https://nextstrain.org/), a platform for real-time public-facing genomic data analysis, primarily focusing on SARS-CoV-2, influenza A and B viruses, RSV A&B, zika virus, but also including some bacterial pathogens like *Mycobacterium tuberculosis* [8]. He highlighted the platform's capability to allow users to share their analyses either publicly or privately. He discussed challenges of data sharing, ranging from concerns about data reuse and “scooping” to the difficulties of uploading data to public repositories. He discussed the benefits of analyzing data pre-submission using Nextclade (https://clades.nextstrain.org/), which provides quality control (QC) and alignment through client-side computation via web assembly. He emphasized the importance of diversifying funding sources beyond grants to ensure the sustainability of such platforms.

2.7. Impact for public health

Sharon Peacock highlighted the broader public health implications of genomic data and its role in informing policy decisions and outbreak responses, emphasizing the importance of using evidence-based use cases to justify funding for sequencing initiatives. Strong use cases were highlighted, including vaccine design and redesign (including tracking of pathogen evolution over time to detect likely vaccine escape); drug development; pro-active surveillance for outbreaks in the community and hospitals; and diagnostic test development. The importance of routine sequencing in hospitals to detect and manage outbreaks (as part of a continuous surveillance in contrast to outbreak-triggered investigations) was highlighted. This detects outbreaks and pseudo-outbreaks and can inform a targeted and cost-effective approach to infection prevention and control. Also highlighted were the potential of sequencing to support personalized medicine, particularly (although not limited to) choice of drug for patients with HIV, and the detection of resistance (including more rapid detection of multidrug resistance compared with conventional methods) in *M. tuberculosis*. She advocated for the integration of metagenomics

Fig. 1. Participants of the conference in front of the Swiss Alps.

in diagnostics and highlighted future applications, including antimicrobial resistance and gut microbiome studies. However, there is a need for cost-benefit studies and the involvement of economists to make a compelling case for the routine use of sequencing in healthcare settings.

2.8. Panel discussion: data sharing, data governance & ethical considerations

The panel of Day 1 discussed the need for developing systematic and harmonized frameworks for data sharing, particularly for research projects and outbreak responses. Legal frameworks should be clarified to facilitate seamless data sharing for surveillance, with ethical committees consistently approving the sharing of bacterial strains, as these belong not to individuals but to the laboratories that isolated them. Incentivization of data sharing was discussed, highlighting the importance of uploading data post-publication and adhering to data management plans. Issues with routine data sharing, especially in non-epidemic contexts, were also discussed. The panel considered Creative Commons licences for broader genome data usage and the need for journals to verify data usage rights. Metadata access and use will depend on whether the sequencing is embedded within a clinical setting or is academic, or a partnership between the two. The discussion concluded with considerations for sustainable data governance, simplifying data access models, and balancing public health needs with academic research.

3. Setting standards

31. Pre-analytical standardization

Jacques Schrenzel emphasized the importance of framing WGS in a clinical context. He discussed the need for recognizing use cases, balancing costs with medical value, and ensuring reproducibility and precision in pre-analytical procedures. He advocated we should think about standardization from the earliest steps of sample collection, especially if storing native samples (as compared to cultured strains directly), ensuring reproducibility. Other key points included data capture and reporting of metadata and clinical information associated with each sample, using a common language (ontologies).

3.2. Analytical standardization

Hege Vangstein Aamot outlined various guidelines for analytical standardization and highlighted key quality control measures throughout the workflow, from DNA extraction to sequencing and assembly quality. She emphasized the importance of validation of workflows and pipelines and outlined best practices for species identification and antimicrobial resistance detection. She raised the topic of using WGS for detection of virulence factors but argued that a proper validation and comprehensive studies are still needed to demonstrate clinical utility.

3.3. Post-analytical standardization

Natacha Couto covered post-analytical standardization, focusing on interpretation of results (e.g. antimicrobial resistance), post-analytical standards and reporting for clinical and public health purposes. She emphasized the importance of effective and timely communication of test results to various stakeholders, including clinicians, infection prevention teams, and public health institutions. She discussed the need for standardized reporting formats, particularly for antimicrobial resistance data and the need for translating that information into clinical actionable outcomes. She discussed several tools for integration, manipulation and visualization of data, as well as maintenance and curation of data in centralized or decentralized databases.

3.4. Quality controls for the wet lab

Claire Bertelli exchanged views on quality controls in the wet lab, covering assay validation, internal and external quality controls, and standardized procedures to ensure reliability and reproducibility in genomic analyses. She discussed the inclusion of positive and negative samples as internal quality controls in terms of necessity, feasibility, costs and frequency. For external quality assessments (EQA), the examples of the Swiss MRSA and SARS-CoV-2 sequencing ring trials were discussed [9,10]. She also stressed the importance of regular proficiency testing to maintain high standards in WGS, and the current lack of adapted schemes in certified organizations, which is only partly mitigated by occasional inter-laboratory quality assessments.

3.5. Quality controls for pipelines

Tim Roloff presented on the quality control of bioinformatic pipelines. He discussed the importance of validating pipelines to ensure they perform as expected and produce reproducible results. He highlighted various QC tools and parameters for different stages of the sequencing process, including read quality assessment, de novo assembly quality, and annotation quality control. He discussed the need for reference datasets to benchmark pipeline performance and ensure the accuracy of sequencing data and consistency across different sequencing platforms and institutions. He also discussed the importance of version control, containerization to ensure reproducibility of data analysis, modularity of pipelines and thorough documentation of each bioinformatic step.

3.6. Bioinformatic challenges with pipelines

Jonas Björkman discussed the challenges associated with design, selection and use of bioinformatic pipelines. He emphasized analysis of WGS data can be done using different pipelines and methods of choice, but the comparability of results needs to be ensured. This could be achieved via EQAs, or by testing pipelines using a set of similar data. He discussed the benefits and drawbacks

of commercial versus open-source software, the need for standardizing data formats and QC measures to facilitate data sharing and interoperability, and stressed the importance of keeping pipelines up-to-date to ensure they remain relevant and effective in the long run.

3.7. Panel discussion: reaching a shared standard

The panel of Day 2 discussed the accreditation of laboratories, emphasizing its role in enforcing standardization for different use cases. They highlighted that while accreditation ensures quality, accrediting bodies may lack expertise in WGS. The burden and cost of accreditation were noted, along with the challenges in data analysis and reporting. Participants discussed the importance of EQAs and internal quality assessments (IQA), and they pointed out that not all pathogens and pipelines require constant quality assurance (QA). Reporting complexities were addressed, with suggestions for simplified, standardized formats and the need for guidelines from ECDC and WHO. The interpretation of reports, especially for public health, was identified as a challenge due to legal implications. The discussion underscored the need for clarity in reporting species identification, AMR, and phylogenetic relationships, with appropriate disclaimers based on current knowledge.

4. Practical use cases

4.1. Foodborne pathogens

Alison Mather started the session with insights into foodborne pathogens and her current research on it. She discussed the application of genomic sequencing in identifying and managing foodborne pathogens, such as *Campylobacter jejuni*, *Pseudomonas aeruginosa*, and *Escherichia coli*. She provided case studies demonstrating the impact of genomic data on food safety and public health; in addition to the need to develop standards, data sharing and pipelines, as discussed on previous days, consideration of sampling strategies is also critical to obtain an accurate and timely understanding of pathogen burden and diversity, and to perform source attribution and outbreak investigation.

4.2. New and emerging pathogens

Lena Feige delivered a comprehensive talk on new and emerging pathogens. She explained the concept of (re-)emerging infectious diseases, which either appear in a population for the first time or rapidly increase in incidence or spread geographically. She highlighted the historical case of the 1976 Ebola virus disease outbreak in Yambuku and how sequencing helped to detect the source of later outbreaks. She also mentioned the critical need for global surveillance and rapid response systems. In addition, the importance of standardised pathogen nomenclature and epidemiological and genomic data integration to enhance outbreak detection and response efforts was emphasized.

4.3. Biomarker discovery

Laurent Jacob discussed advances in microbial genome-wide association studies (GWAS) aiming to identify correlations between genetic variants and traits, such as antibiotic resistance. He described the standard approach of using single nucleotide polymorphisms (SNPs) or short indels and the challenges of capturing genetic variation in the accessory genome. Using k-mers (short sequences) as variants is an agnostic way to capture any type of genetic variation. An own developed method was presented, which uses De Bruijn graphs to structure these k-mers, improving the interpretability of GWAS results and enabling the identification of significant genetic determinants of resistance [11]. He also described how neural networks combined with hypothesis testing could capture more complex genetic information.

4.4. Prediction of resistance in Gram-negative bacteria

Mohammad Razavi covered the prediction of antibiotic resistance in Gram-negative bacteria. He emphasized the significance of detecting resistance to improve treatment efficacy, patient outcomes, and surveillance. He discussed the concept of the resistome (intrinsic and acquired resistance genes), the role of mobile genetic elements in spreading antibiotic resistance genes, and the impact of antibiotic pressure on bacterial evolution. He highlighted the use of WGS and various bioinformatic pipelines, such as the Swedish GMS's Nextflow pipeline “JASEN”, for analysing genomic data to predict resistance mechanisms and inform clinical decisions.

4.5. Prediction of antibiotic resistance in STDs

Mathew Beale focused on key bacterial sexually transmitted infections (STIs), including *Neisseria gonorrhoeae, Treponema pallidum, Chlamydia trachomatis*, and *Mycoplasma genitalium*. Genomics has not yet been demonstrated to be useful for directly influencing patient outcomes in STIs, but can help to develop effective vaccines through phylogenetically informed surveillance of variants. Genomic data are also important for tracking the spread of antimicrobial resistance. He mentioned the example of *T. pallidum,* causative agent of syphilis, and the high prevalence of macrolide resistance in many settings. Using this, resistance could be linked to one specific sublineage in the UK [12,13]. However, a new policy of prophylactic doxycycline for individuals at high risk of STI exposure could lead to resistance emergence in targeted but also untargeted organisms. He also discussed a potential genomic surveillance system to monitor the outcomes of this policy decision.

4.6. Prediction of antibiotic resistance in M. tuberculosis

Stefan Niemann discussed sequencing approaches for predicting drug resistance in *M. tuberculosis*. He highlighted the global challenge of multidrug-resistant (MDR) and extensively drug-resistant (XDR) *M. tuberculosis*. He outlined the use of WGS for rapid molecular drug susceptibility testing (DST) and the development of algorithms to predict resistance [14]. He also emphasised the importance of implementing sequencing-based diagnostics in high-incidence settings and the need for comprehensive workflows to support TB surveillance and treatment.

5. Foster new technologies

5.1. What we need from policy makers

Emma Hodcroft presented on the crucial role of policy makers in supporting genomic surveillance and sequencing for public health. She highlighted the importance of diversified funding sources to avoid reliance on a single stream, citing the example of SPSP, which combines governmental and research grants. She emphasized the importance of setting clear and achievable objectives, developed in consultation with scientists, to effectively guide public health initiatives. She called for clear, international policies for data sharing, particularly for sensitive information. She stressed the importance of workforce training and retention, particularly in genomics and bioinformatics, to maintain a skilled and knowledgeable team

capable of responding to public health crises. Additionally, she advocated for using "downtime" to gather baseline data from environmental and animal samples, which can be critical during outbreaks. Finally, the importance of tailoring pathogen prioritization and surveillance policies to local contexts rather than blindly following international guidelines was emphasized.

5.2. What we need from industry

Gilbert Greub discussed the needs from the industry to improve diagnostics through three primary avenues: automation, artificial intelligence, and bioinformatics. He outlined how industry innovations can drive the development of new sequencing technologies and improve public health outcomes. He also highlighted the challenges posed by the need to comply with In Vitro Diagnostic Regulation (IVDR) [15]. He underscored the importance of maintaining capacity for research and development despite the tendency of some industries to cease investment in research and development for existing tests. He stressed the need for robust technology, stable chemistry, and reliable stock management in routine diagnostics.

5.3. Future data management

Per Sikora emphasized the importance of data management and informatics in precision medicine, which requires integrating large amounts of diverse data, including omics, QC, biometric, treatment, economic, and clinical outcomes. He stressed the need for standardized data organization and machine-to-machine interfaces (APIs) to facilitate easy data access and sharing. He warned against data silos (professional, semantic, organizational, and platform) and advocated for active efforts to break them down. He also discussed the importance of developing national and international data infrastructures to support large-scale genomic projects and improve public health outcomes through better data management.

5.4. New technologies - a data science perspective

Andre Kahles provided a data science perspective on microbial genomics, discussing the potential of new technologies in research such as microbial pangenomes and long real-time sequencing. He introduced the concept of metagraphs, which are vectors of labels that can be compressed to allow efficient querying, enabling effective phenotype–genotype associations. He also highlighted the use of long-read real-time sequencing to enhance the accuracy and resolution of genomic data, emphasizing the importance of these advancements in improving our understanding of microbial diversity and the functional implications of genetic variations.

5.5. Impact of WGS in daily practice

Martin Sundqvist discussed the practical implications of WGS in clinical settings. He highlighted the challenges faced by many laboratories, such as the need for expertise in bioinformatics, and the lack of baseline data for effective local outbreak management. Further, he emphasized that suspected outbreaks sometimes involve unexpected pathogens, underscoring the necessity for continuous sequencing, accessible databases and easy to use tools for data analysis.

6. Workshop: working on a shared guideline

During the conference, a workshop was conducted with all participants to develop shared laboratory guidelines for bacterial WGS across Europe. The participants were divided into three groups, each focusing on one of the key areas: pre-analytical, analytical, and post-analytical challenges and solutions. This collaborative effort aimed to ensure consistency and quality in laboratory practices across Europe and identified several opportunities for standardization (Fig. 2).

6.1. Pre-analytical

The review of the pre-analytical process of laboratory work emphasized the need for standardized extraction protocols for different bacteria, considering whether to use manual or machine-based methods, and for awareness on how variations in workflow and staff can affect the results. The workshop members noted that the quality of samples from pure cultures is generally not a concern. Establishing minimal standards for DNA concentration and quality for WGS sequencing platforms was discussed, along with whether including positive and negative controls is necessary. The participants agreed on standardizing library concentration measurements with instruments like NanoDrop, Qubit, and TapeStation. Post-

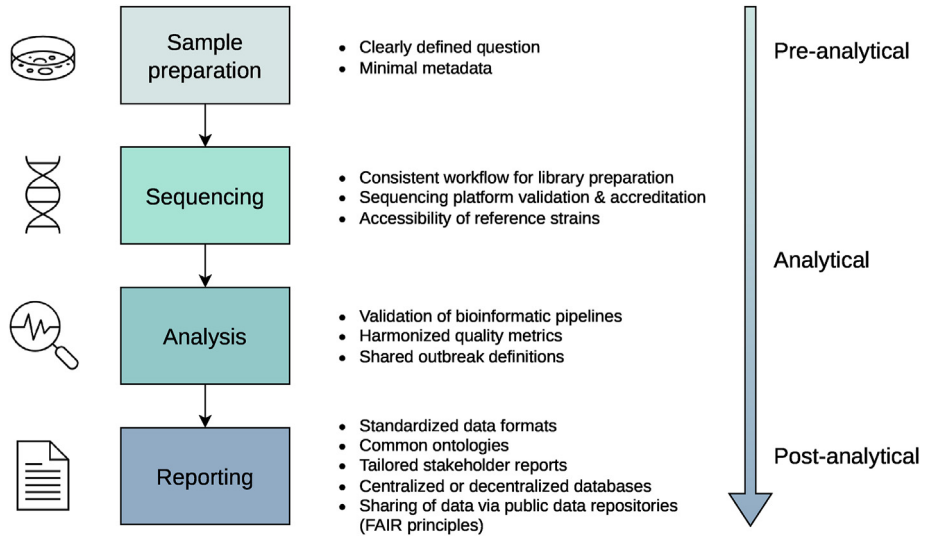

Fig. 2. Whole genome sequencing (WGS) workflow steps and the main standardization opportunities identified during the workshop.

analytical considerations such as normalization, genome size, GC/AT richness and coverage were highlighted, as well as the need for proper metadata management, which some labs lack. Various sequencing machines and their suitability were discussed underlining the need for different methods according to the test indication, urgency and context. Finally, the ability to run sequence based diagnostic facilities cost-efficiently is the biggest challenge for many labs.

6.2. Analytical QC

The key analytical challenges in WGS include the lack of harmonized quality metrics. These are to an extent dependent on the specific questions (species identification, epidemiological typing, AMR prediction), and the pathogen in question. Establishing standardized metrics for these parameters across different pathogens and use cases is essential. The panellists discussed specific recommendations, such as ideal average coverage of 30× for typing, 50× for AMR prediction and 80× for phylogenetics, while minimal coverage in targeted positions may be lower. Downsampling methods can help determine these necessary coverage levels for different species. Additionally, robust assembly techniques should be adopted, such as verifying assembly quality through remapping reads. Quality filters should be implemented to maintain data integrity. Specific guidelines for AMR prediction should address challenges like mixed strains and ensure robust assembly, while maintaining up-to-date AMR databases is crucial. The workshop also emphasized the need for reducing sequencing turnaround times to enhance clinical utility and support effective implementation of standardized QC processes.

6.3. Post-analytical

The post-analytical workshop discussed various aspects of post sequencing including the establishment of reproducible and standardized pipelines, comparison of results, and how to inform the different stakeholders after the NGS results are generated. First, streamlined standards are critical to the validation and standardization of WGS pipelines. Establishing a gold standard for WGS pipeline validation is essential, but challenges include defining this standard and accessing affordable reference strains. Validating pipelines to detect single nucleotide polymorphisms (SNPs), especially those with small coverage proportions, also presents difficulties. Furthermore, ensuring transferrable and comparable results requires standardized data formats and a common language, along with clear cluster definitions for comparing isolates. The necessity of external quality controls was debated, with questions about how pipelines respond to updates, who would provide the controls, and what they would include. The importance of shared outbreak definitions was highlighted to ensure consistent and effective responses across different regions. Finally, the workshop emphasized the need for clear, concise reports, which use a common language (ontologies) and are tailored to different needs and levels of detail, ensuring the information is understandable and useful to all stakeholders. The importance of funding for stakeholder education and combating false information was also highlighted.

7. Conclusion

The "Bacterial Genome Sequencing Pan-European Network" conference successfully brought together experts from across Europe to discuss the latest advancements and challenges in bacterial genome sequencing. Key themes included the importance of standardization, data sharing, and quality control in genomics workflows. The event highlighted the need for harmonized quality metrics and validated pipelines to ensure reliable results. It also emphasized the importance of collaboration between research institutions, public health bodies, and industry to foster innovation and improve public health outcomes. Crucially, the conference underscored the necessity of validating use cases and conducting studies to demonstrate the impact and cost-effectiveness of WGS on clinical and public health practices. Participants left with a renewed sense of purpose and a commitment to ongoing collaboration and innovation driving the future of genomics in clinical microbiology. While this was the first Bacterial Genome Sequencing Pan-European Network Conference, a second edition of the meeting is in planning for March 2025. We aim to expand our current pan-European network to include more European countries, including a more diverse scientific committee, transform it towards a winter school focusing on sequencing, and eventually reach a global scale. By showcasing our discussions and findings in this report, we can engage potential new members, both within and beyond the networks of past participants.

Funding

We want to thank the Swiss National Science Foundation for financial support of travelling costs (IZSEZ0_222471). We also thank the University of Zurich and the Swedish "Nyckelfonden" - Örebro University hospital research Foundation, and Genomic Medicine Sweden for financial support of the conference.

CRediT authorship contribution statement

Zoja Germuskova: Writing – original draft, Project administration, Methodology. **Elisa Sosa:** Writing – original draft, Project administration, Methodology. **Amaya Campillay Lagos:** Writing – original draft, Project administration. **Hege Vangstein Aamot:** Writing – review & editing, Investigation. **Mathew A. Beale:** Writing – review & editing, Investigation. **Claire Bertelli:** Writing – review & editing, Investigation. **Jonas Björkmann:** Writing – review & editing, Project administration, Investigation. **Natacha Couto:** Writing – review & editing, Investigation. **Lena Feige:** Writing – review & editing, Investigation. **Gilbert Greub:** Writing – review & editing, Investigation. **Erika Tång Hallbäck:** Writing – review & editing, Project administration, Investigation, Conceptualization. **Emma B. Hodcroft:** Writing – review & editing, Investigation. **Dag Harmsen:** Writing – review & editing, Investigation. **Laurent Jacob:** Writing – review & editing, Investigation. **Keith A. Jolley:** Writing – review & editing, Investigation. **Andre Kahles:** Writing – review & editing, Investigation. **Alison E. Mather:** Writing – review & editing, Investigation. **Richard A. Neher:** Writing – review & editing, Investigation. **Aitana Neves:** Writing – review & editing, Investigation. **Stefan Niemann:** Writing – review & editing, Investigation. **Oliver Nolte:** Writing – review & editing, Investigation. **Sharon J. Peacock:** Writing – review & editing, Investigation. **Mohammad Razavi:** Writing – review & editing, Investigation. **Tim Roloff:** Writing – review & editing, Investigation. **Jacques Schrenzel:** Writing – review & editing, Investigation. **Per Sikora:** Writing – review & editing, Investigation. **Martin Sundqvist:** Writing – review & editing, Investigation. **Paula Mölling:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Adrian Egli:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of AI and AI-assisted technologies in the writing process

During the preparation of this report, the authors used chatGPT 4o (OpenAI) to convert hand-written notes into text and summarize notes and presentation slides for a first draft. After using this tool, all authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

RAN is a paid consultant for ModernaTX and BioNTech. SJP is a consultant for Next Gen Diagnostics. GGR is scientific advisor of Resitell (Muttenz, Switzerland), a start-up implicated in nanomotion-based measure of antimicrobial resistance. DH is managing director and shareholder of the company Ridom GmbH that develops the SeqSphere + tool. AE is scientific advisor of Sefunda (Muttenz, Switzerland).

Acknowledgment

We want to specifically acknowledge Birke Mebold (Institute of Medical Microbiology, University of Zurich, Switzerland) for excellent organisation and administrative support of the conference.

References

[1] Balloux F, Brynildsrud OB, Van Dorp L, Shaw LP, Chen H, Harris KA, et al. From theory to practice: translating whole-genome sequencing (WGS) into the clinic. Trends Microbiol 2018;26(12):1035–48.
[2] Mutschler E, Roloff T, Neves A, Vangstein Aamot H, Rodriguez-Sanchez B, Ramirez M, et al. Towards unified reporting of genome sequencing results in clinical microbiology. Peer J 2024;12:e17673.
[3] Egli A, Blanc DS, Greub G, Keller PM, Lazarevic V, Lebrand A, et al. Improving the quality and workflow of bacterial genome sequencing and analysis: paving the way for a Switzerland-wide molecular epidemiological surveillance platform. Swiss Med Wkly 2018;148:w14693.
[4] Neves A, Walther D, Martin-Campos T, Barbie V, Bertelli C, Blanc D, et al. The Swiss Pathogen Surveillance Platform - towards a nation-wide One Health data exchange platform for bacterial, viral and fungal genomics and associated metadata. Microb Genom 2023;9(5):mgen001001.
[5] Wegner F, Cabrera-Gil B, Tanguy A, Beckmann C, Beerenwinkel N, Bertelli C, et al. How much should we sequence? An analysis of the Swiss SARS-CoV-2 surveillance effort. Microbiol Spectr 2024;12(5):e0362823.
[6] Edsjö A, Lindstrand A, Gisselsson D, Mölling P, Friedman M, Cavelier L, et al. Genomic Medicine Sweden (GMS). Building a precision medicine infrastructure at a national level: the Swedish experience. Camb Prism Precis Med. 2023;1:e15.
[7] Fioretos T, Wirta V, Cavelier L, Berglund E, Friedman M, Akhras M, et al. Implementing precision medicine in a regionally organized healthcare system in Sweden. Nat Med 2022;28(10):1980–2.
[8] Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics 2018;34(23):4121–3.
[9] Dylus D, Pillonel T, Opota O, Wüthrich D, Seth-Smith HMB, Egli A, et al. NGS-based *S. aureus* typing and outbreak analysis in clinical microbiology laboratories: lessons learned from a Swiss-wide proficiency test. Front Microbiol 2020;11:591093.
[10] Wegner F, Roloff T, Huber M, Cordey S, Ramette A, Gerth Y, et al. External quality assessment of SARS-CoV-2 sequencing: an ESGMD-SSM pilot trial across 15 European laboratories. J Clin Microbiol 2022;60(1):e0169821.
[11] Jaillard M, Lima L, Tournoud M, Mahé P, van Belkum A, Lacroix V, et al. A fast and agnostic method for bacterial genome-wide association studies: bridging the gap between k-mers and genetic events. PLoS Genet 2018;14(11):e1007758.
[12] Beale MA, Thorn L, Cole MJ, Pitt R, Charles H, Ewens M, et al. Genomic epidemiology of syphilis in England: a population-based study. Lancet Microbe 2023;4(10):e770–80.
[13] Lieberman NAP, Reid TB, Cannon CA, Nunley BE, Berzkalns A, Cohen SE, et al. Near-universal resistance to macrolides of *Treponema pallidum* in North America. N Engl J Med 2024;390(22):2127–8.
[14] Sibandze DB, Kay A, Dreyer V, Sikhondze W, Dlamini Q, DiNardo A, et al. Rapid molecular diagnostics of tuberculosis resistance by targeted stool sequencing. Genome Med 2022;14(1):52.
[15] Coste AT, Egli A, Schrenzel J, Nickel B, Zbinden A, Lienhard R, et al. IVDR: analysis of the social, economic, and practical consequences of the application of an ordinance of the in Vitro diagnostic ordinance in Switzerland. Diagnostics 2023;13(18):2910.