

Next-generation sequencing in routine clinical microbiology and infectious diseases: an ESGMD-ESGEM ESCMID postgraduate course

Hege Vangstein Aamot

Department of Microbiology and Infection Control, Akershus University Hospital, Lørenskog, Norway

Eric C. J. Claas

Department of Medical Microbiology, Leiden University Medical Center, Leiden, the Netherlands

Natacha Couto

The Milner Centre for Evolution, Department of Biology and Biochemistry, University of Bath, Bath, United Kingdom and Department of Medical Microbiology and Infection Control, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands

Henrik Westh

Department of Clinical Microbiology, Amager-Hvidovre Hospital, Hvidovre, Denmark and Institute of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark

Gilbert Greub

Institute of Microbiology, University of Lausanne and University Hospital Center, Lausanne, Switzerland

John W. A. Rossen

Department of Medical Microbiology and Infection Control, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands, Laboratory of Clinical Microbiology and Infectious Diseases, Isala Hospital – Zwolle, the Netherlands and Department of Pathology, University of Utah School of Medicine, Salt Lake City, UT, USA

Original Submission: 7 September 2022; **Revised Submission:** 4 November 2022; **Accepted:** 4 November 2022

Article published online: 12 November 2022

Corresponding author. Akershus University Hospital, Dept of Microbiology and Infection Control, P.O. Box 1000, 1478, Lørenskog, Norway.

E-mail: hege.vangstein.aamot@ahus.no

¹Member of the Executive Committee of the ESCMID Study Group for Genomic and Molecular Diagnostics (ESGMD)

²Member of the Executive Committee of the ESCMID Study Group for Epidemiological Markers (ESGEM)

³Member of the Executive Committee of the ESCMID Study Group for Genomic and Molecular Diagnostics (ESGMD)

Dear Editor,

With multiple benefits, next-generation sequencing (NGS) is revolutionising microbiological diagnostics and is on the way to becoming the new gold standard in many settings. In order to highlight the opportunities and challenges with NGS technology and present how it can be used in clinical diagnostic settings, the European Society for Clinical Microbiology and Infectious Diseases (ESCMID) study groups for genomic and molecular diagnostics (ESGMD) and molecular markers (ESGEM) organised a three-days, postgraduate, online course on 27-29 September 2021. Almost 300 epidemiologists, clinical microbiologists, infectious diseases specialists, veterinarians, bioinformaticians and biologists from around the world joined, resulting in a unique meeting around molecular diagnosis, genomics, and microbiota analysis. The online format allowed more participants compared to previous onsite ESCMID workshops/courses, supporting the organisation of more online or hybrid format courses.

The course started with a general introduction to NGS sequencing, the different sequencing technologies and platforms, as well as bioinformatics and visualisations of results. Next, examples of the many clinical applications using NGS sequencing, including amplicon, whole-genome, and metagenomic sequencing, were presented. Possibilities and challenges with NGS sequencing in a clinical setting were discussed. Finally, the benefits of using machine learning in these settings were reviewed. For full program, see Supplement File 1.

The introduction to NGS emphasized that NGS should not be divided into second and third generation but should be referred to as massively parallel sequencing. NGS can be divided into short-read sequencing (SRS) and long-read sequencing (LRS). SRS needs an amplification step, whereas LRS can be performed “real-time” from a single RNA or DNA molecule without prior reverse transcription or amplification. Longer reads are easier to assemble, even without using a reference genome. This is especially helpful for sequencing mobile genetic elements such as plasmids. However, platform choice depends on the application, cost, required throughput and turnaround time.

Bioinformatics is often considered a significant obstacle to performing NGS. Therefore, it is vital to understand which bioinformatics tools to use, how to use them, and how to interpret the results. Several bioinformatics tools, workflows

and resources were presented, and the participants were challenged to overcome any fear of using a command-line interface. The importance of the bioinformatics pipelines' quality control (QC) was also discussed, highlighting using certified reference materials, standardised validation schemes, international proficiency testing programs and accreditation.

The analyses and visualisation of metagenomics data were introduced, discussing the general possibilities and challenges with metagenomic sequencing. As metagenomics can detect all microbial and host nucleic acids in any sample, it is crucial to define the objective for testing and determine how the relevant data should be presented. This impacts every workflow step, from sample collection to the final report. Complex results need to be comprehensively visualised for non-experts. Defining reporting thresholds is also necessary but can be challenging without a gold standard. Using reference material can provide a starting point.

The second day was dedicated to the application of NGS in clinical laboratories. The overall aim is to provide NGS data in a timely frame that has a clinical impact on patient management and costs. Long-read technology was suggested to replace traditional microbiology in the future, with a sample-to-sequence turnaround time of less than an hour and genomic antibiogram and strain typing within a few hours. The added value of reporting virulence factors (the virulome) was also exemplified. However, the lack of standardisation in the databases available to identify virulence factors is challenging.

Different NGS applications were presented. Amplicon sequencing was exemplified by genomic characterization of human papillomavirus for possible use in cervical cancer diagnostics [1] and an improved workflow for 16S rRNA gene sequencing as a tool for bacterial identification [2]. Several applications of whole-genome sequencing (WGS) were given, including outbreak investigations, rare species identification, virulence factors, and antimicrobial resistance (AMR) prediction.

WGS can be applied for outbreak investigations using cgMLST (core genome multilocus sequence typing) or SNP (single nucleotide polymorphism)-based typing. Both approaches have advantages and disadvantages, and the resolution level depends on the approach. Broader typing by MLST/cgMLST is helpful to group genomes. However, an SNP-based approach should be used to analyse the whole genome to improve the resolution in outbreak scenarios.

AMR prediction by WGS can be complicated as minimum inhibitory concentrations (MIC) depend not only on the presence or absence of genes and mutations but also on the permeability of the cell wall and the efflux of drugs. The importance of selecting a curated and updated database and applying machine learning to better predict the MIC by WGS was emphasized.

The last day of the course was dedicated to the applications of metagenomic approaches and future perspectives. "True" metagenomics captures the complete genomic information of a given ecological community and allows an exact taxonomic classification of the microbes present in the samples. In contrast, amplicon- and enrichment-based metagenomics is NGS applied to specific genome regions, thereby capturing only a part of the genomes. It can be taxonomically less accurate but eases the binning of the sequenced amplicons in predefined taxa and is often more sensitive than shotgun metagenomics. The ideal main requirements for "clinicogenomics" were defined as: (i) unbiased non-*a priori* approach to detect any microbes and parasites, (ii) taxonomic assignment of sequences and estimation of their relative abundance, (iii) short time to results, and (iv) validation in terms of sensitivity and specificity. Moreover, to keep costs low (in terms of human resources, i.e., bioinformaticians' time) and to get implemented in many laboratories, the "clinicogenomics" pipeline should be easy to use.

How and when to use shotgun metagenomic approaches on infected tissue was illustrated by application on patients with orthopaedic implant-associated infections [3]. The utility of viral metagenomics was shown on SARS-CoV-2 samples and patients with encephalitis, gastroenteritis and prolonged fever using the MetaMix pipeline (<https://github.com/smorfopoulou>). Metagenomics on bacteraemia was also presented.

Typing is essential for surveillance and outbreak investigation in hospital, regional and national settings. Metagenomics can give insights into pathogenesis, population structure and microbial genetic diversity and allows typing of unculturable microorganisms. For example, cgMLST results for *E. faecium* using WGS and shotgun metagenomics gave identical results at coverage of 50x sequencing depth in shotgun metagenomics.

Future perspectives were given on machine learning (ML), what it is, why we need it, and if it can move the field of clinical microbiology forward. The three machine learning methods were reviewed, i.e., supervised, unsupervised, and reinforcement learning. ML was used to predict resistance from MALDI-TOF results using ~300,000 spectra and 78 antibiotics. A take-home message was that the doubling of knowledge is 73 days, and there is no way a person can follow this – but a computer can. More information on the digitalisation of microbiological workflows to efficiently use big data can be found in "Digital microbiology" [4].

Examples of the use of ML in clinical metagenomics were given, e.g., integrating host response and unbiased microbe detection for lower respiratory tract infection diagnosis in critically ill adults [5]. Although technical improvements and cost reduction of sequencing resulted in more metagenomics data that could be input for developing ML diagnostic approaches, the development of open-access databases of

microbiota and WGS data is a strong asset for the development of ML systems and should be further encouraged. The impact on processes and clinical outcomes should be addressed before using ML in routine clinical practice.

The course ended with a presentation on NGS to meet future challenges in clinical microbiology. Examples were given on the use of WGS for extensively drug-resistant *Pseudomonas aeruginosa*, Gram-negative Enterobacteriaceae and vancomycin-resistant enterococci outbreaks and the role of the environment and healthcare clusters for their spread. Typing for SARS-CoV-2 has created an essential focus on the value of molecular epidemiology, continuous surveillance and data sharing. Having a financial structure for real-time sequencing is essential.

Ethics statement

Nothing to declare. This is a meeting report.

Conflict of interest

John W. A. Rossen is an advisor for IDbyDNA, ARES-genetics and MolZym.

Acknowledgements

The ESCMID postgraduate meeting was organised by Hege Vangstein Aamot and John W. A. Rossen on behalf of ESCMID European Study Group on Genomic and Molecular Diagnostics

(ESGMD), ESCMID Study Group for Epidemiological Markers (ESGEM), Akershus University Hospital (Norway) and University Medical Center Groningen (Netherlands). It was supported by the European Society for Clinical Microbiology and Infectious Diseases (ESCMID). A special thanks to Jeanette Borthen and William Caddy for their help with the organisation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.nmni.2022.101046>.

References

- [1] Lagström S, Umu SU, Lepistö M, Ellonen P, Meisal R, Christiansen IK, et al. TaME-seq: an efficient sequencing approach for characterisation of HPV genomic variability and chromosomal integration. *Scientific Reports* 2019;9(1):524.
- [2] Sabat AJ, van Zanten E, Akkerboom V, Wisselink G, van Slochteren K, de Boer RF, et al. Targeted next-generation sequencing of the 16S-23S rRNA region for culture-independent bacterial identification - increased discrimination of closely related species. *Scientific Reports* 2017;7(1): 3434.
- [3] Noone JC, Helmersen K, Leegaard TM, Skråmm I, Aamot HV. Rapid diagnostics of orthopaedic-implant-associated infections using nanopore shotgun metagenomic sequencing on tissue biopsies. *Microorganisms* 2021;9(1).
- [4] Egli A, Schrenzel J, Greub G. Digital microbiology. *Clin Microbiol Infect* 2020;26(10):1324–31.
- [5] Langelier C, Kalantar KL, Moazed F, Wilson MR, Crawford ED, Deiss T, et al. Integrating host response and unbiased microbe detection for lower respiratory tract infection diagnosis in critically ill adults. *Proc Natl Acad Sci U S A* 2018;115(52). E12353–e62.

