Exploring next-generation sequencing in clinical microbiology

Peker, Nilay

DOI:
10.33612/diss.218469374

Chapter 9

General discussion and future perspectives

General Discussion

Optimizing the diagnostic process is crucial in managing infectious diseases (1). In this respect, clinical microbiology laboratories look for diagnostic tests that provide relevant, clinically useful answers within a clinically actionable time frame to promote appropriate treatment and help prevent the further spread of resistance. Early administration of appropriate antimicrobial therapy is vital for a patient not only for successful treatment but also to prevent side effects. In addition, inappropriate use of antimicrobials is a primary driver of AMR. Moreover, delay in specific clinical diagnosis may also result in increased length of stay, a higher complication rate, and ultimately increased medical costs (2).

In recent years, the introduction of molecular technologies, mainly based on nucleic-acid amplification and next-generation sequencing (NGS), has presented a quantum leap in microbial diagnostics (3), together with the routine use of MALDI-ToF MS, mainly for species identification (4). Specifically, for BSIs, many methods, like the ones described in Chapter 2, have been developed, all with their drawbacks (2). Some have been implemented in routine clinical diagnostics. Appropriate utilization of available diagnostic assays (considering, e.g., panel spectrum, sensitivity, specificity, and cost-effectiveness) is an essential aspect of infection management and constitutes one of several main components of Diagnostic Stewardship (5-7). Therefore, laboratories should clearly define the intended use of the available diagnostic assays depending on their settings/infrastructure, clinical questions and patient population.

The introduction of NGS as a high-throughput sequencing method has enabled simultaneous detection of all pathogen classes, i.e., bacteria, viruses, fungi, and parasites, directly from patient samples (syndromic molecular testing: e.g., panels for neutropenic fever, respiratory, gastrointestinal, and CNS infection panels) comparable to targeted methods such as multiplex PCR panels but with a much broader scope (8,9). Recent developments in NGS technologies, including decreasing cost, faster turnaround times (TAT) and the development of low to medium throughput platforms, have expanded the application of NGS as a diagnostic tool for identification, characterization and typing of pathogens and detection of antimicrobial resistance determinants. NGS techniques allow the identification of uncultivable microorganisms that conventional methods would not detect. In addition, slow-growing bacteria are detected faster (2). For instance, the prolonged time required for phenotypic susceptibility testing for *M. tuberculosis,* which usually requires 21 days to grow in culture and another 28 days for a first-line antimicrobial susceptibility test result (10), has considerably increased the interest in the use of NGS for the characterization

of *M. tuberculosis* and associated resistance patterns with the potential of same-day diagnosis directly from patient samples (11).

WGS and metagenomics sequencing have found their place within clinical microbiology laboratories (12,13) for identification, typing, and/or antimicrobial susceptibility/virulence prediction of pathogens (Fig. 1). However, several challenges, ranging from sampling, nucleic acid extraction, library preparation to sequencing and data analysis, are still faced. Fig.1 explains the NGS workflows for WGS and metagenomics studied in this thesis. NGS workflows can be divided into a wet-lab (sample processing, DNA extraction, and sequencing) and a dry-lab (data analysis and interpretation) part (14). Both parts need improvement for more straightforward implementation into diagnostics. Several aspects of these workflows were further investigated in this thesis.

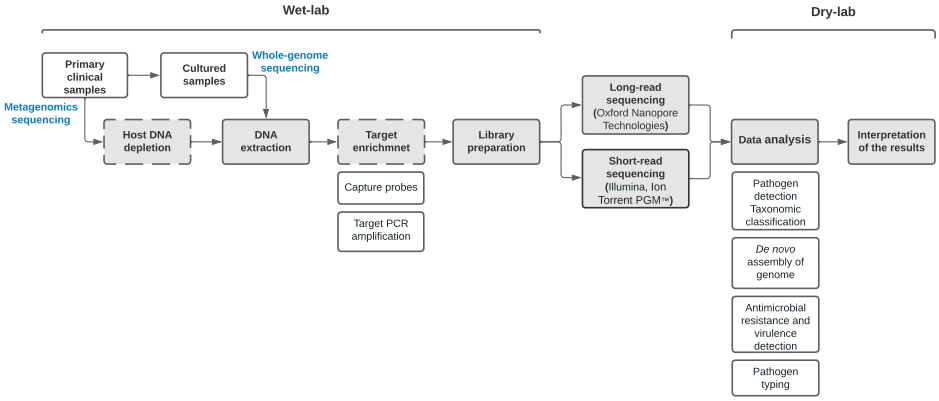

Fig. 1: Next-generation sequencing workflow for whole-genome sequencing (cultured samples) and metagenomics sequencing (primary clinical samples). The dashed boxes indicate optional steps. The workflow is divided into wet-lab (from sampling to sequencing) and dry-lab (data analysis and interpretation) procedures. Grey boxes indicate steps further investigated in Chapters 3, 4, 5, 7 and 8.

1. Wet-lab aspect

1.1. Sampling, storage conditions, standardization, and proper controls of the NGS protocol are crucial for the interpretation of the results

Sample collection and the following laboratory processing are well-known sources of contamination found in the environment and reagents (“kitome”) (15). While the detection of extraneous nucleic acids of commensals or contaminants influences the specificity of NGS methods, one should also be cautious about distinguishing microorganisms causing colonization from the ones causing infection. This is especially important for species that can be both

commensals and pathogens, like, for example, *Cutibacterium acnes* (formerly *Propionibacterium acnes*). In Chapter 5, we presented a method to distinguish between true pathogens and potential contaminating species based on calculating a cut-off value (16), but other methods exist, described in Chapter 6 (13). Several negative controls, e.g., sampling blank, nucleic acid extraction blank, and/or no-template control, are necessary to exclude possible contamination, which can be introduced at any step, from sampling to sequencing and data handling (e.g., non-curated databases that contain erroneous entries). It is essential to highlight that negative controls should be included in every routine sequencing run, not only during the method development. Besides, reference standards (e.g., ATCC® Microbiome Standards (17,18), ZymoBIOMICS Microbial Community Standards (19), MS2 [RNA] and T1 [DNA] bacteriophage (20,21)) of whole microorganisms, viruses, or nucleic acids can be used as external/internal controls to monitor nucleic acid extraction efficiency for different pathogen classes or, when spiked into clinical samples to quantify pathogens. Notably, the concentration of the spike-in controls should be carefully chosen to not compete with pathogens present at low loads while still providing sufficient sensitivity, also for downstream steps in the process (27).

Moreover, the PhiX bacteriophage, commonly used as a quality and calibration control, enables assessing the error rate of a specific run. However, one should also be cautious about "contaminated" genomes, for which PhiX might get integrated as part of the target genome (22). On the other hand, including a well-characterized strain in each WGS run would benefit the downstream data analysis. For example, sequencing a well-characterized *M. tuberculosis* strain together with other *M. tuberculosis* isolates would be a helpful control to check the accuracy of sequencing and identifying their single nucleotide polymorphisms (SNPs) by bioinformatics tools, such as Mykrobe and TBProfiler, studied in Chapter 4. Overall, clinical NGS methodologies need proper external and internal controls, which are essential for standardization and ensuring quality and reproducibility (13).

Storage conditions of the biological samples are another important aspect of the reliability of diagnostic test results. Currently, clinical samples sent for metagenomics sequencing are usually leftover samples that have been frozen and thawed many times, or they represent old samples used for retrospective studies. In general, the sample quality is decreased after freeze-thaw cycles, which complicates the reliability and interpretation of the results. Microbiome studies have revealed that changing storage conditions like temperature, storage time, and the number of freeze-thaw cycles have significant and systematic effects on the microbiome composition (including the relative abundance of taxa) and observed interactions between microbes and the microbiome and the host (23) (24,25). Optimal storage conditions vary

depending on the sample type and are particularly critical for low-biomass samples (26). Consequently, sampling, optimum storage conditions, and proper controls are essential for the reliability of the results and should be carefully planned in future metagenomics studies.

1.2. Nucleic acid extraction is a persistent bias in NGS protocols

The unbiased nature of SMg allows the sequencing of the nucleic acid of all pathogens (including commensal microbes) and the host. However, this unbiased nature of SMg might lower the sensitivity of pathogen detection. As the sequencing library comprises both nucleic acids from the patient and pathogens, the sequence coverage of the pathogen depends on the ratio of host/pathogen nucleic acid present in the sample (27). Host-depletion may help to increase the genome coverage of the pathogen. As mentioned in Chapter 6, several approaches (e.g., differential cell lysis, centrifugation, DNase/RNase treatment, CRISPR-Cas9, capture probes) can be applied pre- and post-lysis to deplete host nucleic acids with pros and cons and varying depletion efficiency, also depending on the sample type (13). Chapters 7 and 8 of this thesis studied the benchmarking of different nucleic acid extraction protocols for DNA isolation from different clinical sample matrices spiked with various bacteria to address the nucleic acid extraction bias. We applied two methods (one commercial and one in house) for differential cell lysis, which led to variable results depending on the sample matrices (Chapter 7, blood and plasma; Chapter 8, sputum). In Chapter 8, the saponin-depletion method was shown to be an efficient option to deplete host DNA (28,29), resulting in a 60-fold decrease and a 12-fold increase in the percentage of host (human) and spiked *Mycobacterium abscessus* reads, respectively. This resulted in 100% genome coverage of *M. abscessus* and > 30x sequencing depth of coverage. Chapters 7 and 8 also demonstrated that no single sample processing (e.g., nucleic acid extraction) approach applies to all different clinical sample types. For instance, the viscosity of blood, affected by plasma composition, varies among individuals and increases in case of infection (30). Consequently, protocols developed for nucleic acid extraction from plasma might be limited by the amount obtained from the whole blood, depending on its biophysical properties. Moreover, one should also consider the sample type and the differences in lysis efficiency of different bacterial populations. For example, a kit used for molecular diagnosis to detect a specific set of pathogens may not serve the purpose of SMg. It has been shown that the bacterial community composition in a sample is highly influenced by the extraction protocol (31). Some protocols may specifically favor the extraction of Gram (+) bacteria or Gram (-) bacteria (32).

All in all, nucleic acid extraction protocols, following host depletion and subsequent downstream analysis steps (e.g., NGS library preparation) should

be well-defined considering the biophysical properties of clinical samples (including the biological variation from the host) to obtain reliable results.

1.3. The longer the fragments, the higher the resolution

The selected library preparation protocol and the sequencing platforms (short- or long-read) ultimately affect the resolution of the diagnostic results. As the widely used Illumina sequencing platform produces short sequences up to 300 bp, sequencing of different sub-regions of the 16S rRNA gene (e.g., V1-V3, V3-V4) has been commonly used to characterize the microbiome and to identify infection-causing pathogens (33,34) with lower cost compared to Sanger sequencing. However, studies have shown that species discrimination often cannot be achieved (reliably) by targeting the variable sub-regions of the 16S rRNA gene with short-read sequencing platforms (33). Likewise, sequencing a larger genomic region, covering the entire 16S-23S rRNA gene, has shown higher resolution for species discrimination in our study in Chapter 5 (16). Illumina sequencing features different sequencing kits, providing increased cluster density (increased number of reads) and extended read length for different platforms (e.g., MiSeq, NextSeq). In our evaluation of the MiSeq 300-cycles, 500-cycles, and 600-cycles reagent kits (Chapter 5), the relative abundance of pathogens identified at the species level using 16S-23S rRNA amplicon sequencing was highest with the MiSeq 600-cycles Reagent Kit (Illumina) known to generate the longest reads (16). Using a technology resulting in reads spanning the whole amplicon region may get an even higher resolution to discriminate between closely related species. This has been the case in a recent microbiome characterization study that revealed a higher resolution for distinguishing between members of particular taxa with long-read compared to short-read sequencing of the full-length 16S rRNA gene (35). Therefore, long-read sequencing (e.g., nanopore sequencing) of the whole 16S-23S rRNA gene holds great promise for discriminating pathogens even at the strain level.

Long-reads also have the advantage of resolving structural variations and variants in repetitive regions, which are poorly resolved by short-reads and are often excluded in bioinformatics analysis. For instance, a better understanding of the role of highly repetitive PE/PPE (proline-glutamate/proline-proline-glutamate) gene families, which comprise approximately 10% of the coding regions in *M. tuberculosis,* in the pathogenesis of the strain can be obtained through long-read sequencing (36). The long reads obtained from nanopore sequencing span a larger part of the genome than short-read sequencing but come with lower accuracy, affecting the error rate of calling a single nucleotide variant, as demonstrated in Chapter 3. The core-genome single nucleotide polymorphism (cgSNP) typing using short-reads resulted in more accurate phylogenies of *M. tuberculosis* isolates than the cgSNP typing using long-reads

(37). On the other hand, the hybrid approach, which combines both short- and long-read sequence data, could overcome the lower accuracy of long-read sequencing while still exhibiting higher genome resolution by spanning repetitive regions.

Overall, the choice of the sequencing platform (short-read vs long-read) and the sequencing chemistry (e.g., library preparation kits) should be based on the resolution of the information needed.

2. *Dry-lab aspect*

2.1. *Bioinformatics tools: "One tool does not fit all", the lack of standardization and the gap between results and interpretation*

While the application of genomics in clinical microbiology has been increasing, the translation of genetic information remains challenging. There are plenty of bioinformatics tools (Chapter 1; Table 2), often open source and usually not further developed/maintained after the initial versions. The constant changes in versions and/or the discontinuation of a bioinformatics tool complicates the standardization of data analysis. In routine settings, automation and standardization of the analysis are significant for the reliability of the diagnostic test results. While some laboratories rely on open-source tools requiring trained laboratory technicians and bioinformaticians/computational biologists, other laboratories use user-friendly specialized software and pipelines (e.g., commercial CLC Genomics Workbench; Ridom Seqsphere+; free Galaxy). These commercialized software/pipelines allow the standardized analysis of batches of isolates within-laboratory and across laboratories (38). However, most laboratories still lack expertise in interpreting these data. There should be more of a multidisciplinary team, including researchers, laboratory technicians, medical molecular microbiologists, and clinical medical microbiologists, to evaluate and put the result into clinical context. Metadata is equally important for getting the correct diagnosis. Considering shotgun metagenomics, which is (essentially) a hypothesis-free approach, one can get unexpected results from the data analysis. Therefore, one may need patient data to interpret and evaluate the results.

Recent examples in the literature highlight how different bioinformatics tools can affect the overall interpretation of the results (39-41). For example, different *de novo* assembly tools yielded varying performances in the contiguity and quality of genome assembly of bacterial species with diverse GC content (42). In Chapter 3, we compared several *de novo* assembly tools for the *M. tuberculosis* genome. In the absence of a polishing step in the assemblies of one of the most used assembly tools, SPAdes (43), variants introduced during assembly were not corrected (37). This is concerning because incorrect genome

assembly will undoubtedly lead to (some) loss of information. In that case, one may lose information about genes of interest, for example, AMR or virulence genes. If the assembly is not good enough, these genes might be broken into different contigs, and one may not be able to identify them when the threshold setting is too high (e.g., 95% identity, 80% coverage for the identification of the gene). Thus, the gene might still be there but is not identified as it is present on several contigs. Therefore, it is crucial to have an accurate assembly and optimize the assembly matrices. Choice of the assembler should be made considering the genome composition of the bacterial species.

Another example is the use of different bioinformatics approaches for the same purpose. In Chapter 5, we used three different bioinformatics approaches, OTU clustering, mapping and *de novo* assembly followed by BLASTN, to identify bacterial pathogens. OTU clustering is the most used approach in microbiome studies (44). However, this method is not able to distinguish between highly similar species. *De novo* assembly followed by BLAST using an in-house database was the optimal data analysis approach with the shortest turnaround time (TAT) and the highest sensitivity.

2.2. Standardization needs the use of public curated databases

The advancements in the genomics field have revolutionized the drug susceptibility prediction from sequencing data even though only known resistance genes and mutations present in the databases can be identified. One of the main challenges is that there is no consistent way of predicting drug susceptibility among several AMR detection tools and databases. As we have shown in the benchmarking study of TB drug resistance prediction tools in Chapter 3 (37), prediction differs depending on the algorithm of the software (the way of SNP calling and scoring) and the comprehensiveness of the associated databases. A recent study highlights the urgent need for standardization of AMR bioinformatics tools, which differ from the type of accepted input data (reads or assembly) to the AMR databases and detection method (either based on assembly or mapping) (45). The CRyPTIC project to combat TB is an excellent example of worldwide collaboration with a unified open-access database to identify genes associated with resistance markers for certain anti-TB drugs (46). The CRyPTIC project has aimed at a faster, targeted treatment provided via sufficiently accurate genetic resistance prediction, which has been recently tested in the large dataset of > 20,000 clinical *M. tuberculosis* isolates with associated minimum inhibitory concentration measures (47). Therefore, standardization of the bioinformatics tools and the databases is essential to aid global surveillance of pathogens and antimicrobial resistance, which can be achieved through (inter-) national joint efforts.

Overall, both standardization and automation of NGS workflows are not entirely in place. Therefore, a routine microbiology laboratory requires suitable technical facilities, trained personnel, and adequate informatics infrastructure (48). Laboratories developing their own (bioinformatics) workflow hinder the inter-laboratory standardization. In this regard, commercial solutions, for example, collaborating with NGS service providers performing sequencing and sending the raw sequencing data back or performing the analyses and/or interpretation, would be an option (13). Alternatively, open-source software, e.g., AMR package (49), can help standardize and improve the quality. Then again, implementing a commercial pipeline from sample preparation to reporting would be a choice for a standardized diagnostics workflow. Lastly, the collaboration between investigators and industry partners would expedite the development and standardization of the workflows.

Future Perspectives

An appropriate diagnostic algorithm is essential for appropriate infection management

The selection of the NGS workflow depends on what diagnostic information is required. Targeted metagenomics, for example, targeting specific regions of a particular kingdom, e.g., 16S rRNA, 16S-23S rRNA for bacteria, or targeting specific resistance genes, has the advantage of increased sensitivity, as the gene of interest is enriched and the background (host sequencing reads) is reduced. However, this approach limits the breadth of detectable pathogens and target genes. Yet, a targeted metagenomics approach designed for specific targets can be highly favorable in outbreak situations, and it can be used as a rapid screening method. Patients with a positive result can be isolated, and the labor-intensive culture-based work can focus on a smaller proportion of positive samples. Although the sensitivity of the NGS method might be lower than the culture-based method, like in our study in Chapter 4, there is a trade-off between the sensitivity and TAT of the technique. Here, NGS reduced the TAT from 4 days to 14 h compared to culture but had lower sensitivity. However, the positive predictive value was 100% for both methods (50). Moreover, optimizing the NGS protocol may improve its sensitivity.

Precision metagenomic approaches designed to enrich common pathogens for each sample type (e.g., blood, urine, sputum) and clinical presentation may be the way to allow a sensitive and accurate clinical diagnosis. Thus, the composition of syndromic panels/probes targeting specific pathogens and resistance genes should be defined based on the disease and specimen type. As shown with the Respiratory Pathogen ID/AMR panel used to diagnose respiratory infections successfully (51), syndromic NGS testing can increase

diagnostic specificity and sensitivity. It enables enrichment of the pathogen, common to a particular specimen and clinical question, present in the panel while maintaining the background to detect pathogens not targeted. Additionally, information obtained from SMg can go beyond identifying the pathogen and can also result in, e.g., detection of virulence factors, identification of antimicrobial resistance (AMR) markers, and typing. This may need deeper sequencing at the expense of longer TAT and higher cost (52). The SMg may be valuable in the case of immunocompromised patients with consistently negative results from routine testing. It can be potentially used to guide personalized treatments, as shown for acute leukaemia patients, by detecting bacterial, fungal, and viral pathogens and AMR genes directly in blood samples at different time points of their antimicrobial treatment (53). Moreover, the advancements in sequencing technologies, notably Oxford Nanopore Technologies (ONT), allow obtaining results within 6 hours following sampling, e.g., for identifying pathogens based on circulating cell-free DNA from blood (54). Consequently, it is important to define on a case-by-case basis when to use metagenomics as a diagnostics tool.

Clinical utility studies should move from one sample to large scale datasets

Despite the continuing integration of NGS-based methods into clinical diagnostics on a routine basis, few studies are accessing the clinical utility of the protocols using large-scale clinical studies (55). Such studies should be conducted to evaluate the potential of a method to be used as a diagnostic tool. For example, the implementation of metagenomics as a diagnostic tool should go beyond spike-in experiments (which sometimes do not reflect the complexity of a patient sample) and sequencing of a single clinical sample at a time. Else, we must continue to confirm the NGS results with other approaches, which is often restricted by insufficient sample material left to perform the retesting. Clinically unvalidated metagenomics approaches cannot be used for accurate diagnosis within an accredited clinical laboratory. Clinical evaluation studies are essential to determine the clinical value for the patient and the benefit for health care systems in general, including adequate cost-effectiveness analysis.

Conclusion and Outlook

Overall, this work demonstrated a wide range of applications of NGS technologies to diagnose infections and obtain several layers of information, covering pathogen detection and species identification, drug susceptibility prediction, and pathogen typing for outbreak investigations. The hurdles of each NGS approach for “wet-lab” and “dry-lab” steps were also extensively

addressed. The latest advances, especially ONT sequencing, hold great promises to transform the relatively slow and laborious conventional culture-based microbial diagnosis into same-day diagnosis. Nevertheless, the main challenge is the lack of standardization in protocols within/among laboratories. To address this, we need joint efforts to globally collaborate and unite to combat infections and improve life expectancy and quality of life. Standardization of NGS workflows and their integration into clinical diagnostics would move faster when both industry-driven and research-driven collaborations were present more frequently and productively. Projects like the CRyPTIC project could serve as a start for this.

References

1. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet. 2020 Oct 17;396(10258):1204–22.
2. Peker N, Couto N, Sinha B, Rossen JW. Diagnosis of bloodstream infections from positive blood cultures and directly from blood samples: recent developments in molecular approaches. Clin Microbiol Infect. 2018 Sep;24(9):944–55.
3. Graf EH, Pancholi P. Appropriate Use and Future Directions of Molecular Diagnostic Testing. Curr Infect Dis Rep. 2020 Feb 6;22(2):5.
4. Torres-Sangiao E, Leal Rodriguez C, García-Riestra C. Application and Perspectives of MALDI–TOF Mass Spectrometry in Clinical Microbiology Laboratories. Microorganisms 2021;9:1539
5. Dik J-WH, Poelman R, Friedrich AW, Panday PN, Lo-Ten-Foe JR, et al. An integrated stewardship model: antimicrobial, infection prevention and diagnostic (AID). Future Microbiology 2016;11:93–102.
6. Dik JH, Poelman R, Friedrich AW, Niesters HGM, Rossen JWA, et al. Integrated Stewardship Model Comprising Antimicrobial, Infection Prevention, and Diagnostic Stewardship (AID Stewardship). J Clin Microbiol 2017;55:3306–3307.
7. Messacar K, Parker SK, Todd JK, Dominguez SR. Implementation of Rapid Molecular Infectious Disease Diagnostics: the Role of Diagnostic and Antimicrobial Stewardship. J Clin Microbiol 2017;55:715–723.
8. Miller S, Chiu C, Rodino KG, Miller MB. Point-Counterpoint: Should We Be Performing Metagenomic Next-Generation Sequencing for Infectious Disease Diagnosis in the Clinical Laboratory? J Clin Microbiol. 2020 Feb 24;58(3):e01739-19.
9. Chiu CY, Miller SA. Clinical metagenomics. Nat Rev Genet. 2019 Jun;20(6):341–55.
10. Forbes BA, Hall GS, Miller MB, Novak SM, Rowlinson M-C, Salfinger M, et al. Practical Guidance for Clinical Microbiology Laboratories: Mycobacteria. Clin Microbiol Rev. 2018 Apr;31(2):e00038-17.
11. Votintseva AA, Bradley P, Pankhurst L, Del Ojo Elias C, Loose M, Nilgiriwala K, et al. Same-Day Diagnostic and Surveillance Data for Tuberculosis via Whole-Genome Sequencing of Direct Respiratory Samples. J Clin Microbiol. 2017 May;55(5):1285–98.
12. Deurenberg RH, Bathoorn E, Chlebowicz MA, Couto N, Ferdous M, García-Cobos S, et al. Application of next generation sequencing in clinical microbiology and infection prevention. J Biotechnol. 2017 Feb;243:16–24.
13. Schuele L, Cassidy H, Peker N, Rossen JWA, Couto N. Future potential of metagenomics in microbiology laboratories. Expert Review of Molecular Diagnostics 2021;21:1273–1285.
14. Gargis AS, Kalman L, Lubin IM. Assuring the Quality of Next-Generation Sequencing in Clinical Microbiology and Public Health Laboratories. J Clin Microbiol 2016;54:2857–2865.
15. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biology. 2014 Nov 12;12(1):87.
16. Peker N, Garcia-Croes S, Dijkhuizen B, Wiersma HH, van Zanten E, Wisselink G, et al. A Comparison of Three Different Bioinformatics Analyses of the 16S–23S rRNA Encoding Region for Bacterial Identification. Frontiers in Microbiology. 2019;10:620.
17. 10 Strain Even Mix Genomic Material | ATCC [Internet]. [cited 2021 Nov 10]. Available from: https://www.atcc.org/products/msa-1000

18. 10 Strain Even Mix Whole Cell Material | ATCC [Internet]. [cited 2021 Nov 10]. Available from: https://www.atcc.org/products/msa-2003

19. ZymoBIOMICS Microbial Community Standards [Internet]. ZYMO RESEARCH. [cited 2021 Nov 10]. Available from: https://www.zymoresearch.com/collections/zymobiomics-microbial-community-standards

20. Miller S, Naccache SN, Samayoa E, Messacar K, Arevalo S, Federman S, et al. Laboratory validation of a clinical metagenomic sequencing assay for pathogen detection in cerebrospinal fluid. Genome Res. 2019 May;29(5):831–42.

21. Bal A, Pichon M, Picard C, Casalegno JS, Valette M, Schuffenecker I, et al. Quality control implementation for universal characterization of DNA and RNA viruses in clinical respiratory samples using single metagenomic next-generation sequencing workflow. BMC Infect Dis. 2018 Dec;18(1):1–10.

22. Mukherjee S, Huntemann M, Ivanova N, Kyrpides NC, Pati A. Large-scale contamination of microbial isolate genomes by Illumina PhiX control. Standards in Genomic Sciences. 2015 Mar 30;10(1):18.

23. Langille MGI. Exploring Linkages between Taxonomic and Functional Profiles of the Human Microbiome. mSystems 2018;3:e00163-17.

24. Poulsen CS, Kaas RS, Aarestrup FM, Pamp SJ. Standard Sample Storage Conditions Have an Impact on Inferred Microbiome Composition and Antimicrobial Resistance Patterns. Microbiology Spectrum. 9(2):e01387-21.

25. Hickl O, Heintz-Buschart A, Trautwein-Schult A, Hercog R, Bork P, Wilmes P, et al. Sample Preservation and Storage Significantly Impact Taxonomic and Functional Profiles in Metaproteomics Studies of the Human Gut Microbiome. Microorganisms. 2019 Sep 19;7(9):367.

26. Marotz C, Cavagnero KJ, Song SJ, McDonald D, Wandro S, Humphrey G, et al. Evaluation of the Effect of Storage Methods on Fecal, Saliva, and Skin Microbiome Composition. mSystems. 6(2):e01329-20.

27. Schlaberg R, Chiu CY, Miller S, Procop GW, Weinstock G, Professional Practice Committee and Committee on Laboratory Practices of the American Society for Microbiology, et al. Validation of Metagenomic Next-Generation Sequencing Tests for Universal Pathogen Detection. Arch Pathol Lab Med. 2017 Jun;141(6):776–86.

28. Charalampous T, Kay GL, Richardson H, Aydin A, Baldan R, Jeanes C, et al. Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. Nat Biotechnol. 2019 Jul;37(7):783–92.

29. Street TL, Barker L, Sanderson ND, Kavanagh J, Hoosdally S, Cole K, et al. Optimizing DNA Extraction Methods for Nanopore Sequencing of *Neisseria gonorrhoeae* Directly from Urine Samples. Journal of Clinical Microbiology. 58(3):e01822-19.

30. Sloop GD, De Mast Q, Pop G, Weidman JJ, St. Cyr JA. The Role of Blood Viscosity in Infectious Diseases. Cureus. 12(2):e7090.

31. McCarthy A, Chiang E, Schmidt ML, Denef VJ. RNA Preservation Agents and Nucleic Acid Extraction Method Bias Perceived Bacterial Community Composition. PLOS ONE. 2015 Mar 23;10(3):e0121659.

32. Dalla-Costa LM, Morello LG, Conte D, Pereira LA, Palmeiro JK, Ambrosio A, et al. Comparison of DNA extraction methods used to detect bacterial and yeast DNA from spiked whole blood by real-time PCR. Journal of Microbiological Methods. 2017 Sep 1;140:61–6.

33. Johnson JS, Spakowicz DJ, Hong B-Y, Petersen LM, Demkowicz P, Chen L, et al. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. Nat Commun. 2019 Nov 6;10(1):5029.

34. Decuypere S, Meehan CJ, Puyvelde SV, Block TD, Maltha J, Palpouguini L, et al. Diagnosis of Bacterial Bloodstream Infections: A 16S Metagenomics Approach. PLOS Neglected Tropical Diseases. 2016 Feb 29;10(2):e0004470.

35. Matsuo Y, Komiya S, Yasumizu Y, Yasuoka Y, Mizushima K, Takagi T, et al. Full-length 16S rRNA gene amplicon analysis of human gut microbiota using MinION™ nanopore sequencing confers species-level resolution. BMC Microbiology. 2021 Jan 26;21(1):35.

36. Bainomugisa A, Duarte T, Lavu E, Pandey S, Coulter C, Marais BJ, et al. A complete high-quality MinION nanopore assembly of an extensively drug-resistant *Mycobacterium tuberculosis* Beijing lineage strain identifies novel variation in repetitive PE/PPE gene regions. Microbial Genomics,. 2018;4(7):e000188.

37. Peker N, Schuele L, Kok N, Terrazos M, Neuenschwander SM, de Beer J, et al. Evaluation of whole-genome sequence data analysis approaches for short- and long-read sequencing of *Mycobacterium tuberculosis*. Microb Genom. 2021 Nov;7(11).

38. Balloux F, Brønstad Brynildsrud O, van Dorp L, Shaw LP, Chen H, Harris KA, et al. From Theory to Practice: Translating Whole-Genome Sequencing (WGS) into the Clinic. Trends in Microbiology. 2018 Dec 1;26(12):1035–48.

39. Liu D, Zhou H, Xu T, Yang Q, Mo X, Shi D, et al. Multicenter assessment of shotgun metagenomics for pathogen detection. EBioMedicine. 2021 Dec 1;74:103649.

40. Coolen JPM, Jamin C, Savelkoul PHM, Rossen JWA, Wertheim HFL, Matamoros SP, et al. Centre-specific bacterial pathogen typing affects infection-control decision making. Microb Genom. 2021 Aug;7(8).

41. Junier T, Huber M, Schmutz S, Kufner V, Zagordi O, Neuenschwander S, et al. Viral Metagenomics in the Clinical Realm: Lessons Learned from a Swiss-Wide Ring Trial. Genes. 2019 Sep;10(9):655.

42. Goldstein S, Beka L, Graf J, Klassen JL. Evaluation of strategies for the assembly of diverse bacterial genomes using MinION long-read sequencing. BMC Genomics. 2019 Jan 9;20(1):23.

43. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes *De Novo* Assembler. Current Protocols in Bioinformatics. 2020;70(1):e102.

44. Nguyen N-P, Warnow T, Pop M, White B. A perspective on 16S rRNA operational taxonomic unit clustering using sequence similarity. npj Biofilms Microbiomes. 2016 Apr 20;2(1):1–8.

45. Hendriksen RS, Bortolaia V, Tate H, Tyson GH, Aarestrup FM, McDermott PF. Using Genomics to Track Global Antimicrobial Resistance. Frontiers in Public Health. 2019;7:242.

46. CRyPTIC Consortium and the 100,000 Genomes Project, Allix-Béguec C, Arandjelovic I, Bi L, Beckert P, Bonnet M, et al. Prediction of Susceptibility to First-Line Tuberculosis Drugs by DNA Sequencing. New England Journal of Medicine. 2018 Oct 11;379(15):1403–15.

47. Fowler PW, Wright C, Spiers H, Zhu T, Baeten EM, et al. BashTheBug: a crowd of volunteers reproducibly and accurately measure the minimum inhibitory concentrations of 13 antitubercular drugs from photographs of 96-well broth microdilution plates. 2021;2021.07.20.453060.

48. Leo S, Cherkaoui A, Renzi G, Schrenzel J. Mini Review: Clinical Routine Microbiology in the Era of Automation and Digital Health. Frontiers in Cellular and Infection Microbiology. 2020;10:706.

49. Berends MS, Luz CF, Friedrich AW, Sinha BNM, Albers CJ, et al. AMR - An R Package for Working with Antimicrobial Resistance Data. 2021;810622.

50. Peker N, Rossen JWA, Deurenberg RH, Langereis PC, Raangs EGC, Kluytmans JA, et al. Evaluation of an Accelerated Workflow for Surveillance of ESBL (CTX-M)-Producing *Escherichia coli* Using Amplicon-Based Next-Generation Sequencing and Automated Analysis. Microorganisms. 2018 Jan 11;6(1):6.

51. Illumina [Internet]. [cited 2021 December]. Available from: https://www.illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-respiratory-pathogen-panel-performance-app-note-470-2020-014/illumina-respiratory-pathogen-panel-performance-app-note-470-2020-014.pdf

52. Dulanto Chiang A, Dekker JP. From the Pipeline to the Bedside: Advances and Challenges in Clinical Metagenomics. The Journal of Infectious Diseases. 2020 Mar 28;221(Supplement_3):S331–40.

53. Gyarmati P, Kjellander C, Aust C, Song Y, Öhrmalm L, Giske CG. Metagenomic analysis of bloodstream infections in patients with acute leukemia and therapy-induced neutropenia. Scientific Reports. 2016 Mar 21;6(1):23532.

54. Grumaz C, Hoffmann A, Vainshtein Y, Kopp M, Grumaz S, Stevens P, et al. Rapid Next-Generation Sequencing–Based Diagnostics of Bacteremia in Septic Patients. The Journal of Molecular Diagnostics. 2020 Mar 1;22(3):405–18.

55. Miao Q, Ma Y, Wang Q, Pan J, Zhang Y, Jin W, et al. Microbiological Diagnostic Performance of Metagenomic Next-generation Sequencing When Applied to Clinical Practice. Clinical Infectious Diseases. 2018 Nov 13;67(suppl_2):S231–40.