METHODOLOGICAL GUIDANCE FOR CLINICAL METAGENOMICS AND ANTIMICROBIAL RESISTANCE RESEARCH

S. De Koster[1], J.P. Rodriguez–Ruiz[1], Y. Glupczynski[1], H. Goossens[1,2], B.B. Xavier[1]

[1]Laboratory of Medical Microbiology, Vaccine & Infectious Disease Institute, Universiteit Antwerpen, Antwerp, Belgium
[2]Laboratory of Clinical Microbiology, Antwerp University Hospital, Edegem, Belgium

Corresponding Author: Basil Britto Xavier, Ph.D; e-mail: basilbritto.xavier@uantwerpen.be

Abstract – In recent years, there has been an increasing interest in the role of the microbiome on human health and disease. Along with this interest, the need for methodological guidance on how to best study the microbiome and its relationship to antibiotic resistance is growing. High-throughput metagenomic sequencing is a powerful tool for investigating the microbiome, and it is becoming increasingly used in clinical research. This review summarizes the existing literature on clinical metagenomics and antimicrobial resistance research and provides guidelines on how to design and conduct studies using this technology. We also discuss some of the challenges and limitations of clinical metagenomics research and suggest future directions for this rapidly evolving field.

Keywords: AMR, Metagenomics, Microbiome, Clinical metagenomics, Kitome, Splashome.

INTRODUCTION: THE MICROBIOME AND ANTIMICROBIAL RESISTANCE

The microbiome is a complex community whose composition varies based on diet, age, geographical location, a specific region in the host, as well as exposure to antibiotics and infections[1-4]. More specifically, antibiotic treatment causes reduced species diversity and altered metabolic activity that may lead to the development of antibiotic-resistant strains. The pool of antibiotic resistance genes is referred to as the resistome[3]. Definitions of the terminology used in this manuscript are summarized in Table I.

Metagenomic sequencing allows for an in-depth characterisation of the microbiota and resistome directly from samples, for example, faecal, food, environmental samples and samples that are recalcitrant to culture[3,5]. Two main culture-independent sequencing methods are amplicon and shotgun metagenomic sequencing. Amplicon sequencing involves sequencing a single marker gene such as the small subunit of bacterial ribosomal RNA (16S rRNA) or the fungal internal transcribed spacers (ITS1/ITS2) used for taxonomic identification and determination of species diversity (metataxonomics). On the other hand, shotgun metagenomic sequencing allows for the identification of all the genes present in the sample without the selection of a specific gene[3,6]. In general, processing the samples remains challenging because of the high heterogeneity, variable

TABLE 1. DEFINITIONS OF TERMS USED IN THE MANUSCRIPT.

	Definition
Assembly	Raw sequencing reads are stitched into larger contiguous sequences known as 'contigs' and extended contigs called 'scaffolds'.
Binning	The clustering of contigs or reads based on feature patterns of the sequences into individual genome bins which represent microbial genomes by machine learning methods.
Contig	A set of overlapping DNA segments that provides a contiguous representation of a genomic region.
Chimeric assembly	Sequences from two or more (sub-)species that are incorrectly joined during the assembly.
Deep sequencing	Sequencing a genomic region multiple times (hundreds or thousands of times) to detect rare microbes, genes or mutations as little as 1% of the original sample.
Genome annotation	Demarcation of a gene or protein coding sequences, and other genetic features such as tRNA, and rRNA in a raw DNA sequence of genome.
Genome coverage or depth	The number of unique reads that are mapped to a given nucleotide in the reconstructed sequence. The number of times a nucleotide is read during sequencing.
Hybrid assembly	Raw sequencing reads from second generation (short read), and third generation (long read) technologies are used to make larger contiguous sequences like contigs, and scaffolds is called a hybrid assembly.
Insertion sequence	Insertion sequence is a short DNA sequence flanked by inverted repeats and act as a transposable element.
Kitome	Contaminating DNA in DNA extraction kits and other laboratory reagents.
Long-read sequencing	Also referred to as third generation sequencing. Sequencing of a single molecule and generating longer lengths (5000 bp->5 kb).
Metagenome	All the genetic material of microorganisms presents in a sample, consisting of the genomes of the microbial community in the sample.
Metagenome-assembled genomes	A single-taxon assembly based on one or more binned metagenomes that has been asserted to be a close representation to an actual individual genome.
Metataxonomics	The study and characterization of the entire microbiota (based on 16S rRNA gene sequencing) to create a metataxonomic tree, which shows the relationships between all sequences.
Metatranscriptome	The collection of all gene transcripts (RNAs) encoded in a community of microorganisms within a sample, which provides a snapshot of the gene expression in a sample at a given moment.
Microbiota	The community of microorganisms (such as bacteria, fungi and viruses) present in a defined environment (refers to the taxonomy of microorganisms present).
Microbiome	The community of microorganisms (such as bacteria, fungi and viruses) and their genes present in a defined environment (refers to the bacteria and their genes).
Mobile genetic element	Segments of DNA that encode enzymes and other proteins that mediate the movement of DNA within genomes or between bacterial cells.
Multi-omics	Multiple omics provides an integrated perspective on the genotype-phenotype-envirotype relationship by integrating diverse omics data (generated from genome, proteome, transcriptome, metabolome and epigenome).

CONTINUED

TABLE 1 (*CONTINUED*). DEFINITIONS OF TERMS USED IN THE MANUSCRIPT.	
	Definition
N50	Parameter to define the quality of the genome assembly by the size and a number of contigs or scaffolds produced by the assembler.
Resistome	The collection of all the antibiotic resistance genes (acquired and intrinsic resistance genes) and their precursors in pathogenic and nonpathogenic bacteria in a given microbial ecosystem.
Short-read sequencing	Also referred to as second generation sequencing. Sequencing of short fragments of DNA (50-500 bp) by synthesis or ligation using a DNA polymerase or ligase enzyme, respectively.
Shotgun metagenomic sequencing	The untargeted sequencing of all genomic DNA present in a sample.
Splashome	Internal or cross-contamination between samples during sample processing.
Targeted metagenomics	Metagenomic approach to target a specific region of a genome (e.g., 16S rRNA or resistance genes) in multiple microorganisms and samples.
U50	Metric for measuring the performance of the assembly based on unique, non-overlapping, target-specific contigs by using a reference genome as baseline. This parameter corrects for the high background noise (i.e., host and other non-targets), which contribute to having a skewed, misrepresented N50 value.

composition of the bacterial community, resistome, and PCR inhibitory substances, for example, bilirubin, complex polysaccharides, and certain lipid types in faecal samples[7]. In particular, the bacterial loads depend on the sampling site; for example, stool samples contain higher bacterial loads than samples from the oropharynx. Thus, the results of microbiome research depend on several methodological factors, such as the sampling method, type of swabs, volume of sample, time of sample collection, sample preservation, whether or not a culturing step is applied, DNA extraction method, the use of positive and negative controls, the sequencing method (e.g., 16s rRNA gene amplicon sequencing, shotgun metagenomic, metatranscriptomic sequencing), contaminations and bioinformatic tools. In this review, we aim to summarise recent methodologies and findings regarding metagenomic sequencing and resistome research from sample collection to bioinformatics analysis (Figure 1).

Sample Collection and Pre-Processing

When collecting samples, consistency in the handling of the sampling is essential to minimise technical variations and avoid misinterpretation[8]. Here, we list the various sampling methods for the different microbiota.

GUT MICROBIOTA

Most intestinal microbiota studies collect faecal samples as a proxy for the distal colon microbiome[2,9]. It is recommended to specify a defined time for collecting stool to minimise the shifts in microbial composition during the day[9]. Following sampling, homogenisation to minimise intrasample variations and preservation to minimise possible confounding variations should be consistent across all study samples[10]. Bacteria in the sample should be inactivated as soon as possible to prevent the overgrowth of certain bacterial species or DNA degradation that can lead to taxonomical biases[7]. Microbial integrity is best preserved when stool samples are frozen at -80°C or with liquid nitrogen

as quick as possible upon collection using a proper transportation container or Cary-Blair transport medium, which is considered the standard method for storing stool samples[9,11]. Stool samples must be fresh, however, this is usually not possible when remote participants are sampled, thus, self-made preservation buffers or commercial preservation kits can be utilised. Self-made preservation buffers include buffers with preservatives such as ethanol[12,13], DMSO-EDTA salt solution[14], EDTA, citrate trisodium salt dihydrate and ammonium sulfate[15]. The commercial kits include DNA/RNA Shield solution (Zymo Research, Irvine, CA, USA), RNAlater (Thermo Fisher, Waltham, MA, USA), OMNIgene-Gut (DNA Genotek, Kanata, ON, Canada), PrimeStore MTM (Longhorn Vaccines and Diagnostics, San Antonia, TX, USA) or Norgen collection kits (Norgen Biotek, Thorold, ON, Canada) that can be held/shipped at room temperature[9]. Of note, PrimeStore MTM (Longhorn Vaccines and Diagnostics, San Antonia, TX, USA), the OMNIgene GUT kit (DNA Genotek, Kanata, ON, Canada) and the DNA/RNA Shield collection (Zymo Research, Irvine, CA, USA) solution were all efficient in conserving faecal samples[7,10,15,16].
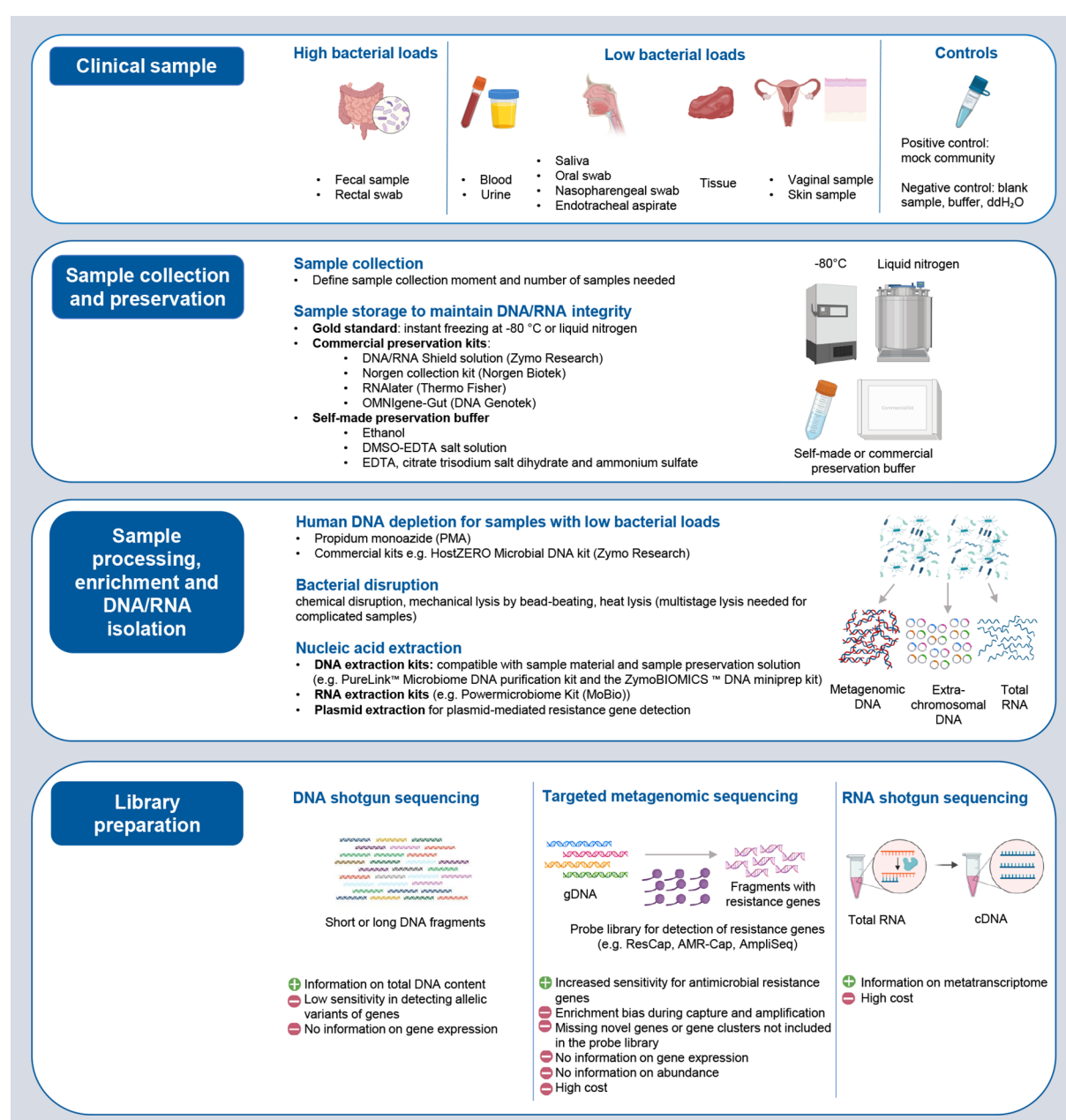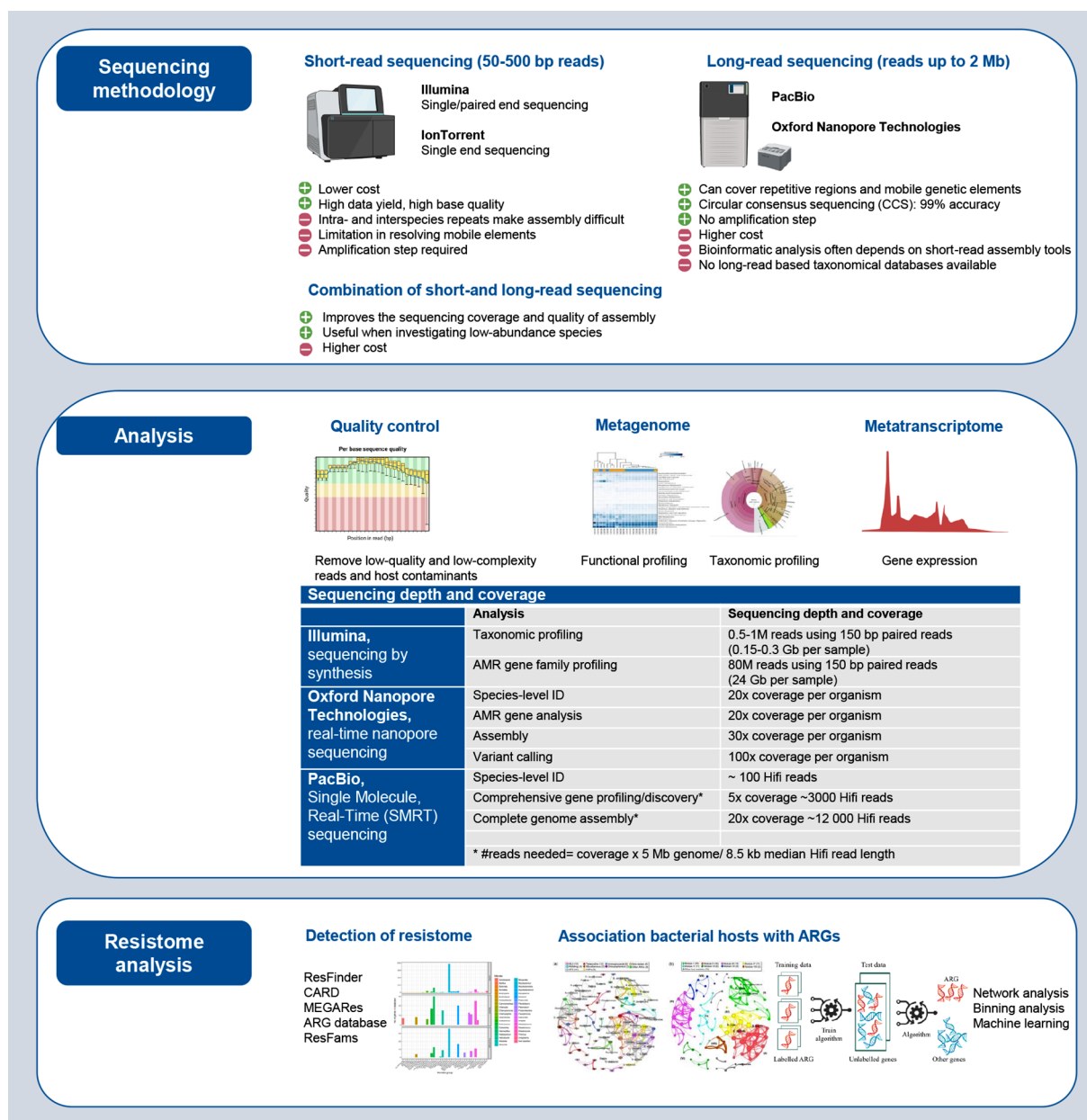

Figure 1.   *CONTINUED*

Figure 1. Overview of sample collection, sample processing and metagenomic and metatranscriptomic sequencing to investigate antimicrobial resistance in the microbiome. Figure of functional profiling from[56]. Figure of taxonomic profiling from[57]. Figure of resistome databases from[46]. Figure of network analysis from[58]. Figure of artificial intelligence adapted from[59].

The abundance of critical resistant pathogens, such as extended-spectrum beta-lactamase (ESBL) or carbapenemase-producing *Enterobacteriaceae*, is often below the detection threshold of direct sequencing. This is a major limitation when characterising the gut resistome[17,18]. Selective culture-enrichment of stool samples can be used to identify low abundance pathogens within the microbiome but hampers the quantification of the resistome because of differential growth rates of bacteria[17]. Merging culture-dependent and- independent techniques could provide more profound resolution and help better understand microbial communities, including low abundant species[6]. However, there is currently a lack of validated enrichment methods, and, additionally, the combined use of such techniques would significantly increase the complexity and costs.

OTHER MICROBIOTA

In contrast to the rich gut microbiota, studying other human body samples (blood, urine, oral, tissue and skin samples) provides obstacles such as low microbial biomass. The samples are dominated by host DNA (90% of human-genome specific reads), challenging the metagenomics analysis[6,19].

The eNAT swab (Copan Diagnostics Inc., Murrieta, CA, USA), containing a lysis buffer to stabilise microbial DNA and RNA, was shown to keep the microbial composition stable for vaginal, skin and saliva samples[20]. In addition, the eSwabs (Copan Diagnostics Inc., Murrieta, CA, USA) yielded high consistency in the population captured from the skin[21]. Samples from skin, tissue, blood, the oral cavity and respiratory tract carrying low bacterial loads might need enrichment. Microbial enrichment can be carried out using human DNA depletion by differential lysis of mammalian cells (osmotic or chemical) and treatment with propidium monoazide (PMA) or using commercial kits prior to DNA extraction of the microbial cells. The HostZERO Microbial DNA kit (Zymo Research, Irvine, CA, USA) appeared most effective for human DNA depletion in vaginal and saliva samples[20]. In contrast, the performance of PMAxx™ (Biotium Inc., CA, USA) for free DNA inactivation in different biological samples (saliva, faeces, urine, vaginal swab) varied extensively and should be further evaluated[22]. Furthermore, culture enrichment on selective and/or non-selective enrichment media reduces the effect of host DNA contamination by allowing the proliferation of microorganisms and improving the detection of promiscuous and fastidious organisms in bacterial communities[6]; however, this will affect the composition of species and genes in the sample.

DNA Extraction for Metagenomics

The DNA extraction method plays a crucial role in detecting complex bacterial communities and is the process where bias is most likely to occur. The DNA extraction methods must be compatible with the sample material and preservation solution[7].

The use of a suspended or pelleted sample depends on the instructions of the DNA extraction kit. Different bacterial disruption methods (chemical disruption, mechanical lysis by bead-beating, heat lysis, sonication) exist. Mechanical lysis by bead-beating combined with other disruption methods (multistage lysis) ensures the recovery of microbial DNA from the most complicated samples, such as Gram-positive bacteria[7].

Kazantseva et al[7] (2021) assessed the DNA extraction of two commercial DNA extraction kits: PureLink™ Microbiome DNA purification kit (Thermo Fisher, Waltham, MA, USA) and the ZymoBIOMICSTM DNA Miniprep Kit (Zymo Research, Irvine, CA, USA). The ZymoBIO-MICS kit exhibited better quality characteristics of isolated DNAs (higher OD 260/230 ratio) and was compatible with the DNA/RNA Shield™ solution (Zymo Research, Irvine, CA, USA)[7]. For resistome assessment, the addition of plasmid extraction to the standard DNA extraction could increase the detection of plasmid-mediated resistance genes[17,23]; however, it can introduce bias since the efficiencies of extraction are related to the plasmid size, i.e., smaller plasmids are extracted more efficiently than large plasmids[17].

Library Preparation for Metagenomic Sequencing

The performance of different library preparation approaches depends on the sample type, microbial diversity and amount of input DNA[24]. Each sequencing platform (Illumina (San Diego, CA, USA), Ion Torrent (Thermo Fisher, Waltham, MA, USA), PacBio (Menlo Park, CA, USA), Oxford Nanopore (Oxford, UK) has specific metagenomics technical guidance available, which describes the sequencing systems, the library preparation methods and downstream analysis pipelines[25-27]. The Illumina Nextera XT and DNA Flex kits (Illumina, San Diego, CA, USA) are commonly used in metagenomic library preparations. They cover various genomes, from small DNA viruses, microbial genomes and amplicons to complex genomes in eukaryotic and human systems[24].

Sequencing Methodologies for the Detection of Antimicrobial Resistance Genes

While 16S rRNA amplicon sequencing generates information on the microbial diversity in the sample, shotgun metagenomic sequencing generates information on the total DNA content of the sample, making it suitable for the study of antimicrobial resistance (AMR) and for identifying the species specific marker gene for taxonomically complex organisms, for instance *rplF* gene for *Neisseria* species. Determining all AMR genes in the sample, referred to as the resistome, will help understand the complex interactions between organisms, genes and their environment. However, for effective AMR surveillance, the focus should be on clinically relevant yet low abundant AMR genes (such as ESBLs and carbapenemases) and horizontal gene transfer events, which are generally rare[28]. In this case, direct metagenomic shotgun sequencing might suffer from low sensitivity in detecting minority populations harbouring resistance genes and/or low specificity in identifying allelic variants[18,29]. All resistance sequences account for less than 1% of the total sequenced DNA, indicating that the proportion of these genes is relatively low, and even deep sequencing may not be able to capture these elements in the total gene pool present in the samples[4,28,30].

Targeted metagenomics using a capture library specific for resistance genes and genes involved in DNA mobilisation can increase sensitivity and improve the identification of resistance genes within a complex metagenome background[4,18]. Target capture is a method to quantify low abundant AMR genes and/or detect allelic variants that alter the susceptibility phenotypes. Target capture enriches resistance genes directly from standard metagenomic DNA extractions and increases the proportion of sequenced reads mapping to resistance genes. A targeted sequence capture platform for resistome analysis, such as ResCap or AMR-cap, significantly improves the detection of resistance genes and plasmids compared to direct metagenomic shotgun sequencing[4,18,30-32]. However, an enrichment bias can occur due to differential capture affinity and amplification rates. In addition to low efficiencies in the probe-hybridisation step, signal overloads, or inhibition of the enzyme-based steps (e.g. amplification during library preparation) may affect the outcome[31]. Noyes et al[28] (2017) were able to increase the detection of resistome sequences, including antimicrobial resistance genes of public health importance, such as ESBLs, using target enrichment while molecular indices were employed to count DNA molecules and correct for enrichment bias[28].

In contrast, Stege et al[4] (2022) observed 32 resistance genes that went undetected by ResCap but were detected using metagenomic shotgun sequencing. A number of genes were not included in the probe library; however, 14 genes included in the ResCap library were also not detected using ResCap. This emphasizes the challenge of continuously updating the probe libraries to include all known resistance genes and shows that genes might be missed using ResCap[4]. Resistance genes that are not present in the reference database when the probe libraries are designed might not be captured and might be missed. However, AMR-cap was able to capture novel antibiotic resistance genes from the CARD database that were unknown when the capture probes were designed[32].

SEQUENCING TECHNOLOGY

Short-read sequencing platforms [represented by Illumina (San Diego, CA, USA) and Ion Torrent (Thermo Fisher, Waltham, MA, USA)] and long-read sequencing platforms [represented by PacBio (Menlo Park, CA, USA) and Oxford Nanopore (Oxford, UK)] are available for metagenome sequencing. Short-read sequencing made assembly driven metagenomics possible, allowing the assembly of composite genomes called metagenome-assembled genomes (MAGs). However, short-read sequencing platforms produce read lengths of 50-300 bp, making high-quality metagenomic assembly more difficult due to intra-and inter-species repeats that confuse short-read assembly algorithms[33]. This results in low recovery of high microdiversity microbes, low recovery of flexible genomes and uncertainty due to potential chimaera generation[34]. MAGs of abundant taxa in the dataset have been prone to chimeric assemblies and have limitations in resolving

mobile elements, information which is of increasing relevance in medical research on antibiotic resistance, toxin encoding genes and bacterial pathogenicity islands encoding virulence factors[35].

In contrast to short-read sequencing, PacBio long-read sequencing can cover repetitive regions and recover complete MAGs (circularised and without gaps) directly from assembled human gut metagenomes[36]. PacBio sequencing can be performed using highly accurate long reads (10 kb HiFi reads or circular consensus sequencing, CCS) and continuous long reads (CLR). PacBio long-read sequencing can provide complete gene profiling; however, CLR has a higher error rate (10-15%). Moreover, dependence on bioinformatics designed for short-read assembly makes gene discovery and assembly for low-abundance organisms difficult. When using long-read sequencing, high coverage is needed to obtain reliable sequence with low error rates, which can be obtained by using CCS, DNA subreads that are sequenced multiple times by the same DNA polymerase before generating a consensus sequence. This increases the accuracy to comparable levels to Sanger and Illumina sequencing (more than 99%)[5,34,37]. Also, the Oxford Nanopore Technologies is a long-read sequencing technique (reads up to 2 Mb) that can be used to obtain in-depth annotations for taxonomical and functional profiles of the metagenome with 97% raw read accuracies[27,38,39].

To improve sequencing coverage, investigating low-abundance species can be done by combining third-generation long-read sequencing with ultra-deep second-generation sequencing[36]. Hybrid assembly, using short and long reads, improves the quality of the assembly (increasing average contig length, contig N50, U50, size and number of large contigs), generates a higher percentage of complete genes and increases the completeness of the genome reconstruction compared to using only short-read or long-read sequencing[5,36]. The quality of assemblies and microbiome profiles can be influenced by sequencing depth[9], the complexity of the community, the sequencing technology and/or the proportions of host DNA contamination[6].

SEQUENCING DEPTH, COMPLEXITY OF THE COMMUNITY AND HOST DNA CONTAMINATION

Identifying new microbial taxa improves with more significant sequencing depths and with lower proportions of host DNA. This is particularly important for detecting very low abundant species[19]. Increasing sequencing depth also increases the number of identified AMR determinants. Similarly, rare antibiotic resistance genes were detected more frequently at higher sequencing depths and were often absent in datasets generated with lower sequencing depths[40]. However, a balance between cost and required sequencing depth needs to be established, and therefore, it is crucial to select the appropriate sequencing method, read length and sequencing analysis tools.

A critical factor for a robust analysis of shotgun sequencing data is the number of sequenced reads. In a study by Durazzi et al[41] (2021), shotgun metagenomes with a low number of reads (<500,000 reads) were characterised as low-quality samples, which show high skewness of the relative species abundance distribution[41]. For large-scale microbiome studies, shallow whole-metagenome shotgun sequencing with a sequencing depth of 1 Gbp can provide more accurate data at the genus and species level compared to 16S rRNA gene amplicon sequencing[42]. However, it will usually not provide data on low-abundance microbial genomes. The sequencing depth of 5-10 Gbp per sample is not enough to capture all the microbes with low abundance. Small sequencing data sizes of 5 Gbp per sample provide limited genome coverage (5 Mbp) for a species of 0.1% of relative abundance in a human gut microbiome. Ultra-deep metagenomic sequencing showed that the size of assembled genomes and N50 continued to increase with the sequencing depth until 40 Gbp[36]. To detect ARGs, increasing sequencing depths were associated with an increasing number of AMR determinants identified. A sequencing depth associated with ≥50 million reads was sufficient for detecting the resistome in bovine faecal samples[40]. Sequencing depth issues might be solved with the introduction of novel techniques such as reduced metagenome sequencing[43], droplet microfluidics[35] or novel functional metagenomics[44].

TOOLS FOR ANALYSIS OF METAGENOMIC DATASETS

A comprehensive review of computational tools for metagenomic sequencing data analysis was recently published by Yang et al[33] (2021). An analysis overview for gut microbiome studies is provided by Gao et al[45] (2021). We refer to these reviews to describe the tools for sequencing quality control, metagenomic assembly, binning, gene prediction, gene annotation, taxonomic classification, and MAG abundance profiling[33,45]. In addition, we list novel bioinformatics tools for the analysis of 16S rRNA, metagenomics and metatranscriptomics data in Table II.

Detection of Resistance Genes in Metagenomic Datasets

Metagenomics enables the analysis of antimicrobial resistance in bacterial communities. Regularly updated and comprehensive resistome databases are crucial for the quality of information obtained. Several reviews describe and compare antimicrobial resistance gene databases. Recently, de Abreu et al[29] (2021) published a minireview to describe the analysis of antibiotic resistance using metagenomic approaches. The review describes the functional annotation of the metagenome to study the resistome and the reference databases available. In brief, the most used ARG databases that allow metagenomic data input are ResFinder, Comprehensive Antibiotic Resistance Database, MEGARes, ARG-database and Resfams[29]. Papp and Solymosi[46] compared well-known antibiotic resistance gene databases based on structure and content. The CARD database has been shown to be the most comprehensive database for acquired antibiotic resistance genes and mutations[46]. Peng et al[47] (2021) published a minireview on the bioinformatic tools for resistome analysis of environmental samples. They showed that applying two or more bioinformatic tools and databases could provide a comprehensive knowledge of ARG profiles in diverse environmental samples[47]. However, the differences in antibiotic classification of the databases show the need for expert knowledge to interpret results. Results should be interpreted with care as not all resistance genes are sufficient to cause resistance alone and might not confer resistance in all bacteria. For example, Margolis et al[48] (2021) reported that the abundance of the *tetX* genes was strongly associated with the relative abundance of Bacteroidetes. However, an increase in *tetX* in this context does not suggest an increase in resistance to tetracycline antibiotics. *TetX,* a flavin-dependent monooxygenase inactivates and confers resistance to tetracyclines only in aerobic bacteria and not in the anaerobic Bacteroidetes[48].

Associating the potential bacterial hosts with ARGs is challenging. However, the potential bacterial hosts and ARGs are closely associated within the microbiome. Several analysis methods like network analysis as well as machine learning techniques have been applied to identify any connections between these two entities that could be relevant for future studies on infection transmission network[49]. A novel sequencing approach combining shotgun metagenomic with chromosome conformation capture (Hi-C) can link bacteria to ARGs, plasmids and phages using Hi-C-based networks. Using this technique, increased details on bacterial gene content and mobile genetic elements can be obtained[50].

Identification of Bias and Contamination: the Use of Controls and in Silico Contaminant Removal

Bias can arise at every step of the microbiome workflow, particularly in the DNA extraction and PCR amplification steps[9,51]. The use of positive controls will help to identify bias introduced by PCR amplification during library preparation or by various sequencing technologies. A positive mock control should be a valid representation for the investigated environment[8]. The DNA extraction of the positive control should be validated with the DNA extraction kit to ensure the extraction of the correct proportions of DNA from the positive control. Secondly, a positive control for sequencing (pre-extracted DNA mix) should be used to ensure that no errors were introduced during sequencing[8].

TABLE 2. NEW BIOINFORMATIC TOOLS TO ANALYZE METAGENOMICS DATA.			
Name	**Functionality**	**Available at**	**Ref.**
Agamemnon	Metagenomics and metatranscriptomics quantification analysis suite providing abundances at genus, species and strain resolution and exploratory data visualisation.	https://github.com/ ivlachos/agamemnon	60
Animalcules	Interactive microbiome analytic toolkit in R for 16S rRNA sequencing data, shotgun DNA metagenomics data and RNA-based metatranscriptomics profiling data. The toolkit combined analytics, vizualization methods and machine learning models.	https://github.com/ compbiomed/animalcules	61
Binnacle	Cluster scaffolds into comprehensive bins and integrate existing binning methods to improve the contiguity and quality of metagenomic bins.	https://github.com/ marbl/binnacle	62
Biobakery 3	Integrate methods for taxonomic, strain-level, functional and phylogenetic profiling of metagenomes to help multi-omics profiling for microbial community studies. It includes sequence-level quality control and contaminant depletion guidelines (KneadData), MatPhlAn 3 for taxonomic profiling, HUMAnN 3 for functional profiling, StrainPhlAn 3 and PanPhlAn 3 for nucleotide-and gene-variant-based strain profiling, and PhyloPhlAn 3 for phylogenetic placement and putative taxonomic assignment of new assemblies.	https://github.com/ biobakery/ biobakery/wiki	63
EukDetect	Identify microbial eukaryotes in shotgun metagenomic sequencing data.	https:// github.com/ allind/EukDetect	64
HOME-BIO	Pipeline for metagenomic shotgun data analysis including a quality control step, assembly of sequences in contigs and taxonomic profiling.	https://github.com/ carlferr/HOME-BIO	65
KAUST Metagenomic Analysis Platform	Comprehensive exploration of shotgun metagenomic data. Annotation of contigs or genes (including ARGs) and sample-to-sample or gene catalog-based comparison.	https: //www.cbrc.kaust.edu.sa/ aamg/kmap.start	66
LueVari	Reference-free SNP caller based on read-colored de Bruijn graphs for identification of SNPS in AMR genes and chromosomal DNA from shotgun metagenomics data.	https://github.com/ baharpan/cosmo/ tree/LueVari.	67
MaxBin 2.0	Recovering individual genomes from metagenomes in a de novo manner.	https: //sourceforge.net/ projects/maxbin/	68
Meta-Apo	Reliable, high-resolution view of microbiome function from 16S amplicon sequencing.	https://github.com/ qibebt-bioinfo/ meta-apo	69
METAnnotatorX2	Integrated analysis of deep and shallow metagenomic data. Extracting taxonomic and function information from metagenomic short sequence reads, assembly of short, long and hybrid read-based metagenomic data sets and species-specific genome reconstruction with gene prediction and associated functional annotation.	http: //probiogenomics.unipr.it/ cmu/	70

Name	Functionality	Available at	Ref.
	TABLE 2 (_CONTINUED_). NEW BIOINFORMATIC TOOLS TO ANALYZE METAGENOMICS DATA.		
metaXplor	Interactive viral and microbial web-interfaced data manager for managing, sharing and exploring metagenomic data.	https://github.com/SouthGreenPlatf orm/metaXplor	71
MetGEMs toolbox	Predicting functional composition of metagenomic samples, annotating putative enzyme functions and metabolic routes r elated in human disease using 16S rRNA sequences.	https://github.com/yumyai/MetGEMs	72
MiDSystem	Online system for de novo assembly and analysis of metagenomic data. Results from embedded tools are visualised in an online summary report.	https://github.com/NTU-CGM/miDSystem/	73
mixtureS	Identify strains, their numbers and abundance from shotgun reads.	http: //www.cs.ucf.edu/ ~xiaoman/mixtureS/	74
nf-core/mag	Pipeline for metagenome hybrid assembly and binning.	https://github.com/nf-core/mag	75
OGRE:Overlap Graph-based metagenomic Read clustEring	Reduce the size of metagenomic datasets by clustering reads into species-specific groups based on their overlaps and facilitate assembly and downstream analyses.	https://github.com/Marleen1/OGRE	76

Commercial positive controls with defined synthetic communities are available[42]. Recently, mock communities for the human gut microbiome were developed and are available from the NITE Biological Resource Center at the National Institute of Technology and Evaluation (NITE, Tokyo, Japan)[52]. However, artificial microbiomes with defined taxonomic information and similar complexity to a real microbial sample are still scarce[42]. If the available controls are found unsuitable for the study, custom positive controls comparable to the investigated microbiome should be designed[8].

Negative controls (reagent-only or blank sampling) should be applied during the sampling and processing of the samples to identify major contaminations from different possible sources: external contamination of sampling equipment, DNA kit contaminations or 'kitome'[7,53-55], internal or cross-contamination between samples during sample processing or 'splashome'[55] and index hopping[53]. _In silico_ decontamination can complement the laboratory approaches to distinguish contaminant microbial DNA from accurate microbial sequences. The open-source R package _decontam_ (https://github.com/benjjneb/decontam) can be used to identify and remove contaminant sequences in a marker gene and metagenomic data[53]. Additionally, the use of bacteriophage PhiX174 DNA as a quality and calibration control in Illumina sequencing may lead to contamination of the sequenced genomes. Therefore, filtering of bacteriophage sequences by mapping demultiplexed reads against the PhiX174 genome is needed prior to the analysis[40].

CONCLUSIONS AND FUTURE PERSPECTIVES

In conclusion, studies have shown that the composition of gut bacteria is important for human health. However, recent advances in sequencing technology and metagenomics are still limited because of the lack of standardised protocols to analyse samples, making it difficult to understand the impact of the microbe on individuals or populations. A multidisciplinary approach, such as the multi-omics strategy, will allow scientists to get a more accurate picture of how the microbiome and its host interact. This multi-omics strategy can also be used to

discover new therapeutic targets for treating microbiome-related illnesses. While the use of multi-omics data is still in its early stages, it holds great promise for improving our understanding of the microbiome and its role in human health. Additionally, this approach can help to identify potential biomarkers for disease. Furthermore, there is a need to develop better-standardised methods or algorithms for identifying and characterising microbial genes and proteins. Finally, more extensive and longitudinal studies are required to understand how the microbiota changes over time. Microbiome research is a rapidly growing field with great potential despite these limitations. Ultimately, this research field will provide insights into the role of microbiota in human health and pave the way for new treatments for various diseases.

Conflict of Interest

None to declare. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgements

Informed Consent

Not applicable.

Authors' Contribution

BX, YG, HG: Conceptualization and supervision. SDK: writing- original draft preparation. SDK, JR, BX: data collection and writing. BX, YG, HG, JR, SDK: writing- review and editing. All authors have read and agreed to the version of the manuscript.

ORCID ID

Sien De Koster: https://orcid.org/0000-0002-5267-5080; Juan Pablo Rodriguez-Ruiz: https://orcid.org/0000-0002-8847-7297; Basil Britto Xavier: https://orcid.org/0000-0002-5897-0240.

Funding

REFERENCES

1) Manor O, Dai CL, Kornilov SA, Smith B, Price ND, Lovejoy JC, Gibbons SM, Magis AT. Health and disease markers correlate with gut microbiome composition across thousands of people. Nat Commun 2020; 11: 1-12.
2) Tang Q, Jin G, Wang G, Liu T, Wang B, Cao H. Current Sampling Methods for Gut Microbiota: A Call for More Precise Devices. Front Cell Infect Microbiol 2020; 10: 151.
3) Ramirez J, Guarner F, Bustos Fernandez L, Maruy A, Sdepanian VL, Cohen H. Antibiotics as Major Disruptors of Gut Microbiota. Front Cell Infect Microbiol 2020; 10: 572912.
4) Stege PB, Hordijk J, Shetty SA, Visser M, Viveen MC, Rogers MRC, Gijsbers E, Dierikx CM, van der Plaats RQJ, van Duijkeren E, Franz E, Willems RJL, Fuentes S, Paganelli FL. Impact of long-term dietary habits on the human gut resistome in the Dutch population. Sci Rep 2022; 12: 1-13.
5) Xie H, Yang C, Sun Y, Igarashi Y, Jin T, Luo F. PacBio Long Reads Improve Metagenomic Assemblies, Gene Catalogs, and Genome Binning. Front Genet 2020; 11: 516269
6) Whelan FJ, Waddell B, Syed SA, Shekarriz S, Rabin HR, Parkins MD, Surette MG. Culture-enriched metagenomic sequencing enables in-depth profiling of the cystic fibrosis lung microbiota. Nat Microbiol 2020; 5: 379-390.
7) Kazantseva J, Malv E, Kaleda A, Kallastu A, Meikas A. Optimisation of sample storage and DNA extraction for human gut microbiota studies. BMC Microbiol 2021; 21: 1-13.
8) Hornung BVH, Zwittink RD, Kuijper EJ. Issues and current standards of controls in microbiome research. FEMS Microbiol Ecol 2019; 95: 1-7.

9) Jones J, Reinke SN, Ali A, Palmer DJ, Christophersen CT. Fecal sample collection methods and time of day impact microbiome composition and short chain fatty acid concentrations. Sci Rep 2021; 11: 1-13.

10) Plauzolles A, Toumi E, Bonnet M, Pénaranda G, Bidaut G, Chiche L, Allerdet-Servent J, Retornaz F, Goutorbe B, Halfon P. Human Stool Preservation Impacts Taxonomic Profiles in 16S Metagenomics Studies. Front Cell Infect Microbiol 2022; 12: 722886.

11) Nagata N, Tohya M, Takeuchi F, Suda W, Nishijima S, Ohsugi M, Ueki K, Tsujimoto T, Nakamura T, Kawai T, Miyoshi-Akiyama T, Uemura N, Hattori M. Effects of storage temperature, storage time, and Cary-Blair transport medium on the stability of the gut microbiota. Drug Discov Ther 2019; 13: 256-260.

12) Hale VL, Tan CL, Knight R, Amato KR. Effect of preservation method on spider monkey (Ateles geoffroyi) fecal microbiota over 8 weeks. J Microbiol Methods 2015; 113: 16-26.

13) Marotz C, Cavagnero KJ, Song SJ, McDonal D, Wandro S, Humphrey G, Bryant M, Ackermann G, Diaz E, Knight R. Evaluation of the Effect of Storage Methods on Fecal, Saliva, and Skin Microbiome Composition. mSystems 2021; 6: e01329-20.

14) Kawada Y, Naito Y, Andoh A, Ozeki M, Inoue R. Effect of storage and DNA extraction method on 16S rRNA profiled fecal microbiota in Japanese adults. J Clin Biochem Nutr 2019; 64: 106-111.

15) Wu C, Chen T, Xu W, Zhang T, Pei Y, Yang Y, Zhang F, Guo H, Wang Q, Wang L, Zhao B. The maintenance of microbial community in human fecal samples by a cost effective preservation buffer. Sci Rep 2021; 11: 1-10.

16) Tap J, Cools-Portier S, Pavan S, Druesne A, Ohman L, Törnblom H, Simren M, Derrien M. Effects of the long-term storage of human fecal microbiota samples collected in RNAlater. Sci Rep 2019; 9: 601.

17) Peto L, Fawcett NJ, Crook DW, Peto TEA, Llewelyn MJ, Walker AS. Selective culture enrichment and sequencing of feces to enhance detection of antimicrobial resistance genes in third-generation cephalosporin resistant Enterobacteriaceae. PLoS One 2019; 14: e0222831.

18) Lanza VF, Baquero F, Martínez JL, Ramos-Ruiz R, Gonzalez-Zorn B, Andremont A, Sanchez-Valenzuela A, Ehrlich SD, Kennedy S, Ruppe E, van Schaik W, Willems RJ, de la Cruz F, Coque TM. In-depth resistome analysis by targeted metagenomics. Microbiome 2018; 6: 1-14.

19) Pereira-Marques J, Hout A, Ferreira RM, Weber M, Pinto-Ribeiro I, van Doorn LJ, Knetsch CW, Figueiredo C. Impact of host DNA and sequencing depth on the taxonomic resolution of whole metagenome sequencing for microbiome analysis. Front Microbiol 2019; 10: 1277.

20) Ahannach S, Delanghe L, Spacova I, Witthouck S, Van Beeck W, De Boeck I, Lebeer S. Microbial enrichment and storage for metagenomics of vaginal, skin, and saliva samples. iScience 2021; 24: 103306.

21) Bjerre RD, Hugerth LW, Boulund F, Seifert M, Johansen JD, Engstrand L. Effects of sampling strategy and DNA extraction on human skin microbiome investigations. Sci Rep 2019; 9: 1-11.

22) Mancabelli L, Milani C, Anzalone R, Alessandri G, Lugli GA, Tarracchini C, Fontana F, Turroni F, Ventura M. Free DNA and Metagenomics Analyses: Evaluation of Free DNA Inactivation Protocols for Shotgun Metagenomics Analysis of Human Biological Matrices. Front Microbiol 2021; 12: 749373.

23) Zhang T, Zhang XX, Ye L. Plasmid metagenome reveals high levels of antibiotic resistance genes and mobile genetic elements in activated sludge. PLoS One 2011; 6: e26041.

24) Gaulke CA, Schmeltzer ER, Dasenko M, Tyler BM, Vega Thurber R, Sharpton TJ. Evaluation of the Effects of Library Preparation Procedure and Sample Characteristics on the Accuracy of Metagenomic Profiles. mSystems 2021; 6: e00440-21.

25) Illumina. Shotgun Metagenomics Methods Guide. Illumina. Published online 2020 [accessed on 14 april 2022].

26) Pacific biosciences. Technical overview: metagenomics shotgun library preparation using SMRTbell express template prep kit 2.0. Published online February 2021: https://www.pacb.com/wp-content/uploads/Metagenomics-Shotgun-Library-Preparation-Using-SMRTbell-Express-TPK-2.0-Customer-Training.pdf [accessed on 14 april 2022].

27) Oxford Nanopore Technologies. Metagenomic sequencing with Oxford Nanopore. Published online 2020: https://nanoporetech.com/sites/default/files/s3/literature/metagenomic-sequencing-guide.pdf [accessed on 14 april 2022].

28) Noyes NR, Weinroth ME, Parker JK, Dean CJ, Lakin SM, Raymond RA, Rovira P, Doster E, Abdo Z, Martin JN, Jones KL, Ruiz J, Boucher CA, Belk KE, Morley PS. Enrichment allows identification of diverse, rare elements in metagenomic resistome-virulome sequencing. Microbiome 2017; 5: 142.

29) de Abreu VAC, Perdigão J, Almeida S. Metagenomic Approaches to Analyze Antimicrobial Resistance: An Overview. Front Genet 2021; 11: 575592.

30) Guitor AK, Raphenya AR, Klunk J, Kuch M, Alcock B, Surette MG, McArthur AG, Poinar HN, Wright GD. Capturing the resistome: A targeted capture method to reveal antibiotic resistance determinants in metagenomes. Antimicrob Agents Chemother 2019; 64: e01324-19.

31) Macedo G, van Veelen HPJ, Hernandez-Leal L, van der Maas P, Heederik D, Mevius D, Bossers A, Schmitt H. Targeted metagenomics reveals inferior resilience of farm soil resistome compared to soil microbiome after manure application. Sci Total Environ 2021; 770: 145399.

32) Beaudry MS, Thomas JC, Baptista RP, Sullivan AH, Norfolk W, Devault A, Enk J, Kieran TJ, Rhodes Jr OE, Perry-Dow KA, Rose LJ, Bayona-Vasquez NJ, Oladeinde A, Lipp EK, Sanchez S, Glenn TC. Escaping the fate of Sisyphus: assessing resistome hybridization baits for antimicrobial resistance gene capture. Environ Microbiol 2021; 23: 7523-7537.

33) Yang C, Chowdhury D, Zhang Z, Cheung WK, Lu Aiping, Bian Z, Zhang L. A review of computational tools for generating metagenome-assembled genomes from metagenomic sequencing data. Comput Struct Biotechnol J 2021; 19: 6301-6314.

34) Haro-Moreno JM, López-Pérez M, Rodriguez-Valera F. Enhanced Recovery of Microbial Genes and Genomes From a Marine Water Column Using Long-Read Metagenomics. Front Microbiol 2021; 12: 708782.

35) Pryszlak A, Wenzel T, Seitz KW, Hildebrand F, Kartal E, Cosenza MR, Benes V, Bork P, Merten CA. Enrichment of gut microbiome strains for cultivation-free genome sequencing using droplet microfluidics. Cell Reports Methods 2022; 2: 100137.

36) Jin H, You L, Zhao F, Li S, Ma T, Kwok LY, Xi H, Sun Z. Hybrid, ultra-deep metagenomic sequencing enables genomic and functional characterization of low-abundance species in the human gut microbiome. Gut Microbes 2022; 14: 2021790.

37) Yu X, Jiang W, Huang X, Lin J, Ye H, Liu B. rRNA Analysis Based on Long-Read High-Throughput Sequencing Reveals a More Accurate Diagnostic for the Bacterial Infection of Ascites. Biomed Res Int 2021; 2021: 6287280.

38) Ciuffreda L, Rodríguez-Pérez H, Flores C. Nanopore sequencing and its application to the study of microbial communities. Comput Struct Biotechnol J 2021; 19: 1497-1511.

39) Ong CT, Ross EM, Boe-Hansen GB, Turni C, Hayes BJ, Tabor AE. Technical note: overcoming host contamination in bovine vaginal metagenomic samples with nanopore adaptive sequencing. J Anim Sci 2022; 100: 1-9.

40) Zaheer R, Noyes N, Ortega Polo R, Cook SR, Marinier E, Van Domselaar G, Belk KE, Morley PS, McAllister TA. Impact of sequencing depth on the characterization of the microbiome and resistome. Sci Rep 2018; 8: 1-11.

41) Durazzi F, Sala C, Castellani G, Manfreda G, Remondini D, De Cesare A. Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota. Sci Rep 2021; 11: 1-10.

42) Xu W, Chen T, Pei Y, Guo H, Li Z, Yang Y, Zhang F, Yu J, Li X, Yang Y, Zhao B, Wu C. Characterization of Shallow Whole-Metagenome Shotgun Sequencing as a High-Accuracy and Low-Cost Method by Complicated Mock Microbiomes. Front Microbiol 2021; 12: 678319.

43) Snipen L, Angell IL, Rognes T, Rudi K. Reduced metagenome sequencing for strain-resolution taxonomic profiles. Microbiome 2021; 9: 1-19.

44) Crofts TS, McFarland AG, Hartmann EM. Mosaic Ends Tagmentation (METa) Assembly for Highly Efficient Construction of Functional Metagenomic Libraries. mSystems 2021; 6: e00524-21.

45) Gao B, Chi L, Zhu Y, Shi X, Tu P, Li B, Yin J, Gao N, Shen W, Schnabl B. An introduction to next generation sequencing bioinformatic analysis in gut microbiome studies. Biomolecules 2021; 11: 530.

46) Papp M, Solymosi N. Review and Comparison of Antimicrobial Resistance Gene Databases. Antibiotics 2022; 11: 339.

47) Peng Z, Mao Y, Zhang N, Zhang L, Wang Z, Han M. Utilizing Metagenomic Data and Bioinformatic Tools for Elucidating Antibiotic Resistance Genes in Environment. Front Environ Sci 2021; 9: 757365.

48) Margolis EB, Hakim H, Dallas RH, Allison KJ, Ferrolino J, Sun Y, Pui CH, Yao J, Chang TC, Hayden RT, Jeha S, Tuomanen EI, Tang L, Rosch JW, Wolf J. Antibiotic prophylaxis and the gastrointestinal resistome in paediatric patients with acute lymphoblastic leukaemia: a cohort study with metagenomic sequencing analysis. The Lancet Microbe 2021; 2: e159-e167.

49) Sun Y, Clarke B, Clarke J, Li X. Predicting antibiotic resistance gene abundance in activated sludge using shotgun metagenomics and machine learning. Water Res 2021; 202: 117384.

50) Ivanova V, Chernevskaya E, Vasiluev P, Ivanov A, Tolstoganov I, Shafranskaya D, Ulyantsev V, Korobeynikov A, Razin SV, Beloborodova N, Ulianov SV, Tyakht A. Hi-C Metagenomics in the ICU: Exploring Clinically Relevant Features of Gut Microbiome in Chronically Critically Ill Patients. Front Microbiol 2022; 12: 770323.

51) McLaren MR, Willis AD, Callahan BJ. Consistent and correctable bias in metagenomic sequencing experiments. Elife 2019; 8: e46923.

52) Tourlousse DM, Narita K, Miura T, Ohashi A, Matsuda M, Ohyama Y, Shimamura M, Furukawa M, Kasahara K, Kameyama K, Saito S, Goto M, Shimizu R, Mishima R, Nakayama J, Hosomi K, Kunisawa J, Terauchi J, Sekiguchi Y, Kawasaki H. Characterization and Demonstration of Mock Communities as Control Reagents for Accurate Human Microbiome Community. Microbiol Spectr 2022; 10: e1915-21.

53) Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome 2018; 6: 1-14.

54) Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, Walker AW. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol 2014; 12: 1-12.

55) Olomu IN, Pena-Cortes LC, Long RA, Vyas A, Krichevskiy O, Luellwitz R, Singh P, Mulks MH. Elimination of "kitome" and "splashome" contamination results in lack of detection of a unique placental microbiome. BMC Microbiol 2020; 20: 1-19.

56) Bicalho MLS, Machado VS, Higgins CH, Lima FS, Bicalho RC. Genetic and functional analysis of the bovine uterine microbiota. Part I: Metritis versus healthy cows. J Dairy Sci 2017; 100: 3850-3862.

57) Osuna-Mascaró C, Doña J, Johnson KP, Esteban R, De Rojas M. Complete mitochondrial genomes and bacterial metagenomic data from two species of parasitic avian nasal-mites (Rhinonyssidae: Mesostigmata). Front Ecol Evol 2020; 8: 142.

58) Ju F, Li B, Ma L, Wang Y, Huang D, Zhang T. Antibiotic resistance genes and human bacterial pathogens: Co-occurrence, removal, and enrichment in municipal sewage sludge digesters. Water Res 2016; 91: 1-10.

59) Lau HJ, Lim CH, Foo SC, Tan HS. The role of artificial intelligence in the battle against antimicrobial-resistant bacteria. Curr Genet 2021; 67: 421-429.

60) Skoufos G, Almodaresi F, Zakeri M, Paulson JN, Patro R, Hatzigeorgiou AG, Vlachos IS. AGAMEMNON: an Accurate metaGenomics And MEtatranscriptoMics quaNtificatiON analysis suite. Genome Biol 2022; 23: 1-27.

61) Zhao Y, Federico A, Faits T, Manimaran S, Segre D, Monti S, Johnson WE. animalcules: interactive microbiome analytics and visualization in R. Microbiome 2021; 9: 1-16.

62) Muralidharan HS, Shah N, Meisel JS, Pop M. Binnacle: Using Scaffolds to Improve the Contiguity and Quality of Metagenomic Bins. Front Microbiol 2021; 12: 638561.

63) Beghini F, McIver LJ, Blanco-Míguez A, Dubois L, Asnicar F, Maharjan S, Mailyan A, Manghi P, Scholz M, Thomas AM, Valles-Colomer M, Weingart G, Zhang Y, Zolfo M, Huttenhower C, Franzosa EA, Segata N. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with biobakery 3. Elife 2021; 10: e65088.

64) Lind AL, Pollard KS. Accurate and sensitive detection of microbial eukaryotes from whole metagenome shotgun sequencing. Microbiome 2021; 9: 1-18.

65) Ferravante C, Memoli D, Palumbo D, Ciaramella P, Di Loria A, D'Agostino Y, Nassa G, Rizzo F, Tarallo R, Weisz A, Giurato G. HOME-BIO (sHOtgun MEtagenomic analysis of BIOlogical entities): a specific and comprehensive pipeline for metagenomic shotgun sequencing data analysis. BMC Bioinformatics 2021; 22: 1-10.

66) Alam I, Kamau AA, Ngugi DK, Gojobori T, Duarte CM, Bajic VB. KAUST Metagenomic Analysis Platform (KMAP), enabling access to massive analytics of re-annotated metagenomic data. Sci Rep 2021; 11: 1-14.

67) Alipanahi B, Muggli MD, Jundi M, Noyes NR, Boucher C. Metagenome SNP calling via read-colored de Bruijn graphs. Bioinformatics 2020; 36: 5275-5281.

68) Wu YW, Singer SW. Recovering Individual Genomes from Metagenomes Using MaxBin 2.0. Curr Protoc 2021; 1: 1-14.

69) Jing G, Zhang Y, Cui W, Liu L, Xu J, Su X. Meta-Apo improves accuracy of 16S-amplicon-based prediction of microbiome function. BMC Genomics 2021; 22: 1-11.

70) Milani C, Lugli GA, Fontana F, Lugli GA, Fontana F, Mancabelli L, Alessandri G, Longhi G, Anzalone R, Viappiani A, Turroni F, van Sinderen D, Ventura M. METAnnotatorX2: a Comprehensive Tool for Deep and Shallow Metagenomic Data Set Analyses. mSystems 2021; 6: e00583-21.

71) Sempéré G, Pétel A, Abbé M, Lefeuvre P, Roumagnac P, Mahe F, Baurens G, Filloux D. MetaXplor: An interactive viral and microbial metagenomic data manager. Gigascience 2021; 10: 1-8.

72) Patumcharoenpol P, Nakphaichit M, Panagiotou G, Senavonge A, Suratannon N, Vongsangnak W. MetGEMs Toolbox: Metagenome-scale models as integrative toolbox for uncovering metabolic functions and routes of human gut microbiome. PLoS Comput Biol 2021; 17: e1008487.

73) Lee CY, Lee YF, Lai LC, Tsai MH, Lu TP, Chuang EY. MiDSystem: A comprehensive online system for de novo assembly and analysis of microbial genomes. N Biotechnol 2021; 65: 42-52.

74) Li X, Hu H, Li X. MixtureS: A novel tool for bacterial strain genome reconstruction from reads. Bioinformatics 2021; 37: 575-577.

75) Krakau S, Straub D, Gourlé H, Gabernet G, Nahnsen S. Nf-Core/Mag: a Best-Practice Pipeline for Metagenome Hybrid Assembly and Binning. NAR Genomics Bioinforma 2022; 4: 1-6.

76) Balvert M, Luo X, Hauptfeld E, Schönhuth A, Dutilh BE. OGRE: Overlap Graph-based metagenomic Read clustEring. Bioinformatics 2021; 37: 905-912.