International Journal of
Molecular Sciences

MDPI

Review

Metagenomic Sequencing for Microbial DNA in Human Samples: Emerging Technological Advances

Yu Shi, Guoping Wang, Harry Cheuk-Hay Lau and Jun Yu *

Institute of Digestive Disease, Department of Medicine and Therapeutics, State Key Laboratory of Digestive Disease, Li Ka Shing Institute of Health Sciences, CUHK Shenzhen Research Institute,
The Chinese University of Hong Kong, Hong Kong, China; shiyu2019@link.cuhk.edu.hk (Y.S.);
wangguoping@link.cuhk.edu.hk (G.W.); harrylau@link.cuhk.edu.hk (H.C.-H.L.)
* Correspondence: junyu@cuhk.edu.hk; Tel.: +852-37636099; Fax:+852-21445330

Abstract: Whole genome metagenomic sequencing is a powerful platform enabling the simultaneous identification of all genes from entirely different kingdoms of organisms in a complex sample. This technology has revolutionised multiple areas from microbiome research to clinical diagnoses. However, one of the major challenges of a metagenomic study is the overwhelming non-microbial DNA present in most of the host-derived specimens, which can inundate the microbial signals and reduce the sensitivity of microorganism detection. Various host DNA depletion methods to facilitate metagenomic sequencing have been developed and have received considerable attention in this context. In this review, we present an overview of current host DNA depletion approaches along with explanations of their underlying principles, advantages and disadvantages. We also discuss their applications in laboratory microbiome research and clinical diagnoses and, finally, we envisage the direction of the further perfection of metagenomic sequencing in samples with overabundant host DNA.

Keywords: high throughput sequencing; human microbiome; host DNA depletion; clinical metagenomics

Citation: Shi, Y.; Wang, G.; Lau, H.C.-H.; Yu, J. Metagenomic Sequencing for Microbial DNA in Human Samples: Emerging Technological Advances. *Int. J. Mol. Sci.* **2022**, *23*, 2181. https://doi.org/10.3390/ijms23042181

Academic Editor: Alessandro Desideri

Received: 14 January 2022
Accepted: 11 February 2022
Published: 16 February 2022

1. Introduction

High throughput sequencing (HTS) is widely used in microbiology research and its applications have been rapidly moved from basic research to clinical practice. Any specimen yielding a sufficient amount of nucleic acid can be subjected to an HTS analysis. HTS can either be targeted—that is, enriching certain genes or genomic regions—or untargeted as in metagenomic sequencing [1] (Figure 1). Targeted HTS approaches include universal amplification by a polymerase chain reaction (PCR), multiplex PCR amplification of specific whole genomes [2] and bait capture enrichment [3]. Targeted approaches increase the number and proportion of reads of interest in the sequence data although they limit the breadth of microorganisms that can be identified. For untargeted metagenomic sequencing, the total DNA of all organisms in a sample is sequenced without using any target-specific primers; thus, it covers all genetic information within the sample and allows the detection of all microorganism kingdoms including bacteria, fungi, viruses and parasites [4]. Moreover, metagenomic sequencing can provide a higher taxonomical resolution [5] and enable gene functional analyses such as virulence factors, antibiotic resistance and metabolic networks compared with targeted HTS. These capacities facilitate its great utilisation in microbiome studies.

However, metagenomic sequencing is still challenging to apply to samples with a high host nucleic acid background. Fortunately, a series of methods for depleting host DNA has been developed. This review explores numerous strategies to deplete non-microbial DNA in various specimens. We also discuss potential solutions to constraints and obstacles when using these methodologies in research and clinical diagnoses.

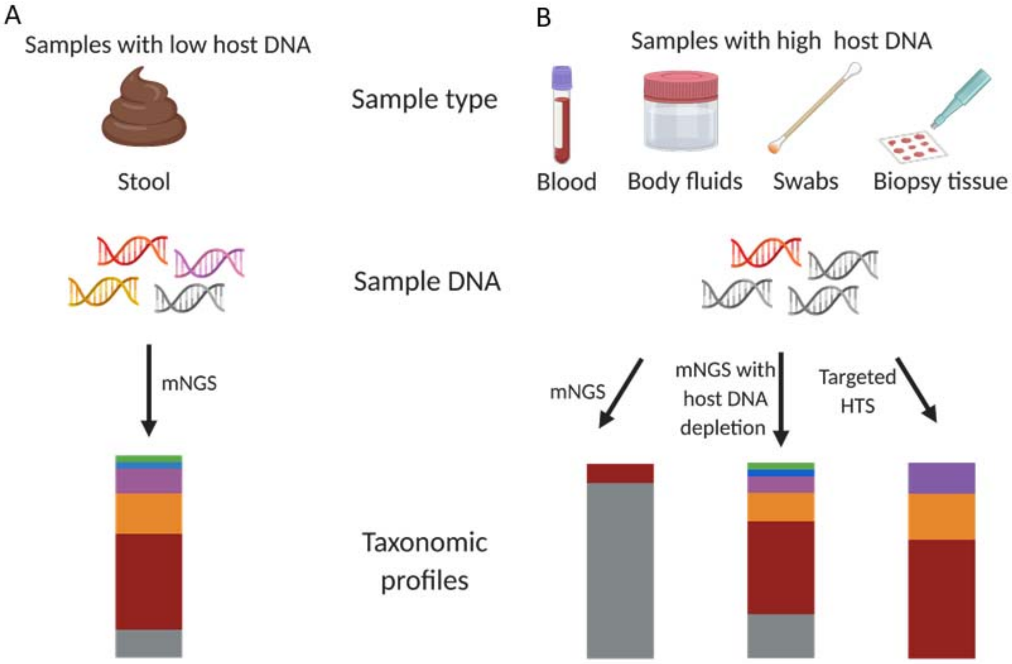

Figure 1. Schematic illustration of untargeted metagenomic sequencing and targeted sequencing in human samples. A variety of human-derived samples can be analysed using untargeted metagenomic sequencing or targeted sequencing. (**A**) For samples with a low amount of host DNA such as faecal samples, a taxonomic profile with a great resolution can be obtained when directly performing untargeted metagenomic sequencing. Gray represents host DNA; red, yellow and purple represent various bacteria; and blue and green represent viruses and archaea, respectively. (**B**) For samples with overabundant human DNA including nasal/oral/skin swabs, body fluids, blood and biopsy tissues, the vast majority of sequencing reads are aligned to the human genome, which can obscure signals from microorganisms when using metagenomic sequencing. As a solution, removal of host DNA before sequencing can improve the resolution of microbial DNA. These samples can also be analysed with targeted sequencing, which can increase the number and proportion of reads of interest in the sequence data although it limits the breadth of microorganisms that can be identified. mNGS, metagenomic next-generation sequencing.

2. The Growing Need for Microbial DNA Enrichment Prior to Metagenomic Sequencing

As microbiome studies have progressed, the scope of investigations has gradually shifted from faeces to microbial communities in various regions of the body and from high biomass samples to samples with a low biomass. A human genome is about a thousand times larger than a microbial genome (for example, human genome 3.2 Gb; the bacteria genome is 3.6 Mb on average [6]); therefore, the presence of just a few human cells could completely inundate the DNA components of microorganisms. For example, when the human cell count is over 200 cells per cubic millilitre in the cerebrospinal fluid, this high host DNA background can overwhelm the pathogen signal and decrease the sensitivity of the metagenomic NGS testing [7]. Given that untargeted metagenomic sequencing comprehensively assays all genetic materials regardless of whether they originate from host cells or microorganisms, host DNA has been shown to totally dominate the number of sequencing reads in human skin, vaginal, nasal and oral metagenomes [8]. Whilst the downstream computational filtering of human genome-mapped reads is a common solution, these sequence reads from the human genome can consume unnecessary sequencing space, reduce the overall sensitivity of the assay, obscure microbial differences and even ignore trace pathogen signals. Theoretically, enhancing the sequencing depth (the number of reads generated per specimen [9]) is a potential solution to increase the microorganism-related reads in metagenomic HTS but the huge sequencing costs and analysis time associated with ultra-deep sequencing are far beyond the capability of many laboratories [10] thus limiting the feasibility of the application across a large number of samples.

Collectively, many host-derived samples are deemed to be unsuitable for direct metagenomic sequencing due to their low data yield and high resource requirements. Metagenomic studies of several host-derived samples could benefit from removing the host DNA. This avoids a waste of resources and ensures that only the DNA of interest is sequenced, thereby reducing the amount of sequencing needed to obtain an adequate coverage and depth. More importantly, the removal of host DNA increases the amount and coverage of the microbial reads in metagenomic sequencing, thereby greatly facilitating the subsequent assembly of the reads and the analysis.

3. Current Approaches to Host DNA Depletion

Certain differences between the host cell (specifically referring to human beings throughout this review) and the microorganism, including the cellular structure and genomic variances, have been exploited as strategies for microbial DNA enrichment in samples with mixed nucleic acid populations.

3.1. Removal of the Host Cells before DNA Extraction

By taking advantage of the disparities in size and density between the microorganisms and eukaryotic host cells, they were filtered with a 5 µm filter, differential centrifugation and a flow cytometry assay, and then were tested to remove the buccal epithelial cells in human saliva [11]. However, no significant difference was observed between any of the host and non-host partitions. In contrast, size-based filtering followed by a deoxyribonuclease (DNase) treatment increased the amount of microbial DNA in sputum samples by 14–33% [12]. These results indicate the critical impact of extracellular DNA originating from host cells when physically separating the host cells and microorganisms. A combination of physical separation with DNase digestion can achieve a greater efficiency in enrichment. On the other hand, enriching virus-like particles by physical methods is a routine pre-treatment of viral metagenomic sequencing. Filtration is frequently used as the first step to purify virus-like particles because viruses (20–500 nm) are much smaller than host cells; hence, viruses can readily be separated from the host cells by physical methods. Typically, filters with pore sizes of 0.2 µm and 0.45 µm are used [13]. Size-based filtering followed by a DNase treatment is also capable of virus enrichment [14]. Due to the compact nature of virions, it is also possible to enrich the amount of virus-derived nucleic acids in samples using density gradient centrifugation (e.g., sucrose, caesium chloride, polyethylene glycol).

A cell wall can be found in the majority of bacteria and fungi and is substantially harder than the plasma membrane of mammalian cells. For viruses, although they have no cell wall, a protein capsid of a naked virus is sufficient to allow viruses to have a resistance against many disinfectants. In this context, a mild lysis buffer could be used to selectively lyse the plasma membrane of the host cells without damaging the microorganisms. The released DNA from the lysed host cells is then degraded by DNase, leaving intact microorganisms for a downstream extraction (Figure 2).

A variety of reagents including sterile water, saponin (a non-ionic detergent belonging to the group of glycosides forms), Triton X-100, Tween 20 and zwitterionic detergent have been assessed for their ability to selectively lyse human cells [15–17]. Among them, saponin has been the most widely used lysis reagent of mammalian cells owing to its high lysis efficiency on host cells with minimised effects on bacteria, fungi and even a non-enveloped "naked" virus [15]. An optimised two-step lysis protocol, in which treating the sample with saponin followed by an osmotic shock using double-distilled water, can further boost its efficiency [18–20].

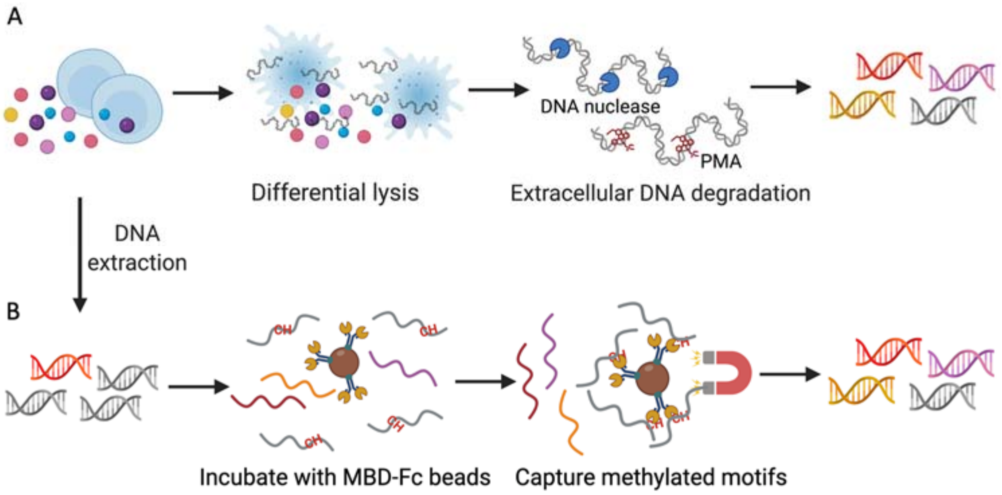

Figure 2. Workflow of typical host DNA depletion approaches. Before metagenomic sequencing, human DNA can be removed by different approaches. (**A**) The mainstream pre-extraction method to remove host DNA is first treating human cells with a selective lysis buffer followed by DNA nuclease or PMA treatment. (**B**) The most commonly used post-extraction methods take advantage of the disparity of the cytosine methylation frequency between eukaryotic and prokaryotic DNA. MBD-Fc-bound magnetic beads can capture methylated human DNA sequences, leaving the unmethylated motifs for downstream library preparation. MBD, methyl-CpG binding domain; HTS, high throughput sequencing.

Once the plasma membrane is destroyed by the lysis reagent, the DNA in the host cells is released. DNase I is commonly employed to degrade extracellular DNA [12,21], yet its activity depends closely on the choice of appropriate reaction buffers and could be limited by the presence of selective lysis reagents. Therefore, Benzonase nuclease has become an optimal approach for host DNA depletion due to its wide range of operating conditions and exceptionally high specificity, which allows the cleavage of exposed nucleic acid into 3–5 bases in length [22–26]. For samples already lysed by saponin and osmotic shock, using heat-labile salt active nuclease (HL-SAN; it has optimal activity at a high salt concentration) can achieve a 1000-fold reduction in the amount of human DNA [20,27]. Propidium monoazide (PMA) was found to have multiple advantages over conventional enzymatic degradation including a lower cost as well as being less time-consuming and having fewer sample processing steps. PMA is a cell membrane-impermeable DNA intercalator that can covalently modify DNA in the dark so that the DNA cannot be further amplified; the remaining PMA can be inactivated by light exposure [28]. These unique features make PMA an alternative to a DNA nuclease without the necessity of including any washing steps that could result in a loss of DNA.

Several commercially available protocols (Table 1) have been developed with applications of a series of lysis reagents and nucleases [11,29–35]. Generally, all these kits (or their modified protocols) exhibit a good efficiency in host DNA depletion and can increase the proportion of microbial DNA in a variety of sample types including blood [36], skin [31,37], bronchoalveolar lavage fluid [22] and tissues [32]. However, for samples with trace amounts of a microbial content, degrading the host nucleic acid using these approaches can lead to insufficient amounts of DNA for the sequencing library preparation. To ensure the amount of microbial DNA, whole genome amplification (WGA) by multiple displacement amplification can be used to amplify the raw DNA from nanograms to micrograms, providing enough DNA for the library preparation and sequencing. In a study with samples of sonicated fluids from a resected arthroplasty component, WGA was adopted to obtain a sufficient quantity of DNA for the library preparation after human DNA depletion by a MolYsis^{TM} Basic kit (see Table 1) [30,38]. However, WGA can introduce an amplification bias [39] and any microbial or cross-contamination among samples could also be magnified [40]. To solve these issues, WGA performed in sub-nanolitre droplets could significantly reduce the amplification bias and contamination [41,42] as this technique reduces competition for the primers and polymerases among the DNA fragments via the partitioning of the template DNA into tiny individual spaces. For example, a study by Shi et al. [12] utilised a microfluidic chip to separate microorganisms in samples and amplify their DNA in droplets to efficiently recover the microbial genome with a markedly reduced amplification bias.

Table 1. Commercial kits for microorganism enrichment.

Kit	Principle	Pros	Cons	Hands-On Time Per Sample	Cost Per Sample (USD)	Ref.
QIAamp DNA Microbiome (Qiagen, Hilden, Germany)	Lysis of host cell by saponin, degrade extracellular DNA with Benzonase nuclease	Ultra-clean columns to minimise contamination risk	Requires fresh sample	160 min	13	[43]
MolYsis™ Complete/Ultra-Deep Microbiome Prep (Molzym, Bremen, Germany)	Chaotropic lysis of host cell, degrade extracellular DNA with MolDNase	Applicable for body fluids, tissue and swab samples. Enrichment of bacterial and fungal DNA	Fresh sample is recommended	120 min	11	[44]
HostZERO Microbial DNA Kit (Zymo Research, Irvine, CA, USA)	Lysis of host cell, degrade extracellular DNA with microbial selection enzyme	Protocols for both tissue and liquid samples are provided	Requires intact (living) bacteria cells	30 min	10	[45]
NEBNext Microbiome DNA Enrichment (New England BioLabs, Ipswich, MA, USA)	Capture methylated host DNA	Can retain cell-free DNA from dead organisms to avoid DNA loss	Requires high molecular weight intact DNA. Bias to high CpG-methylated microbes	30 min *	39 *	[46]
LOOXSTER Enrichment Kit (Analytik Jena GmbH, Jena, Germany)	Capture non-methylated CpG dinucleotides	Can retain cell-free DNA from dead organisms to avoid DNA loss	Requires high molecular weight intact DNA. Bias to high CpG-methylated microbes	75 min *	34 *	[47]

*: DNA extraction step is excluded.

3.2. Separating the Microbial DNA from the Host Background

The recent development of highly multiplexed sequence capture approaches has enabled the enrichment of DNA from hundreds of known viruses and bacteria by selectively amplifying their DNA sequences as capture tools. These probes can be immobilised on solid carriers or biotinylated for sequence-specific hybridisation with the target DNA. Of note, an intrinsic limitation is that a depletion by a microorganism-specific hybridisation capture is not a viable option when the DNA sequence of interest is unknown in advance because the microbial DNA sequence has to be determined to enable the design of specific probes. Alternatively, probes can be designed to capture the host genome for the purpose of studying entire microbial communities harbouring a mixture of microorganisms with known or unknown sequences [48]. However, this hybridisation-based clean-up approach is inefficient due to the large size and complexity of the human genome.

Both prokaryotes and eukaryotes undergo methylation as their major epigenetic event but microbial and human epigenetics have distinct characteristics and functions. For higher eukaryotes, 5-methylcytosine is the dominant form of methylated DNA [49]. In humans, methylated cytosine occurs predominantly in a context of CpG dinucleotide with an estimated frequency of 60% to 90% [50]. The NEBNext Microbiome DNA Enrichment Kit (Table 1) employs a methyl-CpG binding protein domain to pull down vertebrate DNA based on this methylation pattern in eukaryotes. In comparison, a derivative of human CXXC finger protein-1 with a specific affinity to non-methylated CpG dinucleotides [51] was adopted for prokaryotic DNA isolation in a LOOXSTER® enrichment kit [52]. Moreover, several cytosine methylation-dependent endonucleases such as MspJI [53], HpaII and McrB [54] have been reported as feasible tools for microbial DNA enrichment.

One of the most studied functions of DNA methylation in prokaryotes is as a component of the restriction-modification system. Restriction enzymes in prokaryotes can cleave foreign DNA in this system whereas DNA methylation protects the prokaryotic genome from destruction [55]. N6-methyladenine modification is widespread in prokaryotes but has been rarely reported in eukaryotic genomes [56]. DpnI is a methyl-directed restriction endonuclease that restricts DNA only when it is methylated on adenine residues within the GATC sequence [57]. A DpnI-mediated DNA enrichment strategy was introduced in which only bacteria but not human DNA could specifically bind to DpnI immobilised on magnetic beads [58]. This method enabled more than a 100-fold enrichment of the prokaryotic genomes present at 1/10,000 of the level of human DNA.

3.3. Limitation and Controversy

In general, the selective lysis of human cells followed by the degradation of background DNA is shown to be effective in reducing the host DNA. However, the most obvious downside of these methods is a biased recovery among the microbial species due to their unequal sensitivities to the lysing conditions. For example, *Mycoplasma* spp. and parasites are more likely to be destroyed by a selective lysis buffer, leading to a low recovery in sequencing. Freezing and thawing can also potentially disrupt the microorganisms, thus emphasising the preferential use of fresh samples and the avoidance of having multiple freeze–thaw cycles. Moreover, these protocols involve multiple steps of lysis and centrifugation that inevitably cause DNA loss, thus limiting the potential of low biomass samples to be successfully processed [15,59].

The methylomes of only a few bacterial species have been well defined; the methylomes of protists, fungi and viruses are even less characterised. Several bacteria exhibit a similar methylation pattern with the human genome; for instance, *Helicobacter pylori* exhibits a high density of 5 mC modification [60], which is similar to human DNA methylation. Fungi also display a large portion of 5 mC content in their genomes [61] as well as DNA viruses, which demonstrate that complex cytosine methylation is involved in the genome replication state and host environment [62]. However, the selective enrichment of microbial sequences using DNA binding proteins or methylation-dependent endonucleases greatly depends on the methylation state of the target genomes, resulting in an unequal recovery of the microbial reads and distorting the ratio of the different microorganism lineages. For example, although most bacterial genomes could be increased 70- to 200-fold using HpaII, a failure in enriching *Borrelia burgdorferi* has been reported [54]. Thus, methylation-based isolation methods have limited the types of microbes that can be enriched.

Of note, comparing the efficacy of approaches across different samples could be unjustified. The efficacy of an approach could be varied among different type of samples, depending on the original ratio of non-host DNA to host DNA. For example, if the proportion of microbial DNA in a sample is 10%, it can only be enriched 10-fold at most to reach the maximal 100%; if the proportion of microbial DNA is lower than 1%, it can be enriched 100-fold to reach the maximum. The disparities in analyses and sample properties can also lead to controversial results. For example, Marotz et al. concluded that an osmotic lysis followed by a PMA treatment in saliva samples was promising to remove host-derived sequencing reads with only a small taxonomic bias [11]. In contrast, Ganda et al. argued that this method not only decreased the host DNA but also reduced the bacterial DNA extracted from a bovine milk sample in a dose-dependent manner [35]. One explanation is the difference in approaches to assess the enrichment efficacy between these two studies.

The first study used shotgun sequencing to acquire reads from various organisms at a relative proportion whereas the latter was based on qPCR to quantify the exact amount of bacterial DNA. A microbiome study also reported that a PMA treatment could not impact on the human DNA proportion in a sputum sample due to its complexity and viscosity compared with saliva [24]. Nevertheless, despite these limitations, current host DNA depletion methods enable the characterisation of microbial profiles in samples dominated by human DNA and these approaches are being increasingly used in clinical metagenomics.

4. Application in Microbiome Research and Clinical Metagenomics

The majority of current human microbiome research focuses on characterising the microbial profile and its association with host gene functions in health and disease. As previously stated, several studies have succeeded in depleting the host DNA in various human samples. For instance, an osmotic lysis followed by a PMA treatment on human saliva could increase the amount of microbial sequencing reads from 1% to 89% [11]. Moreover, in the field of metagenomic research on non-human hosts, the host DNA depletion is also critical. For example, in a resistome study of the milk production environment, depleting the host DNA before extraction was proven to be an efficient approach to remove the bovine reads, thereby facilitating the identification of the antimicrobial resistance gene [33].

However, not all specimens are suitable to be subjected to host DNA depletion approaches. The major reason is that the amount of depleted human DNA is not enough to effectively change the relative abundance of the reads assigned to the microbial genome in metagenomic HTS studies [35]. In actual practice, the final proportion of the microbial DNA after host DNA depletion deserves greater concern rather than focusing on the fold of the reduction in the host DNA. In a few cases, even if the host DNA is largely eliminated, the ratio of microbial DNA to human DNA is still too low for a metagenomic analysis. For example, our unpublished data showed that the bacteria-to-human DNA ratio in a gastric biopsy was extremely low at 1:1,000,000. Even if the human DNA could be reduced 1000-fold by host DNA depletion, the ratio of the bacteria-to-human DNA was still about 1:1000 or even lower. This indicated that if 15 GB of raw data was obtained from sequencing, only 15 MB of the data belonged to bacterial DNA, which obviously demonstrated that very limited information could be provided. Such an inadequacy could also be the major reason why amplicon-targeted sequencing is predominantly used in microbiome research with low biomass samples.

Clinical metagenomics require a broad identification of known and uncharacterised pathogens, thereby providing genomic information for evolutionary tracing, mutation discovery and drug resistance characterisation [63,64]. The primary goals of clinical metagenomics differ from those of basic research; a microbiome study often focuses on the relationship between microorganisms and disease, necessitating more microorganism-related data and an unbiased presentation of diverse microbial populations whereas clinical metagenomics pay more attention to microbial detection—the presence or absence of pathogens in a more time- and cost-dependent manner than basic research. For untargeted metagenomic sequencing, where a broad spectrum of pathogens can be identified in a single assay [65], the efficient removal of background human genetic material could concentrate the pathogen DNA to increase the sensitivity of the detection. Therefore, host DNA depletion (Table 2) is more widely used in clinical metagenomics compared with basic microbiome research.

Table 2. Case examples of host DNA depletion in clinical metagenomics in the last five years.

Sample Type	Potential Clinical Indication	Sample Size	Depletion Method	Sequencing Platform	Reads Number	Ref.
Cerebrospinal fluid	Infectious aetiology identification	13	Selective lysis by a bead-beater tissue homogeniser followed by a Benzonase nuclease treatment	Ion Torrent PGM	N/A	[23]
Prosthetic joint sonicate fluid	Pathogen identification	408	MolYsis basic kit	Illumina HiSeq	2.8 million, mean	[38]
Urine	Pathogen identification	10	Differential centrifugation and MolYsis kit	MinION	0.026 million, median	[66]
Urine	Antimicrobial resistance marker identification	13	NEBNext microbiome kit	Ion Torrent PGM	N/A	[67]
Sputum	Pathogen detection	6	Microfluidic separation followed by DNase digestion	Illumina HiSeq	36.3 million, mean	[12]
Sputum, bronchoalveolar lavage and endotracheal aspirates	Diagnosis of known and unknown infections	40	Saponin-based differential lysis followed by HL-SAN DNase digestion	MinION	0.041 million, mean	[20]
Cerebrospinal fluid	Diagnosis of known and unknown infections	95	NEB Microbiome Enrichment Kit	Illumina HiSeq	5~10 million	[68]
Endotracheal aspirates	Pathogen identification	22	Saponin-based differential lysis followed by HL-SAN DNase digestion	MinION	6628, median	[69]
Synovial fluid	Pathogen detection	168	MolYsis basic kit	Illumina HiSeq	30 million, mean	[70]
Bone and joint infectious tissue	Pathogen detection and antibiotic susceptibility prediction	24	Ultra-Deep Microbiome Prep kit	Illumina HiSeq	20 million, mean	[71]
Valve tissue	Pathogen identification	1	Ultra-Deep Microbiome Prep kit	Illumina MiSeq	1.4 million, mean	[72]
Hepatic tissue	Diagnosis of unknown infections	1	Ultra-Deep Microbiome Prep kit	Illumina MiSeq	1.1 million, mean	[73]
Blood culture bottles inoculated with prosthetic joint tissue	Pathogen identification	9	MolYsis basic kit	Illumina MiSeq	10.3 million, mean	[74]
Blood	Pathogen detection	8	MolYsis complete kit and WGA	Illumina HiSeq	27.5 million, mean	[75]
Whole blood	Diagnosis of infection	101	MolYsis complete kit	Ion Torrent	N/A	[76]

Table 2. *Cont.*

Sample Type	Potential Clinical Indication	Sample Size	Depletion Method	Sequencing Platform	Reads Number	Ref.
Sputum	*M. tuberculosis* detection and antibiotic susceptibility prediction	40	MolYsis basic kit	Illumina MiSeq and MinION	3.6 million, mean	[77]
Prosthetic joint sonication fluid	Diagnosis of prosthetic joint infections	97	A 5 µm pore size filter	Illumina MiSeq	N/A	[78]
Urine	Pathogen detection and antimicrobial susceptibility prediction	40	NEB Microbiome Enrichment Kit	Ion Proton	N/A	[79]

N/A: data is not publicly available for analysis.

Most clinical samples are processed promptly to avoid the release of both the host and microbial DNA from the freeze–thaw cycle. Therefore, differential centrifugation [80,81] and a pre-lysis followed by a DNA nuclease [20,66] are reliable methods to minimise the amount of host DNA in fresh samples. Unlike mock samples cultured in the laboratory, microorganisms in clinical specimens may have already been harmed by host immune cells or antibiotic treatments before sampling or damaged during the transportation and handling of the sample. These events inevitably result in a certain extent of release of cell-free DNA from pathogens that could be subsequently lost during selective lysis protocols. In comparison, although a methylated-CpG capture-based strategy is less efficient than a selective lysis [30] with a reported biased recovery toward several microbes such as *Neisseria flavescens* [46], such a strategy is capable of retaining cell-free DNA from dead organisms to avoid DNA loss, as observed when using a selective lysis [68,79]. Combining size- or density-based separation methods with a methylated CpG capture strategy could increase the proportion of the host DNA depletion whilst retaining the cell-free microbial DNA.

To ensure accuracy and avoid false-positive results, a threshold should be established prior to the detection [65]; a species or genus should only be considered to be "detected" if it meets the threshold otherwise it should be reported as "not detected". In this context, reporting the presence or absence of a pathogen does not require a large amount of sequencing reads from the HTS testing. Therefore, for a low biomass sample, even the ratio of pathogen-to-human DNA would be about 1:1000 or less; it could still be considered to be positive as long as its signal exceeded the threshold. Therefore, the host DNA depletion could be robustly applied to low biomass samples in clinical metagenomics as the increased amount of the pathogen reads in the patients could simultaneously enhance the sensitivity of the diagnosis. Emerging data have suggested that the host DNA depletion in clinical metagenomics facilitates an improvement in the diagnosis sensitivity as well as the discovery and identification of potential pathogens and gene features [20,68,71]. These approaches are still being challenged by the reproducibility of the results and a potential contamination in the reagents [82]. Proper consideration of these issues is necessary to enable the future application of host DNA depletion in clinical metagenomics.

5. New Strategies to Facilitate Metagenomic Sequencing in Samples with Overabundant Host DNA

In studies of DNA or RNA, a large proportion of nucleic acid molecules that are irrelevant to the question at hand are frequently encountered. Recently, several novel strategies for removing unwanted nucleic acid have been reported that may help to eliminate the host DNA in metagenomic sequencing (Figure 3).

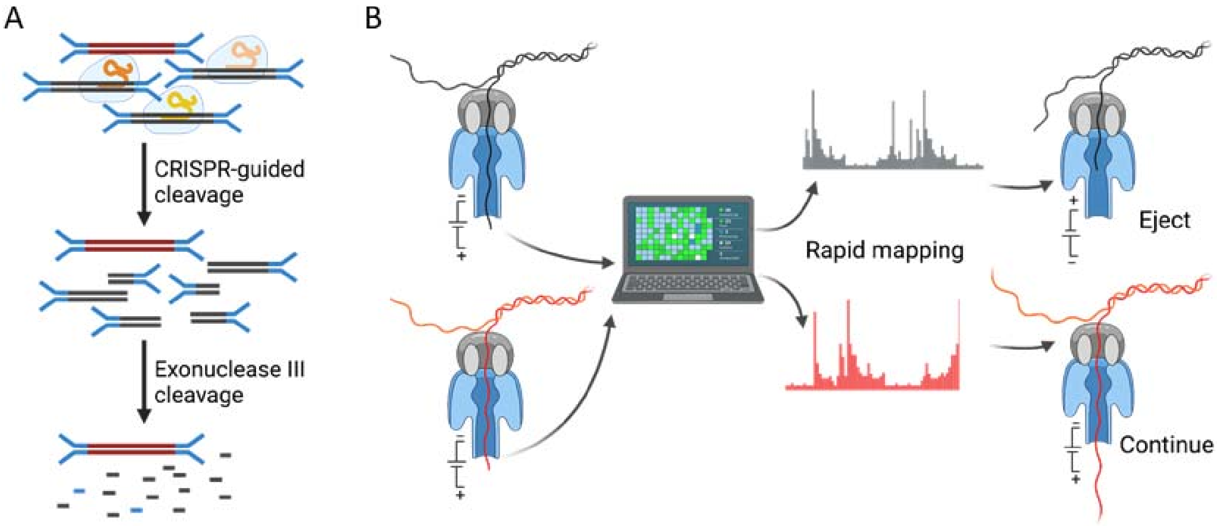

Figure 3. Illustration of strategies for removing unwanted high abundance DNA. (**A**) Sequencing library with Y-shape adapters contacts with a plurality of protein-guide RNA (gRNA) complexes in CRISPR/Cas9 system wherein gRNAs are complementary to the targeted human sequences to allow cleavage. Cleaved host DNA is then degraded by exonuclease III from blunt-ends cleaved by Cas9, leaving other sequences intact for subsequent amplification and sequencing [83]. (**B**) Using a nanopore device and computational approaches, individual double-strand DNA molecules can be selectively sequenced. When the DNA strand is sequenced, its current signal can be rapidly classified with or without base-calling. If the molecule is mapped to the pre-set reference genome such as a human genome, these reads would then be ejected from pores in real-time by revering the voltage polarity; otherwise, the sequencing would continue. Figures created using BioRender (https://biorender.com, accessed on 3 February 2022).

5.1. The Removal of Unwanted High Abundance Species in Sequencing Libraries

The clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein 9 (Cas9) system provides opportunities for depleting the targeted nucleic acid sequences in a sample (Figure 3). It works well in RNA sequencing to remove mitochondrial rRNA, which is the most abundant sequence in cerebral fluid-derived RNA samples [84]. However, to date, there is no comparable efficiency for using such an approach to deplete the host DNA as targeting the entire human DNA genome is impractical due to its requirement of a substantially high cost but with a low efficiency. One potential solution to improve the performance is to design guide RNA-targeting multicopy sequences such as primate-specific Alu elements, which comprise 11% of the entire human genome [85]. On the other hand, a combination of a CRISPR/Cas9 cleavage with long read sequencing can also be a feasible solution. Given that only one Cas9 cleaving site is required per each long DNA fragment, unwanted sequences could be effectively removed.

5.2. Selective Sequencing

The Oxford Nanopore MinION is a portable real-time sequencing device that functions by sensing the change in the current flow through a nanopore when DNA passes through. Nanopore sequencing has the significant advantage of providing real-time data and analyses compared with the conventional Illumina platform, thereby allowing a fast turnaround time of clinical results. Another critical advantage of nanopore sequencing is its unique capability of conducting selective sequencing (Figure 3). Individual molecules can be selectively sequenced (ReadUntil) using only computational methods by allowing nanopore devices to selectively eject reads from the pores in real-time through reversing the voltage polarity across the specified pores for a short period of time (~0.1 s) [86]. However, this requires the rapid classification of the current signal from the first part of a read to determine whether the molecule should be sequenced or removed and replaced with a new molecule. Recently, an open-source mapper named UNCALLED [87] (https://github.com/skovaka/UNCALLED, accessed on 3 February 2022) has been established that can rapidly match the streaming nanopore current signals to a reference sequence without base-calling. Another recently developed toolkit, Readfish [88] (https://www.github.com/looselab/readfish, accessed on 3 February 2022), requires a sufficiently fast base-caller but can also design and control selective sequencing procedures by removing the need to have complex signal mapping algorithms. To date, selective sequencing has been used in a variety of studies including the depletion of known bacterial genomes within a metagenomic community, the enrichment of certain specific human genes associated with hereditary cancers and the enrichment of low abundance genomes from samples with mixed populations [87,88]. Although selective sequencing is yet to be implemented in depleting the whole human genome, this approach could theoretically be an attractive alternative for selectively enriching microbial DNA in human background DNA without a pre-sequencing enrichment during the sample preparation.

Nevertheless, there are always gaps between a newly established method and its popularisation. CRISPR/Cas9 system-assisted methods and selective sequencing by Oxford Nanopore MinION are currently conceptual applications for host DNA depletion; hence, further validation studies are needed to assess whether these strategies could be promising to enrich the microbial reads in samples with overabundant host DNA without a bias.

6. Conclusions and Future Perspectives

A range of host DNA depletion and microbial DNA enrichment methods for metagenomic sequencing have been developed and evaluated in samples with varying characteristics. Among all approaches, one consistent conclusion is that host DNA depletion improves the efficiency and sensitivity of metagenomic sequencing. The main controversy among these studies is the variation in performance across the methods and sample types, highlighting the need for an individual assessment of each host depletion method to determine its desired sample type. Numerous factors should be considered when choosing an optimal method including the sample type (e.g., solid tissues or liquid samples; fresh or frozen), microbial load, pathogens of interest and budget.

Identifying the difference between the host and microbial genome can theoretically broaden possible strategies to achieve a microbial DNA enrichment. Despite the fact that most of these new methods are not commercially available, they are still worth optimising given their promising initial results. A few novel strategies have aimed to eliminate high abundance DNA populations during the library preparation or sequencing steps to achieve a human DNA depletion. Overall, continuous efforts are needed to validate new approaches when using metagenomics to study microbiomes in samples with large amounts of DNA derived from a human host.

Author Contributions: Y.S. collected the data and drafted the manuscript; G.W. and H.C.-H.L. revised the manuscript; J.Y. supervised and revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by National Key R&D Program of China (No. 2020YFA0509200/2020YFA0509203), RGC Theme-based Res Scheme Hong Kong (T21-705/20-N), RGC Collaborative Research Fund (C4039-19GF, C7065-18GF), RGC-GRF Hong Kong (14163817), Vice-Chancellor's Discretionary Fund Chinese University of Hong Kong.

Data Availability Statement: Data are available within the article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chiu, C.Y.; Mille, S.A. Clinical metagenomics. *Nat. Rev. Genet.* **2019**, *20*, 341–355. [CrossRef] [PubMed]
2. Quick, J.; Grubaugh, N.D.; Pullan, S.T.; Claro, I.M.; Smith, A.D.; Gangavarapu, K.; Oliveira, G.; Robles-Sikisaka, R.; Rogers, T.F.; Beutler, N.A.; et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat. Protoc.* **2017**, *12*, 1261–1276. [CrossRef] [PubMed]

3. Metsky, H.C.; Siddle, K.J.; Gladden-Young, A.; Qu, J.; Yang, D.K.; Brehio, P.; Goldfarb, A.; Piantadosi, A.; Wohl, S.; Carter, A.; et al. Capturing sequence diversity in metagenomes with comprehensive and scalable probe design. *Nat. Biotechnol.* **2019**, *37*, 160–168. [CrossRef] [PubMed]
4. Yang, S.; Gao, X.; Meng, J.; Zhang, A.; Zhou, Y.; Long, M.; Li, B.; Deng, W.; Jin, L.; Zhao, S.; et al. Metagenomic analysis of bacteria, fungi, bacteriophages, and helminths in the gut of giant pandas. *Front. Microbiol.* **2018**, *9*, 1717. [CrossRef]
5. Johnson, J.S.; Spakowicz, D.J.; Hong, B.Y.; Petersen, L.M.; Demkowicz, P.; Chen, L.; Leopold, S.R.; Hanson, B.M.; Agresta, H.O.; Gerstein, M.; et al. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat. Commun.* **2019**, *10*, 5029. [CrossRef]
6. Di Cenzo, G.C.; Finan, T.M. The divided bacterial genome: Structure, function, and evolution. *Microbiol. Mol. Biol. Rev.* **2017**, *81*, e00019-17. [CrossRef]
7. Wilson, M.R.; Sample, H.A.; Zorn, K.C.; Arevalo, S.; Yu, G.; Neuhaus, J.; Federman, S.; Stryke, D.; Briggs, B.; Langelier, C.; et al. Clinical metagenomic sequencing for diagnosis of meningitis and encephalitis. *N. Engl. J. Med.* **2019**, *380*, 2327–2340. [CrossRef]
8. The Human Microbiome Project Consortium. A framework for human microbiome research. *Nature* **2012**, *486*, 215–221. [CrossRef]
9. Simner, P.J.; Miller, S.; Carroll, K.C. Understanding the promises and hurdles of metagenomic next-generation sequencing as a diagnostic tool for infectious diseases. *Clin. Infect. Dis.* **2018**, *66*, 778–788. [CrossRef]
10. Zhang, C.; Hu, T.; Xiu, L.; Li, Y.; Peng, J. Use of ultra-deep sequencing in a patient with tuberculous coxitis shows its limitations in extrapulmonary tuberculosis diagnostics: A case report. *Infect. Drug Resist.* **2019**, *12*, 3739–3743. [CrossRef]
11. Marotz, C.A.; Sanders, J.G.; Zuniga, C.; Zaramela, L.S.; Knight, R.; Zengler, K. Improving saliva shotgun metagenomics by chemical host DNA depletion. *Microbiome* **2018**, *6*, 42. [CrossRef] [PubMed]
12. Shi, X.; Shao, C.; Luo, C.; Chu, Y.; Wang, J.; Meng, Q.; Yu, J.; Gao, Z.; Kang, Y. Microfluidics-based enrichment and whole-genome amplification enable strain-level resolution for airway metagenomics. *Msystems* **2019**, *4*, e00198-19. [CrossRef] [PubMed]
13. Liang, G.; Bushman, F.D. The human virome: Assembly, composition and host interactions. *Nat. Rev. Microbiol.* **2021**, *19*, 514–527. [CrossRef] [PubMed]
14. Bal, A.; Pichon, M.; Picard, C.; Casalegno, J.S.; Valette, M.; Schuffenecker, I.; Billard, L.; Vallet, S.; Vilchez, G.; Cheynet, V.; 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**, *18*, 537. [CrossRef]
15. Hasan, M.R.; Rawat, A.; Tang, P.; Jithesh, P.V.; Thomas, E.; Tan, R.; Tilley, P. Depletion of human DNA in spiked clinical specimens for improvement of sensitivity of pathogen detection by next-generation sequencing. *J. Clin. Microbiol.* **2016**, *54*, 919–927. [CrossRef] [PubMed]
16. Yeoh, Y.K. Removing host-derived DNA sequences from microbial metagenomes via mapping to reference genomes. *Methods Mol. Biol.* **2021**, *2232*, 147–153. [CrossRef]
17. Fong, W.; Rockett, R.; Timms, V.; Sintchenko, V. Optimization of sample preparation for culture-independent sequencing of Bordetella pertussis. *Microb. Genom.* **2020**, *6*, e000332. [CrossRef]
18. Zelenin, S.; Hansson, J.; Ardabili, S.; Ramachandraiah, H.; Brismar, H.; Russom, A. Microfluidic-based isolation of bacteria from whole blood for sepsis diagnostics. *Biotechnol. Lett.* **2015**, *37*, 825–830. [CrossRef]
19. Anscombe, C.; Misra, R.V.; Gharbia, S. Whole genome amplification and sequencing of low cell numbers directly from a bacteria spiked blood model. *bioRxiv* **2018**, 153965. [CrossRef]
20. Charalampous, T.; Kay, G.L.; Richardson, H.; Aydin, A.; Baldan, R.; Jeanes, C.; Rae, D.; Grundy, S.; Turner, D.J.; Wain, J.; et al. Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. *Nat. Biotechnol.* **2019**, *37*, 783–792. [CrossRef]
21. Bruggeling, C.E.; Garza, D.R.; Achouiti, S.; Mes, W.; Dutilh, B.E.; Boleij, A. Optimized bacterial DNA isolation method for microbiome analysis of human tissues. *MicrobiologyOpen* **2021**, *10*, e1191. [CrossRef] [PubMed]
22. Wen, Y.; Xiao, F.; Wang, C.; Wang, Z. The impact of different methods of DNA extraction on microbial community measures of BALF samples based on metagenomic data. *Am. J. Transl Res.* **2016**, *8*, 1412–1425. [PubMed]
23. Oechslin, C.P.; Lenz, N.; Liechti, N.; Ryter, S.; Agyeman, P.; Bruggmann, P.; Leib, S.L.; Beuret, C.M. Limited correlation of shotgun metagenomics following host depletion and routine diagnostics for viruses and bacteria in low concentrated surrogate and clinical samples. *Front. Cell Infect. Microbiol.* **2018**, *8*, 375. [CrossRef] [PubMed]
24. Nelson, M.T.; Pope, C.E.; Marsh, R.L.; Wolter, D.J.; Weiss, E.J.; Hager, K.R.; Vo, A.T.; Brittnacher, M.J.; Radey, M.C.; Hayden, H.S.; et al. Human and extracellular DNA depletion for metagenomic analysis of complex clinical infection samples yields optimized viable microbiome profiles. *Cell Rep.* **2019**, *26*, 2227–2240. [CrossRef]
25. Ojo-Okunola, A.; Claassen-Weitz, S.; Mwaikono, K.S.; Gardner-Lubbe, S.; Zar, H.J.; Nicol, M.P.; du Toit, E. The influence of DNA extraction and lipid removal on human milk bacterial profiles. *Methods Protoc.* **2020**, *3*, 39. [CrossRef]
26. Amar, Y.; Lagkouvardos, I.; Silva, R.L.; Ishola, O.A.; Foesel, B.U.; Kublik, S.; Schöler, A.; Niedermeier, S.; Bleuel, R.; Zink, A.; et al. Pre-digest of unprotected DNA by benzonase improves the representation of living skin bacteria and efficiently depletes host DNA. *Microbiome* **2021**, *9*, 123. [CrossRef]
27. Israeli, O.; Makdasi, E.; Cohen-Gihon, I.; Zvi, A.; Lazar, S.; Shifman, O.; Levy, H.; Gur, D.; Laskar, O.; Beth-Din, A. A rapid high-throughput sequencing-based approach for the identification of unknown bacterial pathogens in whole blood. *Future Sci. OA* **2020**, *6*, Fso476. [CrossRef]

28. Fittipaldi, M.; Nocker, A.; Codony, F. Progress in understanding preferential detection of live cells using viability dyes in combination with DNA amplification. *J. Microbiol. Methods* **2012**, *91*, 276–289. [CrossRef]
29. Horz, H.P.; Scheer, S.; Vianna, M.E.; Conrads, G. New methods for selective isolation of bacterial DNA from human clinical specimens. *Anaerobe* **2010**, *16*, 47–53. [CrossRef]
30. Thoendel, M.; Jeraldo, P.R.; Greenwood-Quaintance, K.E.; Yao, J.Z.; Chia, N.; Hanssen, A.D.; Abdel, M.P.; Patel, R. Comparison of microbial DNA enrichment tools for metagenomic whole genome sequencing. *J. Microbiol. Methods* **2016**, *127*, 141–145. [CrossRef]
31. Bjerre, R.D.; Hugerth, L.W.; Boulund, F.; Seifert, M.; Johansen, J.D.; Engstrand, L. Effects of sampling strategy and DNA extraction on human skin microbiome investigations. *Sci. Rep.* **2019**, *9*, 17287. [CrossRef] [PubMed]
32. Heravi, F.S.; Zakrzewski, M.; Vickery, K.; Host, H.H. DNA depletion efficiency of microbiome DNA enrichment methods in infected tissue samples. *J. Microbiol. Methods* **2020**, *170*, 105856. [CrossRef] [PubMed]
33. Rubiola, S.; Chiesa, F.; Dalmasso, A.; Di Ciccio, P.; Civera, T. Detection of Antimicrobial Resistance Genes in the Milk Production Environment: Impact of Host DNA and Sequencing Depth. *Front. Microbiol.* **2020**, *11*, 1983. [CrossRef] [PubMed]
34. Yap, M.; Feehily, C.; Walsh, C.J.; Fenelon, M.; Murphy, E.F.; McAuliffe, F.M.; van Sinderen, D.; O'Toole, P.W.; O'Sullivan, O.; Cotter, P.D. Evaluation of methods for the reduction of contaminating host reads when performing shotgun metagenomic sequencing of the milk microbiome. *Sci. Rep.* **2020**, *10*, 21665. [CrossRef]
35. Ganda, E.; Beck, K.L.; Haiminen, N.; Silverman, J.D.; Kawas, B.; Cronk, B.D.; Anderson, R.R.; Goodman, L.B.; Wiedmann, M. DNA extraction and host depletion methods significantly impact and potentially bias bacterial detection in a biological fluid. *Msystems* **2021**, *6*, e0061921. [CrossRef]
36. Hansen, W.L.; Bruggeman, C.A.; Wolffs, P.F. Pre-analytical sample treatment and DNA extraction protocols for the detection of bacterial pathogens from whole blood. *Methods Mol. Biol.* **2013**, *943*, 81–90. [CrossRef]
37. Avanzi, C.; Del-Pozo, J.; Benjak, A.; Stevenson, K.; Simpson, V.R.; Busso, P.; McLuckie, J.; Loiseau, C.; Lawton, C.; Schoening, J.; et al. Red squirrels in the British Isles are infected with leprosy bacilli. *Science* **2016**, *354*, 744–747. [CrossRef]
38. Thoendel, M.J.; Jeraldo, P.R.; Greenwood-Quaintance, K.E.; Yao, J.Z.; Chia, N.; Hanssen, A.D.; Abdel, M.P.; Patel, R. Identification of prosthetic joint infection pathogens using a shotgun metagenomics approach. *Clin. Infect. Dis.* **2018**, *67*, 1333–1338. [CrossRef]
39. Probst, A.J.; Weinmaier, T.; DeSantis, T.Z.; Domingo, J.W.S.; Ashbolt, N. New perspectives on microbial community distortion after whole-genome amplification. *PLoS ONE* **2015**, *10*, e0124158. [CrossRef]
40. Cheng, M.; Cao, L.; Ning, K. Microbiome big-data mining and applications using single-cell technologies and metagenomics approaches toward precision medicine. *Front. Genet.* **2019**, *10*, 972. [CrossRef]
41. Sidore, A.M.; Lan, F.; Lim, S.W.; Abate, A.R. Enhanced sequencing coverage with digital droplet multiple displacement amplification. *Nucleic Acids Res.* **2016**, *44*, e66. [CrossRef] [PubMed]
42. Ruan, Q.; Ruan, W.; Lin, X.; Wang, Y.; Zou, F.; Zhou, L.; Zhu, Z.; Yang, C. Digital-WGS: Automated, highly efficient whole-genome sequencing of single cells by digital microfluidics. *Sci. Adv.* **2020**, *6*, eabd6454. [CrossRef] [PubMed]
43. Kubicek, J.; Singer, T.; Sander, A.-K. Method, Lysis Solution and Kit for Selectively Depleting Animal Nucleic Acids in a Sample. 2018, QIAGEN GmbH: U.S. Patent No. 10,655,122, 19 May 2020.
44. Lorenz, M.G. Use of Nucleases for Degrading Nucleic Acids in the Presence of Chaotropic Agents and/or Surfactants. U.S. Patent Application No. 11/885,256, 3 July 2008.
45. Radzieta, M.; Sadeghpour-Heravi, F.; Peters, T.J.; Hu, H.; Vickery, K.; Jeffries, T.; Dickson, H.G.; Schwarzer, S.; Jensen, S.O.; Malone, M. A multiomics approach to identify host-microbe alterations associated with infection severity in diabetic foot infections: A pilot study. *NPJ Biofilms Microbiomes* **2021**, *7*, 29. [CrossRef]
46. Feehery, G.R.; Yigit, E.; Oyola, S.O.; Langhorst, B.W.; Schmidt, V.T.; Stewart, F.J.; Dimalanta, E.T.; Amaral-Zettler, L.A.; Davis, T.; Quail, M.A.; et al. A method for selectively enriching microbial DNA from contaminating vertebrate host DNA. *PLoS ONE* **2013**, *8*, e76096. [CrossRef] [PubMed]
47. Osterloh, D.; Felsmann, K. Methods for Separating, Detecting or Enriching Different DNA Species. U.S. Patent Application No.14/056,756, 17 October 2013.
48. Smith, M.; Campino, S.; Gu, Y.; Clark, T.G.; Otto, T.D.; Maslen, G.; Manske, M.; Imwong, M.; Dondorp, A.M.; Kwiatkowski, D.P.; et al. An in-solution hybridisation method for the isolation of pathogen DNA from human DNA-rich clinical samples for analysis by NGS. *Open Genom. J.* **2012**, *5*, 10–2174. [CrossRef] [PubMed]
49. Willbanks, A.; Leary, M.; Greenshields, M.; Tyminski, C.; Heerboth, S.; Lapinska, K.; Haskins, K.; Sarkar, S. The evolution of epigenetics: From prokaryotes to humans and its biological consequences. *Genet. Epigenet.* **2016**, *8*, 825–836. [CrossRef] [PubMed]
50. Bird, A.P. CpG-rich islands and the function of DNA methylation. *Nature* **1986**, *321*, 209–213. [CrossRef]
51. Xu, C.; Bian, C.; Lam, R.; Dong, A.; Min, J. The structural basis for selective binding of non-methylated CpG islands by the CFP1 CXXC domain. *Nat. Commun.* **2011**, *2*, 227. [CrossRef]
52. Glassing, A.; Dowd, S.E.; Galandiuk, S.; Davis, B.; Jorden, J.R.; Chiodini, R.J. Changes in 16s RNA gene microbial community profiling by concentration of prokaryotic DNA. *J. Microbiol. Methods* **2015**, *119*, 239–242. [CrossRef]
53. Oyola, S.O.; Gu, Y.; Manske, M.; Otto, T.D.; O'Brien, J.; Alcock, D.; Macinnis, B.; Berriman, M.; Newbold, C.I.; Kwiatkowski, D.P.; et al. Efficient depletion of host DNA contamination in malaria clinical sequencing. *J. Clin. Microbiol.* **2013**, *51*, 745–751. [CrossRef]
54. Liu, G.; Weston, C.Q.; Pham, L.K.; Waltz, S.; Barnes, H.; King, P.; Sphar, D.; Yamamoto, R.T.; Forsyth, R.A. Epigenetic segregation of microbial genomes from complex samples using restriction endonucleases HpaII and McrB. *PLoS ONE* **2016**, *11*, e0146064. [CrossRef] [PubMed]

55. Blow, M.J.; Clark, T.A.; Daum, C.G.; Deutschbauer, A.M.; Fomenkov, A.; Fries, R.; Froula, J.; Kang, D.D.; Malmstrom, R.R.; Morgan, R.D.; et al. The epigenomic landscape of prokaryotes. *PLoS Genet.* **2016**, *12*, e1005854. [CrossRef] [PubMed]
56. Douvlataniotis, K.; Bensberg, M.; Lentini, A.; Gylemo, B.; Nestor, C.E. No evidence for DNA N^6 methyladenine in mammals. *Sci. Adv.* **2020**, *6*, eaay3335. [CrossRef] [PubMed]
57. Ratel, D.; Ravanat, J.L.; Berger, F.; Wion, D. N6-methyladenine: The other methylated base of DNA. *Bioessays* **2006**, *28*, 309–315. [CrossRef]
58. Barnes, H.E.; Liu, G.; Weston, C.Q.; King, P.; Pham, L.K.; Waltz, S.; Helzer, K.T.; Day, L.; Sphar, D.; Yamamoto, R.T.; et al. Selective microbial genomic DNA isolation using restriction endonucleases. *PLoS ONE* **2014**, *9*, e109061. [CrossRef]
59. Ferretti, P.; Farina, S.; Cristofolini, M.; Girolomoni, G.; Tett, A.; Segata, N. Exerimental metagenomics and ribosomal profiling of the human skin microbiome. *Exp. Dermatol.* **2017**, *26*, 211–219. [CrossRef]
60. Krebes, J.; Morgan, R.D.; Bunk, B.; Spröer, C.; Luong, K.; Parusel, R.; Anton, B.P.; König, C.; Josenhans, C.; Overmann, J.; et al. The complex methylome of the human gastric pathogen Helicobacter pylori. *Nucleic Acids Res.* **2014**, *42*, 2415–2432. [CrossRef]
61. Bewick, A.J.; Hofmeister, B.T.; Powers, R.A.; Mondo, S.J.; Grigoriev, I.V.; James, T.Y.; Stajich, J.E.; Schmitz, R.J. Diversity of cytosine methylation across the fungal tree of life. *Nat. Ecol. Evol.* **2019**, *3*, 479–490. [CrossRef]
62. Hoelzer, K.; Shackelton, L.A.; Parrish, C.R. Presence and role of cytosine methylation in DNA viruses of animals. *Nucleic Acids Res.* **2008**, *36*, 2825–2837. [CrossRef]
63. Snitkin, E.S.; Zelazny, A.M.; Thomas, P.J.; Stock, F.; Henderson, D.K.; Palmore, T.N.; Segre, J.A. Tracking a hospital outbreak of carbapenem-resistant Klebsiella pneumoniae with whole-genome sequencing. *Sci. Transl. Med.* **2012**, *4*, 148ra116. [CrossRef]
64. Gire, S.K.; Goba, A.; Andersen, K.G.; Sealfon, R.S.; Park, D.J.; Kanneh, L.; Jalloh, S.; Momoh, M.; Fullah, M.; Dudas, G.; et al. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* **2014**, *345*, 1369–1372. [CrossRef] [PubMed]
65. Gu, W.; Miller, S.; Chiu, C.Y. Clinical metagenomic next-generation sequencing for pathogen detection. *Annu. Rev. Pathol.* **2019**, *14*, 319–338. [CrossRef] [PubMed]
66. Schmidt, K.; Mwaigwisya, S.; Crossman, L.C.; Doumith, M.; Munroe, D.; Pires, C.; Khan, A.M.; Woodford, N.; Saunders, N.J.; Wain, J.; et al. Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-based metagenomic sequencing. *J. Antimicrob. Chemother.* **2017**, *72*, 104–114. [CrossRef]
67. Graham, R.M.; Doyle, C.J.; Jennison, A.V. Epidemiological typing of *Neisseria gonorrhoeae* and detection of markers associated with antimicrobial resistance directly from urine samples using next generation sequencing. *Sex Transm. Infect.* **2017**, *93*, 65–67. [CrossRef]
68. Miller, S.; Naccache, S.N.; Samayoa, E.; Messacar, K.; Arevalo, S.; Federman, S.; Stryke, D.; Pham, E.; Fung, B.; Bolosky, W.J.; et al. Laboratory validation of a clinical metagenomic sequencing assay for pathogen detection in cerebrospinal fluid. *Genome Res.* **2019**, *29*, 831–842. [CrossRef]
69. Yang, L.; Haidar, G.; Zia, H.; Nettles, R.; Qin, S.; Wang, X.; Shah, F.; Rapport, S.F.; Charalampous, T.; Methé, B.; et al. Metagenomic identification of severe pneumonia pathogens in mechanically-ventilated patients: A feasibility and clinical validity study. *Respir. Res.* **2019**, *20*, 265. [CrossRef]
70. Ivy, M.I.; Thoendel, M.J.; Jeraldo, P.R.; Greenwood-Quaintance, K.E.; Hanssen, A.D.; Abdel, M.P.; Chia, N.; Yao, J.Z.; Tande, A.J.; Mandrekar, J.N.; et al. Direct detection and identification of prosthetic joint infection pathogens in synovial fluid by metagenomic shotgun sequencing. *J. Clin. Microbiol.* **2018**, *56*, e00402-18. [CrossRef] [PubMed]
71. Ruppé, E.; Lazarevic, V.; Girard, M.; Mouton, W.; Ferry, T.; Laurent, F.; Schrenzel, J. Clinical metagenomics of bone and joint infections: A proof of concept study. *Sci. Rep.* **2017**, *7*, 7718. [CrossRef] [PubMed]
72. Kolb, M.; Lazarevic, V.; Emonet, S.; Calmy, A.; Girard, M.; Gaïa, N.; Charretier, Y.; Cherkaoui, A.; Keller, P.; Huber, C.; et al. Next-generation sequencing for the diagnosis of challenging culture-negative endocarditis. *Front. Med.* **2019**, *6*, 203. [CrossRef] [PubMed]
73. Lazarevic, V.; Gaïa, N.; Girard, M.; Leo, S.; Cherkaoui, A.; Renzi, G.; Emonet, S.; Jamme, S.; Ruppé, E.; Vijgen, S.; et al. When bacterial culture fails, metagenomics can help: A case of chronic hepatic brucelloma assessed by next-generation sequencing. *Front Microbiol.* **2018**, *9*, 1566. [CrossRef]
74. Sanabria, A.; Hjerde, E.; Johannessen, M.; Sollid, J.E.; Simonsen, G.S.; Hanssen, A.M. Shotgun-metagenomics on positive blood culture bottles inoculated with prosthetic joint tissue: A proof of concept study. *Front. Microbiol.* **2020**, *11*, 1687. [CrossRef] [PubMed]
75. Vijayvargiya, P.; Jeraldo, P.R.; Thoendel, M.J.; Greenwood-Quaintance, K.E.; Garrigos, Z.E.; Rizwan Sohail, M.; Chia, N.; Pritt, B.S.; Patel, R. Application of metagenomic shotgun sequencing to detect vector-borne pathogens in clinical blood samples. *PLoS ONE* **2019**, *14*, e0222915. [CrossRef] [PubMed]
76. Parize, P.; Muth, E.; Richaud, C.; Gratigny, M.; Pilmis, B.; Lamamy, A.; Mainardi, J.L.; Cheval, J.; de Visser, L.; Jagorel, F.; et al. Untargeted next-generation sequencing-based first-line diagnosis of infection in immunocompromised adults: A multicentre, blinded, prospective study. *Clin. Microbiol. Infect.* **2017**, *23*, e571–e574. [CrossRef] [PubMed]
77. Votintseva, A.A.; Bradley, P.; Pankhurst, L.; Elias, C.d.; Loose, M.; Nilgiriwala, K.; Chatterjee, A.; Smith, E.G.; Sanderson, N.; Walker, T.M.; et al. Same-day diagnostic and surveillance data for tuberculosis via whole-genome sequencing of direct respiratory samples. *J. Clin. Microbiol.* **2017**, *55*, 1285–1298. [CrossRef] [PubMed]

78. Street, T.L.; Sanderson, N.D.; Atkins, B.L.; Brent, A.J.; Cole, K.; Foster, D.; McNally, M.A.; Oakley, S.; Peto, L.; Taylor, A.; et al. Molecular diagnosis of orthopedic-device-related infection directly from sonication fluid by metagenomic sequencing. *J. Clin. Microbiol.* **2017**, *55*, 2334–2347. [CrossRef]
79. Barraud, O.; Ravry, C.; François, B.; Daix, T.; Ploy, M.C.; Vignon, P. Shotgun metagenomics for microbiome and resistome detection in septic patients with urinary tract infection. *Int. J. Antimicrob. Agents* **2019**, *54*, 803–808. [CrossRef]
80. Ji, P.; Zhang, Y.; Wang, J.; Zhao, F. MetaSort untangles metagenome assembly by reducing microbial community complexity. *Nat. Commun.* **2017**, *8*, 14306. [CrossRef]
81. Gu, W.; Deng, X.; Lee, M.; Sucu, Y.D.; Arevalo, S.; Stryke, D.; Federman, S.; Gopez, A.; Reyes, K.; Zorn, K.; et al. Rapid pathogen detection by metagenomic next-generation sequencing of infected body fluids. *Nat. Med.* **2021**, *27*, 115–124. [CrossRef]
82. Strong, M.J.; Xu, G.; Morici, L.; Bon-Durant, S.S.; Baddoo, M.; Lin, Z.; Fewell, C.; Taylor, C.M.; Flemington, E.K. Microbial contamination in next generation sequencing: Implications for sequence-based analysis of clinical samples. *PLoS Pathog.* **2014**, *10*, e1004437. [CrossRef]
83. Carpenter, M.L.; Bustamante, C.D.; Gourguechon, S.B. Compositions and Methods for Targeted Depletion, Enrichment, and Partitioning of Nucleic Acids Using Crispr/Cas System Proteins. U.S. Patent Application No.16/231,338, 21 December 2018.
84. Gu, W.; Crawford, E.D.; O'Donovan, B.D.; Wilson, M.R.; Chow, E.D.; Retallack, H.; DeRisi, J.L. Depletion of abundant sequences by hybridization (DASH): Using Cas9 to remove unwanted high-abundance species in sequencing libraries and molecular counting applications. *Genome Biol.* **2016**, *17*, 41. [CrossRef]
85. Deininger, P. Alu elements: Know the SINEs. *Genome Biol.* **2011**, *12*, 236. [CrossRef]
86. Loose, M.; Malla, S.; Stout, M. Real-time selective sequencing using nanopore technology. *Nat. Methods* **2016**, *13*, 751–754. [CrossRef] [PubMed]
87. Kovaka, S.; Fan, Y.; Ni, B.; Timp, W.; Schatz, M.C. Targeted nanopore sequencing by real-time mapping of raw electrical signal with uncalled. *Nat. Biotechnol.* **2021**, *39*, 431–441. [CrossRef] [PubMed]
88. Payne, A.; Holmes, N.; Clarke, T.; Munro, R.; Debebe, B.J.; Loose, M. Readfish enables targeted nanopore sequencing of gigabase-sized genomes. *Nat. Biotechnol.* **2021**, *39*, 442–450. [CrossRef] [PubMed]