



Article

Nanopore Is Preferable over Illumina for 16S Amplicon Sequencing of the Gut Microbiota When Species-Level Taxonomic Classification, Accurate Estimation of Richness, or Focus on Rare Taxa Is Required

Márton Szoboszlay ^{1,*}, Laetitia Schramm ¹, David Pinzauti ¹, Jeanesse Scerri ¹, Anna Sandionigi ² 
and Manuele Biazzo ¹

¹ The BioArte Ltd., SGN 3000 San Gwann, Malta

² Department of Informatics, Systems and Communication, University of Milan Bicocca, 20126 Milan, Italy

* Correspondence: m.szoboszlay@thebioarte.com

Abstract: Nanopore sequencing is a promising technology used for 16S rRNA gene amplicon sequencing as it can provide full-length 16S reads and has a low up-front cost that allows research groups to set up their own sequencing workflows. To assess whether Nanopore with the improved error rate of the Kit 12 chemistry should be adopted as the preferred sequencing technology instead of Illumina for 16S amplicon sequencing of the gut microbiota, we used a mock community and human faecal samples to compare diversity, richness, and species-level community structure, as well as the replicability of the results. Nanopore had less noise, better accuracy with the mock community, a higher proportion of reads from the faecal samples classified to species, and better replicability. The difference between the Nanopore and Illumina results of the faecal bacterial community structure was significant but small compared to the variation between samples. The results show that Nanopore is a better choice for 16S rRNA gene amplicon sequencing when the focus is on species-level taxonomic resolution, the investigation of rare taxa, or an accurate estimation of richness. Illumina 16S sequencing should be reserved for communities with many unknown species, and for studies that require the resolution of amplicon sequence variants.

Keywords: Nanopore; Illumina; 16S rRNA; gut microbiota; species-level taxonomy



Citation: Szoboszlay, M.; Schramm, L.; Pinzauti, D.; Scerri, J.; Sandionigi, A.; Biazzo, M. Nanopore Is Preferable over Illumina for 16S Amplicon Sequencing of the Gut Microbiota When Species-Level Taxonomic Classification, Accurate Estimation of Richness, or Focus on Rare Taxa Is Required. *Microorganisms* **2023**, *11*, 804. <https://doi.org/10.3390/microorganisms11030804>

Academic Editor: Sandrine P. Claus

Received: 13 February 2023

Revised: 2 March 2023

Accepted: 13 March 2023

Published: 21 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The 16S ribosomal RNA (rRNA) sequence is the best single molecular marker for identifying bacteria, and with the development of high-throughput sequencing methods, 16S rRNA gene amplicon sequencing has become a pivotal tool for the characterization of bacterial diversity and community structure of almost any environment [1,2]. Since the introduction of the MiSeq in 2011, Illumina instruments became the most widely used platforms for 16S rRNA gene sequencing [3]. They offer very low error rates (Phred quality (Q) scores of 30, equivalent to 99.9% accuracy, for the majority of base calls), but are limited in read length, currently capable of reaching only 2×300 bp. Furthermore, the results do not have the full information content of the 16S rRNA gene, as the reads can cover only one, two, or (at most) three out of the nine hypervariable regions. This limits the taxonomic accuracy of the results [4]. In addition, the cost of Illumina devices can be challenging for most research groups. Therefore, instead of performing the sequencing runs in-house, it has become common practice to ship DNA extracts or PCR products to sequencing centers. Having to outsource the sequencing limits the possibility of optimizing the library preparation and run parameters for the needs of each research project, possibly resulting in long turnaround times.

A promising solution to this is Nanopore sequencing technology, which has two crucial properties: long reads and a low up-front cost. Firstly, it can sequence the full 16S

rRNA gene retrieving the information content of all nine hypervariable regions. Secondly, the cost of a MinION sequencer is a small fraction of that of any Illumina or PacBio sequencing system. This enables individual research groups to set up and optimize their own library preparation and sequencing workflows. In addition, it is a small portable device, allowing the processing of samples on site without the need of elaborate laboratory infrastructure. Consequently, there is growing interest in using Nanopore for 16S rRNA gene amplicon sequencing, and recent publications have demonstrated its success with a variety of samples, for example freshwater [5], seawater and algal surfaces [6], soil [7,8], surfaces in the international space station [9], animal feces [10], urine [11], buccal and rectal swabs [12], ophthalmic samples [13], bronchoalveolar lavage fluid, blood, pleural and ascitic fluid, cerebrospinal fluid, and wound secretions [14].

The most important limitation of Nanopore sequencing is its comparatively high error rate. The benefit provided by the higher information content of the long reads for taxonomic classification is lost if the reads have low accuracy. In 2019, sequencing of mock communities of known composition led to the conclusion that species-level classification was not yet reliable with Nanopore [15]. Later studies determined the sequencing accuracy to be 92.08% [5] and 93.17–95.97% [16]. Considering that the threshold to distinguish bacterial species based on their 16S rRNA sequence has traditionally been 97% similarity and later revised to be 98.65% similarity [17], it is expected that species-level taxonomic classification of Nanopore reads is problematic with such sequencing accuracy rates. However, the introduction of the kit 12 chemistry in late 2021 brought a substantial improvement, with >99% (Q20) accuracy becoming attainable [18]. This makes the question of whether, or when, Nanopore should be adopted as the preferred sequencing technology instead of Illumina for 16S rRNA gene amplicon sequencing ever more relevant.

To obtain a clear answer, a comparison of Nanopore and Illumina sequencing results is needed that: (1) includes samples of known composition, such as a mock community; (2) includes real samples of high bacterial diversity; (3) compares the results on diversity, richness, and community structure instead of focusing only on specific taxa, such as potential pathogens; (4) assesses replicability, i.e., the differences between repeated analyses of the same sample; and (5) preferably uses the same bioinformatical tools and taxonomic reference database to process the Nanopore and Illumina data. Publications that included comparisons between Nanopore and Illumina 16S rRNA gene amplicon sequencing results are listed in Table 1. They fulfilled some of the above points; however, covering all of them in a single study using the current Nanopore sequencing chemistry with improved accuracy is missing. Therefore, using faecal samples collected from six individuals and a mock community sample, we compared 16S rRNA gene amplicon sequencing results from Nanopore and Illumina to assess their ability to distinguish samples based on the bacterial community structure at species and higher taxonomic levels, deliver replicable results and good coverage of species richness. Each sample was sequenced in three technical replicates with both sequencing platforms. The reads were processed with the EzBioCloud MTP pipeline and reference database [19].

Table 1. Published studies comparing 16S rRNA gene amplicon sequencing with Illumina and Nanopore.

Samples	Mock Community Included	Nanopore Sequencing	Illumina Sequencing	Difference in Community Structure	Difference in Diversity	Difference in Replicability	References
2 faecal samples from mice	No	V1-9, SQK-MAP006	V3-4	No difference, except at species level	Not compared	Not assessed	[20]
13 water samples	Yes	V1-9, SQK-RAB204	V4	Focus on the detection of potential pathogens and faecal indicators	Not compared	Not assessed	[21]
1 mock community	Yes	V1-9, SQK-RAB201	V1-3, V3-4, V4, V4-5, V4-6, V6, V8, V8-9	Good results with both at genus level	Not compared	Not assessed	[15]
6 tap water and biofilm samples	Yes	V1-9, SQK-RAB204 or SQK-RAB201	V3-4	Only the 15 most abundant genera were considered	Not compared	Not assessed	[22]
59 human nasal swabs	No	V1-9, SQK-RAB201	V5-6	Similar results	Similar diversity	Not assessed	[23]
12 indoor dust samples	No	V1-9, SQK-LSK108	V3-4	Difference at genus and species level, but not at family level	Not compared	Not assessed	[24]
50 human faecal samples	No	V1-9, SQK-16S024	V3-4	Good correlation at genus level but not at species level	Not compared	Not assessed	[25]
Only a mock community was included in the comparison with Illumina	Yes	V1-9 and 16S-ITS-23S, SQK-PBK004	V3-4	Nanopore performed better at genus level	Not compared	Only assessed for 16S-ITS-23S	[26]
Corneal and conjunctival swabs from 2 patients	No	V1-9, kit not specified	V4	Similar results	Not compared	Not assessed	[27]
6 human faecal samples	Yes	V1-9 and V3-4, SQK-PBK004	V3-4	Only the 15 most abundant genera were considered	Not compared	Not assessed	[28]
8 human endometrial tissue samples	No	V1-9, SQK-RAB204	6 primer pairs pooled to cover V1-9	Similar results	Not compared	Not assessed	[29]
8 mock communities	Yes	V1-9, SQK-LSK108	V1-2, V1-3, V3, V4	Nanopore showed overall higher bias	Not compared	Not assessed	[30]
31 surface swabs from intensive care units	No	V1-9, SQK-16S024	V4	Focused on the detection of potential pathogens	Not compared	Not assessed	[31]
A mixture of DNA from 72 cheek skin swabs	Yes	V1-9 and 16S-ITS-23S, SQK-LSK110	V1-3	Similar results	Not compared	Good correlation between duplicates	[32]
8 soil samples from 2 sites	Yes	V1-9 and V3-4, SQK-LSK109	V3-4	Significant effect, but not as strong as the differences between sites	Difference in richness estimates	Not assessed	[33]

2. Materials and Methods

2.1. Samples

This study included the ZymoBIOMICS Gut Microbiome Standard (Zymo Research, Irvine, CA, USA), hereafter referred to as the mock community, which is a mixture of 14 bacterial, 1 archaeal, and 2 fungal species, as well as faecal samples collected from six individuals: sample A1 from a 2-year-old male child, samples AN1 and S1X from adult females, samples B1 and MD01 from adult males, and sample S2P from an adult male diagnosed with irritable bowel syndrome (IBS). The samples were obtained with the informed consent of the participants or their parents, and were anonymized. The work was carried out from microbial material and no human data was used. The faecal material was sampled with Danastool Sample Collection Microbiome Kit (Danagen, Barcelona, Spain). The samples were homogenized by vigorous shaking and transported to the laboratory at room temperature.

DNA was extracted on the day of sample collection. The Danastool tubes were vortexed and from each sample, a 200 µL aliquot was extracted with MagMAX Microbiome Ultra Nucleic Acid Isolation Kit (Applied Biosystems, Waltham, MA, USA), using a King-Fisher Flex Purification System (Thermo Fisher Scientific, Waltham, MA, USA). In the case of the mock community, a 75 µL aliquot was extracted with the same protocol. The DNA concentration in the extracts was determined with the Qubit 1X dsDNA High-Sensitivity Kit (Invitrogen, Waltham, MA, USA).

2.2. Nanopore Sequencing

Each DNA extract was processed in three technical replicates. The 16S rRNA gene was amplified with primers 27f (5'-TTTCTGTTGGTGCTGATATTGC-AGRGTTYGATYMTGGCTCAG-3') and 1492r (5'-ACTTGCCGTGTCGCTCTATCTTC-CGGTTACCTTGTTACGACTT-3'). The 22 nt 5' tails were added to the primers to serve as the annealing sites for the barcoding primers. The PCRs were carried out in 12.5 µL total volume containing 6.25 µL LongAmp Taq 2X Master Mix (New England Biolabs, Ipswich, MA, USA), the primers in 400 nM concentration, and 0.05 ng template DNA. The reactions were run in a T100 thermal cycler (BioRad, Hercules, CA, USA). The initial denaturation at 95 °C for 4 min was followed by 30 cycles of 95 °C for 20 s, 51 °C for 30 s, and 65 °C for 4 min. The final extension was 5 min at 65 °C. The PCR products were checked on a 2% agarose gel and cleaned with 1 × Agencourt AMPure XP (Beckman Coulter, Indianapolis, IN, USA). The purity and yield of the PCR products were assessed with a NanoDrop 8000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA).

Barcodes and adapters were added in a second PCR using primers from the Nanopore PCR Barcoding Expansion Kit EXP-PBC001 (Oxford Nanopore Technologies, Oxford, UK). The reactions contained in 25 µL total volume 12.5 µL LongAmp Taq 2X Master Mix, 0.5 µL barcoding primer mix, and 25 ng 16S rRNA gene PCR product. The reactions were run in a T100 thermal cycler with initial denaturation at 95 °C for 3 min; 12 cycles of 95 °C for 15 s, 62 °C for 15 s, and 65 °C for 4 min; and a final extension for 5 min at 65 °C. The PCR products were checked on a 2% agarose gel, cleaned with 1 × Agencourt AMPure XP, and quantified with a NanoDrop 8000 spectrophotometer.

Since the EXP-PBC001 Kit contains 12 different barcodes, the PCR products had to be split between two libraries, one containing the technical replicates of samples A1, AN1, B1, and the mock community, and the other containing the technical replicates of samples MD01, S1X, and S2P. The libraries were pooled and subjected to DNA repair and end-prep with the NEBNext Companion Module for ONT Ligation Sequencing (New England Biolabs, Ipswich, MA, USA) in a 60 µL volume containing 1 µg DNA. The reactions were incubated in a T100 thermal cycler at 20 °C for 5 min, followed by 65 °C for 5 min. The libraries were cleaned with 1 × Agencourt AMPure XP and prepared for sequencing with the Nanopore Ligation Sequencing Kit SQK-LSK112. The two libraries were sequenced on separate R9.4.1 flow cells loading 5 fmol DNA. The flow cells were run on a GridION sequencer with real-time super accurate base-calling in MinKNOW 21.11.7 with Guppy 5.1.13 and a filter threshold of average Q-score 11. Barcoding was completed post-run in

MinKNOW with the default settings. The reads are accessible in the European Nucleotide Archive under the accession number PRJEB56380.

2.3. Illumina Sequencing

Three aliquots of each DNA extract were sent to Novogene (Cambridge, UK). The V4 region of the 16S rRNA gene was amplified with primers 515f (5'-GTGCCAGCMGCCGCGGTAA-3') and 806r (5'-GGACTACHVGGGTWTCTAAT-3'), and the PCR products were sequenced on a NovaSeq 6000 generating 250 nt paired-end reads. The reads were demultiplexed and the primer sequences were trimmed by Novogene. The reads were deposited in the European Nucleotide Archive under the accession number PRJEB56380.

2.4. Data Processing

Both the Illumina and Nanopore reads were processed with the EzBioCloud Microbial Taxonomic Profiling (MTP) pipeline and the PKSSU4.0 database [19,34] to remove non-target, low-quality, and chimeric reads, and to perform the taxonomic classification of the filtered sequences. Briefly, the pipeline merges paired reads, discards reads that are <100 bp, or >2000 bp, or have an average Q-score < 25, or are not predicted to be 16S sequences. The dataset is then de-replicated and taxonomic assignment is done with VSEARCH [35] against the EzBioCloud 16S database. Reads that do not match any reference sequence are subjected to chimera detection with UCHIME [36] using a manually curated chimera-free reference database. Sequences that do not have a match in the EzBioCloud 16S database with at least 97% similarity are clustered into 97% OTUs with UCLUST [37]. Singletons OTUs are discarded.

From each sample the EzBioCloud server allows uploading up to 100,000 reads to the pipeline which were randomly selected in case more reads were available from a sample. Since the pre-processing steps of the EzBioCloud MTP pipeline were not designed for Nanopore data, it was necessary to trim and filter the Nanopore reads beforehand. Trimming was done with Cutadapt 3.5 [38] to remove adapter, barcode, and primer sequences. The presence of at least 15 bases of each primer sequence was required with no more than 0.2 error. Reads that did not have recognizable primer sequences on both ends were discarded. The filterAndTrim function of the dada2 1.18.0 R package [39] was used to discard reads that were shorter than 1300 bp or longer than 1600 bp, had ambiguous bases, or had more than 25 expected errors.

The results of the EzBioCloud MTP pipeline were organized into three MTP sets: one containing the Nanopore data, one containing the Illumina data, and one merging the data from both sequencing platforms. Hereafter, these are referred to as the Nanopore MTP set, the Illumina MTP set, and the merged MTP set, respectively. The EzBioCloud 16S reference database has different versions to match various regions of 16S sequences. In these, taxa that are indistinguishable by the given 16S region are merged into groups. The merged MTP set was created with a version of the EzBioCloud 16S reference database that joins species that either the V4 or V1–V9 regions can't distinguish into groups. Consequently, the classification results from the Nanopore data were not identical in the Nanopore MTP set and the merged MTP set. In the analysis of the results, the merged MTP set was used only when it was inevitable to have the Nanopore and Illumina data in a single data matrix.

For comparison, the Illumina reads were also processed with two other pipelines. In the first, the Divisive Amplicon Denoising Algorithm 2 (DADA2) [39] was used to quality filter, trim, denoise, and merge read pairs. Chimeric sequences were then removed using the consensus method. The taxonomic assignment of the amplicon sequence variants (ASVs) was carried out using the feature-classifier2 plugin [40], implemented in QIIME2 against the SILVA SSU non-redundant database (138 release) [41] adopting a consensus confidence threshold of 0.8. In the second pipeline, the reads were processed in mothur [42] following the MiSeq SOP [43] with the RDP taxonomy 18 reference database [44].

2.5. Statistical Analysis

The merged MTP set was used only in the cluster analysis and the permutational multivariate analysis of variance (PERMANOVA). All other results were obtained from the Nanopore and Illumina MTP sets. The analyses were carried out in R 4.2.1 (<https://www.r-project.org>) with functions from the vegan package 2.6-2 [45]. Cluster analysis was performed with the hclust function. The data matrix was relativized before calculating Bray–Curtis dissimilarities. Clusters were found with the unweighted pair group method with arithmetic mean (UPGMA). PERMANOVA based on Bray–Curtis dissimilarities was used with the adonis function to test the marginal effects of sequencing method and the differences between faecal samples. To quantify the proportion of variance explained by these factors, ω^2 values were calculated from the PERMANOVA results with the adonis_OmegaSq function of the MicEco package 0.9.17 [46]. To compare the differences in bacterial community structure between faecal samples to the differences between technical replicates of the same sample, Bray–Curtis dissimilarities were calculated from the Nanopore and Illumina MTP sets and compared with Welch two-sample t-tests. Shannon and Simpson diversity indices were calculated with the diversity function of the vegan package. Rarefaction curves were generated with the rarecurve function.

3. Results

3.1. Sequencing Yield

Nanopore sequencing generated 4,289,400 reads in total (Table S1) with median quality score 17.6 and 17.5 in the two libraries and a median read length of 1637 nt in both libraries. 90.7% of the reads passed trimming and 55.7% (2,387,309 in total) the quality filter corresponding to $113,681 \pm 16,289$ reads per sample. Illumina sequencing yielded 2,509,425 read pairs in total or $119,496 \pm 11,038$ reads per sample. The read length was 253 nt and 93% of the bases had quality scores >30 . From each sample, up to 100,000 Nanopore reads and Illumina read pairs were processed with the EzBioCloud MTP pipeline. 1.7% of the Nanopore reads were found to be non-16S, in total 33 reads to be non-bacterial, and 19.8% to be chimeric (Table S2). $77,438 \pm 5496$ reads per sample passed the pipeline, 98.7% of which could be classified into species with valid names, phylotypes (putative species represented by full length, high-quality 16S rRNA gene sequences in the taxonomic reference database), or species groups (species that cannot be distinguished based on the sequenced region of the 16S rRNA gene). In comparison, 0.7% of the Illumina reads were low-quality or non-16S, 0.4% non-bacterial, and 8.2% chimeric (Table S3). $90,817 \pm 1481$ reads per sample passed the pipeline, 94.5% of which could be classified into species, phylotypes, or species groups.

3.2. Mock Community

The mock community contained 18 bacterial strains representing 14 species. With Nanopore sequencing, 35, 38, and 39 species-level taxa (species with valid names, phylotypes, and species groups) were identified in the three technical replicates of the mock community. In contrast, 2076, 2268, and 2548 species-level taxa were found in the Illumina data. Due to this gross overestimation of richness, we tried several approaches to remove noise from the Illumina data. Keeping only taxa that were detected in all three technical replicates left 1073 species-level taxa in the dataset, while discarding 4.2%, 5.0%, and 5.2% of the reads in the replicates. Removing species-level taxa that did not have at least 60 reads in total across the three replicates left 195, 195, and 204 species-level taxa but discarded 9.6%, 10.6%, and 12.4% of the reads. Deleting taxa that did not reach at least 0.1% relative abundance in any of the replicates left 87, 88, and 92 species-level taxa but removed 14.7%, 15.3%, and 17.8% of the reads. The Illumina data were also denoised with dada2 in the QIIME2 pipeline which found 949, 1212, and 1411 amplicon sequence variants in the three replicates of the mock community representing 471, 526, and 586 genera. With the mothur MiSeq SOP, 690, 705, and 749 genera were found in the three replicates.

Clostridium perfringens and *Enterococcus faecalis* were not detected either with Nanopore or with Illumina sequencing, and the Illumina data also missed *Salmonella enterica* (Table 2). This is explained by the very low abundance of these species in the mock community. In the Nanopore data, *Faecalibacterium prausnitzii* was represented by only a few reads, but many reads were classified into two *Faecalibacterium* phylotypes: *Faecalibacterium* GG697149 and *Faecalibacterium* NMTZ. *Faecalibacterium* GG697149 was also abundant in the Illumina data, as well as *Roseburia cecicola* and *Veillonella dispar*. These are very likely results of the misclassification of *Faecalibacterium prausnitzii*, *Roseburia hominis*, and *Veillonella rogosae* reads into species with very similar 16S rRNA sequences. When these misclassified reads were included, 96.2%, 96.3%, and 96.4% of the Nanopore reads from the three replicates were from species that were part of the mock community (Table 2). Furthermore, nearly 100% of the reads were classified into genera represented in the mock community (Table S4). In contrast, only 64.7%, 68.4%, and 72.5% of the Illumina reads were classified into the species of the mock community (Table 2). This improved somewhat when the results were classified at higher taxonomic levels: 69.8%, 72.7%, and 76.3% of the Illumina reads of the three replicates were from genera (Table S4) and 77.8%, 81.0%, and 82.6% from families that were present in the mock community. Removing species that did not reach at least 0.1% relative abundance in any of the replicates increased the proportion of Illumina reads classified into the species of the mock community to 75.9%, 83.2%, and 85.6% (Table 2).

Table 2. The relative abundance (%) of bacterial species in the mock community compared to the Nanopore and Illumina results.

Species	Composition		Nanopore		Illumina Unfiltered			Illumina Filtered		
<i>Akkermansia muciniphila</i>	0.97	1.61	1.68	1.58	0.67	0.58	0.39	0.79	0.69	0.47
<i>Bacteroides fragilis</i>	9.94	13.97	14.21	13.88	7.76	6.85	12.60	9.10	8.09	15.33
<i>Bifidobacterium adolescentis</i>	8.78	6.66	7.20	6.92	8.11	9.54	8.17	9.51	11.27	9.93
<i>Clostridioides difficile</i>	2.62	5.40	5.05	5.41	1.69	2.70	2.15	1.98	3.19	2.61
<i>Clostridium perfringens</i>	0.0002	0	0	0	0	0	0	0	0	0
<i>Enterococcus faecalis</i>	0.0009	0	0	0	0	0	0	0	0	0
<i>Escherichia coli</i> group	12.12	8.72	8.79	8.51	5.41	8.42	7.53	6.34	9.94	9.16
<i>Faecalibacterium prausnitzii</i>	17.63	0.0035	0.0048	0.0039	11.56	15.94	15.69	13.55	18.83	19.08
<i>Fusobacterium nucleatum</i>	7.49	7.57	7.33	7.03	9.08	13.60	4.47	10.65	16.06	5.44
<i>Lactobacillus fermentum</i>	9.63	5.67	5.65	6.00	2.30	3.11	2.25	2.69	3.68	2.73
<i>Prevotella corporis</i>	4.98	6.57	6.42	6.47	5.55	0.94	4.14	6.51	1.11	5.03
<i>Roseburia hominis</i>	9.89	7.10	7.07	7.06	4.08	5.49	4.79	4.79	6.48	5.83
<i>Salmonella enterica</i>	0.009	0.018	0.024	0.012	0	0	0	0	0	0
<i>Veillonella rogosae</i>	15.87	18.44	18.69	19.36	2.69	1.71	1.94	3.16	2.02	2.35
Total	99.9	81.7	82.1	82.2	58.9	68.9	64.1	69.1	81.4	78.0
<i>Faecalibacterium</i> GG697149		4.39	4.54	4.42	1.44	0.93	1.15	1.68	1.09	1.40
<i>Faecalibacterium</i> NMTZ		9.62	9.15	9.31	0	0	0	0	0	0
<i>Roseburia cecicola</i>		0.043	0.043	0.046	1.43	1.1	1.12	1.68	1.29	1.36
<i>Veillonella dispar</i>		0.45	0.43	0.39	2.92	1.58	2.02	3.43	1.87	2.46
Total		99.9	96.2	96.3	64.7	72.5	68.4	75.9	85.6	83.2

The Illumina filtered results were obtained by removing taxa that did not reach at least 0.1% relative abundance in any of the replicates.

Apart from the misclassification of the *Faecalibacterium prausnitzii* reads, the Nanopore results were close to the composition of the mock community (Table 2). The largest deviations were found in the cases of *Akkermansia muciniphila* and *Clostridioides difficile*, the relative abundances of which were overestimated. The differences between the three replicates were minimal. In contrast, the Illumina results showed large variations between the replicates, especially in the cases of *Bacteroides fragilis*, *Fusobacterium nucleatum*, and *Prevotella corporis*. These differences were also present in the genus-level classification results (Table S4). The QIIME2 and mothur pipelines gave similar results (Table S5).

3.3. Faecal Samples

With both sequencing methods, nearly 100% of the reads from the faecal samples could be classified to genera- or genus-level phylotypes. At species-level classification, however, 82.1–97.4% of the Nanopore reads but only 30.8–53.8% of the Illumina reads could be assigned to species with valid names or phylotypes (Table S6). The remaining reads were classified into species groups or could only be classified at a higher taxonomic level.

The Illumina data indicated very high bacterial richness in the faecal samples. It identified 187–292 species, 258–411 species groups, and 791–1236 phylotypes in the samples (Table S7). In comparison, 82–127 species, 9–32 species groups, and 46–249 phylotypes were found with Nanopore. Similarly, the samples had 353–483 genera according to the Illumina data, but only 63–94 based on Nanopore (Table S8). However, technical replicates prepared from the same faecal sample were similar to each other in their number of species, species groups, phylotypes, and genera in both the Nanopore and Illumina data. Rarefaction analysis indicated that with Nanopore, our sequencing effort was enough to capture almost the complete bacterial richness in the samples, while with Illumina more sequences would be needed to obtain the same coverage (Figure S1).

Despite the large difference in the number of species-level taxa detected with Nanopore and Illumina sequencing, the difference in Shannon and Simpson diversity was relatively small. In fact, Simpson diversity was higher according to the Nanopore data than based on the Illumina results in samples AN1 and B1 (Figure 1).

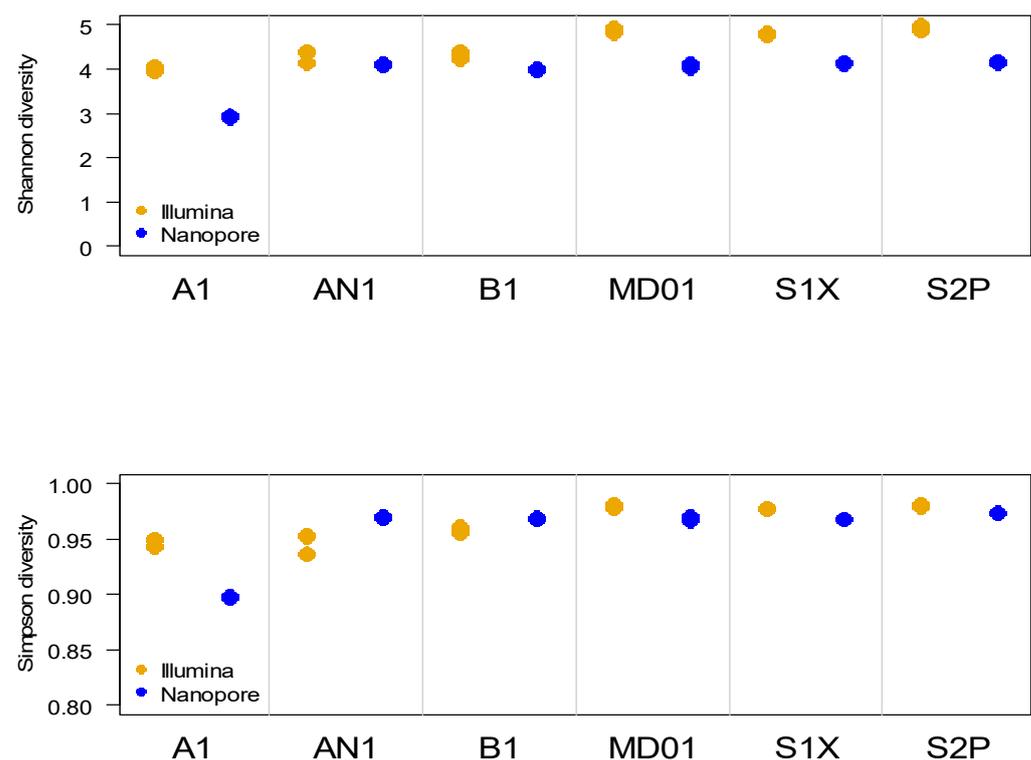


Figure 1. Shannon and Simpson diversity in the faecal samples.

The six stool samples were clearly distinguished based on their bacterial community structures in both the Illumina and Nanopore data (Figure 2). The technical replicates prepared from the same faecal sample with the same sequencing method were very similar. In the Nanopore data, the replicates clustered together at shorter distances, indicating less replicate-to-replicate variation than in the Illumina data. In the case of every faecal sample, the technical replicates divided into two clusters according to the sequencing method. This shows that the sequencing method clearly affected the bacterial community structure results, but this effect was smaller than the differences between the six faecal

samples. In fact, the difference between the Nanopore and Illumina results was significant ($p < 0.001$) and accounted for 10.9% of the variation in the bacterial community structure at species-level resolution and 10.0% at genus-level resolution. However, the differences between the six faecal samples explained 81.3% and 82.4% of the variation at species- and genus-level, respectively. The clustering results obtained from species- and genus-level taxonomy were very similar (Figure S2).

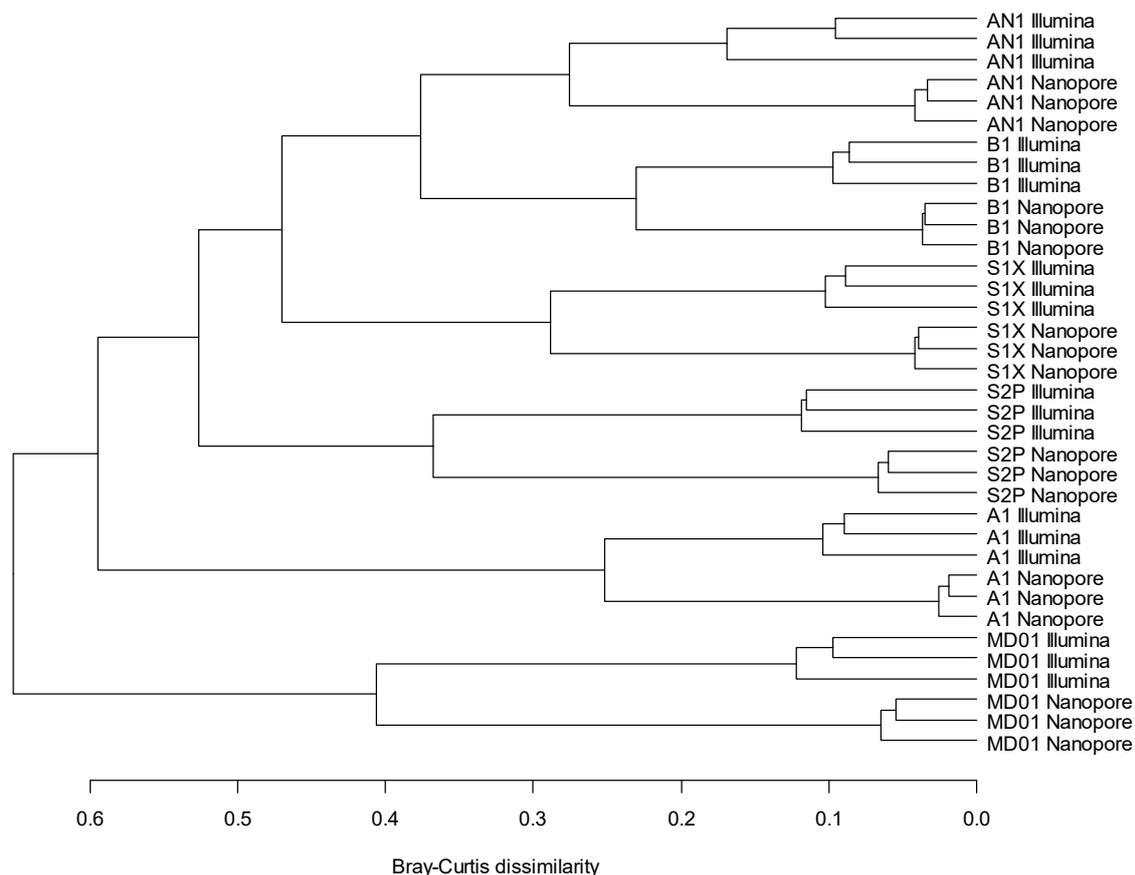


Figure 2. Clustering of the faecal samples based on their species-level bacterial community structure.

Differences between the Nanopore and Illumina results were present even at the highest taxonomic level. Illumina identified 46 phyla in the dataset, while Nanopore detected only 10. Regarding the most abundant phyla, the relative abundance of *Firmicutes*, and consequentially *Firmicutes/Bacteroidetes* and *Firmicutes/Actinobacteria* ratios, were higher with Nanopore than with Illumina in every faecal sample (Table 3). On the other hand, the relative abundance of *Proteobacteria* was consistently measured to be lower with Nanopore. In the case of the other abundant phyla, however, some samples had very similar results in the Illumina and Nanopore datasets (e.g., *Actinobacteria* in sample A1 or *Bacteroidetes* in sample S2P), while others were very different (e.g., *Actinobacteria* in sample S2P or *Bacteroidetes* in sample MD01). The effect of the sequencing method on the results depended on the faecal sample. With the small number of samples included in this study, this interaction impeded the characterization of the differences in the taxonomic composition results between the Nanopore and Illumina data. Nevertheless, it is clear from the data that even at phylum-level classification, the differences between the Nanopore and Illumina results cannot be removed by applying simple conversion factors.

Table 3. Relative abundance (% \pm SD) of the dominant phyla and their ratios in the faecal samples.

Sample	Sequencing	Firmicutes	Bacteroidetes	Actinobacteria	Proteobacteria	Verrucomicrobia	Firmicutes/ Bacteroidetes	Firmicutes/ Actinobacteria
A1	Nanopore	79.1 \pm 0.3	7.6 \pm 0.1	12.5 \pm 0.4	0.5 \pm 0.01	0.25 \pm 0.04	10.3 \pm 0.15	6.4 \pm 0.23
	Illumina	71.6 \pm 1.3	9.5 \pm 0.3	12.2 \pm 1.0	4.6 \pm 0.2	0.20 \pm 0.03	7.5 \pm 0.36	5.9 \pm 0.57
AN1	Nanopore	64.9 \pm 0.2	14.2 \pm 0.3	18.8 \pm 0.2	2.2 \pm 0.2	0.0004 \pm 0.0008	4.6 \pm 0.06	3.5 \pm 0.06
	Illumina	50.2 \pm 1.2	19.0 \pm 1.6	23.5 \pm 1.8	4.3 \pm 0.2	0.39 \pm 0.05	2.6 \pm 0.25	2.2 \pm 0.23
B1	Nanopore	73.1 \pm 0.4	10.6 \pm 0.2	15.5 \pm 0.4	0.8 \pm 0.1	0.012 \pm 0.002	6.9 \pm 0.17	4.7 \pm 0.10
	Illumina	62.4 \pm 0.6	11.4 \pm 0.8	18.6 \pm 0.8	5.8 \pm 0.4	0.22 \pm 0.03	5.5 \pm 0.46	3.4 \pm 0.12
MD01	Nanopore	77.8 \pm 0.8	1.7 \pm 0.1	11.3 \pm 0.6	3.3 \pm 0.1	5.05 \pm 0.17	46.7 \pm 2.05	6.9 \pm 0.46
	Illumina	60.9 \pm 2.4	12.7 \pm 0.3	14.8 \pm 1.1	4.2 \pm 0.5	4.55 \pm 1.68	4.8 \pm 0.26	4.1 \pm 0.51
S1X	Nanopore	66.9 \pm 0.8	23.1 \pm 0.6	8.7 \pm 0.3	0.7 \pm 0.02	0.095 \pm 0.02	2.9 \pm 0.10	7.7 \pm 0.31
	Illumina	58.3 \pm 1.1	21.6 \pm 0.8	11.0 \pm 0.4	6.2 \pm 0.2	0.42 \pm 0.02	2.7 \pm 0.15	5.3 \pm 0.26
S2P	Nanopore	66.9 \pm 1.0	28.5 \pm 0.6	1.0 \pm 0.2	3.7 \pm 0.3	0.0012 \pm 0.0012	2.4 \pm 0.06	71.5 \pm 12.23
	Illumina	53.8 \pm 1.2	26.8 \pm 2.0	6.7 \pm 0.6	9.9 \pm 0.7	0.55 \pm 0.07	2.0 \pm 0.20	8.1 \pm 0.56

To assess the replicability of the results from the two sequencing methods, we compared the differences in bacterial community structure among technical replicates prepared from the same faecal sample to the differences between the six faecal samples. Bray–Curtis dissimilarities calculated from the Nanopore and Illumina MTP sets showed that with both sequencing methods, the differences in bacterial community structure between technical replicates were minor compared to the differences between the faecal samples (Figure 3). This was the case with both species- and genus-level taxonomy. The difference between the variation among technical replicates from the same sample and the variation among faecal samples was larger in the Nanopore data. Furthermore, the Bray–Curtis dissimilarities between the replicates were significantly smaller with Nanopore than with Illumina (Welch *t*-test $p < 0.0001$ with both species- and genus-level taxonomy). Thus, the replicability was good with both sequencing methods, but better with Nanopore.

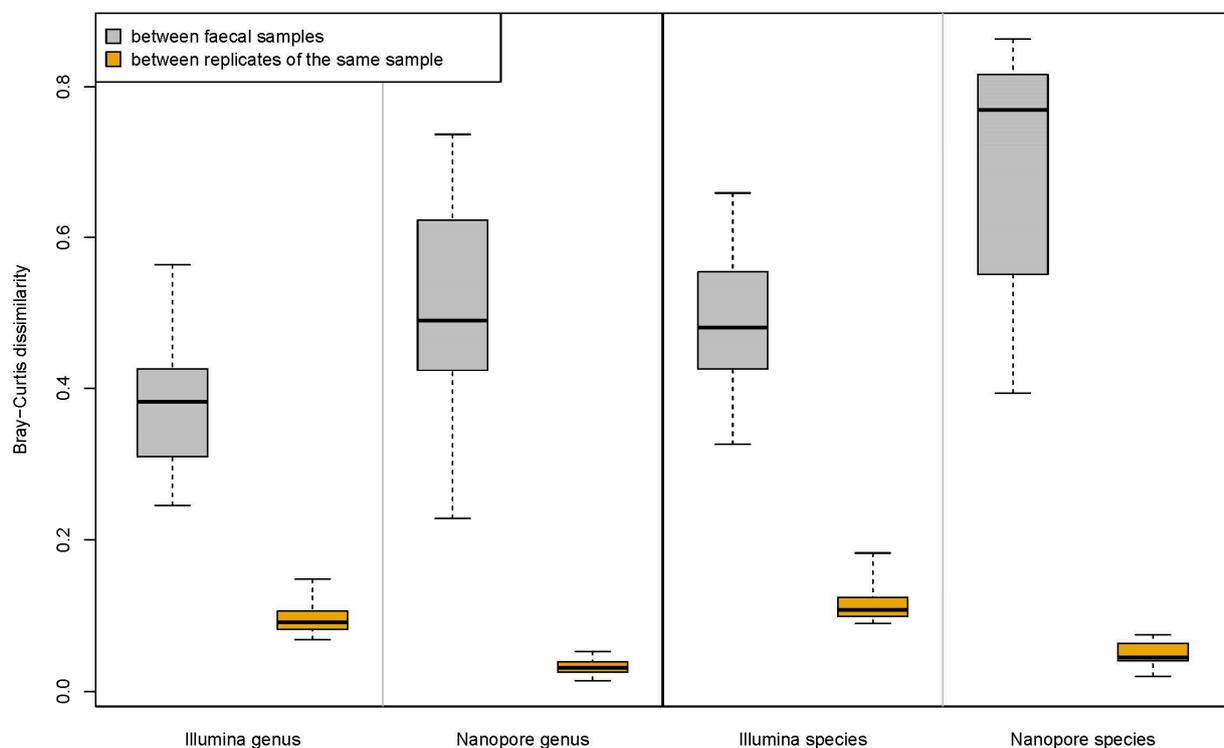


Figure 3. Dissimilarity in bacterial community structure at species- and genus-level taxonomy between technical replicates of the same sample and different faecal samples. The bold lines show the median values, the boxes cover the first to the third quartiles, and the whiskers extend to the data extremes.

4. Discussion

The most important concern about Nanopore sequencing is its higher error rate which could undermine the accuracy of the taxonomic classification of the reads. Therefore, our aim was to compare the Nanopore results with the Illumina setup that delivers the highest possible sequencing quality. Hence, we selected the NovaSeq 6000 system and targeted the V4 region of the 16S rRNA gene with which the 2×250 bp read pairs have a near complete overlap. This means that most base pairs were sequenced twice, generating an improved consensus sequence. In consequence, hardly any of the Illumina reads had to be removed due to insufficient quality during the data processing. The kit 12 chemistry brought a substantial improvement to the sequence quality of Nanopore; however, a very large portion (44.3%) of the Nanopore reads were still lost at the quality filtering step. With the recently introduced Kit 14 chemistry and R10.4.1 flow cells, it is likely that future studies will not have to expect losing this much data due to insufficient quality. The retained reads, however, had over 98.08% accuracy as estimated from the Q-scores. Since this accuracy is higher than the commonly used 97% similarity threshold to distinguish species, it was sensible to attempt the species-level classification of the reads.

There are no established best practices for the processing of Nanopore 16S rRNA gene amplicon reads and published studies used a wide range of methods and reference databases [47]. We selected the EzBioCloud MTP pipeline as it allowed us to process the Nanopore and Illumina data with the same toolset, except for the primer trimming and initial quality filtering step, and because the EzBioCloud 16S database is especially suitable for the taxonomic classification of full-length 16S reads at species-level. In brief, it is extensive, manually curated, contains only full-length 16S sequences, includes the complete taxonomy of every sequence, and is frequently updated (the current version was released on 7 July 2021). A comparison to the SILVA and GreenGenes [48] databases using mock community data found that the EzBioCloud 16S reference gave better genus- and species-level classification results with lower false-positive and false-negative rates [49]. Furthermore, the EzBioCloud MTP pipeline takes into account which hypervariable regions of the 16S rRNA gene are covered by the reads by merging the sequences in the reference database that are indistinguishable by the given 16S region and labelling them as a species group. Since the Nanopore reads covered all the hypervariable regions of the 16S rRNA gene while the Illumina reads covered only one, these groups were different for the two datasets. For example, with V1 – V9 reads, *Veillonella rogosae* can be distinguished from other *Veillonella* species but with only the sequence of the V4 region it cannot be distinguished from *V. parvula*, *V. atypica*, *V. rodentium*, *V. denticariosi*, and *V. tobetsuensis* [50]. Thus, when processing V4 reads, these six species are joined under the name “*V. parvula* group” in the EzBioCloud reference database. The merged MTP set that joined the Nanopore and Illumina results into a single data matrix was created with a version of the EzBioCloud 16S reference database that joins species that either the V4 or V1–V9 regions cannot distinguish into groups. If the Nanopore and Illumina results were directly merged without this, several taxa present in the Nanopore data would have been found missing from the Illumina data, not because Illumina sequencing failed to detect their 16S rRNA gene sequences, but because the short reads could not distinguish them from the sequences of related taxa. This would have artificially increased the effect of the sequencing method on the bacterial community structure.

16S rRNA amplicon sequencing results are known to be burdened with noise which may originate from contamination, PCR and sequencing errors, undetected chimeras, and barcode crosstalk. To mitigate the noise and improve the accuracy of the results, removal of singleton and rare OTUs has been recommended [51,52], and denoising methods such as dada2 [39] and UNOISE2 [53] were developed for Illumina reads. Our Illumina results from the mock community showed an extreme overestimation of richness and a large portion of the reads did not classify into taxa present in the mock community, not even at the family level. This indicates that the data had a considerable amount of noise. We arrived at these results with three different data processing pipelines that employ different quality filtering

methods, denoising, reference databases, and classification algorithms. Therefore, it is unlikely that the noise was a product of the bioinformatic toolset. Removing rare taxa improved the results, but even applying an abundance threshold that discarded 14.7–17.8% of the data left a 6.5-fold overestimation of species richness and still a large number of reads classified into taxa not present in the mock community. The source of the noise is unclear. Contamination and sequencing errors were not likely to be the cause since there were no detectable products in the negative controls and all run metrics indicated very high sequencing quality. Barcode crosstalk is a possibility as the library included a large number of samples from other clients of the sequencing center. Pooling samples from different projects is common practice, and with the increasing use of large-scale Illumina sequencers, such as the NovaSeq 6000, samples are sequenced in ever-larger libraries, which might make barcode crosstalk a significant source of error. It has been suggested to use the results from a mock community to determine the abundance threshold to discard rare OTUs or taxa and then apply this threshold to remove noise from all the samples [51]. This approach, however, seems questionable in our case. We identified fewer species-level taxa in the Illumina data from the faecal samples than from the mock community replicates, suggesting that the amount of noise was not even across the samples and a filtering threshold that works well for one sample might not be optimal for another. It is possible that low-diversity samples, such as a mock community, are more prone to generating noisy data.

With Nanopore, our results from the mock community were close to its actual composition and consistent between replicates. Over 99.9% of the reads were classified into the genera present in the mock community, indicating that the data were practically noise-free. Not having to apply an abundance threshold to mitigate noise is a major advantage for studies investigating the rare biosphere [54]. It also improves the performance of richness estimators that rely on the number of singleton and doubleton OTUs or taxa such as the Chao1 [55] and ACE [56] indices or breakaway [57]. The *F. prausnitzii* AP34BHI strain that was present in the mock community has six copies of the 16S rRNA gene in its genome which cluster into two groups by similarity, one consisting of two copies and the other of four copies (Figure S3). This ratio is similar to the *F. GG697149:F. NMTZ* relative abundance ratios in our results which were 1:2.19, 1:2.02, and 1:2.11 in the three replicates. Thus, it is likely that the intra-genomic variation in the *F. prausnitzii* AP34BHI 16S rRNA gene sequence led to the classification error.

In the case of the faecal samples, both sequencing methods delivered replicable results and could clearly distinguish the samples based on their bacterial community structures. The results with Nanopore and Illumina were not identical, but the effect of the sequencing method on the bacterial community structure was small compared to the differences between faecal samples. Differences between the Nanopore and Illumina results were present at every taxonomic level. However, these can be attributed to the different primer pairs and PCR protocols, the effects of which on 16S rRNA gene sequencing results are well documented in the literature [58,59]. Therefore, our investigation focused on the power to distinguish faecal samples and the replicability of these results instead of identifying differentially abundant taxa.

The Nanopore data outperformed the Illumina data with respect to the proportion of reads from the faecal samples that could be classified to species. This highlights the improved taxonomic resolution of the long reads that has been previously shown with *in silico* analysis and PacBio sequencing [4]. Due to the better taxonomic resolution, we expected that the Nanopore reads would identify a larger number of species in the faecal samples than the short Illumina reads. Yet, more species were found in the Illumina data. The fact that the samples had somewhat fewer Nanopore than Illumina reads might explain a small part of the difference in species richness between the two datasets. However, the rarefaction curves showed that the coverage of the Nanopore data was much better despite the lower read counts. The true richness of the samples is unknown, but based on the results from the mock community, we suspect that the Illumina data overestimated the number of species in the faecal samples due to noise.

In this study, Nanopore 16S rRNA gene amplicon sequencing showed good accuracy and good replicability at species-level taxonomic resolution. Thus, the data quality provided by the current sequencing chemistry and flow cells is clearly sufficient for characterizing the species composition of complex bacterial communities, such as the human gut microbiota. Methods developed to further decrease the error rate by amplicon self-ligation and rolling circle amplification [60], or with using molecular tags to build consensus sequences [61], are not required to achieve this. However, they might still be necessary if the aim is to reach the resolution of strains and amplicon sequence variants. A novel pipeline, Emu [62], is also a promising tool to correct errors and improve the accuracy of Nanopore 16S rRNA gene amplicon sequencing, but as the expectation–maximization algorithm inflates the number of rare species and requires the application of an abundance threshold, Emu is only applicable when rare taxa and the estimation of richness and diversity are not in the focus.

To our knowledge, this is the first study to utilize the same bioinformatics pipeline and reference database to compare the accuracy of species-level classification of a mock community, the obtained diversity, richness, and community structure of real samples, and the technical replicability of Oxford Nanopore and Illumina technology simultaneously. Our results show that Nanopore long-read sequencing makes up for its higher error rate for species-level resolution. In conclusion, Nanopore is a better choice than Illumina for 16S rRNA gene amplicon sequencing of the gut microbiota when the focus is on assessing bacterial community structure at species-level taxonomic resolution, on the investigation of rare taxa, or on obtaining an accurate estimation of richness. Using Illumina 16S sequencing should be reserved for bacterial communities with many unknown species that are thus not represented in reference databases, and for studies that require the resolution of amplicon sequence variants. With the frequent release of improved sequencing chemistries and flow cells, the advantages of Nanopore can only be expected to grow in the near future.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/microorganisms11030804/s1>. Table S1: Sequencing yield. Table S2: The number of Nanopore reads at each stage of the EzBioCloud MTP pipeline. Table S3: The number of Illumina reads at each stage of the EzBioCloud MTP pipeline. Table S4: Relative abundance (%) of the bacterial genera in the mock community compared to the Nanopore and Illumina results from the EzBioCloud MTP pipeline. Table S5: Relative abundance (%) of the bacterial genera in the mock community compared to the Illumina results from the QIIME2 and mothur pipelines. Table S6: The percentage of reads classified into species and genera in the faecal samples. Table S7: The number of species-level taxa detected in the faecal samples. Table S8: The number of genera detected in the faecal samples. Figure S1: Rarefaction curves. Figure S2: Clustering of the faecal samples based on their genus-level bacterial community structure. Figure S3: Average distance clustering of the 16S rRNA gene sequences of the *Faecalibacterium prausnitzii* AP34BHI genome after Clustal WS alignment.

Author Contributions: M.S.: conceptualization, methodology, investigation, formal analysis, writing—original draft; L.S.: conceptualization and writing—review and editing; D.P.: conceptualization and writing—review and editing; J.S.: conceptualization and writing—review and editing; A.S.: conceptualization and writing—review and editing; M.B.: funding acquisition, conceptualization, and writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This study was financed through the Research and Development Scheme 2020 Eureka Network Project 15839/NP/HealthyAgeing granted by Malta Enterprise.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: All sequencing data are accessible in the European Nucleotide Archive under the accession number PRJEB56380.

Acknowledgments: We owe thanks for the valuable discussion on this project to the entire team at The BioArte Ltd., and to James Schofield and Erika Parkinson. We would like to thank Christine Podrini for her comments on the manuscript and Maria D’Aguzzo for her work on project administration.

Conflicts of Interest: Laetitia Schramm, David Pinzauti, Jeanesse Scerri, and Manuele Biazzo are co-inventors of patent application EP22213544.4. Márton Szoboszlai and Anna Sandionigi declare no conflicts of interest.

References

- Janda, J.M.; Abbott, S.L. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: Pluses, perils, and pitfalls. *J. Clin. Microbiol.* **2007**, *45*, 2761–2764. [[CrossRef](#)] [[PubMed](#)]
- Pollock, J.; Glendinning, L.; Wisedchanwet, T.; Watson, M. The madness of microbiome: Attempting to find consensus “best practice” for 16S microbiome studies. *Appl. Environ. Microbiol.* **2018**, *84*, e02627-17. [[CrossRef](#)]
- Pichler, M.; Coskun, Ö.K.; Ortega-Arbulú, A.S.; Conci, N.; Wörheide, G.; Vargas, S.; Orsi, W.D. A 16S rRNA gene sequencing and analysis protocol for the Illumina MiniSeq platform. *Microbiologyopen* **2018**, *7*, e00611. [[CrossRef](#)]
- 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](#)] [[PubMed](#)]
- Urban, L.; Holzer, A.; Baronas, J.J.; Hall, M.B.; Braeuninger-Weimer, P.; Scherm, M.J.; Kunz, D.J.; Perera, S.N.; Martin-Herranz, D.E.; Tipper, E.T.; et al. Freshwater monitoring by nanopore sequencing. *ELife* **2021**, *10*, e61504. [[CrossRef](#)]
- van der Loos, L.M.; D’hondt, S.; Willems, A.; De Clerck, O. Characterizing algal microbiomes using long-read nanopore sequencing. *Algal Res.* **2021**, *59*, 102456. [[CrossRef](#)]
- Chavan, S.; Sarangdhar, V.; Vigneshwaran, N. Nanopore-based metagenomic analysis of the impact of nanoparticles on soil microbial communities. *Heliyon* **2022**, *8*, e09693. [[CrossRef](#)] [[PubMed](#)]
- Li, X.; Kong, P.; Daughtrey, M.; Kosta, K.; Schirmer, S.; Howle, M.; Likins, M.; Hong, C. Characterization of the Soil Bacterial Community from Selected Boxwood Gardens across the United States. *Microorganisms* **2022**, *10*, 1514. [[PubMed](#)]
- Stahl-Rommel, S.; Jain, M.; Nguyen, H.N.; Arnold, R.R.; Aunon-Chancellor, S.M.; Sharp, G.M.; Castro, C.L.; John, K.K.; Juul, S.; Turner, D.J.; et al. Real-time culture-independent microbial profiling onboard the international space station using nanopore sequencing. *Genes* **2021**, *12*, 106. [[CrossRef](#)]
- Beale, D.J.; Bissett, A.; Nilsson, S.; Bose, U.; Nelis, J.L.D.; Nahar, A.; Smith, M.; Gonzalez-Astudillo, V.; Braun, C.; Baddiley, B.; et al. Perturbation of the gut microbiome in wild-caught freshwater turtles (*Emydera macquarii macquarii*) exposed to elevated PFAS levels. *Sci. Total Environ.* **2022**, *838*, 156324. [[CrossRef](#)] [[PubMed](#)]
- Zhang, L.; Huang, W.; Zhang, S.; Li, Q.; Wang, Y.; Chen, T.; Jiang, H.; Kong, D.; Lv, Q.; Zheng, Y.; et al. Rapid Detection of Bacterial Pathogens and Antimicrobial Resistance Genes in Clinical Urine Samples with Urinary Tract Infection by Metagenomic Nanopore Sequencing. *Front. Microbiol.* **2022**, *13*, 858777. [[CrossRef](#)] [[PubMed](#)]
- Ammer-Herrmann, C.; Pfisterer, N.; van den Berg, T.; Gavrilova, I.; Amanzada, A.; Singh, S.K.; Khalil, A.; Alili, R.; Belda, E.; Clement, K.; et al. Comprehensive wet-bench and bioinformatics workflow for complex microbiota using Oxford Nanopore technologies. *mSystems* **2021**, *6*, e00750-21. [[CrossRef](#)] [[PubMed](#)]
- Omi, M.; Matsuo, Y.; Araki-Sasaki, K.; Oba, S.; Yamada, H.; Hirota, K.; Takahashi, K. 16S rRNA nanopore sequencing for the diagnosis of ocular infection: A feasibility study. *BMJ Open Ophthalmol.* **2022**, *7*, e000910. [[CrossRef](#)]
- Fu, Y.; Chen, Q.; Xiong, M.; Zhao, J.; Shen, S.; Chen, L.; Pan, Y.; Li, Z.; Li, Y. Clinical Performance of Nanopore Targeted Sequencing for Diagnosing Infectious Diseases. *Microbiol. Spectr.* **2022**, *10*, e00270-22. [[CrossRef](#)]
- Winand, R.; Bogaerts, B.; Hoffman, S.; Lefevre, L.; Delvoeye, M.; Van Braekel, J.; Fu, Q.; Roosens, N.H.; De Keersmaecker, S.C.; Vanneste, K. Targeting the 16s rRNA gene for bacterial identification in complex mixed samples: Comparative evaluation of second (illumina) and third (oxford nanopore technologies) generation sequencing technologies. *Int. J. Mol. Sci.* **2019**, *21*, 298. [[CrossRef](#)] [[PubMed](#)]
- Delahaye, C.; Nicolas, J. Sequencing DNA with nanopores: Troubles and biases. *PLoS ONE* **2021**, *16*, e0257521. [[CrossRef](#)] [[PubMed](#)]
- Kim, M.; Oh, H.S.; Park, S.C.; Chun, J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int. J. Syst. Evol. Microbiol.* **2014**, *64*, 346–351. [[CrossRef](#)]
- Yoon, S.H.; Ha, S.M.; Kwon, S.; Lim, J.; Kim, Y.; Seo, H.; Chun, J. Introducing EzBioCloud: A taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int. J. Syst. Evol. Microbiol.* **2017**, *67*, 1613. [[CrossRef](#)]
- Oxford Nanopore Technologies. Oxford Nanopore Broadens Its Q20+ Early Access Programme and Releases New Nanopore, R10.4, into This Early Access Programme. 2021. Available online: <https://nanoporetech.com/about-us/news/oxford-nanopore-broadens-its-q20-early-access-programme-and-releases-new-nanopore> (accessed on 23 September 2022).
- Shin, J.; Lee, S.; Go, M.J.; Lee, S.Y.; Kim, S.C.; Lee, C.H.; Cho, B.K. Analysis of the mouse gut microbiome using full-length 16S rRNA amplicon sequencing. *Sci. Rep.* **2016**, *6*, 29681. [[CrossRef](#)]
- Acharya, K.; Khanal, S.; Pantha, K.; Amatya, N.; Davenport, R.J.; Werner, D. A comparative assessment of conventional and molecular methods, including MinION nanopore sequencing, for surveying water quality. *Sci. Rep.* **2019**, *9*, 15726.

22. Fujiyoshi, S.; Muto-Fujita, A.; Maruyama, F. Evaluation of PCR conditions for characterizing bacterial communities with full-length 16S rRNA genes using a portable nanopore sequencer. *Sci. Rep.* **2020**, *10*, 12580.
23. Heikema, A.P.; Horst-Kreft, D.; Boers, S.A.; Jansen, R.; Hiltmann, S.D.; de Koning, W.; Kraaij, R.; de Ridder, M.A.; van Houten, C.B.; Bont, L.J.; et al. Comparison of Illumina versus nanopore 16S rRNA gene sequencing of the human nasal microbiota. *Genes* **2020**, *11*, 1105. [[CrossRef](#)] [[PubMed](#)]
24. Nygaard, A.B.; Tunsjø, H.S.; Meisal, R.; Charnock, C. A preliminary study on the potential of Nanopore MinION and Illumina MiSeq 16S rRNA gene sequencing to characterize building-dust microbiomes. *Sci. Rep.* **2020**, *10*, 3209. [[CrossRef](#)] [[PubMed](#)]
25. Wei, P.L.; Hung, C.S.; Kao, Y.W.; Lin, Y.C.; Lee, C.Y.; Chang, T.H.; Shia, B.C.; Lin, J.C. Characterization of fecal microbiota with clinical specimen using long-read and short-read sequencing platform. *Int. J. Mol. Sci.* **2020**, *21*, 7110. [[CrossRef](#)] [[PubMed](#)]
26. Kinoshita, Y.; Niwa, H.; Uchida-Fujii, E.; Nukada, T. Establishment and assessment of an amplicon sequencing method targeting the 16S-ITS-23S rRNA operon for analysis of the equine gut microbiome. *Sci. Rep.* **2021**, *11*, 11884. [[CrossRef](#)] [[PubMed](#)]
27. Low, L.; Fuentes-Utrilla, P.; Hodson, J.; O'Neil, J.D.; Rossiter, A.E.; Begum, G.; Suleiman, K.; Murray, P.I.; Wallace, G.R.; Loman, N.J.; et al. Evaluation of full-length nanopore 16S sequencing for detection of pathogens in microbial keratitis. *PeerJ* **2021**, *9*, e10778. [[CrossRef](#)]
28. Matsuo, Y.; Komiya, S.; Yasumizu, Y.; Yasuoka, Y.; Mizushima, K.; Takagi, T.; Kryukov, K.; Fukuda, A.; Morimoto, Y.; Naito, Y.; et al. Full-length 16S rRNA gene amplicon analysis of human gut microbiota using MinION™ nanopore sequencing confers species-level resolution. *BMC Microbiol.* **2021**, *21*, 35. [[CrossRef](#)]
29. Oberle, A.; Urban, L.; Falch-Leis, S.; Ennemoser, C.; Nagai, Y.; Ashikawa, K.; Ulm, P.A.; Hengstschläger, M.; Feichtinger, M. 16S rRNA long-read nanopore sequencing is feasible and reliable for endometrial microbiome analysis. *Reprod. Biomed. Online* **2021**, *42*, 1097–1107.
30. Park, C.; Kim, S.B.; Choi, S.H.; Kim, S. Comparison of 16S rRNA gene based microbial profiling using five next-generation sequencers and various primers. *Front. Microbiol.* **2021**, *12*, 715500.
31. De Siqueira, G.M.V.; Pereira-dos-Santos, F.M.; Silva-Rocha, R.; Guazzaroni, M.E. Nanopore sequencing provides rapid and reliable insight into microbial profiles of intensive care units. *Front. Public Health* **2021**, *9*, 710985. [[CrossRef](#)] [[PubMed](#)]
32. Rozas, M.; Brillet, F.; Callewaert, C.; Paetzold, B. MinION™ Nanopore Sequencing of Skin Microbiome 16S and 16S-23S rRNA Gene Amplicons. *Front. Cell. Infect. Microbiol.* **2022**, *11*, 1317. [[CrossRef](#)] [[PubMed](#)]
33. Stevens, B.; Creed, T.; Reardon, C.; Manter, D. Comparison of Oxford Nanopore Technologies and Illumina MiSeq sequencing with mock communities and agricultural soil. *Res. Sq.* **2022**. [[CrossRef](#)]
34. The EzBioCloud Team. MTP-Primary Analysis User Guide. 2019. Available online: <https://help.ezbiocloud.net/mtp-pipeline/#mtp-basic-concept> (accessed on 23 September 2022).
35. Rognes, T.; Flouri, T.; Nichols, B.; Quince, C.; Mahé, F. VSEARCH: A versatile open source tool for metagenomics. *PeerJ* **2016**, *4*, e2584. [[CrossRef](#)] [[PubMed](#)]
36. Edgar, R.C.; Haas, B.J.; Clemente, J.C.; Quince, C.; Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **2011**, *27*, 2194–2200.
37. Edgar, R.C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **2010**, *26*, 2460–2461. [[CrossRef](#)]
38. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **2011**, *17*, 10–12. [[CrossRef](#)]
39. Callahan, B.J.; McMurdie, P.J.; Rosen, M.J.; Han, A.W.; Johnson, A.J., A.; Holmes, S.P. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **2016**, *13*, 581–583. [[CrossRef](#)] [[PubMed](#)]
40. Bokulich, N.A.; Kaehler, B.D.; Rideout, J.R.; Dillon, M.; Bolyen, E.; Knight, R.; Huttley, G.A.; Gregory Caporaso, J. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* **2018**, *6*, 90. [[CrossRef](#)]
41. Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glöckner, F.O. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **2012**, *41*, D590–D596.
42. Schloss, P.D.; Westcott, S.L.; Ryabin, T.; Hall, J.R.; Hartmann, M.; Hollister, E.B.; Lesniewski, R.A.; Oakley, B.B.; Parks, D.H.; Robinson, C.J.; et al. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **2009**, *75*, 7537–7541. [[CrossRef](#)] [[PubMed](#)]
43. Kozich, J.J.; Westcott, S.L.; Baxter, N.T.; Highlander, S.K.; Schloss, P.D. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* **2013**, *79*, 5112–5120. [[CrossRef](#)] [[PubMed](#)]
44. Cole, J.R.; Wang, Q.; Fish, J.A.; Chai, B.; McGarrell, D.M.; Sun, Y.; Brown, C.T.; Porras-Alfaro, A.; Kuske, C.R.; Tiedje, J.M. Ribosomal Database Project: Data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* **2014**, *42*, D633–D642. [[CrossRef](#)] [[PubMed](#)]
45. Oksanen, J.; Simpson, G.; Blanchet, F.; Kindt, R.; Legendre, P.; Minchin, P.; O'Hara, R.; Solymos, P.; Stevens, M.; Szoecs, E.; et al. R Package Version 2.6-2. 2022. Available online: <https://CRAN.R-project.org/package=vegan> (accessed on 20 September 2022).
46. Russel, J. Russel88/MicEco: v0.9.15 (v0.9.15), Zenodo 2021. Available online: <https://doi.org/10.5281/zenodo.4733747> (accessed on 10 October 2022).
47. Kerkhof, L.J. Is Oxford Nanopore sequencing ready for analyzing complex microbiomes? *FEMS Microbiol. Ecol.* **2021**, *97*, fiab001. [[CrossRef](#)] [[PubMed](#)]

48. DeSantis, T.Z.; Hugenholtz, P.; Larsen, N.; Rojas, M.; Brodie, E.L.; Keller, K.; Huber, T.; Dalevi, D.; Hu, P.; Andersen, G.L. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **2006**, *72*, 5069–5072. [[CrossRef](#)] [[PubMed](#)]
49. Park, S.C.; Won, S. Evaluation of 16S rRNA databases for taxonomic assignments using a mock community. *Genom. Inform.* **2018**, *16*, e24. [[CrossRef](#)]
50. MTP Taxonomy for Genus *Veillonella*. Available online: <https://www.ezbiocloud.net/mtp/taxonomy?db=PKSSU4.0&tn=Veillonella&depth=2&rg=v4> (accessed on 23 September 2022).
51. Caporaso, J.G.; Lauber, C.L.; Walters, W.A.; Berg-Lyons, D.; Lozupone, C.A.; Turnbaugh, P.J.; Fierer, N.; Knight, R. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 4516–4522. [[CrossRef](#)] [[PubMed](#)]
52. Auer, L.; Mariadassou, M.; O'Donohue, M.; Klopp, C.; Hernandez-Raquet, G. Analysis of large 16S rRNA Illumina data sets: Impact of singleton read filtering on microbial community description. *Mol. Ecol. Resour.* **2017**, *17*, e122–e132. [[CrossRef](#)]
53. Edgar, R.C. UNOISE2: Improved error-correction for Illumina 16S and ITS amplicon sequencing. *BioRxiv* **2016**. [[CrossRef](#)]
54. Lynch, M.D.; Neufeld, J.D. Ecology and exploration of the rare biosphere. *Nat. Rev. Microbiol.* **2015**, *13*, 217–229. [[CrossRef](#)]
55. Chao, A. Nonparametric estimation of the number of classes in a population. *Scand. J. Stat.* **1984**, *11*, 265–270.
56. Chao, A.; Lee, S.M. Estimating the number of classes via sample coverage. *J. Am. Stat. Assoc.* **1992**, *87*, 210–217. [[CrossRef](#)]
57. Willis, A.; Bunge, J. Estimating diversity via frequency ratios. *Biometrics* **2015**, *71*, 1042–1049. [[CrossRef](#)] [[PubMed](#)]
58. Klindworth, A.; Pruesse, E.; Schweer, T.; Peplies, J.; Quast, C.; Horn, M.; Glöckner, F.O. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* **2013**, *41*, e1. [[CrossRef](#)]
59. Sze, M.A.; Schloss, P.D. The impact of DNA polymerase and number of rounds of amplification in PCR on 16S rRNA gene sequence data. *mSphere* **2019**, *4*, e00163-19. [[CrossRef](#)]
60. Calus, S.T.; Ijaz, U.Z.; Pinto, A.J. NanoAmpli-Seq: A workflow for amplicon sequencing for mixed microbial communities on the nanopore sequencing platform. *Gigascience* **2018**, *7*, giy140. [[CrossRef](#)] [[PubMed](#)]
61. Karst, S.M.; Dueholm, M.S.; McIlroy, S.J.; Kirkegaard, R.H.; Nielsen, P.H.; Albertsen, M. Retrieval of a million high-quality, full-length microbial 16S and 18S rRNA gene sequences without primer bias. *Nat. Biotechnol.* **2018**, *36*, 190–195. [[CrossRef](#)]
62. Curry, K.D.; Wang, Q.; Nute, M.G.; Tyshaieva, A.; Reeves, E.; Soriano, S.; Wu, Q.; Graeber, E.; Finzer, P.; Mendling, W.; et al. Emu: Species-level microbial community profiling for full-length Nanopore 16S reads. *Nat. Methods* **2022**, *19*, 845–853. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.