

# Abundance tracking by long-read nanopore sequencing of complex microbial communities in samples from 20 different biogas/wastewater plants

Christian Brandt (✉ [christian.jena@gmail.com](mailto:christian.jena@gmail.com))

Swedish University of Agricultural Science, Department Molecular Science <https://orcid.org/0000-0002-7199-3957>

Erik Bongcam-Rudloff

Swedish University of Agricultural Science, Dept. Animal Breeding and Genetics

Bettina Müller

Swedish University of Agricultural Science, Department Molecular Science

---

## Methodology

**Keywords:** Nanopore, Sequencing, DNA extraction, Metagenome, Abundance, GTDB

**Posted Date:** November 25th, 2019

**DOI:** <https://doi.org/10.21203/rs.2.17734/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.  
[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Applied Sciences on October 26th, 2020.  
See the published version at <https://doi.org/10.3390/app10217518>.

# Abstract

Anaerobic digestion (AD) has long been critical technology for green energy, but the majority of the microorganisms involved are unknown and not cultivable, which makes abundance tracking difficult. Developments in nanopore sequencing make it a promising approach for monitoring microbial communities via metagenomic sequencing. For reliable monitoring of AD via long reads, a robust protocol for obtaining less fragmented, high-quality DNA, while preserving bacterial composition, was established. Samples from 20 different biogas/waste-water reactors were investigated and a median of 20 Gb sequencing data per flow cell were retrieved for each reactor. Using the GTDB index allowed sufficient characterisation of abundance of bacteria and archaea in biogas reactors. A dramatic improvement (1.8- to 13-fold increase) in taxonomic classification was achieved using the GTDB-based index compared with the RefSeq index. Ongoing efforts in GTDB to achieve more phylogenetically coherent taxonomic species definitions, including meta-assembled genomes, give a clear advantage over conventional classification databases such as RefSeq. Unlike conventional 16S rRNA studies, metagenomic read classification allows abundance of the unknown microbial fraction to be monitored.

## Background

Anaerobic digestion (AD) has been a key technology in Europe and worldwide for many decades and is vital for full transition from a fossil fuel-based economy to a sustainable bio-economy. During AD, various metabolic and functional bottlenecks that affect process stability and efficiency can occur. Process efficiency depends on the composition and activity of the microbial community, and therefore much effort has been devoted to understanding how the microbial community correlates with process parameters and performance [1, 2]. In AD, many microorganisms form complex consortia to link and combine metabolic activities, and each member has different requirements for nutrients and physical conditions. However, the majority of the microorganisms involved are unknown [3], because isolation and characterisation of these microorganisms is time-consuming and many are not cultivable. Moreover, the metabolism of pure cultures often differs from that of a microbial consortium. Thus, the function of many AD microorganisms and their importance for conversion of organic material have not yet been fully explored.

Meta-barcoding via 16S rRNA gene amplicon sequencing is widely used for analysis of the composition, structure and dynamics of microbial assemblages [4, 5]. However, this approach has two main drawbacks: 1) The experimental set-up does not allow prompt monitoring; and 2) inherent PCR amplification bias (e.g. primer mismatches, different gene copy number, sequence- and primer-dependent polymerase efficiency, choice of hypervariable region) [6–9] reduces the accuracy, as read abundance does not correlate with species abundance [10, 11]. An alternative approach to circumvent the latter is to sequence the entire microbial DNA (metagenomics). Classifying the reads obtained from metagenomic sequencing gives insights into bacterial diversity and absolute abundance. Metagenomics is usually performed via high-throughput sequencing (e.g. Illumina) [12]. However, long-read sequencing (e.g. nanopore sequencing) may provide a more accurate alternative, since the ratio of correctly classified

reads to all reads is as high or even higher with nanopore reads as with Illumina [13]. Although long reads are less accurate, they may allow better overall taxonomic classification due to the higher information content per read [14, 15]. This is particularly useful because taxonomic read classifiers use the lowest common ancestor (LCA) method. Another crucial advantage is that the actual absolute abundance of an organism is preserved during sequencing and reflected in the sequencing yield [16].

However, long-read sequencing of samples with such high microbial complexity, as found in AD systems, has two bottlenecks: i) It requires less-fragmented, high-quality DNA while preserving bacterial composition and ii) choice of index database has a significant influence on read classification performance. For pure cultures or filtered bacteria (e.g. water samples), several isolation protocols based on chemical cell lysis have been developed. For complex microbiota, as found in biogas reactors, these protocols are usually not applicable, because chemical cell lysis is not equally effective for all bacteria due to very diverse cell wall structures. Mechanical approaches such as bead beating allow more uniform cell disruption, especially of Gram-positive bacteria [17], which are very abundant in AD processes [5, 18]. However, the high shearing forces significantly reduce the length of the DNA fragments. Besides, AD reactor samples may contain various particles/sediments, inhibitory components, cell debris and partly digested DNA that further reduce the final quality and length of the recovered DNA, and can significantly affect the sequencing performance.

Here we present a DNA isolation protocol adjusted to the properties of digestate from AD processes that results in high-quality DNA suitable for nanopore sequencing. Since operational parameters such as feedstock and temperature strongly affect microbial composition and physicochemical properties of the digestate, we tested the protocol on digestate samples retrieved from 20 different methanogenic AD processes. We compared DNA quantity and sequence performance and analysed the taxonomic affiliation and abundance in all 20 reactors directly via read classification using different databases.

## Results

### Digestate samples and DNA isolation

We isolated DNA from digestates retrieved from 17 biogas plants, two wastewater treatment plants (WWTP) and one laboratory-scale reactor, all located in either Sweden (SW) or Germany (GER). These AD processes were chosen as they use common feedstocks/substrates such as organic household waste, slaughterhouse waste, manure, sewage sludge and agricultural products (Table 1). Thirteen operate under mesophilic temperature (37-42°C), three under hyper-mesophilic temperature (44°C) and four under thermophilic temperature (48-52°C) (Table 1). The digestate samples, hereafter referred to simply as samples, were taken directly from the methanogenic reactors without any further storage. All German and one Swedish sludge sample(s) were pre-frozen for transportation.

The DNA isolation protocol described in this work was extensively pre-tested and optimised on sample 01-SW before being applied to the other 19 reactor samples. Sample 01-SW was from a laboratory-scale reactor mimicking the operation of a large-scale biogas plant. The protocol was optimised for the largest possible DNA fragment length and high DNA purity, for optimal nanopore activity and sequencing yield. The full protocol, with each step described in detail, is available online under at protocols.io [19]. The protocol includes all necessary and optimised steps in order to improve DNA yield and length, with brief explanations. The most significant improvements in overall DNA yield and length are shown in Figure 1.

In brief, we first compared the reactor sample with controlled-growth culture samples using the DNA isolation kit from the FastDNA Spin Kit for Soil (MP Biomedicals) according to the manufacturer. As expected, we observed more digested DNA in the sludge samples, with significantly lower DNA length to yield ratio than in the cultures (Figure 1A). To improve overall DNA yield and fragment length, we modified the washing steps and the total amount of sludge sample used. Increasing the amount of sludge used slightly increased the total DNA yield. Performing an additional humic acid (HA) washing step improved the yield and reduced the number of smaller DNA fragments. Both steps combined yielded the best overall results (Figure 1B). More additional washing steps with the default buffer did not improve the results further (data not shown).

The reactor samples contained large amounts of degraded DNA, due to the high biological activity in AD. These small fragments had a negative impact on the overall sequencing performance and the total sequencing yield. Centrifugation of the samples and replacing the supernatant with distilled water significantly reduced the amount of degraded DNA (Figure 1C). DNase treatment before isolation resulted in reduced smear, but in one case also reduced the overall yield during DNA isolation (sample lane 4 in Figure 1C). Thus, DNase treatment seems unsuitable for achieving a stable and reproducible protocol in this case.

Another influencing factor was the bead beating settings. We achieved the best results using a force of 6 m/s once for 20 s with a FastPrep -24™ Classic Instrument (MP Biomedicals™) (Figure 1D). We strongly recommend re-evaluating the force if another device is used for performing beat beating. To remove small DNA fragments (<1000 bp) produced during bead beating, the DNA sample was cleaned up using magnetic beads (0.35:1 ratio of beads to sample). This step also removed other impurities that may remain after DNA isolation.

Figure 1: Effect on DNA quality of different steps during isolation. White boxes indicate the chosen approach for the DNA isolation protocol. A) Default manufacturer's protocol applied to (left) sludge and (right) controlled growth cultures. B) DNA yield/length after introduction of a humic acid removal wash (HA-wash) step. Arrows indicate use of 0.4 mL sludge instead of 0.2 mL. C) Pre-preparation of sludge and effect on DNA yield/length. (supern.: supernatant removed and replaced with water; DNase: sample pretreated with DNase; control: no pre-preparation; both: supernatant replaced with water and incubated with DNase). D) Impact of different bead beating settings on DNA yield/fragment length.

# Sequencing yield and quality of metagenomic DNA

The optimised DNA isolation protocol [19] was applied to all 20 digestate samples (Table 1), using a MinION device (Oxford Nanopore Technologies) with a single flow cell for each DNA sample. We initially used the LSK-108 Kit, but changed later to the LSK-109 Kit as we observed better sequencing performance. Figure 2 summarises the sequencing results and the quality scores obtained.

Figure 2: Summary of all 20 sequence runs of biogas reactor samples using the MinION device. Different colours (red/blue) indicate properties that influenced quality parameters.

We observed a slight decrease in N50 and median read length for pre-frozen samples compared with fresh samples. Sequencing runs via LSK-108 Kits were also performed using flow cells stored for two months or longer, which is likely the main reason for the decrease in median Q-score. Total sequencing yield varied between 18.2 and 28.5 Gb for most samples and, across all samples, we achieved a median of 20.5 Gb. Total sequencing yield was below 10 Gb for only two of the 20 samples (03-SW and 04-SW). Thus the DNA isolation protocol presented here yields sufficient DNA of appropriate quality for nanopore sequencing and an acceptable number of long reads in all cases tested.

## Abundance estimation and taxonomic read classification

Classifying as many reads as possible is essential to achieve the best characterisation and abundance estimates for the microbial community. To this end, we chose centrifuging for the taxonomic classification, because it allows for smaller kmer size to consider the higher error profile of long reads [20]. Moreover, the availability of pre-built databases and less RAM consumption in comparison with other classifiers made it a favourable choice. We also applied three different taxonomic indices: the default NCBI RefSeq supplied by the centrifuge manufacturer (<https://ccb.jhu.edu/software/centrifuge/manual.shtml>); the GTDB index (GTDB version r89, de-replicated to 54k genomes) from Méric *et al.* [21]; and an GTDB index based on 24,706 bacterial and archaeal representative species from Parks *et al.* [22]. We used Illumina reads (paired-end 150 bp) for each reactor to compare classification capability. Figure 3 gives an overview of the proportion of reads classified for each taxonomic index at phylum, genus and species level.

Figure 3: Proportion of taxonomically classified reads of each reactor sample at phylum, genus and species level. The four columns for each reactor represent different read classification approaches. The first bar represents Illumina reads using the GTDB-Méric index, for direct comparison against nanopore sequencing.

In all cases, we achieved a dramatic increase (range 1.8- to 13-fold) in the proportion of classified reads when using the GTDB index rather than the RefSeq index. In particular, improvements were obtained for samples 05-GER, 04-SW, 04-GER and 03-GER (11.5- to 13.2-fold increase in classified reads). We did not observe any noticeable decrease in classified reads when comparing phylum to genus or species level. The average proportion of classified reads was 8.8% of all reads on genus level using the RefSeq database and 42.11% using the 'GTDB-Méric' index. These results correspond to a median 4.8-fold increase in classified reads across all samples. For the reactors with a 10-fold or higher increase, more meaningful observation of the actual bacterial/archaeal composition was possible. Based on these results, all samples were subjected to further abundance analysis based on the 'GTDB-Méric' index, with abundance calculated based on the proportion of classified taxonomic reads in all classified reads. The results obtained at phylum level are shown in Figure 4 and the results at genus level in Supplementary Figure S1.

Figure 4: Summary of abundance for samples from all 20 reactors, calculated based on all classified reads. Only phyla with at least 0.1% abundance are shown (circles). The size and colour of each circle correspond to the abundance of each phylum in each reactor. Taxonomic names are based on GTDB version r89 (<https://gtdb.ecogenomic.org>), which includes "placeholder" names like Firmicutes\_A. DTU030 corresponds to the NCBI phylum Firmicutes. UBP3 corresponds to "bacteria" classification via NCBI. The reactor samples are described in detail in Table 1.

In general, we identified a wide range of different phyla, especially for Firmicutes, due to the more phylogenetically coherent taxonomic definitions by GTDB. This increased taxonomic granularity at phylum level provided better resolution, especially for Firmicutes, thus improving tracking of changes. As known for biogas processes, Bacteroidota (NCBI: Bacteroidetes) and Firmicutes\_G and A (NCBI: Firmicutes) were the dominant phyla in most samples [23]. Thermotogota, a phylum frequently found in systems operating at higher temperatures [23], was highly abundant in three of the four thermophilic reactors (02-SW, 09-SW and 08-GER). Euryarchaeota, including methanogens, represented less than 10% of the total classified reads, as frequently described previously for biogas processes [23].

Samples 07-GER and 09-GER represented reactors operated in parallel by the same facility using the same operational parameters and feedstock. Both reactors had very similar microbial composition and structure on both phylum and genus level, but with slight abundance variations (e.g. some very low-abundance organisms were not common to both reactors). Some small abundance variations also occurred at genus level (Supplementary Figure S1).

## Discussion

The overarching goal of this study was to develop a DNA purification protocol for evaluating whether classified reads obtained by a single nanopore sequencing run, as an alternative to 16S rRNA studies,

sufficiently characterise microbial community abundance in AD reactors. We evaluated and adjusted all the steps necessary, from DNA extraction to abundance estimation.

## DNA isolation and sequencing

The protocol devised was able to produce a sufficient amount and purity of DNA for nanopore sequencing. The intended use of the protocol is to estimate the microbial abundance of any given AD process, enabling conclusions or correlations between microorganisms and reactor stability and process efficiency to be identified. Therefore, consistent DNA isolation and retrieval were favoured, which is usually achieved by mechanical shearing. The presence of matter, sediments, solids and other particles in samples from AD reactors may affect the overall shearing force during bead beating. The resulting variation in total DNA fragment length may impact total yield during sequencing. Using a chemical approach for “matter absent” cultures could be more favourable if the aim is to obtain longer DNA fragments for a particular species of interest.

We observed that sequencing runs were heavily negatively impacted if many very small reads were present and sequenced. They had a negative impact on library preparation and the subsequent sequencing run. We therefore introduced a magnetic bead-based purification step directly after DNA isolation, to remove these very small DNA fragments. Due to the already small size of DNA fragments in general (read median ~2500 bp), other solutions to remove short DNA fragments, such as the Short Read Eliminator Kit (Circulomics) or Fire Monkey Kit (RevoluGen), were not considered.

## Database choice and abundance estimation

We applied the metagenomic approach to samples from 20 different reactors, to estimate the classification performance for different AD types. In general, the phylum-level estimates obtained in this work are in line with findings in 16S rRNA studies and MAG studies pointing to Firmicutes and Bacteroidetes as the main phyla for most AD processes [5, 18, 23]. They are also in line with Campanaro *et al.*, who identified 790 Firmicutes MAGs among 1600 public available MAGs from ADs [3]. Sun *et al.* identified mainly Bacteroidetes (17-70%), followed by Firmicutes (5-10%), in five samples from WWTP [24]. The main drawback of 16S rRNA studies is the lack of information on how much of the population might be missing after untargeted PCR amplification. Therefore, it is difficult to compare the performance of 16S rRNA with that of a metagenomic approach outside standard microbial community samples.

While we observed varying performance during sequencing, we were able to use all runs to classify from 25% to 66% of all reads using the GTDB database from Méric *et al.* [21]. The RefSeq index, on the other hand, did not yield a sufficient amount of taxonomic classification for most samples and is therefore not suitable for abundance tracking in AD. The additional taxonomic definitions of GTDB also improved the overall resolution, particularly for Firmicutes. In general, indices based on large, phylogenetically coherent taxonomic species definitions significantly increase the number of classified reads [21]. Publicly available

indices are becoming increasingly important for metagenomic research. De-replication of such an enormous number of genomes is time-consuming and difficult to compute without access to computation clusters or cloud computing.

Eukaryotic cells from e.g. fungi and plants are not included in GTDB and possibly represent part of the unclassified number of reads. Inclusion of eukaryotic cells, including anaerobic fungi [25], in taxonomic read classification remains challenging, mainly due to the large genome size and the resulting index size. DNA remains of cells (e.g. plants) could be expected within AD reactor samples, due to the organic substrate used. Moreover, phage DNA or plasmid DNA would not be classified in most cases [26]. We suspect that there are also other currently unknown bacteria or archaea hidden within the unclassified reads.

Some limitations of the study must be noted. We probably observed false-positive hits for closely related species while validating the centrifuge parameters for taxonomic read classification on datasets of ZymoClean mock communities (GridION data from <https://lomanlab.github.io/mockcommunity/>). False-positive hits with less than 0.1% abundance probably contributed to higher read error rate. A plausible explanation is that higher error rates introduce erroneous kmers that are more similar to another species of the same genus. In any case, we included a taxonomic level only if the total amount of taxonomic classifications for a given taxonomic level amounted to more than 0.1% of all classified reads, in order to mitigate false-positive results.

## Conclusions

Nanopore sequencing using the pocket-sized MinION device provides relatively cheap and readily available access to sequencing. Typical issues like raw-read error rates have been improved in recent years. Our results demonstrated that “shallow sequenced” metagenomics is possible by maximising extraction of taxonomic information via more suitable indices that allow abundance estimations of AD. Sequencing at lower depth thus enables much cheaper comparison of multiple metagenomic samples. Metagenomes with many uncultivable microorganisms are especially appropriate for long-read sequencing technology, because these reads can achieve better completeness of a given genome. This facilitates recovery of complete microbial genomes, which is of high interest if the metagenome remains mainly unclassified when using indices such as GTDB. Applying other indices or methods (developed in the future) to identify eukaryotes would enable abundance tracking closer to the reality of any given microbial community. Further improvements in raw read quality in nanopore sequencing would give more reliable species identification at lower abundances and could facilitate reactor monitoring with even less data.

The method requires a minimum of technical equipment and technician training, and enables community monitoring across different time frames or process parameter changes. It may also reveal whether the number of unclassified reads increases or decreases under certain conditions, indicating whether the



currently classified population size is increasing or decreasing. In the case of a decreasing classified population, another currently unclassified, but important, organism might be present.

The method provides a promising approach for monitoring biogas processes, because it can reveal actual and absolute changes in the microbial community in response to operational parameters and feedstock. Use of automated and biogas-reactor specific monitoring tools, e.g. via machine learning, could help to predict upcoming inefficiency and disturbances.

## Methods

The practical procedure of metagenomic DNA sequencing is covered by two protocols (DNA Extraction and Library Preparation): the “Long-read DNA preparation for Metagenomic samples” protocol described and developed here [19] and the “One-pot Library preparation” published by Josh Quick [27] for the LSK-108 Kit, or the nanopore protocol for the LSK-109 Kit, with some slight modifications. We increased the incubation times for the DNA repair and end preparation with the LSK-109 Kit to 20 min each, to improve the overall sequencing performance.

## DNA extraction

We developed a DNA isolation protocol for complex metagenomic samples for biogas or wastewater treatment plants that is a modified and extended version of the FastDNA Spin Kit for Soil (MP Biomedicals). This protocol, called “Long-read DNA preparation for Metagenomic samples”, is stored in detail as a step-by-step protocol for better experimental reproducibility at protocols.io [19].

DNA isolation was performed and tested on fresh or frozen substrate/sludge samples. DNA extraction was performed according to the manufacturer's protocol, with various exceptions and additions to improve DNA fragment length while maintaining high yield and high enough purity to avoid interferences with the nanopores inside each sequencing flow cell.

In brief, 400 µL sample (substrate or sludge) were initially centrifuged for 5 min at 20,000 g. The supernatant was removed and the sludge/substrate was re-suspended in nuclease-free water and then added to Lysing Matrix E (FastDNA Spin Kit for Soil). This step was included to remove short DNA fragment remains, e.g. from dead cells. The samples were homogenised in a MP Biomedicals FastPrep-24 Classic Instrument (Fischer Scientific) at 20 s and 6 m/s and immediately placed on ice. Increased shearing time (e.g. 2 x 20s) led to decreased DNA fragment length, while not visibly improving the total yield on agarose gel. Decreased shearing force (e.g. 4 or 5 m/s) led to decreased yield, with no or minor fragment length improvements (decreased yield could also indicate incomplete DNA extraction). DNA yield was judged based on electrophoresis results, to favour high yield in longer DNA fragment ranges.

Before proceeding with protein precipitation, we added an additional 5-min RNase incubation step. Furthermore, we tested two different washing solutions for the washing steps: “HA-wash solution” (see

[19]) and SEWS-M (FastDNA Spin Kit for Soil). Introducing the additional HA-wash solution step improved the overall DNA fragment length and total yield. More washing steps significantly reduced the total yield. After DNA retrieval, a subsequent “pre-cleaning” step was added using magnetic beads (either AMPure or Highprep beads) to further reduce short fragments and possible impurities from the biogas or wastewater samples before library preparation.

Samples were temporarily stored at 4 °C before library preparation (freezing was omitted as it causes DNA shearing).

## Library preparation

Library preparation with the LSK-108 Kit was performed according to the protocol [27]. This a modified version of the standard nanopore library preparation by ligation for the LSK-108 Kit, which served here as an additional reduction in DNA shearing in comparison with the default protocol for library preparation. Library preparation with the SQK-LSK109 Kit was carried out according to the ‘Genomic DNA by Ligation (SQK-LSK109)’ protocol from [community.nanoporetech.com](https://community.nanoporetech.com). We increased the incubation times for ‘DNA repair’ and ‘end prep’ to 20 min each.

We used 1-2 µg of input DNA for library preparation, assuming an average DNA length of 5,000 bp for the DNA fragments after bead beating. Higher DNA amounts led to a decrease in pore activity and a decrease in total yield.

## Sequencing

All samples were sequenced using a MinION Sequencer for 72 hours or until no sequencing activity was observed, using either an R.4.9.1 or R.4.9 flow cell (FLO-MIN106) for each sample. We used the MinKNOW software with active channel selection enabled and basecalling deactivated. We performed a ‘flow cell-refuel’ step after approx. 18-20 hours of runtime by adding 75 µL of a 1:1 water-SQB-Buffer mixture to the flow cell via the SpotON port. SQB-Buffer is part of the Oxford-Nanopore SQK-LSK109 Kit. Illumina sequencing was performed on a MiSeq Sequencer using exactly the same DNA material as was used for nanopore sequencing.

## Bioinformatic tools

Basecalling was performed using the GPU accelerated guppy basecaller with the high accuracy model and adapter trimming (available at <https://nanoporetech.com>). Read quality was analysed using nanoplots (<https://github.com/wdecoster/NanoPlot>). Taxonomic classification of each read was performed using centrifuge [20] with the NCBI RefSeq index from (<https://ccb.jhu.edu/software/centrifuge/>), the GTDB index from [https://monash.figshare.com/articles/GTDB\\_r89\\_54k/8956970](https://monash.figshare.com/articles/GTDB_r89_54k/8956970) [21] and an index from 24,706 bacterial

and archaeal representative species [22]. The centrifuge output was filtered by only including read classifications that met the centrifuge score of at least 250 and length of at least 150 bp. The results were converted via “centrifuge-kreport” and plotted via R using ggballoonplot (ggpubr) to compare abundance across biogas reactors.

## Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

The DNA isolation protocol is accessible via [dx.doi.org/10.17504/protocols.io.5feg3je](https://dx.doi.org/10.17504/protocols.io.5feg3je)

All the nanopore and Illumina reads are available on ENA (European Nucleotide Archive) under project number PRJEB34573.

Competing interest

None to declare.

Funding

This study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) BR 5692/1-1 and BR 5692/1-2. The material is based upon work supported by Google Cloud.

Author's contribution

CB – wrote, designed, and executed the practical and in silico experiments, drafted the manuscript and prepared the diagrams.

EB - critical reading and correction regarding the bioinformatic approach.

BM – critical reading and correction regarding the biogas reactors.

Acknowledgment

We are grateful to Michael Lebuhn, Anna Schnürer and Sabine Kleinsteuber for providing us with the reactor samples for this work.

# References

1. Campanaro S, Treu L, Kougias PG, Luo G, Angelidaki I. Metagenomic binning reveals the functional roles of core abundant microorganisms in twelve full-scale biogas plants. *Water Res.* 2018;140:123–34.
2. Ziganshin AM, Ziganshina EE, Kleinstaub S, Nikolausz M. Comparative Analysis of Methanogenic Communities in Different Laboratory-Scale Anaerobic Digesters. *Archaea Vanc BC.* 2016;2016:3401272.
3. Campanaro S, Treu L, Rodriguez-R LM, Kovalovszki A, Ziels RM, Maus I, et al. The anaerobic digestion microbiome: a collection of 1600 metagenome-assembled genomes shows high species diversity related to methane production. *bioRxiv.* 2019;;680553.
4. Krause L, Diaz NN, Edwards RA, Gartemann K-H, Krömeke H, Neuweiger H, et al. Taxonomic composition and gene content of a methane-producing microbial community isolated from a biogas reactor. *J Biotechnol.* 2008;136:91–101.
5. Stolze Y, Bremges A, Rumming M, Henke C, Maus I, Pühler A, et al. Identification and genome reconstruction of abundant distinct taxa in microbiomes from one thermophilic and three mesophilic production-scale biogas plants. *Biotechnol Biofuels.* 2016;9:156.
6. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* 2013;41:e1.
7. Suzuki MT, Giovannoni SJ. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Appl Environ Microbiol.* 1996;62:625–30.
8. Acinas SG, Sarma-Rupavtarm R, Klepac-Ceraj V, Polz MF. PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Appl Environ Microbiol.* 2005;71:8966–9.
9. Teng F, Darveekaran Nair SS, Zhu P, Li S, Huang S, Li X, et al. Impact of DNA extraction method and targeted 16S-rRNA hypervariable region on oral microbiota profiling. *Sci Rep.* 2018;8. doi:10.1038/s41598-018-34294-x.
10. Louca S, Doebeli M, Parfrey LW. Correcting for 16S rRNA gene copy numbers in microbiome surveys remains an unsolved problem. *Microbiome.* 2018;6. doi:10.1186/s40168-018-0420-9.
11. Krehenwinkel H, Wolf M, Lim JY, Rominger AJ, Simison WB, Gillespie RG. Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. *Sci Rep.* 2017;7:17668.
12. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. *Nat Biotechnol.* 2017;35:833–44.
13. Pearman WS, Freed NE, Silander OK. The advantages and disadvantages of short- and long-read metagenomics to infer bacterial and eukaryotic community composition. *bioRxiv.* 2019;;650788.
14. Sanderson ND, Street TL, Foster D, Swann J, Atkins BL, Brent AJ, et al. Real-time analysis of nanopore-based metagenomic sequencing from infected orthopaedic devices. *BMC Genomics.*

2018;19:714.

15. Somerville V, Lutz S, Schmid M, Frei D, Moser A, Irmeler S, et al. Long read-based de novo assembly of low complex metagenome samples results in finished genomes and reveals insights into strain diversity and an active phage system. *bioRxiv*. 2018;:476747.
16. Nicholls SM, Quick JC, Tang S, Loman NJ. Ultra-deep, long-read nanopore sequencing of mock microbial community standards. *GigaScience*. 2019;8.
17. de Boer R, Peters R, Gierveld S, Schuurman T, Kooistra-Smid M, Savelkoul P. Improved detection of microbial DNA after bead-beating before DNA isolation. *J Microbiol Methods*. 2010;80:209–11.
18. Güllert S, Fischer MA, Turaev D, Noebauer B, Ilmberger N, Wemheuer B, et al. Deep metagenome and metatranscriptome analyses of microbial communities affiliated with an industrial biogas fermenter, a cow rumen, and elephant feces reveal major differences in carbohydrate hydrolysis strategies. *Biotechnol Biofuels*. 2016;9:121.
19. Brandt C. Long-read DNA preparation for metagenomic samples. 2019. doi:10.17504/protocols.io.5feg3je.
20. Kim D, Song L, Breitwieser FP, Salzberg SL. Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Res*. 2016;26:1721–9.
21. Méric G, Wick RR, Watts SC, Holt KE, Inouye M. Correcting index databases improves metagenomic studies. *bioRxiv*. 2019;:712166.
22. Parks DH, Chuvochina M, Chaumeil P-A, Rinke C, Mussig AJ, Hugenholtz P. Selection of representative genomes for 24,706 bacterial and archaeal species clusters provide a complete genome-based taxonomy. *bioRxiv*. 2019;:771964.
23. Westerholm M, Schnürer A. Microbial Responses to Different Operating Practices for Biogas Production Systems. *Anaerob Dig*. 2019. doi:10.5772/intechopen.82815.
24. Sun L, Liu T, Müller B, Schnürer A. The microbial community structure in industrial biogas plants influences the degradation rate of straw and cellulose in batch tests. *Biotechnol Biofuels*. 2016;9:128.
25. Kazda M, Langer S, Bengelsdorf FR. Fungi open new possibilities for anaerobic fermentation of organic residues. *Energy Sustain Soc*. 2014;4:6.
26. Stalder T, Press MO, Sullivan S, Liachko I, Top EM. Linking the resistome and plasmidome to the microbiome. *ISME J*. 2019;13:2437–46.
27. Quick J. One-pot ligation protocol for Oxford Nanopore libraries. 2018. doi:10.17504/protocols.io.k9acz2e.

## Tables

Table 1: Main substrate and temperature conditions in the 20 anaerobic digestion processes analysed in this work

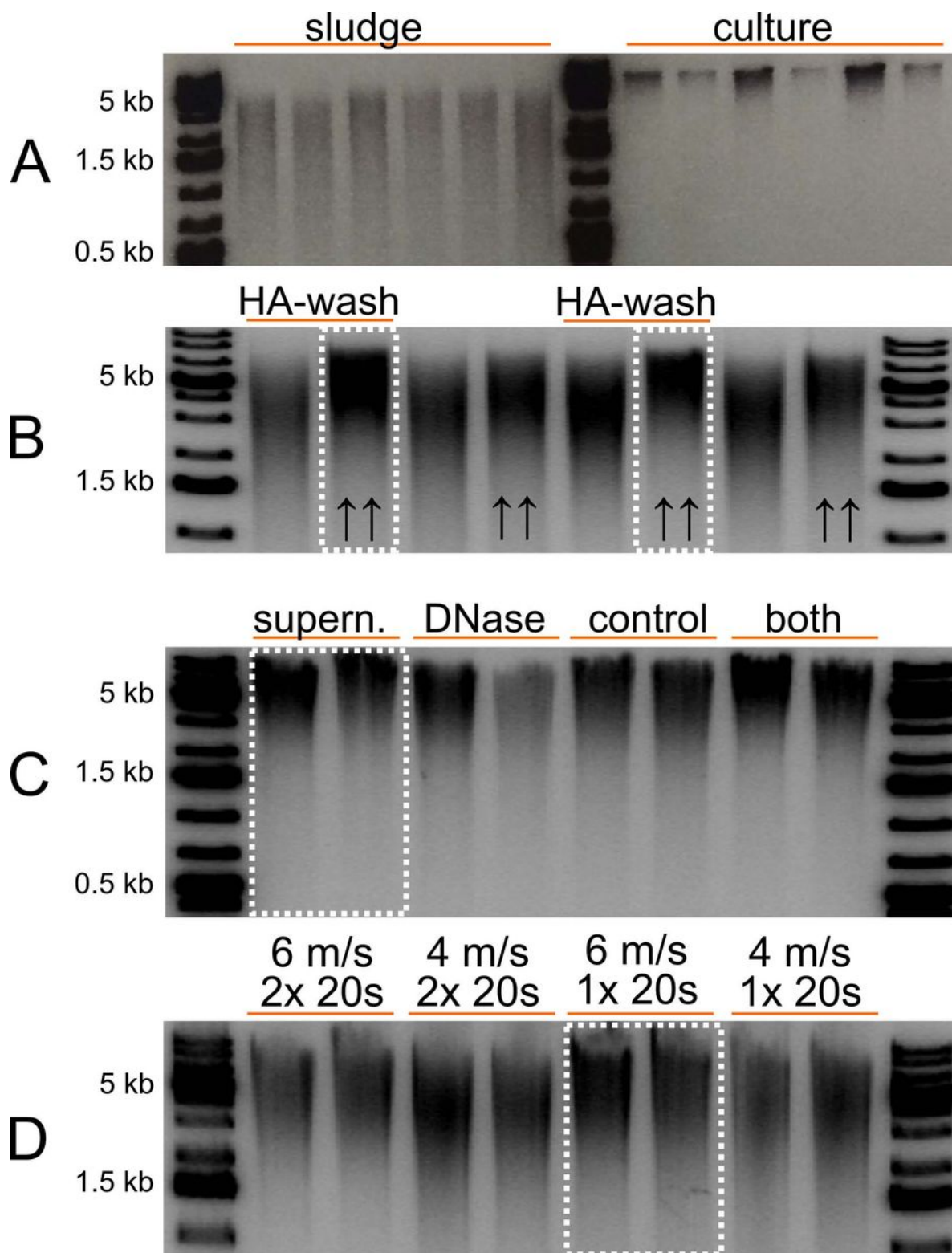
	Sample ID	Substrate					Operation	Status
		<i>Organic household waste</i>	<i>Slaughter-house waste</i>	<i>Manure</i>	<i>Sewage sludge</i>	<i>Green<sup>+</sup>-based</i>		
Sweden	01-SW <sup>#</sup>	o					Mesophilic	Fresh
	02-SW	o	o				Thermophilic	Fresh
	03-SW				o		Mesophilic	Fresh
	04-SW			o			Mesophilic	Fresh
	05-SW					o	Mesophilic	Frozen
	06-SW					o	Mesophilic	Fresh
	07-SW	o	o				Mesophilic	Fresh
	08-SW				o		Mesophilic	Fresh
	09-SW	o	o				Thermophilic	Fresh
	10-SW			o			Mesophilic	Fresh
Germany	01-GER					o	Mesophilic	Frozen
	02-GER			o		o	Hyper-mesophilic	Frozen
	03-GER			o		o	Hyper-mesophilic	Frozen
	04-GER			o		o	Thermophilic	Frozen
	05-GER			o		o	Hyper-mesophilic	Frozen
	06-GER			o		o	Mesophilic	Frozen
	07-GER*			o		o	Mesophilic	Frozen
	08-GER	o				o	Thermophilic	Frozen
	09-GER*			o		o	Mesophilic	Frozen
	10-GER			o		o	Mesophilic	Frozen

<sup>#</sup>Laboratory-scale reactor.

<sup>+</sup>Mainly crop residues, grass and/or silages.

\*Digestate samples retrieved from reactors operated in parallel at the same biogas plant

## Figures



**Figure 1**

Effect on DNA quality of different steps during isolation. White boxes indicate the chosen approach for the DNA isolation protocol. A) Default manufacturer's protocol applied to (left) sludge and (right) controlled growth cultures. B) DNA yield/length after introduction of a humic acid removal wash (HA-

wash) step. Arrows indicate use of 0.4 mL sludge instead of 0.2 mL. C) Pre-preparation of sludge and effect on DNA yield/length. (supern.: supernatant removed and replaced with water; DNase: sample pretreated with DNase; control: no pre-preparation; both: supernatant replaced with water and incubated with DNase). D) Impact of different bead beating settings on DNA yield/fragment length.

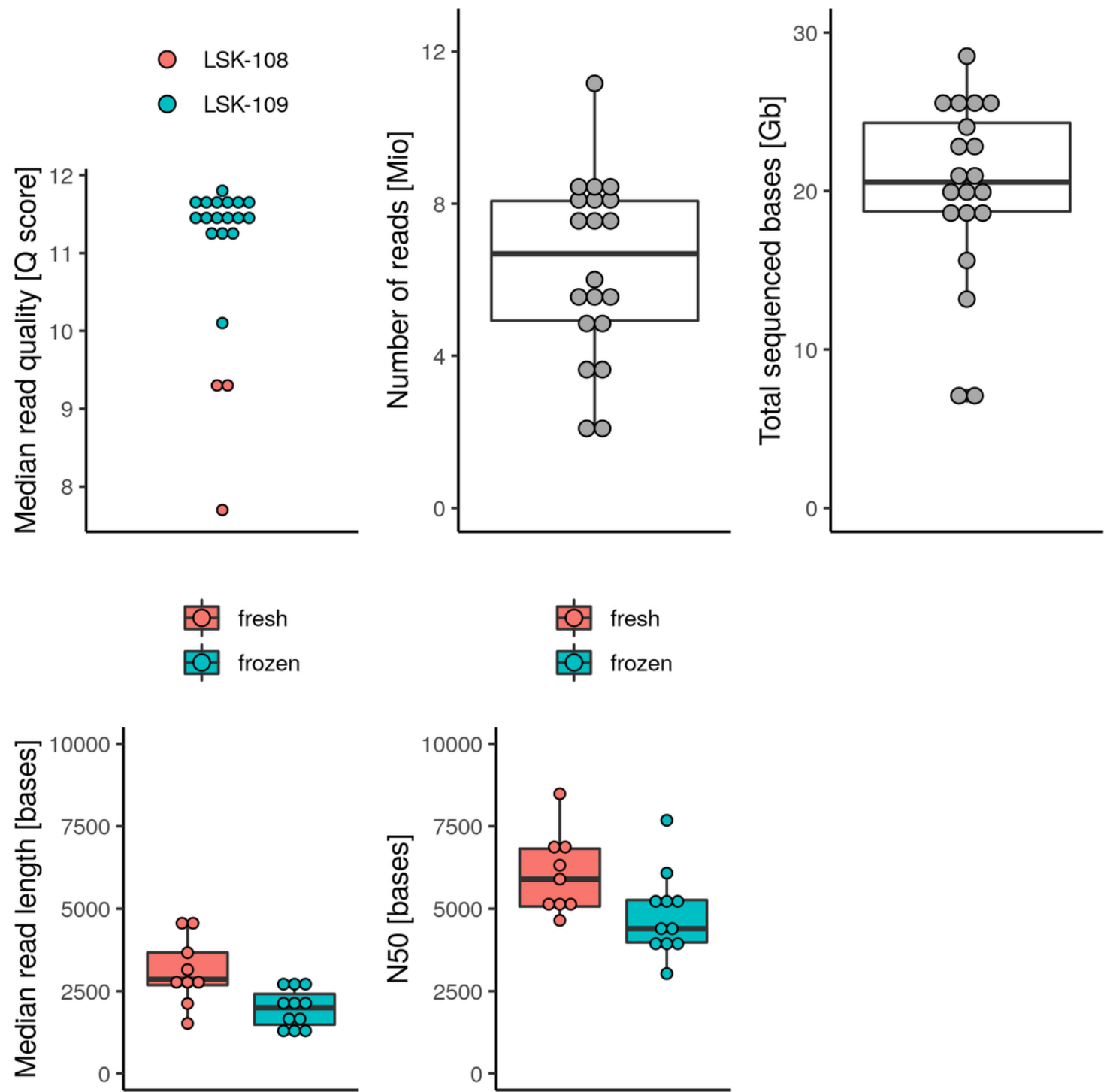
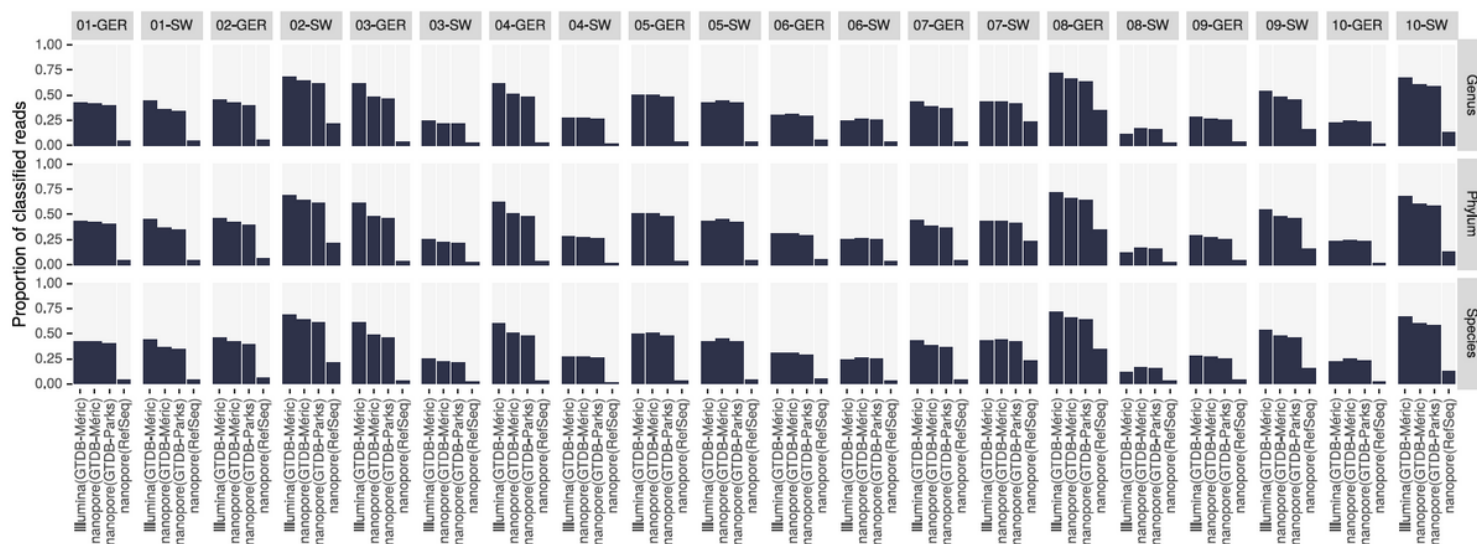


Figure 2

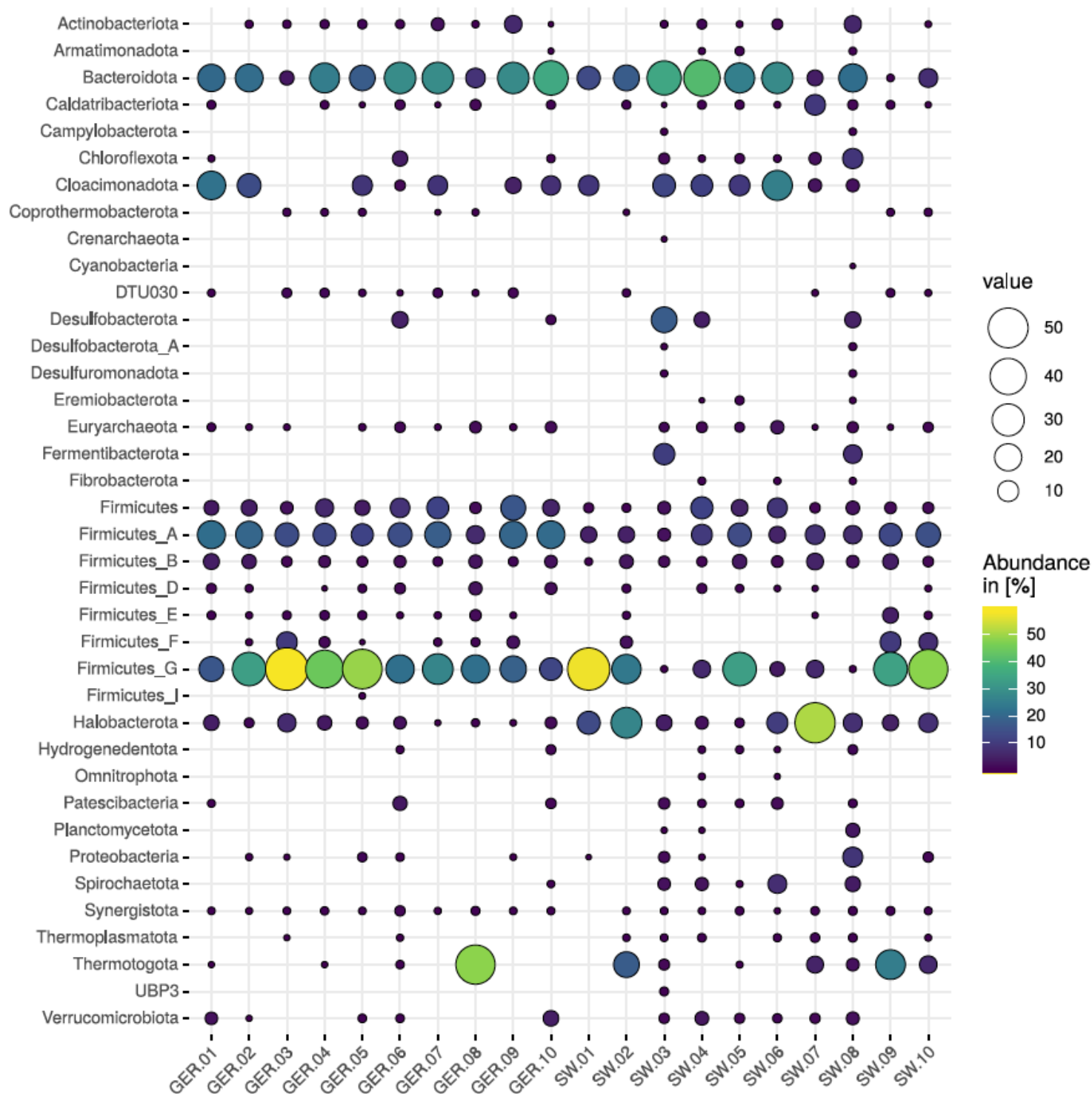
Summary of all 20 sequence runs of biogas reactor samples using the MinION device. Different colours (red/blue) indicate properties that influenced quality parameters.





**Figure 3**

Proportion of taxonomically classified reads of each reactor sample at phylum, genus and species level. The four columns for each reactor represent different read classification approaches. The first bar represents Illumina reads using the GTDB-Mérid index, for direct comparison against nanopore sequencing.



**Figure 4**

Summary of abundance for samples from all 20 reactors, calculated based on all classified reads. Only phyla with at least 0.1% abundance are shown (circles). The size and colour of each circle correspond to the abundance of each phylum in each reactor. Taxonomic names are based on GTDB version r89 (<https://gtdb.ecogenomic.org>), which includes “placeholder” names like Firmicutes\_A. DTU030 corresponds to the NCBI phylum Firmicutes. UBP3 corresponds to “bacteria” classification via NCBI. The reactor samples are described in detail in Table 1.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplfig1genus.pdf](#)