

ARTICLE OPEN



Microbiome-based environmental monitoring of a dairy processing facility highlights the challenges associated with low microbial-load samples

Aoife J. McHugh^{1,2}, Min Yap^{1,2}, Fiona Crispie^{1,3}, Conor Feehily^{1,3}, Colin Hill^{1,3} and Paul D. Cotter^{1,3}✉

Efficient and accurate identification of microorganisms throughout the food chain can potentially allow the identification of sources of contamination and the timely implementation of control measures. High throughput DNA sequencing represents a potential means through which microbial monitoring can be enhanced. While Illumina sequencing platforms are most typically used, newer portable platforms, such as the Oxford Nanopore Technologies (ONT) MinION, offer the potential for rapid analysis of food chain microbiomes. Initial assessment of the ability of rapid MinION-based sequencing to identify microbes within a simple mock metagenomic mixture is performed. Subsequently, we compare the performance of both ONT and Illumina sequencing for environmental monitoring of an active food processing facility. Overall, ONT MinION sequencing provides accurate classification to species level, comparable to Illumina-derived outputs. However, while the MinION-based approach provides a means of easy library preparations and portability, the high concentrations of DNA needed is a limiting factor.

npj Science of Food (2021)5:4; <https://doi.org/10.1038/s41538-021-00087-2>

INTRODUCTION

Dairy processing environments harbour microorganisms that have the potential to contaminate food before and during processing^{1–5}. Some of these microorganisms have the potential to cause spoilage or be pathogenic^{6–9}. Routine environmental monitoring is carried out in food processing environments for this reason, and usually involves the use of swabbing and agar plating to determine total numbers of general (e.g., total bacteria count) or specific (generally potentially spoilage-associated or pathogenic species) categories of microorganisms⁷. These analyses frequently involve selective and phenotype-based agar assays¹⁰, some of which can yield high false positive numbers^{10,11}. These approaches are further limited by the fact that they do not provide information about non-targeted species or indeed the microbial population as a whole.

DNA sequencing methods have recently been applied to dairy and environmental samples to determine the microbial population composition and enable source tracking^{7,12–14}. High throughput metagenomic sequencing can provide greater insights into the taxonomic composition of populations present in these environments than culture based methods. Specifically it uncovers information relating to the functional potential of species and strains present, including virulence and spoilage properties. Despite these benefits, high throughput metagenomic sequencing approaches typically require expensive reagents and platforms as well as personnel skilled in molecular biology, data generation and interpretation. These requirements limit their routine implementation in manufacturing facilities. Some of these issues have the potential to be addressed through use of portable DNA sequencing devices such as the Oxford Nanopore Technologies (ONT) MinION. The MinION's portability and workflows are designed to facilitate their use by less experienced personnel and could allow easier detection and identification of the causative agents of microbial contamination. Such approaches have recently

been tested in a clinical setting to identify causative agents of disease from metagenomic samples¹⁵, including studies where the results were compared with those generated through Illumina sequencing^{16,17} or culture-based analysis¹⁸. This approach has yet to be applied to food processing settings for environmental monitoring.

As a proof-of-concept, we conducted a study to determine the ability of MinION-based rapid sequencing to correctly classify a simple, four-strain, mock community of highly related spore-forming microorganisms of relevance to the dairy processing chain. Prompted by this initial analysis, we proceeded to compare the outputs of MinION-based rapid sequencing to Illumina-based, and culture-based methods to characterise the microbiota of environmental swabs collected from a food processing facility. Overall, MinION-based approaches were comparable to the Illumina sequencing equivalent in terms of species level taxonomic classification. However, the requirement of high concentration and quality input DNA for the routine implementation of MinION sequencing was a limitation due to the environment tested. To overcome this, random amplification of template DNA was required. Regardless, the potential benefits of the routine application of metagenomic sequencing to food processing environments were clear.

RESULTS

MinION sequencing accurately identified species in mock community

Metagenomic DNA representing a simple mock community of 4 related (Supplementary Figure 1) dairy processing-associated, spore-forming contaminants, i.e., *Bacillus cereus*, *Bacillus thuringiensis*, *Bacillus licheniformis*, *Geobacillus stearothermophilus*, was sequenced using ONT MinION rapid sequencing kits. This proof-of-concept exercise was performed to determine the extent to which

¹Food Bioscience Department, Teagasc Food Research Centre, Cork, Ireland. ²School of Microbiology, University College Cork, Cork, Ireland. ³APC Microbiome Ireland, Cork, Ireland. ✉email: Paul.Cotter@teagasc.ie

MinION-based sequencing could identify, and discriminate between, related, and in some cases difficult to distinguish, microorganisms found in dairy processing environments. Full length amplicon 16S rRNA gene-based sequencing of the simple mock metagenomic DNA using the ONT 16S barcoding kit SQK-RAB204 resulted in 996,441 reads following rebasecalling by albacore. These reads contained a total of 1,454,835,092 bases with an average read length of 1460 bp and a median read length of 1561 bp. 16S rRNA reads aligned by BLASTn to the Silva 16S database (version 132) with MEGAN 6 classification resulted in successful identification of three out of the 4 species. The fourth strain, *G. stearothermophilus* DSM 458, was correctly identified to the genus level only (Fig. 1a).

Rapid whole metagenome sequencing (WMGS) of the mock community using the SQK-RAD004 kit resulted in 97,503 reads following rebasecalling by albacore and adaptor removal. These

97,503 reads contained a total of 750,359,905 bases with an average read length of 7696 bases and a median of 5762 bases. LAST alignment against the nr database followed by MEGAN long read (LR) lowest common ancestor (LCA) analysis resulted in 74.76% bases being classified to some taxonomic level. Of these, 42.63% were classified to species level, 46.28% classified to species group level and 8.15% classified to genus level, accounting for 97.06% of classified reads. 64.37% of bases classified to genus level only were attributed to *Geobacillus*, with the remaining 35.63% classified as *Bacillus* (Fig. 1a). Of the sequences classified to the species level, 57.26% of bases were attributed to *Bacillus thuringiensis*, 14.74% were attributed to *Bacillus licheniformis*, 13.98% were attributed to *Bacillus cereus*, 13.8% were attributed to *Geobacillus stearothermophilus* and, 0.21% misassigned as *Bacillus paralicheniformis* (Fig. 1a). De novo assembly of raw reads from the rapid sequencing reads using the canu (version 1.7) assembler¹⁹

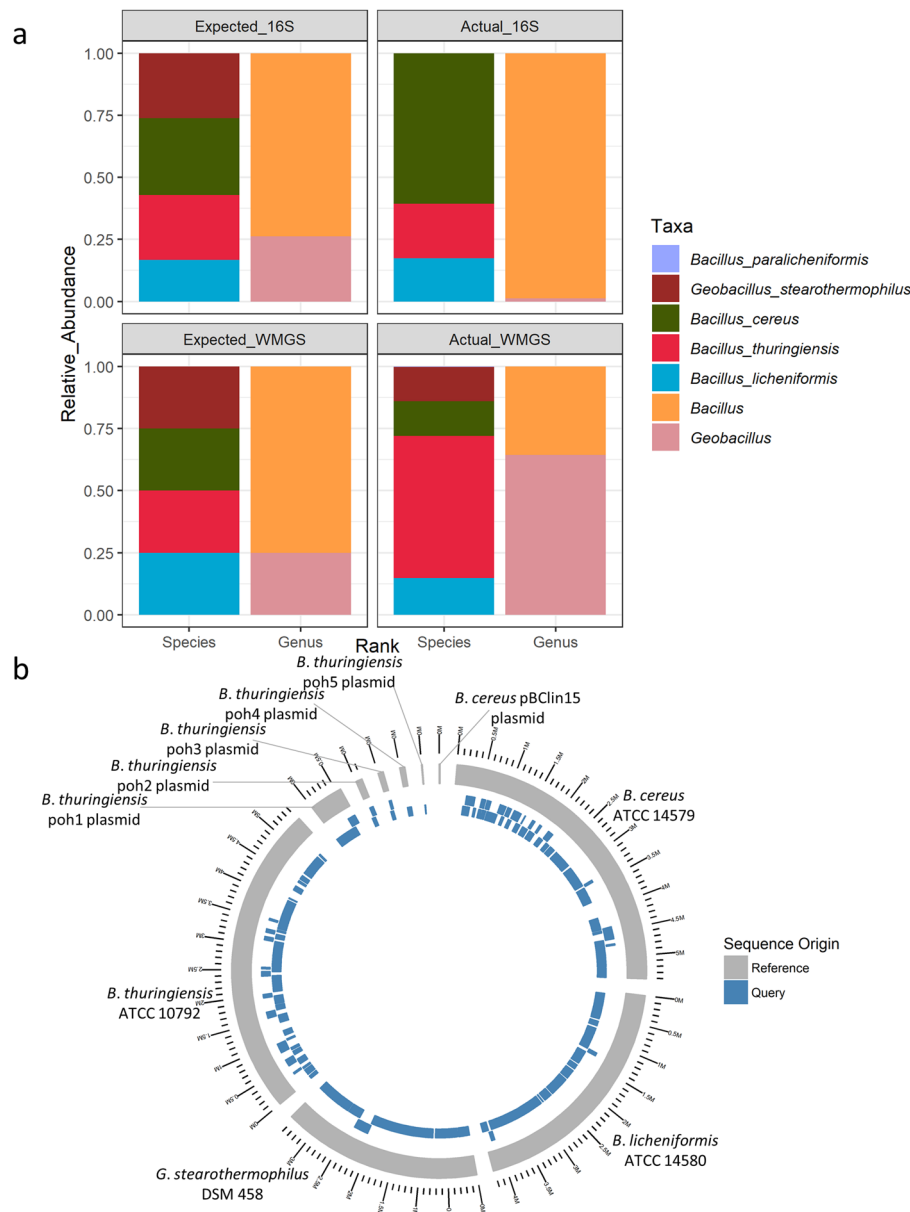


Fig. 1 Mock community analysis. MinION sequencing followed by MEGAN taxonomic classification of a simple mock community. **a** Taxonomic classification following 16S and whole metagenome (WMGS) sequencing. Also shown are expected relative abundances. **b** De novo assembly of genomes by the canu assembler, followed by mapping back to original known genomes, to illustrate coverage at 97% identity. 4 genomes, with 6 plasmids illustrated, of which 4 genomes and 5 plasmids had sequences aligned at 97% identity.

resulted in 104 contigs and mapping back of reads to references resulted in good coverage up to 97% identity (Fig. 1b). The 4 reference strains genomes included 6 plasmids, corresponding to 10 contiguous stretches of DNA. Nine of these 10 contigs were identified following sequence assembly, the exception being pBClin15, a 15 kb plasmid from *B. cereus* (Fig. 1b). A total of 99.59% of the assembled bases aligned to the reference genomes and, of the reference genomes, 98.27% aligned to the assembled MinION sequences (Supplementary Table 1).

MinION and NextSeq sequencing provided comparable classifications

Prompted by the successful use of MinION-based sequencing to characterise the mock metagenomic community DNA, the technology was applied to study the microbiota of a food processing facility and to compare outputs with those derived through NextSeq (Illumina)-based sequencing. Eight locations in a single processing facility were swabbed on three different days across October, November, and December 2018, each after cleaning in place (CIP) but before the next round of dairy processing (Fig. 2). These eight locations comprised a table, door, wall, gaskets/flow plate seals, external surface of dryer balance tank, internal surface of dryer balance tank, external surface of evaporator, and drain beside evaporator. These swabs were prepared for sequencing, along with a series of negative controls

and a positive control, consisting of the simple mock metagenomic community used previously. For MinION sequencing, rapid sequencing of multiple displacement amplification (MDA)-generated template DNA from 36 samples, used to address the relatively high quantities of DNA required for library preparation, was carried out using the SQK-RBK004 rapid barcoding sequencing kit. After processing, a total of 899,306 reads were generated, containing a total of 1,648,724,928 bases with an average read length of 1,833 bases and median of 926 bases per read (and an average of 45,797,915 bases and 24,980.7 reads per sample). LAST alignment against the nr database followed by MEGAN long read (LR) lowest common ancestor (LCA) analysis resulted in 62% of bases being classified to some taxonomic level. Of these, 29.11% were classified to species level and 38.36% classified to genus level, accounting for 67.47% of classified reads. A total of 59 species were detected at > 5% relative abundance in at least one sample by MEGAN (Supplementary Figure 2).

Other shotgun sequencing-based approaches were employed to study the microbiomes of these environmental samples for comparative purposes. These included Illumina-based sequencing of MDA and non-MDA DNA, as well as of metagenomic DNA extracted from easily culturable microorganisms to allow a comparison with the species that grow when traditional culturing-based approaches are employed. This Illumina (NextSeq)-based sequencing of 93 samples produced 734,909,370 reads

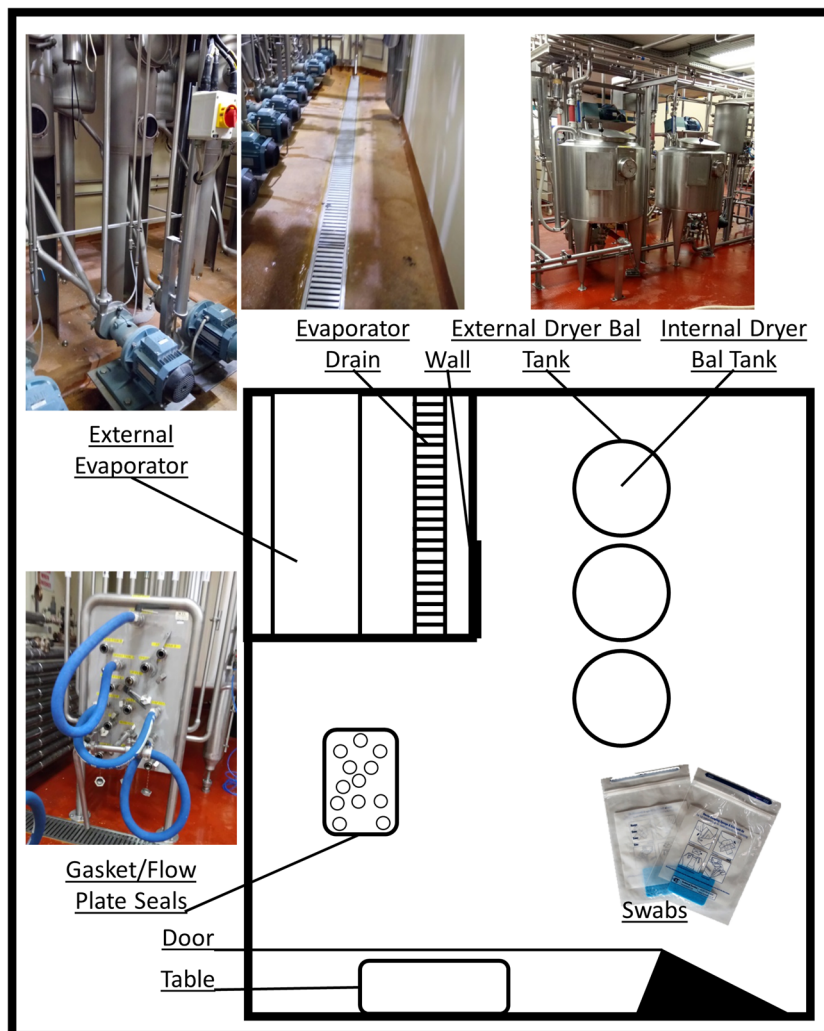


Fig. 2 Schematic of dairy processing facility sampling areas. Dairy processing facility schematic includes the 8 areas sampled in each of October, November, and December 2018. Areas were sampled post CIP and prior to the recommencement of processing.

containing 150 bases each with an average of 7,902,251 reads per sample. To allow a comparison with MinION outputs, and to avoid discrepancies through use of different bioinformatic pipelines, Diamond alignment against the nr database followed by MEGAN 6 lowest common ancestor (LCA) analysis was employed and resulted in 78% reads being classified to some taxonomic level. Of these, 10.8% were classified to the species level and 39.6% classified to the genus level, accounting for 50.3% of classified bases. In comparison, Kraken2 and Bracken classification resulted in 61% reads classified to some taxonomic level, with 99% of those classified being classified to species level. This approach did not correctly classify the composition of the mock community (positive control) (Supplementary Figure 3). Similarly, MetaPhlan2 did not correctly classify all of the species of the mock community (Supplementary Figure 4), with both classifiers incorrectly classifying at least one species. Interestingly, both classifiers misclassified different species, whereby Bracken misclassified *B. licheniformis* as a *Bacillus* phage, and MetaPhlan2 did not differentiate between *B. cereus* and *B. thuringiensis*. Additionally, MetaPhlan2 only classified the *G. stearothermophilus* to genus level.

Using the MEGAN classification, which correctly classified the simple mock community, 108 species were identified at > 5% relative abundance in at least one sample from all MinION and NextSeq sequenced samples (Fig. 3). Species level classification by MEGAN revealed consistencies between corresponding NextSeq- and MinION-sequenced samples (Fig. 3). Overall, reads corresponding to *Kocuria sp.* WRN011 were detected at the highest relative abundance. This taxon was detected in multiple locations, at each time-point, in both the MinION, and corresponding NextSeq, MDA-generated samples. Its relative abundance was highest in the evaporator drain samples at each time point. *Kocuria sp.* ZOR0020 was present in high relative abundance in external dryer balance tank swabs in both MinION- and NextSeq-MDA sequenced MDA samples (Fig. 3). Other dominant species included *Acinetobacter johnsonii* in gasket/flow plate seals (MinION and Illumina), *Micrococcus luteus* in evaporator drain (MinION and Illumina sequenced samples), *Enterococcus faecium* from the inside of the dryer balance tank as well as many other October and November samples (MinION and MDA amplified Illumina sequencing), *Klebsiella pneumonia* in many December samples regardless of sequencing approach and *Enterococcus casseliflavus* in many samples from October and November (high relative abundance in MinION sequenced samples and at lower abundance in the corresponding MDA Illumina sequenced samples) (Fig. 3). *Exiguobacterium sibiricum* was also detected in high relative abundance in MinION sequenced October and November door samples. It was also at lower relative abundances in many other October and November samples and in the corresponding Illumina sequenced door samples.

There were some notable sequencing platform-dependent differences. *Exiguobacterium sp.* S3.2 and *Pseudochrobactrum sp.* B5 were present at higher relative abundance in October and November MDA Illumina NextSeq sequences compared to MinION sequences and *Enterobacter sp.* HK169 was detected in December MinION samples, but not corresponding Illumina samples (Fig. 3). Species level taxonomic identification was performed on negative controls also. Many species were specific to negative controls, including *Kribbia dieselivorans* and *Cytophagales bacterium* B6, detected at a high relative abundance in MinION sequenced MDA negative controls, and *Paenibacillus fonticola*, detected at high relative abundance in both MinION and Illumina sequenced MDA negative controls. There was also a high relative abundance of *Escherichia coli* in MDA negative controls, with *Salmonella enterica* in the December samples, in both MinION sequences and corresponding Illumina sequences. *Ralsonia insidiosa* was also seen above 0.2% exclusively in negative controls. However, there was some overlap with species identified in negative controls also identified in environmental samples. In particular, the swab

negative control for both MDA MinION and MDA NextSeq from each month are similar to results generated from swabbing of the internal of the dryer balance tank, which are the environmental samples with the lowest environmental load (Supplementary Table 2). *Kocuria sp.*, *Acinetobacter johnsonii*, *Enterococcus casseliflavus*, *Klebsiella pneumoniae*, *Exiguobacterium sibiricum*, *Enterococcus casseliflavus*, *Pseudochrobactrum sp.* B5, *Enterobacter sp.* HK169, and *Raoultella planticola* are all seen in negative controls (Fig. 3). These findings highlight the risks of relying on data from samples with a low microbial load and the importance of including negative controls.

Metagenome-assembled genomes (MAGs) were extracted from assemblies of combined Illumina MDA and MinION MDA sequences. This resulted in 162 bins, of which 10 were high quality at > 80% complete and < 10% contamination (Table 1). In total 7 of the 10 MAGs were from environmental isolates, with 3 out of 10 being the positive control species used. From the remaining MAGs, 3 out of 7 environmental isolates could not be definitively assigned at the species level, being assigned as each of a number of species at similar levels of relative abundances. These MAGs were assigned at the genus level as *Planococcus*, *Exiguobacterium* and *Kocuria* and were sourced from the October evaporator drain, gasket/flow plate seal and external dryer balance tank, respectively. The MAGs that were assigned at the species level were an *Enterococcus casseliflavus* from the October table swab sample, a *Paracoccus chinensis* from the November evaporator drain, a *Macrocooccus caseolyticus* from the November gasket/flow plate seal and a *Nesterenkonia massiliensis* from the November external of dryer balance tank sample (Table 1).

MDA amplification introduced bias towards some species

In order to determine the potential for bias arising from MDA pre-processing, outputs from MDA-generated NextSeq sequencing were compared to non-MDA derived NextSeq (NPP). Higher relative abundances of *Pseudochrobactrum sp.* B5 and *Pseudochrobactrum sp.* AO18b were seen in October and November NPP samples compared to the MDA-amplified equivalents (Fig. 3). Overall, the NPP samples we found to be less diverse than their MDA counterparts (Fig. 4a).

Culture-based analyses introduced a selection bias

In order to determine to what extent culture-dependent and -independent approaches provided different outputs, a comparison between NPP NextSeq-generated sequences and those resulting from sequencing of pools of easily cultured colonies (Plate samples) was performed. Sequences generated from Plate samples were noted to be significantly less diverse (Fig. 4a), however many of the Plate samples clustered with the non-cultured samples when beta-diversity was analysed (Fig. 4b). A number of the species detected were similar to species identified in the corresponding non-cultured samples (NPP and MDA amplified). Overall, *Kocuria sp.* WRN011, was detected in all samples in which it had previously been identified through culture-independent approaches. *Enterococcus faecium*, the species found at highest relative abundance in all internal dryer balance tank samples from November (i.e., MDA MinION, NextSeq MDA, NPP and Plate; Fig. 3) was also detected. Pre-culturing enriched some species that had been identified at low relative abundance in metagenomic NPP and MDA samples. These included *Planococcus massiliensis* (October door sample), *Microbacterium oxydans* (November Table sample), *Acinetobacter baumannii* (November external dryer balance tank) and *Lysinibacillus sp.* B2A1 (December internal dryer balance tank; Fig. 3).

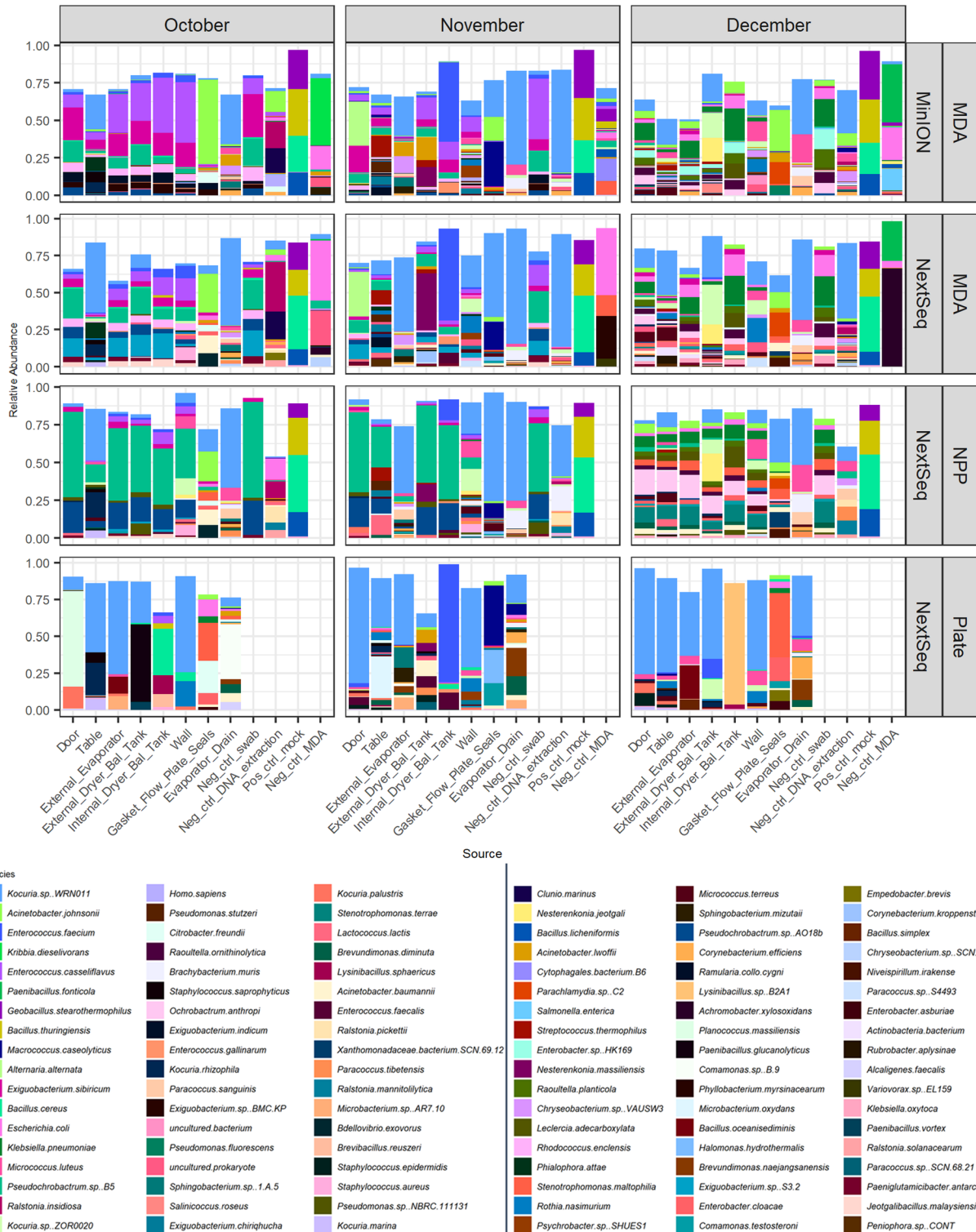


Fig. 3 Species level classification of MinION and NextSeq sequenced environmental samples. Taxonomic assignment of MinION and NextSeq sequenced samples generated following the use of different pre-processing and sequencing methods. Pre-processing methods include MDA amplification, no pre-processing (NPP), and spread plating on BHI before washing colonies, pelleting, and treating as a metagenomic sample (Plate). Species level classification was performed using LAST (for MinION) and Diamond (for NextSeq) alignment of reads against the NR database and classification with MEGAN (LR for MinION). Species present in at least 5% in at least one sample are shown.

Genus level classification highlighted further selection bias

As some genera could not be distinguished at species level, genus level assignments were also investigated and compared. MEGAN LCA analysis identified sequences that could not be more

accurately classified to species level, and assigned these as far as genus level only. A combined 56 genera were identified between MinION, NextSeq (both at > 5% relative abundance) and Sanger sequencing. Fifteen of these 56 genera were identified in

Table 1. High quality MAGs.

Month	Sample	Bin	Kaiju assignment	Megan assignment	Percent Complete	Percent Contamination
October	Positive control (mock)	2	<i>Geobacillus stearothermophilus</i>	<i>Geobacillus stearothermophilus</i>	91.37	1.1235
October	Evaporator Drain	5	<i>Planococcus plaktortidis</i> / <i>Planococcus maitriensis</i> / <i>Planococcus maritimus</i>	<i>Planococcus plaktortidis</i> / <i>Planococcus maritimus</i> / <i>Planococcus rifietoensis</i>	82.41	0.6622
October	Gasket/Flow Plate Seal	5	<i>Exiguobacterium acetyllicum</i> / <i>Exiguobacterium</i> sp. RIT341/ <i>Exiguobacterium indicum</i>	<i>Exiguobacterium indicum</i> / <i>Exiguobacterium acetyllicum</i>	81.9	0.6578
October	Table	14	<i>Enterococcus casseliflavus</i>	<i>Enterococcus casseliflavus</i>	82.77	1.4622
November	Positive control (mock)	1	<i>Bacillus licheniformis</i>	<i>Bacillus licheniformis</i>	94.23	0.4149
November	Evaporator Drain	7	<i>Paracoccus chinensis</i>	<i>Paracoccus chinensis</i>	93.03	1.1235
November	Gasket/Flow Plate Seal	1	<i>Macroccoccus caseolyticus</i>	<i>Macroccoccus caseolyticus</i>	90.35	1.1049
November	External Dryer Balance Tank	3	<i>Nesterenkononia massiliensis</i>	<i>Nesterenkononia massiliensis</i>	91.32	1.07
December	Positive control (mock)	3	<i>Bacillus licheniformis</i>	<i>Bacillus licheniformis</i>	96.29	0.0829
December	External Dryer Balance Tank	1	<i>Kocuria</i> sp. CPCC 104605 / <i>Kocuria</i> sp. ZOR0020	<i>Kocuria</i> sp. ZOR0020	81.35	0.6578

Taxonomy was assigned to metaBAT2 binned contigs by Megan LR and open reading frames of these contigs by Kaiju. If more than one species assigned, the bold species represents the top hit per classifier. Bin quality determined by checkM.

samples from all 3 sequencing types (Supplementary Fig. 5). Sanger sequencing involved partial 16S rRNA sequencing of morphologically different colonies from BHI plates, including total spread plate (TBC), thermophilic enriched spore pasteurised (ST) and mesophilic enriched spore pasteurised (SM) tests (Supplementary Table 2; Fig. 5). There was agreement between Sanger sequencing of isolates and next generation sequencing of plate samples with respect to *Kocuria*, *Acinetobacter* and *Lysinibacillus* (Fig. 5). Some genera identified in Plate NextSeq samples and Sanger sequences had not been seen in high relative abundance in corresponding culture-independent NextSeq or MinION sequencing. These included *Microbacterium* in the November table sample and *Lysinibacillus* in the December internal dryer balance tank (Fig. 5).

Overall, Sanger sequencing of 16S variable region of TBC isolates corresponded well with NextSeq 'Plate' sequencing but fewer genera were identified per sample. This may in part be due to only very morphologically distinct isolates being selected for Sanger sequencing. Counts per swab are also included. At all time points the gasket/flow plate seals and the evaporator drains had the highest CFU/swab, with on average 3.18×10^7 CFU/swab and 1.82×10^8 CFU/swab each. These two areas also had the highest mesophilic spore count with an average of 1.17×10^4 CFU/swab and 3.64×10^4 CFU/swab each (Fig. 5, Supplementary Table 2).

Relatively few taxonomic classification significant differences

Overall, only 6 out of 108 species had significantly different relative abundance between environmental samples (excluding controls) due to sample processing or sequencing method, based on Pairwise Wilcoxon rank sums test using Benjamini Hochberg *p*-value correction analysis of sequential pairs (Supplementary Figure 6). *Enterococcus casseliflavus*, *Acinetobacter lwoffii* and *Acinetobacter johnsonii* had significantly higher relative abundance in MDA MinION sequenced samples than MDA NextSeq sequenced samples, whereas *Kocuria* sp. WRN011 was identified at significantly higher relative abundance in MDA NextSeq samples than MDA MinION samples. *Pseudochrobactrum* sp B5 was detected at significantly higher relative abundance in NPP NextSeq samples than MDA processed NextSeq samples, whereas *Exiguobacterium sibiricum* was detected at significantly higher relative abundance in MDA NextSeq samples compared to NPP NextSeq samples (Fig. 6a). Genera that had significantly different relative abundances, depending on whether MinION or NextSeq sequencing approaches were used, were also identified. In this case a greater number of significantly different taxa was observed, with 24 genera out of a total of 46 being significantly different as a consequence of the sample processing or sequencing method used (Fig. 6b, Supplementary Figure 7). Six genera differed significantly between more than one pairwise group (Fig. 6b). *Pseudochrobactrum* was present at significantly different relative abundances across all 3 pairwise groups (i.e., MDA MinION and MDA NextSeq, MDA NextSeq and NPP NextSeq, and NPP NextSeq and Plate NextSeq). *Exiguobacterium* and *Planococcus* were present at significantly different relative abundances between MDA MinION and MDA NextSeq as well as MDA NextSeq and NPP NextSeq. *Bacillus*, *Staphylococcus* and *Ochrobactrum* were present at significantly different relative abundances between MDA NextSeq and NPP NextSeq as well as NPP NextSeq and Plate NextSeq. The remaining 18 genera only differed across one pair of analyses (Fig. 6b).

DISCUSSION

16S rRNA rapid barcoding-based MinION sequencing of a simple mock community coupled with MEGAN classification by aligning with BLAST against a Silva database provided species level classification to 3 of the 4 species in a mock community and

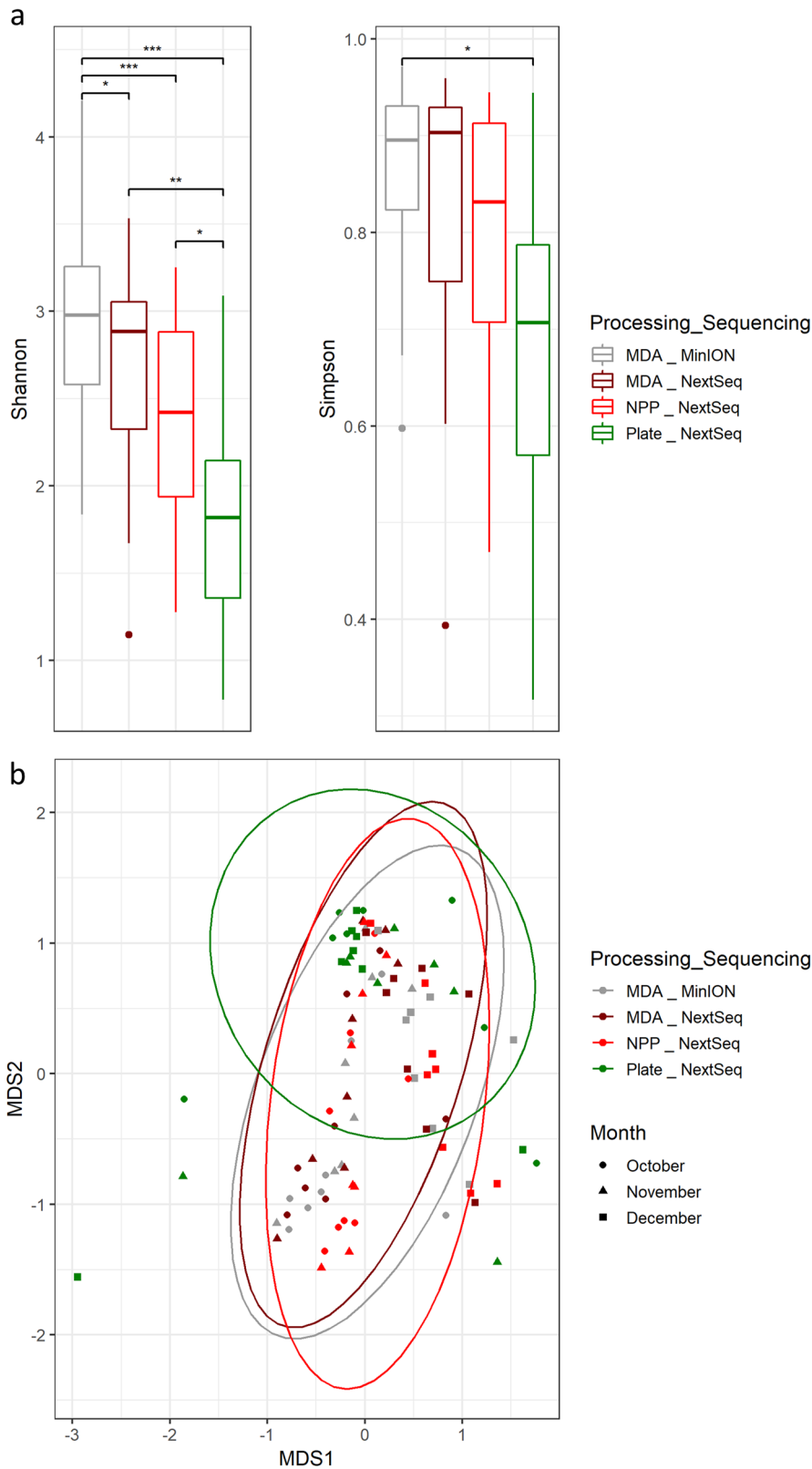


Fig. 4 Diversity analysis. **a** Shannon and Simpson alpha diversity analysis. **b** Bray-Curtis multidimensional scaling (MDS) beta diversity analysis. Boxplot centre line, median; box limits, upper and lower quartiles; whiskers, $\times 1.5$ interquartile range; points, outliers (** $= p < 0.01$, * $= p < 0.05$). Controls are excluded from these calculations and figures. MDA MinION, refers to samples whose DNA was subject to MDA before sequenced on an Oxford Nanopore MinION sequencer. MDA NextSeq, refers to samples whose DNA was subject to MDA before sequenced on a Illumina NextSeq sequencer. NPP NextSeq, refers to samples whose DNA had no pre-processing (NPP) in terms of amplification before sequencing on an Illumina NextSeq sequencer. Plate NextSeq, refers to samples whose metagenomic DNA was extracted from easily cultured microorganisms before sequencing on an Illumina NextSeq sequencer.

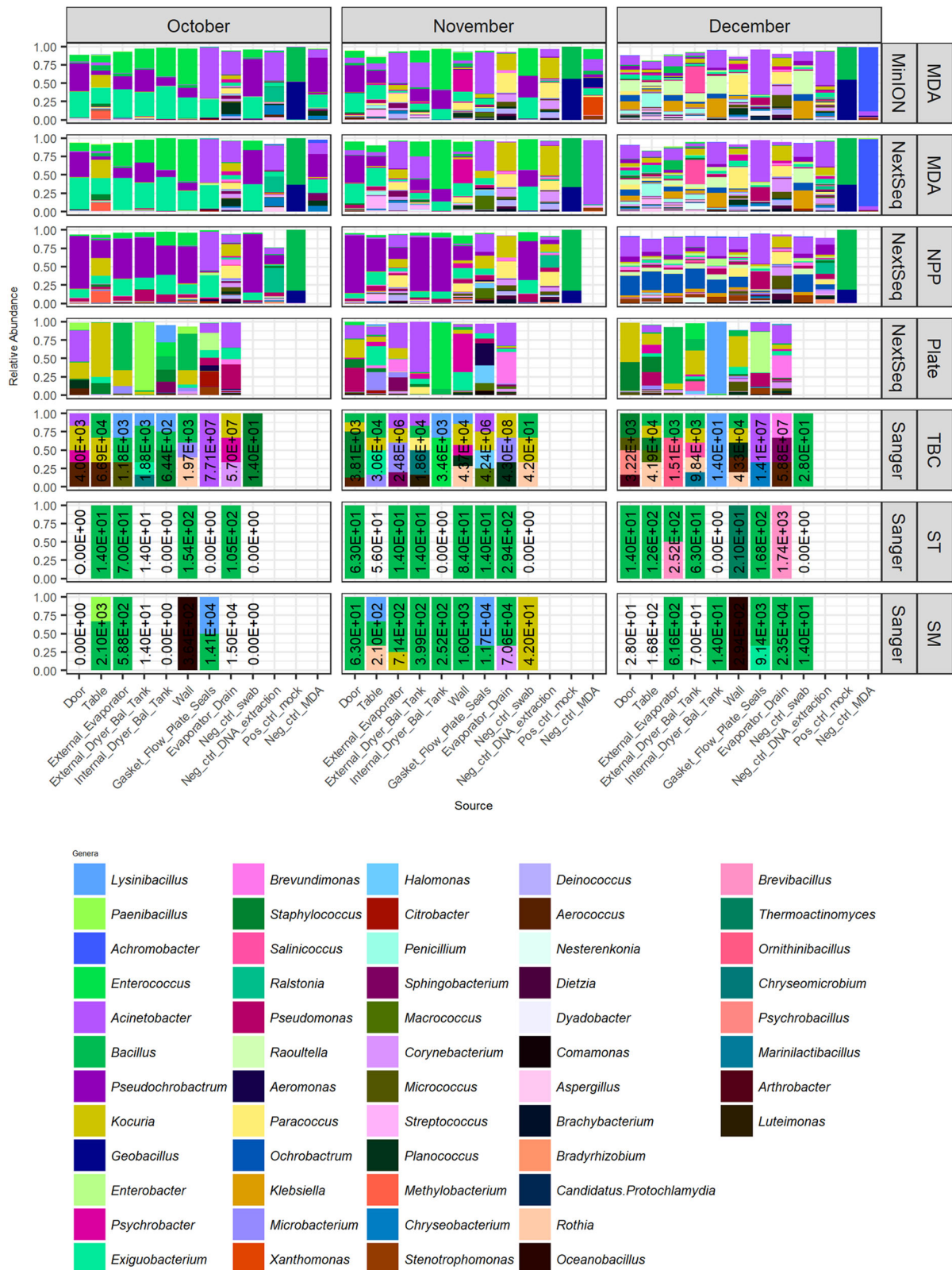


Fig. 5 Genus level classification of environmental samples and controls following different pre-processing methods and sequencing methods. MEGAN LCA based genera level classification of MinION and NextSeq sequences. Also depicted are Sanger results to genus level for morphologically different colonies from each sample (TBC) along with thermophilic sporeformer enriched (ST) and mesophilic sporeformer enriched (SM) counts. Also included are CFU/swab counts for each culturing type. Sanger results represent relative abundance of a subset of morphologically distinct isolates rather than total isolates.

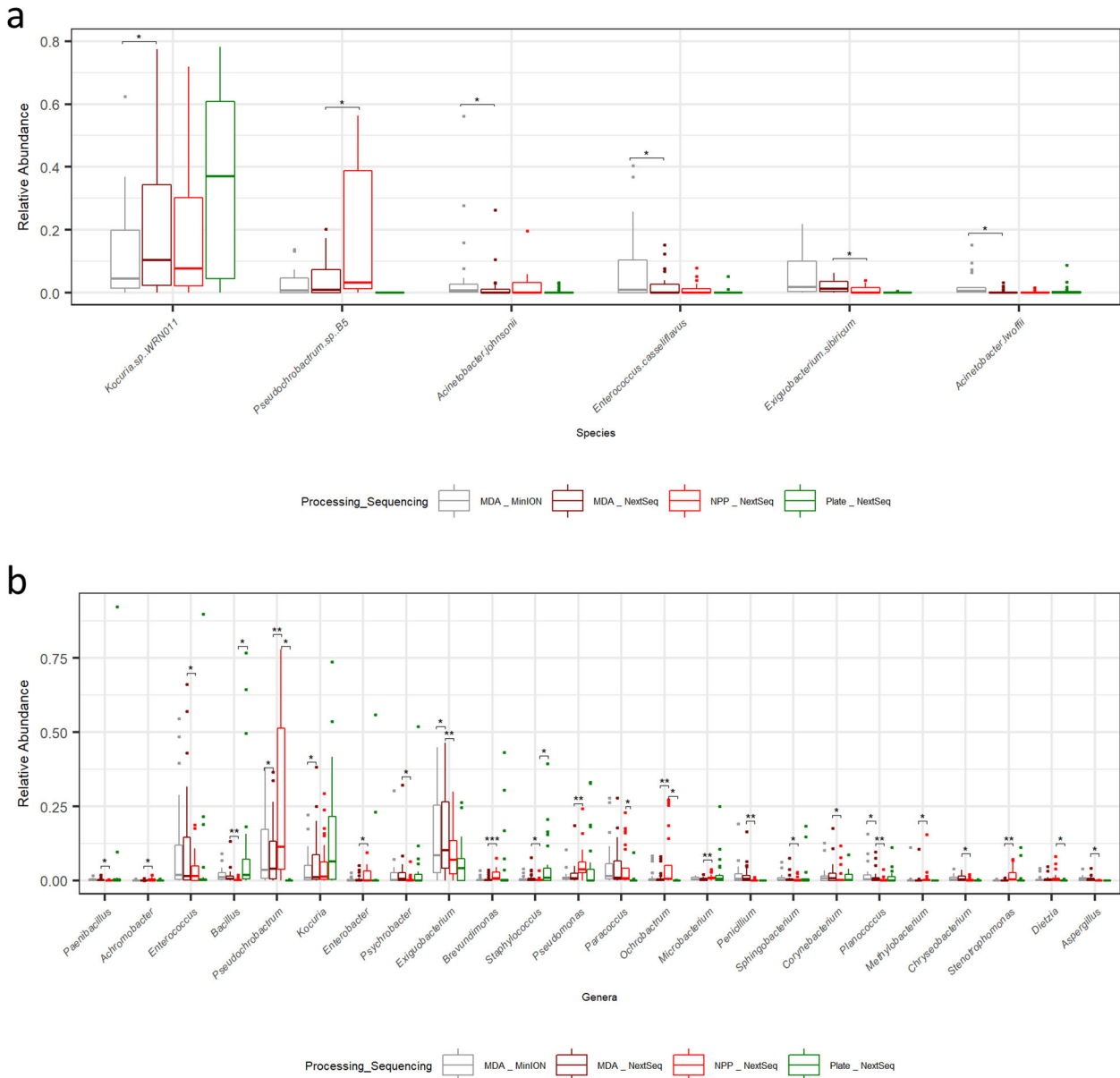


Fig. 6 Significant differences in the relative abundance of taxa based on processing and sequencing method. **a** Significant species level differences due to sequencing and processing methods on environmental samples. Controls are excluded from these calculations and figures. **b** Significant genera level differences in relative abundance due to sequencing and processing methods on environmental samples. Controls are excluded from these calculations and figures. Boxplot centre line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers (***) = $p < 0.001$, ** = $p < 0.01$, * = $p < 0.05$).

correctly identified both genera present. Indeed, this sequencing kit is marketed for genus level bacterial identification by ONT. The rapid sequencing kit-based shotgun sequencing on the MinION platform coupled with LAST alignment against the NR database and MEGAN taxonomic classification resulted in correct classification of all four taxa but with low level false positive detection of *B. paralicheniformis*, a close relative of *B. licheniformis*. Thus, in this regard, MinION rapid WGS performed better than MinION 16S sequencing for species level classification of related species, and could be further improved by reducing/eliminating false positives by exercising a stricter cut off and only focusing on species detected at high relative abundance.

Environmental DNA samples subject to MDA resulted in MinION sequencing reads that were shorter, with lower output than the high quality, high quantity, pooled mock metagenomic DNA generated. This is a particular issue for sequencing of low biomass

environmental samples where, without the use of MDA, the quantities of DNA would not suffice for current rapid protocols, even after pooling of multiple swabs. The multiplexing of poorer quality DNA from environmental samples resulted in saturation of flow cells, resulting in lower output compared to the mock sequencing run. Despite this, MinION sequencing of environmental samples did perform well and was comparable to other methods when all factors were considered. Some of the most abundant species identified included *Kocuria* sp. WRN011, *Enterococcus casseliflavus* and *Enterococcus faecium*. *Kocuria* sp. WRN011 is a saline alkaline soil isolate, and is perhaps selected for due to the unfavourable conditions within a food processing environment arising after cleaning in place (CIP). Both *Enterococcus faecium* and *Enterococcus casseliflavus* are common dairy microorganisms^{20,21}, with *Enterococcus* sp. been known to also be capable of growth at high pH and in the presence of NaCl²².

However, caution is needed when interpreting the results, particularly from low biomass areas. There did appear to be some cross over between environmental sequences and negative controls particularly in environmental samples with low molecular loads. It must be considered that results for species classified in these samples could be false positives from cross over or contamination of sequences from other samples at any stage of swabbing, extraction, amplification or sequencing. As this occurrence was noted in both MinION and NextSeq generated sequences it is unlikely to be due to barcode misassignment or index swapping alone.

MAG analysis revealed 10 good quality genomes from combined MinION and MDA Illumina sequence reads. Seven of the 10 genomes originated from environmental swab samples, with the other 3 corresponding to positive controls. This form of analysis can, if carried out on a larger scale in the future and with greater sequencing depth, be used to bridge discrepancies in taxonomic classification.

There were also significant differences in the relative abundance of species due to the pre-processing and sequencing approaches taken. MinION sequencing indicated greater relative abundances of *Enterococcus casseliflavus*, *Acinetobacter lwoffii* and *Acintobacter johnsonii* than was suggested by MDA NextSeq sequencing. NextSeq MDA appeared to preferentially sequence *Kocuria sp.* WRN011 compared to MinION. *Pseudochrobactrum sp.* B5 abundances appeared lower in MDA (MinION and NextSeq) and easily culturable NextSeq plate samples than NPP samples. From a culture-based perspective, it is noted that this species is known to reduce hexavalent chromium²³ and it may not grow well on the BHI agar used. *Exiguobacterium sibiricum* was detected in higher relative abundance in MDA amplified samples, with significantly higher relative abundance in MDA NextSeq samples compared to NPP NextSeq samples. This suggests it is preferentially amplified by multiple displacement amplification, leading to an overestimation of its relative abundance in these samples.

There were also significant differences in the relative abundances of genera that could not be assigned at the species level. This was most apparent when MDA NextSeq and NPP NextSeq outputs were compared. As well as plate sequences having lower levels of *Pseudochrobactrum*, they were also a lot less diverse than those generated through culture-independent approaches, suggesting culturing at the conditions used was less sensitive. Many species were seen in higher relative abundance in NextSeq plate samples than samples not subject to pre-culturing, including *Planococcus massiliensis*, *Microbacterium oxydans*, *Acinetobacter baumannii* and *Lysinibacillus sp.* B2A1, presumably as a consequence of being better suited to growth in these conditions.

Small, portable, real-time DNA sequencers provide the first steps towards real-time industry paced microbial classification and analysis, which could allow the implementation of process change to counteract microbial issues. Although DNA sequencing has been used sporadically for source tracking^{12,14} and monitoring the microbiota through various seasons and environmental conditions²⁴, there are currently limited numbers of publications and datasets relating to food chain and processing facility microbiomes. While Oxford Nanopore sequencing accuracy is constantly improving²⁵, this in itself provides another hurdle to routine implementation in food processing environments, due to often lack of back compatibility with kits, hardware, software and analysis pipelines. More importantly, the need for high quality, high quantity DNA from swabs of an area that actively aims to have low bacterial loads is a challenge, further highlighting the need for adequate controls. Ideally, future forms of portable technologies can be implemented with a rapid kit, without a need for amplification. Despite these challenges, this study and the data generated will aid further attempts to characterise the microbiotas across the food chain, leading to an acceleration towards routine implementation. This is particularly true regarding the generation

of MAGs from MDA amplified DNA, resulting in good quality MAGs for 7 environmental isolates, for which relatively few genomes are already available. Notably, in some cases it was difficult to assign some of these MAGs to an existing species, suggesting that the genomes isolated were from related, but previously unclassified species. While *Exiguobacterium sp.* and *Kocuria sp.* have previously been reported in food processing environments^{26,27}, *Planococcus sp.*, although not well characterised with few genomes available, are regarded as halotolerant, water-associated microorganisms, rather than food processing contaminants²⁸. The generation of this MAG and further generation of MAGs, will accelerate the identification of food chain microbes through sequencing-based approaches in the future.

Ultimately, while this study highlights issues relating to sourcing sufficient template DNA, inconsistencies across sequencing approaches and platforms, and challenges with assigning taxa, the considerably great potential merits of applying metagenomic approaches to monitor the microbiology of the food chain are clear.

MATERIALS AND METHODS

Mock community

DNA from 4 target strains, *Bacillus cereus* DSM 31/ATCC 14579, *Bacillus thuringiensis* DSM 2046/ATCC 10792, *Bacillus licheniformis* DSM 13/ATCC 14580 and *Geobacillus stearothermophilus* DSM 458 (Accession numbers GCF_000007825.1, GCF_002119445.1, GCF_000011645.1, and GCF_002300135.1, respectively), was combined to represent a 'mock' metagenomic sample of spore-forming bacteria. Genomic DNA was purchased (latter two strains, DSMZ) or extracted from in-house stocks (former two strains). Where necessary, DNA extraction was performed using the GenElute Bacterial Genomic DNA extraction kit (Sigma Aldrich, NA2110) according to manufacturer's instructions for Gram positive bacteria DNA extraction except that DNA was eluted in 75 µl elution solution. DNA concentrations were determined using the Qubit double-stranded DNA (dsDNA) high sensitivity (HS) assay kit (BioSciences) and ran on 1% agarose gel to check quality. DNA was diluted to 24 pM and pooled equimolar. 16S rRNA gene amplicon sequencing, using the 16S rapid barcoding kit SQK-RAB204, as well as rapid whole metagenome sequencing (WMGS), using the rapid sequencing kit SQK-RAD004, was performed using the Oxford Nanopore MinION sequencer. These kits required 10 ng and 400 ng of DNA input, respectively. More specifically, the SQK-RAB204 16S rapid barcoding kit was used for library preparation according to manufacturer's instructions with barcode 01. DNA was sequenced on FloMIN 106 R9 version flowcell mk1 with minKNOW version 1.7.14 according to manufacturer's instructions. The SQK-RAD004 rapid sequencing kit was used to prepare the DNA according to manufacturer's instructions, DNA was sequenced on FloMIN 106 R9 version flowcell mk1 with minKNOW version 1.11.5 according to manufacturer's instructions.

Bioinformatic analysis of mock community metagenomic DNA

Genome sequences for the 4 strains represented in the mock metagenomic community were downloaded from NCBI RefSeq and aligned in a pairwise manner using the Artemis comparison tool (ACT)²⁹ (Supplementary Figure 1). 16S DNA sequences were rebasecalled using Albacore (version 2.2.6). FastQC was used to check sequence length and quality. IDBA fq2fa was used to convert fastq files to fasta format. BLASTn alignment³⁰ of sequences against 16S Silva database (release 132)^{31,32} was performed with taxonomic classification by MEGAN (version 6.12.3)³³. Genus and species levels of classification were determined, and relative abundances calculated and plotted using R ggplot2³⁴. Following basecalling with Albacore, Porechop (version 0.2.4) was used to remove adaptors from rapid WMGS reads before FastQC was used to check sequence length and quality and IDBA fq2fa was used to convert fastq format to fasta format³⁵. LAST alignment of reads^{36,37} was performed against the NR database (March 2018)^{38,39} with MEGAN long read (LR) (MEGAN version 6.12.3)⁴⁰ taxonomic classification. Ranks were split, relative abundances calculated and plotted using R ggplot2³⁴.

The assembly of contigs from metagenomic reads was performed using Canu version 1.7¹⁹ with -nanopore-raw flag. MUMmer alignment was performed on the assembled contigs against the 4 known species

genomes from RefSeq, with dnadiff used to highlight differences between assemblies and reference genomes^{41,42}. The resulting comparisons were visualised using R ggbio and GenomicRanges^{43,44}.

Environmental sample collection and processing

Environmental swabbing was performed in a commercial dairy processing pilot plant. Eight locations were swabbed during the course of a single day, after cleaning in place (CIP) had been completed and before the next round of dairy processing (Fig. 2). These eight locations included a table, door, wall, gaskets/flow plate seals, external surface of dryer balance tank, internal surface of dryer balance tank, external surface of evaporator, and drain beside evaporator. Overall, these eight locations were swabbed over three different months (October, November, December), at a frequency of once per month. Swabbing was performed using Technical Service Consultants Ltd. sponges in neutralising buffer (Sparks Lab Supplies, SWA2023). A total of 5 swabs were performed per surface. Swabbing was performed according to manufacturer's instructions (Hygiene sponge sampling kits swabbing procedure). Briefly, a stomacher bag containing a pre-moistened sponge was shaken to bring the sponge to the bottom of the bag. The bag was torn open above the zip lock; then, holding the sponge at the bottom from the outside, the bag was carefully peeled back, from above the zip lock over the gloved hand, taking care not to touch the inside of the bag or sponge. The exposed sponge was used to swab an area of 360 cm² swabbing vertically with one side of the sponge and horizontally over the same area with the other side of the sponge. The bag was then carefully reverted to its original position, without touching the inside and the zip lock sealed. A total of 5 swabs of each area to be sampled were performed in this way. The surface area was then wiped with disinfectant to remove neutralising buffer.

In the laboratory, 5 sponges for each area were pooled aseptically into the stomacher bag of one. Each bag of 5 sponges was subjected to stomaching at 260 rpm for 1 min. The liquid was then removed, yielding 21 ml for each sample of 5 sponges. A total of 20 ml was prepared for DNA extraction. 1 ml was used for culturing. Two × 15 ml falcon tubes for each sample holding a total of 20 ml were centrifuged at 4,500 × g for 20 min at 4 °C. The supernatant was discarded, and pellet resuspended in 500 µl UV treated, autoclaved phosphate buffered saline (PBS). The two resuspended pellets for each area were pooled into a 2 ml microfuge tube. This tube was centrifuged at 13,000 × g for 2 min and the supernatant was discarded. The pellet was stored at −80 °C for up to 1 month before DNA extraction. Swab negative controls were also processed in the same way for each sampling day. Briefly, 5 swabs were pooled, subjected to stomaching, liquid collected, 1 ml split for culturing, 20 ml pelleted, washed and frozen.

Culture analysis

Of the 1 ml of liquid recovered from each stomacher bag, 100 µl was plated on BHI agar in triplicate. Another 100 µl was used for serial tenfold dilution to 10^{−6} and spread plate on BHI agar in triplicate. All agar plates incubated at 30 °C for 48 h. 600 µl of liquid was subjected to spore pasteurisation by heating to 80 °C for 12 min in a heating block. This heat treated liquid was then spread plated on BHI in triplicate for incubation at both 30 °C and 55 °C for 48 h, after which time colonies were counted to determine colony forming units (CFU).

For each sample, the colonies from one agar plate, onto which the neat stomacher bag liquid had been plated, were removed by washing and pelleted to facilitate DNA extraction to represent metagenomic DNA from easy to culture environmental microorganisms. To this end, 5 ml PBS was added to the agar plate, and swirled around, before colonies were scraped off with a sterile Lazy-L spreader (Sigma-Aldrich) and 4 ml recovered into a sterile 15 ml falcon tube. This was centrifuged at 4,500 × g for 20 min at 4 °C before removing supernatant. The resulting pellet was resuspended in 1 ml PBS and transferred to a 2 ml microfuge tube. The tube was centrifuged at 13,000 × g for 2 min at room temperature and supernatant removed. The pellet was stored at −80 °C for up to three months before DNA extraction. From other agar plates, isolated colonies with obviously different morphologies from each sample were picked, restreaked for purity, inoculated in BHI broth and stocked at −20 °C in a final concentration of 25% glycerol.

DNA extraction and MDA amplification

The Qiagen PowerSoil Pro kit was used for DNA extractions from both environmental sample pellets, and easily culturable washed plate pellets. Easily culturable pellets were removed from −80 °C storage and

resuspended in 1 ml PBS. A total of 200 µl (or 500 µl for 9 smaller pellets, corresponding door, external evaporator and internal dryer balance tank samples for all 3 months) was removed and centrifuged at 12,000 × g for 2 min. The supernatant was discarded and the pellet retained. These pellets, and those sourced directly from environmental swabbing, i.e., without culture, were resuspended in 800 µl CD1 and transferred to a Powerbead Pro tube. Powerbead Pro tubes were secured in a tissue lyser set at 20 Hz for 10 min before centrifuging and following the rest of the PowerSoil Pro kit manufacturer's instructions, eluting in a smaller volume, of 35 µl. For each sampling day, negative controls, involving unused swabs, were also prepared by following an identical extraction protocol and additional negative controls, to detect kit contaminants, were generated whereby an extraction was performed using the kit reagents alone, starting with 800 µl solution CD1.

Whole metagenome amplification was performed using multiple displacement amplification (MDA) with the REPLI-g Single Cell kit (Qiagen, 150345). MDA was performed using DNA from environmental samples and controls for each day. These controls consisted of swab negative control, DNA extraction kit negative control, blank MDA preparation as a MDA negative control and mock metagenomic community (section 1.6.1) as a positive control. DNA concentrations were determined using Qubit dsDNA HS kit. Samples with high DNA concentrations were diluted such that all samples had a final concentration of <10 ng in 2.5 µl. MDA amplification was performed according to manufacturer's instructions for 12 sample amplifications at a time (8 environmental samples, 1 positive control, 3 negative controls (swab, extraction, MDA)). Briefly, 500 µl of H₂O was added to buffer DLB, mixed well and centrifuged briefly, storing at −20 °C for up to 6 months. Buffer D1 and N1 were prepared according to instructions on the day of use. 2.5 µl buffer D1 was added to 2.5 µl DNA, this was mixed by vortexing and centrifuged briefly. Samples were incubated at room temperature for 3 min. 5 µl buffer N1 was added, mixed by vortexing and centrifuged briefly before storing on ice. Mastermix was prepared on ice according to manufacturer's instructions. For each amplification 40 µl master mix was added to 10 µl denatured DNA. This was incubated on thermocycler at 30 °C for 8 h. The polymerase was then inactivated by heating samples to 65 °C for 3 min also in the thermocycler. Thermocycler settings (2 × 4 h holds at 30 °C and 1 × 3 min hold at 65 °C). Amplified DNA was then stored at −20 °C.

Library preparation and sequencing

For MinION sequencing library preparation DNA concentrations of 36 MDA samples were measured using both the Qubit dsDNA broad range (BR) and HS assays and diluted to 400 ng in 7.5 µl. Three libraries were prepared, containing 12 samples each (8 environmental MDA samples, 3 MDA negative controls (swab, extraction and MDA kit negative controls) and a MDA mock community positive control) per flow cell. The SQK-RBK004 rapid barcoding kit was used to prepare the DNA according to manufacturer's instructions, including an optional Ampure XP clean up step, directly prior to sequencing. DNA was sequenced on FloMIN 106 R9 version flowcell mk1 with MinKNOW version 18.12.4 according to manufacturer's instructions.

For Illumina sequencing library preparation the DNA concentrations of MDA ($n = 36$), non-MDA (i.e., metagenomic DNA not subjected to pre-processing (NPP)) ($n = 33$), and easily culturable (Plate) ($n = 24$) metagenomic DNA samples was measured using the Qubit HS dsDNA kit and diluted. DNA was prepared for Illumina sequencing following Illumina Nextera XT Library Preparation Kit guidelines except that tagmentation was performed for 7 min. DNA tagmentation was visualised using Agilent Bioanalyzer high sensitivity DNA analysis, and average fragment size calculated. The DNA concentration was measured by Qubit HS dsDNA assay and the concentration then calculated, before diluting and pooling at equimolar ratios. The DNA library was sequenced on Illumina NextSeq at the Teagasc DNA sequencing facility, with a NextSeq (500/500) High Output 300 cycles v2.5 kit (Illumina 20024908).

For 16S rDNA sanger sequencing of isolated colonies 16S colony PCR was performed using universal primers 27F and 338R for 16S gene (AGAGTTTGTATCTGGCTCAG and CATGCTGCCTCCCGTAGGAGT, respectively). Colonies were picked, and mixed in 50 µl PCR water, before microwaving on full power for 1 minute to disrupt cells. Master mix consisting of 5 µl AccuTaq LA 10x buffer, 2.5 µl 10 mM dNTP mix, 1 µl DMSO, 2 µl 10 µM Forward primer, 2 µl 10 µM reverse primer, 32 µl PCR water, 0.5 µl AccuTaq LA DNA polymerase (Sigma Aldrich, D8045) per amplification was made. 45 µl mastermix was added to each tube of 5 µl disrupted colony, before centrifuging briefly to mix and placing on pre-

programmed thermocycler with 95 °C × 5 min, 25 cycles of 95 °C × 30 s, 55 °C × 30 s, 72 °C × 30 s and a final 72 °C × 5 min, before holding at 4 °C. PCR products were run on a 1% agarose gel, before cleaning with 1.8× Ampure XP. 5 µl of each cleaned up PCR product was aliquoted into a 96 well plate and 5 µl of forward primer added on top at 5 µM according to GATC requirements. A unique barcode was added to each plate and sent to GATC Biotech (Germany) for Sanger sequencing. A subset of amplicons were also sequenced with the reverse primer to ensure accuracy.

Bioinformatic analysis of environmental metagenomic DNA

For analysis of MinION data, Guppy basecalled reads obtained from MinKnow (version 18.12.4) were demultiplexed using Guppy barcoder version (2.1.3) to produce a barcoding summary text file. This contained the percentage match of each read to their barcodes with a minimum score of 60, the default). All fastq files produced by MinKnow were concatenated and guppy_bcsplit.py (https://github.com/ms-gx/guppy_bcsplit) allowed demultiplexing of reads based on their barcode assigned in the barcoding summary text file. Porechop (version 0.2.4) was used to remove adaptors from rapid kit sequence reads before Fastqc was used to check sequence length and quality. IDBA fq2fa was used to convert fastq to fasta³⁵. LAST alignment of fasta files^{36,37} against the NR database (March 2018)^{38,39} was performed with the MEGAN LR classification (MEGAN version 6.12.3)⁴⁰. Files were merged, ranks were split, total number of bases sequenced, and classified were calculated. Relative abundances calculated and plotted using R ggplot2³⁴.

For analysis of NextSeq data, BCI2fastq was used to convert raw sequence reads from Illumina NextSeq to fastq format. Kneaddata from bioBakery⁴⁵ used trimmomatic for quality filtering and trimming paired end files⁴⁶ with BMTagger to remove human and bovine reads. FastQC was used to visualise sequence length and quality. IDBA converted fastq to fasta³⁵. Diamond alignment⁴⁷ of fasta files was performed against the NR database (march 2018)^{38,39} with MEGAN classification (MEGAN version 6.12.3)⁴⁰. Files were merged, ranks were split, total number of bases sequenced and classified calculated and relative abundances calculated and plotted using R ggplot2³⁴. Illumina data were also analysed using Kraken2 and Bracken^{48,49} for taxonomy classification as well as using MetaPhlan2⁵⁰ for taxonomy classification for the purpose of comparison.

In order to generate Metagenome-assembled genomes, MDA amplified sequences from both Illumina and Oxford Nanopore sequencing were assembled using OPERA-MS⁵¹. Illumina reads were then mapped against assemblies using bowtie2⁵² and bam files sorted using samtools⁵³. Depth was calculated and Metabat2 ran on assembled contigs to produce bins^{54,55}. Checkm was used to determine the quality of the metagenome assembled genomes (MAGs). Prokka⁵⁶ was used to generate.ffn files from bins, Kaiju⁵⁷-based taxonomic classification was performed on the open reading frames from prokka. Megan LR⁴⁰ was also used on the whole bins for taxonomic classification of high quality MAGs.

In order to analyse cultures and 16S rDNA sanger sequences, CFUs were determined on the basis of an average of three agar plates per sample. CFU per swab was calculated by dividing by 5 (5 swabs = 1 sample, and each swab covered area 360 cm²). 16S rRNA Sanger sequences resulting from morphologically different isolates per sample were blasted using BLASTn against the 16S ribosomal RNA (Bacteria and Archaea) database on NCBI, with top hits recorded, and genus level classification analysed.

In all comparisons Pairwise Wilcoxon rank sums test using Benjamini Hochberg *p*-value correction analysis was used to compare sample groups, including investigations of the impact of sequencer type on taxonomy classification with MinION MDA-treated and NextSeq MDA-treated samples. The impact of MDA amplification was also investigated in this way through comparison between NextSeq MDA treated samples and NextSeq no pre-processing (NPP) samples. Differences in taxonomy classification between sequences derived from environmental metagenomic DNA versus those sourced from easy to culture microorganisms was shown by comparing NextSeq NPP and NextSeq easy to culture (plate) sequences. Diversity analysis was performed in R with vegan package. Shannon and Simpson alpha diversity metrics were calculated along with Bray Curtis Nonmetric Multidimensional Scaling beta diversity metrics. Pairwise Wilcoxon rank sums test using Benjamini Hochberg *p*-value correction was used to compared samples groups based on sequencing and processing methods used, controls were excluded from these calculations.

DATA AVAILABILITY

Sequence data have been deposited in the European Nucleotide Archive (ENA) under the study accession number PRJEB39267.

Received: 18 September 2020; Accepted: 13 January 2021;

Published online: 15 February 2021

REFERENCES

- Gleeson, D., O'Connell, A. & Jordan, K. Review of potential sources and control of thermophilic bacteria in bulk-tank milk. *Ir. J. Agric. Food Res.* **52**, 217–227 (2013).
- Doyle, C. J., Gleeson, D., O'Toole, P. W. & Cotter, P. D. High-throughput sequencing highlights the significant influence of seasonal housing and teat preparation on the raw milk microbiota. *Appl. Environ. Microbiol.* **83**, 02694–16 (2016).
- Faillie, C. et al. Sporulation of *Bacillus* spp. within biofilms: a potential source of contamination in food processing environments. *Food Microbiol.* **40**, 64–74 (2014).
- Wang, B. et al. Bacterial composition of biofilms formed on dairy-processing equipment. *Prep. Biochem. Biotechnol.* **49**, 477–484 (2019).
- Fysun, O., Kern, H., Wilke, B. & Langowski, H.-C. Evaluation of factors influencing dairy biofilm formation in filling hoses of food-processing equipment. *Food Bioprod. Process.* **113**, 39–48 (2019).
- Doyle, C. J. et al. Anaerobic sporeformers and their significance with respect to milk and dairy products. *Int. J. Food Microbiol.* **197**, 77–87 (2015).
- Cho, T. J. et al. New insights into the thermophilic spore-formers in powdered infant formula: Implications of changes in microbial composition during manufacture. *Food Control* **92**, 464–470 (2018).
- Sadiq, F. A. et al. The heat resistance and spoilage potential of aerobic mesophilic and thermophilic spore forming bacteria isolated from Chinese milk powders. *Int. J. Food Microbiol.* **238**, 193–201 (2016).
- Burgess, S. A., Lindsay, D. & Flint, S. H. Thermophilic bacilli and their importance in dairy processing. *Int. J. Food Microbiol.* **144**, 215–225 (2010).
- Tallent, S. M., Kotewicz, K. M., Strain, E. A. & Bennett, R. W. Efficient isolation and identification of *Bacillus cereus* group. *J. AOAC Int* **95**, 446–451 (2012).
- Doyle, C. J., O'Toole, P. W. & Cotter, P. D. Genomic characterization of sulphite reducing bacteria isolated from the dairy production chain. *Front. Microbiol.* **9**, 1507 (2018).
- Doyle, C. J., Gleeson, D., O'Toole, P. W. & Cotter, P. D. Impacts of seasonal housing and teat preparation on raw milk microbiota: a high-throughput sequencing study. *Appl. Environ. Microbiol.* **83** (2017).
- McHugh, A. J. et al. Mesophilic Sporeformers Identified in Whey Powder by Using Shotgun Metagenomic Sequencing. *Appl. Environ. Microbiol.* **84**, e01305–18 (2018).
- Fretin, M. et al. Bacterial community assembly from cow teat skin to ripened cheeses is influenced by grazing systems. *Sci. Rep.* **8**, 200 (2018).
- Charalampous, T. et al. Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. *Nat. Biotechnol.* **37**, 783–792 (2019).
- Quick, J. et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat. Protoc.* **12**, 1261 (2017).
- Kafetzopoulou, L. E. et al. Assessment of metagenomic Nanopore and Illumina sequencing for recovering whole genome sequences of chikungunya and dengue viruses directly from clinical samples. *Euro Surveill* **23**, 1800228 (2018).
- Sanderson, N. D. et al. Real-time analysis of nanopore-based metagenomic sequencing from infected orthopaedic devices. *BMC genomics* **19**, 714 (2018).
- Koren, S. et al. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
- Rivas, F. P., Castro, M. P., Vallejo, M., Marguet, E. & Campos, C. A. Antibacterial potential of *Enterococcus faecium* strains isolated from ewes' milk and cheese. *LWT - Food Sci. Technol.* **46**, 428–436 (2012).
- Gelsomino, R., Vancanneyt, M., Cogan, T. M., Condon, S. & Swings, J. Source of enterococci in a farmhouse raw-milk cheese. *Appl. Environ. Microbiol.* **68**, 3560 (2002).
- Khedid, K., Faid, M., Mokhtari, A., Soulaymani, A. & Zinedine, A. Characterization of lactic acid bacteria isolated from the one humped camel milk produced in Morocco. *Microbiological Res.* **164**, 81–91 (2009).
- Ge, S., Dong, X. & Zhou, J. Comparative evaluations on bio-treatment of hexavalent chromate by resting cells of *Pseudochrobactrum* sp. and *Proteus* sp. in wastewater. *J. Environ. Manag.* **126**, 7–12 (2013).
- Li, N. et al. Variation in raw milk microbiota throughout 12 months and the impact of weather conditions. *Sci. Rep.* **8**, 2371 (2018).
- Watson, M. & Warr, A. Errors in long-read assemblies can critically affect protein prediction. *Nat. Biotechnol.* **37**, 124–126 (2019).

26. Vishnivetskaya, T. A. & Kathariou, S. Putative transposases conserved in Exiguobacterium isolates from ancient Siberian permafrost and from contemporary surface habitats. *Appl. Environ. Microbiol.* **71**, 6954–6962 (2005).
27. Røder, H. L. et al. Interspecies interactions result in enhanced biofilm formation by co-cultures of bacteria isolated from a food processing environment. *Food Microbiol.* **51**, 18–24 (2015).
28. Waghmode, S. et al. Genomic insights of Halophilic Planococcus maritimus SAMP MCC 3013 and detail investigation of its biosurfactant production. *Front. Microbiol.* **10**, 235–235 (2019).
29. Carver, T. J. et al. ACT: the Artemis comparison tool. *Bioinformatics* **21**, 3422–3423 (2005).
30. Altschul, S., Gish, W., Miller, W., Myers, E. & Lipman, D. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
31. Pruesse, E. et al. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **35**, 7188–7196 (2007).
32. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2012).
33. Huson, D., Auch, A., Qi, J. & Schuster, S. MEGAN analysis of metagenomic data. *Genome Res.* **17**, 377–386 (2007).
34. Wickham, H. ggplot2: Elegant Graphics for Data Analysis. (Springer Publishing Company, Incorporated, 2009).
35. Peng, Y., Leung, H. C. M., Yiu, S. M. & Chin, F. Y. L. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**, 1420–1428 (2012).
36. Kielbasa, S. M., Wan, R., Sato, K., Horton, P. & Frith, M. C. Adaptive seeds tame genomic sequence comparison. *Genome Res.* **21**, 487–493 (2011).
37. Sheettlin, S. L., Park, Y., Frith, M. C. & Spouge, J. L. Frameshift alignment: statistics and post-genomic applications. *Bioinformatics* **30**, 3575–3582 (2014).
38. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **33**, D501–D504 (2005).
39. Pruitt, K. D., Tatusova, T., Brown, G. R. & Maglott, D. R. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res.* **40**, D130–D135 (2012).
40. Huson, D. H. et al. MEGAN-LR: new algorithms allow accurate binning and easy interactive exploration of metagenomic long reads and contigs. *Biol. Direct* **13**, 6 (2018).
41. Kurtz, S. et al. Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12–R12 (2004).
42. Delcher, A. L., Phillippy, A., Carlton, J. & Salzberg, S. L. Fast algorithms for large-scale genome alignment and comparison. *Nucleic acids Res.* **30**, 2478–2483 (2002).
43. Yin, T., Cook, D. & Lawrence, M. ggbio: an R package for extending the grammar of graphics for genomic data. *Genome Biol.* **13**, R77 (2012).
44. Lawrence, M. et al. Software for computing and annotating genomic ranges. *PLOS Computational Biol.* **9**, e1003118 (2013).
45. Mclver, L. J. et al. bioBakery: a meta'omic analysis environment. *Bioinformatics* **34**, 1235–1237 (2018).
46. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
47. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using. *Diamond. Nat. Meth* **12**, 59–60 (2015).
48. Lu, J., Breitwieser, F. P., Thielen, P. & Salzberg, S. L. Bracken: estimating species abundance in metagenomics data. *PeerJ Computer Sci.* **3**, e104 (2017).
49. Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **15**, R46 (2014).
50. Truong, D. T. et al. MetaPhlan2 for enhanced metagenomic taxonomic profiling. *Nat. Meth* **12**, 902–903 (2015).
51. Bertrand, D. et al. Hybrid metagenomic assembly enables high-resolution analysis of resistance determinants and mobile elements in human microbiomes. *Nat. Biotechnol.* **37**, 937–944 (2019).
52. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Meth* **9**, 357–359 (2012).
53. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
54. Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **3**, e1165 (2015).
55. Kang, D. D. et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359–e7359 (2019).
56. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
57. Menzel, P., Ng, K. L. & Krogh, A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat. Commun.* **7**, 11257 (2016).

ACKNOWLEDGEMENTS

This research was funded by the Department of Agriculture, Food and the Marine (DAFM), under the FIRM project SACCP, reference number 14/F/883. Research in the Cotter laboratory is also funded by Science Foundation Ireland (SFI) under grant numbers SFI/12/RC/2273 (APC Microbiome Ireland) and SFI/16/RC/3835 (Vistamilk) and by the European Commission under the Horizon 2020 program under grant number 818368 (Master).

AUTHOR CONTRIBUTIONS

A.J.M. performed swabbing, culture-based analysis, DNA extraction, DNA library preparations, MinION sequencing, bioinformatics and statistical analysis, drafted and edited manuscript. M.Y. helped A.J.M. with swabbing and culture-based lab work. F.C. helped A.J.M. with MinION sequencing and performed Illumina sequencing. C.F., C.H., and P.D.C. supervised A.J.M. throughout. All authors revised, edited, approved the manuscript and are accountable.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41538-021-00087-2>.

Correspondence and requests for materials should be addressed to P.D.C.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021