FEMS Microbiology Letters, 364, 2017, fnx139

doi: 10.1093/femsle/fnx139
Advance Access Publication Date: 29 June 2017
Research Letter

RESEARCH LETTER – Environmental Microbiology

Comparison of four DNA extraction methods for comprehensive assessment of 16S rRNA bacterial diversity in marine biofilms using high-throughput sequencing

Natàlia Corcoll[1,*,†,‡], Tobias Österlund[2,†], Lucas Sinclair[3], Alexander Eiler[3], Erik Kristiansson[2], Thomas Backhaus[1] and K. Martin Eriksson[4]

[1]Department of Biological and Environmental Sciences, University of Gothenburg, SE-405 30 Gothenburg, Sweden, [2]Department of Mathematical Sciences, Chalmers University of Technology, SE-41296 Gothenburg, Sweden, [3]Department of Ecology and Genetics, Limnology, and Science for Life Laboratory, Uppsala University, SE-752 36 Uppsala, Sweden and [4]Department of Mechanics and Maritime Sciences, Chalmers University of Technology, SE-41296 Gothenburg, Sweden

*Corresponding author: Department of Biological and Environmental Sciences, University of Gothenburg, Box 461, SE-405 30 Gothenburg, Sweden. Tel: (+46) 31 7864807; E-mail: natalia.corcoll@gu.se
†Both authors contributed equally to this work.
One sentence summary: The choice of DNA extraction method for marine peripnyton biofilms significantly influences the bacterial community profiles generated.
Editor: David Clarke
‡Natàlia Corcoll, http://orcid.org/0000-0002-1660-9422

ABSTRACT

High-throughput DNA sequencing technologies are increasingly used for the metagenomic characterisation of microbial biodiversity. However, basic issues, such as the choice of an appropriate DNA extraction method, are still not resolved for non-model microbial communities. This study evaluates four commonly used DNA extraction methods for marine periphyton biofilms in terms of DNA yield, efficiency, purity, integrity and resulting 16S rRNA bacterial diversity. Among the tested methods, the Plant DNAzol® Reagent (PlantDNAzol) and the FastDNA® SPIN Kit for Soil (FastDNA Soil) methods were best suited to extract high quantities of DNA (77–130 μg g wet wt^{-1}). Lower amounts of DNA were obtained (<37 μg g wet wt^{-1}) with the Power Plant® Pro DNA Isolation Kit (PowerPlant) and the Power Biofilm® DNA Isolation Kit (PowerBiofilm) methods, but integrity and purity of the extracted DNA were higher. Results from 16S rRNA amplicon sequencing demonstrate that the choice of a DNA extraction method significantly influences the bacterial community profiles generated. A higher number of bacterial OTUs were detected when DNA was extracted with the PowerBiofilm and the PlantDNAzol methods. Overall, this study demonstrates the potential bias in metagenomic diversity estimates associated with different DNA extraction methods.

Keywords: DNA; extraction methods; bacteria; 16S amplicon sequencing; biofilms; metagenomics

Received: 17 March 2017; **Accepted:** 27 June 2017

INTRODUCTION

The use of high-throughput DNA sequencing technologies, in particular the amplicon sequencing of the bacterial 16S rRNA gene, has been established as a fundamental tool for understanding the structure and diversity of microbial communities (Stoeck *et al.* 2010; Timoner *et al.* 2014) as well as for biomonitoring and ecotoxicological studies (Yergeau *et al.* 2012; Pascault *et al.* 2014). Analysis of the resulting sequences provides a holistic view of the studied microbial community, including the vast majority of organisms that are hard to culture using standard protocols (Ward, Weller and Bateston 1990). The applicability of high-throughput DNA sequencing is constantly growing due to the rapidly decreasing costs for DNA sequencing and breakthroughs in the bioinformatics analysis (Caporaso *et al.* 2010; Jonsson *et al.* 2016). Nevertheless, fundamental aspects such as the choice of a comprehensive and reliable genomic DNA extraction method for prokaryotic microorganisms are still not resolved for specific microbial communities, such as the marine periphyton biofilms that were used as test material in this study.

Periphyton biofilms are microbial communities that live attached to submerged substrata in shallow areas of aquatic ecosystems and host a very high diversity of microorganisms, including bacteria, cyanobacteria, algae, protozoa, fungi and metazoa, among others (Lock 1993; Salta *et al.* 2013; Sanli *et al.* 2015). Biofilms represent the dominant form of microbial communities in many aquatic systems and have a key role in ecological processes such as biogeochemical cycles or as a food source for higher trophic levels (Battin *et al.* 2003; Sundbäck *et al.* 2004). Given their ecological relevance and sensitivity to several environmental stressors, they are commonly used in biomonitoring and environmental risk assessments (Sabater *et al.* 2007; Blanck *et al.* 2009; Corcoll *et al.* 2014). Biofilm microorganisms live closely together, embedded in a self-produced matrix of extracellular polymeric substances (EPS) that provides stability to the biofilm. EPS comprise polysaccharides, proteins, humic substances and lipids (Flemming and Wingender 2010). All these compounds can bind to nucleic acids during the DNA extraction process and decrease the yield and purity of the extracted DNA. DNA contaminated with proteins, polysaccharides, salts or metals may hamper subsequent downstream analyses, such as polymerase chain reaction (PCR) amplification and/or DNA library preparation (Michiels *et al.* 2003; Krizman *et al.* 2006). Mechanical methods used for cell lysis, such as sonication, bead-beating, homogenisation or freeze–thaw cycles, can increase the efficiency of DNA extraction, but can also shear the DNA, resulting in smaller DNA fragments that might be less suitable for further downstream analyses. Furthermore, methods for extracting DNA from microbial communities should extract DNA from all species with the same efficiency. Otherwise the analyses may be biased against or in favour of particular groups of microorganisms (Bürgmann *et al.* 2001; Carrigg *et al.* 2007; Feinstein, Woo and Blackwood 2009). Any bias introduced during the DNA extraction will affect the ability to correctly describe the composition and biodiversity of a bacterial community, and will thus also hamper the comparison of different samples.

Studies have shown for various microbial communities, such as human gut microbiome, mock communities, activated sludge, soils or sea-water that DNA yield, DNA purity and, to a lesser extent, bacterial diversity assessments vary depending on the DNA extraction methodology used (de Lipthay *et al.* 2004; Morgan, Darling and Eisen 2010; Koid *et al.* 2012; Yuan *et al.* 2012; Albertsen *et al.* 2015; Fouhy *et al.* 2016). To our knowledge, for biofilms no study has addressed how the choice of DNA extraction method affects the measurement of community composition and diversity using high-throughput sequencing technologies. The aim of this study is to evaluate the extraction efficiency and estimates of prokaryotic community composition and diversity for four commonly used genomic DNA extraction methods: the Plant DNAzol® Reagent (Invitrogen, USA) that has previously been used for marine periphyton biofilms (Eriksson *et al.* 2009; Sanli *et al.* 2015), the FastDNA® SPIN Kit for Soil (MP Biomedicals, USA) that has previously been used for marine and freshwater biofilms (Corcoll *et al.* 2015; Hellal *et al.* 2016), the Power Biofilm® DNA Isolation Kit (MoBio Laboratories, USA) that has previously been used for freshwater biofilms (Rosi-Marshall *et al.* 2013; Tlili *et al.* 2017) and the Power Plant® Pro DNA Isolation Kit (MoBio Laboratories, USA) that has been recommended for biofilms by the manufacturer because of the high abundance of microalgae and cyanobacteria. Each method was tested in replicate samples of marine periphyton biofilms in terms of DNA yield, efficiency, purity, integrity and bacterial community composition and diversity using amplicon sequencing of the 16S rRNA gene on the Illumina MiSeq platform.

MATERIAL AND METHODS

Biofilm sampling

Marine periphyton biofilms were sampled at Blåbärsholmen (58.254211°N, 11.48505°S), just outside the Sven Lovén Centre for Marine Sciences—Kristineberg at the Gullmar Fjord, on the west coast of Sweden, in October 2013. For the sampling of the periphyton biofilms, three polyethylene-sampling racks were used. These were hanging in the water column from separate buoys at 1.5 m below the surface and connected to the sea bottom on separate ropes (Blanck and Wangberg 1988). Each rack hence represents an independent biological replicate. A total of 17 rectangular glass slides were mounted in polyethylene holders on each side of the racks; in total 34 slides on each rack, corresponding to an area of 3.3 dm^2. Biofilms were allowed to colonise the rectangular glass slides for 2 weeks. Colonised slides were then transported to the laboratory in seawater collected at the site and protected from strong sunlight. The sampling was performed by randomly picking slides from within each replicate rack and scraping off the biofilm from both sides of the slides using sterile cell scrapers (Merck Millipore, Darmstadt, Germany). The biofilm from each replicate rack was hence pooled and subsequently mixed. The pooled biofilms were aliquoted in tubes, pelleted by centrifuged at 6500 g for 10 min, the supernatant was removed and the resulting pellets were snap-frozen in liquid nitrogen and stored at –80°C.

DNA extraction

The following four commonly used DNA extraction methods were compared: (1) the Plant DNAzol® Reagent (Invitrogen, USA), referred to as PlantDNAzol in the following, which is optimised for plant tissues; (2) the FastDNA® SPIN Kit for Soil (MP Biomedicals, USA), referred to as FastDNA Soil in the following, which targets plant and animal tissues from soil-living organisms as well as soil-living bacteria, algae and fungi; (3) the Power Plant® Pro DNA Isolation Kit (MoBio Laboratories, USA), referred to as Power Plant in the following, which targets plant cells and tissues; (4) the Power Biofilm® DNA Isolation Kit (MoBio Laboratories, USA), referred to as PowerBiofilm in the following, which targets various biofilms such as periphyton and microbial mats. The PlantDNAzol protocol is based on grinding the sample in liquid nitrogen after which DNA is separated and purified. The

other three methods subject the samples to bead beating, after which DNA is separated and purified.

For all methods, DNA extraction was conducted according to the manufacturer's instructions, except that the bead beating intensity was increased to 2×30 m s^{-1} for 60 s in order to enhance cell disruption and DNA extraction. Bead beating was performed in a Fast Prep FP120 (MP Biomedicals, USA). The DNA was treated with 3 μL of RNase A (QIAGEN). The starting biofilm biomass ranged between 0.05 and 0.12 g wet weight (Table S1, Supporting Information). Each DNA extraction was performed in triplicate. However, one of the PowerPlant replicates failed during the DNA sequencing, so that data are only available for duplicate samples.

Determination of DNA yield, efficiency, purity and integrity

DNA yield (ng μl^{-1}) was quantified spectrophotometrically with a NanoDrop 2000c Spectrophotometer (Thermo Fisher Scientific, USA). It was also quantified fluorometrically in a TapeStation 2200 using Genomic DNA ScreenTapes (Agilent, USA) and in a Qubit 2.0 fluorometer (Invitrogen, USA). DNA extraction efficiency was calculated as the total amount of DNA extracted per biomass (μg g wet wt^{-1}). NanoDrop was also used to estimate the purity of the extracted DNA. Low absorption ratios at 260/280 nm (<1.7) were used as an indicator of protein impurities, and low absorption ratios at 260/230 nm (<2) were used as an indicator of contamination from humic acids and polysaccharides (Stach *et al.* 2001; Zipper *et al.* 2003). The TapeStation system performs electrophoresis in so-called ScreenTapes and outputs images of DNA integrity as well as a DNA Integrity Number (DIN) based on the sizes of the isolated DNA. The DIN ranges from 1 to 10, and a high DIN indicates large DNA fragments whereas a low DIN indicates more fragmented DNA. DIN determines the fragmentation of a genomic DNA sample by assessing the distribution of signal across the size range using a proprietary algorithm.

Amplicon and Illumina sequencing of bacterial 16S rRNA genes

Amplicon sequences of bacterial 16S rRNA genes were obtained as previously described (Sinclair *et al.* 2015) with some modifications. In short, each sample was first amplified in duplicate using the bacterial primers Bakt_341F (CCTACGGGNGGCWGCAG) and Bakt_805R (GACTACHVGGGTATCTAATCC) with a length of 464 bp (Andersson, Riemann and Bertilsson 2010). These primers target the variable regions V3 and V4 and are equipped with parts of the Thruplex Illumina sequencing adapter (forward primers 5′-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-NNNN-CCTACGGGNGGCWGCAG-3′ and reverse primers 5′-AGACGTGTGCTCTTCCGATCT-GACTACHVGGGTATCTAATCC-3′). Duplicates were pooled and then purified using Agencourt AMPure XP (Beckman Coulter) as recommended by the manufacturer. They were then used as templates in a second PCR step using primers equipped with a 7-bp index to obtain amplicons with complete Thruplex adapters for Illumina sequencing. Normalisation of DNA concentrations was made in the second PCR step and before library preparation, but not in the first PCR step, following the approach of Degnan and Ochman (2012), Sinclair *et al.* (2015) and Frey *et al.* (2017). After another purification using the Agencourt AMPure XP kit and fluorometric quantification in the PicoGreen assay (Quant-iT PicoGReen, Invitrogen), samples were pooled in equimolar amounts. The pooled samples were sequenced at the SciLifeLab SNP/SEQ next-generation sequencing facility (Uppsala University campus) using an Illumina MiSeq with 2×300 bp chemistry.

Bioinformatics

Raw sequence data were analyzed with a custom tailored pipeline for demultiplexing and sequence-pair joining. The forward and reverse reads were joined with the PANDAseq algorithm. This process detects the location of overlaps and recomposes full-length amplicons (Masella *et al.* 2012). Any sequences with missing primers, a low-scoring overlapping region or unassigned base pairs were removed. Next, sequences were filtered using a 10-bp sliding window average PHRED score with a threshold of 25. Downstream taxonomic analyses were conducted using the software package QIIME v.1.8.0 (Caporaso *et al.* 2010) and the Vegan package v.2.3–4 (Dixon 2003). Sequences were grouped into operational taxonomic units (OTUs) based on a 97% identity threshold using UCLUST (Edgar 2010). Low abundant OTUs were filtered out, requiring each OTU to be present at least 10 times among all samples. Taxonomic annotation was performed in QIIME using the Greengenes Core 16S rRNA reference database. The number of OTUs present in each sample was calculated after randomly rarefying each sample down to 6789 sequences (corresponding to the lowest sequencing depth) to minimise any bias due to differences in number of sequences between the samples. The Vegan package was again used to calculate richness (number of observed OTUs) as an indicator for α-diversity. 16S sequences that were classified as coming from chloroplast or eukaryotic chloroplasts were removed before further analysis. The raw data have been deposited to the NCBI short read archive (SRA) with BioProject ID PRJNA378915, https://www.ncbi.nlm.nih.gov/Traces/study/?acc=SRP101769.

Statistical analysis

One-way ANOVAs were used to assess differences between the extraction methods for DNA yield, DNA efficiency, DNA purity and DNA integrity. Post hoc comparisons between the four tested methods were made using the Tukey HSD test. OTUs or OTUs pooled at phylum, class, order, family or genera level with different abundances were identified using a generalized linear model where the counts follow an overdispersed Poisson distribution (Kristiansson, Hugenholtz and Dalevi 2009; Jonsson *et al.* 2016). The p-values were corrected for multiple testing using the false discovery rate (FDR) method. The OTU abundance was used for principal component analysis (PCA). Shared OTUs between DNA extraction methods were graphically visualised in Venn diagrams using the corresponding OTU tables exported from QIIME. The hypergeometric distribution was used to test the distribution of gram negatives and gram positives among the taxa identified with the respective four DNA extraction methods. Pearson correlations were used to test for correlations between descriptors of DNA quantity and quality (Table 1), and descriptors of taxonomic diversity (Table 2). The statistical significance for all the analyses was set to $P < 0.05$ or FDR< 0.05. All statistical analyses were carried out using the R v.3.2.0 software (R Core Team 2013).

RESULTS AND DISCUSSION

DNA yield, efficiency, purity and integrity

We found that the choice of a DNA extraction method influenced DNA yield, extraction efficiency, purity and integrity substantially (Table 1, Fig. S1, Supporting Information). As shown in Table 1, the Plant DNAzol and the FastDNA soil methods produced

Table 1. DNA yield (ng DNA μL^{-1}), DNA extraction efficiency (μg DNA g biofilm wet wt^{-1}), DNA purity (absorbance ratios of 260/280 nm and 260/230 nm) and DNA integrity (DIN) from the four extraction methods tested.

	FastDNA Soil	PowerPlant	PowerBiofilm	PlantDNAzol	P-values
DNA yield					
Nanodrop	195 ± 13.4	18.2 ± 4.1	21.8 ± 3	331 ± 83.9	$P < 0.05$
Qubit	107 ± 5.5	14.9 ± 5	21.8 ± 2.1	143 ± 47.5	$P < 0.05$
TapeStation	50.7 ± 8.8	16.7 ± 2.6	12.5 ± 4.2	74.7 ± 6.7	$P < 0.05$
DNA extraction efficiency					
Nanodrop	142 ± 29.2	22.5 ± 5.5	37.6 ± 6.6	298 ± 76.9	$P < 0.05$
Qubit	77.4 ± 10.2	18.4 ± 5.9	37.5 ± 5.3	129 ± 43.3	$P < 0.05$
TapeStation	37.2 ± 2.6	20.2 ± 1.2	22.1 ± 2.5	67.3 ± 2.1	$P < 0.05$
DNA purity					
Abs 260 nm/280 nm	2 ± 0.03	2.4 ± 0.5	2.2 ± 0.04	1.8 ± 0.1	ns
Abs 260 nm/230 nm	0.4 ± 0.1	1.1 ± 0.4	2.6 ± 1.5	0.5 ± 0.05	$P < 0.05$
DNA integrity					
DIN	4 ± 0.2	7 ± 0.15	6.2 ± 0.15	4.3 ± 1.3	$P < 0.05$

Each value represents the arithmetic mean ± standard error of the mean (n = 3). Statistical significance between extraction methods is denoted as $P < 0.05$ (ANOVA). ns indicates no statistically significant differences between extraction methods.

Table 2. Detected 16S rRNA richness and biodiversity from marine periphyton biofilm DNA extracted with the four studied methods.

	FastDNA Soil	PowerPlant	PowerBiofilm	PlantDNAzol	
n	3	2	3	3	P-values
No. of OTUs	666 ± 42	704 ± 58	809 ± 11	791 ± 7	$P < 0.05$
No. of phyla	17 ± 1	17 ± 1	17 ± 1	18 ± 0	ns
No. of classes	39 ± 1	40 ± 1	40 ± 2	41 ± 1	ns
No. of orders	68 ± 3	70 ± 3	71 ± 2	72 ± 1	ns
No. of families	91 ± 2	95 ± 6	104 ± 3	106 ± 2	$P < 0.05$
No. of genera	141 ± 4	145 ± 12	159 ± 4	162 ± 3	$P < 0.05$

Each value represents the arithmetic mean ± standard error of the mean. n: number of replicates. Statistical significance between extraction methods is denoted as $P < 0.05$ (ANOVA).
ns: indicates no statistically significant differences between extraction methods.

much higher DNA yields than the PowerPlant and PowerBiofilm methods. However, for all DNA extraction methods, the obtained DNA yields were higher than 10 ng μL^{-1}, indicating that further PCR amplification and/or library preparation for DNA sequencing was possible. The Plant DNAzol and the FastDNA Soil methods also had the highest DNA extraction efficiency (Table 1). DNA extraction efficiencies were not correlated to the starting biomass, which was comparable between samples and following manufacture recommendations (Table S1). Extracted DNA applies to both prokaryotic and eukaryotic DNA, while further amplicon sequencing analysis only applies to the prokaryotes.

Estimates of the absolute amounts of DNA extracted differed depending on the quantification method used. NanoDrop estimations of DNA yield were higher compared to those obtained with fluorometric methods (Qubit and TapeStation system) for samples with low 260/230 absorbance ratios (around 0.4 and 0.5; Table 1). This is caused by the UV absorbance measurements in the Nanodrop being affected by contaminants, such as humic acids, polysaccharides or metals, and single-stranded DNA, and because the NanoDrop provides only non-discriminatory concentrations for DNA and RNA (Nakayama *et al.* 2016). In contrast, the Qubit and TapeStation methods use a dye that specifically binds to double-stranded DNA and both these methods are therefore less impacted by contaminants and provide more accurate DNA concentration estimates (Georgiou and Papapostolou 2006; Nakayama *et al.* 2016). Based on the Qubit results, the most reproducible DNA extraction methods in terms of DNA yield were the PowerBiofilm, the PowerPlant and the FastDNA Soil methods with standard errors of 2, 5 and 5.5, respectively, compared to 47.5 with PlantDNAzol. A lower standard error for PlantDNAzol was detected when DNA yield was quantified with TapeStation (6.7). The reason for this difference in reproducibility between the quantification methods for PlantDNAzol is, however, not known.

Different DNA extraction methods gave different purification efficiency of the DNA, with different contaminants left in the extract (Table 1). For all samples, the 260/280 absorbance ratio was above 1.7, indicating a low contamination with proteins (Table 1). The average 260/280 absorbance ratios detected for the PowerPlant and PowerBiofilm kits were above 2.2 which, in spite of the use of 3 μL RNase A per sample, might suggest RNA contamination from these methods. Marked differences in 260/230 absorbance ratios were detected among DNA extraction methods (Table 1). DNA extracted with the FastDNA Soil and the PlantDNAzol methods had very low 260/230 ratios (0.4 and 0.5 respectively; Table 1), indicating a high contamination with polysaccharides and salts, which can inhibit subsequent downstream analyses including PCR amplification (Michiels *et al.* 2003; Krizman *et al.* 2006). Differences in DNA size fragmentation were also observed among methods (Fig. S1). Less small-sized DNA, and hence higher DNA integrity values (DIN > 6), were detected for DNA extracted with the PowerPlant and the PowerBiofilm methods (Table 1). Comparatively, more small-sized DNA was detected after extraction with the FastDNA Soil and the PlantDNAzol methods (Fig. S1). Whether the FastDNA Soil and the PlantDNAzol methods actually shear the DNA during extraction, or if

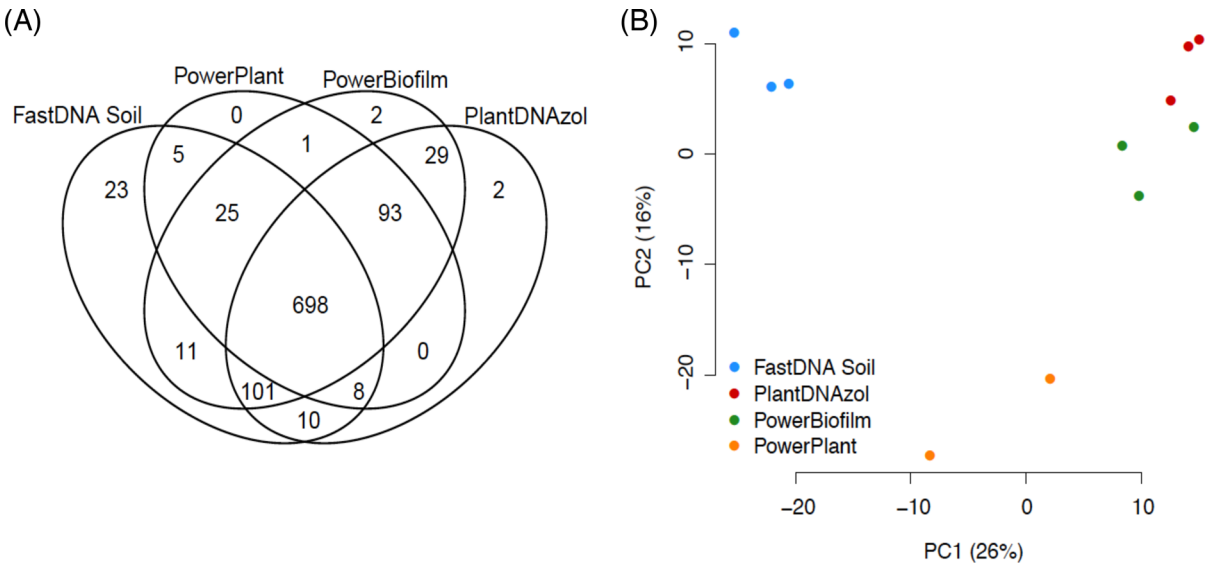

Figure 1. (**A**) Venn diagram of 16S rRNA OTUs. Numbers within overlapping ellipses show numbers of OTUs shared between methods; numbers in non-overlapping ellipses indicate OTUs unique for the indicated method. (**B**) PCA plot of 16S rRNA OTU abundance in all samples.

the PowerPlant and the PowerBiofilm methods fail to capture naturally occurring small-sized DNA fragments remains to be clarified. It is known that polysaccharides in the biofilm EPS bind DNA (Das, Sehar and Manefield 2013). Small-size DNA, such as DNA in various stages of degradation and transferable genetic elements (e.g. plasmids), might therefore bind to a larger extent to polysaccharides, and might thus be discarded from the extraction together with the polysaccharides. If so, the methods that efficiently remove polysaccharides and other contaminants would produce DNA extractions in which only the longer DNA fragments are kept, and in which, consequently, DNA concentrations are lower. In general, DNA extracts with high integrity are preferable (Michiels *et al.* 2003), but differences in DNA size fragmentation should be unproblematic for PCR amplification-based approaches, as long as the length of the amplicon of interest is shorter than the shortest fragments in the extracted DNA.

Overall, our study shows that DNA yield, DNA purity and DNA integrity are affected substantially by the extraction method chosen. The methods extracting the highest amounts of DNA in marine periphyton biofilms (i.e. the FastDNA Soil and the PlantDNAzol methods) are also the methods that produce the lowest DNA purity and integrity. The PowerBiofilm and PowerPlant methods, in contrast, produce high purity DNA with low fragmentation. However, when faced with small biofilm samples or with samples that are difficult to extract, the FastDNA Soil kit and the PlantDNAzol methods might be more suitable due to their higher DNA extraction efficiency.

Composition and diversity of 16S rRNA genes

Overall, the results obtained in this study show that the choice of the DNA extraction method influences the estimated bacterial diversity of marine periphyton biofilms (Table 2, Figs 1 and 2). A higher richness and diversity of 16S OTUs (i.e. number of OTUs, number of families and number of genera) was detected after extraction with the PowerBiofilm and the PlantDNAzol methods compared to extractions with FastDNA Soil and PowerPlant (Table 2). The corresponding Venn diagram shows a high number of shared OTUs between the PlantDNAzol and PowerBiofilm methods (Fig. 1A). In addition to the 698 OTUs that were found with all methods, 29 OTUs were only detected with these two methods. PlantDNAzol and PowerBiofilm do not only produce the most similar OTU composition estimates, they also result in the highest richness and diversity estimates (Table 2). The FastDNA Soil method produced a similar number of OTUs as the PowerPlant method (Table 2). However, the Venn diagram (Fig. 1A) shows that the FastDNA Soil method isolated 23 specific OTUs, which were not detected in extracts from any other method (Fig. 1A, Table S2, Supporting Information). These findings are also reflected in the PCA that was used to visualise the extraction method-induced differences in OTU composition (Fig. 1B). The first axis of the PCA explains 26% of the variance and grouped the samples from PlantDNAzol and PowerBiofilm on the right side of the axis, while samples from FastDNA soil grouped on the left side of the axis. The second axis of the PCA explained 16% of the variance and primarily separated the samples from the PowerPlant method (Fig. 1B). Since the sampling protocol included pooling of all biofilm from each replicate, the differences in OTU composition reflects the overall extraction efficiency for all prokaryotes in the community. However, since no negative controls were included in this study, it does not address the question whether the observed differences could come from contaminating DNA in the reagents or buffers supplied for the different methods. Such contamination is, however, unlikely to cause the observed differences since only fresh reagents and buffers were used and we strictly followed sterile working conditions and the manufacturers' recommendations.

Phylogenetic results show that the marine periphyton biofilms contained a total of 17 phyla and 40 classes of bacteria (Fig. 2). Only small differences were observed between the four DNA extraction methods when comparing abundances at phylum and class levels (Fig. 2). Members of the phyla Bacteriodetes, Proteobacteria, Verrucomicrobiota, Cyanobacteria and Planctomycetes dominated the bacterial biofilm community irrespectively of the DNA extraction method used. These findings are in line with a previous metagenomic study of marine biofilms from the same area (Sanli *et al.* 2015). The proportion of gram-negative bacteria (90%) was higher than the proportion of gram-positive bacteria (2%–3%), something expected for biofilm communities (Moreno-Paz *et al.* 2010; Felczykowska *et al.* 2015).

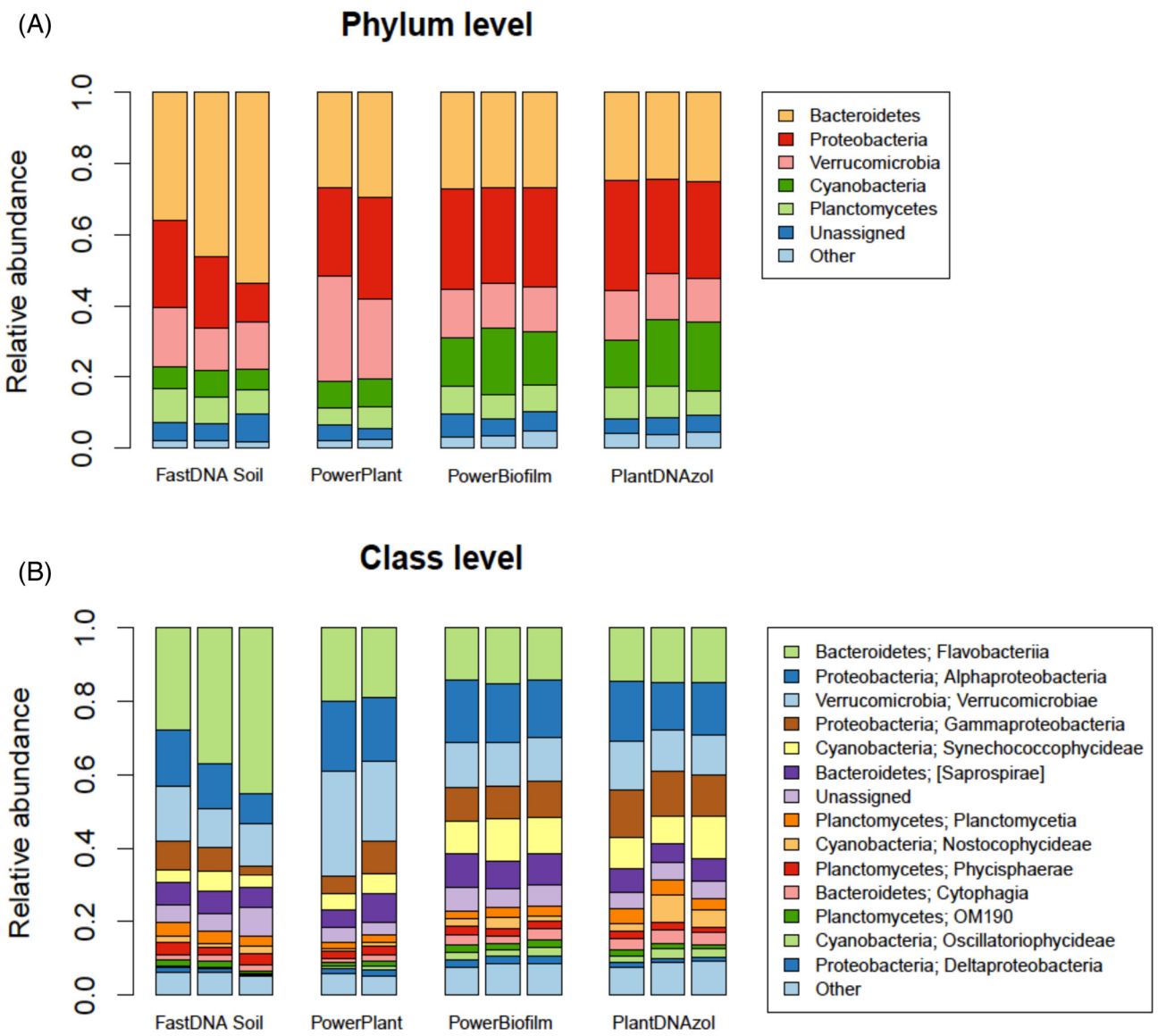

Figure 2. Comparison of relative abundances of 16S rRNA OTUs isolated from the four different extracts, at the phylum (**A**) and class (**B**) levels. Each bar represents one replicate sample.

Another reason that could not be discarded and partially contribute to the observed high abundance of gram-negative bacteria is that gram-negative cells are more susceptible to cell lysis during DNA extraction procedures than gram-positive cells (Theron and Cloete 2000; Felczykowska *et al.* 2015). Although the four different DNA extracts indicated similar abundances for most genera (Table 2), significant differences in presence and/or relative abundance of 20 genera were found (Fig. 3). Genera from the *Verrucomicrobita* class, such as *Rubritalea* and *Persicirhabdus*, or from the *Proteobacteria* class, such as *Rhodobacteraceae*, *Phyllobacteriaceae* and *Arcobacter*, were underrepresented in extracts from the FastDNA Soil method (Fig. 3). Instead, genera from *Cyanobacteria* (i.e. *Nostocaeae*) or *Bacteriodetes* (i.e. *Gaetbulibacter*) were overrepresented in the FastDNA Soil extract. In principle, these differences in community composition might be linked to a varying efficiency of the PCR amplification from extracts with different DNA amounts and/or types of contaminants. This would assume that the amounts and/or types of contaminants differentially affect the binding between primer and template for specific sequences. In our study, however, we have not found any link between DNA yield, contamination or integrity and sequencing results. We hypothesise that the observed differences in estimated community composition might therefore be better explained by dissimilar capacities of the four methods to extract DNA from microorganisms with different cell walls and membrane structures (Bürgmann *et al.* 2001; Carrigg *et al.* 2007).

This study demonstrates that the final estimates of community composition and diversity, derived from 16S sequencing, depend on the DNA extraction method chosen. Similar results have been observed for microbial communities in samples of soil (de Lipthay *et al.* 2004), human microbiome (Yuan *et al.* 2012) or *in vitro* mock communities (Morgan, Darling and Eisen 2010). For biofilms, no clear differences in bacterial diversity among DNA extraction methods were observed when using denaturing gradient gel electrophoresis (DGGE) fingerprints as criteria for comparison (Lyautey *et al.* 2005). However, it is known that high-throughput sequencing provides much deeper insights into the diversity of microbial communities than

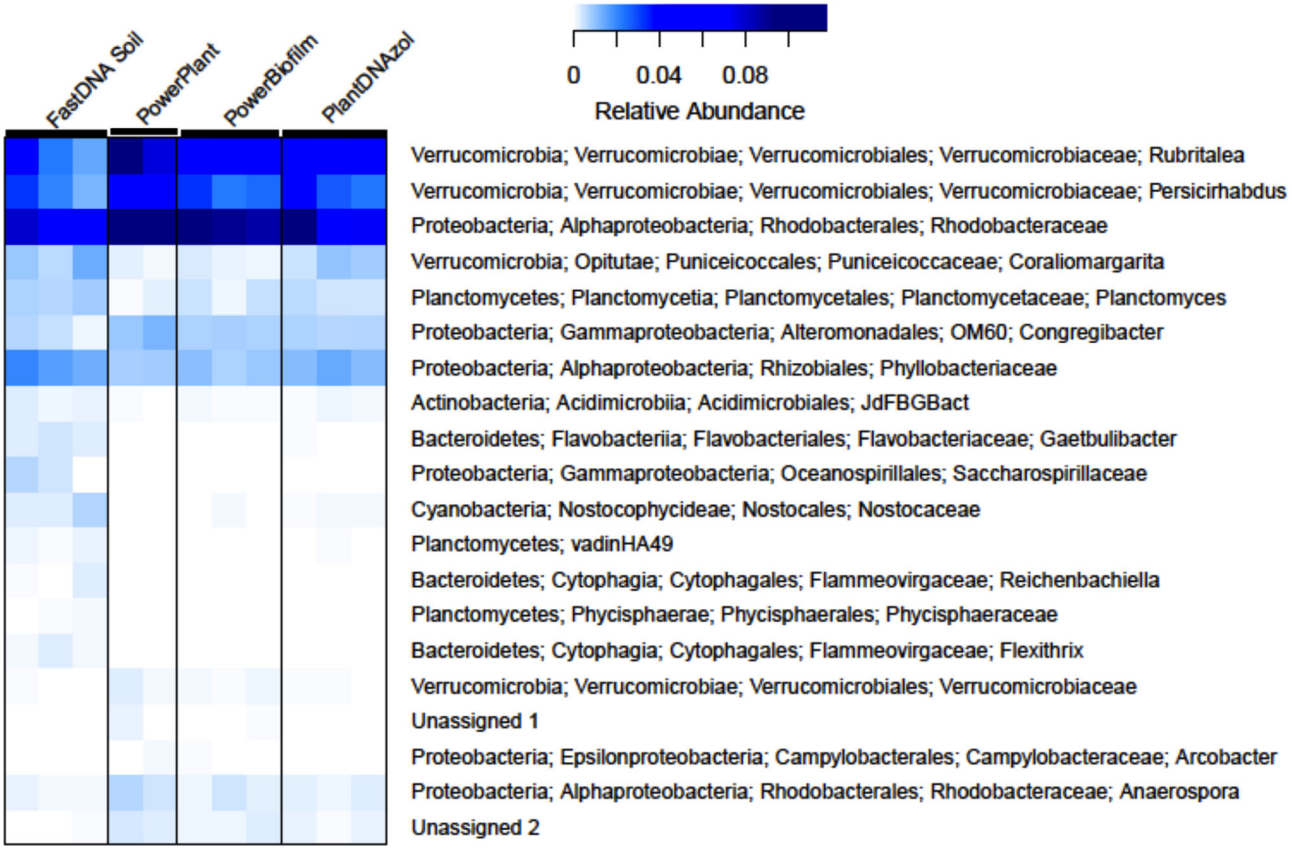

Figure 3. Relative abundances of the 20 bacterial genera showing significant differences ($P < 0.05$ after ANOVA test) between extraction methods. Each lane represents one replicate sample.

fingerprinting techniques such as TRFLP, ARISA or DGGE (Degnan and Ochman 2012). The 'best' DNA extraction method can be defined as the one that yields the highest number of OTUs, while not excluding detection of any taxonomic groups. Such a method generates the highest biodiversity estimates and hence the most complete description of the taxonomic composition of a given community, and it minimises the bias of not extracting DNA from specific taxonomic groups. Failing to extract DNA from specific taxonomic groups is particularly problematic for comparative assessments of community composition, e.g. when performing site-specific assessments or tracing effects of pollution, since changes in specific taxonomic groups can go undetected. Accordingly, based on our results, the PowerBiofilm and the PlantDNAzol methods are the most suitable methods for describing the bacterial community composition in marine biofilm samples using 16S sequencing. The PowerBiofilm method has the additional advantages that it is easier to use, has a higher reproducibility and produces less fragmented and less contaminated DNA. However, at the same time it also produces lower DNA concentrations and could potentially fail to capture small-sized DNA fragments.

In summary, this study shows that the choice of an appropriate DNA extraction method is a critical step in establishing a laboratory protocol for the microbial community characterisation based on high-throughput sequencing techniques or similar DNA-based methods. Important selection criteria to be considered include the amount of available sample to begin with, the DNA fragment size required for downstream analyses, amount and type of impurities expected in the samples. Trial runs with the microbial community of interest, in order to describe potential biases and extraction characteristics of the various DNA extraction methods, are recommended prior to implementing full studies.

SUPPLEMENTARY DATA

Supplementary data are available at *FEMSLE* online.

ACKNOWLEDGEMENT

The authors acknowledge the support by all staff at the Sven Lovén Centre of Marine Sciences–Kristineberg.

FUNDING

This work has been founded by the Swedish research council Formas (NICE, Project No. 2011–1733 and HerbEvol, Project No. 2015–1464).

Conflict of interest. None declared.

REFERENCES

Albertsen M, Karst SM, Ziegler AS *et al*. Back to basics—the influence of DNA extraction and primer choice on phylogenetic analysis of activated sludge communities. *PLoS One* 2015;**10**:e0132783.

Andersson AF, Riemann L, Bertilsson S. Pyrosequencing reveals contrasting seasonal dynamics of taxa within Baltic Sea bacterioplankton communities. *ISME J* 2010;**4**:171–81.

Battin TJ, Kaplan LA, Denis Newbold J *et al*. Contributions of microbial biofilms to ecosystem processes in stream mesocosms. *Nature* 2003;**426**:439–42.

Blanck H, Eriksson KM, Grönvall F *et al*. A retrospective analysis of contamination and periphyton PICT patterns for the antifoulant irgarol 1051, around a small marina on the Swedish west coast. *Mar Pollut Bull* 2009;**58**: 230–7.

Blanck H, Wangberg SA. Validity of an ecotoxicological test system: Short-term and long-term effects of arsenate on marine periphyton communities in laboratory systems. *Can J Fish Aquat Sci* 1988;**45**:1807–15.

Bürgmann H, Pesaro M, Widmer F *et al*. A strategy for optimizing quality and quantity of DNA extracted from soil. *J Microbiol Methods* 2001;**45**:7–20.

Caporaso JG, Kuczynski J, Stombaugh J *et al*. Correspondence QIIME allows analysis of high- throughput community sequencing data Intensity normalization improves color calling in SOLiD sequencing. *Nat Publ Gr* 2010;**7**:335–6.

Carrigg C, Rice O, Kavanagh S *et al*. DNA extraction method affects microbial community profiles from soils and sediment. *Appl Microbiol Biot* 2007;**77**:955–64.

Corcoll N, Acuña V, Barceló D *et al*. Pollution-induced community tolerance to non-steroidal anti-inflammatory drugs (NSAIDs) in fluvial biofilm communities affected by WWTP effluents. *Chemosphere* 2014;**112**:185–93.

Corcoll N, Casellas M, Huerta B *et al*. Effects of flow intermittency and pharmaceutical exposure on the structure and metabolism of stream biofilms. *Sci Total Environ* 2015;**503**:159–70.

Das T, Sehar S, Manefield M. The roles of extracellular DNA in the structural integrity of extracellular polymeric substance and bacterial biofilm development. *Environ Microbiol Rep* 2013;**5**:778–86.

de Lipthay JR, Enzinger C, Johnsen K *et al*. Impact of DNA extraction method on bacterial community composition measured by denaturing gradient gel electrophoresis. *Soil Biol Biochem* 2004;**36**:1607–14.

Degnan PH, Ochman H. Illumina-based analysis of microbial community diversity. *ISME J* 2012;**6**:183–94.

Dixon P. VEGAN, a package of R functions for community ecology. *J Veg Sci* 2003;**14**:927–30.

Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 2010;**26**:2460–1.

Eriksson KM, Antonelli A, Nilsson RH *et al*. A phylogenetic approach to detect selection on the target site of the antifouling compound irgarol in tolerant periphyton communities. *Environ Microbiol* 2009;**11**:2065–77.

Feinstein LM, Woo JS, Blackwood CB. Assessment of bias associated with incomplete extraction of microbial DNA from soil. *Appl Environ Microb* 2009;**75**:5428–33.

Felczykowska A, Krajewska A, Zielińska S *et al*. Sampling, metadata and DNA extraction - important steps in metagenomic studies. *Acta Biochim Pol* 2015;**62**:151–60.

Flemming H-C, Wingender J. The biofilm matrix. *Nat Rev Microbiol* 2010;**8**:623–33.

Fouhy F, Clooney AG, Stanton C *et al*. 16S rRNA gene sequencing of mock microbial populations- impact of DNA extraction method, primer choice and sequencing platform. *BMC Microbiol* 2016;**16**:123.

Frey B, Rime T, Phillips M *et al*. Microbial diversity in European alpine permafrost and active layers. *FEMS Microbiol Ecol* 2017;**92**:1–17.

Georgiou CD, Papapostolou I. Assay for the quantification of intact/fragmented genomic DNA. *Anal Biochem* 2006; **358**:247–56.

Hellal J, Michel C, Barsotti V *et al*. Representative sampling of natural biofilms: influence of substratum type on the bacterial and fungal communities structure. *Springerplus* 2016;**5**: 822.

Jonsson V, Österlund T, Nerman O *et al*. Statistical evaluation of methods for identification of differentially abundant genes in comparative metagenomics. *BMC Genomics* 2016; **17**:78.

Koid A, Nelson WC, Mraz A *et al*. Comparative analysis of eukaryotic marine microbial assemblages from 18S rRNA gene and gene transcript clone libraries by using different methods of extraction. *Appl Environ Microb* 2012;**78**: 3958–65.

Kristiansson E, Hugenholtz P, Dalevi D. ShotgunFunctionalizeR: an R-package for functional comparison of metagenomes. *Bioinformatics* 2009;**25**:2737–8.

Krizman M, Jakse J, Baricevic D *et al*. Robust CTAB-activated charcoal protocol for plant DNA extraction. *Acta Agric Slov* 2006;**87**:427–33.

Lock MA. Attached microbial communities in rivers. In: Ford TE (ed.) *Aquatic Microbiology: An Ecological Approach*. Oxford: Blackwell Scientific Publications, 1993, 113–38.

Lyautey E, Lacoste B, Ten-Hage L *et al*. Analysis of bacterial diversity in river biofilms using 16S rDNA PCR-DGGE: methodological settings and fingerprints interpretation. *Water Res* 2005;**39**:380–8.

Masella AP, Bartram AK, Truszkowski JM *et al*. PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics* 2012;**13**:31.

Michiels A, Van Den Ende W, Tucker M *et al*. Extraction of high-quality genomic DNA from latex-containing plants. *Anal Biochem* 2003;**315**:85–9.

Moreno-Paz *et al*. Environmental transcriptome analysis reveals physiological differences between biofilm and planktonic modes of life of the iron oxidizing bacteria Leptospirillum spp. in their natural microbial community. *BMC Genomics* 2010;**11**:404.

Morgan JL, Darling AE, Eisen JA. Metagenomic sequencing of an in vitro-simulated microbial community. *PLoS One* 2010;**5**:1–10.

Nakayama Y, Yamaguchi H, Einaga N *et al*. Pitfalls of DNA quantification using dnabinding fluorescent dyes and suggested solutions. *PLoS One* 2016;**11**:1–12.

Pascault N, Roux S, Artigas J *et al*. A high-throughput sequencing ecotoxicology study of freshwater bacterial communities and their responses to tebuconazole. *FEMS Microbiol Ecol* 2014;**90**:563–74.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R foundation for statistical computing [WWW Document], 2013. http//www.R-project.org/.

Rosi-Marshall EJ, Kincaid DW, Bechtold HA *et al*. Pharmaceuticals suppress algal growth and microbial respiration and alter bacterial communities in stream biofilms. *Ecol Appl* 2013;**23**:583–93.

Sabater S, Guasch H, Ricart M *et al*. Monitoring the effect of chemicals on biological communities. The biofilm as an interface. *Anal Bioanal Chem* 2007;**387**:1425–34.

Salta M, Wharton JA, Blache Y *et al*. Marine biofilms on artificial surfaces: Structure and dynamics. *Environ Microbiol* 2013;**15**:2879–93.

Sanli K, Bengtsson-Palme J, Nilsson RH *et al*. Metagenomic sequencing of marine periphyton: taxonomic and functional insights into biofilm communities. *Front Microbiol* 2015;**6**:1–14.

Sinclair L, Osman OA, Bertilsson S *et al*. Microbial community composition and diversity via 16S rRNA gene amplicons: evaluating the Illumina platform. *PLoS One* 2015;**10**:e0116955.

Stach JEM, Bathe S, Clapp JP *et al*. PCR-SSCP comparison of 16S rDNA sequence diversity in soil DNA obtained using diferent isolation and purification methods. *FEMS Microbiol Ecol* 2001;**36**:139–51.

Stoeck T, Bass D, Nebel M *et al*. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol Ecol* 2010;**19**:21–31.

Sundbäck K, Linares F, Larson F *et al*. Benthic nitrogen fluxes along a depth gradient in a microtidal fjord: The role of denitrification and microphytobenthos. *Limnol Oceanogr* 2004;**49**:1095–1107.

Theron J, Cloete TE. Molecular Techniques for determining microbial diversity and community structure in natural environments. *Crit Rev Microbiol* 2000;**26**:37–57.

Timoner X, Borrego CM, Acuña V *et al*. The dynamics of biofilm bacterial communities is driven by flow wax and wane in a temporary stream. *Limnol Oceanogr* 2014;**59**:2057–67.

Tlili A, Hollender J, Kienle C *et al*. Micropollutant-induced tolerance of in situ periphyton: establishing causality in wastewater-impacted streams. *Water Res* 2017;**111**:185–94.

Ward DM, Weller R, Bateston MM. 16S rRNA sequences reveak numerous uncultured microorganisms in a natural community. *Nature* 1990;**345**:63–5.

Yergeau E, Lawrence JR, Sanschagrin S *et al*. Next-generation sequencing of microbial communities in the Athabasca River and its tributaries in relation to oil sands mining activities. *Appl Environ Microb* 2012;**78**:7626–37.

Yuan S, Cohen DB, Ravel J *et al*. Evaluation of methods for the extraction and purification of DNA from the human microbiome. *PLoS One* 2012;**7**:e33865.

Zipper H, Buta C, Lämmle K *et al*. Mechanisms underlying the impact of humic acids on DNA quantification by SYBR Green I and consequences for the analysis of soils and aquatic sediments. *Nucleic Acids Res* 2003;**31**:e39.