



Phospholipid fatty acid (PLFA) analysis as a tool to estimate absolute abundances from compositional 16S rRNA bacterial metabarcoding data

Natascha Lewé^{a,g}, Syrie Hermans^b, Gavin Lear^b, Laura T. Kelly^c, Georgia Thomson-Laing^c, Barbara Weisbrod^d, Susanna A. Wood^c, Robert A. Keyzers^{e,f}, Julie R. Deslippe^{a,g,*}

^a School of Biological Sciences, Victoria University of Wellington, Wellington 6012, New Zealand

^b School of Biological Sciences, University of Auckland, Auckland, New Zealand

^c Coastal and Freshwater Group, Cawthron Institute, 98 Halifax Street, East Nelson, 7010, New Zealand

^d Human and Environmental Toxicology, Department of Biology, Universität Konstanz, 78457 Konstanz, Germany

^e School of Chemical and Physical Sciences, Victoria University of Wellington, Wellington 6012, New Zealand

^f Centre for Biodiscovery, Victoria University of Wellington, Wellington 6012, New Zealand

^g Centre for Biodiversity and Restoration Ecology, Victoria University of Wellington, Wellington 6012, New Zealand

ARTICLE INFO

Keywords:

eDNA
Microbial biomass
PLFA
Environmental monitoring
Biomarker
Environmental substrates

ABSTRACT

Microbial biodiversity monitoring through the analysis of DNA extracted from environmental samples is increasingly popular because it is perceived as being rapid, cost-effective, and flexible concerning the sample types studied. DNA can be extracted from diverse media before high-throughput sequencing of the prokaryotic 16S rRNA gene is used to characterize the taxonomic diversity and composition of the sample (known as metabarcoding). While sources of bias in metabarcoding methodologies are widely acknowledged, previous studies have focused mainly on the effects of these biases within a single substrate type, and relatively little is known of how these vary across substrates. We investigated the effect of substrate type (water, microbial mats, lake sediments, stream sediments, soil and a mock microbial community) on the relative performance of DNA metabarcoding in parallel with phospholipid fatty acid (PLFA) analysis. Quantitative estimates of the biomass of different taxonomic groups in samples were made through the analysis of PLFAs, and these were compared to the relative abundances of microbial taxa estimated from metabarcoding. Furthermore, we used the PLFA-based quantitative estimates of the biomass to adjust relative abundances of microbial groups determined by metabarcoding to provide insight into how the biomass of microbial taxa from PLFA analysis can improve understanding of microbial communities from environmental DNA samples. We used two sets of PLFA biomarkers that differed in their number of PLFAs to evaluate how PLFA biomarker selection influences biomass estimates. Metabarcoding and PLFA analysis provided significantly different views of bacterial composition, and these differences varied among substrates. We observed the most notable differences for the Gram-negative bacteria, which were overrepresented by metabarcoding in comparison to PLFA analysis. In contrast, the relative biomass and relative sequence abundances aligned reasonably well for Cyanobacteria across the tested freshwater substrates. Adjusting relative abundances of microbial taxa estimated by metabarcoding with PLFA-based quantification estimates of the microbial biomass led to significant changes in the microbial community compositions in all substrates. We recommend including independent estimates of the biomass of microbial groups to increase comparability among metabarcoding libraries from environmental samples, especially when comparing communities associated with different substrates.

1. Introduction

Microorganisms play crucial roles in all habitats, especially as drivers of decomposition and nutrient cycling (Aislabie and Deslippe, 2013). In

soil and sediments, they constitute the bulk of the biomass (Balestrini et al., 2011; Torsvik and Øvreås, 2002). Studies of the composition of microbial communities along biotic and abiotic gradients, between habitats or in response to experimental treatments or management

* Corresponding author at: School of Biological Sciences, Victoria University of Wellington, Wellington 6012, New Zealand.

E-mail address: Julie.deslippe@vuw.ac.nz (J.R. Deslippe).

<https://doi.org/10.1016/j.mimet.2021.106271>

Received 23 May 2021; Accepted 14 June 2021

Available online 17 June 2021

0167-7012/© 2021 The Authors.

Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

practices, can help resolve which taxa drive ecosystem processes or are affected by environmental stress (e.g., pH, temperature) (Deslippe et al., 2005; Evans and Wallenstein, 2014). However, to achieve these insights, tools that provide reproducible and comparable data are required. Over the past decade, the application of next-generation sequencing (NGS) to study microbial communities in environmental samples has become increasingly popular (Cristescu and Hebert, 2018). Typically, metabarcoding is utilised, whereby short regions of DNA, especially the 16S rRNA gene region, are analysed by high-throughput DNA sequencing, followed by taxonomic identification. This approach has the potential to characterize complex microbial communities in detail, a significant improvement on culture-based approaches, which may characterize fewer than 1% of bacteria in a sample (Torsvik et al., 1990). Metabarcoding is utilised for a range of applications, including monitoring the occurrence of taxa of interest, characterising biodiversity (Bohmann et al., 2014; Cristescu and Hebert, 2018; Knudsen et al., 2019) and examining changes in community structure along environmental gradients (Berry et al., 2019; Evans and Wallenstein, 2014; Lacoursière-Roussel et al., 2018). Due to its increasing cost-efficiency, particularly for the analysis of relatively large numbers of samples, DNA metabarcoding analysis has become broadly adopted for microbial ecological research.

Despite the widespread use of metabarcoding of the 16S rRNA gene, questions about its reliability as a tool for biodiversity assessment remain (Bonk et al., 2018; Brooks et al., 2015; Harper et al., 2019). Each step of the metabarcoding process can introduce bias (Lear et al., 2018). For example, cell lysis and DNA extraction methods are a significant source of uncertainty (Dopheide et al., 2019; Hermans et al., 2018). Gram-positive bacteria are often thought to be underrepresented in metabarcoding studies because the cell wall of Gram-positive bacteria is composed of a thick peptidoglycan layer, which is more difficult to lyse (Jacobs-Wagner and Cabeen, 2005; Mahalanabis et al., 2009; Martzy et al., 2019). During polymerase chain reaction (PCR), parameters like the condition of the DNA template, GC-content of the template (Nichols et al., 2018), the choice of primers (Elbrecht and Leese, 2015; Parada et al., 2016) and polymerases, and the PCR conditions can induce amplification biases (Dopheide et al., 2019; Kanagawa, 2003; Kennedy et al., 2014). Ideally, each step, from sampling to bioinformatics, requires calibration and validation. To this end, the parallel application of the workflow to a mock microbial community standard and the sample set is frequently used in environmental microbiology (Hermans et al., 2018). Typically, these mock communities are defined mixtures of diverse microbial taxa of varying sizes, GC-content and cell wall composition, used as reference standards for metabarcoding to reveal whether certain taxa are under- or overestimated during the analysis (Hardwick et al., 2017). A further issue concerns the presence of relic DNA which can be found in various states of degradation in most environments (Carini et al., 2016). Depending on the environmental conditions, relic DNA may persist for many years after cell death. It is co-isolated and amplified by polymerase chain reaction (PCR) along with DNA derived from living microbes, potentially biasing biodiversity assessment (Jo et al., 2019; Lennon et al., 2018).

While several studies have assessed the effects of sample processing methods (e.g., DNA extraction, PCR protocols) on metabarcoding results for microbial communities (Bonk et al., 2018; Dopheide et al., 2019; Hermans et al., 2018), none have compared metabarcoding and analysis of PLFA across substrates from different habitats. Sequencing analyses of microbial communities from marine or freshwater sediments and soil or microbial mats can provide vastly differing views of their composition and structure (Lozupone and Knight, 2007), which is expected due to the intrinsically different physicochemical properties of those substrates. However, there is cause to suspect that some portion of the variation in these microbial communities may be induced by the different protocols used to sample and process substrates (Pollock et al., 2018). For example, the presence of humic substances (Brooks et al., 2015; Harper et al., 2019) and the type and size of particles (Probandt et al., 2018;

Stoeckle et al., 2017) are known to influence metabarcoding results, but the strength and direction of these biases are likely to vary among substrates (Buxton et al., 2017). Likewise, relic DNA occurs in all environments, but its turnover depends on environmental factors such as temperature, light exposure, and chemical properties of the substrate. In soil, relic DNA can persist for years (Nielsen et al., 2007), whereas the degradation to smaller oligonucleotides occurs within days or weeks in sediments from freshwater habitats (Carini et al., 2016; Lennon et al., 2018; Sirois and Buckley, 2019). Several authors have concluded that substrate type and environmental factors need to be taken into account when using DNA-based techniques (Buxton et al., 2017; Sirois and Buckley, 2019; Stoeckle et al., 2017) but to our knowledge, few direct comparisons of metabarcoding biases across substrates from different habitats exist (Hermans et al., 2018). As a consequence, comparisons of multiple metabarcoding studies when samples originate from different habitats is currently fraught with uncertainties, even when a universal methodology for the sample processing had been applied.

Phospholipid fatty acid (PLFA) analysis is used to measure microbial biomass and community composition, especially in sediments and soils (Frostegård et al., 1991; White et al., 1979) but also to characterize bacterial communities in water (Glucksman et al., 2000; Green and Scow, 2000). PLFAs are a major structural constituent of biological membranes, and their fatty acid (FA) components vary in composition among taxa. Some PLFAs are specific to a single microbial taxon and can therefore be used as taxonomic biomarkers, albeit with low taxonomic resolution (Frostegård et al., 2011; Willers et al., 2015; Zelles et al., 1992). One advantage of PLFA analysis in comparison to molecular methods is that it produces quantitative outputs of the PLFAs, which can be applied as proxies for the biomass of the microbial groups affiliated with their respective biomarkers (Kirk et al., 2004). Secondly, because cell membranes are rapidly degraded and metabolised after cell death (Dunfield, 2007), the total amount of PLFAs is an important indicator of the biomass of living microorganisms (Hill et al., 2000; Zhang et al., 2019) in a sample. Furthermore, PLFA analysis is not subject to some of the biases that affect DNA-based approaches, such as variable amounts of relic DNA or humic substances in substrates (Nielsen and Petersen, 2000). However, while some PLFAs are considered as signature biomarkers, many PLFAs are common across taxa, and some PLFAs that are routinely used as biomarkers are found outside their indicated taxon (Frostegård et al., 2011; Ruess and Chamberlain, 2010; Willers et al., 2015; Zelles, 1999). Despite these limitations, the variable concentrations of numerous PLFAs provide quantifiable outputs at a resolution well suited to distinguish changes in the PLFA profile between samples (Ramsey et al., 2006). These advantages make PLFA analysis a useful tool to test for biases in sequencing-based approaches that might emerge from the various physicochemical properties of different substrates.

Including measurements of microbial load when examining microbial community composition strengthens our ability to compare between substrates and studies. In a study on soils, Zhang et al. (2017) quantified bacterial cell numbers by several methods including PLFA and applied these absolute estimates to relative abundances from metabarcoding. They reported that cell estimates correlated well between PLFA analysis and other methods of quantification like flow cytometry and qPCR (Zhang et al., 2017). Similarly, Vandeputte et al. (2017) determined bacterial biomass in stool samples and rarefied the samples to equal sampling depth based on their biomass instead of modifying metabarcoding read count numbers based on the minimum sequencing depth. They proposed that this correction might prove especially useful for comparisons of samples obtained from ecosystems with significantly different species compositions and biomass (Vandeputte et al., 2017). Analogously, PLFA values can be used to correct sequencing read counts to obtain absolute abundances of taxa. A difficulty when working with sequencing data is that it is inherently compositional, that is, changes in the abundance of taxa are always relative, which complicates their interpretation. For example, while changes in the relative abundance of taxonomic groups are evident, it is not possible to resolve whether this

change reflects the growth of one group or the reduction of co-occurring taxa in response to treatments or time (Morton et al., 2019). By correcting for the initial bacterial load per microbial group, for example using PLFA as a proxy for their microbial biomass, count data of sequencing reads becomes more valuable for many applications, particularly where samples need to be compared across studies.

Here, we investigate the influence of different substrate types on metabarcoding outputs by applying PLFA analysis in parallel. Our approach was to test how well PLFA- and metabarcoding-based abundance estimates correlate across substrates with different properties. Furthermore, we compare two methods of PLFA analysis that differ in their number of PLFA biomarkers to examine how the number of PLFA biomarkers affect the alignment between PLFA and metabarcoding-based data across different substrates. Finally, we demonstrate how the quantitative measurement of microbial biomass by PLFA analysis can be combined with the relative abundances of microbial taxa obtained by 16S rRNA gene metabarcoding to allow for a meaningful comparison of samples with different microbial loads.

2. Methods

2.1. Sample collection

To compare the relative ability of metabarcoding of the 16S rRNA marker gene and PLFA analysis to characterize bacterial communities we sampled five different substrates: forest soil, lake sediment, stream sediment, river microbial mats, and stream water (Fig. 1). Soil, stream water and stream sediment samples were sampled in the Kelly Stream catchment, West Auckland, New Zealand ($-36^{\circ}53'53.8''S$, $174^{\circ}32'17.3''E$), as described in Hermans et al. (2019). The Kelly Stream catchment comprises of a podocarp-broadleaf forest dominated by kauri (*Agathis australis* (D. Don) Loudon) and includes a ~ 1.4 km reach of a stream. Briefly, forest soil samples were collected with a 2.5 cm diameter soil corer to a sampling depth of 10 cm, after removing leaf litter and plant biomass, and each sample thoroughly mixed in a sealed plastic bag. All sampling sites were at least 50 m apart. Microbial cells from water samples were obtained by filtering 1 L stream water per sample through a 0.22 μm polyvinylidene difluoride filter (Merck Millipore) for lipid extraction, whereas for each DNA extraction, 0.6 L of stream water was filtered through a 0.22 μm polyethersulfone filter (Merck Millipore). Stream sediment samples were obtained by using 50 mL centrifugation tubes to scoop surface sediment to a depth of ~ 5 cm. To obtain benthic

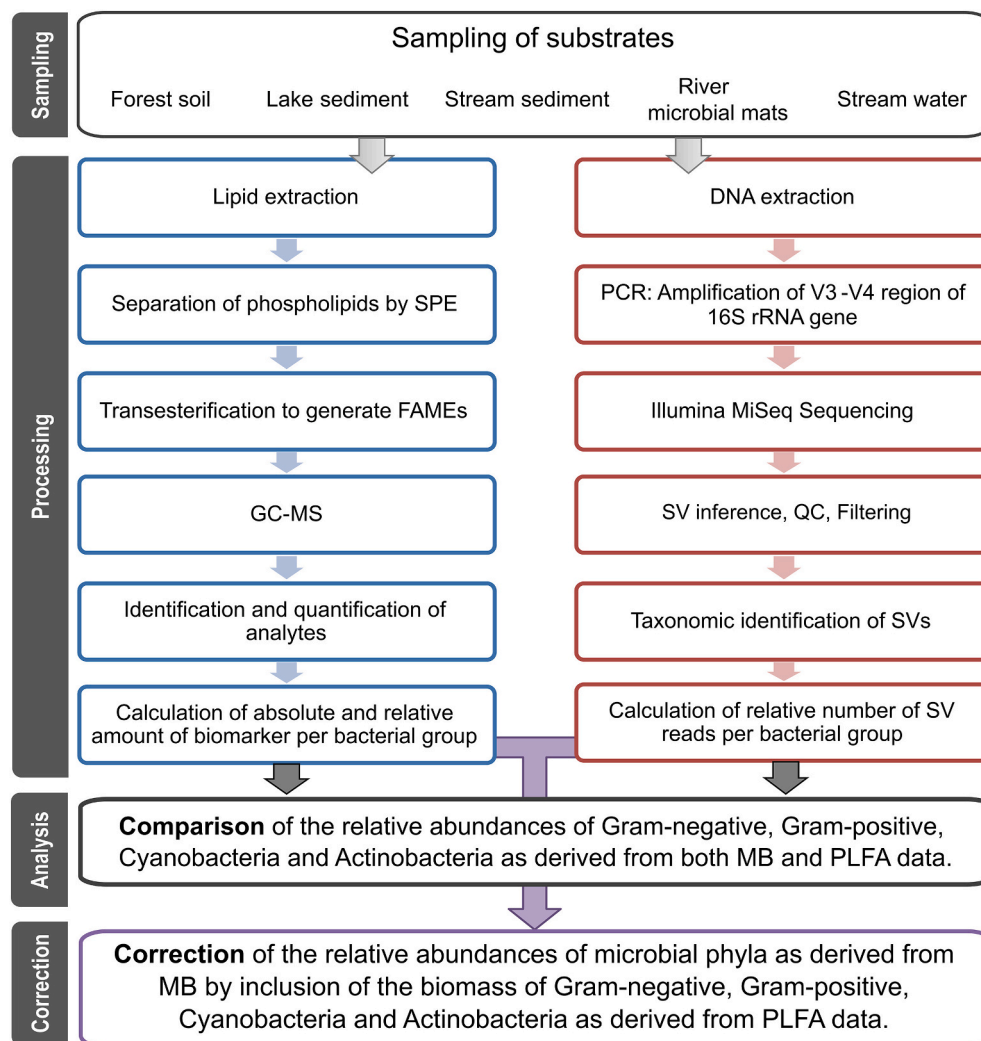


Fig. 1. Overview of the sampling, processing, and analysis of environmental microbes by phospholipid fatty acid (PLFA) and metabarcoding analyses. SPE = solid-phase-extraction, FAME = Fatty acid methyl esters, GC-MS = gas chromatography-mass spectrometry, PCR = polymerase chain reaction, QC = quality control, SV = sequence variants, MB = metabarcoding.

microbial mat samples, a 2 cm² portion of the mat was scraped from rocks in the Hutt River, New Zealand (−41°13′60.00″S, 174°53′59.99″E), and each placed into individual 15 mL centrifugation tubes, as described in Thomson-Laing et al. (2020). Five replicates were obtained at each of the six sites situated along a length of 23 km of the Hutt River. Using a sediment core sampler (UWITEC, Mondsee, Austria), sediment samples from 12 sites covering the area of Lake Rotorua, Canterbury, New Zealand (−42.24′21″94″S, 173°34′52.76″E) were obtained. As described in Weisbrod et al. (2020) the top 0.5 cm layer from the cores were separated for further processing. Lake sediment samples were collected in three replicates at each site. All samples were transported to the laboratory on ice. Subsamples of each substrate were aliquoted for either DNA extraction or lipid extraction and stored at −20 °C until required. Additionally, three replicates of a microbial community standard (mock community) consisting of a mix of known composition of eight bacterial strains and two yeasts were included as samples (ZymoBIOMICS Microbial Community Standard, Zymo Research, US).

2.1.1. Phospholipid fatty acid analysis

2.1.1.1. Sample preparation. All samples were lyophilized overnight (FreeZone 2.5 L Benchtop Freeze Dryer, Labconco, US) and their dry weight determined before applying a high-throughput PLFA method, modified from Buyer and Sasser (2012). Briefly, soil, microbial mat and sediment samples were weighed directly into 10 mL screw-cap glass tubes. For the soil and sediment samples, about 500 mg of substrate was used, whereas we used about 50 mg samples of microbial mats to account for their greater biomass contribution per unit weight. Each entire filter for the freshwater samples was placed into the screw-cap glass tubes for lipid extraction. For the three replicates of the mock community, 150 µL were measured into the glass tubes. Lipids were extracted from each substrate using a chloroform:methanol:phosphate-buffer (1:2:0.8, v/v/v, pH 7.4) mixture, that was spiked with the phospholipid 1,2-dinonadecanoyl-*sn*-glycero-3-phosphocholine (Avanti Polar Lipids, Inc., US) as an internal standard (2.5 nmol per sample). The samples were vortexed, sonicated for 15 min at 35 kHz (Kudos SK_GT Heating Ultrasonic Cleaner, Alphatec, NZ) and shaken for 2 h at 25 rpm (Rocker II, model 260350, Boekel Scientific, USA). After addition of equal volumes of water and chloroform, the lipid-containing phase was removed and dried under a stream of nitrogen gas. The total lipid extract was dissolved in 1 mL chloroform and fractionated into neutral lipids, glycolipids and phospholipids with chloroform, acetone and a 5:5:1 methanol:chloroform:water solution (Findlay, 2004) on a silica column (50 mg silica per column, column capacity = 1 mL, Thermo Fisher, NZ). The eluted phospholipids were derivatized by alkaline methanolysis (0.1 M KOH, 0.2 mL) to generate fatty acid methyl esters (FAMES). After neutralisation with acetic acid, FAMES were extracted twice with 0.4 mL chloroform and the two extracts combined. The resulting extract was concentrated under a nitrogen stream until dry and then dissolved in 75 µL hexane for further analysis.

2.1.1.2. Gas chromatography-mass spectrometry (GC-MS). FAMES were separated on a Restek RXi5-Sil (30 m × 0.25 mm × 0.25 µm) capillary column in a Shimadzu gas chromatograph equipped with a mass spectrometer (Shimadzu GC-MS-QP2010 Plus). The injection temperature was 260 °C and a split ratio of 1:10 was used, with ultra-high purity helium as carrier gas at a flow rate of 1.06 mL·min^{−1} and a linear velocity of 39.6 cm·s^{−1}. A linear temperature gradient with an initial oven temperature of 140 °C increasing to 240 °C at a rate of 3 °C·min^{−1}, followed by a hold of 3 min was applied. The electron impact (EI) mass spectrometer operated at an ionisation energy of 70 eV and scanned a range from 40 to 600 m·z^{−1} (0.3 s·cycle^{−1}). GC-MS solutions version 4.44 (Shimadzu) was used to control the instrument, and for the qualitative and quantitative analysis of the resulting chromatograms.

2.1.1.3. Identification of fatty acid methyl esters (FAMES). FAMES were designated using the ω-reference system as described by Bååth et al. (1992) according to the position of the first double bond, counted from the methyl end (ω end) of the carbon chain. For a detailed description of the designation system and a list of FAMES resolved in this study, see Table S1.

To identify the FAMES, the mass spectra and retention times were compared with those of 37 commercially available standards (Matreya, USA; Nu-Chek, USA). Calibration curves for 25 of these, consisting of representatives from each structural group (saturated, isobranched, anteiso-branched, cyclic, monounsaturated, polyunsaturated) were analysed to obtain the response factor (RF) for each FAME. The RF is required to calculate the quantity of the analytes in response to the detector used (Dodds et al., 2005), in this case a mass spectrometer. For FAMES that could not be commercially obtained, the RF from the structurally most related FAME standard was applied for the correction. For each FAME, the ratio of its RF to the RF of the FAME 19:0 was calculated to obtain the relative response factor (RRF) of the FAME of interest. Because lipid is lost at each step of the PLFA analysis, the RRF is required to determine the concentration of the PLFA of interest in the original sample. FAMES were identified using the GC-MS solutions version 4.44 (Shimadzu) and the amount of analyte per gram dry weight sample was calculated (nmol·g^{−1} dry weight substrate). For FAMES that could not be identified, we determined the percentage peak area per sample to approximate the amount of unidentified FAMES in each profile. It was not possible to calculate the exact amounts in nmol for unidentified FAMES because an RRF could not be determined for them. Therefore, we used the chromatogram peak area to assess the total amount of the FAMES that could not be identified to evaluate how much of the PLFA profile remained unused in our analysis and whether this value differed among substrates. All calculations were performed in R v.3.6.1 (R Core Team, 2016). Although FAMES are analysed with GC-MS, the substances of interests are FAs, derived from extracted phospholipids. Therefore, we use only the term PLFA to describe these substances hereafter.

2.1.1.4. Biomarker designation. To assess whether the number of PLFAs resolved affects the alignment among PLFA and DNA metabarcoding views of microbial community structure, we compared two common approaches of PLFA analysis that differed in the number of PLFA biomarkers (Table 1). The first set, hereafter “PLFA1”, is a conservative approach that utilizes a small set of clearly designated biomarker PLFAs, i.e., those that are both common and abundant in single taxonomic groups (Zelles, 1997; Zelles, 1999). The second set, hereafter “PLFA2”, contains additional, mostly monounsaturated, FAs. This expanded set of PLFAs enables a higher resolution of the PLFA profile by taking into consideration the large variability in the composition of microbial phospholipids across all taxa. However, a drawback to the inclusion of these additional PLFAs is that sample origin must be taken into account because the designation of a PLFA as a biomarker for a microbial group may differ among substrates. For this reason, the number and designation of biomarkers for set PLFA1 and PLFA2 differ depending on the sample origin, which in this study is either forest soil or a freshwater ecosystem. For example, typical fungal PLFA biomarkers like 18:2ω6 and 18:3ω3 are also common in the phylum Cyanobacteria. These PLFAs may therefore be used as biomarkers of Cyanobacteria in samples that are known to contain negligible fungal biomass. Likewise, 18:1ω7c is a well-known biomarker for Gram-negative bacteria in soil but can be found in high proportions in members of the phylum Cyanobacteria as well. Here, we used 18:1ω7c as a biomarker for Gram-negative bacteria in biomarker set PLFA1 for both soil and freshwater samples, and also in set PLFA2 for soil. In biomarker set PLFA2 for freshwater samples, the PLFA is used for Cyanobacteria (Willers et al., 2015).

2.1.1.5. Statistical analyses. To compare the microbial biomass of each

Table 1

Phospholipid fatty acid (PLFA) biomarker designation for soil and freshwater samples. PLFA1 and PLFA2 represent two common approaches to quantify microbial biomass in environmental substrates. PLFA1 is the more conservative approach, which employs only PLFAs as biomarkers that are either unique to, or distinctly more common in, a specific taxonomic group. In comparison, the PLFA2 set of biomarkers attempts to employ a large proportion of the microbial PLFAs to increase the taxonomic resolution per sample. Depending on the sample origin and PLFA approach, some PLFA have diverging designations (shown in bold).

PLFA approach	Soil		Freshwater (water or sediment)	
	PLFA1	PLFA2	PLFA1	PLFA2
Bacteria - general marker	17:0	14:0	17:0	14:0
		15:0		15:0
		16:0		16:0
		17:0		17:0
		18:0		18:0
Gram-positive bacteria	<i>a</i> 15:0	<i>i</i> 15:0	<i>i</i> 15:0	<i>i</i> 15:0
	<i>i</i> 15:0	<i>a</i> 15:0	<i>a</i> 15:0	<i>a</i> 15:0
	<i>a</i> 16:0	<i>i</i> 16:0	<i>i</i> 16:0	<i>i</i> 16:0
	<i>i</i> 16:0	<i>a</i> 16:0	<i>a</i> 16:0	<i>a</i> 16:0
	<i>a</i> 17:0	<i>i</i> 17:0	<i>i</i> 17:0	<i>i</i> 17:0
	<i>i</i> 17:0	<i>a</i> 17:0	<i>a</i> 17:0	<i>a</i> 17:0
Gram-negative bacteria	18:1ω7c	2OH10:0	18:1ω7c	2OH10:0
	<i>cy</i> 17:0	2OH12:0	<i>cy</i> 17:0	2OH12:0
	<i>cy</i> 19:0	3OH12:0	<i>cy</i> 19:0	3OH12:0
		2OH14:0		2OH14:0
		3OH14:0		3OH14:0
		2OH16:0		2OH16:0
		3OH16:0		3OH16:0
		16:1ω7c		16:1 ω 7t
		16:1 ω 7t		16:1 ω 5t
		16:1 ω 5t		16:1 ω 5c
		18:1ω7c		18:1 ω 7t
		18:1 ω 7t		19:1 ω 9c
		19:1 ω 9c		<i>cy</i> 17:0
	<i>cy</i> 17:0		<i>cy</i> 19:0	
	<i>cy</i> 19:0			
Actinobacteria	10Me16:0	10Me16:0	10Me16:0	10Me16:0
	10Me17:0	10Me17:0	10Me17:0	10Me17:0
	10Me18:0	10Me18:0	10Me18:0	10Me18:0
Cyanobacteria	–	–	16:3 ω 4	16:1ω7c
			18:2 ω 6	18:1ω7c
			18:3 ω 3	18:1 ω 9c
				18:2 ω 6
				16:3 ω 4
References	(Ahlgren et al., 1992, Bååth et al., 1992, Dijkman et al., 2010, Francisco et al., 2016, Frostegård et al., 1993, Willers et al., 2015, Zelles, 1997)			18:3 ω 3

substrate, we determined the amount of PLFA biomarkers per microbial group (Gram-negative, Gram-positive, Actinobacteria and Cyanobacteria) and of all PLFAs (total PLFA), using the PLFA2 method which allowed for the integration of the largest number of PLFAs. For each sample, the amounts in nmol per g dry weight substrate of all biomarkers per taxonomic group were summed. The mean and standard deviation of the total PLFA per substrate were calculated in R 3.6.3 (R Core Team, 2016). As an indication of potential information loss, we also calculated the approximate percentage of the amount of PLFAs that could not be identified or those that were identified but could not be placed in any biomarker category based on their percentage peak area per substrate.

2.1.2. 16S rRNA gene metabarcoding analysis

2.1.2.1. DNA extraction, PCR and sequencing.

DNA was extracted from each sample aliquot using either the DNeasy PowerSoil DNA Isolation Kit (Qiagen, US) or Water DNA Isolation Kit (Qiagen, US) following the manufacturers' protocols with minor changes for soil, stream sediment and water as described in Hermans et al. (2016). PCR amplification for next generation sequencing was performed using bacterial specific

primers to obtain a c. 400 base pair (bp) amplicon of the V3-V4 region of the 16S rRNA gene. The primers were 341F (5'-TCGTCGGCAGCGTCAG ATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG-3) and 785R (5'-TCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC-3') (Herlemann et al., 2011; Klindworth et al., 2012); and included Illumina adapter sequences (**bold**) that are required for downstream sequencing. All PCR products were sequenced on the Illumina MiSeq platform (Auckland Genomics, New Zealand). For a detailed description of the DNA extraction and the PCR parameters of the respective substrates, see Hermans et al. (2019), Thomson-Laing et al. (2020) and Weisbrod et al. (2020).

2.1.2.2. Bioinformatics.

Bioinformatic analysis of sequence data was performed in QIIME2 2019.4 (Bolyen et al., 2019). First, primers and adapters were removed using cutadapt (Martin, 2011). DNA sequences were denoised and quality filtered with the dada2 algorithm using default parameters (Callahan et al., 2017; Callahan et al., 2016). After trimming, paired end reads were merged to resolve representative exact sequence variants (SVs). Exact sequence variants (also called amplicon sequence variants (ASVs)) benefit from a higher resolution than operational taxonomic units (OTUs), utilizing the high quality of modern Illumina sequencing techniques and are less prone to spurious sequences (Callahan et al., 2016; Edgar, 2017). Samples which contained less than 5% of the mean read abundance of their respective dataset were excluded. A naïve Bayesian classifier was trained on the relevant 16S rRNA gene region of the Greengenes database (McDonald et al., 2012), and used in taxonomic classification of the SVs. Sequences that were ascribed to Archaea, mitochondria or chloroplasts were excluded from further analysis. Furthermore, sequences that could not be assigned Gram affiliation, either because they were unidentified at the phylum level or identified as an undescribed candidate phylum, were removed from further analyses.

2.1.3. Comparison of metabarcoding and PLFA analysis approaches

2.1.3.1. Ratio of Gram-positive and Gram-negative bacteria (GP:GN).

We quantified microbial community composition in the different substrates by three methods: PLFA1, PLFA2 and 16S rRNA gene metabarcoding (MB), generating three views of the microbial community. To evaluate how well these three views align across the substrate types, the relative abundances of microbial groups were calculated using two different approaches. The first was based on the ratio of Gram-positive to Gram-negative bacteria (GP:GN) because it is an important metric that is often used in microbial studies of soil (Fanin et al., 2019; Fierer et al., 2003) and freshwater sediment bacteria (Córdova-Kreylos et al., 2006; Rajendran et al., 1994). The use of the GP:GN metric is common in PLFA approaches (Fanin et al., 2019; Wang and Wang, 2018), but has more recently been applied in metabarcoding approaches (Kumar et al., 2019; Orwin et al., 2018). For the 16S rRNA gene metabarcoding data, the identification of the Gram affiliation was determined at the phylum level, and the relative abundance of both microbial groups was calculated on a per sample basis separately for each method and substrate. For PLFA1 and PLFA2 data, we calculated the amount of the relevant PLFA biomarker (see Table 1) for both microbial groups and divided it by their respective total PLFA, resulting in a proportional view of Gram-positive and Gram-negative bacteria with units of mol%. This value was directly applied as a proxy for the relative abundance of each microbial group (Orwin et al., 2018). A non-parametric Wilcoxon rank sum test was applied to assess whether the three different approaches (PLFA1, PLFA2, MB) produced different views of the GP:GN in the various substrates.

2.1.3.2. Mean abundances of microbial groups.

Microbial communities vary strongly by substrate type. Therefore, the second approach we used was designed to evaluate how well PLFA1, PLFA2 and metabarcoding views align across substrate types, focused on the relative abundances of

microbial groups that are differentially represented across habitats. For soil samples, we focused on the phylum Actinobacteria, whereas in freshwater samples, we quantified the phylum Cyanobacteria. The relative abundances of Actinobacteria, Cyanobacteria, Gram-positive and Gram-negative bacteria were calculated in a similar way as described above. In contrast to the first approach, we deducted the contribution of Actinobacteria and Cyanobacteria from the sequence read counts from the groups they are affiliated with, i.e., Gram-positive bacteria for Actinobacteria and Gram-negative bacteria for Cyanobacteria. We did this to enhance comparability with the PLFA data, for which Actinobacteria and Cyanobacteria were defined by different PLFAs than those used as biomarkers of Gram-positive and negative bacteria generally. For example, the descriptor “Gram-positive bacteria” therefore denotes the remaining Gram-positive bacteria after deducting the proportion of Actinobacteria. To estimate the relative abundance of microbial taxa for the metabarcoding data set, the number of sequence reads per taxonomic group was determined at the phylum level. For each sample, we summed sequence reads per taxonomic group and divided them by the total number of reads per sample as an estimate of the relative abundances of bacterial taxa. For PLFA1 and PLFA2, we calculated the relative abundance of each microbial group in an analogous fashion, as per the biomarker affiliation in Table 1. The resulting values were applied directly as proxy for the relative abundance of each of the microbial groups (Deagle et al., 2019). We applied a non-parametric Wilcoxon rank sum test to reveal if the three different approaches yielded significantly different views of the proportional abundances of these four taxonomic groups of bacteria. All calculations were done in R3.6.3, and the package rstatix used for the statistical tests (Kassambara, 2020).

2.1.4. PLFA-correction of the relative abundances of major phyla

The relative abundances of phyla obtained by 16S rRNA gene metabarcoding do not reflect their absolute abundances. To evaluate if the PLFA-derived biomass estimation can be used to improve metabarcoding data, we used the amount of PLFA as a proxy for the biomass to adjust the relative abundances of the microbial taxa identified in the mock community. We defined the PLFA-corrected relative abundance as the product of the relative abundance of each genus per corresponding microbial group multiplied with the proportion of the biomass of the corresponding microbial group. We compared the resulting PLFA-corrected abundances to the pre-defined theoretical marker gene abundances (hereafter DNA_{th}) of the mock community. To investigate the effects of substrate type on the bias of metabarcoding methodology, we used the biomass of the microbial groups Gram-positive, Gram-negative, Actinobacteria and Cyanobacteria as determined by PLFA1 to correct the relative abundances of microbial phyla in the corresponding metabarcoding dataset. We defined the PLFA-corrected relative abundance as the product of the relative abundance of each phylum per corresponding microbial group multiplied with the proportion of the biomass of the corresponding microbial group. All PLFA-corrected relative abundances were calculated on a per sample basis, and the mean values were calculated for each substrate. We compiled PLFA-corrected proportions of the ten most abundant phyla per environmental substrate and summed the proportions of the remaining phyla into the category “other”.

As a second approach, we focused on how the PLFA-correction changed the bacterial composition among samples of a common substrate. In this analysis, we only used data from the river microbial mats and the lake sediments, for which there were replicates for each sampling location (“site”). For each substrate, we calculated the metabarcoding-based relative abundances of all phyla and the PLFA-corrected relative abundance for each sample as already described. We then multiplied the PLFA-corrected relative abundances with the total PLFA1 value of the sample and calculated the mean values per sampling site. The resulting value is the PLFA-corrected *absolute* abundance which can be used as a proxy for the biomass of the respective microbial phyla.

3. Results

3.1. Sequence data

Across all substrates, 113 metabarcoding samples were obtained after removal of two samples of insufficient sequencing depth. The remaining samples had on average 1089 ± 1089 SVs (mean \pm sd) per sample, of which an average of 3% SVs were excluded as they were non-bacterial. The number of SVs per substrate varied significantly (Kruskal-Wallis rank sum test: $\chi^2 = 66.06$, $df = 4$, $p < 0.001$; Table 2), with stream water yielding significantly more SVs than all other substrates. A summary of the sequence data per substrate type is shown in Table 2, and a detailed overview for each sample is provided in Table S2.

3.2. PLFA data

The substrates varied significantly in their PLFA content and compositions. The total amount of FAs per gram of lyophilized substrate varied strongly depending on its origin. Stream water samples had the lowest total PLFA content with less than $0.01 \pm 0.00 \mu\text{mol}\cdot\text{L}^{-1}$ (mean \pm sd), followed by stream sediments ($0.06 \pm 0.05 \mu\text{mol}\cdot\text{g}^{-1}$) and soil ($0.20 \pm 0.05 \mu\text{mol}\cdot\text{g}^{-1}$). The lake sediments had a total PLFA content of $1.18 \pm 0.35 \mu\text{mol}\cdot\text{g}^{-1}$, and microbial mats had the highest PLFA contents at $4.56 \pm 0.25 \mu\text{mol}\cdot\text{g}^{-1}$. Across all substrates, a total 30 of 35 possible microbial PLFA biomarkers were detected but the number of PLFA biomarkers differed dramatically among substrates (Table 3). Stream water samples contained the smallest number of PLFA biomarkers of all environmental substrates with only 4 of 16 possible PLFAs identified for PLFA1, and 7 of 34 FAs for PLFA2. Furthermore, only three samples contained biomarkers other than the saturated PLFAs, which are indicators for bacteria in general. River microbial mats were somewhat more diverse in their PLFA composition with 10 of 16 possible FAs identified for PLFA1, and 19 of 34 PLFAs for PLFA2. In addition, unidentified PLFAs accounted for an average peak area of 17% of the PLFA profiles of river microbial mat samples. In contrast, freshwater sediments and soil samples contained a much greater diversity of PLFAs. Lake sediment samples contained 14 of 16 PLFAs for PLFA1 and 27 of 34 for PLFA2, and unidentified PLFAs constituted about 25% of the peak area in those profiles. Similarly, the stream sediment samples contained 14 of the 16 PLFA1 biomarkers, and 20 of 34 PLFA2 biomarkers, with 21% of the total peak area comprised of unidentified PLFAs. Soil samples contained 12 of the 13 biomarkers used in the PLFA1 methodology and we detected 21 of the 29 PLFA2 biomarkers. Soil samples had the largest proportion of unidentified PLFAs, comprising, on average, 28% of the total PLFA peak area. The mock community samples were the simplest, with only 7 of 16 PLFAs present using the PLFA1 methodology, and 12 of 35 PLFA2 markers. Unidentified PLFAs accounted for an estimated 18%

Table 2

Summary of the 16S rRNA gene metabarcoding datasets. For each substrate, the number of samples (N), their mean number of sequence variants (SVs), the total number of exact sequence variants per substrate and their read counts are shown. Different letters in superscript denote significant differences among substrates, as determined by a pairwise Wilcoxon rank sum test (Table S3).

Substrate	N	Mean SVs per sample	Number of unique SVs for each substrate	Read counts per substrate after removal of non-bacterial reads
River microbial mats	23	429 ^a	3523	2,155,029
Lake sediment	26	1365 ^c	13,636	1,645,925
Stream sediment	19	768 ^b	9197	473,701
Soil	28	836 ^{ab}	17,414	1,076,835
Stream water	14	2832 ^d	24,682	1,869,716
‘Zymo’ mock community	3	26	41	302,514

Table 3

Phospholipid fatty acid (PLFA) concentrations for five environmental substrates and a mock microbial community standard.

Substrate	River microbial mat	Lake sediment	Stream sediment	Soil	Stream water	Mock community
N	24	26	23	28	14	3
	Concentration of PLFA					
Biomarker	nmol g ⁻¹				nmol L ⁻¹	nmol mL ⁻¹
General bacterial biomarkers						
14:0	141.6 ± 60.1	49.0 ± 13.2	0.2 ± 0.8	1.3 ± 1.2	–	–
15:0	24.2 ± 33.9	60.6 ± 50.8	0.1 ± 0.4	2.0 ± 2.1	–	–
16:0	1179.3 ± 653.3	190.3 ± 44.8	13.4 ± 12.5	37.6 ± 18.5	4.02 ± 1.8	65.9 ± 4.4
17:0	20.2 ± 25.9	35.2 ± 21.4	0.3 ± 0.7	3.2 ± 2.3	–	–
18:0	141.3 ± 85.7	36.7 ± 8.3	4.8 ± 2.8	8.1 ± 3.3	2.2 ± 0.9	9.7 [‡]
SUM	1506.5 ± 768.1	371.9 ± 120.2	18.8 ± 15.7	52.2 ± 24.3	6.2 ± 2.2	70.7 ± 11.2
Gram-positive bacterial biomarkers						
a15:0	9.9 ± 14.2	93.1 ± 29.6	4.3 ± 6.0	17.1 ± 5.9	0.7 [‡]	16.2 ± 2.8
i15:0	153.2 ± 109.8	58.8 ± 29.9	4.8 ± 4.6	19.1 ± 7.1	–	5.7 ± 0.4
a16:0	–	8.3 ± 22.6	0.3 ± 0.8	–	–	–
i16:0	–	30.4 ± 21.6	1.4 ± 1.71	6.6 ± 4.2	0.3 [‡]	3.0 ± 0.2
a17:0	12.0 ± 16.7	27.3 ± 10.5	1.5 ± 1.83	8.1 ± 3.9	–	13.8 ± 1.7
i17:0	12.7 ± 19.0	17.6 ± 8.4	0.9 ± 1.3	5.3 ± 2.3	–	8.8 ± 1.5
SUM	187.8 ± 127.6	235.6 ± 57.4	13.2 ± 14.5	56.3 ± 15.9	1.0	47.6 ± 6.6
Gram-negative bacterial biomarkers						
16:1ω5c	–	2.6 ± 12.2	–	NA	–	–
16:1ω5t	–	11.0 ± 11.2	–	3.3 ± 5.0	–	–
18:1ω7c	–	0.2 ± 1.2 ¹	0.5 ± 0.7 ¹	4.1 ± 1.3	–	–
16:1ω5	41.0 ± 196.5 ^{ur}	3.5 ± 11.8 ^{ur}	3.9 ± 5.4 ^{ur}	–	1.2 ± 1.1 ^{ur}	28.9 ± 4.7 ^{ur}
16:1ω7c	NA	NA	NA	6.1 ± 4.2	NA	–
16:1ω7t	187.8 ± 130.5	64.0 ± 20.4	3.2 ± 6.7	2.1 ± 6.4	–	–
18:1ω7t	5.6 ± 6.1	2.5 ± 2.9	0.6 ± 0.8	4.1 ± 1.3	–	–
19:1ω9c	0.1 ± 0.5	3.7 ± 5.8	–	0.6 ± 0.5	–	19.9 ± 1.9
cy17:0	13.8 ± 17.1	18.2 ± 17.4	1.6 ± 1.4	6.3 ± 2.5	–	22.7 ± 1.8
cy19:0	1.1 ± 1.7	2.9 ± 1.3	5.0 ± 4.2	37.0 ± 8.6	0.4 [‡]	6.9 ± 1.0
3OH14:0	–	1.8 ± 3.6	–	0.3 ± 1.8	–	–
2OH16:0	–	–	–	0.1 ± 0.2	–	–
SUM	249.4 ± 207.8	110.2 ± 37.8	14.4 ± 13.9	64.0 ± 14.0	2.0 ± 2.4	78.3 ± 9.5
Actinobacterial biomarkers						
10Me16:0	–	89.2 ± 78.1	4.6 ± 4.3	14.2 ± 5.5	–	–
10Me17:0	–	1.7 ± 2.7	0.1 ± 0.2	1.9 ± 1.0	–	–
10Me18:0	–	9.1 ± 6.2	0.8 ± 1.3	6.6 ± 2.7	–	–
SUM	–	100.1 ± 86.1	5.5 ± 5.1	22.6 ± 6.9	–	–
Cyanobacterial biomarkers						
16:1ω7c	347.0 ± 185.8	226.9 ± 56.1	3.5 ± 4.5	NA	0.79 ± 0.4	–
18:1ω7c	–	0.2 ± 1.2 ¹	0.5 ± 0.7 ¹	NA	–	–
18:1ω9c	157.8 ± 112.7	94.6 ± 27.2	4.4 ± 4.2	NA	0.4 ± 0.2	–
16:3ω4	337.0 ± 234.2	–	–	–	–	–
18:3ω3	1430.6 ± 1023.5	–	–	–	–	–
18:2ω6	345.0 ± 335.2	44.9 ± 15.6	1.2 ± 1.4	NA	1.0 ± 0.4	–
SUM	2617.4 ± 1501.8	366.7 ± 80.7	9.6 ± 9.9	NA	1.2 ± 0.9	–
Total PLFA1	2335.5 ± 1350.2	442.5 ± 178.5	27.2 ± 26.4	130 ± 31.8	1.7 ± 1.4	77.1 ± 9.4
Total PLFA2	5561.1 ± 2467.5	1184.4 ± 348.3	61.3 ± 54.6	195.1 ± 52.3	6.9 ± 3.2	196.5 ± 27.2
% peak area of excluded FAs	17.3 ± 6.2	24.7 ± 6.2	20.6 ± 8.1	28.4 ± 9.6	15.5 [‡]	18.2 ± 4.5

Notes: NA not applicable: depending on the origin of the substrate, PLFA biomarker have different designation, see designations for PLFA in Table 1. - not detected, ‡ detected in one sample only, ^{ur} unresolved identification regarding stereochemistry, ¹the values for PLFA 18:1ω7c are given twice in the table, for the phylum Cyanobacteria and Gram-negative bacteria, because the PLFA is a designated biomarker for both microbial groups (see Table 1). The hydroxy fatty acids 2OH10:0, 2OH12:0, 3OH12:0, 2OH14:0 and 3OH16:0 were not detected.

of the total peak area in the mock microbial communities.

The PLFA analysis revealed that the substrates differed strongly in their PLFA composition, reflecting differences in the composition of their microbial communities (Fig. 2). For example, the biomarkers for Gram-negative bacteria were the most abundant PLFAs in soil samples, while they were among the least abundant in lake sediment samples where Cyanobacterial and general bacterial biomarkers were dominant. The most pronounced difference in microbial composition among substrates manifested for the phylum Cyanobacteria. For example, in microbial mats the amount of Cyanobacterial biomarkers was up to an order of magnitude greater than the mean amount of biomarker for Gram-positive and Gram-negative bacteria (Table 3). In contrast, the

mean amount of the biomarkers from all microbial groups was similar for the stream sediment. Substrates also varied with respect to which PLFA biomarker was dominant within the set of biomarkers for a microbial group. In the river microbial mats, the polyunsaturated PLFA 18:3ω3 dominated the Cyanobacterial biomarkers and the overall PLFA profile, comprising on average 25.7 mol% of the total PLFA amount. In contrast, soil, stream sediment and lake sediment all had a high mean content of the saturated PLFA 16:0, a common biomarker for bacteria, but differed regarding the relative contents of common mono-unsaturated PLFAs like 18:1ω9t and 16:1ω7 as well as the cyclic PLFAs cy17:0 and cy19:0.

Comparing the total PLFA retained in the PLFA1 and PLFA2 datasets

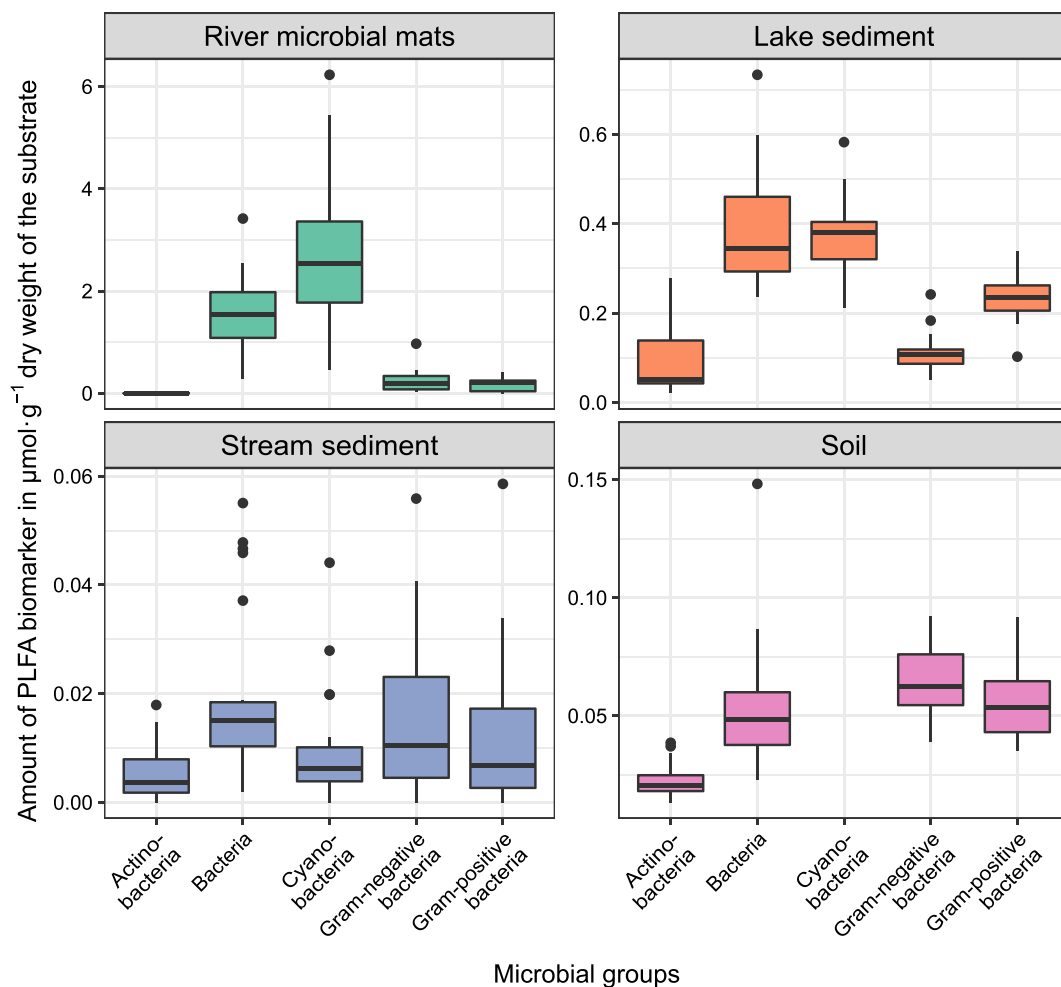


Fig. 2. Total amount of phospholipid fatty acid (PLFA) biomarker for five microbial groups in each substrate. The amount of the PLFAs for each group of biomarkers was calculated in μmol per gram of lyophilized substrate on a per sample basis. Bacterial PLFAs without association to any specific taxonomic group were combined in the group “bacteria”. No PLFA biomarker for Actinobacteria could be detected in the microbial mat samples. In soil, the phylum Cyanobacteria was not determined. Horizontal lines show median values, boxes denote the interquartile range (IQR), while whiskers show $1.5\cdot\text{IQR}$ ranges. Outliers are indicated by dots. Note different y-axis scales among figure panels.

revealed how much of the PLFA profile remained unused when applying the smaller biomarker set. In soil, the PLFA1 approach covered 66% of the total PLFA2 amount, which was the highest coverage in all substrates. In contrast, in water only 25% of the total PLFA2 amount was also included in PLFA1. All sediments showed intermediate percentages of coverage of around 40%. The PLFA1 biomarker set covered 42% of the total amount of PLFA measured by the PLFA2 approach in river microbial mats, 37% in lake sediments and 44% in stream sediments. A more detailed calculation of the percentage areas of the chromatograms that were unused is provided in Table S5.

3.3. Comparison of metabarcoding and PLFA datasets

3.3.1. Ratio of Gram-positive and Gram-negative bacteria (GP:GN)

The estimated relative abundances of Gram-positive and Gram-negative bacteria varied strongly among substrates, as well as between the methods applied for their quantification (Fig. 3 and Table S4). For samples of the mock microbial communities, metabarcoding and PLFA methods provided similar views of the relative abundances of Gram-positive and Gram-negative bacteria. The theoretical relative abundance based on the 16S rRNA marker gene (DNA_{H}) for Gram-negative bacteria in the mock microbial community was 26.4%; metabarcoding and PLFA methods overestimated the relative abundance of Gram-negative bacteria by 12 to 27 percentage points. Whereas the PLFA1

method agreed best with the theoretical value of the proportion of the Gram-negative bacteria, assessment by metabarcoding overestimated this value by 27 percentage points. However, the metabarcoding measurement and that of PLFA2 analysis were similar with values of 54% and 51% respectively, meaning both approaches equally overestimated the proportion of the Gram-negative bacteria in the mock community. In the environmental substrates, metabarcoding and PLFA1 and PLFA2 methodologies yielded significantly different values of the relative abundances of Gram-positive and -negative taxa. For example, the proportions of Gram-negative bacteria estimated to be present in the environmental samples differed by between 31 and 93 percentage points when determined by either metabarcoding or PLFA analysis. In general, metabarcoding analysis returned much smaller predicted proportions of Gram-positive bacteria (Actinobacteria included) in all environmental substrates, and this difference was greatest in river microbial mats (up to 93 percentage points). In sediments, Gram-positive bacteria (Actinobacteria included) were predicted to make up less than 6% of the microbial community when assessed by metabarcoding. In contrast, this group was assessed to comprise between 42 and 94% of the microbial community, depending on the PLFA method and the origin of the sediment. In all substrates, PLFA1 resulted in the highest proportions of Gram-positive bacteria, differing most sharply from the metabarcoding approach. In contrast, the PLFA2 approach aligned better with the relative abundances as estimated by metabarcoding. The removal of SVs

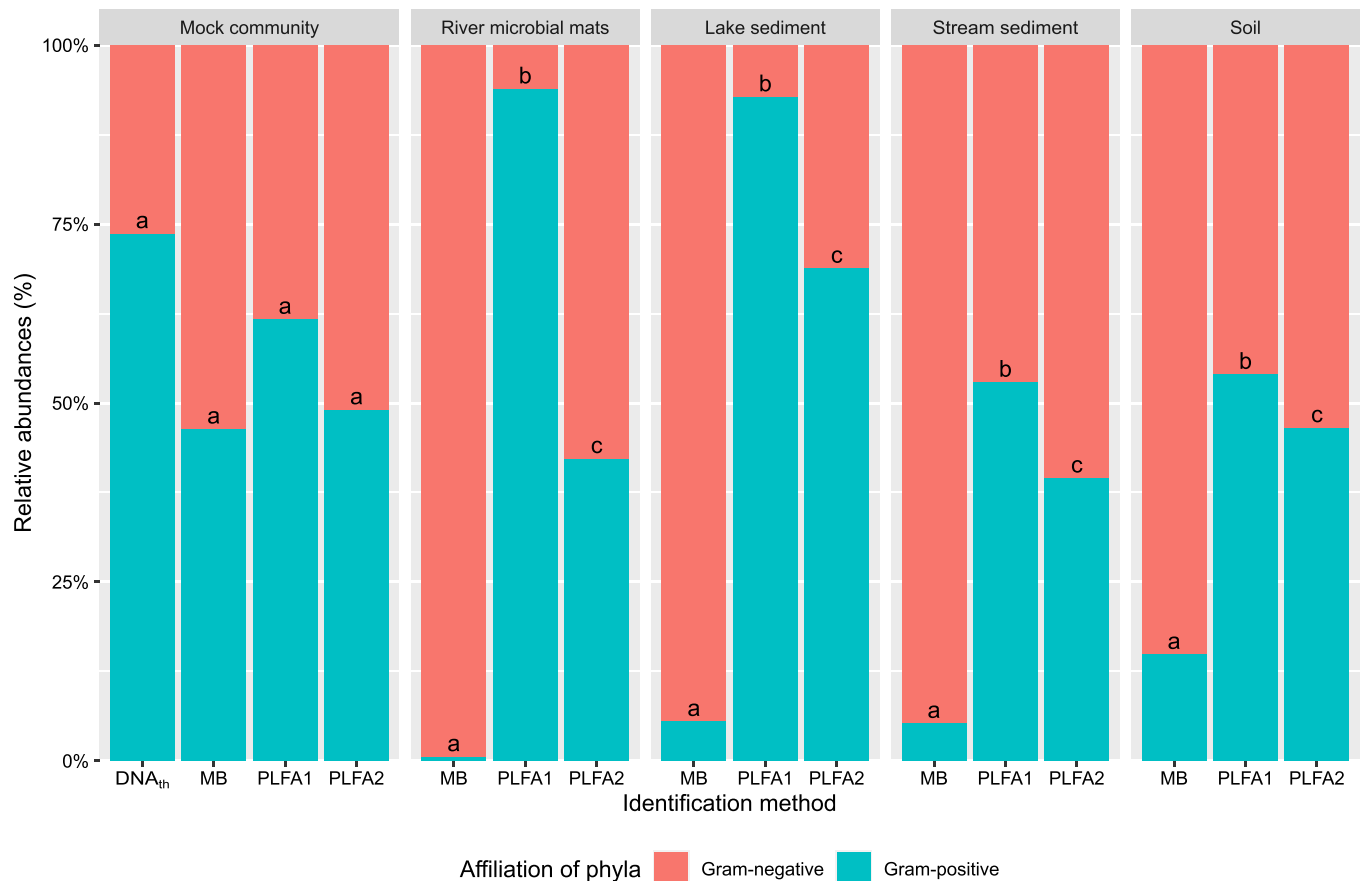


Fig. 3. Mean relative abundances of Gram-positive and Gram-negative bacteria in five substrates as determined using metabarcoding (MB) and phospholipid fatty acid (PLFA) analysis. The first bar on the left (DNA_{th}) indicates the true mock community composition (ZymoBIOMICS Microbial Community Standard) which is based on the relative abundance of 16S rRNA genes expected to be present for each group (Gram-positive and Gram-negative). The other bars represent the relative amount of Gram-positive and Gram-negative bacteria as determined either by MB or by two PLFA approaches (PLFA1, PLFA2) using different sets of PLFAs as biomarkers. For each substrate separately, different letters indicate significant differences between the relative abundances of the bacteria as estimated by the different methods. Detailed results of the applied Wilcoxon rank sum test are shown in Table S3.

that could not be assigned a Gram status because they could not be identified on the phylum level or were affiliated with candidate phyla accounted for $3.0 \pm 3.8\%$ (mean \pm sd) of the total read count (Table S2).

3.3.2. Mean abundances of microbial groups

The extent to which metabarcoding and PLFA views aligned differed strongly among substrates when the microbial community was partitioned into the microbial groups Actinobacteria, Gram-positive bacteria, Cyanobacteria and Gram-negative bacteria (Fig. 4). For all substrates, the relative predicted abundances of Gram-negative bacteria (Cyanobacteria excluded from the group) were much higher when calculated from metabarcoding data than by one of the PLFA methods. Except for samples from river microbial mats, the metabarcoding data suggest that Gram-negative bacteria were dominant in all substrates. However, Gram-negative bacteria were much less abundant when measured by either of the PLFA approaches. For example, in lake sediments, the mean relative abundance of Gram-negative bacteria was 5.6% by PLFA1 and 10% by PLFA2, which is in striking contrast to the view of these samples based on the metabarcoding methodology, yielding a mean abundance of 83.7% Gram-negative bacteria. Analogously, both the yields of Actinobacteria and of Gram-positive bacteria (Actinobacteria excluded) were highest when measured by PLFA1, whereas metabarcoding measurements resulted in the smallest yields for these two microbial groups. In river microbial mats, metabarcoding and PLFA methodologies all indicated a strong dominance by members of the phylum Cyanobacteria.

However, mean percentages ranged from 73.7% (MB) up to 91.6% (PLFA1), with the PLFA2 approach indicating an intermediate value of 86.5%. Overall, our results indicate that metabarcoding and PLFA analyses result in highly different compositional data for microbial groups, as we demonstrate for the Actinobacteria, Cyanobacteria, Gram-positive and Gram-negative bacteria which differ in their estimated relative abundances in some instances by up to an order of magnitude. Except for Cyanobacteria, the results from the PLFA2 approach aligned somewhat better with the metabarcoding data than the PLFA1 approach did.

3.4. Relative abundances: PLFA-correction of metabarcoding-derived estimates of microbial abundances

The correction of the estimated relative abundances of microbial taxa in metabarcoding datasets, which was achieved by applying the PLFA-derived proportion of biomass for the corresponding microbial group, led to changes in the estimated proportions of the major genera in the mock community and also to marked changes in the estimated proportions of microbial phyla in all environmental substrates (Figs. 5, 6 and S1). In the mock community, PLFA1-correction of the metabarcoding data resulted in relative abundances of genera that were more closely aligned with their theoretical values. In the environmental substrates, prior to PLFA-correction, the most abundant phyla were all Gram-negative bacteria (Fig. 6 and S1). Most notable were the Proteobacteria, Bacteroidetes and Verrucomicrobia, especially in lake

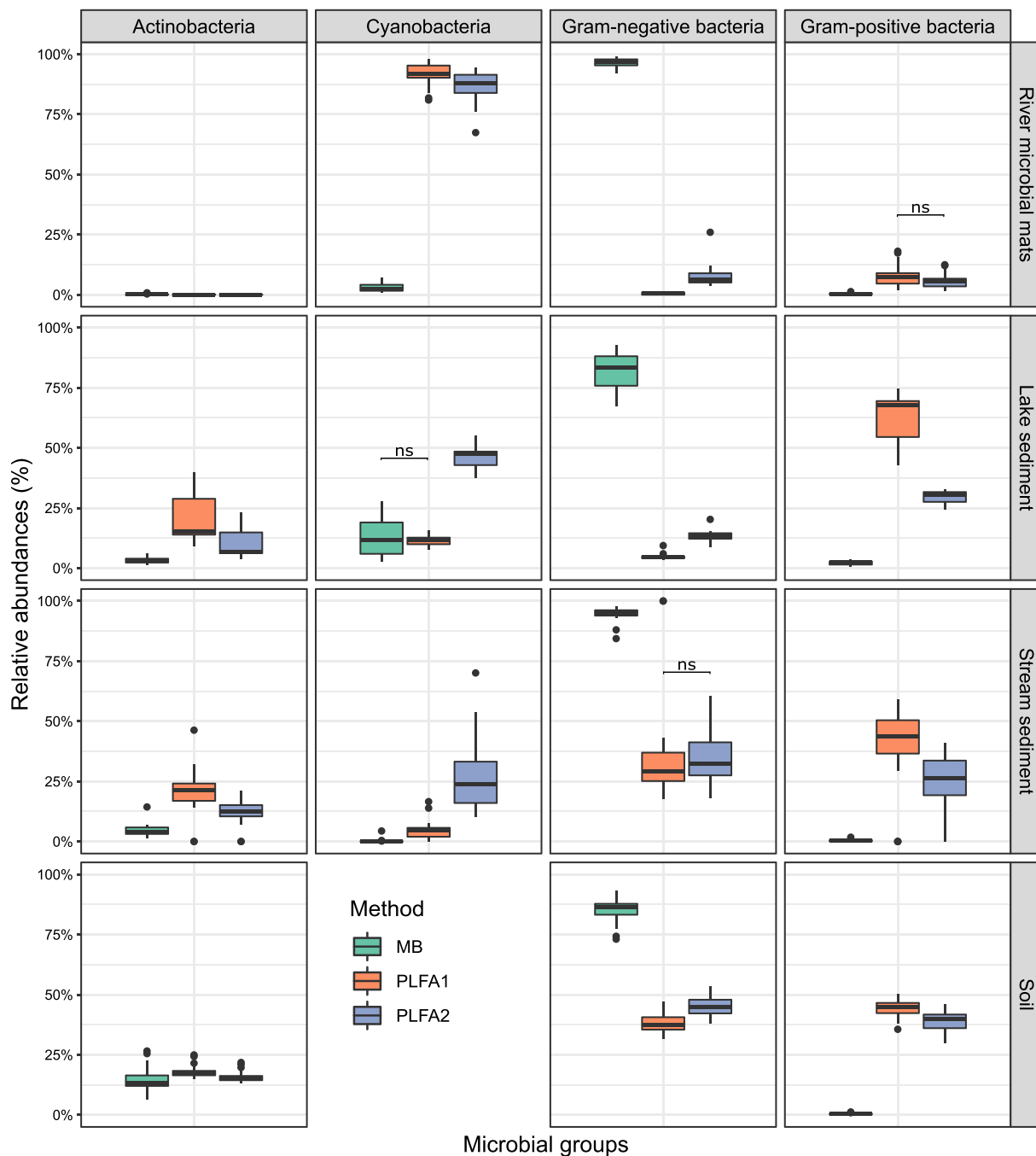


Fig. 4. The estimated relative abundances of four microbial groups (Actinobacteria, Cyanobacteria, Gram-negative bacteria and Gram-positive bacteria) in four substrates as determined by three methods (metabarcoding, PLFA1, PLFA2). Cyanobacteria were excluded from the soil analysis. In river microbial mats, Actinobacteria could not be detected by PLFA analysis. Horizontal lines indicate median values, boxes denote the interquartile range (IQR), while whiskers show the 1.5*IQR ranges. Outliers are indicated by dots. 'ns' indicates a non-significant difference in the predicted relative abundance of microbial groups between two methods as calculated by Wilcoxon rank sum test, all other comparisons are significantly different (see Table S6).

sediments, soils and stream sediments. After PLFA1-correction, however, Gram-positive phyla became the most dominant groups in these three substrates. For example, in the lake sediment samples, the dominant phylum changed from the Gram-negative Proteobacteria to the Gram-positive Firmicutes with PLFA1-correction. In soil and stream sediments, the phyla TM7 and Firmicutes increased notably in their abundance after PLFA1-correction. However, we observed relatively small changes in the composition of microbial communities in river microbial mats when we applied PLFA-corrections to metabarcoding dataset. In river microbial mats Cyanobacteria remained the dominant group after PLFA1-correction, and their estimated abundance increased. PLFA1-

correction also resulted in an exchange of the next most abundant groups from Gram-negative phyla (Proteobacteria and Bacteroidetes) to the Gram-positive phyla Firmicutes and TM7 in the river microbial mat samples. The results of the application of the PLFA2 data are shown in the supplementary material (Fig. S1). Briefly, the correction by PLFA2 also resulted in an increase in the proportion of the Gram-positive bacteria, especially in lake sediments, and in a smaller increase in soil and stream sediment samples. Overall, the shifts in the proportions generated by PLFA2-correction were smaller to those that resulted from the PLFA1-correction, but still considerable. However, the PLFA2-correction resulted in an increase in the estimations of the relative

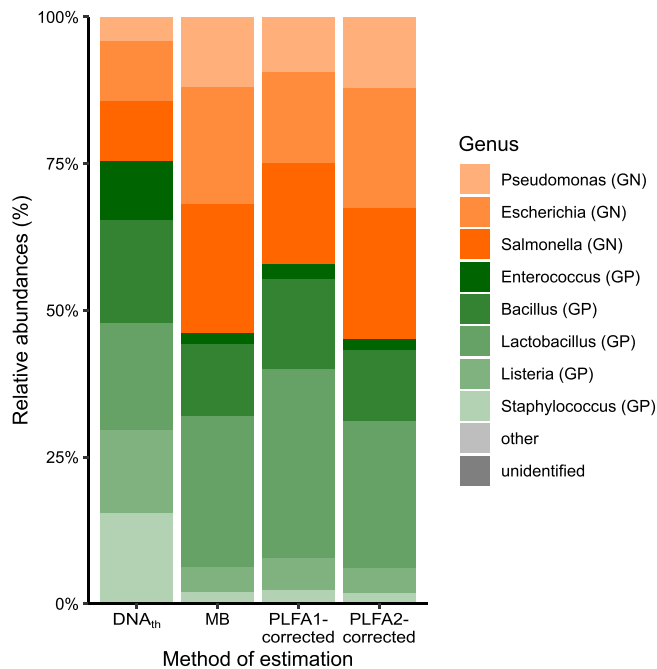


Fig. 5. Relative abundances of the genera from the mock community before and after correction. The theoretical composition of the mock community of the ZymoBIOMICS Microbial Standard is shown on the left (DNA_{th}). The estimated relative abundances as derived from the metabarcoding are depicted in the second bar (MB). The last two bars show the change in genera abundances after biomass correction by either PLFA1 or PLFA2 data. Note: the mock community did not contain any Actinobacteria or Cyanobacteria. MB = metabarcoding, DNA_{th} = theoretical composition of the mock community, based on the marker gene count number, PLFA = phospholipid fatty acid, GP = Gram-positive bacteria, GN = Gram-negative bacteria.

abundance of the Cyanobacteria in all freshwater samples, a shift that did not occur as notably through PLFA1-correction (see Fig. 5).

3.5. Absolute abundances: PLFA1-correction of metabarcoding-derived abundance estimates to explore variation among samples of a common substrate

To gain insight to how spatial variation in microbial biomass may affect metabarcoding views of community composition, we explored how PLFA-correction changed the predicted bacterial composition among samples in common substrates. We focused on river microbial mats and lake sediments, for which we had samples from multiple sites (see Table S2 and details in Thomson-Laing et al. (2020) and Weisbrod et al. (2020)). The PLFA1-correction of absolute abundances of microbial phyla in samples resulted in considerable change in the presumed bacterial composition in both substrates (Figs. 7 and 8). In particular, differences in total microbial biomass among sites restricted to comparison of only proportional abundance data led to false conclusions regarding the similarity of microbial communities among sampling sites in river microbial mats and lake sediments. For example, in river microbial mats, comparison of the bacterial relative abundances among sampling sites would lead to the conclusion that the four sites were highly similar. However, after combining the metabarcoding results with microbial biomass approximation by PLFA1, sites A, B and C showed a higher similarity in their bacterial profile, whereas site D displayed a significant smaller biomass per dry weight substrate. At site D, Gram-positive Firmicutes could not be detected by metabarcoding, but were the second most abundant phylum in sites A, B and C (Fig. 7). Likewise, in lake sediment samples, the inclusion of the PLFA1 data as a biomass proxy resulted in a significant difference in biomass between site A and that of sites B, C and D (Fig. 8). This difference reflected a

pronounced increase in Actinobacterial biomass in site A. In comparison, the estimated relative abundances of Actinobacteria were similar over all four sampling sites before data correction by PLFA1 as a biomass proxy. Another difference in the bacterial profile of the sediment samples is the change in proportion of the Cyanobacteria, for which the estimated relative abundances differed distinctly among the four sampled sites. After PLFA-correction, however, Cyanobacteria showed very similar absolute abundances at all sites. Although the relative abundances of the measured microbial groups showed different trends when compared by PLFA1 and PLFA2, the overall trends when comparing the sites per substrate stayed consistent (see supplementary Figs. S2 and S3). PLFA2-correction produced identical trends as PLFA1 in both microbial mats and lake sediments.

4. Discussion

We investigated the influence of different substrate types on the relative performance of metabarcoding in parallel with PLFA analysis. Metabarcoding and two different PLFA approaches were used to estimate the composition of the microbial communities of five different environmental substrates. The ratios of bacterial groups as determined by metabarcoding and PLFA analysis varied widely among substrates, and these differences were largest for the Gram-negative bacteria. In addition, two different PLFA biomarker sets were used that differed in their number of PLFAs to examine if the increased resolution of the larger biomarker set would lead to a better alignment of the PLFA-based data with the metabarcoding-based data. The PLFA approach based on the smaller set of biomarkers (PLFA1) aligned best with the theoretical abundance values of a mock microbial community and also aligned well with metabarcoding data for the phylum Cyanobacteria in lake sediments. In environmental samples, however, the larger set of biomarkers (PLFA2) aligned better with the results from metabarcoding across most microbial groups and substrates. In what follows, we discuss these results in more detail and consider how the total biomass of the substrate and the particularities of a PLFA profile may influence the resulting view of the microbial community composition as determined by PLFA analysis. Despite the wide variation in bias across substrates when generating metabarcoding views of microbial community composition from environmental samples, and despite additional considerations imposed by methodological choices in PLFA analysis, we argue that insight into the biomass of microbial taxa from PLFA analysis can be used to constrain the pitfalls of assessing microbial community composition from metabarcoding data.

4.1. Metabarcoding vs PLFA analysis: relative abundances of bacteria

In all substrates, the estimated relative abundances of Gram-negative and Gram-positive bacteria differed significantly among methods. These differences were most striking in microbial mats, which consist mainly of Gram-negative Cyanobacteria. Compared to other Gram-negative bacteria, these microbes contain high proportions of characteristic polyunsaturated and monounsaturated PLFAs (Ahlgren et al., 1992; Dijkman et al., 2010). Therefore, our approach to calculate the Gram-positive and Gram-negative bacteria based on their respective PLFA biomarkers only instead of including Cyanobacterial PLFAs as biomarkers for Gram-negative bacteria, does not seem to represent the microbial community of microbial mats well. Future studies that wish to apply PLFA methodology to correct the metabarcoding-derived bacterial composition of microbial mats or of other substrates primarily comprised of Cyanobacteria may prefer to include the Cyanobacterial biomarkers in calculations of Gram-negative bacterial biomass. Our results from sediment samples, however, reflect those from others studies reporting values between 60 and 90% for the percentage abundance of Gram-positive bacteria based on PLFA analysis (Wang and Wang, 2018), and between 10 and 30% based on metabarcoding data (Lee et al., 2020; Wang et al., 2012; Wang and Wang, 2018). In soil, results from

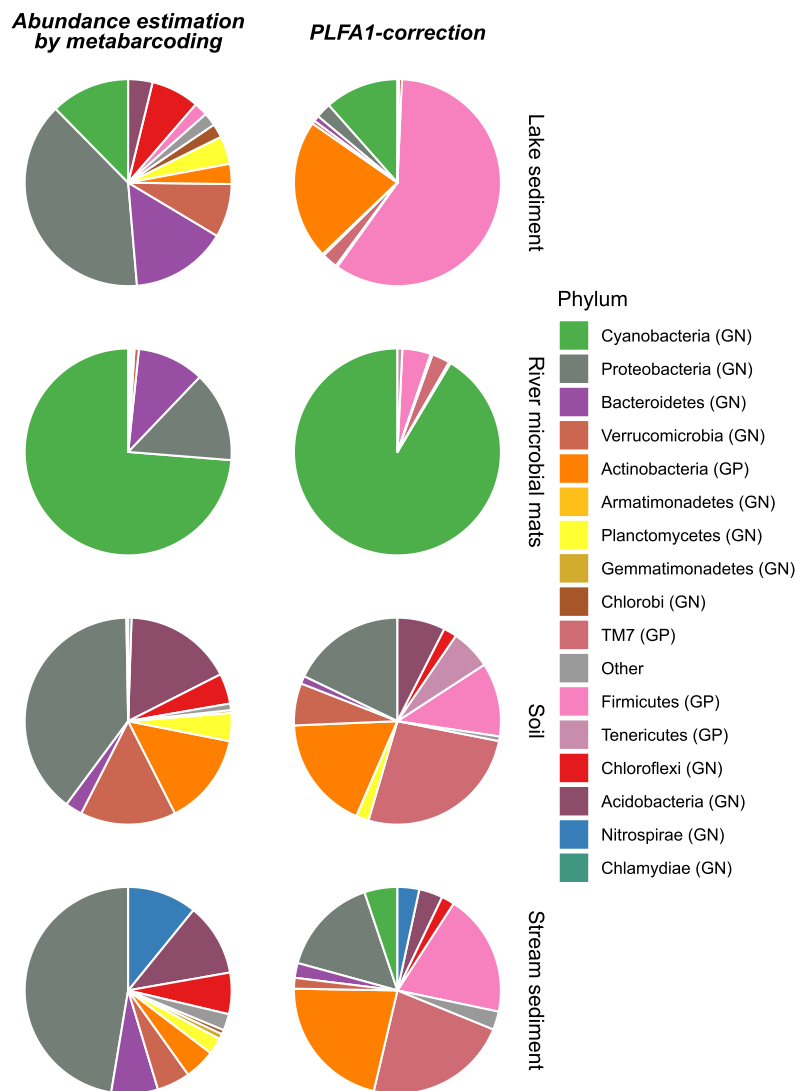


Fig. 6. Predicted relative abundances of major phyla in four substrates as determined by metabarcoding only and corrected by the total biomass of the microbial groups Actinobacteria, Cyanobacteria, Gram-positive and Gram-negative bacteria. For each substrate, the estimated relative abundances as determined by metabarcoding (left) and corrected by the estimated relative biomasses of the microbial groups as measured by PLFA1 (right) are compared. GP = Gram-positive bacteria.

metabarcoding and the PLFA approaches agree better, a result that aligns with other soil surveys (Orwin et al., 2018).

When Cyanobacteria and Actinobacteria were measured independently from other bacteria, their proportions as determined by metabarcoding and PLFA aligned better than for the Gram-negative and Gram-positive bacteria. In each freshwater substrate, metabarcoding and PLFA1 data of the estimated relative abundances of Cyanobacteria were similar, regardless of the dominance of Cyanobacteria in the substrate (e.g., river microbial mats vs. stream sediment samples). These observations indicate that our selection of the PLFA biomarkers for Cyanobacteria were a reasonable choice as a biomass proxy in sediments and microbial mats from freshwater systems. Furthermore, our findings that metabarcoding and PLFA data somewhat align, agrees with others studies, e.g., for microbial mats from hot springs dominated by Cyanobacteria (Zhang et al., 2007). Taken together with other studies, our data suggest that metabarcoding methods are likely to underestimate the Gram-positive bacteria in environmental samples, and that PLFA analyses can help to provide an estimate of this bias.

Previous studies evaluating the validity of using metabarcoding of the 16S rRNA gene to approximate the biomass of taxonomic groups have described several possible biases associated with each step from the

DNA extraction to the bioinformatic analysis (Bonk et al., 2018; Pollock et al., 2018). Biases that have been associated with the Gram status of the bacteria are either based on the higher G + C-richness of the DNA of Gram-positive bacteria or are ascribed to the structural composition of their cells (Frostegård et al., 1999; Nichols et al., 2018). The cell walls of Gram-positive bacteria are more difficult to disrupt compared to those of Gram-negative bacteria, which is a possible source of DNA extraction bias (Miller et al., 1999; Moré et al., 1994; Wu et al., 2010). This could have been a significant contributory factor to our observation that metabarcoding seems to underestimate the abundance of Gram-positive bacteria, regardless of the origin of the samples and the PLFA approach against which the metabarcoding data were compared. In addition, several other sources of error in metabarcoding exist that might impact on the relative abundances of bacterial taxa (Bonk et al., 2018; Dopheide et al., 2019; Stoeckle et al., 2017). Extracted DNA templates might be amplified with varying efficiency due to factors like mismatches between primers and target DNA templates, the G + C content of the template DNA and PCR conditions (Boers et al., 2019; Dopheide et al., 2019). For example, in metabarcoding studies in marine samples (Parada et al., 2016) and soils (Beckers et al., 2016; Thijs et al., 2017) large variation in community composition were found due to difference

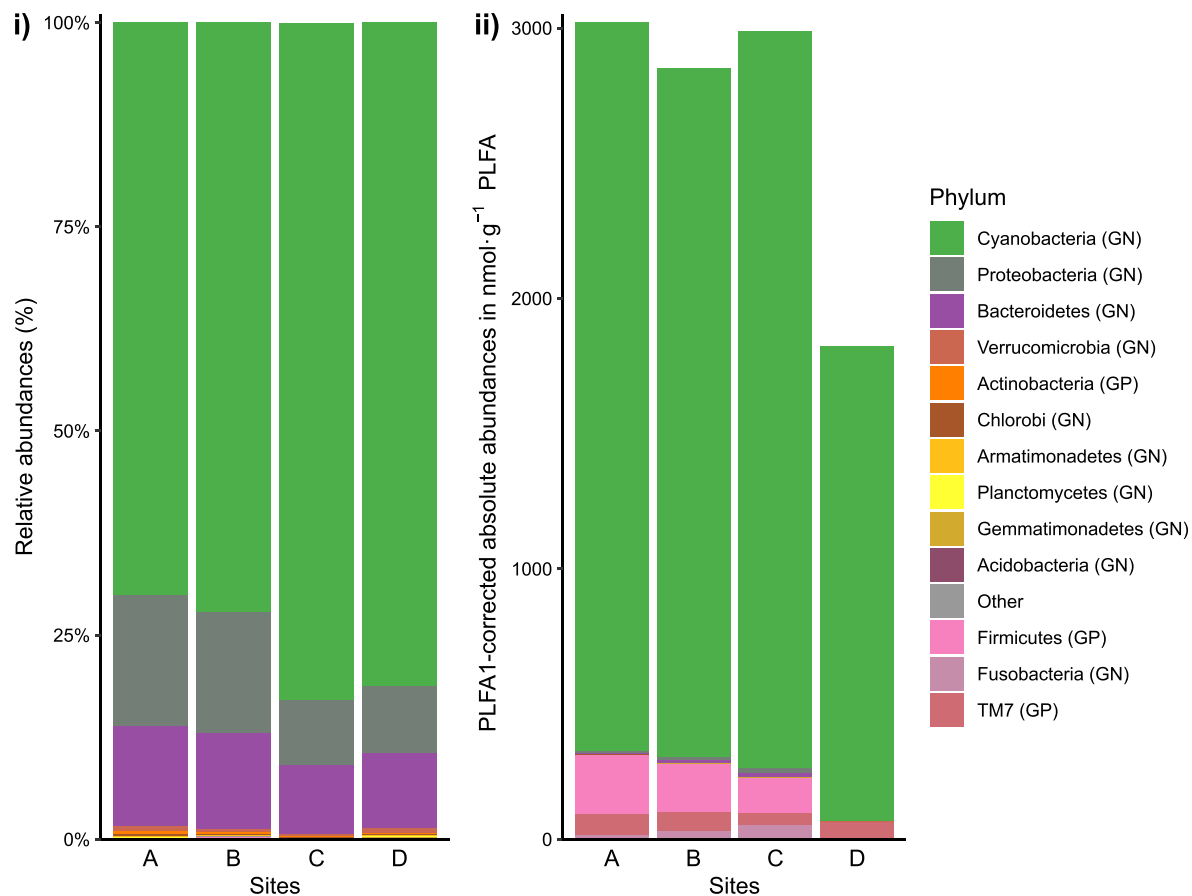


Fig. 7. Relative and absolute abundances of major phyla at sampling sites of river microbial mats. i) relative abundances as estimated by metabarcoding ii) PLFA1 corrected absolute abundances. The values obtained by PLFA1 for Actinobacteria, Gram-positive bacteria, Cyanobacteria and Gram-negative bacteria were included to estimate the absolute abundance of each bacterial phylum. Values are in nmol PLFA per g dry weight of the substrate. Five replicate samples were taken at each site. GP = Gram-positive bacteria, GN = Gram-negative bacteria.

among primer pairs, with some primers greatly under- or overestimating taxa abundances. Similarly, variation in the G + C content of the substrates (Laursen et al., 2017) and variation in PCR conditions (Sze and Schloss, 2019) can lead to substantial differences in metabarcoding views of microbial community composition. Although we found large differences in the relative abundances of Gram-positive bacteria in samples when determined by either PLFA or metabarcoding, we could not determine whether G + C rich (i.e., Actinobacteria) or G + C low Gram-positive bacteria (i.e., Firmicutes) were differentially impacted by the metabarcoding approach. Our study used a polymerase with low error rates when amplifying G + C regions (Nichols et al., 2018), therefore pre-emptively decreasing the chance of PCR bias resulting from choice of the polymerase (Laursen et al., 2017). Taken together, the well-known extraction bias against Gram-positive cells walls seems the most parsimonious explanation for the underestimation of this bacterial group, as reported in this study.

In the mock community, 16S rRNA gene metabarcoding and PLFA methods provided similar views of the relative abundances of the Gram-positive and Gram-negative bacteria, but metabarcoding still underestimated the abundances of Gram-positive bacteria significantly. In contrast, PLFA1 values aligned reasonably well with the theoretical values of the relative abundances. Our results from a mock community seem to reflect the bias against Gram-positive bacteria, which have been described for a range of extraction methods (Hermans et al., 2018), including the commercially available extraction kit used in this study. For example, the theoretical relative abundance of Gram-negative

bacteria in a simple mock community was overestimated by about 20 percentage points when determined by metabarcoding (Hermans et al., 2018), a value that corroborates our own results. However, we found that metabarcoding and PLFA analysis of mock communities aligned much better than that of environmental samples which showed stronger disagreement between metabarcoding and PLFA values. Interestingly, Parada et al. (2016) also reported a stronger bias for Gram-negative bacteria in their environmental samples compared to the mock community they used - a result that supports our findings. These discrepancies in methodological biases between assessments of the mock community and the various environmental samples suggest that the inclusion of a simple mock community to identify biases introduced by the metabarcoding techniques may be of limited use in environmental studies. Mock communities are often included in sequencing approaches to test for biases during method development and are a helpful standard against which run-to-run variations can be measured (Yeh et al., 2018). Readily available, pre-defined mock communities, however, lack the complexity of the microbiota of environmental samples. Ideally, mock communities should be prepared with the sample in mind, i.e., containing a representation of the expected taxa of the specific sample to be suitable to test for biases. Because this is seldom feasible for microbial community profiling in environmental studies, correction of these data following PLFA analysis could help alleviate bias between runs and treatments.

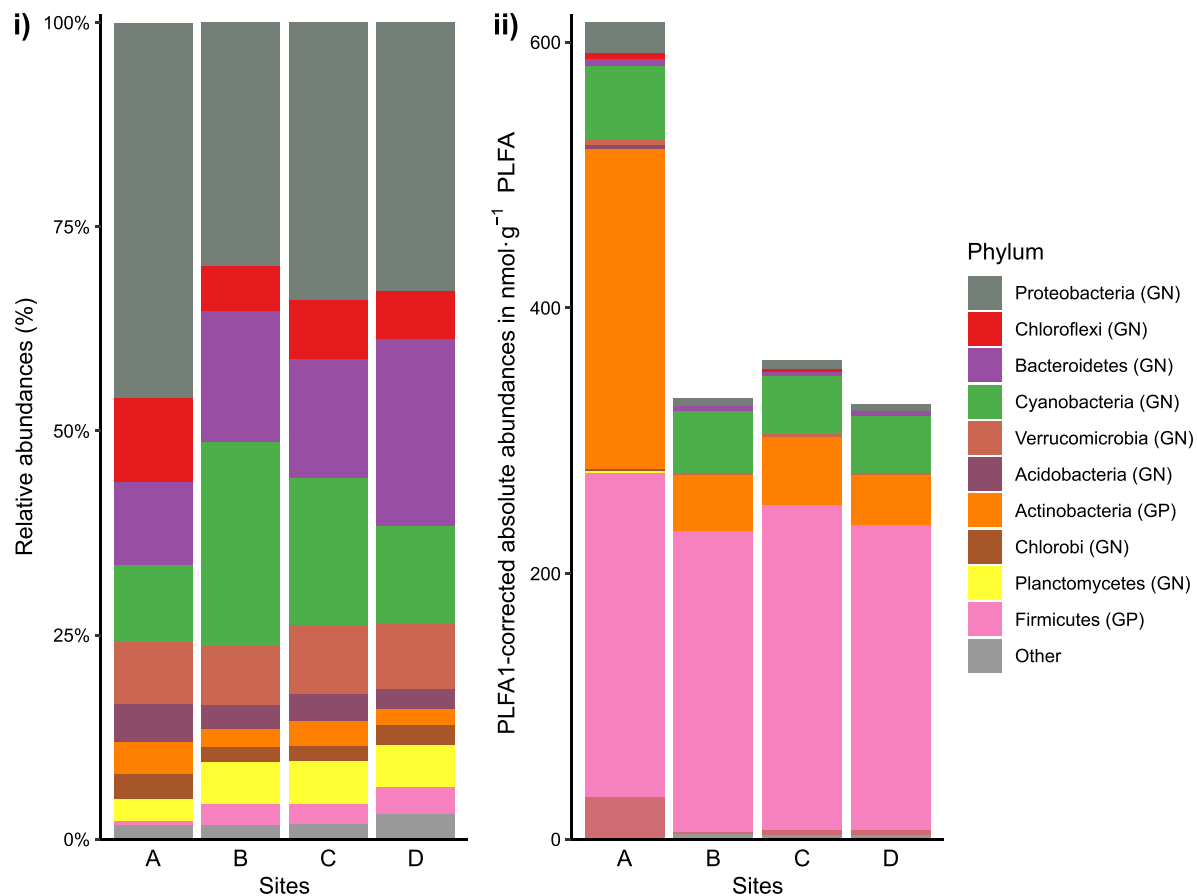


Fig. 8. Relative and absolute abundances of the major phyla at four sampling sites of lake sediment. i) relative abundances as estimated by metabarcoding. ii) PLFA1-corrected absolute abundances. The values obtained by PLFA1 for Actinobacteria, Cyanobacteria, Gram-negative and Gram-positive bacteria were included to estimate the absolute abundance of each bacterial phylum. Values are in nmol PLFA per g dry weight of the substrate. Three replicates were sampled at each site. GP = Gram-positive bacteria, GN = Gram-negative bacteria.

4.2. PLFA analysis: Influence of biomarker selection and origin of substrate

The measurement of PLFA has long been accepted as a quantitative method to determine the biomass of viable microbes in environmental samples (Ritz et al., 2009; Watzinger, 2015). Our determination of total PLFA content per g of each dry substrate indicated that microbial biomass varied by up to two orders of magnitude among the substrates studied. These results were not unexpected due to the inherent properties of the substrates and the specific origin of the samples. Whereas the stream sediment and the stream water samples, which yielded the lowest microbial biomass estimates, originated from an oligotrophic stream (Auckland Council, 2018), the lake sediment was sampled from a hypertrophic lake known to have frequent bacterial blooms (Wood et al., 2017). Lastly, microbial mats formed by photoautotrophs are known to consist almost entirely of microorganisms (Stahl and Klug, 1996). It is therefore consistent that we found the highest PLFA content in these samples and that they were dominated by PLFA biomarkers for Cyanobacteria. Although we can conclude that our results for total microbial biomass fall well within expected ranges for the substrates studied and agree with published data for soils (de Gannes et al., 2016; Francisco et al., 2016) and diverse freshwater sediments (Pratt et al., 2012; Steger et al., 2011), PLFA results are inherently influenced by the number and selection of PLFA biomarkers.

While the selection of specific PLFA biomarkers for microbial groups is common practice in PLFA approaches (Francisco et al., 2016; Rajendran et al., 1994), it is associated with a loss of information from the

PLFA profile, which has the potential to reduce biological insight. To evaluate how much of the PLFA profile remained unused in our analysis, we assessed the percent concentration of the PLFAs that could not be identified. This allowed us to evaluate if the selection of the biomarker PLFAs had the potential to lead to unequal biases across different substrates. For example, in soil, an average of 28% of the peak area remained unidentified, whereas this value was smaller in the different types of sediments, ranging from 17% to 25%. Unidentified PLFAs typically consist of a mixture of many structurally different PLFAs, each at very small concentrations (Kohring et al., 1994; Zelles, 1997). Moreover, these PLFAs are unlikely to be specific to one microbial group, but rather present in small quantities in many taxa (Zelles, 1997). Therefore, in our study, it is unlikely that resolution of the unidentified PLFAs would have substantially altered our view of the overall composition of microbial communities in environmental samples. Nevertheless, our finding that different quantities of unidentified PLFAs exist in different substrates suggests that the comparison of microbial biomass among environmental substrates is, at best, semi-quantitative.

Although the different types of substrates varied greatly in their total PLFA content, this did not appear to influence the detectability of PLFAs over a threshold value. For example, we only detected 19 PLFA biomarkers in the river microbial mats and 20 in the stream sediments, although the total PLFA of the former was more than 90× higher than that of the latter. In contrast, 27 biomarkers were detected in the lake sediment samples, which had a total PLFA content of only about a fifth of the biomass of the river microbial mats. Conversely, in several of the oligotrophic stream water samples, we detected only saturated PLFAs

(16:0 and 18:0), which are typically found in large proportions in bacteria (Zelles, 1997). Thus, it appears that 1 L of filtered stream water provided insufficient biomass to fully resolve the active microbial community in this substrate using PLFA analysis. However, for the environmental substrates with sufficient biomass per sample, we conclude that the number of PLFAs detected was not strongly influenced by the microbial biomass of the sample, but instead depended on the microbial community of the substrate type.

Comparison among studies that employ PLFA analysis is often hampered by inconsistencies in the biomarkers used. While some studies have identified small numbers of signature PLFAs, for example 10 PLFAs as bacterial biomarkers (Orwin et al., 2018), others have employed more than 50 PLFAs to assess the composition of microbial communities (Zhao et al., 2019). One goal of this study was to evaluate how much the interpretation of PLFA profiles varies as a result of the selection of biomarkers, and if the substrate type influences this variation. We applied two sets of PLFA biomarkers that differed in the number of PLFAs resolved and compared the proportions of Gram-positive and Gram-negative bacteria by both methods. This revealed that the number of PLFAs resolved in a PLFA analysis can lead to significantly different views of the composition of a microbial community in all substrates tested. Moreover, we showed that the discrepancy between the PLFA approaches varies depending on the substrate type. For example, while the results of PLFA1 and PLFA2 provided reasonably similar views of the Gram-positive bacteria in soil samples (PLFA1: 54% Gram-positive bacteria, PLFA2: 46%), the values differed by 13 percentage points for the stream sediment, 24 for the lake sediments and a sizeable 51 percentage points for the river microbial mats. In all substrates, the application of a larger number of PLFAs as biomarkers in PLFA2 increased the amount of resolved PLFAs significantly. Because the PLFA2 approach added additional biomarkers for Gram-negative bacteria, the relative abundances of Gram-negative bacteria increased in PLFA2 in comparison to the PLFA1 approach, which led to a better alignment with the metabarcoding data in all substrates. However, the increase of the resolution due to an increased number of PLFAs differed among the soil and freshwater samples. For example, in freshwater samples, the PLFA1 set of biomarkers resolved an average of 41% of the total amount of PLFAs resolved by PLFA2 in freshwater. This difference between the resolved amounts of total PLFAs in PLFA1 and PLFA2 decreased considerably in soil, where PLFA1 comprises of 66% of the total amount of resolved PLFAs in PLFA2. This disagreement between soil and freshwater samples suggests that a higher resolution of the PLFA method is more beneficial when freshwater samples are analysed. Although it contrasts our PLFA results based on the mock community, we argue that a wide range of biomarkers, encompassing a larger portion of the total PLFA profile, should be used to characterize complex microbial communities based on their PLFA composition.

In all freshwater substrates, the differences between the PLFA1 and PLFA2-based composition of the bacterial community were driven by two specific PLFAs: 16:1 ω 7t and 16:1 ω 5, both of which are defined as biomarkers for Gram-negative bacteria in method PLFA2. Both PLFAs were present in high proportions in these samples, which resulted in a prominent increase in the proportion of Gram-negative bacteria as assessed using PLFA2. It seems that while the inclusion of a larger number of PLFA biomarkers should decrease the loss of information, it might also introduce sources of error. For example, the PLFA 16:1 ω 7 is also a prominent constituent of the lipids of diatoms (Dijkman et al., 2010), which can occur in all sampled substrates, but are mainly found in aquatic habitats (Antonelli et al., 2017). Furthermore, the differences between PLFA1 and PLFA2 might have been influenced by the amount of Cyanobacteria per substrate type. We assessed the biomass of Cyanobacteria separately from the Gram-negative bacteria in our approach, based on polyunsaturated PLFAs and some monounsaturated fatty acids. The PLFA 16:1 ω 7c has been specifically described as a biomarker for Cyanobacteria (Potts et al., 1987), whereas its stereoisomer 16:1 ω 7t has been recommended as a biomarker for Gram-negative bacteria in the

literature (Hill et al., 2000; Zelles, 1999). Some PLFA studies on Cyanobacteria, however, do not differentiate between the isomers, instead naming 16:1 ω 7 as an important constituent of Cyanobacterial phospholipids (Ahlgren et al., 1992). Furthermore, bacteria are known to switch between their PLFA stereoisomers depending on the environmental conditions (Heipieper et al., 1996; Kaur et al., 2005). In our study, 16:1 ω 7t was notably concentrated in the substrates that had high concentrations of 16:1 ω 7c as well. Taken together, this leads us to suspect that the large proportion of Cyanobacteria, in combination with the use of the trans isomer of 16:1 ω 7 for Gram-negative bacteria in general, instead of for Cyanobacteria, led to the extreme differences between the two PLFA methods when applied to freshwater substrates. Contrastingly, Cyanobacteria do not play an important role in surface soil and the inclusion of these PLFAs as Gram-negative biomarkers had only a small impact on the difference between the profiles of the PLFA1 and PLFA2 approaches. We conclude that the origin of the substrate might bias the results of PLFA analysis, especially if single taxonomic groups dominate the microbial community composition. If this is the case, the careful selection of the PLFAs for biomarker analysis becomes important to avoid skewed views of the microbial community.

4.3. Application: Biomass-correction of metabarcoding proportions by PLFA

We combined PLFA data for microbial groups with their relative abundances as estimated by a metabarcoding approach. In the mock community, the application of the PLFA1-based estimation of the biomass led to an improved view of the metabarcoding data. However, the mock community did not include Actinobacteria and Cyanobacteria which made it impossible for us to evaluate if the additional application of the estimated biomasses of these two microbial groups would positively influence the resulting estimates of the PLFA-correction. Our application of the PLFA1 values as a biomass-correction factor resulted in notable changes in the perceived composition of the microbial community in all environmental substrates. In lake sediments, the PLFA1-correction led to a striking change in the identity of the most abundant phylum, from Gram-negative Proteobacteria to Gram-positive Firmicutes, a phylum that can dominate hypertrophic lake sediment (Wood et al., 2017; Zhang et al., 2020). In contrast, Cyanobacteria remained the dominant group in microbial mats before and after PLFA1-correction of the metabarcoding data. In both soil and stream sediment samples, the changes in the composition of the microbial community due to the PLFA-correction were relatively small. Overall, the changes induced by PLFA1-correction were similar to those of PLFA2-correction. However, our results highlight that the magnitude of the change of the bacterial profile through PLFA-correction depended entirely on the results of the PLFA analysis, i.e., how different the respective PLFA values are from their metabarcoding data. This observation stresses the importance of careful PLFA biomarker selection according to the origin of the substrate, as we previously discussed. Nonetheless, we argue that PLFA-correction, whether by a small set of PLFA biomarkers, or using a more comprehensive set like PLFA2, is a useful tool to deal with the pitfalls of compositional data resulting from metabarcoding approaches. In both PLFA applications, we were able to resolve relevant trends in the absolute abundances of microbial groups that led to distinctly different views of the sampling sites. These biomass-based views of the microbial community enabled description of the absolute trends among sites in our study.

A variety of methods to quantify bacteria exists, but their applicability depends on the sample type and the objective of the study. For example, absolute abundances have been determined by applying quantitative PCR (qPCR) in tandem with other methods of quantification of microbial load (Galazzo et al., 2020; Zhang et al., 2017). In accordance with our results, Galazzo et al. (2020) reported that while quantification by qPCR and flow cytometry of a mock community correlated strongly, large disagreement between the methods were

found when complex faecal samples were used and they argued that the process of the DNA extraction might have introduced the bias, be it due to DNA saturation, fragmentation or incomplete lysis. Flow cytometry quantifies living cells and measures their characteristics, and it is therefore able to distinguish between bacterial taxa (Prest et al., 2014; Props et al., 2017), but questions about its applicability for environmental samples remain, especially for samples of different properties (Khalili et al., 2019). Other studies used DNA (Hardwick et al., 2018; Tkacz et al., 2018; Wang et al., 2021) or bacterial strains (Lou et al., 2018) as internal standards added to the sample to obtain absolute quantitative results from NGS, but these methods do not account for possible biases in DNA extraction steps. Furthermore, bioinformatic approaches exist that either correct for inherent differences among bacterial taxa or can be used to detect changes in the abundances of taxa between samples. The 16S rRNA gene copy number varies among different bacteria, and tools for correcting (= gene copy normalization [GCN]) for that difference exist (Angly et al., 2014; Kembel et al., 2012). However, this approach still lacks a comprehensive description of the gene copy number of most organisms, and has been found to not reliably improve NGS sequencing abundances of mock communities (Starke et al., 2020). In our study, the GCN might have had increased the differences between metabarcoding and PLFA-based estimates of the abundances, because Gram-positive Firmicutes have comparably high gene copy numbers (Kembel et al., 2012) which would have led to a further decrease in their relative abundance in our results. If microbial biomass information is not available, or not of relevance, the application of mathematical methods like differential abundance analysis (Morton et al., 2019) can be used to infer changes in taxon abundances between samples. Each method provides additional information and has its advantages and disadvantages, and a careful selection of the most appropriate method is important.

5. Conclusion

Using PLFA in parallel with metabarcoding when measuring microbial communities can strengthen comparisons between samples, substrates and studies, especially when mock communities and other means of correction fall short in environmental studies. Our data demonstrates that PLFA-correction adds bacterial biomass as an ecologically meaningful dimension to the data, therefore enabling researchers to avoid common problems arising from compositional sequencing results. However, the origin of the samples, and their microbial composition needs to be carefully considered when selecting appropriate PLFA biomarkers.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

NL was financially supported by a Victoria University of Wellington Doctoral Scholarship. Funding for this research was supplied by National Science Challenge New Zealand NSC3 A National Framework for Biological Heritage Assessment across Natural and Production Landscapes grant E3049-2940 to JRD. The Centre of Biodiversity and Restoration Ecology at Victoria University of Wellington funds to support field work to NL (student award 2016). NL would like to thank Ian Vorster, Michael Jackson and Christine Stockum for their support in GC-MS work. SAW, BW and GTL acknowledge the New Zealand Ministry of Business, Innovation and Employment research programme - Our Lakes' Health; past, present and future (C05X1707) for the funding for the collection and analysis of the Lake Rotorua sediment samples.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.mimet.2021.106271>.

References

- Ahlgren, G., Gustafsson, I.-B., Boberg, M., 1992. Fatty acid content and chemical composition of freshwater microalgae. *J. Phycol.* 28, 37–50.
- Aislabie, J., Deslippe, J., 2013. Soil microbes and their contribution to soil services. In: Dymond, J.R. (Ed.), *Ecosystem Services in New Zealand – Conditions and Trends*. Manaaki Whenua Press, Lincoln, New Zealand, pp. 143–161.
- Angly, F.E., Dennis, P.G., Skarshewski, A., Vanwonterghem, I., Hugenholtz, P., Tyson, G.W., 2014. CopyRighter: a rapid tool for improving the accuracy of microbial community profiles through lineage-specific gene copy number correction. *Microbiome*. 2, 11.
- Antonelli, M., Wetzel, C.E., Ector, L., Teuling, A.J., Pfister, L., 2017. On the potential for terrestrial diatom communities and diatom indices to identify anthropic disturbance in soils. *Ecol. Indic.* 75, 73–81.
- Auckland Council, 2018. State of the Waitakere Ranges Heritage Area 2018. Auckland Council, Auckland, New Zealand, pp. 202–220.
- Bååth, E., Frostegård, Å., Fritze, H., 1992. Soil bacterial biomass, activity, phospholipid fatty acid pattern, and pH tolerance in an area polluted with alkaline dust deposition. *Appl. Environ. Microbiol.* 58, 4026–4031.
- Balestrini, R., Bianciotto, V., Bonfante, P., Schloter, M., Sharat, S., Thorn, R.G., Williamson, K.E., Wommack, K.E., 2011. *Microbiota*. CRC Press, Handbook of soil sciences, pp. 1–40.
- Beckers, B., Op De Beeck, M., Thijs, S., Truyens, S., Weyens, N., Boerjan, W., Vangronsveld, J., 2016. Performance of 16S rDNA primer pairs in the study of rhizosphere and endosphere bacterial microbiomes in metabarcoding studies. *Front. Microbiol.* 7, 650.
- Berry, T.E., Saunders, B.J., Coghlan, M.L., Stat, M., Jarman, S., Richardson, A.J., Davies, C.H., Berry, O., Harvey, E.S., Bunce, M., 2019. Marine environmental DNA biomonitoring reveals seasonal patterns in biodiversity and identifies ecosystem responses to anomalous climatic events. *PLoS Genet.* 15, e1007943.
- Boers, S.A., Jansen, R., Hays, J.P., 2019. Understanding and overcoming the pitfalls and biases of next-generation sequencing (NGS) methods for use in the routine clinical microbiological diagnostic laboratory. *Eur. J. Clin. Microbiol. Infect. Dis.* 38, 1059–1070.
- Bohmann, K., Evans, A., Gilbert, M.T.P., Carvalho, G.R., Creer, S., Knapp, M., Yu, D.W., de Bruyn, M., 2014. Environmental DNA for wildlife biology and biodiversity monitoring. *Trends Ecol. Evol.* 29.
- Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., Alexander, H., Alm, E.J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J.E., Bittinger, K., Brejnrod, A., Brislawn, C.J., Brown, C.T., Callahan, B.J., Caraballo-Rodríguez, A.M., Chase, J., Cope, E.K., Da Silva, R., Diener, C., Dorrestein, P.C., Douglas, G.M., Durall, D.M., Duvallet, C., Edwardson, C.F., Ernst, M., Estaki, M., Fouquier, J., Gauglitz, J.M., Gibbons, S.M., Gibson, D.L., Gonzalez, A., Gorlick, K., Guo, J., Hillmann, B., Holmes, S., Holste, H., Huttenhower, C., Huttley, G.A., Janssen, S., Jarmusch, A.K., Jiang, L., Kaehler, B.D., Kang, K.B., Keefe, C.R., Keim, P., Kelley, S.T., Knights, D., Koester, I., Kosciolk, T., Kreps, J., Langille, M.G.L., Lee, J., Ley, R., Liu, Y.-X., Loftfield, E., Lozupone, C., Maher, M., Marotz, C., Martin, B.D., McDonald, D., McIver, L.J., Melnik, A.V., Metcalf, J.L., Morgan, S.C., Morton, J.T., Naimey, A.T., Navas-Molina, J.A., Nothias, L.F., Orchanian, S.B., Pearson, T., Peoples, S.L., Petras, D., Preuss, M.L., Pruesse, E., Rasmussen, L.B., Rivers, A., Robeson, M.S., Rosenthal, P., Segata, N., Shaffer, M., Shiffer, A., Sinha, R., Song, S.J., Spear, J.R., Swafford, A.D., Thompson, L.R., Torres, P.J., Trinh, P., Tripathi, A., Tuzubaugh, P.J., Ul-Hasan, S., van der Hooft, J.J.J., Vargas, F., Vázquez-Baeza, Y., Vogtmann, E., von Hippel, M., Walters, W., Wan, Y., Wang, M., Warren, J., Weber, K.C., Williamson, C.H.D., Willis, A.D., Xu, Z.Z., Zaneveld, J.R., Zhang, Y., Zhu, Q., Knight, R., Caporaso, J.G., 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* 37, 852–857.
- Bonk, F., Popp, D., Harms, H., Centler, F., 2018. PCR-based quantification of taxa-specific abundances in microbial communities: quantifying and avoiding common pitfalls. *J. Microbiol. Methods* 153, 139–147.
- Brooks, J.P., Edwards, D.J., Harwich, M.D., Rivera, M.C., Fettweis, J.M., Serrano, M.G., Reris, R.A., Sheth, N.U., Huang, B., Girerd, P., Strauss, J.F., Jefferson, K.K., Buck, G.A., Vaginal Microbiome, C., 2015. The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies. *BMC Microbiol.* 15, 66.
- Buxton, A.S., Groombridge, J.J., Griffiths, R.A., 2017. Is the detection of aquatic environmental DNA influenced by substrate type? *PLoS One* 12, e0183371.
- Buyer, J.S., Sasser, M., 2012. High throughput phospholipid fatty acid analysis of soils. *Appl. Soil Ecol.* 61, 127–130.
- Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., Holmes, S.P., 2016. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581.
- Callahan, B.J., McMurdie, P.J., Holmes, S.P., 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J.* 11, 2639.
- Carini, P., Marsden, P.J., Leff, J.W., Morgan, E.E., Strickland, M.S., Fierer, N., 2016. Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. *Nat. Microbiol.* 2, 16242.
- Córdova-Kreylos, A.L., Cao, Y., Green, P.G., Hwang, H.-M., Kuivila, K.M., LaMontagne, M.G., Van De Werfhorst, L.C., Holden, P.A., Scow, K.M., 2006.

- Diversity, composition, and geographical distribution of microbial communities in California salt marsh sediments. *Appl. Environ. Microbiol.* 72, 3357–3366.
- Cristescu, M.E., Hebert, P.D.N., 2018. Uses and misuses of environmental DNA in biodiversity science and conservation. *Annu. Rev. Ecol. Syst.* 49, 209–230.
- de Gannes, V., Bekele, I., Dipchansingh, D., Wuddivira, M.N., De Cairies, S., Boman, M., Hickey, W.J., 2016. Microbial community structure and function of soil following ecosystem conversion from native forests to teak plantation forests. *Front. Microbiol.* 7, 1976.
- Deagle, B.E., Thomas, A.C., McInnes, J.C., Clarke, L.J., Vesterinen, E.J., Clare, E.L., Kartzinel, T.R., Eveson, J.P., 2019. Counting with DNA in metabarcoding studies: how should we convert sequence reads to dietary data? *Mol. Ecol.* 28, 391–406.
- Deslippe, J.R., Egger, K.N., Henry, G.H.R., 2005. Impacts of warming and fertilization on nitrogen-fixing microbial communities in the Canadian High Arctic. *FEMS Microbiol. Ecol.* 53, 41–50.
- Dijkman, N.A., Boschker, H.T.S., Stal, L.J., Kromkamp, J.C., 2010. Composition and heterogeneity of the microbial community in a coastal microbial mat as revealed by the analysis of pigments and phospholipid-derived fatty acids. *J. Sea Res.* 63, 62–70.
- Dodds, E.D., McCoy, M.R., Rea, L.D., Kennish, J.M., 2005. Gas chromatographic quantification of fatty acid methyl esters: flame ionization detection vs. Electron impact mass spectrometry. *Lipids* 40, 419–428.
- Dopheide, A., Xie, D., Buckley, T.R., Drummond, A.J., Newcomb, R.D., 2019. Impacts of DNA extraction and PCR on DNA metabarcoding estimates of soil biodiversity. *Methods Ecol. Evol.* 10, 120–133.
- Dunfield, K.E., 2007. Lipid-Based Community Analysis. In: M. R. Carter, E. G. Gregorich (Eds.). CRC Press, Soil sampling and methods of analysis, p. 587.
- Edgar, R.C., 2017. Accuracy of microbial community diversity estimated by closed- and open-reference OTUs. *PeerJ* 5, e3889.
- Elbrecht, V., Leese, F., 2015. Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass—sequence relationships with an innovative metabarcoding protocol. *PLoS One* 10, e0130324.
- Evans, S.E., Wallenstein, M.D., 2014. Climate change alters ecological strategies of soil bacteria. *Ecol. Lett.* 17, 155–164.
- Fanin, N., Kardol, P., Farrell, M., Nilsson, M.-C., Gundale, M.J., Wardle, D.A., 2019. The ratio of gram-positive to gram-negative bacterial PLFA markers as an indicator of carbon availability in organic soils. *Soil Biol. Biochem.* 128, 111–114.
- Fierer, N., Schimel, J.P., Holden, P.A., 2003. Variations in microbial community composition through two soil depth profiles. *Soil Biol. Biochem.* 35, 167–176.
- Findlay, R.H., 2004. Determination of microbial community structure using phospholipid fatty acid profiles. In: Kowalchuk, G.A., Bruijn, F.J.D., Head, I.M., Akkermans, A.D. L., Elsas, J.D.V. (Eds.), *Molecular microbial ecology manual*. Kluwer Academic Publishers, Dordrecht, London, pp. 983–1004.
- Francisco, R., Stone, D., Creamer, R.E., Sousa, J.P., Morais, P.V., 2016. European scale analysis of phospholipid fatty acid composition of soils to establish operating ranges. *Appl. Soil Ecol.* 97, 49–60.
- Frostegård, Å., Tunlid, A., Bååth, E., 1991. Microbial biomass measured as total lipid phosphate in soils of different organic content. *J. Microbiol. Methods* 14, 151–163.
- Frostegård, Å., Bååth, E., Tunlid, A., 1993. Shifts in the structure of soil microbial communities in limed forests as revealed by phospholipid fatty acid analysis. *Soil Biol. Biochem.* 25, 723–730.
- Frostegård, Å., Courtois, S., Ramišse, V., Clerc, S., Bernillon, D., Le Gall, F., Jeannin, P., Nesme, X., Simonet, P., 1999. Quantification of bias related to the extraction of DNA directly from soils. *Appl. Environ. Microbiol.* 65, 5409–5420.
- Frostegård, Å., Tunlid, A., Bååth, E., 2011. Use and misuse of PLFA measurements in soils. *Soil Biol. Biochem.* 43, 1621–1625.
- Galazzo, G., van Best, N., Benedikter, B.J., Janssen, K., Bervoets, L., Driessen, C., Oomen, M., Lucchesi, M., van Eijck, P.H., Becker, H.E.F., Hornef, M.W., Savelkoul, P. H., Stassen, F.R.M., Wolfs, P.F., Penders, J., 2020. How to count our microbes? The Effect of Different Quantitative Microbiome Profiling Approaches. *Front. Cell. Infect. Microbiol.* 10, 403.
- Glucksman, A.M., Skipper, H.D., Brigmon, R.L., Domingo, J.W.S., 2000. Use of the MIDI-FAME technique to characterize groundwater communities. *J. Appl. Microbiol.* 88, 711–719.
- Green, C.T., Scow, K.M., 2000. Analysis of phospholipid fatty acids (PLFA) to characterize microbial communities in aquifers. *Hydrogeol. J.* 8, 126–141.
- Hardwick, S.A., Deveson, I.W., Mercer, T.R., 2017. Reference standards for next-generation sequencing. *Nat. Rev. Genet.* 18, 473.
- Hardwick, S.A., Chen, W.Y., Wong, T., Kanakamedala, B.S., Deveson, I.W., Ongley, S.E., Santini, N.S., Marcellin, E., Smith, M.A., Nielsen, L.K., Lovelock, C.E., Neilan, B.A., Mercer, T.R., 2018. Synthetic microbe communities provide internal reference standards for metagenome sequencing and analysis. *Nat. Commun.* 9, 3096.
- Harper, L.R., Buxton, A.S., Rees, H.C., Bruce, K., Brys, R., Halfmaerten, D., Read, D.S., Watson, H.V., Sayer, C.D., Jones, E.P., Priestley, V., Mächler, E., Múrria, C., Garcés-Pastor, S., Medupin, C., Burgess, K., Benson, G., Boonham, N., Griffiths, R.A., Lawson Handley, L., Hänfling, B., 2019. Prospects and challenges of environmental DNA (eDNA) monitoring in freshwater ponds. *Hydrobiologia* 826, 25–41.
- Heipieper, H.J., Meulenbeld, G., van Oirschot, Q., de Bont, J., 1996. Effect of environmental factors on the trans/cis ratio of unsaturated fatty acids in *Pseudomonas putida* S12. *Appl. Environ. Microbiol.* 62, 2773–2777.
- Herlemann, D.P.R., Labrenz, M., Jürgens, K., Bertilsson, S., Waniek, J.J., Andersson, A.F., 2011. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *The ISME Journal* 5, 1571–1579.
- Hermans, S.M., Buckley, H.L., Case, B.S., Curran-Cournane, F., Taylor, M., Lear, G., 2016. Bacteria as emerging indicators of soil condition. *Appl. Environ. Microbiol.* 83 (1), e02826–e02916.
- Hermans, S.M., Buckley, H.L., Lear, G., 2018. Optimal extraction methods for the simultaneous analysis of DNA from diverse organisms and sample types. *Mol. Ecol. Resour.* 18, 557–569.
- Hermans, S.M., Buckley, H.L., Case, B.S., Lear, G., 2019. Connecting through space and time: catchment-scale distributions of bacteria in soil, stream water and sediment. *Environ. Microbiol.* 22, 1000–1010.
- Hill, G.T., Mitkowski, N.A., Aldrich-Wolfe, L., Emele, L.R., Jurkonie, D.D., Ficke, A., Maldonado-Ramirez, S., Lynch, S.T., Nelson, E.B., 2000. Methods for assessing the composition and diversity of soil microbial communities. *Appl. Soil Ecol.* 15, 25–36.
- Jacobs-Wagner, C., Cabeen, M.T., 2005. Bacterial cell shape. *Nat. Rev. Microbiol.* 3, 601–610.
- Jo, T., Murakami, H., Yamamoto, S., Masuda, R., Minamoto, T., 2019. Effect of water temperature and fish biomass on environmental DNA shedding, degradation, and size distribution. *Ecology and Evolution* 9, 1135–1146.
- Kanagawa, T., 2003. Bias and artifacts in multitemplate polymerase chain reactions (PCR). *J. Biosci. Bioeng.* 96, 317–323.
- Kassambara, A., 2020. Rstatix: pipe-friendly framework for basic statistical tests, R package version 0.5.0.
- Kaur, A., Chaudhary, A., Kaur, A., Choudhary, R., Kaushik, R., 2005. Phospholipid fatty acid - a bioindicator of environment monitoring and assessment in soil ecosystem. *Curr. Sci.* 89, 1103–1112.
- Kemmel, S.W., Wu, M., Eisen, J.A., Green, J.L., 2012. Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. *PLoS Comput. Biol.* 8, e1002743.
- Kennedy, K., Hall, M.W., Lynch, M.D.J., Moreno-Hagelsieb, G., Neufeld, J.D., 2014. Evaluating bias of Illumina-based bacterial 16S rRNA gene profiles. *Appl. Environ. Microbiol.* 80, 5717–5722.
- Khalili, B., Weihe, C., Kimball, S., Schmidt, K.T., Martiny, J.B.H., 2019. Optimization of a method to quantify soil bacterial abundance by flow Cytometry. *mSphere* 4 e00435-00419.
- Kirk, J.L., Beaudette, L.A., Hart, M., Moutoglou, P., Klironomos, J.N., Lee, H., Trevors, J. T., 2004. Methods of studying soil microbial diversity. *J. Microbiol. Methods* 58, 169–188.
- Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., Glöckner, F.O., 2012. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* 41, e1.
- Knudsen, S.W., Ebert, R.B., Hesselsoe, M., Kuntke, F., Hassingboe, J., Mortensen, P.B., Thomsen, P.F., Sigsgaard, E.E., Hansen, B.K., Nielsen, E.E., Møller, P.R., 2019. Species-specific detection and quantification of environmental DNA from marine fishes in the Baltic Sea. *J. Exp. Mar. Biol. Ecol.* 510, 31–45.
- Kohring, L.L., Ringelberg, D.B., Devereux, R., Stahl, D.A., Mittelman, M.W., White, D.C., 1994. Comparison of phylogenetic relationships based on phospholipid fatty acid profiles and ribosomal RNA sequence similarities among dissimilatory sulfate-reducing bacteria. *FEMS Microbiol. Lett.* 119, 303–308.
- Kumar, S., Suyal, D.C., Yadav, A., Shouche, Y., Goel, R., 2019. Microbial diversity and soil physicochemical characteristic of higher altitude. *PLoS One* 14, e0213844.
- Lacoursière-Roussel, A., Howland, K., Normandeau, E., Grey, E.K., Archambault, P., Deiner, K., Lodge, D.M., Hernandez, C., Leduc, N., Bernatchez, L., 2018. eDNA metabarcoding as a new surveillance approach for coastal Arctic biodiversity. *Ecology and evolution* 8, 7763–7777.
- Laursen, M.F., Dalgaard, M.D., Bahl, M.I., 2017. Genomic GC-content affects the accuracy of 16S rRNA gene sequencing based microbial profiling due to PCR bias. *Front. Microbiol.* 8.
- Lear, G., Dickie, I., Banks, J., Boyer, S., Buckley, H., Buckley, T., Cruickshank, R., Dopheide, A., Handley, K., Hermans, S., Kamke, J., Lee, C., Macdiarmid, R., Morales, S., Orlovich, D., Smissen, R., Wood, J., Holdaway, R., 2018. Methods for the extraction, storage, amplification and sequencing of DNA from environmental samples. *N. Z. J. Ecol.* 42, 10–50A.
- Lee, A.H., Lee, J., Noh, J., Lee, C., Hong, S., Kwon, B.-O., Kim, J.-J., Kim, J.S., 2020. Characteristics of long-term changes in microbial communities from contaminated sediments along the west coast of South Korea: Ecological assessment with eDNA and physicochemical analyses. *Mar. Pollut. Bull.* 160.
- Lennon, J.T., Muscarella, M.E., Placella, S.A., Lehmkuhl, B.K., 2018. How, when, and where relic DNA affects microbial diversity. *mBio* 9 e00637-00618.
- Lou, J., Yang, L., Wang, H., Wu, L., Xu, J., 2018. Assessing soil bacterial community and dynamics by integrated high-throughput absolute abundance quantification. *PeerJ* 6, e4514.
- Lozupone, C.A., Knight, R., 2007. Global patterns in bacterial diversity. *Proc. Natl. Acad. Sci. U. S. A.* 104, 11436–11440.
- Mahalanabis, M., Al-Muayad, H., Kulinski, M.D., Altman, D., Klapperich, C.M., 2009. Cell lysis and DNA extraction of gram-positive and gram-negative bacteria from whole blood in a disposable microfluidic chip. *Lab Chip* 9, 2811–2817.
- Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads, 17, p. 3.
- Martzy, R., Bica-Schröder, K., Pálvölgyi, Á.M., Kolm, C., Jakwerth, S., Kirschner, A.K.T., Sommer, R., Krška, R., Mach, R.L., Farnleitner, A.H., Reischer, G.H., 2019. Simple lysis of bacterial cells for DNA-based diagnostics using hydrophilic ionic liquids. *Sci. Rep.* 9, 13994.
- McDonald, D., Price, M.N., Goodrich, J., Nawrocki, E.P., DeSantis, T.Z., Probst, A., Andersen, G.L., Knight, R., Hugenholtz, P., 2012. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* 6, 610–618.
- Miller, D.N., Bryant, J.E., Madsen, E.L., Ghiorse, W.C., 1999. Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. *Appl. Environ. Microbiol.* 65, 4715–4724.

- Moré, M.I., Herrick, J.B., Silva, M.C., Ghiorse, W.C., Madsen, E.L., 1994. Quantitative cell lysis of indigenous microorganisms and rapid extraction of microbial DNA from sediment. *Appl. Environ. Microbiol.* 60, 1572–1580.
- Morton, J.T., Marotz, C., Washburne, A., Silverman, J., Zaramela, L.S., Edlund, A., Zengler, K., Knight, R., 2019. Establishing microbial composition measurement standards with reference frames. *Nat. Commun.* 10, 2719.
- Nichols, R.V., Vollmers, C., Newsom, L.A., Wang, Y., Heintzman, P.D., Leighton, M., Green, R.E., Shapiro, B., 2018. Minimizing polymerase biases in metabarcoding. *Mol. Ecol. Resour.* 18, 927–939.
- Nielsen, P., Petersen, S.O., 2000. Ester-linked polar lipid fatty acid profiles of soil microbial communities: a comparison of extraction methods and evaluation of interference from humic acids. *Soil Biol. Biochem.* 32, 1241–1249.
- Nielsen, K.M., Johnsen, P.J., Bensasson, D., Daffonchio, D., 2007. Release and persistence of extracellular DNA in the environment. *Environ. Biosaf. Res.* 6, 37–53.
- Orwin, K.H., Dickie, I.A., Holdaway, R., Wood, J.R., 2018. A comparison of the ability of PLFA and 16S rRNA gene metabarcoding to resolve soil community change and predict ecosystem functions. *Soil Biol. Biochem.* 117, 27–35.
- Parada, A.E., Needham, D.M., Fuhrman, J.A., 2016. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.* 18, 1403–1414.
- Pollock, J., Glendinning, L., Wisedchanwet, T., Watson, M., 2018. The madness of microbiome: attempting to find consensus “best practice” for 16S microbiome studies. *Appl. Environ. Microbiol.* 84 (e02627–02617).
- Potts, M., Olie, J.J., Nickels, J.S., Parsons, J., White, D.C., 1987. Variation in phospholipid ester-linked fatty acids and carotenoids of desiccated *Nostoc commune* (Cyanobacteria) from different geographic locations. *Appl. Environ. Microbiol.* 53, 4–9.
- Pratt, B., Riesen, R., Johnston, C.G., 2012. PLFA analyses of microbial communities associated with PAH-contaminated riverbank sediment. *Microb. Ecol.* 64, 680–691.
- Prest, E.I., El-Chakhtoura, J., Hammes, F., Saikaly, P.E., van Loosdrecht, M.C.M., Vrouwenvelder, J.S., 2014. Combining flow cytometry and 16S rRNA gene pyrosequencing: a promising approach for drinking water monitoring and characterization. *Water Res.* 63, 179–189.
- Probandt, D., Eickhorst, T., Ellrott, A., Amann, R., Knittel, K., 2018. Microbial life on a sand grain: from bulk sediment to single grains. *The ISME Journal.* 12, 623–633.
- Props, R., Kerckhof, F.-M., Rubbens, P., De Vrieze, J., Hernandez Sanabria, E., Waegeman, W., Monsieurs, P., Hammes, F., Boon, N., 2017. Absolute quantification of microbial taxon abundances. *The ISME Journal.* 11, 584–587.
- R Core Team, 2016. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.** <http://www.R-project.org/>.
- Rajendran, N., Matsuda, O., Urushigawa, Y., Simidu, U., 1994. Characterization of microbial community structure in the surface sediment of Osaka Bay, Japan, by phospholipid fatty acid analysis. *Appl. Environ. Microbiol.* 60, 248–257.
- Ramsey, P.W., Rillig, M.C., Feris, K.P., Holben, W.E., Gannon, J.E., 2006. Choice of methods for soil microbial community analysis: PLFA maximizes power compared to CLPP and PCR-based approaches. *Pedobiologia.* 50, 275–280.
- Ritz, K., Black, H.L.J., Campbell, C.D., Harris, J.A., Wood, C., 2009. Selecting biological indicators for monitoring soils: a framework for balancing scientific and technical opinion to assist policy development. *Ecol. Indic.* 9, 1212–1221.
- Ruess, L., Chamberlain, P.M., 2010. The fat that matters: soil food web analysis using fatty acids and their carbon stable isotope signature. *Soil Biol. Biochem.* 42, 1898–1910.
- Sirois, S.H., Buckley, D.H., 2019. Factors governing extracellular DNA degradation dynamics in soil. *Environ. Microbiol. Rep.* 11, 173–184.
- Stahl, P.D., Klug, M.J., 1996. Characterization and differentiation of filamentous fungi based on fatty acid composition. *Appl. Environ. Microbiol.* 62, 4136–4146.
- Starke, R., Pylro, V.S., Morais, D.K., 2020. 16S rRNA gene copy number normalization does not provide more reliable conclusions in Metataxonomic surveys. *Microb. Ecol.* 81, 535–539.
- Steger, K., Premke, K., Gudasz, C., Sundh, I., Tranvik, L.J., 2011. Microbial biomass and community composition in boreal lake sediments. *Limnol. Oceanogr.* 56, 725–733.
- Stoeckle, B.C., Beggel, S., Cerwenka, A.F., Motivans, E., Kuehn, R., Geist, J., 2017. A systematic approach to evaluate the influence of environmental conditions on eDNA detection success in aquatic ecosystems. *PLoS One* 12, e0189119.
- Sze, M.A., Schloss, P.D., 2019. The impact of DNA polymerase and number of rounds of amplification in PCR on 16S rRNA gene sequence data. *mSphere* 4 e00163-00119.
- Thijs, S., Op De Beeck, M., Beckers, B., Truyens, S., Stevens, V., Van Hamme, J.D., Weyens, N., Vangronsveld, J., 2017. Comparative evaluation of four bacteria-specific primer pairs for 16S rRNA gene surveys. *Front. Microbiol.* 8, 494.
- Thomson-Laing, G., Puddick, J., Laroche, O., Fulton, S., Steiner, K., Heath, M.W., Wood, S.A., 2020. Broad and fine scale variability in bacterial diversity and cyanotoxin quotas in benthic cyanobacterial mats. *Front. Microbiol.* 11.
- Tkacz, A., Hortala, M., Poole, P.S., 2018. Absolute quantification of microbiota abundance in environmental samples. *Microbiome.* 6, 110.
- Torsvik, V., Øvreås, L., 2002. Microbial diversity and function in soil: from genes to ecosystems. *Curr. Opin. Microbiol.* 5, 240–245.
- Torsvik, V., Goksoyr, J., Daee, F.L., 1990. High diversity in DNA of soil bacteria. *Appl. Environ. Microbiol.* 56, 782.
- Vandeputte, D., Kathagen, G., D’hoë, K., Vieira-Silva, S., Valles-Colomer, M., Sabino, J., Wang, J., Tito, R.Y., De Commer, L., Darzi, Y., Vermeire, S., Falony, G., Raes, J., 2017. Quantitative microbiome profiling links gut community variation to microbial load. *Nature.* 551, 507–511.
- Wang, Y., Wang, Z.-L., 2018. Shifts of sediment microbial community structure along a salinized and degraded river continuum. *J. Coast. Res.* 34, 443–450.
- Wang, Y., Sheng, H.-F., He, Y., Wu, J.-Y., Jiang, Y.-X., Tam, N.F.-Y., Zhou, H.-W., 2012. Comparison of the levels of bacterial diversity in freshwater, intertidal wetland, and marine sediments by using millions of Illumina tags. *Appl. Environ. Microbiol.* 78, 8264.
- Wang, S., Wu, Q., Han, Y., Du, R., Wang, X., Nie, Y., Du, X., Xu, Y., 2021. Gradient internal standard method for absolute quantification of microbial amplicon sequencing data. *mSystems* 6 e00964-00920.
- Watzinger, A., 2015. Microbial phospholipid biomarkers and stable isotope methods help reveal soil functions. *Soil Biol. Biochem.* 86, 98–107.
- Weisbrod, B., Wood, S.A., Steiner, K., Whyte-Wilding, R., Puddick, J., Laroche, O., Dietrich, D.R., 2020. Is a central sediment sample sufficient? Exploring spatial and temporal microbial diversity in a small lake. *Toxins (Basel)* 12, 580.
- White, D.C., Davis, W.M., Nickels, J.S., King, J.D., Bobbie, R.J., 1979. Determination of the sedimentary microbial biomass by extractable lipid phosphate. *Oecologia.* 40, 51–62.
- Willers, C., van Rensburg, P.J.J., Claessens, S., 2015. Phospholipid fatty acid profiling of microbial communities—a review of interpretations and recent applications. *J. Appl. Microbiol.* 119, 1207–1218.
- Wood, S.A., Borges, H., Puddick, J., Biessy, L., Atalah, J., Hawes, I., Dietrich, D.R., Hamilton, D.P., 2017. Contrasting cyanobacterial communities and microcystin concentrations in summers with extreme weather events: insights into potential effects of climate change. *Hydrobiologia.* 785, 71–89.
- Wu, G.D., Lewis, J.D., Hoffmann, C., Chen, Y.-Y., Knight, R., Bittinger, K., Hwang, J., Chen, J., Berkowsky, R., Nessel, L., Li, H., Bushman, F.D., 2010. Sampling and pyrosequencing methods for characterizing bacterial communities in the human gut using 16S sequence tags. *BMC Microbiol.* 10, 206.
- Yeh, Y.-C., Needham, D.M., Sieradzki, E.T., Fuhrman, J.A., 2018. Taxon disappearance from microbiome analysis reinforces the value of mock communities as a standard in every sequencing run. *mSystems* 3 e00023-00018.
- Zelles, L., 1997. Phospholipid fatty acid profiles in selected members of soil microbial communities. *Chemosphere.* 35, 275–294.
- Zelles, L., 1999. Fatty acid patterns of phospholipids and lipopolysaccharides in the characterisation of microbial communities in soil: a review. *Biol. Fertil. Soils* 29, 111–129.
- Zelles, L., Bai, Q.Y., Beck, T., Beese, F., 1992. Signature fatty acids in phospholipids and lipopolysaccharides as indicators of microbial biomass and community structure in agricultural soils. *Soil Biol. Biochem.* 24, 317–323.
- Zhang, C.L., Huang, Z., Li, Y.-L., Romanek, C.S., Mills, G.L., Gibson, R.A., Talbot, H.M., Wiegel, J., Noakes, J., Culp, R., White, D.C., 2007. Lipid biomarkers, carbon isotopes, and phylogenetic characterization of bacteria in California and Nevada hot springs. *Geomicrobiol. J.* 24, 519–534.
- Zhang, Z., Qu, Y., Li, S., Feng, K., Wang, S., Cai, W., Liang, Y., Li, H., Xu, M., Yin, H., Deng, Y., 2017. Soil bacterial quantification approaches coupling with relative abundances reflecting the changes of taxa. *Sci. Rep.* 7, 4837.
- Zhang, Y., Zheng, N., Wang, J., Yao, H., Qiu, Q., Chapman, S.J., 2019. High turnover rate of free phospholipids in soil confirms the classic hypothesis of PLFA methodology. *Soil Biol. Biochem.* 135, 323–330.
- Zhang, W., Gu, P., Zhu, W., Wang, N., Jiang, M., He, J., Zheng, Z., 2020. Phenotype changes of cyanobacterial and microbial distribution characteristics of surface sediments in different periods of cyanobacterial blooms in Taihu Lake. *Aquat. Ecol.* 54, 591–607.
- Zhao, C., Long, J., Liao, H., Zheng, C., Li, J., Liu, L., Zhang, M., 2019. Dynamics of soil microbial communities following vegetation succession in a karst mountain ecosystem. *Southwest China. Scientific Reports.* 9, 2160.