Journal of Microbiological Methods 151 (2018) 28–34

Contents lists available at ScienceDirect

Journal of Microbiological Methods

journal homepage: www.elsevier.com/locate/jmicmeth

Does universal 16S rRNA gene amplicon sequencing of environmental communities provide an accurate description of nitrifying guilds?

Vaibhav Diwan, Hans-Jørgen Albrechtsen, Barth F. Smets, Arnaud Dechesne*

Department of Environmental Engineering, Technical University of Denmark, Bygningstorvet, Building 115, 2800 Kgs, Lyngby, Denmark

ARTICLE INFO

Keywords:
Diversity
Drinking water treatment
Wastewater treatment
Nitrifiers
AOB
Nitrospira

ABSTRACT

Universal (i.e., targeting most bacteria/prokaryotes) 16S rRNA gene based amplicon sequencing is widely used for assessing microbial communities due to its low cost, time efficiency, and ability to provide a full overview of the community. However, it is currently unclear if it can yield reliable information on specific microbial guilds, as obtained by using primer sets targeting functional genes or specific16S rRNA gene sequences. Here, we compared the relative abundance, diversity, richness, and composition of selected guilds (nitrifiers), obtained from universal 16S rRNA gene based amplicon sequencing and from guild targeted approaches. The universal amplicon sequencing provided 1) accurate estimates of nitrifier composition, 2) clustering of the samples based on these compositions consistent with sample origin, 3) estimates of the relative abundance of the guilds correlated with those obtained from the targeted approaches and within ~1.2 orders of magnitude of them, but with measurable bias that should be considered when comparing estimates from both approaches. In contrast, the diversity and richness estimations using the universal 16S rRNA based amplicon sequencing were likely limited by the sequencing depth; therefore, we suggest preferring targeted approaches for assessing nitrifiers diversity and richness or using sequencing depth larger than those currently typically practiced. Overvall, we conclude that universal amplicon sequencing provides, in a single analysis, useful information on the abundance and composition of diverse guilds in complex environmental communities.

1. Introduction

Community-wide high throughput amplicon sequencing of one or multiple hypervariable regions of the 16S rRNA gene has become a routine tool to describe the composition and diversity of microbial communities (Bartram et al., 2011; Caporaso et al., 2011). Beyond such an overall community assessment, it is often desirable to quantify and characterize specific constituent guilds in terms of abundance, composition, and diversity. For many guilds, the 16S rRNA gene can be informative because phylogenetic conservation of functional traits is common (Philippot et al., 2010; Martiny et al., 2013). In fact, if the microbial guild of interest consists of one or a few monophyletic clades (for example, ammonia-, nitrite-, or methane-oxidizers), it is possible to utilize 16S rRNA gene primers to specifically target the guild (Degrange and Bardin, 1995; Kowalchuk et al., 1997; Hermansson and Lindgren, 2001; Graham et al., 2007). Additionally, if conserved signature functional genes exist, targeting them can be a strategy. For example, *amoA*, which codes for a subunit of the ammonia monooxygenase can be used to target ammonia oxidizing prokaryotes; and *nxrB,* which codes for the beta subunit of the nitrite oxidoreductase can be used to target nitrite oxidizing bacteria (NOB) (Norton, 2011; Pester et al., 2013). Targeting functional genes can be advantageous to access within-guild diversity, due to their high rate of evolution compared to the 16S rRNA gene (Dopheide et al., 2015).

In contrast to a non-specific approach such as universal (i.e., designed to target most bacteria) 16S rRNA gene amplicon sequencing, any guild-specific approach provides information only on the targeted guild but none on the rest of the community, which is an obvious limitation (Xue et al., 2013; Dopheide et al., 2015). The tradeoff is that universal approaches give a good overview of the whole microbial community, but provide only limited information on non-dominant guilds because of their low contribution to the total sequence pool.

Therefore, it is currently unclear whether universal 16S rRNA gene amplicon sequencing is sufficient to obtain reliable information on specific microbial guilds and, more precisely, whether it correctly differentiates between samples with high and low guild diversity and whether it provides sufficient compositional information to identify samples with similar guild composition.

An additional concern regarding microbial community analysis using amplicon sequencing lies in its ability to provide reliable

* Corresponding author at: Department of Environmental Engineering, Technical University of Denmark, Bygningstorvet, Building 115, 2800 Kgs, Lyngby, Denmark.
E-mail addresses: vdiw@env.dtu.dk (V. Diwan), hana@env.dtu.dk (H.-J. Albrechtsen), bfsm@env.dtu.dk (B.F. Smets), arde@env.dtu.dk (A. Dechesne).

https://doi.org/10.1016/j.mimet.2018.05.025
Received 16 March 2018; Received in revised form 29 May 2018; Accepted 29 May 2018
Available online 30 May 2018

estimates of the relative abundance of the community constituents. Indeed, the quantitative nature of amplicon sequencing has been questioned (Zhou et al., 2011). Studies using mock communities have shown that the results from 16S rRNA gene based amplicon can be biased based on choice of primers and at different in-vitro (DNA extraction) and *in-silico* (sequencing and taxonomic assignment) experimental stages (Brooks et al., 2015; Parada et al., 2016; Thijs et al., 2017). Despite known biases, 16S rRNA gene based amplicon sequencing has been used successfully for quantitative assessment of the ubiquitous taxa in the bacterial community (Ibarbalz et al., 2014). However, to our knowledge, the literature provides no assessment of the reliability of universal amplicon sequencing for quantifying specific guilds. Therefore, one of the aims of this study was to fill this gap.

Here, using nitrifiers as model guilds, we compared bacterial 16S rRNA gene amplicon sequencing (further referred to as "universal approach") to a guild-targeted approach in their abilities to infer relative abundance, diversity, richness, and composition. Nitrifying microbial guilds are a key to ammonium oxidation in many natural and engineered ecosystems. They were traditionally strictly divided into ammonia-oxidizing prokaryotes and nitrite oxidizing bacteria (NOB) before the recent discovery of *Nitrospira* types that can fully oxidize ammonia (Comammox; Daims et al., 2015; van Kessel et al., 2015; Palomo et al., 2016). To target key nitrifying guilds, we used the functional genes *amoA* for ammonia oxidizing bacteria (AOB) and archaea (AOA); and *nxrB* for the main NOB genera *Nitrospira* and *Nitrobacter* (Arp and Stein, 2003; Francis et al., 2007; Vanparys et al., 2007; Pester et al., 2013).

2. Material and methods

2.1. Biomass sampling and DNA extraction

The biomass originated from the top (0–10 cm) layer of the after filter (AF) of 4 drinking water treatment plants (DWTP) in Denmark and from the nitrifying reactor (NR) of wastewater treatment plants (WWTP) from Denmark and Sweden and from an anammox reactor (Sjolunda (AR)) in a WWTP from Sweden (Fig. S1). Three samples for each plant were used for qPCR. For the 16S rRNA gene amplicon sequencing, five samples were analyzed for three DWTP (Glostrup, Hillerod, and Odense (AF)) and single samples for all other plants.

DNA was extracted from 0.5 g of wet drained sand from DWTP and 0.5 g sludge from WWTP using FastDNA™ spin kit for soil (MP Biomedicals, Solon, OH, USA) according to manufacturer's instructions. Duplicate DNA extractions were done for each sample. DNA concentration was estimated using NanoDrop ND-1000 UV-VIS Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and the extracts were stored at −20 °C until further analysis.

2.2. qPCR, PCR and amplicon sequencing

Quantitative PCR (qPCR) analyses were conducted on a Chromo4 thermocycler using Opticon Monitor version 3 (BioRad). Each qPCR reaction contained 12.5 μL of 2× iQ SYBR green Supermix (Bio-Rad Laboratories, Hercules, California, United States), 500 mM primer, DNA template (10 ng), and DNA/RNA-free water (Qiagen, Hilden, Germany) to 25 μL. For each sample, DNA with average concentration ~26 and ~68 ng/μl for sand and sludge samples respectively was sent for 16S rRNA gene, *amoA*, and *nxrB* PCR, purification and amplicon sequencing (2 × 300 nucleotides) using the Illumina MiSeq platform at the DTU Multi Assay Core Center (Kgs. Lyngby, DK). For the universal 16S rRNA gene sequencing, primers (Bakt_341F and Bakt_805R, spanning the V3-V4 regions) and PCR conditions were from Herlemann et al., 2011. For *amoA* AOB sequencing and qPCR, primers (amoA-1F and amoA-2R primers) and PCR conditions were from Rotthauwe et al., 1997a, 1997b. For *nxrB Nitrospira* sequencing and qPCR, primers (nxrB169f and nxrB638r primers) and PCR conditions were from Pester et al., 2013. Primers (CTO189fA-B, CTO189fC, and RT1r) and PCR conditions for AOB 16S rRNA gene qPCR were from Kowalchuk et al., 1997. Quantitative PCR efficiencies and correlation coefficients obtained from the standard curves are in Table S1.

2.3. Bioinformatics analysis

Sequences generated as paired FASTQ files were processed using DADA2 (Version - 1.4; Callahan et al., 2016). DADA2 was preferred over other clustering-based methods as it recovers sequence variants that can vary by as little as a single nucleotide and thus avoids aggregation of sequences at arbitrary identity cutoff. DADA2 was used for quality filtering, trimming, de-replicating the reads, for inferring sequence variation through default error model parameters, for merging paired reads, for removing chimera, and for assigning taxonomy using Silva reference database v123 for 16S rRNA gene and custom taxonomy files for *amoA* and *nxrB*. A very similar pipeline was used for *amoA* and *nxrB*, except that for *amoA* only the forward reads (237 bp after trimming for quality) were analyzed due to the amplicon length. We applied frameshift correction (from package DECIPHER version 2.6; Wright, 2016) and discarded sequences that were too distant from our reference sequences. The data obtained from DADA2 was analyzed using Phyloseq R Package (Version - 1.7.12; McMurdie and Holmes, 2013). Raw sequence files were deposited into the sequence read archive at GenBank under the study accession number SRP127282.

2.4. Relative abundance estimation of AOB and Nitrospira

Universal 16S rRNA gene amplicon sequencing data was used to perform CaRcone analysis to obtain the average 16S rRNA gene copies per genome in the amplicon libraries (R script https://github.com/ardagulay/CaRcone—Community-average-rRNA-gene-copy-nr-estimator). The total numbers of bacteria per gram of biomass were derived by normalizing the gene copy number from qPCR with 1.75 16S rRNA gene copies per genome obtained from the CaRcone analysis. The abundance of AOB and *Nitrospira* cells per gram of biomass was obtained by assuming 2 *amoA* and 2 *nxrB* copies per genome (McTavish et al., 1993; Lücker et al., 2010). Thus, comparing *nxrB* for *Nitrospira* and *amoA* for AOB with 16S rRNA gene was feasible as the abundance values obtained from two genes were normalized per genome.

The relative fractions of AOB and *Nitrospira* were calculated by dividing the normalized abundance estimates obtained from *amoA* and *nxrB* based qPCR by that obtained from 16S rRNA gene based total bacteria qPCR. Similarly, AOB and *Nitrospira* abundance estimates obtained from *amoA* and *nxrB* based sequencing were compared to their respective values obtained after dividing reads assigned to AOB and *Nitrospira* in the 16S rRNA gene based total bacteria sequencing by the total number of reads.

2.5. Statistical analysis

The alpha diversity metrics, Shannon diversity and observed richness were calculated using the "estimate_richness" function in the Phyloseq R package version (Version - 1.7.12; McMurdie and Holmes, 2013). The estimates of Shannon diversity after rarefaction were calculated by performing rarefaction using the minimal number of sequences in each gene library, as we would lose a large amount of sequences from *amoA* and *nxrB* libraries if we rarified to the minimum number of sequences assigned to AOB and *Nitrospira* from the 16S rRNA gene library. For phylogenetic diversity calculation, first the sequences were aligned using the "AlignSeqs" function in the DECIPHER R package (Version 2.0.2; Wright, 2015), then a neighbor joining (NJ) tree was constructed using phangorn R package (Version 2.3.1; Schliep, 2011), phylogenetic diversity (Faith's PD; the sum of the branches of the phylogenetic tree) was then calculated using the PhyloMeasures R package (Version 2.1; Faith, 1992; Tsirogiannis and Sandel, 2016).