Marine Genomics 29 (2016) 39–43

Contents lists available at ScienceDirect

Marine Genomics

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

Method paper

Choice of molecular barcode will affect species prevalence but not bacterial community composition

Karen Lebret [a,b], Joanna Schroeder [a], Cecilia Balestreri [a], Andrea Highfield [a], Denise Cummings [c], Tim Smyth [c], Declan Schroeder [a,*]

[a] Marine Biological Association, Citadel Hill, Plymouth, UK
[b] Department of Ecology and Genetics/Limnology, Uppsala University, Uppsala, Sweden
[c] Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth, UK

ARTICLE INFO

Article history:
Received 16 May 2016
Received in revised form 5 September 2016
Accepted 6 September 2016
Available online 16 September 2016

Keywords:
Marine bacterial community
16S
Pyrosequencing
Illumina

ABSTRACT

The rapid advancement of next generation sequencing protocols in recent years has led to the diversification in the methods used to study microbial communities; however, how comparable the data generated from these different methods are, remains unclear. In this study we compared the taxonomic composition and seasonal dynamics of the bacterial community determined by two distinct 16s amplicon sequencing protocols: sequencing of the V6 region of the 16s rRNA gene using 454 pyrosequencing vs the V4 region of the 16s rRNA gene using the Illumina Hiseq 2500 platform. Significant differences between relative abundances at all taxonomic levels were observed; however, their seasonal dynamics between phyla were largely consistent between methods. This study highlights that care must be taken when comparing datasets generated from different methods.

1. Introduction

In recent years, studies based on large next generation sequencing datasets have unveiled the extensive diversity and complex structure of microbial communities in the oceans, and the potential consequences on the functioning of the ecosystem such as impact on biogeochemical cycles (Lima-Mendez et al., 2015; Sunagawa et al., 2015). Such studies provide unprecedented insights into the key players and the biological processes in the open ocean.

Underpinning this progress is the rapid advancement in sequencing technologies, leading to the rapid shift in the methodology used to study the structure of microbial communities. Early studies are dominated by the use of the 454 pyrosequencing technology, but in recent years, the Illumina platform have been preferred over 454 pyrosequencing due to its lower cost, lower error rate, and higher throughput (Glenn, 2011). In addition, new primer sets have been developed to cover distinct regions of the 16s rRNA gene allowing better taxonomic assignments and better coverage of the taxonomic diversity (Apprill et al., 2015; Caporaso et al., 2011; Parada et al., 2016). Several 16s rRNA regions (V4, V6, V7, or V9 for instance) have been targeted in studies investigating the composition of bacterial communities. Recent studies have shown that the sequencing of the V4 region was the most reliable to describe the diversity and composition of the bacterial communities (Ghyselinck et al., 2013; Tremblay et al., 2015). Thus protocols have evolved rapidly, however, little is known about our ability to compare microbial communities analyses across distinct sequencing protocols.

A few studies have investigated the impact of samples processing (primers and sequencing technology) on the microbial community structure (Caporaso et al., 2012; Claesson et al., 2010; Nelson et al., 2014). However, these studies usually focus on the analyses of a single sample or very different samples, and have not investigated the impact of the methods on the observed temporal dynamics in microbial communities.

In this study we compared the bacterial community structure and dynamic over a year at one sampling location using two 16s rRNA sequencing protocols to evaluate the impact of the sequencing method on the relative abundances and seasonal dynamics.

2. Methods

2.1. Sampling

Water samples were collected from the surface at the L4 sampling site (50°15.00′N, 4°13.02′W) of the Western Channel Observatory (http://www.westernchannelobservatory.org.uk, accessed on March 31st 2016) between January 2008 and December 2008. The sampling

* Corresponding author.
E-mail address: dsch@mba.ac.uk (D. Schroeder).

http://dx.doi.org/10.1016/j.margen.2016.09.001

was performed on 17 sampling dates during this period (Table 1). For each sampling occasion, 5 l of water were filtered through a 0.22 μm Sterivex cartridge (Millipore), which was then stored at −80 °C until processing.

2.2. DNA extraction and sequencing

DNA was extracted for each sample from the filters according to Neufeld et al. (2007) except that the filter was first removed from the Sterivex casing and transferred to a sterile 2 ml container. In order to compare two sequencing protocols, 12 samples were processed for sequencing of the V6 region of the 16s rRNA gene using 454 sequencing pyrosequencing (hereafter referred to as V6-454 samples) according to Gilbert et al. (2009) and Huber et al. (2007), and 11 samples (Table 1) were processed and barcoded according to Caporaso et al. (2011) (hereafter referred to as V4–Illumina samples). Within these samples, 6 were sequenced using both methods independently.

For the V6-454 samples, the V6 region was amplified using a pool of 5 forward primers (967F-PP 5′-gcctccctcgcgccatcagCNACGCGAAGAACCTTANC-3′; 967F-UC1 5′-gcctccctcgcgccatcagCAACGCGAAAAACCTTACC-3′; 967F-UC2 5′-gcctccctcgcgccatcagCAACGCGCAGAACCTTACC-3′; 967F-UC3 5′-gcctccctcgcgccatcagATACGCGARGAACCTTACC-3′; 967F-AQ 5′-gcctccctcgcgccatcagCTAACCGANGAACCTYACC-3′) and 4 reverse primers (1046R 5′-gccttgccagcccgctcagCGACAGCCATGCANCACCT-3′; 1046R-PP 5′-gccttgccagcccgctcagCGACAACCATGCANCACCT-3′; 1046R-AQ1 5′-gccttgccagcccgctcagCGACGGCCATGCANCACCT-3′; 1046R-AQ2 5′-gccttgccagcccgctcagCGACGACCATGCANCACCT-3′) according to Huber et al. (2007). The samples were sequenced using the 454 GS-flx platform and the LR70 kit.

For the V4–Illumina samples, the V4 region of the 16s rRNA gene was amplified using the forward primer 515F (5′-GTGCCAGCMGCCGCGGTAA-3′) and the reverse primer 806R (5′-GGACTACHVGGGTWTCTAAT-3′) according to Caporaso et al. (2011). Replica real-time PCRs were run alongside the samples destined for Illumina sequences to ensure that the Illumina PCR samples were removed during the log amplification phase of the PCR. The multiplex sequencing of the V4 region of the 16s rRNA gene was performed using Illumina HiSeq 2500 at the University of Exeter sequencing service facility.

For the V4–Illumina samples, the raw sequences have been deposited and are available at the European Nucleotide Archive (ENA) under the accession number PRJEB14618. For the V6–454 samples, the raw sequences are available at ENA under the accession number ERP000118.

Table 1
Sequencing method used for each sampling date (x indicates the samples used for the sequencing).

Sampling date	V6–454	V4–Illumina
28/01/2008		x
20/02/2008	x	x
05/03/2008	x	
17/03/2008	x	x
21/04/2008	x	
06/05/2008	x	
28/05/2008		x
02/06/2008	x	
23/06/2008	x	x
21/07/2008	x	x
20/08/2008	x	x
22/09/2008	x	x
21/10/2008		x
27/10/2008	x	
17/11/2008		x
08/12/2008	x	
22/12/2008		x

2.3. Sequence processing and data analyses

The sequences obtained from the 454 sequencing were processed as described in (Gilbert et al., 2010; Gilbert et al., 2012). For the present study a subsampled OTU table (to 4101 sequences per samples) was used for further analyses.

For the V4-Illumina sequences, the quality of the HiSeq 2500 pair-end sequences was checked using Fastqc. Due to low quality of Read 2, only Read 1 was processed further. The primer and adaptor sequences were removed from the reads, and the sequences were trimmed to the same length (80 bases). The sequences were then processed using Qiime 1.8 (Caporaso et al., 2010), OTUs cluster with 97% identity, and the Silva reference database (release 119) was used to perform the taxonomic annotation. OTUs assigned to Archaea, mitochondria or chloroplast were removed for downstream analyses. To avoid bias due to differences in sequencing depth among samples, the OTU table was subsampled to the lowest number of sequences of a sample (79,302 sequences). To characterize the total diversity at the order level of the V4-Illumina dataset, the 200 most abundant OTUs (representing approximately 93% of the sequences) were extracted, the number of reads for each order were summed among samples to obtain the total yearly abundances and the diversity at order level was visualized using Krona (Ondov et al., 2011).

To compare the community composition and dynamics over the year 2008 of both datasets, the relative abundances of main bacterial groups (phylum) were calculated on the datasets containing all the bacterial OTUs. For each bacterial group a polynomial-fitting curve was applied to the data represented over time; this was done to highlight the seasonal dynamics of each group.

3. Results and discussion

The Illumina Hiseq method allowed for a greater sequencing depth than 454 pyrosequencing (increased by approximately 1830%), resulting in the number of OTUs identified to increase by 100% (Table 2). The increase sequencing depth has highlighted the presence of a hidden diversity in the 2008 samples sequenced using the 454 pyrosequencing method. In addition, the increased sequencing depth has allowed for the reduction in the percentage of singletons in the data set; a reduction of nearly half the singletons (Table 2). Increased sequencing depth allowed not only a reduction in the number of spurious OTUs (Pinto and Raskin, 2012) but it permits a better identification of true rare OTUs over sequencing errors.

For the V4-Illumina method, the 200 most abundant OTUs were dominated by the class of the Bacteroidetes (with the order of Flavobacteriales), followed by the Alphaproteobacteria (with the orders of Rhodobacterales and Rickettsiales with SAR11) and the Gamma Proteobacteria (Fig. 1). These taxa were also the taxa that dominated the V6-454 bacterial community as reported by (Gilbert et al., 2012), although some differences in the relative abundances were observed in some of these taxa (Fig. 2). For both sequencing strategies, the relative abundances of Verrucomicrobia and Cyanobacteria were similar.

Table 2
Summary of the OTU dataset obtained with the V6–454 and V4–Illumina methods.

Method	Level of sub-sampling (number of sequences)	Number of OTUs	% singletons (in the subsampled dataset)
V6–454[a]	4101	1459	48.5%
V4–Illumina	79,302	2919	25.6%

[a] Taken from (Gilbert et al., 2010; Gilbert et al., 2012).