Journal of Microbiological Methods 135 (2017) 77–84

Contents lists available at ScienceDirect

Journal of Microbiological Methods

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

A quantitative SMRT cell sequencing method for ribosomal amplicons

Bethan M. Jones [a,*,1], Adam B. Kustka [a]

[a] Department of Earth and Environmental Sciences, Rutgers University-Newark, Newark, NJ 07102, United States

ARTICLE INFO

Article history:
Received 24 October 2016
Received in revised form 27 January 2017
Accepted 28 January 2017
Available online 1 February 2017

Keywords:
18S
qSMRT
Quantitative
Ribosomal
Ross Sea
Sequencing
Silicoflagellate
SMRT

ABSTRACT

Advances in sequencing technologies continue to provide unprecedented opportunities to characterize microbial communities. For example, the Pacific Biosciences Single Molecule Real-Time (SMRT) platform has emerged as a unique approach harnessing DNA polymerase activity to sequence template molecules, enabling long reads at low costs. With the aim to simultaneously classify and enumerate in situ microbial populations, we developed a quantitative SMRT (qSMRT) approach that involves the addition of exogenous standards to quantify ribosomal amplicons derived from environmental samples. The V7–9 regions of 18S SSU rDNA were targeted and quantified from protistan community samples collected in the Ross Sea during the Austral summer of 2011. We used three standards of different length and optimized conditions to obtain accurate quantitative retrieval across the range of expected amplicon sizes, a necessary criterion for analyzing taxonomically diverse 18S rDNA molecules from natural environments. The ability to concurrently identify and quantify microorganisms in their natural environment makes qSMRT a powerful, rapid and cost-effective approach for defining ecosystem diversity and function.

1. Introduction

The enumeration and classification of organisms in the environment is crucial for our understanding of ecosystem structure and function. Over the past two decades, advances in genomic technologies have revolutionized our abilities to characterize microbial communities. Previously, 16S and 18S ribosomal gene cloning and Sanger sequencing (Sanger et al., 1977) were used to discover representatives of many uncultured clades from environmental libraries, revealing extensive bacterial and eukaryotic diversity (e.g. Giovannoni et al., 1990; Pace, 1997; Lopez-Garcia et al., 2001; Moon-van der Staay et al., 2001; Rappé and Giovannoni, 2003). Second-generation sequencing techniques, such as 454 pyrosequencing (Margulies et al., 2005), Illumina sequencing (Bentley et al., 2008) and Ion Torrent (Rothberg et al., 2011) have since yielded sequence data from environmental microbial communities on an unprecedented scale (e.g. Sogin et al., 2006; Gilbert and Dupont, 2011). It is now possible to generate millions of sequences from hundreds of samples (e.g. de Vargas et al., 2015), giving us insights into the cosmopolitan nature of microbial communities and their diverse functions (Worden et al., 2015).

Despite our newfound ability to rapidly characterize microbial communities, it is still surprisingly difficult to obtain highly resolved abundance information for many groups of organisms in the marine environment. For example, HPLC data on photosynthetic pigments can only provide broad phylogenetic resolution. Additionally, chlorophyll-based enumeration techniques are often inappropriate because chlorophyll content does not always reflect biomass concentration (Geider, 1987; Halsey and Jones, 2015). Culture-based methods are often not suitable because most microbes are not amendable to laboratory cultivation (Staley and Konopka, 1985). Flow cytometry has been used to address fundamental questions regarding the abundance and distribution of microbial communities (Legendre et al., 2001), but autofluorescence and side scatter cannot finely resolve taxa. Microscopic enumeration is time-consuming and challenging, considering most prokaryotes and many protists can only be discriminated by cellular characteristics that often are non-distinct with standard bright-field microscopy (Simon et al., 1995; Massana, 2011). Likewise, automated imaging systems (e.g. Flow Cytobot, Olson et al., 2003; Imaging Flow Cytobot, Olson and Sosik, 2007) provide excellent opportunities for in situ imaging and enumeration but many phytoplankton species, particularly picoeukaryotes, are cryptic and difficult to solely identify using image-based analyses. Fluorescent probes (Amann et al., 1995) and q-PCR based techniques (e.g. Howard et al., 2012) have been used to reveal fine-scale microbial population dynamics but are not amenable to high-throughput studies and require a priori knowledge of nucleotide sequences.

Understanding species abundance is fundamental to constrain rate-based measurements, assisting interpretations of carbon sequestration and the flow of energy through ecosystems. Quantification of microbial communities using high-throughput sequencing techniques would provide a rapid and powerful approach for enumeration, coupling abundance

* Corresponding author.
E-mail address: bethan.jones@science.oregonstate.edu (B.M. Jones).
[1] Present address: Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR 97331, United States of America.

http://dx.doi.org/10.1016/j.mimet.2017.01.017
0167-7012/

and composition metrics. However, sequence proportions derived from second-generation sequencing platforms do not always correlate with the actual proportional abundance of species in the environment (Amend et al., 2010; Deagle et al., 2013). Second-generation techniques are amplification-based, so quantitative inaccuracies may be a result of amplicon bias during sequencing (Quail et al., 2012). Additionally, quantification artifacts may result if amplicon-generating or sequencing reactions are stopped during the quasi-linear or plateau stage of amplification.

There is no amplification step during Single Molecule Real Time (SMRT) sequencing. Instead, sequencing data is directly obtained as nucleotides are incorporated into single template molecules (Eid et al., 2009). The Pacific Biosciences SMRT platform has a single pass read error rate of ~15% due to misincorporation of incorrect nucleotides (Eid et al., 2009; Koren et al., 2012), but this is overcome by ligating DNA molecules into hairpin structures to form circularized fragments (SMRT bells), enabling multiple pass coverage. Because the error is random, this process, termed Circular Consensus Sequencing (CCS), improves empirical quality scores (Travers et al., 2010). A handful of studies have already applied SMRT cell ribosomal phylotyping to characterize bacterial communities (e.g. Marshall et al., 2012; Mosher et al., 2013, 2014; Singer et al., 2016).

We applied the SMRT CCS method to characterize and quantify protistan sequences from environmental samples collected in the Ross Sea between January–February 2011. Following DNA extraction, environmentally-derived 18S rDNA molecules were quantified via qPCR. Three exogenous standards were subsequently added to the unamplified environmental DNA. PCR amplicons of the hypervariable V7–9 region of 18S rDNA (454–615 bp) were generated from the resulting DNA mixture (85% environmental DNA: 15% standards). Amplification was stopped before PCR reactions entered a non-linear stage and amplicons were sequenced using the Pacific Biosciences RS II platform. Standards were quantitatively recovered using bioinformatic techniques, thus allowing the back-calculation of OTU-specific copy numbers of 18S rDNA per milliliter of seawater. This new quantitative SMRT sequencing approach (termed "qSMRT") provides a reproducible and low-cost method for the concurrent identification and quantification of rDNA derived from microbial communities.

2. Materials and methods

2.1. Sample collection

Samples were collected at various locations in the Ross Sea in January–February 2011 during research conducted onboard the R/V Nathaniel B. Palmer (Fig. 1). This was part of the larger, multidisciplinary Slocam Enhanced Adaptive Fe Algal Research in the Ross Sea (SEAFAReRS) project (cruise NBP 11–01). Known volumes of seawater (1 − 2 L) were taken directly from 12 L Go-Flo bottles attached to a trace metal clean rosette (Measures et al., 2008), gently filtered onto 47 mm Supor 200 polyethersulfone filters (Pall Life Sciences, Ann Arbor, MI) and stored at −80 °C prior to DNA extraction. Ancillary data for all samples collected can be found in Kustka et al. (2015).

Table 1
Sequences of primers developed and used to amplify the genomic hypervariable V7–9 regions of 18S SSU rDNA.

	Sequence	Reference
Forward primer (V7–9F)	5′-ATG GCC GTT CTT AGT TGG TGG-3′	This study
Reverse primer (V7–9R)	5′-CCT TCY GCA GGT TCA CCT AC-3′	Medlin et al. (1988), Amaral-Zettler et al. (2009)
Forward primer barcode 1 (V7–9F1)	5′-TGTAGTCT-ATG GCC GTT CTT AGT TGG TGG-3′	This study
Forward primer barcode 2 (V7–9F2)	5′-GTTCTCTT-ATG GCC GTT CTT AGT TGG TGG-3′	This study
Reverse primer barcode 1 (V7–9R1)	5′-ATGC-CCT TCY GCA GGT TCA CCT AC-3′	This study
Reverse primer barcode 2 (V7–9R2)	5′-TTTT-CCT TCY GCA GGT TCA CCT AC-3′	This study

2.2. Primers

We targeted a continuous section of the 18S SSU rRNA gene that encompasses the hypervariable V7–9 region of the molecule. Primer design followed recommended guidelines for highly efficient and reproducible amplification (Bustin, 2000). Full-length 18S rRNA gene sequences (2383) that aligned to the reverse primer sequences (V7–9R) of Medlin et al. (1988) and Amaral-Zettler et al. (2009) were compiled from v108 of the eukaryotic ARB-SILVA database (Quast et al., 2013). Following alignment using the ClustalX algorithm (Larkin et al., 2007), the forward primer (V7–9F) was designed by locating conserved regions (>99% similarity of all sequences) on the 5′ end of the V7 region. Sequences from the ARB-SILVA database targeted by our primers were 447–553 bp in length.

To enable duplexed SMRT cell sequencing, 4 bp barcodes were placed on the 5′- end of primer V7–9R (V7–9R1 and V7–9R2; Table 1) for the first 16 samples sequenced. For the remaining samples, 8 bp barcodes were placed on the 5′ – end of primer V7–9F (V7–9F1 and V7–9F2; Table 1). During this initial phase (2012), the SMRT compatible barcodes designed by Pacific Biosciences were not available. Approximately 90% of sequences were recovered after processing, irrespective of barcode sequence.

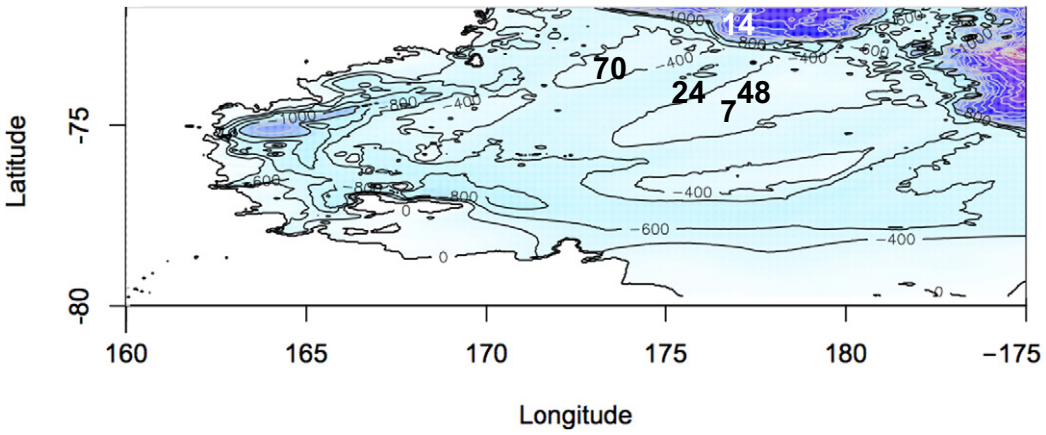

Fig. 1. Bathymetric map of the Ross showing location of the stations sampled for qSMRT cell sequencing. Samples were collected at Pennell Bank (station 7), north Pennell Bank (station 48), Joides Basin (station 24; referred to in Kustka et al. [2015] as Joides Trough), the shelf break (station 14) and Mawson Bank (station 70). Map plotted using marmap (Pante and Simon-Bouhet, 2013).