Journal of Microbiological Methods 91 (2012) 527–536

Contents lists available at SciVerse ScienceDirect

Journal of Microbiological Methods

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

# A combined sequence-based and fragment-based characterization of microbial eukaryote assemblages provides taxonomic context for the Terminal Restriction Fragment Length Polymorphism (T-RFLP) method

Diane Y. Kim [a,*], Peter D. Countway [b], Warren Yamashita [a], David A. Caron [a]

[a] Department of Biological Sciences, University of Southern California, 3616 Trousdale Parkway, Los Angeles, CA 90089-0371, USA
[b] Bigelow Laboratory for Ocean Sciences, 60 Bigelow Drive, East Boothbay, ME 04544, USA

ARTICLE INFO

*Article history:*
Received 23 August 2012
Received in revised form 21 September 2012
Accepted 22 September 2012
Available online 3 October 2012

*Keywords:*
T-RFLP
18S rDNA
Microbial eukaryotes
Diversity
Protist

ABSTRACT

Microbial eukaryotes in seawater samples collected from two depths (5 m and 500 m) at the USC Microbial Observatory off the coast of Southern California, USA, were characterized by cloning and sequencing of 18S rRNA genes, as well as DNA fragment analysis of these genes. The sequenced genes were assigned to operational taxonomic units (OTUs), and taxonomic information for the sequence-based OTUs was obtained by comparison to public sequence databases. The sequences were then subjected to *in silico* digestion to predict fragment sizes, and that information was compared to the results of the T-RFLP method applied to the same samples in order to provide taxonomic context for the environmental T-RFLP fragments. A total of 663 and 678 sequences were analyzed for the 5 m and 500 m samples, respectively, which clustered into 157 OTUs and 183 OTUs. The sequences yielded substantially fewer taxonomic units as *in silico* fragment lengths (*i.e.*, following *in silico* digestion), and the environmental T-RFLP resulted in the fewest unique OTUs (unique fragments). Bray–Curtis similarity analysis of protistan assemblages was greater using the T-RFLP dataset compared to the sequence-based OTU dataset, presumably due to the inability of the fragment method to differentiate some taxa and an inability to detect many rare taxa relative to the sequence-based approach. Nonetheless, fragments in our analysis generally represented the dominant sequence-based OTUs and putative identifications could be assigned to a majority of the fragments in the environmental T-RFLP results. Our empirical examination of the T-RFLP method identified limitations relative to sequence-based community analysis, but the relative ease and low cost of fragment analysis make this method a useful approach for characterizing the dominant taxa within complex assemblages of microbial eukaryotes in large datasets.
© 2012 Elsevier B.V. All rights reserved.

## 1. Introduction

Protists are ubiquitous in marine environments and are responsible for a range of ecological roles including primary production, nutrient regeneration, symbioses and the transfer of energy and carbon to higher trophic levels (Sherr et al., 2007; Caron et al., 2012). The diversity and functional roles of protists have been topics of interest since microbial eukaryotes were first observed by Antonie van Leeuwenhoek in the 17th century. Morphological features have constituted the 'gold standard' for the taxonomy of these species since that time but DNA sequence information increasingly has been employed to augment morphology-based taxonomic schemes (Adl et al., 2005), to elucidate phylogenetic relationships among protistan lineages (Tekle et al., 2009), and most recently to complement morphology-based approaches for the study of protistan diversity and ecology (Caron et al., 2009; Marande et al., 2009).

The development and application of a DNA-based taxonomy as a tool to study protistan ecology are alluring because of the potential to characterize natural protistan assemblages using a single method. Protistan communities in nature are composed of species that span several orders of magnitude in size, and many can be morphologically nondescript (especially minute forms) and/or present at abundances that are below the detection limits of traditional methods. Characterization of the entire protistan assemblage in a sample by traditional methods is therefore exceedingly difficult, involving multiple methods for collection, preservation and observation, thus limiting the number of samples that can be analyzed and the completeness of the assessment (Caron, 2009). However, ecological studies to understand the diversity and spatiotemporal variability of protistan assemblages require the collection and examination of large numbers of samples. Sequence-based approaches offer the possibility of a practical and potentially widely applicable alternative to traditional approaches for characterizing the diversity and taxonomic composition of protistan assemblages. The application of these methodologies within the last decade has revealed an unexpectedly large diversity of protists in numerous marine environments (Diez et al., 2001b; Lopez-Garcia et al., 2001; Amaral-Zettler

* Corresponding author. Tel.: +1 213 821 1800; fax: +1 213 740 8123.
*E-mail address:* dianekim@usc.edu (D.Y. Kim).

0167-7012/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.mimet.2012.09.026

et al., 2002; Edgcomb et al., 2002; Moreira and Lopez-Garcia, 2002; Lopez-Garcia et al., 2003; Stoeck and Epstein, 2003; Gast et al., 2004; Massana et al., 2004; Romari and Vaulot, 2004; Countway et al., 2005; Worden, 2006).

A variety of genetic methods have recently been developed for assessing the diversity of natural microbial assemblages. These methods can be grouped broadly into approaches that rely on the analysis of sequences from genes such as small subunit ribosomal RNA or cytochrome oxidase, and the analysis of DNA fragment sizes from genes or other genomic regions. Analyses of community structure and diversity using DNA sequence information rely on grouping sequences obtained from environmental samples into operational taxonomic units (OTUs) based on sequence similarity or phylogenetic relatedness. For microbial eukaryotes, some studies have attempted to delineate OTUs in a manner that approximates morphological species distinctions (Caron et al., 2009; Nebel et al., 2011). One recent analysis employed well-curated 18S rRNA gene sequences from GenBank for approximating protistan species using an automated OTU calling program, Microbial Eukaryote Species Assignment (MESA) (Caron et al., 2009).

Two popular DNA fragment approaches that have been employed are Terminal Restriction Fragment Length Polymorphism (T-RFLP) and Automated Ribosomal Intergenic Spacer Analysis (Liu et al., 1997; Fisher and Triplett, 1999). T-RFLP has been commonly used to characterize microbial eukaryote assemblages (Countway et al., 2005; Dopheide et al., 2008; Euringer and Lueders, 2008; Vigil et al., 2009; Joo et al., 2010; Steele et al., 2011; Balzano et al., 2012). OTU calling for the T-RFLP method is predicated on the assumption that different species will yield DNA fragments of unique length. Limitations of the method relate to the small number of taxonomic units (unique fragment sizes) that can be detected in T-RFLP patterns, and the lack of taxonomic identity of the species giving rise to particular fragments (Dunbar et al., 2000; Osborn et al., 2000). These caveats reduce the usefulness of the method (and DNA fragment-based approaches in general) for assessing the taxonomic composition of microbial assemblages. However, the relatively low cost and rapid analysis relative to sequencing make T-RFLP a useful method for analyzing large ecological datasets.

DNA fragment-based and sequence-based analyses of microbial eukaryote assemblages were carried out in order to provide taxonomic context for the T-RFLP method. Samples were collected from 5 m and 500 m as a part of the USC Microbial Observatory at the San Pedro Ocean Time-series (SPOT) station located in the eastern North Pacific, where 237 samples from multiple water column depths have been analyzed using the T-RFLP method (Kim et al., in review). Environmental clone libraries of 18S rRNA genes were constructed and sequenced, and OTUs were assembled and taxonomies assigned. The sequences were then used to create *in silico* restriction digests in order to assign taxonomic identity to fragments observed in the T-RFLP patterns obtained from the same samples. Environmental T-RFLP fragments in our analysis generally represented the dominant sequence-based OTUs, and fragment lengths obtained from the *in silico* digests of sequence-based OTUs allowed confident assignments of taxonomic identities to many of the fragments from the T-RFLP analysis.

## 2. Materials and methods

### 2.1. Sample collection

Seawater samples were collected from 2 depths (5 m and 500 m) at the San Pedro Ocean Time-series (SPOT) station in the eastern North Pacific (118° 24′ W, 33° 33′ N) on October 29, 2001 using 10 l Niskin bottles attached to a CTD sampler rosette. Seawater samples were pre-screened through 200 μm and 80 μm mesh screening by gravity-filtration to reduce the contribution of metazoa to subsequent DNA analyses. The <80 μm filtrate was collected in acid-washed 20 l polycarbonate carboys and 2 l of the filtrate from each depth was vacuum filtered (~5 hg) onto 47 mm GF/F filters (Whatman) on board the ship. The filters were immediately transferred to 15 ml BD Falcon tubes (BD Biosciences, San Jose, CA) containing 2 ml of 2× lysis buffer (100 mM Tris (pH 8), 40 mM EDTA (pH 8), 100 mM NaCl, 1% SDS) and flash frozen in liquid nitrogen.

### 2.2. DNA extraction and purification

DNA was extracted from samples after their return to the lab. A combination of chemical and mechanical methods described in Countway et al. (2005) was used to extract, isolate and purify nucleic acids. DNA pellets were resuspended in TE buffer (10 mM Tris, 1 mM EDTA, pH 7.5) and frozen at −20 °C until further processing.

### 2.3. Cloning and sequencing of full-length 18S rRNA genes

Cloning for the two environmental samples was carried out according to Countway et al. (2005) using Euk-A and Euk-B universal primers (Medlin et al., 1988). Sequencing was carried out by the Joint Genome Institute (JGI, Walnut Creek, CA) using two vector primers (M13F and M13R) and one internal primer (Euk-570F). The resulting contigs were assembled at the JGI, and resulted in a total of 1094 and 823 full-length (or near full-length) 18S rRNA gene sequences for the 5 m and 500 m samples, respectively. The sequences were generated as a part of a larger global survey of protistan diversity and used in the current study to provide taxonomic context for T-RFLP analysis performed on the same samples. The raw assembled reads were screened to trim vector sequences and also checked for correct orientation using Geneious (Drummond et al., 2011) and MOTHUR (Schloss et al., 2009). The sequences were further screened for the Euk-A and the Euk-570R primer sites (allowing up to 3 mismatches for each), to ensure that only sequences that had the T-RFLP primer regions were used to construct a fragment database for assigning identifications to specific fragment lengths, and for comparison to T-RFLP-based characterizations of the microbial eukaryote assemblages. The resulting sequences were checked for chimeras using a local implementation of Pintail (Ashelford et al., 2005) and the SILVA (v102) database for eukaryotes (Pruesse et al., 2007). These quality control measures resulted in 663 sequences for the 5 m sample and 678 sequences for the 500 m sample for further analyses (total of 1341 sequences). The sequences (from Euk-A to Euk-570R) can be found in the GenBank database (http://www.ncbi.nlm.nih.gov/genbank/; Accessions JX841335–JX842675).

### 2.4. OTU calling and taxonomy assignment of 18S rRNA gene sequences

The sequenced regions from Euk-A to Euk-570R were extracted from the 1341 18S rDNA sequences passing quality control, and clustered into operational taxonomic units (OTUs) using the Microbial Eukaryote Species Assignment (MESA) program at a 95% sequence similarity (Caron et al., 2009). The selection of the 95% sequence similarity threshold in MESA is discussed in detail by Caron et al. (2009), and is designed to provide approximately species-level, automated OTU calling for protistan sequences. Each sequence was assigned a taxonomic identification based on the best BLAST+ match to the SILVA (v108) database (Pruesse et al., 2007). BLAST+ results against GenBank (Benson et al., 2010) largely corroborated the results from the SILVA (v108) database and also provided identities to novel or unknown groups such as Group I and Group II alveolates that were not found in SILVA (v108). All sequences within an OTU represented the same high-level phylogenetic group. Genus and species-level identifications were assigned to each sequence-based OTU based on a majority basis.

### 2.5. Terminal Restriction Fragment Length Polymorphism (T-RFLP) analysis

DNA extracts were quantified using PicoGreen (Invitrogen) on a BioRad fluorometer, and approximately 10 ng of DNA template was used for each PCR reaction for T-RFLP analysis. PCR reactions consisted of final concentrations of 0.5 μM of each primer (Euk-A labeled with D4 fluorochrome, 5′-D4-AAC CTG GTT GAT CCT GCC AGT-3′ (Medlin et al., 1988) and unlabeled Euk-570R, 5′GCT ATT GGA GCT GGA ATT AC-3′ (Elwood et al., 1985)), 1× GoTaq Flexi Colorless Buffer (Promega), 2.5 mM $MgCl_2$, 250 μM dNTPs, 300 ng/μl BSA (Sigma), 2.5 U of GoTaq (Promega) in a total reaction volume of 50 μl. The following thermal protocol was used for each PCR: 1× (95 °C for 2 min), 35× (95 °C for 30 s, 55 °C for 30 s, 72 °C for 2 min), and 1× (72 °C for 7 min). Three PCRs were performed for each sample, PCR products were visualized on a 1.2% SeaKem LE agarose gel and the triplicate PCRs were pooled before being purified and concentrated using a QIAquick PCR Purification Kit (Qiagen).

Mung bean nuclease (10 U) and mung bean nuclease buffer (New England Biolabs) were added to the cleaned and concentrated PCR products and incubated for 60 min at 30 °C to remove single-stranded PCR artifacts (Egert and Friedrich, 2003). Products digested with mung bean nuclease were purified and concentrated using the QIAquick PCR purification kit and quantified using PicoGreen on a VersaFluor fluorometer (BioRad). Approximately 300 ng of DNA was used for restriction digests. The *HaeIII* and *MnlI* enzymes were selected based on their ability to produce a large number of unique fragments from *in silico* digestions of 18S rRNA gene sequences obtained from publicly available databases (Countway et al., 2005). The restriction digestion reactions comprised: 10 U each of either *HaeIII* or *MnlI* enzyme (New England Biolabs), Buffer 2 (New England Biolabs), BSA (*MnlI* only) and ~300 ng of DNA in a total reaction volume of 20 μl. All restriction enzymatic reactions were incubated at 37 °C for ~15 h and the enzymes were inactivated by heating at 80 °C and 65 °C for 20 min for the *HaeIII* and *MnlI* enzymes, respectively.

The digested DNA products were precipitated using 20 μg of glycogen (Roche), 2 μl of 3 M sodium acetate (Sigma-Aldrich) and 50 μl of 95% ice-cold ethanol and centrifugation at 14,000 RPM for 15 min. The supernatant was decanted and DNA pellets were rinsed twice with 100 μl of 70% ice-cold ethanol and centrifugation at 14,000 RPM for 5 min. The supernatant was decanted after each round and finally the DNA pellets were air-dried and then resuspended in 40 μl of Sample Loading Solution (SLS; a deionized formamide solution, Beckman-Coulter). Approximately 5–10 μl of the resuspended samples were combined with 0.5 μl of the DNA-Size-Standard-Kit-600 (Beckman Coulter) and SLS up to a volume of 40 μl. The resulting terminally labeled fragments (T-RFs) were detected on a Beckman CEQ8000 capillary gel-electrophoresis system (with a detection range between 60 and 640 bp with 1 bp resolution). The CEQ8000 protocol consisted of: 2.9 kV at 60 °C for 70 s and injection times ranged between 9 and 15 s. The volume for each sample and injection times were adjusted so that maximum relative fluorescent units (rfu) for each sample ranged between 80,000 and 125,000 rfu.

T-RFs from cultures of *Ostreococcus* sp. and *Phaeocystis globosa* were used as positive controls for verifying fragment sizes in the T-RFLP analyses. The 18S rRNA genes from these two species were cloned, and purified plasmids containing the genes were amplified using T-RFLP primers, digested with *HaeIII* and *MnlI* enzymes, and run as standards in parallel with the environmental samples. Fragment sizes of these two species on the CEQ8000 were compared to fragment sizes predicted from *in silico* digestion of the complete, sequenced genes to ensure the accuracy of fragment sizes determined by the CEQ8000.

Raw fragment data were analyzed using the Fragment Analysis module of the CEQ8000 software package (Beckman Coulter) using 1 bp-wide bins and a threshold of 0.5% peak area. Fragment peaks were normalized to the total peak area for each sample to allow comparisons of relative abundance of fragments between samples (Kaplan and Kitts, 2004).

### 2.6. Taxonomy assignment to DNA fragments

A perl script was written to perform an *in silico* digestion of the 1341 sequences obtained by cloning and sequencing, using the *HaeIII* (GG′CC) and *MnlI* ($CCTC(N)_7{}'$ or $'(N)_6GAGG$) restriction sites. The fragments resulting from the *in silico* digestions were grouped according to fragment size (*i.e.* into fragment-based OTUs), and taxonomic identities of these *in silico* generated fragment OTUs were assigned based on the identities assigned to the sequences as described above. In cases where more than a single identity was assigned to a fragment size (the case for the majority of fragment sizes), identity was based on the predominant taxonomic identity of the sequences that generated that fragment size. In addition, each of the ten most populated sequence-based OTUs from each sample was examined to determine if one or more fragment sizes were generated for each sequence-based OTU and to provide a comparison of taxonomic identities assigned to the fragment- and sequence-based OTUs. Most sequence-based OTUs yielded a predominant fragment size, but most of the OTUs also yielded one or more minor fragments.

Taxonomic identities were assigned to fragments generated by the environmental T-RFLP analysis of the two samples using the taxonomies generated from the sequence-based identifications assigned to fragments obtained from the *in silico* digestions of the 1341 sequences. T-RF databases (including the taxonomic assignments for each fragment) generated from the sequences from the two samples were constructed using Microsoft Access. The fragment databases generated from *in silico* digestion of the sequences from 5 and 500 m, and the assignment of putative taxonomies to the T-RFs generated by the environmental T-RFLP analysis, were handled separately for the two samples.

### 2.7. Multivariate analysis

Fragment and sequence data were square-root transformed to down-weight the influence of the most dominant taxa in each sample before Bray–Curtis similarities were calculated in PRIMER (v6) & PERMANOVA+β18 (PRIMER-E Ltd.).

## 3. Results

### 3.1. Sequence-based OTU calling

A total of 663 and 678 full-length (or near full-length) 18S rRNA gene sequences were analyzed from the 5 m and 500 m samples in this study, respectively. The Euk-A to Euk-570R region of these 1341 sequences was extracted, pooled and used for OTU-calling using MESA in order to correspond with the environmental T-RFLP analysis. A total of 307 OTUs were obtained for the combined dataset (Table 1), and the sequence-based OTUs yielded a rank abundance curve with a characteristic shape composed of relatively few abundant taxa and a large number of rare taxa (Fig. 1A). The sequences in the 5 m sample clustered into 157 OTUs while the 500 m sequences clustered into 183 OTUs. The individual rank abundance curves for both samples were similar in shape to the curve including all data, although only 33 OTUs (~18% of the 5 m OTUs and ~21% of the 500 m OTUs) were shared between the two samples (Fig. 1B and C). Community similarity analysis between the two depths based on Bray–Curtis similarity of the sequence-based OTUs was only 19%.

The ten most abundant sequence-based OTUs in the 5 m sample (which comprised 261 or 39% of the 5 m sequences) consisted of five dinoflagellate taxa (*Gyrodinium* spp., *Heterocapsa* spp., two unknown syndiniales and a dinoflagellate most closely related to a species described from the Ross Sea, Antarctica (Gast et al., 2006)), two ciliates (*Strombidium cf. basimorphum* and an unknown ciliate), a chlorophyte

**Table 1**
Summary of the data used to compare sequence-based and fragment-based OTUs: the number of sequence-based operational taxonomic units (OTUs, called at 95% sequence similarity in MESA); the number of *in silico* fragments that resulted from *in silico* digestion of the 5 m and 500 m sequences; the number of fragments arising from environmental T-RFLP analysis of the samples, and the number of matches to the *in silico* database. Summaries are shown for each depth separately and combined. Values in brackets under the combined analysis column are the numbers of OTUs that were found in both samples.

| Method used to derive OTUs | Number of OTUs from 5 m | Number of OTUs from 500 m | Number of OTUs from combined analysis of 5 m and 500 m (shared OTUs in brackets) |
|---|---|---|---|
| Sequence-based | 157 | 183 | 307 (33) |
| *In silico* fragment-based | 104 | 98 | 146 (56) |
| Environmental T-RFLP-based | 38 | 37 | 61 (14) |
| Environmental T-RFLP-based OTUs also found in the *in silico* fragment-based OTUs | 31 | 22 | na |

(*Bathycoccus prasinos*), a chlorarachniophyte (*Partenskyella glossopodia*), and a copepod (Table 2). The ten most abundant sequence-based OTUs in the 500 m sample (which constituted 287 or 42% of the 500 m sequences) consisted of 2 dinoflagellates (*Gyrodinium* spp. and *Karlodinium* spp.), 4 Group II alveolates, 2 polycystines (*Arachnosphaera myriacantha* and *Larcopyle butschlii*), a ciliate (*Varistrombidium kielum*) and a hydrozoan (*Apolemia* sp.) (Table 2).

### 3.2. Predicting fragment sizes from in silico digestion of 18S rDNA sequences

The 1341 18S rRNA gene sequences were subjected to *HaeIII* digestion *in silico* in order to predict and compare the DNA fragment sizes observed in the T-RFLP analysis of the same samples. *HaeIII in silico* digestion resulted in the detection of 34% fewer OTUs (*i.e.* unique fragment sizes) in the 5 m sample (104 *in silico HaeIII* fragments) and 46% fewer OTUs in the 500 m sample (98 *in silico HaeIII* fragments) relative to sequence-based OTUs (Table 1, Fig. 2A and B). *In silico* digestions of the 1341 sequences were also conducted using the *MnlI* restriction site, and yielded 94 fragments for the 5 m sequences and 91 fragments for the 500 m sequences (data not shown). Only results from the analyses conducted using *HaeIII* are reported due to the larger number of unique fragment sizes produced by that enzyme. The distribution of fragment sizes produced by *HaeIII in silico* digestion of the 18S sequences for the two samples resembled a typical rank abundance curve (Fig. 2A). The numbers of *in silico* fragments that represented ≥0.5% of the total number of fragments in the two samples were 34 and 44, similar to the number of fragments generally detected in environmental T-RFLP (see below). Community similarity analysis between the two depths based on Bray–Curtis similarity analysis of the relative abundance of *in silico HaeIII* fragments in the two datasets was 53%.

*In silico* digestion of some of the 18S rRNA gene sequences resulted in either a lack of fragmentation or fragment sizes that were below the detection range for the environmental T-RFLP method used in this study (60 bp). There were 74 and 34 sequences from the 5 m and 500 m samples, respectively, that did not have a *HaeIII* restriction site (in which cases the entire length of the Euk-A and Euk-570R region was used in the *in silico* T-RFLP database).

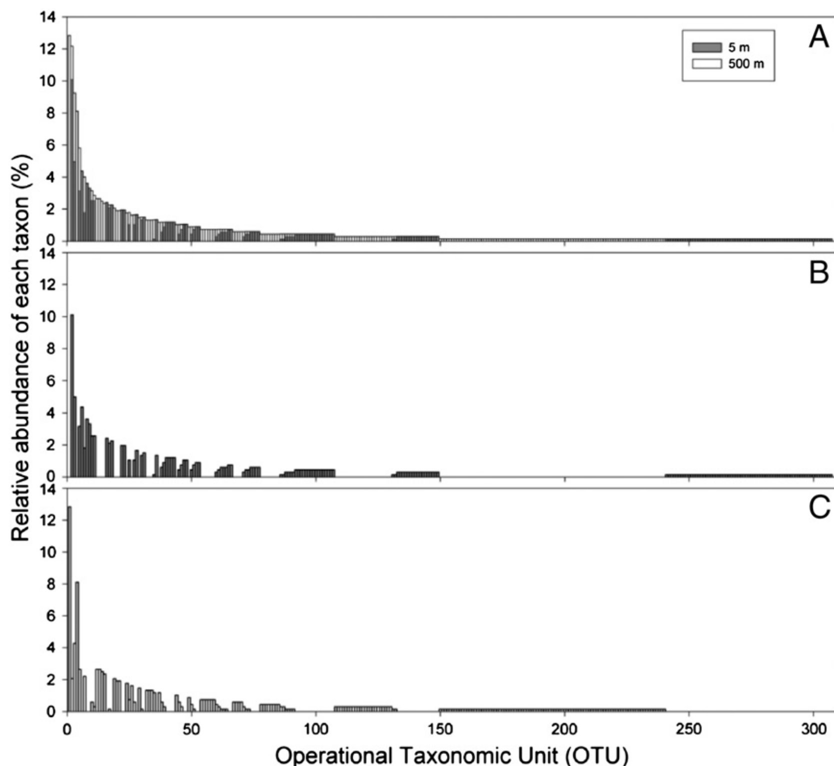


**Fig. 1.** Rank abundance curves for microbial eukaryote assemblages collected from 5 m and 500 m at the USC Microbial Observatory site, based on partial sequences of 18S rRNA genes. Sequences from the total dataset (1341 18S rDNA sequences) were combined for OTU-calling (A), and then separated by sampling depth (B, 5 m; C, 500 m). OTUs in the total dataset were ranked from highest to lowest relative abundance (A), and the x-axes are the same for the three rank abundance curves.

**Table 2**
Summary of the ten most abundant OTUs based on partial 18S sequences obtained for the 5 m and 500 m samples, including the terminally labeled fragment lengths predicted by *in silico* digestion for those sequences, and a comparison of taxonomic identities assigned to the sequence-based OTUs and the dominant fragment size for each OTU (see Fig. 3A and B). The number of sequences in these sequence-based OTUs ranged between 14 and 87. Taxonomic identifications and phylogenetic groups were determined from the sequences in each OTU as described in the Materials and methods. All sequences within an OTU were then subjected to *in silico HaeIII* restriction digestion. Not all sequences within a sequence-based OTU yielded the same fragment size, in which case the dominant fragment size predicted for each OTU was reported ('Predicted *in silico* fragment size'; also reported in Fig. 3). Taxonomies were assigned to the dominant fragment from each OTU based on the taxonomic identity of the sequences yielding that fragment length, as well as the major phylogenetic group for each fragment ('Fragment-based identification'). The phylogenetic group determined for a dominant fragment for each OTU was generally but not always the same (see parentheses in last column).

| Depth (m) | Seq-based OTU (rank abundance) | Number of sequences | Sequence-based identification | | Predicted *in silico* fragment size (bp) | Fragment-based identification | |
|---|---|---|---|---|---|---|---|
| | | | Taxonomy | Phylogenetic group | | Taxonomy | Phylogenetic group |
| 5 | 1 | 67 | *Gyrodinium* spp. | Dinoflagellate | 239 | *Gyrodinium* spp. | Dinoflagellate (49/51) |
| | 2 | 33 | *Strombidium cf. basimorphum* | Ciliate | 273 | *Strombidium* spp. | Ciliate (18/19) |
| | 3 | 29 | *Bathycoccus prasinos* | Chlorophyte | 269 | *Bathycoccus prasinos* | Chlorophyte (27/27) |
| | 4 | 24 | *Paracalanus parvus* | Copepod | 499 | *Paracalanus parvus* | Copepod (14/14) |
| | 5 | 22 | unknown ciliate | Ciliate | 587 | *Strombidium* sp. | Ciliate (20/21) |
| | 6 | 21 | uncultured syndiniales | Dinoflagellate | 279 | uncultured syndiniales | Dinoflagellate (24/26) |
| | 7 | 17 | *Dinophyceae* sp. *RS-24* | Dinoflagellate | 339 | *Pentapharsodinium tyrrhenicum* | Dinoflagellate (42/47) |
| | 8 | 17 | *Heterocapsa* spp. | Dinoflagellate | 337 | *Heterocapsa* spp. | Dinoflagellate (28/32) |
| | 9 | 16 | uncultured syndiniales | Dinoflagellate | 279 | uncultured syndiniales | Dinoflagellate (24/26) |
| | 10 | 15 | *Partenskyella glossopodia* | Cercozoa | 282 | *Partenskyella glossopodia* | Cercozoa (10/13) |
| 500 | 1 | 87 | Group II alveolate | Dinoflagellate | 335 | Group II alveolate | Dinoflagellate (139/139) |
| | 2 | 55 | *Arachnosphaera myriacantha* | Polycystine | 333 | *Arachnosphaera myriacantha* | Polycystine (47/50) |
| | 3 | 29 | *Varistrombidium kielum* | Ciliate | 273 | *Strombidium basimorphum* | Ciliate (12/12) |
| | | | | | 274 | *Novistrombidium* sp. | Ciliate (5/6) |
| | 4 | 18 | Group II alveolate | Dinoflagellate | 281 | Group II alveolate | Dinoflagellate (8/9) |
| | | | | | 337 | Group II alveolate | Dinoflagellate (27/28) |
| | 5 | 18 | *Apolemia* sp. | Cnidaria | 279 | *Apolemia* sp. | Cnidaria (16/17) |
| | 6 | 18 | *Larcopyle butschlii* | Polycystine | 181 | *Larcopyle butschlii* | Rhizaria (8/12) |
| | | | | | 270 | *Arachnosphaera myriacantha* | Rhizaria (9/10) |
| | 7 | 17 | Group II alveolate | Dinoflagellate | 337 | Group II alveolate | Dinoflagellate (27/28) |
| | | | | | 338 | Group II alveolate | Dinoflagellate (22/22) |
| | 8 | 16 | Group II alveolate | Dinoflagellate | 340 | Group II alveolate | Dinoflagellate (35/35) |
| | 9 | 15 | *Karlodinium* spp. | Dinoflagellate | 335 | Group II alveolate | Dinoflagellate (139/139) |
| | 10 | 14 | *Gyrodinium* spp. | Dinoflagellate | 239 | *Gyrodinium* spp. | Dinoflagellate (13/13) |

### 3.3. Comparison of in silico fragments and sequence-based OTUs

*In silico* digestion of the sequences in each of the ten most abundant sequence-based OTUs generally resulted in one dominant fragment size (Fig. 3). Additional fragment sizes were present in all but one OTU (OTU-500 m-1 resulted in a single fragment size), but the relative abundance of the additional fragment lengths was minor in most cases. There were four cases in which *in silico* digestion of the sequences in a sequence-based OTU resulted in two fragment sizes of relatively equal proportions (OTU-500 m-3, 4, 6 and 7), in which case both fragment sizes were reported (Table 2; also, significant contributions of secondary fragments were indicated in Fig. 2 by a '+').

The dominant fragment sizes that corresponded with the 10 most abundant sequence-based OTUs were also some of the most abundant fragment sizes following the *in silico* digestion of all the 5 m or 500 m sequences (Fig. 2A and B; fragment OTUs marked with '*' and '+'). For example, the most abundant fragment length resulting from *in silico* digestion of all the sequences from the 5 m sample (which constituted ~8% of the total *in silico* fragments in that dataset, see Fig. 2A) was generated largely by sequences from the most abundant sequence-based OTU (~93%). However, some fragments with high relative abundance were derived from *in silico* digestions of several sequence-based OTUs. For example, the *in silico* fragment size of 335 bp in the 500 m sample (OTU-5m-1) (Table 2) was generated by the digestion of sequences that clustered into 21 separate sequence-based OTUs (although all the sequences had best BLAST+ matches to dinoflagellate sequences).

Putative taxonomic identifications and phylogenetic groupings were assigned to the *in silico* fragment sizes (*i.e.* fragment-based OTUs, right side of Table 2) associated with the most abundant sequence-based OTUs (left side of Table 2) by determining the best BLAST+ matches of all the 5 m or 500 m sequences giving rise to that particular fragment size (see 'Fragment-based identification', right side of Table 2). In general, there was good correspondence between the taxonomy derived for the dominant fragment size and the taxonomy of the sequence-based OTU (left *vs.* right side of Table 2). However, some fragment sizes had minor contributions from sequences that yielded different phylogenetic affiliations (note parentheses, right side of Table 2). For example, the fragment size of 239 bp was generated from sequences that had highest similarities to dinoflagellates except 2 sequences that had high affinity to chlorophyte sequences. This situation was not the case for the sequence-based OTUs, which yielded the same higher-level phylogenetic affiliations for all sequences within an OTU.

### 3.4. Comparison of environmental T-RFLP fragments and in silico fragments

T-RFLP analysis of the 5 m and 500 m environmental samples resulted in 38 and 37 uniquely sized fragments using *HaeIII* restriction digests, respectively (Table 1, Fig. 4A and B). These results are consistent with an analysis of 237 T-RFLP profiles for samples collected from multiple depths over the course of ~10 years at the USC Microbial Observatory, which resulted in an average of ~35 fragments per sample (Kim et al., in review). A total of 61 unique and 14 shared *HaeIII* fragment sizes were obtained between the 5 m and 500 m samples (Fig. 4A and B). The combined *MnlI* T-RFLP analysis resulted in a total of 58 unique fragment sizes, 10 of which were shared between the two depths (data not shown). Community similarity between the two depths based on Bray–Curtis similarity analysis of the relative abundance of fragments in the T-RFLP datasets was ~40% with *HaeIII* digestion, compared to ~53% obtained using the *in silico HaeIII* fragments and 19% obtained using the sequence-based OTUs.

Approximately 79% (30/38) of the fragment sizes from the environmental T-RFLP analysis of the 5 m sample were found in the *in silico HaeIII* fragment database of the 5 m sequences (Fig. 5A and B). Many of the common *in silico* fragments were observed at high relative abundance in the T-RFLP analysis of the environmental sample,

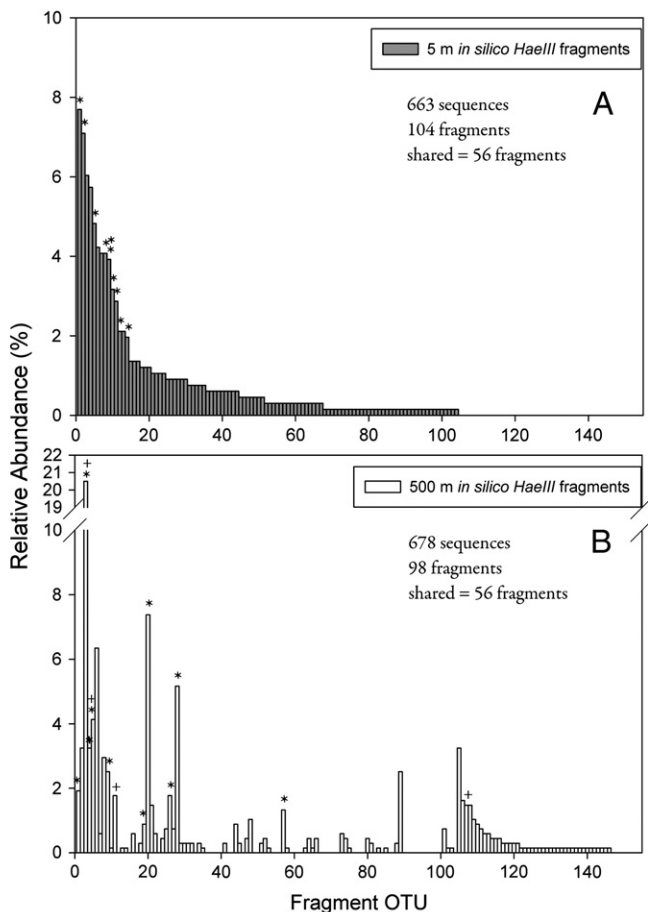


**Fig. 2.** Rank abundance curves depicting the relative abundance of unique fragment sizes (*i.e.* fragment OTUs) resulting from *HaeIII in silico* digestion of 18S rDNA sequences from two samples collected at the USC Microbial Observatory site. Rank abundance curves of fragment OTUs for the 5 m (A) and 500 m (B) samples were generated using the same sequences from Fig. 1. The x-axes are arranged from highest to lowest abundance of the 5 m fragments. Asterisks (*) indicate the fragments resulting from the *in silico* digestion of the ten most abundant sequence-based OTUs in each sample (also see Fig. 3). Three of the most abundant sequence-based OTUs yielded fragments of the same size (stacked asterisks). The *in silico* digestion of four sequence-based OTUs from the 500 m sample yielded two fragment lengths of relatively equal proportions (secondary fragments are indicated by '+'). There is a break in the y-axis for the 500 m rank abundance curve to facilitate comparison between the samples.

including 5 of the fragment sizes associated with the most abundant sequence-based OTUs (marked with '*' in Fig. 5B). The two most abundant 5 m *in silico HaeIII* fragments were not detected in the environmental T-RFLP results, however. There were also discrepancies between the relative abundance of some of the *in silico* fragments and T-RFs from the T-RFLP analysis.

### 3.5. T-RFLP analysis and quantitative PCR for characterizing the dynamics of Ostreococcus

The temporal dynamics of the prasinophyte *Ostreococcus* sp. in surface waters at the USC Microbial Observatory site was examined and compared using fragment analysis and quantitative PCR (qPCR). The relative abundance in T-RFLP patterns of a 259 bp fragment, which was identified as *Ostreococcus* using the T-RF database in this study, was compared to results from a previous study that developed and applied a qPCR approach for *Ostreococcus* (Countway and Caron, 2006). The relative abundance of the prasinophyte reflected by fragment analysis in samples collected from a depth of 5 m at approximately monthly intervals between Sept 2000 and Sept 2002 at the USC Microbial Observatory correlated well with abundance determined by qPCR (Fig. 6). Three major peaks in abundance (May 2001, December 2001, June 2002) were documented using both methods. Extending the T-RFLP results to the end of September 2003 revealed an additional peak in relative abundance in May of 2003 for the fragment size identified as *Ostreococcus*.

## 4. Discussion

### 4.1. T-RFLP analysis of environmental samples

Genetic approaches have yielded significant new insight into the immense diversity of microbial eukaryotes from a variety of geographic locations, depths and times (Massana and Pedros-Alio, 2008). There are few studies, however, that have characterized extensive spatial or temporal scales of variation within microbial eukaryote assemblages. This is, in part, due to the logistical challenges of collecting large numbers of samples over space or time, as well as the cost and time required for generating DNA sequence information and analyzing large datasets. Thus, despite significant advances in microbial and molecular eukaryote ecology, significant gaps in our understanding of spatial and temporal dynamics within assemblages of microbial eukaryotes still remain.

T-RFLP provides a relatively rapid and inexpensive method for the characterization of the dominant taxa within microbial eukaryote assemblages, as demonstrated by the comparison of results from fragment-based and sequence-based analyses of the same samples in the current study (Figs. 2A and B; 5A and B). Fragment-based approaches have been used to characterize natural assemblages of microbial eukaryotes including pico-eukaryotes from different oceanic regimes (Diez et al., 2001a), identify distinct sea-ice and water assemblages of protistan taxa from the Ross Sea in Antarctica (Gast et al., 2004), establish rapid changes in a protistan assemblage during a bottle incubation experiment (Countway et al., 2005), characterize ciliate communities in stream biofilms (Dopheide et al., 2008), document rapid shifts in dominant taxa in an estuary system (Vigil et al., 2009), and monitor the dynamics of potentially harmful algal bloom causing species (Joo et al., 2010). A primary advantage of the T-RFLP method in these studies has been the ability to analyze the large numbers of samples that are often required for ecological studies.

Analysis of DNA fragment data is also relatively straight-forward and is predicated on the assumption that different species will generate unique DNA fragment sizes. Although there are caveats associated with this assumption, the formation of OTUs from fragment data is less subjective than clustering sequence data into OTUs by employing one of a range of similarity values presently in use. Indeed, a number of similarity thresholds have been applied for grouping SSU rRNA gene sequences from environmental surveys of microbial eukaryotes into OTUs (Caron et al., 2009; Nebel et al., 2011). The difficulty in establishing an ecologically meaningful threshold for forming OTUs includes fundamental complexities associated with the species concepts that have been applied to protists.

We employed the conservative threshold of 95% derived by Caron et al. (2009) in this study. The availability of morphologically well-defined protistan species and associated sequences in GenBank in that study facilitated the analysis of 211 complete and partial 18S rRNA gene sequences to determine a sequence similarity threshold for delineating approximately species-level operational taxonomic units (OTUs) for microbial eukaryotes (Caron et al., 2009). The comparison of sequence similarities between strains of the same species as well as different species of the same genus resulted in an average similarity of 95% for delineating approximately species-level OTUs. The authors reported that although this value is generally lower than those employed by other studies (which have ranged between 97% and 99%) and most likely overlooks physiological variability in some OTUs, it provides a conservative estimate of species (*i.e.* OTU) richness for analyzing 18S rRNA gene sequence data.

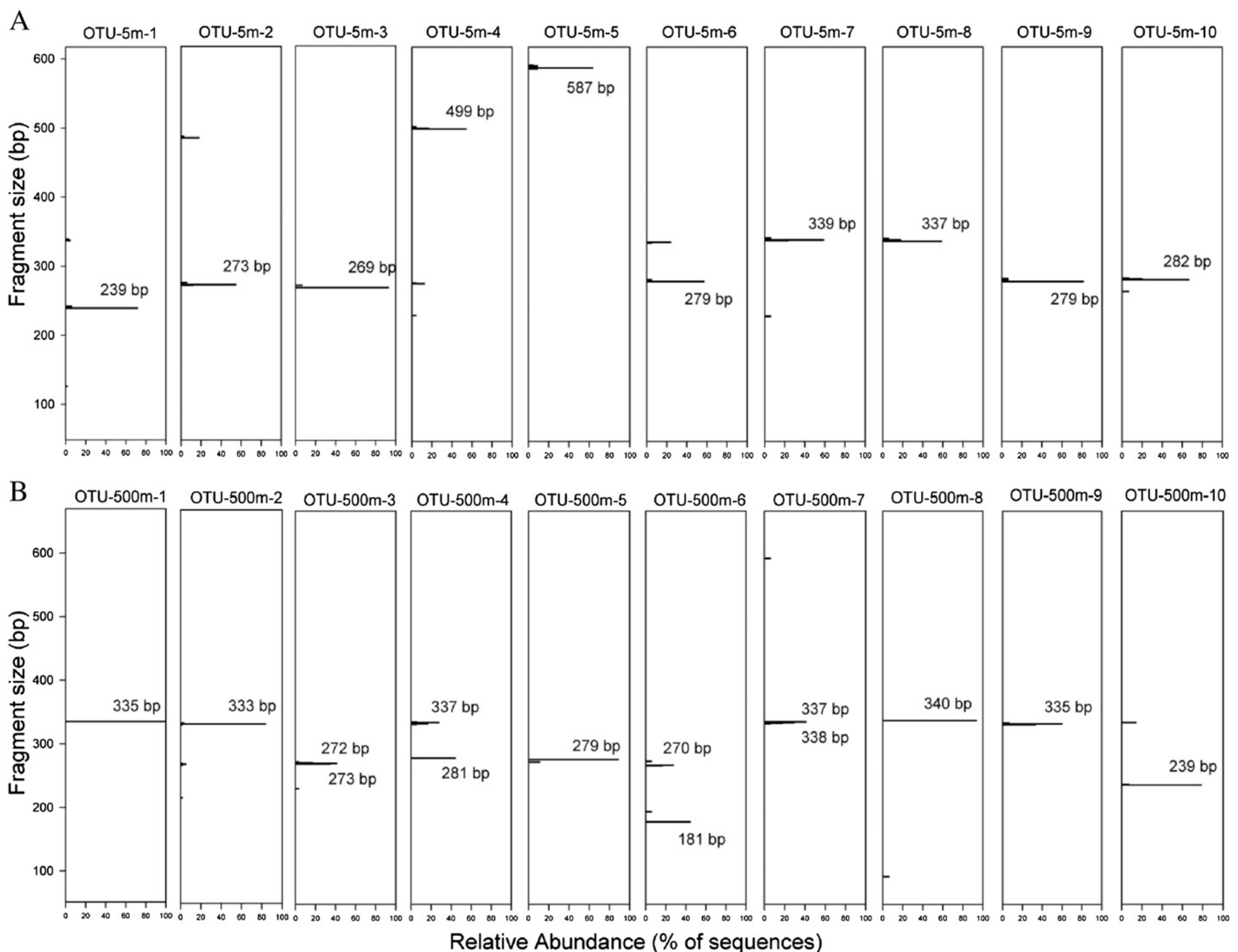


**Fig. 3.** Distributions of the fragment sizes resulting from *HaeIII in silico* digestions of the ten most abundant sequence-based OTUs in the 5 m (A) and 500 m (B) samples. The ten most abundant sequence-based OTUs comprised 39% and 42% of the total number of sequences in the 5 m and 500 m samples, respectively, and are depicted from most abundant on the left (OTU-5 m-1) to least abundant on the right (OTU-5 m-10). The dominant fragment lengths associated with each sequence-based OTU are labeled. Four of the sequence-based OTUs obtained from the 500 m sample yielded two fragments of approximately similar relative abundance.

Increasing the sequence similarity (%) used to cluster sequences into taxonomic units can substantially inflate the number of OTUs detected in a sample (Caron et al. 2009), which can arguably lead to a gross overestimation of species richness or diversity. Conversely, fragment-based approaches provide a limited perspective of species richness (only characterizing the dominant taxa), but may provide an assessment of diversity that is intermediate between sequencing and morphological methods (Dopheide et al., 2008).

### 4.2. Limitations of T-RFLP

Since the introduction of the T-RFLP method in microbial ecology for the assessment of 16S rRNA genes (Liu et al., 1997), limitations of the method as well as methodological improvements and assessments have been reported (Lukow et al., 2000; Egert and Friedrich, 2003; Lueders and Friedrich, 2003; Abdo et al., 2006; Pandey et al., 2007; Schutte et al., 2008; Zhang et al., 2008; Orcutt et al., 2009). One limitation of the T-RFLP method is an inability to characterize members of the 'rare biosphere.' This seemingly ubiquitous feature of microbial eukaryote assemblages (Fig. 1A–C) has been hypothesized to play an ecologically important role in community response and reassembly (Pedros-Alio, 2007; Caron and Countway, 2009). The functional roles and specific relationships of rare taxa within protistan assemblages are presently unknown, but it is possible that at least some rare taxa play critical functional roles (*e.g.*, keystone species). Such taxa that remain perpetually rare in a natural assemblage, but might still be ecologically significant, would be undetected by T-RFLP analysis. For this reason, fragment-based results are not appropriate for estimating total species richness (Dunbar et al., 2000). Fragment-based approaches can also lead to higher estimates of community similarity than sequenced-based approaches presumably due to the inability to differentiate some taxa and detect rare taxa (Table 1; also, compare Figs. 1B and C to 2A and B as well as 4A and B). On the positive side, changes in the dominant taxa of protistan assemblages can occur within hours to days in response to changes in environmental conditions (Countway et al., 2005; Kim et al., 2011), and these exchanges can be captured by fragment analysis.

### 4.3. Improving T-RFLP analysis by providing taxonomic context using sequence-based analyses

DNA fragment analyses are most informative for characterizing changes within microbial eukaryote assemblages when taxonomic context is provided for the fragments. The inability to differentiate

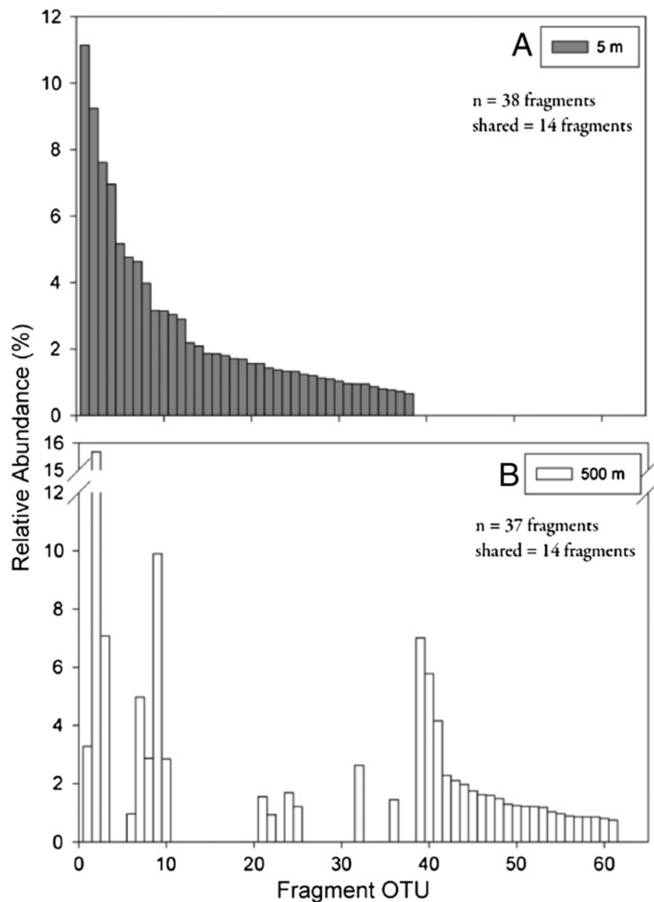


**Fig. 4.** Rank abundance curves of *HaeIII* T-RFLP results obtained from environmental samples collected at 5 m (A) and 500 m (B). The x-axes have been normalized to the relative abundance of fragments in the 5 m sample to highlight unique and shared fragment sizes between the samples. There is a break in the y-axis of the 500 m rank abundance curve to facilitate comparison between the samples.

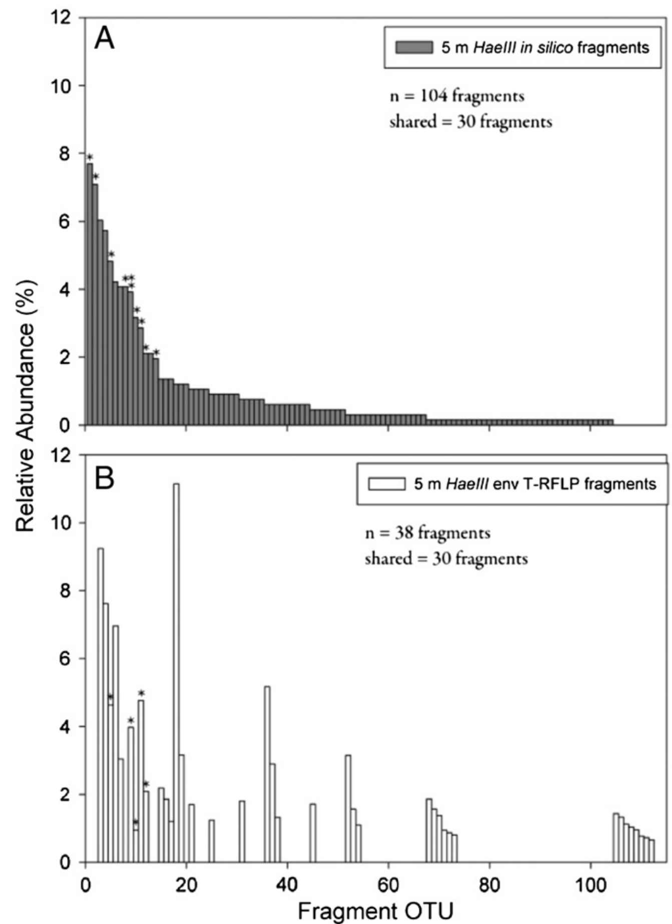


**Fig. 5.** Comparison of rank abundance curves of (A) a fragment dataset obtained from *HaeIII in silico* digestion of 663 partial 18S sequences from a 5 m sample (same data from Fig. 2A), and (B) environmental T-RFLP fragments generated by *HaeIII* digestion of DNA from the same sample (same data from Fig. 4A, rearranged for comparison to the *in silico* digestion in A). Both x-axes are ordered by relative abundance of *in silico* fragments. Asterisks mark the fragment sizes generated by the ten most abundant sequence-based OTUs (from Fig. 3A).

taxa, however, can complicate the assignment of taxonomic identifications to fragment sizes. Fragment-based studies that have utilized publicly available sequences from GenBank or SILVA to generate *in silico* T-RFLP databases or have selectively sequenced clones with unique RFLP signatures for identifying fragments (Diez et al., 2001a; Fernandez-Guerra et al., 2010) have noted this caveat. Here, we directly compared results from sequence-based and fragment-based analyses of the same samples to provide taxonomic context for the T-RFLP results.

Our combined sequence-based and fragment-based analyses of the same samples provided confident assignment of taxonomic identities to many of the fragments resulting from T-RFLP analysis (Fig. 5A and B). Furthermore, our highly comparable results of the qPCR and T-RFLP for characterizing the dynamics of *Ostreococcus* (Fig. 6) indicated that at least some of the fragments in the current study were identified accurately using the combined analysis approach, and that fragment data could be useful for investigating the spatiotemporal distributions of some protists. We observed a strong correlation between absolute 18S rRNA gene abundance (as measured by qPCR) of *Ostreococcus* and the relative abundance of the fragment identified as *Ostreococcus* by T-RFLP. We suspect that this tight relationship may have been in part due to the generally oligotrophic conditions of the USC Microbial Observatory station, where monthly changes in bacterial or protistan assemblages are generally not substantial (Fuhrman et al., 2006; Kim et al., in review). A comparison of results from sequence-based and fragment-based analyses also affirmed that fragments observed by T-RFLP generally represented the most abundant sequence-based OTUs (Figs. 3A and B; 5A and B).

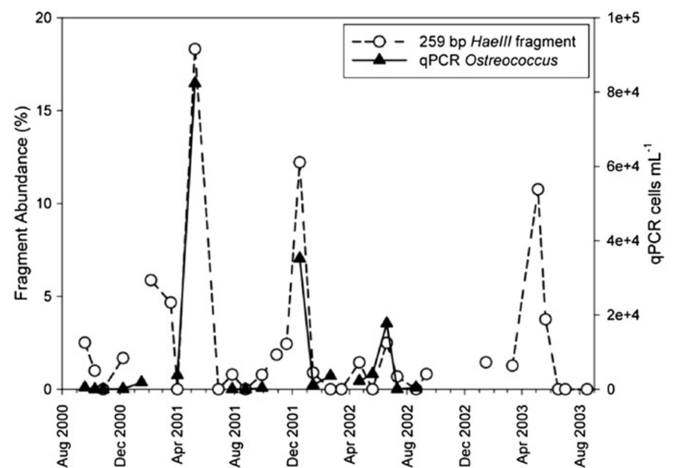


**Fig. 6.** Time-series of changes in the absolute abundance of *Ostreococcus* sp. (filled triangles; solid lines) as detected by quantitative PCR (Countway and Caron, 2006) in samples collected at approximately monthly intervals between September 2000 and September 2002 from the USC Microbial Observatory site compared to relative abundance of a 259 bp T-RFLP fragment (identified as *Ostreococcus* sp. from the database in this study) in the same samples (open circles; dotted line). The time-series for the fragment dataset was extended to September 2003.

## 5. Conclusions

T-RFLP can provide a relatively rapid and inexpensive snapshot of the dominant taxa within complex assemblages of microbial eukaryotes. The analysis of fragment data is relatively straight-forward compared to the computational power required to analyze sequences. These characteristics provide advantages for analyzing large numbers of samples that are often required for ecological studies of microbial eukaryote assemblages. The combined sequence-based and fragment-based analyses of the same samples in this study resulted in the confident assignment of taxonomic identification to many of the fragments observed by T-RFLP and affirmed that fragments generally represented the dominant sequence-based OTUs. Limitations of T-RFLP exist and were demonstrated as a part of the current study, most significantly the inability of T-RFLP to differentiate some taxa and detect rare taxa. The combined analysis of sequence-based and fragment-based analyses of the same samples in the current study provided taxonomic context for the analysis of T-RFLP data from approximately monthly samples collected from multiple depths at the USC Microbial Observatory between 2000 and 2010 (Kim et al., in review).

## Acknowledgments

We would like to thank Troy Gunderson, Michael Newman, as well as the captain and crew of the R/V SeaWatch from the Southern California Marine Institute for assistance with the collection of samples. This work was funded by NSF grants MCB-0703159 and OCE-1136818, and a grant from the Gordon and Betty Moore Foundation.

## References

Abdo, Z., Schuette, U.M.E., Bent, S.J., Williams, C.J., Forney, L.J., Joyce, P., 2006. Statistical methods for characterizing diversity of microbial communities by analysis of terminal restriction fragment length polymorphisms of 16S rRNA genes. Environ. Microbiol. 8, 929–938.

Adl, S.M., Simpson, A.G.B., Farmer, M.A., Andersen, R.A., Anderson, O.R., Barta, J.R., et al., 2005. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. J. Eukaryot. Microbiol. 52, 399–451.

Amaral-Zettler, L., Gomez, F., Zettler, E., Keenan, B.G., Amils, R., Sogin, M., 2002. Eukaryotic diversity in Spain's River of Fire. Nature 417, 137.

Ashelford, K.E., Chuzhanova, N.A., Fry, J.C., Jones, A.J., Weightman, A.J., 2005. At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. Appl. Environ. Microbiol. 71, 7724–7736.

Balzano, S., Marie, D., Gourvil, P., Vaulot, D., 2012. Composition of the summer photosynthetic pico and nanoplankton communities in the Beaufort Sea assessed by T-RFLP and sequences of the 18S rRNA gene from flow cytometry sorted samples. ISME J.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Sayers, E.W., 2010. GenBank. Nucleic Acids Res. 38, D46–D51.

Caron, D.A., 2009. New accomplishments and approaches for assessing protistan diversity and ecology in natural ecosystems. Bioscience 59, 287–299.

Caron, D.A., Countway, P.D., 2009. Hypotheses on the role of the protistan rare biosphere in a changing world. Aquat. Microb. Ecol. 57, 227–238.

Caron, D.A., Countway, P.D., Savai, P., Gast, R.J., Schnetzer, A., Moorthi, S.D., et al., 2009. Defining DNA-based operational taxonomic units for microbial-eukaryote ecology. Appl. Environ. Microbiol. 75, 5797–5808.

Caron, D.A., Countway, P.D., Jones, A.C., Kim, D.Y., Schnetzer, A., 2012. Marine protistan diversity. Ann. Rev. Mar. Sci. 4, 467–493.

Countway, P.D., Caron, D.A., 2006. Abundance and distribution of *Ostreococcus* sp in the San Pedro Channel, California, as revealed by quantitative PCR. Appl. Environ. Microbiol. 72, 2496–2506.

Countway, P.D., Gast, R.J., Savai, P., Caron, D.A., 2005. Protistan diversity estimates based on 18S rDNA from seawater incubations in the western North Atlantic. J. Eukaryot. Microbiol. 52, 95–106.

Diez, B., Pedros-Alio, C., Marsh, T.L., Massana, R., 2001a. Application of denaturing gradient gel electrophoresis (DGGE) to study the diversity of marine picoeukaryotic assemblages and comparison of DGGE with other molecular techniques. Appl. Environ. Microbiol. 67, 2942–2951.

Diez, B., Pedros-Alio, C., Massana, R., 2001b. Study of genetic diversity of eukaryotic picoplankton in different oceanic regions by small-subunit rRNA gene cloning and sequencing. Appl. Environ. Microbiol. 67, 2932–2941.

Dopheide, A., Lear, G., Stott, R., Lewis, G., 2008. Relative diversity and community structure of ciliates in stream biofilms according to molecular and microscopy methods. Appl. Environ. Microbiol. 75, 5261–5272.

Drummond, A.J., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Duran, C., et al., 2011. Geneious v5.4. available from http://www.geneious.com/.

Dunbar, J., Ticknor, L.O., Kuske, C.R., 2000. Assessment of microbial diversity in four southwestern United States soils by 16S rRNA gene terminal restriction fragment analysis. Appl. Environ. Microbiol. 66, 2943–2950.

Edgcomb, V.P., Kysela, D.T., Teske, A., Gomez, A.D., Sogin, M.L., 2002. Benthic eukaryotic diversity in the Guaymas Basin hydrothermal vent environment. Proc. Natl. Acad. Sci. U. S. A. 99, 7658–7662.

Egert, M., Friedrich, M.W., 2003. Formation of pseudo-terminal restriction fragments, a PCR-related bias affecting terminal restriction fragment length polymorphism analysis of microbial community structure. Appl. Environ. Microbiol. 69, 2555–2562.

Elwood, H.J., Olsen, G.J., Sogin, M.L., 1985. The small-subunit ribosomal RNA gene sequences from the hypotrichous ciliates Oxytricha nova and Stylonychia pustulata. Mol. Biol. Evol. 2, 399–410.

Euringer, K., Lueders, T., 2008. An optimised PCR/T-RFLP fingerprinting approach for the investigation of protistan communities in groundwater environments. J. Microbiol. Methods 75, 262–268.

Fernandez-Guerra, A., Buchan, A., Mou, X.Z., Casamayor, E.O., Gonzalez, J.M., 2010. T-RFPred: a nucleotide sequence size prediction tool for microbial community description based on terminal-restriction fragment length polymorphism chromatograms. BMC Microbiol. 10.

Fisher, M.M., Triplett, E.W., 1999. Automated approach for ribosomal intergenic spacer analysis of microbial diversity and its application to freshwater bacterial communities. Appl. Environ. Microbiol. 65, 4630–4636.

Fuhrman, J.A., Hewson, I., Schwalbach, M.S., Steele, J.A., Brown, M.V., Naeem, S., 2006. Annually reoccurring bacterial communities are predictable from ocean conditions. Proc. Natl. Acad. Sci. U. S. A. 103, 13104–13109.

Gast, R.J., Dennett, M.R., Caron, D.A., 2004. Characterization of protistan assemblages in the Ross Sea, Antarctica, by denaturing gradient gel electrophoresis. Appl. Environ. Microbiol. 70, 2028–2037.

Gast, R.J., Moran, D.M., Beaudoin, D.J., Blythe, J.N., Dennett, M.R., Caron, D.A., 2006. Abundance of a novel dinoflagellate phylotype in the Ross Sea, Antarctica. J. Phycol. 42, 233–242.

Joo, S., Lee, S.R., Park, S., 2010. Monitoring of phytoplankton community structure using terminal restriction fragment length polymorphism (T-RFLP). J. Microbiol. Methods 81, 61–68.

Kaplan, C.W., Kitts, C.L., 2004. Bacterial succession in a petroleum land treatment unit. Appl. Environ. Microbiol. 70, 1777–1786.

Kim, D.Y., Countway, P.D., Gast, R.J., Caron, D.A., 2011. Rapid shifts in the structure and composition of a protistan assemblage during bottle incubations affect estimates of total protistan species richness. Microb. Ecol. 62, 383–398.

Kim, D.Y., Countway, P.D., Yamashito, W., Tung, C., Caron, D.A., in review. Monthly, seasonal and interannual variability of microbial eukaryote assemblages within and below the euphotic zone in the eastern North Pacific.

Liu, W.T., Marsh, T.L., Cheng, H., Forney, L.J., 1997. Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. Appl. Environ. Microbiol. 63, 4516–4522.

Lopez-Garcia, P., Rodriguez-Valera, F., Pedros-Alio, C., Moreira, D., 2001. Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. Nature 409, 603–607.

Lopez-Garcia, P., Philippe, H., Gail, F., Moreira, D., 2003. Autochthonous eukaryotic diversity in hydrothermal sediment and experimental microcolonizers at the Mid-Atlantic Ridge. Proc. Natl. Acad. Sci. U. S. A. 100, 697–702.

Lueders, T., Friedrich, M.W., 2003. Evaluation of PCR amplification bias by terminal restriction fragment length polymorphism analysis of small-subunit rRNA and mcrA genes by using defined template mixtures of methanogenic pure cultures and soil DNA extracts. Appl. Environ. Microbiol. 69, 320–326.

Lukow, T., Dunfield, P.F., Liesack, W., 2000. Use of the T-RFLP technique to assess spatial and temporal changes in the bacterial community structure within an agricultural soil planted with transgenic and non-transgenic potato plants. FEMS Microbiol. Ecol. 32, 241–247.

Marande, W., Lopez-Garcia, P., Moreira, D., 2009. Eukaryotic diversity and phylogeny using small- and large-subunit ribosomal RNA genes from environmental samples. Environ. Microbiol. 11, 3179–3188.

Massana, R., Pedros-Alio, C., 2008. Unveiling new microbial eukaryotes in the surface ocean. Curr. Opin. Microbiol. 11, 213–218.

Massana, R., Castresana, J., Balague, V., Guillou, L., Romari, K., Groisillier, A., et al., 2004. Phylogenetic and ecological analysis of novel marine stramenopiles. Appl. Environ. Microbiol. 70, 3528–3534.

Medlin, L.K., Elwood, H.J., Stickel, S., Sogin, M.L., 1988. The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. Genetica 71, 491–499.

Moreira, D., Lopez-Garcia, P., 2002. The molecular ecology of microbial eukaryotes unveils a hidden world. Trends Microbiol. 10, 31–38.

Nebel, M., Pfabel, C., Stock, A., Dunthorn, M., Stoeck, T., 2011. Delimiting operational taxonomic units for assessing ciliate environmental diversity using small-subunit rRNA gene sequences. Environ. Microbiol. Rep. 3, 154–158.

Orcutt, B., Bailey, B., Staudigel, H., Tebo, B.M., Edwards, K.J., 2009. An interlaboratory comparison of 16S rRNA gene-based terminal restriction fragment length polymorphism and sequencing methods for assessing microbial diversity of seafloor basalts. Environ. Microbiol. 11, 1728–1735.

Osborn, A.M., Moore, E.R.B., Timmis, K.N., 2000. An evaluation of terminal-restriction fragment length polymorphism (T-RFLP) analysis for the study of microbial community structure and dynamics. Environ. Microbiol. 2, 39–50.

Pandey, J., Ganesan, K., Jain, R.K., 2007. Variations in T-RFLP profiles with differing chemistries of fluorescent dyes used for labeling the PCR primers. J. Microbiol. Methods 68, 633–638.

Pedros-Alio, C., 2007. Dipping into the rare biosphere. Science 315, 192–193.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W.G., Peplies, J., et al., 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. 35, 7188–7196.

Romari, K., Vaulot, D., 2004. Composition and temporal variability of picoeukaryote communities at a coastal site of the English Channel from 18S rDNA sequences. Limnol. Oceanogr. 49, 784–798.
Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., et al., 2009. Introducing MOTHUR: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. 75, 7537–7541.
Schutte, U.M.E., Abdo, Z., Bent, S.J., Shyu, C., Williams, C.J., Pierson, J.D., et al., 2008. Advances in the use of terminal restriction fragment length polymorphism (T-RFLP) analysis of 16S rRNA genes to characterize microbial communities. Appl. Microbiol. Biotechnol. 80, 365–380.
Sherr, B.F., Sherr, E.B., Caron, D.A., Vaulot, D., Worden, A.Z., 2007. Oceanic protists. Oceanogr. 20, 130–134.
Steele, J.A., Countway, P.D., Xia, L., Vigil, P.D., Beman, J.M., Kim, D.Y., et al., 2011. Marine bacterial, archaeal and protistan association networks reveal ecological linkages. ISME J. 5, 1414–1425.
Stoeck, T., Epstein, S., 2003. Novel eukaryotic lineages inferred from small-subunit rRNA analyses of oxygen-depleted marine environments. Appl. Environ. Microbiol. 69, 2657–2663.
Tekle, Y.I., Parfrey, L.W., Katz, L.A., 2009. Molecular data are transforming hypotheses on the origin and diversification of eukaryotes. Bioscience 59, 471–481.
Vigil, P., Countway, P.D., Rose, J., Lonsdale, D.J., Gobler, C.J., Caron, D.A., 2009. Rapid shifts in dominant taxa among microbial eukaryotes in estuarine ecosystems. Aquat. Microb. Ecol. 54, 83–100.
Worden, A.Z., 2006. Picoeukaryote diversity in coastal waters of the Pacific Ocean. Aquat. Microb. Ecol. 43, 165–175.
Zhang, R., Thiyagarajan, V., Qian, P.Y., 2008. Evaluation of terminal-restriction fragment length polymorphism analysis in contrasting marine environments. FEMS Microbiol. Ecol. 65, 169–178.