J. Eukaryot. Microbiol., 52(2), 2005 pp. 95–106

Protistan Diversity Estimates Based on 18S rDNA from Seawater Incubations in the Western North Atlantic[1]

PETER D. COUNTWAY,[a] REBECCA J. GAST,[b] PRATIK SAVAI[a] and DAVID A. CARON[a]

[a]*Department of Biological Sciences, University of Southern California, Los Angeles, California 90089-0371, USA, and*
[b]*Biology Department, Woods Hole Oceanographic Institution, Woods Hole, Masssachusetts 02543, USA*

ABSTRACT. Cloning/sequencing and fragment analysis of ribosomal RNA genes (rDNA) are becoming increasingly common methods for the identification of microbial taxa. Sequences of these genes provide many additional taxonomic characters for species that otherwise have few distinctive morphological features, or that require involved microscopy or laboratory culture and testing. These same approaches are now being applied with great success in ecological studies of natural communities of microorganisms. Extensive information on the composition of natural microbial assemblages is being amassed at a rapid pace through genetic analyses of environmental samples and comparison of the resulting genetic information with well-established (and rapidly growing) public databases. We examined microbial eukaryote diversity in a natural seawater sample from the coastal western North Atlantic Ocean using two molecular biological approaches: the cloning and sequencing of rRNA genes and by fragment analysis of these genes using the terminal restriction fragment length polymorphism (T-RFLP) method. A simple experiment was carried out to examine changes in the overall eukaryote (largely protistan) diversity and species composition (phylotype diversity) of a natural microbial assemblage when a seawater sample is placed in a container and incubated at ambient light and temperature for 72 h. Containment of the natural seawater sample resulted in relatively minor changes in the overall eukaryote diversity (species richness) obtained by either molecular method at three time points (time-zero, time-24 h, time-72 h). However, substantial changes in the dominance of particular eukaryote phylotypes took place between the three sampling times. Only 18% of the total number of phylotypes observed in the study were observed at all three time points, while 65% (108 of 165) phylotypes were observed only at a single time point (54 unique phylotypes initially, 37 more unique phylotypes at 24 h, and 17 more at 72 h). The results of this study indicate that a high diversity of protistan taxa existed in the original seawater sample at very low abundance, and thus were not observed in the initial characterization of community structure. Containment resulted in significant shifts in the dominance of these taxa, enabling the presence of previously unobserved phylotypes to be documented after 24 or 72 h of incubation.

Key Words. Atlantic Ocean, coastal, diversity, DNA sequence, fragment analysis, incubation, molecular, protist, 18S rRNA, T-RFLP.

THE number and scope of microbial biodiversity studies have increased dramatically during the past 15 years. This effort was stimulated at that time by the recognition that the diversity of many microbial communities had not been (and still is not) adequately characterized by species present in extant culture collections (Delong 1992; Fuhrman, McCallum, and Davis 1992; Giovannoni et al. 1990). Since then, culture-independent studies using molecular approaches have resulted in a large-scale re-evaluation of microbial biodiversity in natural ecosystems across all domains of life (Archaea, Bacteria, and Eukarya) (Pace 1997; Woese, Kandler, and Wheelis 1990). This work has dramatically changed our understanding of microbial diversity, and has even resulted in recent attempts to characterize the genomic diversity of microbes from mixed environmental assemblages in oceanic ecosystems across the globe (Shreeve 2004; Venter et al. 2004). These latter studies may be able to provide insight into the overall breadth of the biogeochemical potential of microbial activities in nature as revealed by whole genome analyses. Additionally, these studies contribute to an ever expanding database of environmental DNA sequences (Benson et al. 2004) that facilitate further studies of microbial diversity and activity.

A major justification for documenting microbial diversity has been a desire to answer basic ecological questions. How many and what types of taxa are present in particular environments (Patterson 1999)? How is this diversity important to ecosystem function, stability, and resilience (McGrady-Steed, Harris, and Morin 1997; Naeem and Li 1997)? Microbial biomass measurements and biogeochemical rate measurements developed and applied in aquatic ecosystems during the 1970s and 1980s indicated an abundant and active microbial community. Microscopy of microbiota from diverse environments, and isotopic tracer-based studies such as those used for measuring heterotrophic bacterial productivity and phytoplankton primary productivity indicated that much of the living biomass and most of the biogeochemical processes taking place in these ecosystems were comprised of, and conducted by, microorganisms. However, these data were generated using methods that did not provide much detail regarding the species responsible, as most microbial taxa were lumped into large heterogeneous groupings by these approaches (Azam et al. 1983; Sieburth, Smetacek, and Lenz 1978).

Based on information obtained using these ''classical'' methods, and complemented with modern molecular biological approaches, it is now well established that Archaea, Bacteria, and microbial Eukarya play fundamental ecological roles in virtually all aquatic systems. Culture-independent molecular techniques, in particular, are being increasingly adopted as a standard means for describing community structure within these groups. This is especially true for studies of archaeal and bacterial assemblages, whose members lack distinct morphological characteristics and for which molecular identification is the only viable option for creating taxonomic lists.

On the other hand, studies of protistan genetic diversity have lagged behind bacterial and archaeal studies in part due to a well-established morphology-based taxonomy (many protistan taxa are relatively large and possess sufficient morphological characters to identify species by light microscopy). This tradition of describing protists based on morphological characteristics has slowed the recognition of the usefulness of molecular-based approaches for studying the diversity and ecology of natural assemblages of protists. However, as the use of molecular methods for studying the diversity of protistan assemblages gains acceptance, the gap between protistan and bacterial/archaeal research increasingly diminishes. For example, recent applications of molecular approaches to natural protistan assemblages have uncovered a large and previously undescribed diversity of protists (Dawson and Pace 2002; Diez, Pedros-Alio, and Massana 2001; Edgcomb et al. 2002a; Fawley, Fawley, and Buchheim 2004; Guillou et al. 2004; López-García et al. 2001, 2003; Massana et al. 2002, 2004; Moon-van der Staay, De Wachter, and Vaulot 2001; Romari and Vaulot 2004; Stoeck, Taylor, and Epstein 2003). Indeed, morphology-based studies of protistan diversity conducted over the past two centuries have revealed tremendous protistan diversity, but this information has been acquired piecemeal. The

[1]Paper presented June 2, 2004, as part of the Symposium, ''Advances in the Molecular Ecology of Protists'' during the 56th Annual Meeting of the Society of Protozoologists, June 2–6, Bryant College, Smithfield, RI.

Corresponding Author: P. D. Countway—Telephone number: 213-821-2123; FAX number: 213-740-8123; E-mail: countway@usc.edu

analysis of entire protistan assemblages for more than a few samples is prohibitive due to the diverse array of methods for sampling, preserving, and identifying protistan taxa (Caron, Countway, and Brown 2004). Studies using culture-independent molecular techniques have begun to address these difficulties, and in particular have revealed high levels of protistan diversity from a variety of ecosystems (Dawson and Pace 2002; Diez, Pedros-Alio, and Massana 2001; Edgcomb et al. 2002a; Fawley, Fawley, and Buchheim 2004; Gast, Dennett, and Caron 2004; Massana et al. 2002; Moon-van der Staay, De Wachter, and Vaulot 2001; Stoeck and Epstein 2003) and in some cases have provided supporting evidence for revising morphology-based taxonomic classifications (Amaral Zettler and Caron 2000; Edgcomb et al. 2002b; Leipe et al. 1994; Marin et al. 2003).

Similarly, molecular techniques have dramatically increased the speed, accuracy, and sensitivity with which some protistan taxa can be enumerated from natural samples (Bowers et al. 2000; Coyne et al. 2001; Galluzzi et al. 2004; Popels et al. 2003). Unlike microscopy-based analyses, which normally assess cell abundance and diversity (typically just broad taxonomic groups) on relatively small-volume water samples (usually less than 1 liter of sample collected on a filter), PCR-based assays are extremely sensitive. DNA can be extracted from large water samples (several to 10s of liters), allowing the detection of specific taxa at very low abundance, which would likely be missed by microscopy (Coyne and Cary 2005; Edvardsen et al. 2003; Marin et al. 2001; Rublee et al. 2005). Culture-independent molecular approaches thereby allow for a more thorough characterization of total protistan diversity in a relatively short amount of time. In addition, PCR-based techniques such as real-time quantitative PCR yield accurate estimates of target DNA over several orders of magnitude, and thus provide approaches to characterize specific taxa over tremendous ranges of abundance in natural samples.

This study was designed to examine microbial eukaryote assemblage diversity (largely protistan), focusing on small subunit (18S) ribosomal RNA genes. Two culture-independent PCR-based methods including terminal restriction fragment length polymorphism (T-RFLP) (Avaniss-Aghajani et al. 1996; Liu et al. 1997) and cloning/sequencing (Giovannoni et al. 1990; Moon-van der Staay, De Wachter, and Vaulot 2001; Rappe, Kemp, and Giovannoni 1995; Schmidt, Delong, and Pace 1991) were used to assess protistan diversity in a coastal ecosystem and to examine the response of the community to containment during 72-h incubations. Overall, both approaches yielded comparable results with respect to the general response of the community to containment. Eukaryote diversity (total phylotype richness and evenness) assessed by either method remained relatively high during the 3-day incubation. However, the species (phylotype) composition of the assemblage at each time point changed substantially over the course of the incubation. Cloning/sequencing provided greater resolution of phylotype diversity, but the labor-intensive nature of this approach hampers its general applicability to ecological studies. Conversely, T-RFLP provided a rapid assessment of dominant phylotypes. The latter method constitutes a useful approach for comparative studies designed to examine the factors controlling assemblage composition and diversity.

MATERIALS AND METHODS

Experimental design. The goal of this experiment was to assay the diversity of a natural protistan assemblage using two molecular biological approaches, and to compare how this assemblage changed during incubation in a bottle. This comparison consisted of baseline (time-zero) measurements of protistan diversity, and additional analyses of diversity following containment of the assemblage for 24 h (time-24 h) and 72 h (time-72 h). The rationale for this experimental design stemmed from the common use of bottle incubations for making process-oriented measurements of microbial activity. Bottle incubations of 1–3 days are a standard component of many ecological studies of aquatic ecosystems; yet, very little information exists on the potential for shifts in community composition (and thus community activity) during such incubations. Turnover times of microbial taxa are often less than the length of incubation experiments, and thus substantial changes in community diversity and composition might be expected within the course of these experiments. Documenting the extent of these changes is beyond the scope of most ecological studies using traditional, extant methodology. This study was designed to test the efficacy of two approaches for characterizing microbial (protistan) diversity and changes in community composition that might take place during bottle incubations.

Sample collection and experimental setup. Protistan community structure was sampled at a station off the coast of North Carolina along the U.S. continental shelf (36°21′N, 75°14′W) during a cruise in August of 2000 aboard the R/V Endeavor (University of Rhode Island). Seawater samples were collected using Niskin bottles (General Oceanics Inc., Miami, FL) from a depth equivalent to 30% of incident solar irradiance within the surface mixed layer. Seawater was transferred directly from Niskin bottles to acid-washed (5% HCl) polycarbonate carboys via gravity filtration through an inline filter apparatus, fitted with a 200-μm Nitex mesh (Sefar America, Monterey Park, CA). The pre-filtration was designed to reduce the contribution of metazoa to the seawater incubations and concomitant DNA analyses. The $<$200-μm seawater filtrate was partitioned into eight 4-liter polycarbonate bottles for subsequent incubation in flow-through incubators on the deck of the research vessel. Incubation chambers were shaded with several layers of neutral density screening to reduce incoming solar irradiance to levels similar to those measured at the time and depth of collection. Incubation bottles were sacrificially sampled immediately after partitioning water at time-zero (two bottles), and after 24 h (three bottles) and 72 h (three bottles) of incubation.

DNA collection and extraction. Samples collected at each of the three time points were filtered through 47-mm GF/F filters (Whatman International Ltd., Florham Park, NJ), which were loosely rolled, placed into 4-ml Cryo-vials (Nalge Nunc, Rochester, NY), and frozen until extraction. DNA samples were processed aboard ship by adding 1 ml of hot ($\sim$ 70 °C) lysis buffer (100 mM Tris (pH 8), 40 mM EDTA (pH 8), 100 mM NaCl, 1% SDS) and 200 μl of 0.5-mm zircon beads (for bead-beating). Samples were bead-beat for 30 s on a vortexer set to highest speed and then heated in a 70 °C water bath for 5 min, repeating this procedure for a total of three times. Lysates were transferred from the Cryo-vials and separated from filter debris by carefully pipetting the solution into a 2-ml micro-centrifuge tube. Sample lysates were adjusted to a final concentration of 0.7 M NaCl and 1% CTAB (Hexadecyltrimethyl-Ammonium Bromide, Sigma) and incubated for 10 min at 70 °C to fully dissolve any precipitate in the solution. Samples were extracted with chloroform and nucleic acids were precipitated from the aqueous layer with isopropanol (Gast, Dennett, and Caron 2004). Nucleic acids were pelleted by centrifugation in a micro-centrifuge at top speed ($>$20,000 *g*) for 15 min. All liquid was decanted after centrifugation and the pellet was air-dried. Pellets were resuspended in sterile water and stored frozen at $-$ 20 °C until thawed for use in a confirmatory PCR analysis aboard ship. This procedure ensured our ability to generate full-length PCR amplicons from all samples with the eukaryotic primer set listed in the PCR section below. PCR products were electrophoresed on 1.2% E-Gels (Invitrogen, Carlsbad, CA) containing ethidium bromide and visualized on a UV transilluminator.

PCR for cloning and sequencing. Full-length eukaryotic small subunit ribosomal RNA genes were amplified from environmental genomic DNA extracts using universal eukaryote primers, Euk-A (5′-AACCTGGTTGATCCTGCCAGT-3′) and Euk-B (5′-GATCCTTCTGCAGGTTCACCTAC-3′), to generate products for TA cloning and subsequent sequencing (Medlin et al. 1988). PCR reagents were mixed at the following final concentrations: 0.5 μM of each primer, 1× buffer B (Promega, Madison, WI), 2.5 mM $MgCl_2$ (Promega), 250 μM dNTPs (Promega), 300 ng/μl BSA (Sigma A-7030 (Kirchman et al. 2001)), 2.5 U of Taq in buffer B (Promega), and 1–2 μ of DNA template. Thermal cycling was carried out on either a Bio-Rad (Hercules, CA) iCycler or MyCycler using the following thermal protocol: one cycle at 95 °C for 2 min, 35 cycles at 95 °C for 30 s, 50 °C for 30 s, 72 °C for 2.5 min, a final extension at 72 °C for 7 min, and a hold at 4 °C. Three to four replicate PCR samples were run for each sample to ensure production of an adequate amount of product for cloning.

Cloning and sequencing. PCR amplicons were separated on 1.2% SeaKem LE agarose gels (Cambrex, Rockland, ME) and products of the appropriate size were excised from the gels using sterile razor blades. DNA was extracted from the gel slices for subsequent ligation and cloning with a gel extraction kit (Zymo Research Labs Inc., Orange, CA). Purified DNA was eluted from the Zymo columns with sterile water and quantified by PicoGreen fluorescence (Molecular Probes, Eugene, OR) on a VersaFluor fluorometer (Bio-Rad). Ligation reactions were set up using the pGEM-T Easy Vector kit (Promega) with 30 ng of PCR product (insert to vector ratio of 1:1). Ligation reactions were performed overnight at 4 °C to maximize the efficiency of transformations. Ligation reactions were cleaned with StrataClean resin (Stratagene, La Jolla, CA) prior to transformation. Approximately 5 ng of ligation product was mixed with 40 μl of Electro10Blue electrocompetent cells (Stratagene) and electroporated on a Gene Pulser Xcell (Bio-Rad) using the following exponential protocol: 1,700 V, 0.10-cm gap-width cuvette, with resistance set to 600 Ω and capacitance set to 25 μF. Immediately after delivering the shock, 960 μl of sterile SOC media was added to the cuvette before transferring the entire cell solution to a sterile 15-ml BD Falcon tube (BD Biosciences, San Jose, CA) for outgrowth at 37 °C for 90 min and shaking at 250 rpm. Samples of electroporated cells (50–200 μl) were plated onto S-Gal (Sigma) color-selective agar plates containing ampicillin (100 μg/ml) for overnight growth at 37 °C.

Distinct white colonies were picked the following day with sterile toothpicks and grown for 18–24 h in deep-well culture blocks containing 1.25 ml of TB medium and ampicillin (100 μg/ml). Glycerol stocks of all clones were prepared and archived at − 80 °C prior to spin-down and collection of bacterial pellets. Plasmid DNA was extracted from bacterial pellets with the Promega Wizard SV96 kit following the manufacturer's protocols. DNA for sequencing was eluted from the Promega plates with 100 μl of sterile water. DNA sequencing was carried out on a Beckman-Coulter (Fullerton, CA) CEQ8000 automated DNA sequencer. Sequencing reactions were conducted with Euk-570F (5′-GTAATTCCAGCTCCAATAGC-3′, (Weekers et al. 1994)) mixed with the Beckman-Coulter Dye Terminator Cycle Sequencing quick-start solution using one-quarter of the recommended DTCS reagent. Partial sequences obtained in this study, generally with lengths of 400–700 bp, have been deposited in GenBank (Accessions AY937465–AY938434).

Phylotype assignment and diversity estimation. DNA sequences were trimmed for quality assurance from both the 5′ and 3′ ends using the Beckman-Coulter CEQ8000 Sequence Analysis Software set at the ''Medium'' trim selection. Chromatograms were visually inspected at selected regions of conserved DNA sequence using Chromas (Technelysium) to ensure the accuracy of base-calling and to remove low-quality sequence reads. This procedure yielded a total of 970 sequences distributed across the three experimental time points as follows: time-zero (493 sequences), time-24 h (265 sequences), and time-72 h (212 sequences). Sequences were exported from Chromas as FASTA files and comparative sequence analysis was conducted by BLAST (Altschul et al. 1997) against both the NCBI (Benson et al. 2004) and ARB databases (Ludwig et al. 2004) to obtain taxonomic identity of the clones (to the degree possible). Pairwise comparisons of all sequences among the 970 sequences of our clone library were also conducted in order to establish operational taxonomic units within the library (OTUs established at 95% similarity).

Rarefaction curves relating sequencing effort to numbers of observed phylotypes (S_{obs} Mao Tau) detected within the eukaryote assemblage at each time point were calculated using EstimateS, version 7 (Colwell 2004). The overall diversity of the assemblage was estimated by pooling information from all three time points into one data set containing all 970 clone sequences. The distribution of phylotypes within this large data set provided an estimate of total phylotype diversity (e.g. the sampled plus non-sampled phylotypes) of the microbial eukaryote assemblage. Specifically, the non-parametric richness estimator, Chao1 (Chao 1984, 1987) was calculated with EstimateS using 100 randomizations, sampling without replacement.

PCR for T-RFLP. PCR products for analysis by T-RFLP were amplified using the primers D4-Euk-A (5′-D4-AA CCTGGTTGATCCTGCCAGT-3′) and Euk-570R (5′-GCTATT GGAGCTGGAATTAC-3′). The ''D4'' at the 5′ end of primer Euk-A indicates the fluorescent dye modification for labeling T-RFLP products. PCR reagents were mixed at the following final concentrations: 0.5 μM of each primer, 1× buffer B (Promega), 2.5 mM $MgCl_2$ (Promega), 250 μM dNTPs (Promega), 300 ng/μl BSA (Sigma A-7030 (Kirchman et al. 2001)), 2.5 U of Taq in buffer B (Promega), and 10 ng of DNA template. PCR was performed with the following thermal protocol: one cycle at 95 °C for 2 min, 35 cycles at 95 °C for 30 s, 50 °C for 30 s, 72 °C for 1.5 min, and a final extension at 72 °C for 5 min with a hold at 4 °C. Three to five replicate amplifications were performed for each DNA extract to provide adequate amounts of PCR product for subsequent analyses. Amplifications from the two duplicate time-zero samples and the three replicate samples collected at 24 and 72 h of incubation were pooled within a time point to average between-bottle variability and to obtain sufficient DNA for the analysis.

T-RFLP analysis. Multiple T-RFLP amplifications for a particular time point were cleaned and concentrated into a single 50-μl aliquot with the UltraClean PCR Clean-up Kit (Mo Bio Labs, Carlsbad, CA). Concentrated PCR products were band-isolated on SeaKem agarose gels in 1× TBE buffer to reduce the possibility of downstream analysis of non-specific PCR products. Amplicons were excised from gels as described above and purified with the UltraClean 15 DNA Purification Kit (Mo Bio Labs). Purified and concentrated T-RFLP products were digested with 10 U of mung bean nuclease (New England Biolabs, Beverly, MA) to remove any partially single-stranded amplicons that might confound T-RFLP analysis (Egert and Friedrich 2003). Mung bean digests were cleaned and purified with the UltraClean PCR Clean-up Kit described above. The resulting blunt-ended, dye-labeled T-RFLP products were quantified by PicoGreen (Molecular Probes) fluorescence. For each sample, a single restriction enzyme digest was performed on 300 ng of T-RFLP product using 10 U of each of four different enzymes (*Hae*III, *Hha*I, *Mnl*I, and *Rsa*I). Enzymes were chosen for this analysis based on the results of an in silico T-RFLP study, which examined the number of unique fragments obtained from a large set of publicly available protistan 18S rDNA sequences over the region of the gene that we amplified. In practice, for the samples analyzed in this study and the four enzymes used, the method resolved 30–70 fragments per sample.

Restriction digests were precipitated with a mixture of 1 µl of 20 mg/ml glycogen (Roche), 2 µl of 3 M sodium acetate (Sigma), and 2.5 times the digest volume of ice-cold 95% ethanol. Samples were mixed well and then centrifuged at 20,000 *g* and 4 °C for 15 min. The liquid phase was decanted from the centrifuge tubes and the pellets were washed twice with 100 µl ice-cold 70% ethanol, with each wash cycle consisting of a spin at 20,000 *g* at 4 °C for 5 min. The resulting DNA pellet was resuspended in 40 µl of deionized formamide (Beckman-Coulter). A fraction of the resuspended sample (5–10 µl) plus 0.5 µl of 600-bp size-standard (Beckman-Coulter) was diluted out with additional formamide to a final volume of 40 µl. T-RFLP fragment samples were separated on a CEQ8000 capillary-based DNA sequencer (Beckman-Coulter) following general guidelines for T-RFLP (Grüntzig et al. 2002). Sample results were analyzed with the Fragment Analysis module of the CEQ8000 Genetic Analysis software (Beckman-Coulter) and binned at single base pair resolution using the CEQ AFLP binning application. Fragment data were further analyzed with a peak-area normalization macro program written by Chris Kitts (California Polytechnic State University, San Luis Obispo, CA) to allow relative peak (phylotype) comparisons among samples (Kaplan and Kitts 2004).

RESULTS

Estimates of eukaryote diversity based on cloning/sequencing. The 18S rDNA clone library established at time-zero from our coastal ocean sample was strongly dominated by alveolate sequences (Fig. 1). These sequences were compared to the ARB database, which contains approximately 6,000 eukaryotic sequences. The ARB database is not exhaustive, but it has the advantage that the sequences have been given species names (i.e. for protists they can be related to morphologically defined taxa). GenBank contains a much richer rDNA sequence database, and our sequences matched many of the GenBank sequences with higher similarity than they matched sequences in the ARB database. However, many of the GenBank sequences are currently classified as ''uncultured'', ''unclassified,'' or ''environmental'' and thus often did not help resolve the taxonomic affinities of our clones.

When matched against the ARB database, more than half of the clones sequenced at time-zero had highest sequence similarity to alveolate sequences, with approximately equal portions of these sequences indicating affinity with the ciliates and dinoflagellates of 25% and 28%, respectively (Fig. 1). When matched to GenBank sequences, the percentages of clones falling into these cat-

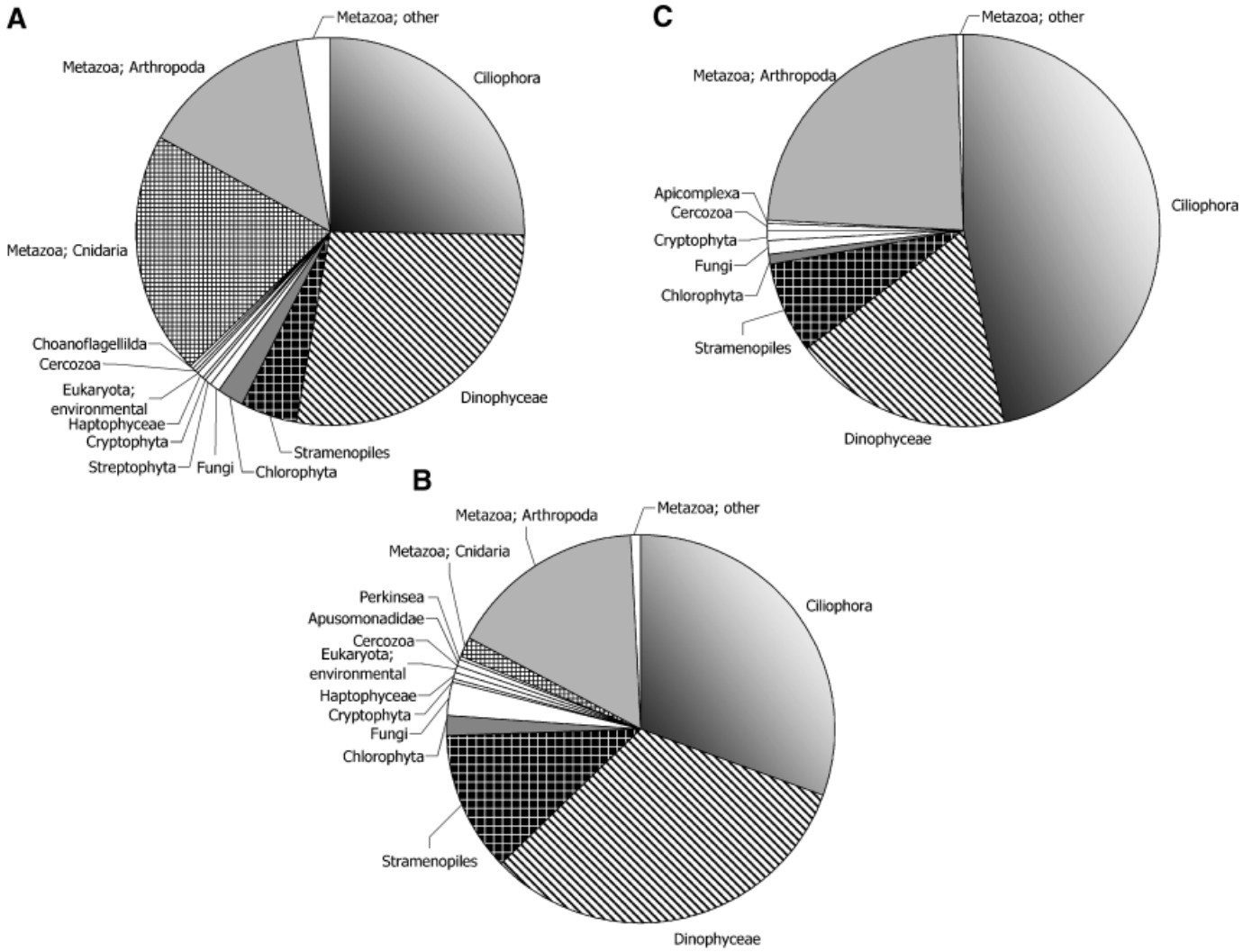

Fig. **1.** Major taxonomic groupings of microbial eukaryote sequences at time-zero, $n = 493$ sequences **(A)**, time-24 h, $n = 265$ sequences **(B)**, and time-72 h, $n = 212$ sequences **(C)** for samples from the coastal North Atlantic. Groupings are based on comparison of our sequences with all 18S rDNA sequences in the ARB database (Ludwig et al. 2004).

egories were much lower (18% ciliates, 7% dinoflagellates; data not shown). Many of the alveolate sequences more closely matched ''environmental eukaryote'', ''unclassified environmental,'' or ''unclassified alveolate'' sequences in GenBank. These latter sequences may represent ciliate or dinoflagellates that have not yet been placed in the ARB database, or they may represent phyla from as yet morphologically uncharacterized, alveolate groups (Diez, Pedros-Alio, and Massana 2001; López-García et al. 2001; Massana et al. 2002).

Metazoan sequences also contributed significantly to our clone library at time-zero (Fig. 1). Arthropod sequences (predominantly copepods) constituted 14% of the library, while cnidarian sequences (presumably fragments of medusae and siphonophores) constituted 20% of the clones. Clones from a variety of other phyla were present at low abundances in the library. Stramenopile and chlorophyte sequences constituted 2–5% of the clone library while other taxa were present at $<2\%$ (including Cercozoa, cryptophytes, choanoflagellates, fungi, haptophytes).

Minor shifts in eukaryote community composition during the 3-day incubations were apparent at the highest taxonomic levels. Mostly notably, cnidarian sequences decreased rapidly and were nearly undetectable after 24 h of incubation. It is not surprising that these sequences did not persist in the bottles since we assume that only pieces of these delicate gelatinous organisms were present in the <200-μm size fraction. Decreases in the relative abundance of the cnidarian sequences after 24 and 72 h of incubation were matched by increases in several protistan taxa. Ciliate sequences increased to 32% at time-24 h (Fig. 1B) and to 47% at time-72 h (Fig. 1C). The relative contribution of stramenopile sequences nearly tripled during the first 24 hours of incubation but decreased during the last 48 h of incubation to approximately 8%. Dinoflagellate sequences also constituted a greater percentage of the clone libraries at time-24 h relative to time-zero or time-72 h. Sequences from a variety of taxa continued to constitute minor percentages of the clone libraries after 24 and 72 h of incubation.

In contrast to relatively modest changes in the composition of the assemblage at higher taxonomic groupings, the composition of the eukaryote community changed substantially during the 3-day incubations when examined with greater taxonomic resolution. This latter finding was established through pairwise comparisons of the clone sequences from each of the three time points (Fig. 2). The 970 sequences obtained in the study grouped into 165 unique phylotypes at the 95% similarity level. Of these, only approximately 18% of the phylotypes were present at all three time points, 17% were observed at two of the sampling times, while 108 phylotypes were observed at only one time point (54 unique phylotypes at time-zero, 37 unique phylotypes at time-24 h, and 17 unique phylotypes at time-72 h). These unique phylotypes constituted 8–14% of the number of clones sequenced at each time point and 65% of all phylotypes observed (108 out of 165). Many phylotypes observed among the 970 clones obtained in this study were rare. More than half of the phylotypes were detected only once or twice, yielding a rank abundance curve characterized by a long ''tail'' of phylotypes occurring at low abundance (Fig. 2). The most common phylotype constituted approximately 7% of the total number of clones.

Estimation of the total diversity of the protistan assemblage was examined by constructing rarefaction curves (Fig. 3) of the clones distributed among the observed phylotypes (Colwell, Mao, and Chang 2004) using EstimateS version 7 (Colwell 2004). Rarefaction curves provide an estimate of phylotype diversity relative to

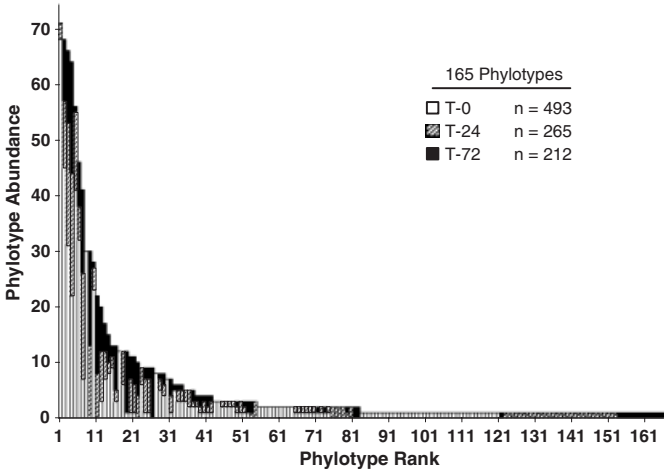

Fig. **2.** Phylotype rank abundance curve for 970 sequences obtained from a coastal North Atlantic sample. Sequences from three time points of a 72-h incubation have been combined (see Fig. 1 for clones obtained at each time point). Phylotypes were grouped into phylotypes at a similarity level of 95% based on pairwise analysis of all 970 sequences using ClustalX (Thompson et al. 1997).

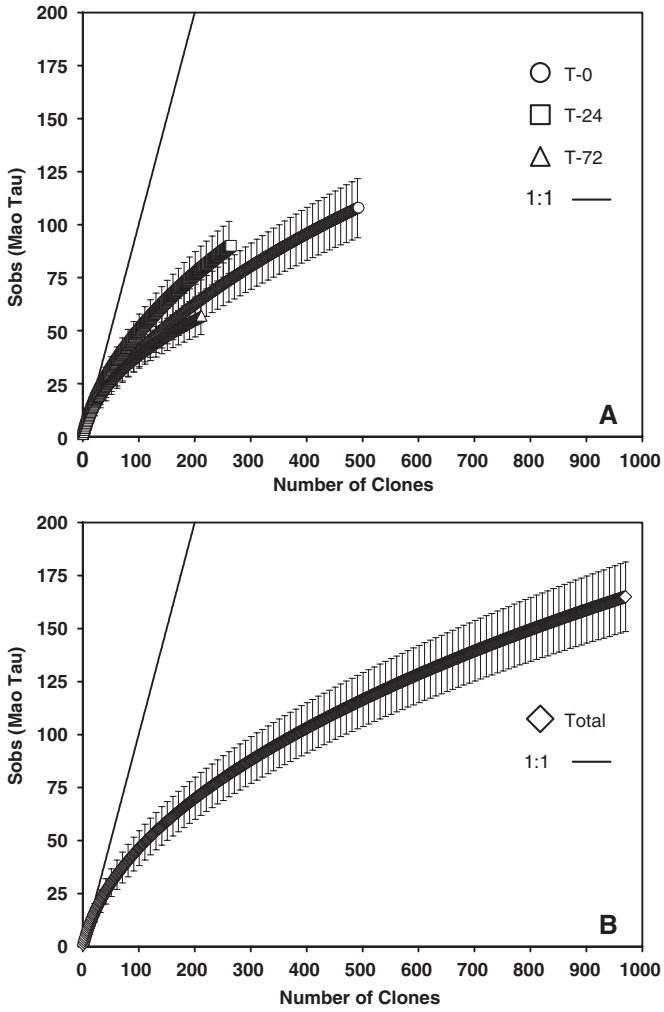

Fig. **3.** Rarefaction curve analysis of S_{obs} (Mao Tau) vs. sequencing effort calculated using EstimateS version 7 (Colwell 2004). Rarefaction curves for samples collected at three time points of a 72-h incubation are presented separately **(A)** and for the combined set of sequences detected across all time points **(B)**.

sampling effort (i.e. number of clones sequenced). Curves for the three time points sampled showed a minor difference in their trajectories, with the curve for the time-24 h time point slightly above the curves for time-zero and time-72 h (Fig. 3A). None of the curves approached an asymptote (i.e. total assemblage diversity was not estimated) for the number of clones sequenced. A single rarefaction curve generated by pooling data from all three time points also did not indicate saturation of our sequencing effort (Fig. 3B). This finding was consistent with the high occurrence of rare phylotypes observed in the rank abundance curve (Fig. 2).

Diversity of the protistan assemblage was estimated using the Chao1 estimator (Chao 1984, 1987) for the three time points combined (Fig. 4) as calculated by EstimateS version 7 (Colwell 2004). This estimator is a non-parametric species richness estimator based on the frequency of singletons and doubletons (phylotypes occurring only once or twice) in the sequence library. The statistic is designed to estimate the probable diversity of the community from which the sample is drawn. One-hundred randomizations of sample selection were performed both with and without replacement of sequences from the population of 165 phylotypes. This procedure yielded estimates of Chao1 (bias corrected) for diversity of the eukaryote assemblage of approximately 162 (144–202, 95% CI) phylotypes using ''with replacement'' (Fig. not shown) to 282 (229–381, 95% C.I.) phylotypes using ''without replacement'' (Fig. 4). These estimates, based on the pooled data from all three time points, were higher than estimates for any of the time points individually.

Estimates of eukaryote diversity based on T-RFLP. T-RFLP analysis provides an estimate of eukaryote assemblage diversity based on discrimination among DNA fragments of different lengths with a resolution of 1-bp difference between 60 and 640 bp (Fig. 5). The analysis provides a ''snapshot'' of assemblage diversity in a single chromatogram, allowing for direct comparisons of changes in the relative contribution of phylotypes (fragments of a specific size) between samples.

An underlying assumption of the T-RFLP method is that each fluorescently labeled DNA fragment detected represents a unique phylotype represented by a single species in the microbial eukaryote assemblage. This assumption is violated if two taxa possess restriction sites in the 18S rRNA gene at positions that yield labeled fragments of identical length. It is anticipated that only closely related species will have identically sized terminal restriction fragments (TRFs) because sequence variability (insertions or deletions of nucleotides) among more distantly related taxa, and/or mutations affecting the position of recognition sites within the gene, should cause differences in the size of labeled fragments. For any given restriction enzyme, it is also possible that some subset of the amplified protistan assemblage will lack any recognition sites causing a cluster of peaks at the upper size-range of the fragments.

The height and peak-area of a TRF are assumed to be indicators of the relative contribution of a phylotype to the overall community composition. This assumption is affected by differences in copy number of the 18S rDNA gene among taxa (this is unknown for most taxa) and differences among taxa in the amplification efficiency of the targeted region of the gene with particular primer sets (also unknown for mixed environmental samples). Thus, peak height in a T-RFLP chromatogram does not correlate directly with cell abundance in a given sample. For these reasons, T-RFLP is not believed to yield an accurate estimate of total phylotype richness and abundance within a sample, but it is a rapid method for comparing differences in community composition for samples collected from different environments or subjected to different conditions (such as bottle incubation).

The four restriction enzymes used to digest rDNA amplified from our samples provided somewhat different estimates of total assemblage diversity (Fig. 6). Summed over all three time points, *Hae*III yielded 76 unique phylotypes, *Mnl*I and *Hha*I yielded 62 and 60 phylotypes, respectively, while *Rsa*I yielded only 46 phylotypes (Table 1). The *Rsa*I data indicated that many eukaryotes possessed a restriction site for the enzyme near the 5′ end of the ssu rRNA gene (many small fragments) while others had a recognition site near the 3′ end or lacked a recognition site within the gene (many very large fragments). This distribution of fragment sizes makes *Rsa*I less useful for diversity analysis.

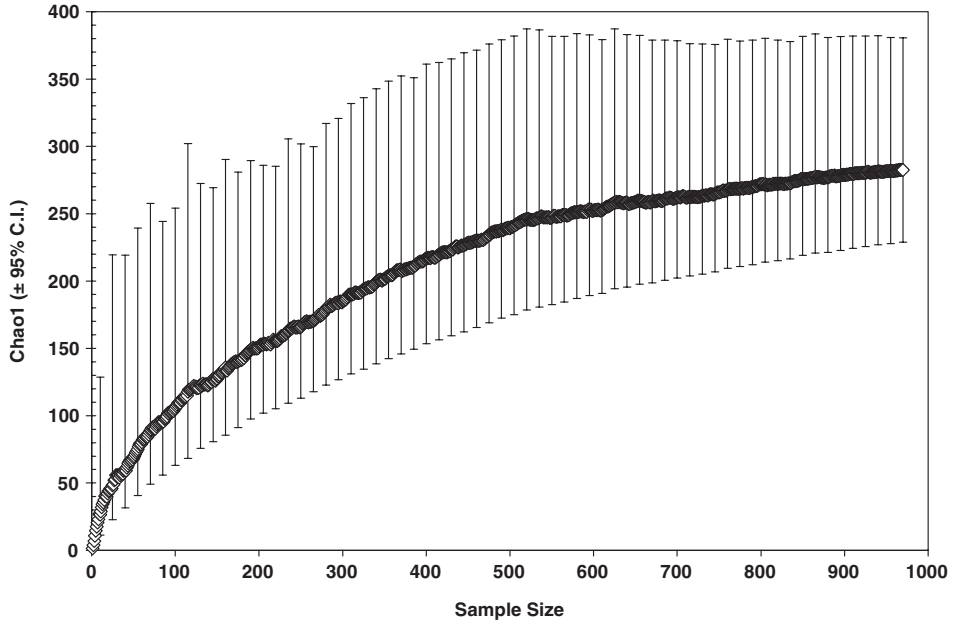

Fig. **4.** Chao1 phylotype richness estimator (100 randomizations) calculated without replacement. Error bars represent the 95% confidence estimates around each richness estimate.

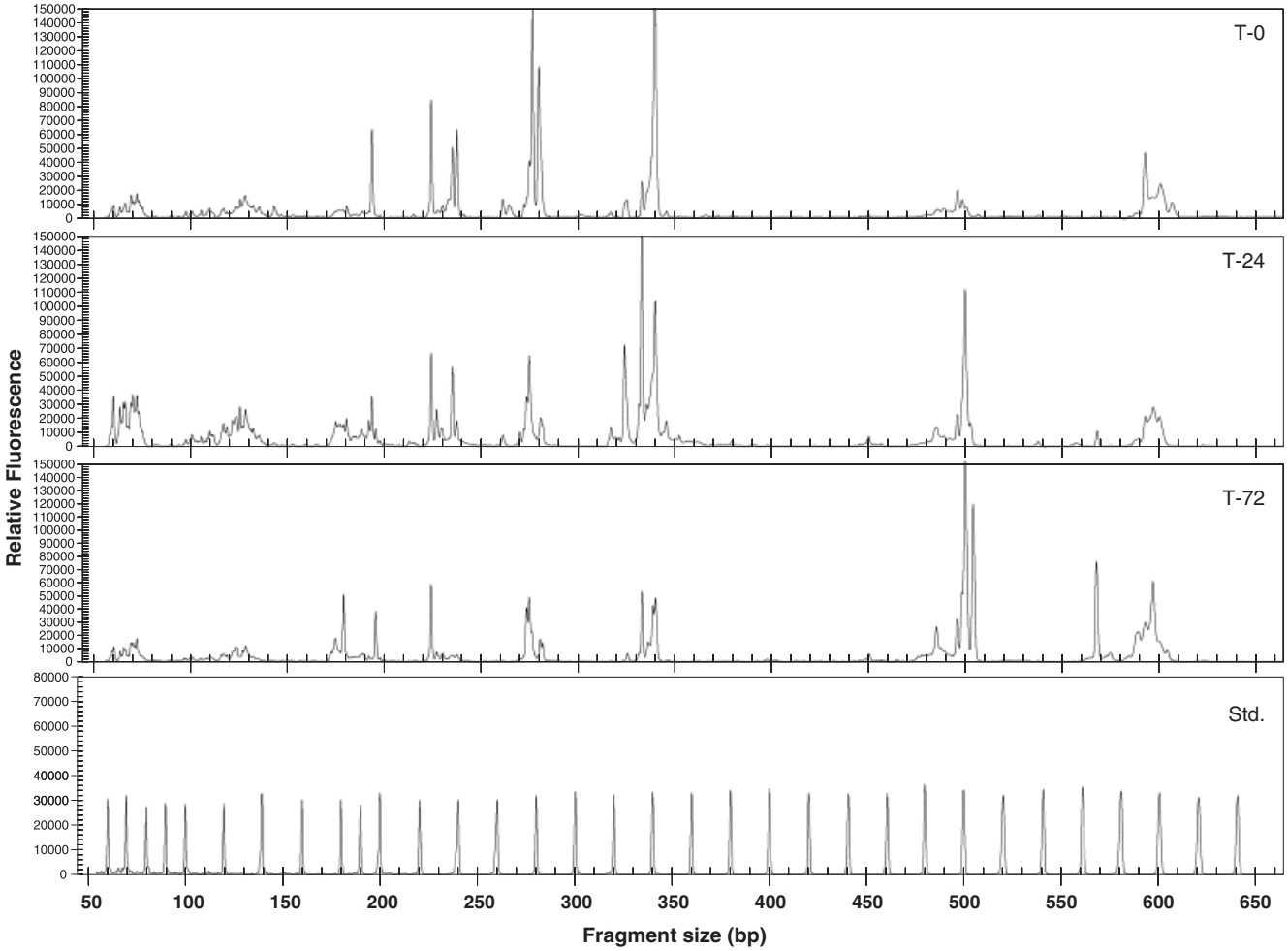

Fig. **5.** Microbial eukaryote assemblage (<200-μm size fraction) diversity represented by T-RFLP chromatograms of *Hae*III digests of 18S rDNA. Chromatograms depict relative fluorescence (*Y*-axis) vs. fragment size in bp (*X*-axis) from time-zero (T-0), time-24 h (T-24), and time-72 h (T-72) samples of a 72-h incubation of a natural sample from the coastal North Atlantic. Fragment sizes were determined by comparison with DNA fragment size standards (Std.). Binning of T-RFLP fragments was set to 1-bp resolution across the displayed size-range (60–640 bp) using peak heights >0.5% of the total peak area within a chromatogram as the threshold value for detection of phylotypes. Sizing of DNA fragments was accomplished with the Beckman CEQ 8000 fragment analysis module. Peak-area data were normalized with an Excel Macro program written by Chris Kitts (California Polytechnic State University, San Luis Obispo, CA), which allowed comparison of peaks both within and across time points.

T-RFLP analysis based on digestion with all four enzymes indicated a substantial shift in the composition of the eukaryote assemblage during the 72-h incubation of the natural sample (Table 1). This finding is in general agreement with shifts in phylotype diversity estimated by sequencing and pairwise comparisons of sequences (Fig. 1). For example, the *Hae*III digest yielded 46 phylotypes at time-zero of which 9 phylotypes (20%) did not appear in T-RFLP patterns at time-24 h or time-72 h. Similarly, 12 of 52 phylotypes (23%) were present only at time-24 h, and 5 of 41 phylotypes (12%) were present only at time-72 h. These results indicate that a substantial number of taxa increased or decreased in relative abundance during the intervals between sampling. Ubiquitous phylotypes (fragments present at all three time points) only constituted approximately half of the phylotypes.

Trends in the diversity estimates obtained from T-RFLP analysis at each time point were similar for the four enzymes (Table 1). In general, phylotype diversity at time-zero and time-72 h was similar for a particular enzyme, while phylotype diversity at time-24 h for the same enzyme was always greater than at time-zero or time-72 h. For example, 46 phylotypes were observed at time-zero for the *Hae*III fragment profile, increasing to 59 phylotypes at time-24 h and then decreasing again to 41 phylotypes at time-72 h. Unique phylotypes (phylotypes appearing at only one time point; parentheses in Table 1) followed the same trend as total fragment number (i.e. highest numbers of unique phylotypes at time-24 h relative to time-zero and time-72 h for each enzyme).

The Shannon Diversity Index (Shannon 1948) was calculated for each T-RFLP profile. This index combines information on the total number of phylotypes in a sample and its relative proportion to the total amount of amplified DNA. Trends in this index for the three time points for each enzyme were similar to trends observed for total phylotypes (higher values at time-24 h relative to time-zero or time-72 h).

DISCUSSION

Molecular estimates of microbial diversity. Cloning/sequencing of small subunit ribosomal RNA genes (16S, 18S) has rapidly become the ''gold standard'' for culture-independent,

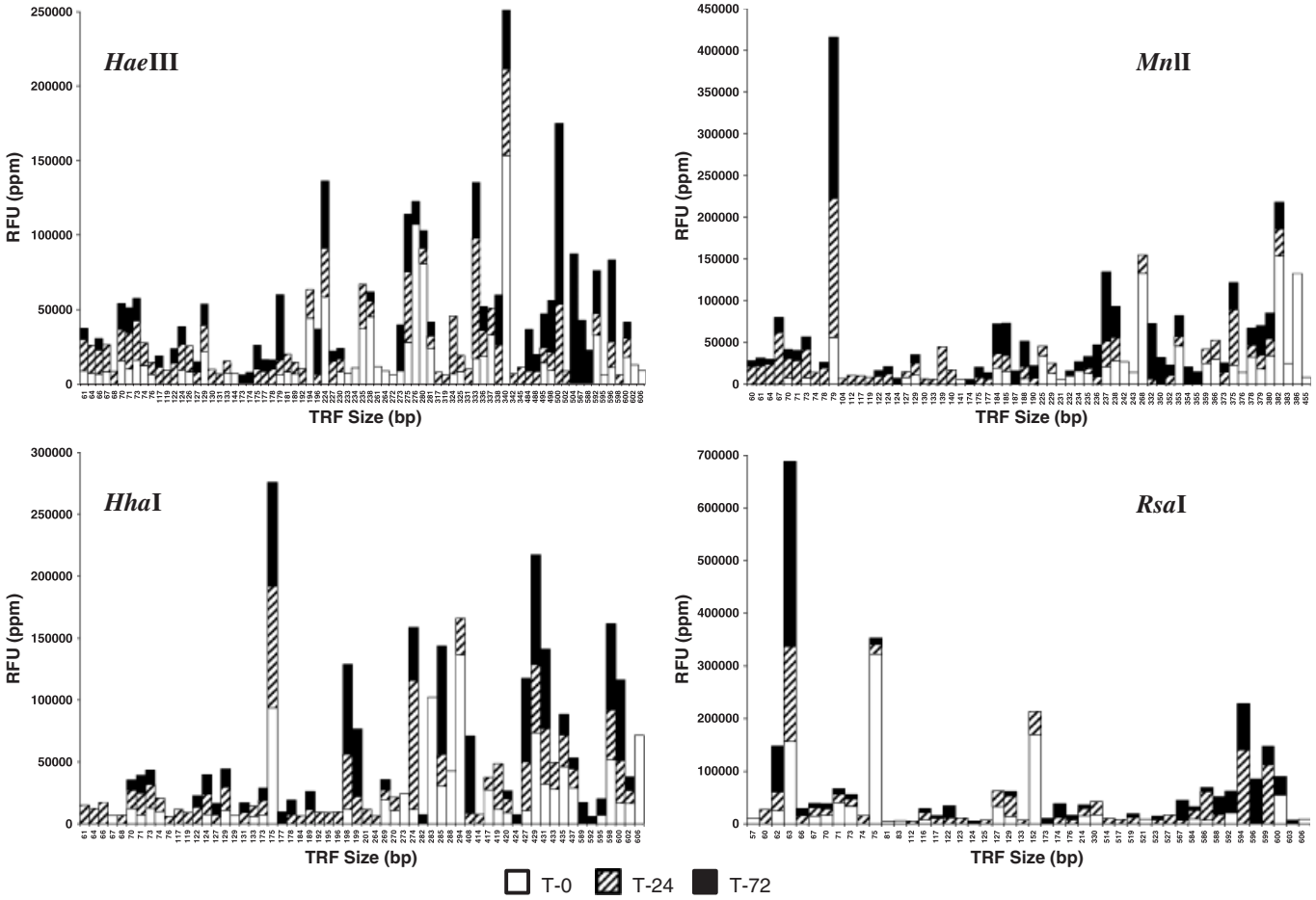

Fig. **6.** Contribution of particular microbial eukaryote T-RFLP phylotypes at each time point (different shading) for each of the four restriction enzymes (*Hae*III, *Mnl*I, *Hha*I, and *Rsa*I) utilized in this study. The total peak area has been normalized to 1 million units of fluorescence and individual peaks scaled accordingly to allow for relative comparisons within a particular digest.

molecular analyses of microbial biodiversity and phylogeny (DeLong and Pace 2001; Giovannoni and Rappe 2000; Sogin 1989). While by no means foolproof, this approach possesses the advantage that it is now relatively straightforward to amplify, clone, and sequence these genes from environmental samples (many well-tested primers and protocols exist). In addition, the DNA sequences that it yields provide information on species richness (i.e. numbers of unique phylotypes) as well as information on the identity of individual taxa (to the degree that these phylotypes are represented and identified in public databases). Ribosomal RNA genes possess sufficient variability that distinctions among phylotypes at or near the species level are possible even when considering only partial gene sequences.

While powerful, analysis of microbial community structure through the cloning and sequencing of ribosomal RNA genes is time-consuming and costly. Ultimately, the sensitivity of the

Table **1.** T-RFLP summary for the four restriction enzymes (*Hae*III, *Mnl*I, *Hha*I, and *Rsa*I) used in this study.

	*Hae*III		*Mnl*I		*Hha*I		*Rsa*I	
	n	*H′*	*n*	*H′*	*n*	*H′*	*n*	*H′*
Time-zero	46 (9)	1.44	32 (8)	1.30	34 (6)	1.34	22 (5)	0.99
Time-24 h	59 (12)	1.67	46 (9)	1.50	48 (14)	1.54	35 (9)	1.31
Time-72 h	41 (5)	1.47	39 (6)	1.40	33 (5)	1.34	29 (4)	1.01
Total TRF phylotypes	76		62		60		46	
Ubiquitous TRFs	22		16		20		12	

The total number of unique TRF phylotypes detected by each enzyme is listed at the top of each panel (e.g. 76 phylotypes across all three time points for *Hae*III), with a breakdown of TRFs by time point (*n*). Numbers in parentheses indicate the number of unique phylotypes within a time point, while ubiquitous phylotypes were detected at all three time points. These data imply a moderately diverse protistan community (50–70 total phylotypes),and similar diversity (similar numbers of phylotypes) at the start and end of the experiment, with a peak diversity at *T*-24 h. The Shannon Diversity Index (*H′*), which accounts for the total number of phylotypes in a sample plus the relative proportion of each TRF, was computed from the relative peak abundance ratios.

T-RFLP, terminal restriction fragment length polymorphism; TRFs, terminal restriction fragments.

approach is limited more by the practical limitation of the number of clone libraries that can be established and sequenced than by the inherent sensitivity of the method to document phylotype diversity in a sample. ''Exhaustive'' sequencing of clones is rarely attempted (because of cost), and most studies to date have typically characterized not more than approximately 100 clones of any single environmental library.

Prohibitive cost and the necessity for ecologists to assay numerous samples in most ecological studies have stimulated the development of a number of genetic approaches based on the analysis of DNA fragment length polymorphisms. DGGE, ARISA, T-RFLP, and other methods have been used to obtain an index of microbial diversity in natural samples relatively quickly (and inexpensively) compared with large sequencing programs (Diez et al. 2001; Fisher and Triplett 1999; Fuhrman, Griffith, and Schwalbach 2002; Gast, Dennett, and Caron 2004; Massana and Jurgens 2003). These methods provide a powerful ecological tool for characterizing the dominant phylotypes in a sample, and a means of performing comparative studies between different samples. However, they typically do not resolve the phylogenetic affinities of the taxa that represent the DNA fragments, nor do they adequately document the presence of rare phylotypes. Thus, fragment analysis possesses advantages that are somewhat complementary to the cloning and sequencing of genes from environmental samples.

Protistan diversity and the effect of containment. The objectives of this study were two-fold. First, this work constitutes a component of a large-scale study that is presently underway to characterize 18S rRNA gene diversity in environmental samples from diverse ecosystems (Dawson and Pace 2002; Diez, Pedros-Alio, and Massana 2001; Edgcomb et al. 2002a; Fawley, Fawley, and Buchheim 2004; Guillou et al. 2004; López-García et al. 2001, 2003; Massana et al. 2002, 2004; Moon-van der Staay, De Wachter, and Vaulot 2001; Romari and Vaulot 2004; Stoeck and Epstein 2003). One broad goal of this work is to plumb the depths of protistan diversity using methodologies that do not possess the same biases and shortcomings of culture and microscopy. Thus far, this work has uncovered significant numbers of ''unknown'', ''uncultured'' eukaryote phylotypes whose identities are presently under study, and it has generated skepticism regarding the extent to which we have fully characterized the diversity of natural microbial assemblages (Dawson and Pace 2002; Habura et al. 2004; Holzmann et al. 2003; López-García et al. 2001; Massana et al. 2004).

In addition, DNA sequence information provides many additional characters for identifying microbial taxa that can be assessed quickly and accurately in complex assemblages of microorganisms. It thereby provides the potential to develop novel approaches for conducting ecological studies of microbial taxa in nature. As noted previously, species identification and quantification are formidable tasks at present because of the varied methodologies that must be used for collecting, preserving, examining, and identifying protistan taxa. If DNA sequence information is going to be applied effectively in microbial ecological studies, it is necessary to determine the extent of phylotype diversity in natural ecosystems and how that diversity relates to classical (morphology based) taxonomic schemes of microbial eukaryotes (Modeo et al. 2003).

A second objective of this study was stimulated by the desire to examine the effects of containment on protistan diversity and taxonomic composition of the assemblage. The propensity for containment to alter the composition of a natural bacterial community has been recognized for more than 60 yr (Zobell 1943). Nevertheless, the incubation of water samples in bottles remains a necessary and fundamental approach for making measurements of microbial physiological rates such as primary production (photosynthesis), bacterial production, microbial feeding activities, and population growth (Caron 2000; Ducklow 2000; Falkowski 1994; Strom 2000). Very little has been done to characterize changes in the protistan assemblage during incubation, and the length of incubations can be quite long (1–3 days) for these assemblages relative to their reproductive rates. Therefore, our goal was to assess how the community structure of a natural water sample changes during incubation in a typical ecological experiment.

Characterization of the diversity of the <200-μm microbial assemblage in our coastal ocean sample from the western North Atlantic indicated a tremendous diversity of eukaryotes at the time and depth of our study. Our cloning/sequencing effort was designed to survey the extent of phylotype diversity in our samples rather than provide high-resolution phylogenetic relationships among our clones. The region of the 18S rRNA gene chosen for sequencing has generally been found to provide at least genus-level differentiation among clones (Romari and Vaulot 2004). Therefore, we did not attempt to obtain full-length information for our 18S rDNA clones, but rather we obtained partial sequence from a larger number of clones than has been undertaken in previous ecological studies.

A total of 165 unique phylotypes was observed among the 970 clones that were sequenced in this study, and many of the clones were represented by only one or two clones (Fig. 2). Both rarefaction and the ChaoI diversity estimators indicated the presence of hundreds of phylotypes in our sample based on our grouping of sequences into phylotypes at the 95% similarity level (Fig. 3, 4). It is worth noting that defining phylotypes at this level of similarity is conservative in that species-level distinctions for small subunit ribosomal RNA genes are often delineated at the 97% to 98% similarity level. Thus, our estimates presumably represent lower limit estimates of the true species diversity present in the sample.

Overall phylotype diversity did not change dramatically within the incubated sample during the 3-day experiment but the specific mixture of phylotypes in the clone libraries was markedly different at each sampling time. Sixty-five percent of the total number of phylotypes (108 out of 165) was observed at only one of the three sampling times. This result indicates that containment and incubation had a substantial effect on the taxonomic composition of the seawater sample as shown by our molecular analyses. Furthermore, it implies that the sample contained a large number of eukaryote phylotypes that were present at background abundances in the sample, and that numerous taxa increased above the detection threshold during the experiment. This observation is in agreement with studies that have shown that some protistan taxa previously thought of as common are actually present in most natural samples at low abundance and are encountered at high abundance only after enrichment (Lim, Dennett, and Caron 1999). These latter species represent 'weed' species that often dominate culture collections but that may not be highly representative of species abundant under most conditions in natural ecosystems.

Most studies to date have only cursorily examined protistan diversity on samples collected at discrete times and locations. While these previous studies have been extremely informative, our results indicate that other studies have most likely not revealed anything close to the true protistan diversity in a parcel of water. For example, if we had only analyzed our time-zero sample, we would have detected only 106 phylotypes instead of the 165 obtained after 72 h of incubation. To appropriately test the question ''Is everything everywhere?'' (Finlay and Fenchel 2004) using a molecular approach, it may therefore be necessary to routinely perturb protistan assemblages with different physical (temperature, light) and chemical (inorganic and organic nutrients, trace-metals) regimes to draw out the otherwise minor components of these assemblages. Recent work varying traditional microbial culture conditions has shown promise for detecting new lineages of previously undetected marine bacteria (Connon and Giovannoni 2002). The results of our

study are analogous in that they revealed an increased number of protistan phylotypes simply by incubating a water sample at in situ temperature and light conditions for several days.

Very few studies have undertaken multiple molecular-based analyses to assess protistan assemblage diversity in marine systems, with the notable exceptions of Casamayor et al. (2002), Diez et al. (2001), and Massana and Jurgens (2003). Diez et al. (2001) employed DGGE, cloning/sequencing, and T-RFLP to examine the diversity of a picoeukaryote ($<5\,\mu m$) assemblage at a site in the Mediterranean Sea, while Casamayor et al. (2002) sampled the eukaryote diversity in saltwater evaporation ponds by DGGE, ARISA, and T-RFLP at a site in Spain. Massana and Jürgens (2003) studied changes in North Sea flagellate assemblages ($<2\,\mu m$) in chemostat cultures by DGGE and T-RFLP. Each of the previous studies demonstrated the utility of using a combined approach for effectively assessing protistan diversity. Yet, the effectiveness of these approaches was dependent on the specific conditions of the method.

Diez et al. (2001) detected approximately two-thirds as many TRF-based OTUs (19 TRFs) as sequence-based OTUs (29). Lower overall phylotype diversity determined by T-RFLP in the present study likewise was substantially less than phylotype diversity obtained by sequencing rRNA genes, but our total number of phylotypes by either method was substantially greater than in the latter study. The restriction enzyme used for T-RFLP (*Hha*I) by Diez et al. (2001) yielded only the third highest diversity estimate of the four enzymes tested in our study (Table 1). Additionally, sequence-based OTUs were defined by Diez et al. (2001) based on identifying unique RFLP patterns among clones, grouping the clones into these putative phylotypes prior to sequencing, and then sequencing representatives of each RFLP pattern. This procedure is generally undertaken to reduce sequencing costs, but it can underestimate the actual sequence diversity because different phylotypes will sometimes yield the same RFLP pattern.

In our study, we defined phylotypes of our sequenced clones through pairwise sequence comparisons of all clones. We also sequenced a significantly larger number of clones than previous studies (970). The highest estimate of TRF-based phylotype diversity in our study was 76 TRFs, which comprised slightly less than half the number of sequence-based phylotypes ($n = 165$, Fig. 2). Therefore, differences in the absolute number of TRFs and sequence-based phylotypes between our study and Diez et al. (2001) may be a consequence of differences in methodological details of the molecular analyses, differences in the size-classes of assemblages examined ($<$ 200 vs. $<$ 5 μm), or inherent differences in the diversity of protistan assemblages.

Nevertheless, if our comparison and that of Diez et al. (2001) are representative of the effectiveness of these two methods, to document eukaryote diversity, then one might expect to detect approximately 40–70% of the sequence-based diversity in a corresponding clone library using T-RFLP analysis. Thus, while more costly, cloning/sequencing was necessary to better represent the total taxonomic diversity of the sample.

Although T-RFLP analysis did not yield diversity estimates similar to those obtained by cloning/sequencing, the method is still extremely versatile for the rapid comparison of relative changes in TRF frequency and abundance in seawater incubation experiments. Massana and Jürgens (2003) discovered that in contrast to bacterial TRF-diversity (which remained relatively constant), the TRF-based diversity of picoeukaryotic flagellates in chemostat cultures shifted dramatically among eight major phylotypes over the course of a 5-wk monitoring period. The previous result is in complete agreement with our observations based on TRF analysis of substantial shifts in the composition of the microbial eukaryote assemblage while maintaining similar levels of overall diversity at each sampling time. Moreover, when monitored alongside the sequence-based phylotypes from clone libraries, the relative proportion of TRF phylotypes roughly corresponds to the relative contribution of identified taxa within a library based on sequence identities and expected TRF fragments. Diez et al. (2001) reported that 16% of a clone library comprised of Prasinophyceae compared with 24% of total TRF peak area for the same putative phylotypes. Although we have made no attempt to identify TRFs in the present study, our previous work has suggested that the prasinophyte *Ostreococcus* spp. comprised 10% of a clone library from a coastal Pacific Ocean study site while the same putative phylotype comprised 9–11% of TRF peak area (Caron, Countway, and Brown 2004).

Few molecular studies of microbial eukaryote diversity have considered the contribution of metazoa to clone libraries or fragment analyses. Diez et al. (2001) made putative assignments of TRF identity, where possible. Despite attempts to limit their study to $<$ 5 μm protists, one of the largest signals in their T-RFLP profiles (TRF comprising 27% of the total peak area) was consistent with the expected TRF size of the metazoan *Oikopleura*. Our clone library results indicated a relatively large (and expected) contribution from metazoan taxa, especially arthropods and cnidarians, given the size-class of microeukaryotes included in our samples ($<$ 200 μm; Fig. 1). Cnidarian sequences declined to undetectable levels during the 72-h incubation, presumably an indication that either these species did poorly in containers or that our initial sample contained dead or damaged cnidarian biomass. In contrast, arthropod phylotypes increased to approximately 25% of the total number of phylotypes present at 72 h. As important predators of both heterotrophic and phototrophic protistan taxa, these metazoa may have significantly impacted species composition of the protistan assemblage at 24 and 72 h. The effect of metazoa (specifically Arthropoda) on protistan assemblage structure in seawater incubations is presently being investigated by Schnetzer et al. (in prep.).

The Chao1 species richness estimator calculated for the distribution of observed rDNA phylotypes has been applied to an increasing number of microbial diversity studies (Hill et al. 2003; Hughes et al. 2001; Kemp and Aller 2004). Similar statistics are only beginning to be applied to protistan rDNA clone libraries (this study) to assess the level of diversity in size-fractions that capture most of the protistan assemblage present in planktonic communities. In contrast, previous studies have examined primarily prokaryotic or picoeukaryotic assemblages. The distribution of microbial eukaryote phylotypes in this study was approximately log normal, which is typical of a dynamic community (Curtis, Sloan, and Scannell 2002). The estimate of Chao1 (without replacement) for the combined dataset in this study of 282 (229–381, 95% CI) represents more than 100 additional phylotypes than were detected among the 970 sequenced clones. Non-replacement of samples into the dataset ensures that all of the detected diversity is used to estimate the total diversity of the community. Randomization of sampling without replacement is the more appropriate method (relative to sampling with replacement) for calculating Chao1 with our sequence-based data because it does not assume that the entire ''universe'' of protistan sequences has been sampled (Colwell, R. K., pers. commun.).

Our demonstration of a highly diverse protistan assemblage that responds rapidly to containment (or changing environmental conditions) is a first step in beginning to examine relationships between microbial diversity and community function. The emergence of new methodology that allows estimation of protistan diversity in natural ecosystems will help enable investigators to begin to ask and answer questions relating to the significance of microbial diversity. Experimental studies that address community-level diversity, and the importance of that diversity for community stability, functional redundancy, resilience, and

predictability are rare, and have largely used defined (artificial) communities due to present limitations in the methodology available for characterizing the diversity of natural microbial communities. (McGrady-Steed, Harris, and Morin 1997; Naeem and Li 1997). These and similar experimental studies will benefit greatly from the application of modern genetic approaches for assaying community diversity and species composition of natural assemblages.

ACKNOWLEDGMENTS

The authors would like to thank the following people for assistance with this study: Mark Dennett, Dawn Moran, David Beaudoin, Rebecca Shaffner, and Julie Rose for their participation on the research cruise aboard R/V Endeavor; Matt Travao, Astrid Schnetzer, and Stephanie Moorthi for help with 18S rDNA cloning and sequencing; and Patrick Vigil for assistance with T-RFLP sample processing. Chris Kitts generously shared his T-RFLP research experience with us and provided a macro program for the normalization of TRF peak areas generated by our CEQ8000. Rory Lentz and Richard Palmieri of Beckman Coulter provided valuable application assistance with DNA sequencer-based work. This project was supported by NSF Grant numbers OCE-9818953 awarded to David A. Caron, Rebecca J. Gast and MCB-0084231, awarded to Jed A. Fuhrman and David A. Caron.

LITERATURE CITED

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**:3389–3402.

Amaral Zettler, L. A. & Caron, D. A. 2000. New insights into the phylogeny of the Acantharea based on SSU rRNA gene sequencing. *Eur. J. Protistol.*, **36**:34–39.

Avaniss-Aghajani, E., Jones, K., Holtzman, A., Aronson, T., Glover, N., Boian, M., Froman, S. & Brunk, C. F. 1996. Molecular technique for rapid identification of mycobacteria. *J. Clin. Microbiol.*, **34**:98–102.

Azam, F., Fenchel, T., Field, J. G., Gray, J. S., Meyer-Reil, L. A. & Thingstad, F. 1983. The ecological role of water-column microbes in the sea. *Mar. Ecol. Progr. Ser.*, **10**:257–263.

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Wheeler, D. L. 2004. GenBank: update. *Nucleic Acids Res.*, **32**:D23–26.

Bowers, H. A., Tengs, T., Glasgow, H. B. Jr., Burkholder, J. M., Rublee, P. A. & Oldach, D. W. 2000. Development of real-time PCR assays for rapid detection of *Pfiesteria piscicida* and related dinoflagellates. *Appl. Environ. Microbiol.*, **66**:4641–4648.

Caron, D. A. 2000. Protistan herbivory and bacterivory. *In*: Paul, J. (ed.), Marine Microbiology: Methods in Microbiology. Academic Press, London. **30**:289–315.

Caron, D. A., Countway, P. D. & Brown, M. V. 2004. The growing contributions of molecular biology and immunology to protistan ecology: molecular signatures as ecological tools. *J. Eukaryot. Microbiol.*, **51**:38–48.

Casamayor, E. O., Massana, R., Benlloch, S., Ovreas, L., Diez, B., Goddard, V. J., Gasol, J. M., Joint, I., Rodriguez-Valera, F. & Pedros-Alio, C. 2002. Changes in archaeal, bacterial and eukaryal assemblages along a salinity gradient by comparison of genetic fingerprinting methods in a multipond solar saltern. *Environ. Microbiol.*, **4**:338–348.

Chao, A. 1984. Non-parametric estimation of the number of classes in a population. *Scand. J. Stat.*, **11**:265–270.

Chao, A. 1987. Estimating the population size for capture-recapture data with unequal catchability. *Biometrics*, **43**:783–791.

Colwell, R. K. 2004. Estimates: Statistical Estimation of Species Richness and Shared Species from Samples. Version 7. User's Guide and Application. Published at: *http://purl.oclc.org/estimates*

Colwell, R. K., Mao, C. X. & Chang, J. 2004. Interpolating, extrapolating, and comparing incidence-based species accumulation curves. *Ecology*, **85**:2717–2727.

Connon, S. A. & Giovannoni, S. J. 2002. High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. *Appl. Environ. Microbiol.*, **68**:3878–3885.

Coyne, K. J. & Cary, S. C. 2005. Molecular approaches to the investigation of viable dinoflagellate cysts in natural sediments from estuarine environments. *J. Eukaryot. Microbiol.*, **52**:(in press).

Coyne, K. J., Hutchins, D. A., Hare, C. E. & Cary, S. C. 2001. Assessing temporal and spatial variability in *Pfiesteria piscicida* distributions using molecular probing techniques. *Aquat. Microb. Ecol.*, **24**:275–285.

Curtis, T. P., Sloan, W. T. & Scannell, J. W. 2002. Estimating prokaryotic diversity and its limits. *Proc. Natl. Acad. Sci. USA*, **99**:10494–10499.

Dawson, S. C. & Pace, N. R. 2002. Novel kingdom-level eukaryotic diversity in anoxic environments. *Proc. Natl. Acad. Sci. USA*, **99**:8324–8329.

Delong, E. F. 1992. Archaea in coastal marine environments. *Proc. Natl. Acad. Sci. USA*, **89**:5685–5689.

DeLong, E. E. & Pace, N. R. 2001. Environmental diversity of Bacteria and Archaea. *Syst. Biol.*, **50**:470–478.

Diez, B., Pedros-Alio, C., Marsh, T. L. & Massana, R. 2001. 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. 2001. 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.

Ducklow, H. W. 2000. Bacterial production and biomass in the oceans. *In*: Kirchman, D. L. (ed.), Microbial Ecology of the Oceans. Wiley-Liss, Inc., New York. p. 85–120.

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

Edgcomb, V. P., Simpson, A. G. B., Zettler, L. A., Nerad, T. A., Patterson, D. J., Holder, M. E. & Sogin, M. L. 2002b. Pelobionts are degenerate protists: insights from molecules and morphology. *Mol. Biol. Evol.*, **19**:978–982.

Edvardsen, B., Shalchian-Tabrizi, K., Jakobsen, K. S., Medlin, L. K., Dahl, E., Brubak, S. & Paasche, E. 2003. Genetic variability and molecular phylogeny of *Dinophysis* species (Dinophyceae) from Norwegian waters inferred from single cell analyses of rDNA. *J. Phycol.*, **39**:395–408.

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

Falkowski, P. G. 1994. The role of phytoplankton photosynthesis in global biogeochemical cycles. *Photosynth. Res.*, **39**:235–258.

Fawley, M. J., Fawley, K. P. & Buchheim, M. A. 2004. Molecular diversity among communities of freshwater Microchlorophytes. *Microb. Ecol.*, **48**:489–499.

Finlay, B. J. & Fenchel, T. 2004. Cosmopolitan metapopulations of free-living microbial eukaryotes. *Protist*, **155**:237–244.

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., Griffith, J. F. & Schwalbach, M. S. 2002. Prokaryotic and viral diversity patterns in marine plankton. *Ecol. Res.*, **17**:183–194.

Fuhrman, J. A., McCallum, K. & Davis, A. A. 1992. Novel major archaebacterial group from marine plankton. *Nature*, **356**:148–149.

Galluzzi, L., Penna, A., Bertozzini, E., Vila, M., Garces, E. & Magnani, M. 2004. Development of a real-time PCR assay for rapid detection and quantification of *Alexandrium minutum* (a dinoflagellate). *Appl. Environ. Microbiol.*, **70**:1199–1206.

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.

Giovannoni, S. J., Britschgi, T. B., Moyer, C. L. & Field, K. G. 1990. Genetic diversity in Sargasso Sea bacterioplankton. *Nature*, **345**:60–63.

Giovannoni, S. J. & Rappe, M. 2000. Evolution, diversity, and molecular ecology of marine prokaryotes. *In*: Kirchman, D. L. (ed.), Microbial Ecology of the Oceans. Wiley & Liss, Inc., New York. p. 47–84.

Grüntzig, V., Stres, B., Ayala del Río, H. L. & Tiedje, J. M. 2002. Improved protocol for T-RFLP by capillary electrophoresis. *rdp.cme.msu.edu*

Guillou, L., Eikrem, W., Chretiennot-Dinet, M. J., Le Gall, F., Massana, R., Romari, K., Pedros-Alio, C. & Vaulot, D. 2004. Diversity of picoplanktonic prasinophytes assessed by direct nuclear SSU rDNA sequencing of environmental samples and novel isolates retrieved from oceanic and coastal marine ecosystems. *Protist*, **155**:193–214.

Habura, A., Pawlowski, J., Hanes, S. D. & Bowser, S. S. 2004. Unexpected foraminiferal diversity revealed by small-subunit rDNA analysis of Antarctic sediment. *J. Eukaryot. Microbiol.*, **51**:173–179.

Hill, T. C. J., Walsh, K. A., Harris, J. A. & Moffett, B. F. 2003. Using ecological diversity measures with bacterial communities. *FEMS Microbiol. Ecol.*, **43**:1–11.

Holzmann, M., Habura, A., Giles, H., Bowser, S. S. & Pawlowski, J. 2003. Freshwater foraminiferans revealed by analysis of environmental DNA samples. *J. Eukaryot. Microbiol.*, **50**:135–139.

Hughes, J. B., Hellmann, J. J., Ricketts, T. H. & Bohannan, B. J. M. 2001. Counting the uncountable: statistical approaches to estimating microbial diversity. *Appl. Environ. Microbiol.*, **67**:4399–4406.

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

Kemp, P. F. & Aller, J. Y. 2004. Bacterial diversity in aquatic and other environments: what 16S rDNA libraries can tell us. *FEMS Microbiol. Ecol.*, **47**:161–177.

Kirchman, D. L., Yu, L. Y., Fuchs, B. M. & Amann, R. 2001. Structure of bacterial communities in aquatic systems as revealed by filter PCR. *Aquat. Microb. Ecol.*, **26**:13–22.

Leipe, D. D., Wainright, P. O., Gunderson, H. J., Porter, D., Patterson, D. J., Valois, F., Himmerich, S. & Sogin, M. L. 1994. The stramenopiles from a molecular perspective: 16S-like rRNA sequences from *Labyrinthuloides minuta* and *Cafeteria roenbergensis*. *Phycologia*, **33**: 369–377.

Lim, E. L., Dennett, M. R. & Caron, D. A. 1999. The ecology of *Paraphysomonas imperforata* based on studies employing oligonucleotide probe identification in coastal water samples and enrichment cultures. *Limnol. Oceanogr.*, **44**:37–51.

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.

López-García, 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. USA*, **100**:697–702.

López-García, P., Rodríguez-Valera, F., Pedrós-Alió, C. & Moreira, D. 2001. Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature*, **409**:603–607.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, Buchner, A., Lai, T., Steppi, S., Jobb, G., Forster, W., Brettske, I., Gerber, S., Ginhart, A. W., Gross, O., Grumann, S., Hermann, S., Jost, R., Konig, A., Liss, T., Lussmann, R., May, M., Nonhoff, B., Reichel, B., Strehlow, R., Stamatakis, A., Stuckmann, N., Vilbig, A., Lenke, M., Ludwig, T., Bode, A. & Schleifer, K.-H. 2004. ARB: a software environment for sequence data. *Nucleic Acids Res.*, **32**:1363–1371.

Marin, B., Palm, A., Klingberg, M. & Melkonian, M. 2003. Phylogeny and taxonomic revision of plastid-containing euglenophytes based on SSU rDNA sequence comparisons and synapomorphic signatures in the SSU rRNA secondary structure. *Protist*, **154**:99–145.

Marin, I., Aguilera, A., Reguera, B. & Abad, J. P. 2001. Preparation of DNA suitable for PCR amplification from fresh or fixed single dinoflagellate cells. *Biotechniques*, **30**:88.

Massana, R. & Jürgens, K. 2003. Composition and population dynamics of planktonic bacteria and bacterivorous flagellates in seawater chemostat cultures. *Aquat. Microb. Ecol.*, **32**:11–22.

Massana, R., Guillou, L., Diez, B. & Pedros-Alio, C. 2002. Unveiling the organisms behind novel eukaryotic ribosomal DNA sequences from the ocean. *Appl. Environ. Microbiol.*, **68**:4554–4558.

Massana, R., Castresana, J., Balague, V., Guillou, L., Romari, K., Groisillier, A., Valentin, K. & Pedros-Alio, C. 2004. Phylogenetic and ecological analysis of novel marine stramenopiles. *Appl. Environ. Microbiol.*, **70**:3528–3534.

McGrady-Steed, J., Harris, P. M. & Morin, P. J. 1997. Biodiversity regulates ecosystem predictability. *Nature*, **390**:162–165.

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

Modeo, L., Petroni, G., Rosati, G. & Montagnes, D. J. S. 2003. A multidisciplinary approach to describe protists: Redescriptions of *Novistrombidium testaceum anigstein* 1914 and *Strombidium inclinatum* Montagnes, Taylor, and Lynn 1990 (Ciliophora, Oligotrichia). *J. Eukaryot. Microbiol.*, **50**:175–189.

Moon-van der Staay, S. Y., De Wachter, R. & Vaulot, D. 2001. Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature*, **409**:607–610.

Naeem, S. & Li, S. B. 1997. Biodiversity enhances ecosystem reliability. *Nature*, **390**:507–509.

Pace, N. R. 1997. A molecular view of microbial diversity and the biosphere. *Science*, **276**:734–740.

Patterson, D. J. 1999. The diversity of eukaryotes. *Am. Nat.*, **154**:S96–S124.

Popels, L. C., Cary, S. C., Hutchins, D. A., Forbes, R., Pustizzi, F., Gobler, C. J. & Coyne, K. J. 2003. The use of quantitative polymerase chain reaction for the detection and enumeration of the harmful alga *Aureococcus anophagefferens* in environmental samples along the United States East Coast. *Limnol. Oceanogr. Methods*, **1**:92–102.

Rappe, M. S., Kemp, P. F. & Giovannoni, S. J. 1995. Chromophyte plastid 16s ribosomal RNA genes found in a clone library from Atlantic Ocean seawater. *J. Phycol.*, **31**:979–988.

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.

Rublee, P. A., Remington, D., Schaefer, E. & Marshall, M. M. 2005. Detection of the dinozoans *Pfiesteria piscicida* and *P. shumwayae*: a review of detection methods and geographic distribution. *J. Eukaryot. Microbiol.*, **52**:(in press).

Schmidt, T. M., Delong, E. F. & Pace, N. R. 1991. Analysis of a marine picoplankton community by 16S ribosomal-RNA gene cloning and sequencing. *J. Bacteriol.*, **173**:4371–4378.

Shannon, C. E. 1948. A Mathematical Theory of Communication. *Bell Syst. Tech. J.*, **27**:379–423, 423–656.

Shreeve, J. 2004. Craig Venter's epic voyage to redefine the origin of the species. *Wired*, **12**:104–113, 146–151.

Sieburth, J. M., Smetacek, V. & Lenz, J. 1978. Pelagic ecosystem structure: heterotrophic compartments of the plankton and their relationship to plankton size fractions. *Limnol. Oceanogr.*, **23**:1256–1263.

Sogin, M. L. 1989. Evolution of eukaryotic microorganisms and their small subunit ribosomal RNAs. *Am. Zool.*, **29**:487–499.

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.

Stoeck, T., Taylor, G. T. & Epstein, S. S. 2003. Novel eukaryotes from the permanently anoxic Cariaco Basin (Caribbean sea). *Appl. Environ. Microbiol.*, **69**:5656–5663.

Strom, S. L. 2000. Bacterivory: interactions between bacteria and their grazers. *In*: Kirchman, D. L. (ed.), Microbial Ecology of the Oceans. Wiley-Liss, Inc., New York. p. 351–386.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*, **25**:4876–4882.

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., Wu, D., Paulsen, I., Nelson, K. E., Nelson, W., Fouts, D. E., Levy, S., Knap, A. H., Lomas, M. W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y.-H. & Smith, H. O. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. *Science*, **304**:66–74.

Weekers, P. H. H., Gast, R. J., Fuerst, P. A. & Byers, T. J. 1994. Sequence variations in small-subunit ribosomal-RNAs of *Hartmannella vermiformis* and their phylogenetic implications. *Mol. Biol. Evol.*, **11**:684–690.

Woese, C. R., Kandler, O. & Wheelis, M. L. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. USA*, **87**:4576–4579.

Zobell, C. E. 1943. The effect of solid surfaces upon bacterial activity. *J. Bacteriol.*, **46**:39–56.

Received: 01/24/05; accepted: 01/24/05