Low Evolutionary Diversification in a Widespread and Abundant Uncultured Protist (MAST-4)

Raquel Rodríguez-Martínez,*[,1] Gabrielle Rocap,[2] Ramiro Logares,[1] Sarah Romac,[3] and Ramon Massana*[,1]

[1]Institut de Ciències del Mar, Consejo Superior de Investigaciones Científicas (CSIC), Barcelona, Spain

[2]School of Oceanography, University of Washington

[3]Station Biologique de Roscoff, Centre National de la Recherche Scientifique (CNRS) et Universite Pierre et Marie Curie (UPMC), Roscoff, France

***Corresponding author:** E-mail: raquelr@icm.csic.es; ramonm@icm.csic.es.

Associate editor: Douglas Crawford

Abstract

Recent culture-independent studies of marine planktonic protists have unveiled a large diversity at all phylogenetic scales and the existence of novel groups. MAST-4 represents one of these novel uncultured lineages, and it is composed of small (~2 μm) bacterivorous eukaryotes that are widely distributed in marine systems. MAST-4 accounts for a significant fraction of the marine heterotrophic flagellates at the global level, playing key roles in the marine ecological network. In this study, we investigated the diversity of MAST-4, aiming to assess its limits and structure. Using ribosomal DNA (rDNA) sequences obtained in this study (both pyrosequencing reads and clones with large rDNA operon coverage), complemented with GenBank sequences, we show that MAST-4 is composed of only five main clades, which are well supported by small subunit and large subunit phylogenies. The differences in the conserved regions of the internal transcribed spacers 1 and 2 (ITS1 and ITS2) secondary structures strongly suggest that these five clades are different biological species. Based on intraclade divergence, ITS secondary structures and comparisons of ITS1 and ITS2 trees, we did not find evidence of more than one species within clade A, whereas as many as three species might be present within other clades. Overall, the genetic divergence of MAST-4 was surprisingly low for an organism with a global population size estimated to be around 10^{24}, indicating a very low evolutionary diversification within the group.

Key words: MAST-4, low evolutionary diversification, uncultured protist, pyrosequencing, ITS secondary structure.

Research article

Introduction

Microbes have vital roles for the functioning of the biosphere (Falkowski et al. 2008), but currently, we are far from having acceptable estimates of their diversity. Furthermore, it is unclear how microbial diversity is distributed in space and time, and how diversity ranks are translated into ecologically meaningful interactions or processes. The marine protists of very small size, the picoeukaryotes, are among the underexplored microbes with large ecological importance (Massana 2011). Picoeukaryotes have key ecological roles in the oceans as primary producers, bacterial grazers, or parasites. They are found in all planktonic marine samples at concentrations ranging between 10^3 and 10^4 cells ml^{-1}. During the last 10 years, molecular tools based on sequencing environmental 18S ribosomal DNA (rDNA) genes have revealed a wide diversity of microeukaryote assemblages as well as the existence of novel and uncultured lineages (Díez et al. 2001; López-García et al. 2001; Moon-van der Staay et al. 2001). Still, most of this diversity remains poorly known.

The assignation of this novel and uncultured diversity to taxonomic groups is a challenging task. An approach to address this issue is to explore the correspondence between genetic divergence and species limits using cultured strains and then use that data as a proxy to investigate species limits in uncultured strains. Studies combining molecular and morphological data have been done within different taxonomic groups, such as prasinophytes (Slapeta et al. 2006), prymnesiophytes (Lange et al. 2002), diatoms (Amato et al. 2007; Evans et al. 2007; Casteleyn et al. 2008; Rynearson et al. 2009; Sorhannus et al. 2010), and dinoflagellates (Montresor et al. 2003; Litaker et al. 2007; Lowe et al. 2010). Gene markers used in the mentioned studies generally involve the 18S rDNA and other more variable genes (such as *rbcL* or *cox1*) since the former may be too conserved to differentiate among related but different species (Edvardsen et al. 2000; Logares et al. 2007). For uncultured protists detected in 18S rDNA surveys, the obvious loci for increasing phylogenetic resolution are the contiguous internal transcribed spacer (ITS) regions (ITS1 and ITS2). The above-mentioned functional genes, proposed as more robust phylogenetic markers (Álvarez and Wendel 2003), are currently inaccessible for uncultured microorganisms. ITS regions are noncoding loci that display high sequence variability but also key functionally constrained positions since transcripts need to fold into a secondary structure to permit their own splicing and the correct processing of the rDNA genes (Schlötterer et al. 1994; Côté et al. 2002). They have been proposed as the best tool for barcoding in diatoms (Moniz and Kaczmarska 2010) and are useful for species and genus phylogenetic inferences (Coleman 2003).

The secondary structure of the ITS2 region has been used for delimiting biological species. Compensatory base changes (CBCs) in particular regions of the secondary structure have been associated with sexual incompatibility (Coleman 2007, 2009). Taxa exhibiting at least one CBC in these conserved regions most likely belong to different biological species (Amato et al. 2007). Significant progress has been made in identifying such relevant positions in Volvocaceae, Haliotis, and Fagales (Coleman 2000, 2003; Coleman and Vacquier 2002; Müller et al. 2007). In addition, this hypothesis has been subjected to a large-scale testing using the ITS2 database containing 100,000 secondary structures (Schultz et al. 2006; Selig et al. 2007) and has been supported in 93% of the cases (Müller et al. 2007). However, this is a one-way diagnostic; a lack of CBCs does not mean that organisms are members of the same species.

In this study, we investigate an important and poorly known uncultured picoeukaryote group, the MAST-4 (Massana et al. 2004). This protist group is widespread in surface marine waters (except polar systems), where it represents approximately 9% of heterotrophic flagellates (Massana, Terrado, et al. 2006; Rodríguez-Martínez et al. 2009). Although MAST-4 remains uncultured, it is easily detected in environmental samples using molecular tools. So far, only the 18S rDNA of MAST-4 has been sequenced. To understand the genetic structure and evolutionary patterns of this uncultured model picoeukaryote, we sequenced a large fragment of the rDNA operon, including the ITS region and the beginning of the 28S (using Sanger sequencing) as well as the V4 region of the 18S (454 pyrosequencing). We have also compiled and analyzed all publicly available MAST-4 18S rDNA sequences. The emerging scenario is that despite being hugely abundant and widely distributed, this lineage has experienced a limited evolutionary diversification. A more detailed study of the biogeography of the group will appear elsewhere (Rodríguez-Martínez R, unpublished data).

Materials and Methods

Compilation of Published MAST-4 Sequences

BLAST (basic local alignment search tool) searches against NCBI-nr were done using as seeds different regions (1–500, 501–1000, and 1001–1700) of the published MAST-4 18S rDNA sequences ME1.19, ME1.20 (Díez et al. 2001), UEPACCp4, UEPAC05Cp2 (Worden 2006), and SSRPD78 (Not et al. 2007). Best hits (sorted in decreasing order by identity) were selected until a sequence classified to another group appeared. The retrieved 134 sequences were screened to remove chimeras, identical sequences from the same study, and sequences that did not cover the V4–V5 regions, leaving 72 sequences that aligned at least ~550 bp. Seven partial GenBank sequences, of clones from our own libraries, were completely sequenced. This resulted in 17 complete MAST-4 sequences ("small subunit [SSU]–complete" data set). The remaining 55 partial sequences (between 461 and 1,266 bp) formed the "SSU-partial" data set.

Retrieval of MAST-4 Using 454 Pyrosequencing

Seawater samples were collected through the BioMarKs consortium (http://www.biomarks.org/) in several European coastal stations (offshore Oslo, Naples, Blanes, Roscoff, Gijon, and Varna) with Niskin bottles attached to a conductivity, temperature, and depth rosette at surface and deep chlorophyll maximum depths. Water samples were prefiltered through 20 μm. Afterward, they were sequentially filtered through 3 and 0.8 μm 142 mm polycarbonate filters. Filters were flash frozen and stored at −80 °C. Total DNA and RNA were extracted simultaneously from the same filter using the NucleoSpin RNA L kit (Macherey-Nagel) and quantified using a Nanodrop ND-1000 Spectrophotometer. Extract quality was checked on a 1.5% agarose gel. To remove contaminating DNA from RNA, we used the TurboDNA kit (Ambion). Extracted RNA was immediately reverse transcribed to DNA using the RT Superscript III_random primers kit (Invitrogen). The universal primers TAReuk454FWD1 and TAReukREV3 were used to amplify the V4 region (~380 bp) of eukaryotic 18S rDNA (Stoeck et al. 2010). The primers were adapted for 454 using the manufacturers specifications and had the configuration A adapter Tag (8 bp)–forward primer and B adapter–reverse primer. Polymerase chain reactions (PCRs) were performed in 25 μl and consisted 1× MasterMix Phusion High-Fidelity DNA Polymerase (Finnzymes), 0.35 μM of each primer, and 3% dimethyl sulfoxide. We added a total of 5 ng of template DNA/cDNA to each PCR reaction. PCR reactions consisted of an initial denaturation step at 98 °C during 30 s, followed by 10 cycles of 10 s at 98 °C, 30 s at 53 °C, and 30 s at 72 °C, and afterward by 15 cycles of 10 s at 98 °C, 30 s at 48 °C, and 30 s at 72 °C. Amplicons were checked in a 1.5% agarose gel for successful amplification. Triplicate amplicons were pooled and purified using the NucleoSpin Extract II (Macherey-Nagel). Purified amplicons were eluted in 30 μl of elution buffer and quantified again using a Nanodrop ND-1000 Spectrophotometer. The total final amount of pooled amplicons for 454 tag sequencing was approximately of 5 μg. Amplicon sequencing was carried out on a 454 GS FLX Titanium system (454 Life Sciences, USA) installed at Genoscope (http://www.genoscope.cns.fr/spip/, France).

Only reads having exact forward and reverse primers and an estimated error of ≤0.1% were kept (682,390 reads) and were annotated using a custom made and curated 18S rDNA database (Guillou L, unpublished data). Sequences with the MAST-4 as the closest group (similarity >90%) were extracted (2,808 reads). Identical reads were removed with Mothur (Schloss et al. 2009) and then clustered at 0.0049 distance, resulting in 169 unique sequences. Subsequently, 81 chimeras were removed with the Chimera Slayer algorithm (Haas et al. 2011) as implemented in Mothur, using a custom-made protist database as a template. Sixteen remaining chimeras were removed manually, after partial sequence BLASTs against NCBI-nr. The final 72 sequences of ~380 bp formed the "SSU-pyrosequencing" data set.

Clone Libraries Covering the 18S to 28S rDNA Regions

Offshore surface samples were selected from separate oceanographic cruises in the Indian Ocean (IND70),

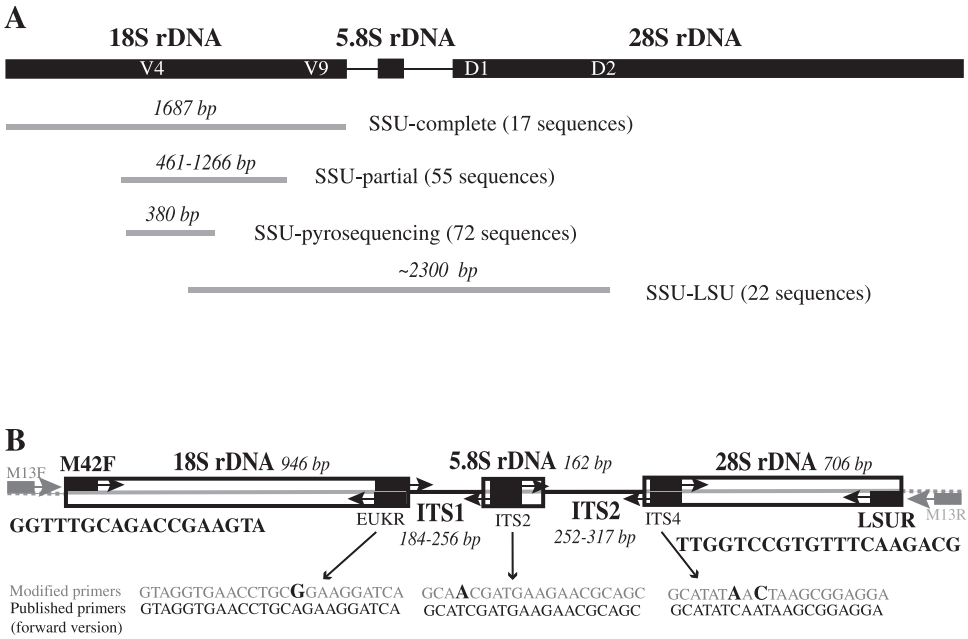

FIG. 1. Map of the rDNA operon showing the covered region of each sequence data set (*A*) and a detailed diagram of the ~2,300 bp MAST-4 rDNA amplicons (*B*). For this last data set, the positions and sequences of primers used are presented.

Sargasso Sea (BE3), North Pacific (WE7), and Mediterranean Sea (BL43, taken on August 2004). These sites correspond to stations INO3, ATL7, PAC1, and MED as shown in Massana, Terrado, et al. (2006). The 0.2–3 μm microbial fraction of surface seawater was collected by peristaltic filtration. A fifth sample was selected (OA4), derived from the peak of heterotrophic flagellates in an unamended incubation from the MED station (March 2006) processed as in Massana, Guillou, et al. (2006). DNA extraction was done using enzymatic and sodium dodecyl sulfate digestion plus phenol purification (Massana et al. 2000).

PCR amplification was done with the MAST-4-specific primer M42F (5′-GGTTTGCAGACCGAAGTA-3′) located in the 18S rDNA (after the V4 region) and the universal eukaryotic primer LSUR (5′-TTGGTCCGTGTTTCAAGACG-3′) located in the 28S rDNA (in the middle of the D2 region) (Jerome and Lynn 1996). Primer M42F is the reverse sequence of the FISH probe NS4 (Massana et al. 2002) and has a good specificity for MAST-4 (Massana, Terrado, et al. 2006). Primer LSUR matches 183 of 187 stramenopile large subunit (LSU) sequences extracted from SILVA database (Pruesse et al. 2007). Primers were checked for formation of primer dimers, GC content, and theoretical melting temperature in the website www.operon.com, using the Oligo Analysis & Plotting Tool. This primer set gave an amplicon size of ~2,300 bp covering the end of the SSU gene, the whole ITS region (ITS1-5.8S-ITS2), and the beginning of the LSU gene (fig. 1).

The PCR mixture (25 μl) contained 1 μl of DNA template, 0.5 μM of each primer, 200 μM of each dNTP, 3 mM $MgCl_2$, 1.25 units of a proofreading *Taq* polymerase (ACCUZYME), and the enzyme buffer. PCR cycling, carried out in an MJ Research thermal cycler, was initial denaturation at 94 °C for 5 min, 30 cycles with denaturation at 94 °C for 1 min, annealing at 60 °C for 1 min, and extension at 72 °C for 3 min and a final extension at 72 °C for 10 min. We added a reconditioning PCR step to eliminate heteroduplexes from mixed-template PCR products (Thompson et al. 2002). The PCR reaction was diluted 5-fold into fresh reaction mixture and cycled three times as above. We tested the $MgCl_2$ concentration (from 1.5 to 3 mM) and the annealing temperature (from 57 to 64 °C) and chose the more stringent conditions giving the expected band. The PCR product from four parallel reactions per sample was pooled and reduced to 25 μl by ethanol precipitation or vacuum concentration and run in a 1% agarose gel electrophoresis. Bands of 2,000–3,000 bp were cut and purified with the QIAquick Gel Extraction kit (QIAGEN). We added 3′ A-overhangs to the final PCR product and cloned it using the TOPO-TA cloning kit (Invitrogen) with the vector (pCR4) following manufacturer's recommendations. Putative positive colonies were picked and transferred to a new Luria–Bertani (LB) plate and finally into LB-glycerol solution for frozen stocks (−80 °C).

Presence of correct insert was checked by PCR reamplification with vector primers M13F and M13R using a small aliquot of culture as template. Amplicons with the right insert size were sequenced in both directions at the Macrogen sequencing service (Korea) with eight primers (fig. 1). After inspecting the first sequences, we modified primers EUKR, ITS2, and ITS4 for a perfect match with MAST-4 sequences (fig. 1). Chromatograms were examined with 4Peaks (A. Griekspoor and T. Groothuis, http://www.mekentosj.com), and sequences for each clone were assembled with Geneious (Drummond et al. 2010), which also allows careful inspection of chromatograms and sequence editing. A total of 22 sequences from clone libraries were

used for subsequent analyses ("SSU-LSU" data set). These sequences were aligned with MAFFT v6.853 (Katoh and Toh 2008) with the E-INS-I algorithm, and the alignment was inspected visually. Boundaries of rDNA genes were determined by comparison with published reference sequences belonging to closely related organisms, resulting in five separate DNA regions: 3′ end of the SSU gene (946 bp), ITS1 (184–256 bp), 5.8S gene (162 bp), ITS2 (252–317 bp), and 5′ beginning of the LSU gene (706 bp). Sequences have been deposited in GenBank under accession numbers JN836289–JN836310.

Sequence Analyses

Sequences from the SSU-complete data set were aligned together with a MAST-7 outgroup using MAFFT as specified above. This alignment (1688 positions) was used as a skeleton, and shorter sequences from GenBank (SSU-partial data set) or from pyrosequencing (SSU-pyrosequencing data set) were incorporated into it using the "-add" option of MAFFT. Alignments with 5.8S and 28S regions were done using *Phytophthora infestans* as outgroup (GenBank accession numbers HQ191489 and EU079637, respectively). Due to the large sequence variability in ITS regions, ITS1 and ITS2 alignments were done separately for each SSU-defined clade, using secondary structure models for alignment improvement (Rocap et al. 2002; Wang et al. 2007; Tippery and Les 2008). All these alignments were used to calculate sequence divergences (uncorrected pairwise distances) using Mothur (Schloss et al. 2009).

Maximum likelihood (ML) phylogenetic trees were reconstructed using RAxML v7.0.4 MPI version (Stamatakis 2006), using the General Time Reversible model of nucleotide substitution and a Gamma distributed rate of variation across sites (GTR+G). As suggested in RaxML, we did not estimate the proportion of invariable sites, and missing data were not considered (i.e., treated as missing data). The shape parameter (α) of the gamma distribution was estimated from the data set using default options. Phylogenies were reconstructed at both the University of Oslo Bioportal (www.bioportal.uio.no) and the Instituto Astrofísico de Canarias (IAC) computer cluster. One thousand alternative ML trees were run, and the tree with the best likelihood was selected and visualized in FigTree v1.3.1 (Rambaut 2009) or iTOL (Letunic and Bork 2007). Bootstrap analyses were run with 1,000 pseudoreplicates, and a consensus tree was constructed with MrBayes (Huelsenbeck and Ronquist 2001).

ITS1 and ITS2 Secondary Structures

ITS1 and ITS2 sequences extracted from the SSU-LSU data set were folded in mFOLD (Zuker 2003), which generates multiple possible secondary structures. We used default settings for a linear molecule with a folding temperature fixed at 37 °C and 1 M NaCl with no divalent ions for ionic conditions. The best conformation for each sequence was the one that possessed the previously defined ITS hallmarks and was also similar between related clones. This generally coincided with the minimum free energy configuration. For ITS2 models, we searched for the familiar four-helix domain seen in eukaryotic taxa, such as green algae and flowering plants (Mai and Coleman 1997), dinoflagellates (Gottschling 2004), and metazoans (Joseph et al. 1999; Coleman and Vacquier 2002; Müller et al. 2007; Wiemers et al. 2009). The core structure and hallmarks for the ITS1 secondary structure are less clear (see Discussion). Exported secondary structures in Vienna format (http://www.tbi.univie.ac.at/~ivo/RNA/) were aligned and visualized as a consensus of each clade with 4SALE version 1.5 (Seibel et al. 2006). Structural models were further analyzed for the presence of CBCs (e.g., a change of paired G-C into paired A-U) in conserved regions (Gutell et al. 1994; Coleman et al. 1998). We used the models proposed by Coleman to identify the ITS2 conserved regions having a biological meaning (Coleman 2003, 2007, 2009).

Results

Low Diversity within the MAST-4 18S rDNA Gene

The phylogenetic tree with the distinct MAST-4 18S rDNA sequences retrieved from our thorough GenBank search (SSU-complete plus SSU-partial data sets) displayed the complete MAST-4 variability published so far. MAST-4 diversity was limited to only five clades (A–E), each one containing at least two complete sequences and being well supported (except clade B) by bootstrap values above 80% (fig. 2). Only clone IND31.115 (Massana, Terrado, et al. 2006) did not belong to a given clade. The intraclade sequence divergence (calculated in the SSU-complete data set) was typically below 0.010, whereas among clades, the average divergence was 0.030 (table 1), with a maximum of 0.044. In addition, a BLAST search of MAST-4 sequences against NCBI-nr displayed a maximum of 91% similarity to the closest outgroup sequence, which belonged to MAST-7 or MAST-8. Sequences with intermediate similarity (i.e., between 91% and 96%) were chimeras.

High-throughput sequencing approaches allow deeper sampling of environmental diversity than traditional cloning and Sanger sequencing. In order to determine whether additional MAST-4 clades were present in the marine plankton, we analyzed 454 reads (~380 bp) obtained from several coastal locations around Europe using eukaryotic universal primers. After processing an initial set of 2,808 MAST-4 sequences, the 72 distinct sequences of the SSU-pyrosequencing data set were used for phylogenetic reconstructions together with a subset of the GenBank sequences (31 remaining sequences after clustering the SSU-complete and SSU-partial data sets at a 0.0049 distance) (fig. 3). Pyrosequences distributed among the five clades reported before, and, most interestingly, no additional clades appeared. The sequence IND31.115 still remained alone.

Analysis of Other rDNA Regions Support Five Main Clades

We obtained good quality sequences of ~2,300 bp (SSU-LSU data set) for 22 clones derived from four

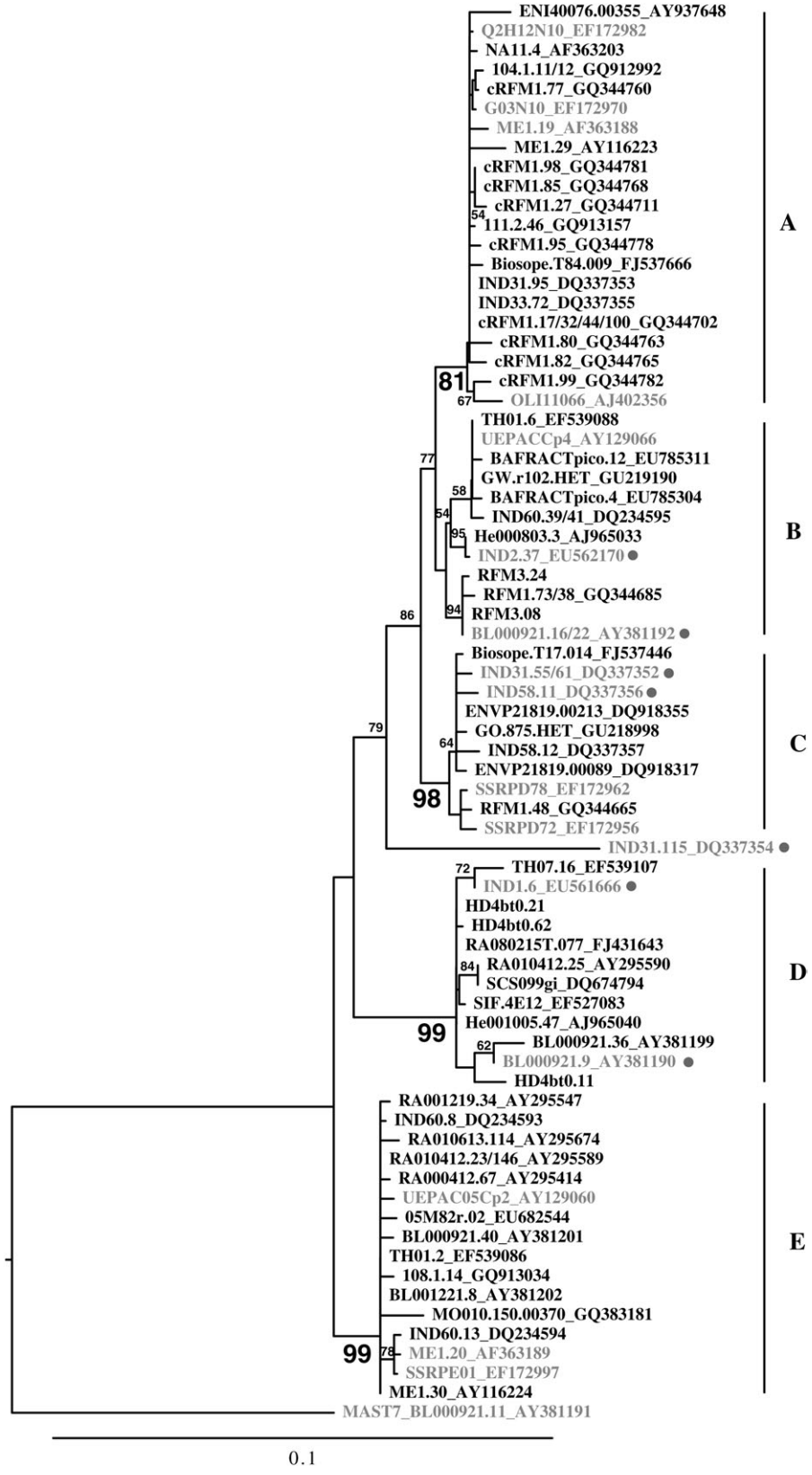

Fig. 2. ML phylogenetic tree of complete (gray) and partial (black) 18S rDNA GenBank sequences affiliating to MAST-4, showing the five main clades (labeled A–E) and their bootstrap support. Complete 18S rDNA sequenced in this study is indicated with a gray dot.

oceanographic regions, the Indian Ocean (six clones), the North Pacific (four clones), the Sargasso Sea (three clones), and the Mediterranean Sea (nine clones, four of them from an unamended enrichment). These were separated into the three genes and the two internal spacers of the rDNA operon for more exhaustive phylogenetic analyses (fig. 4). The

Table 1. Sequence Divergence (uncorrected p-distance) within and among MAST-4 Clades, Shown as Average (minimum–maximum).

	Data Set SSU-Complete		Data Set SSU-LSU						
Clade	*n*	18S rDNA	*n*	18S rDNA	5.8S rDNA	28S rDNA	ITS1	ITS2	ITS
A	4	0.007 (0.002–0.012)	7	0.001 (0–0.002)	0	0.005 (0.001–0.010)	0.044 (0.005–0.071)	0.036 (0.004–0.057)	0.031 (0.011–0.050)
B	3	0.008 (0.008–0.009)	2	0.001	0	0.021	0.204	0.197	0.163
C	4	0.008 (0.004–0.011)	6	0.005 (0–0.010)	0	0.020 (0–0.033)	0.104 (0.005–0.156)	0.121 (0–0.195)	0.078 (0.002–0.117)
D	2	0.011	1		0				
E	3	0.006 (0.002–0.008)	6	0	0	0.003 (0–0.006)	0.020 (0.004–0.032)	0.021 (0–0.031)	0.016 (0.010–0.013)
Interclade	16	0.030 (0.011–0.044)	22	0.027 (0.010–0.047)	0.024 (0–0.049)	0.161 (0.074–0.218)	0.420 (0.200–0.559)	0.418 (0.298–0.562)	0.270 (0.185–0.416)

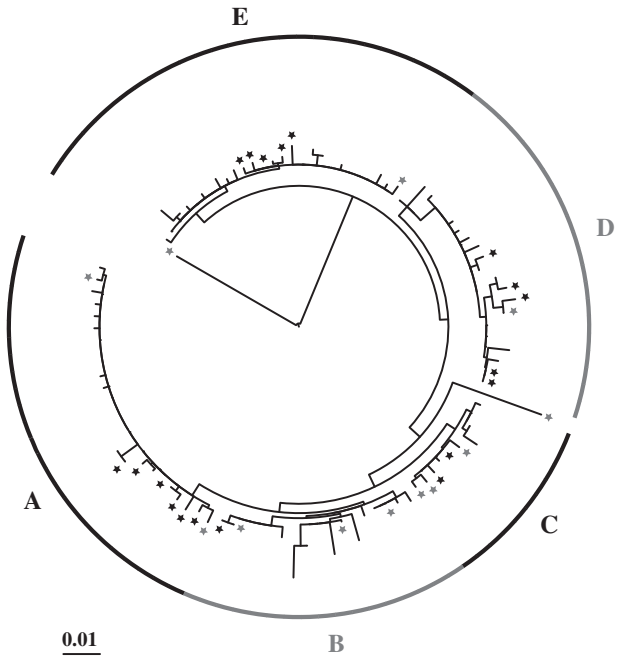

Fig. 3. ML phylogenetic tree of MAST-4 18S rDNA with a subset of GenBank sequences shown in figure 2 (indicated with stars) plus additional unique pyrosequences obtained in our study.

18S rDNA tree (fig. 4*A*) was consistent with that shown in figure 2. The 28S rDNA tree displayed the same five clades as before, but here, the clades appeared better resolved and separated with longer phylogenetic distances (fig. 4*D*). The 5.8S was the least informative of the three genes, since all sequences within clades A–C were identical (fig. 4*C*). LSU rDNA sequences were also used to place MAST-4 within the stramenopiles (tree not shown) and revealed a position consistent with previous 18S rDNA trees (Massana et al. 2004). Giving the variability of ITS regions, we did not attempt to construct a tree with all sequences. Instead, ITS1 and ITS2 trees were done separately for each clade and used to contrast their topology (the order and relative branching of the different clones) (fig. 4*B*). Clade C exhibited a consistent topology, with clones branching in the same way in both ITS trees, whereas clones within clade A were widely mixed when comparing both trees. Clade E would be an intermediate case of the two previous examples.

The averaged 18S rDNA sequence distance between clades for the partial sequences in the SSU-LSU data set was 0.027, very similar to the distance estimated with complete sequences (table 1). The additional resolution provided by 28S rDNA partial sequences was clear, as the interclade distance using this gene was 0.161. Interclade distances using ITS sequences were much higher: 0.420 for ITS1, 0.418 for ITS2 (0.270 for the whole ITS region), although these values were less certain due to the inherent difficulty in aligning ITS regions. It was also clear that not all clades had a comparable intraclade variability, which was low in clades A and E and significantly larger in clades B and C. In fact, the maximal intraclade ITS distance in clade B or C was similar to the minimal interclade distance.

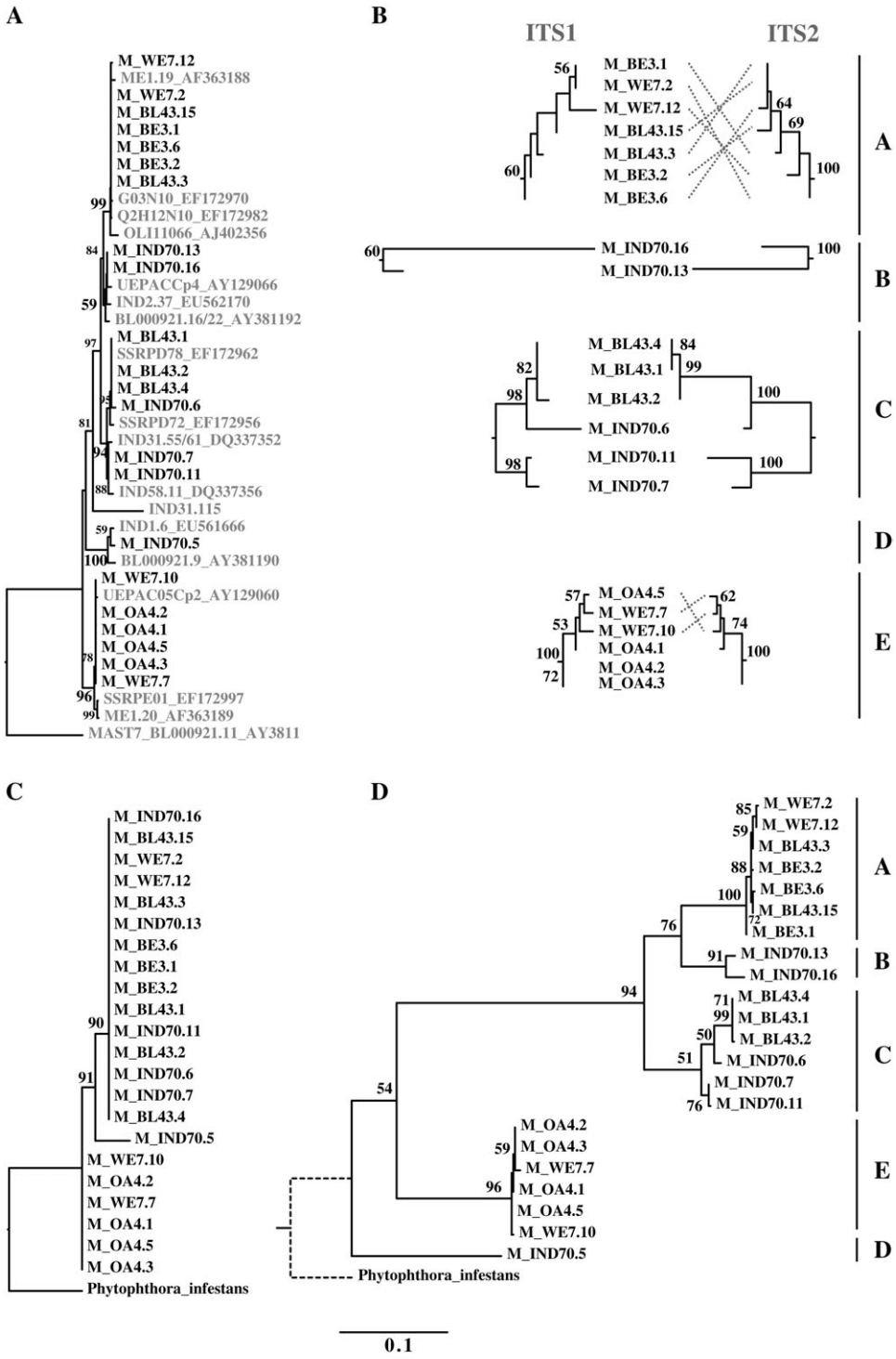

Fig. 4. ML phylogenetic trees constructed with 22 MAST-4 clones considering the partial 18S rDNA sequences (*A*), contrasted ITS1 and ITS2 sequences (*B*), complete 5.8S rDNA sequences (*C*), and partial 28S rDNA sequences (*D*). Tree in *A* also includes complete 18S rDNA sequences (in gray). The scale bar applies to all trees and indicates substitutions per position. Bootstrap values above 50 are shown.

Exploring Intraclade Diversity Using ITS Secondary Structures

ITS secondary structures allow differentiating between constrained or neutral changes in ITS sequences. ITS2 secondary structures of MAST-4 contained the four-helix domains known in many eukaryotic taxa (fig. 5). Helix II included the universal pyrimidine–pyrimidine (U–U) mismatch and had an initial stem of five base pairs that was conserved within each clade, with the exception of one position in clade C (fig. 5, helix II). Differences between clades in this section were supported by compensatory base changes (CBCs). In the 5′-side of helix III, the most conserved part in the ITS2, clades A and E exhibited 30 conserved positions, whereas 24 and 22 positions were conserved in clades B and C, respectively (fig. 5, helix III). The TGGT motif in the middle of the helix III conserved region (Coleman 2007) was observed

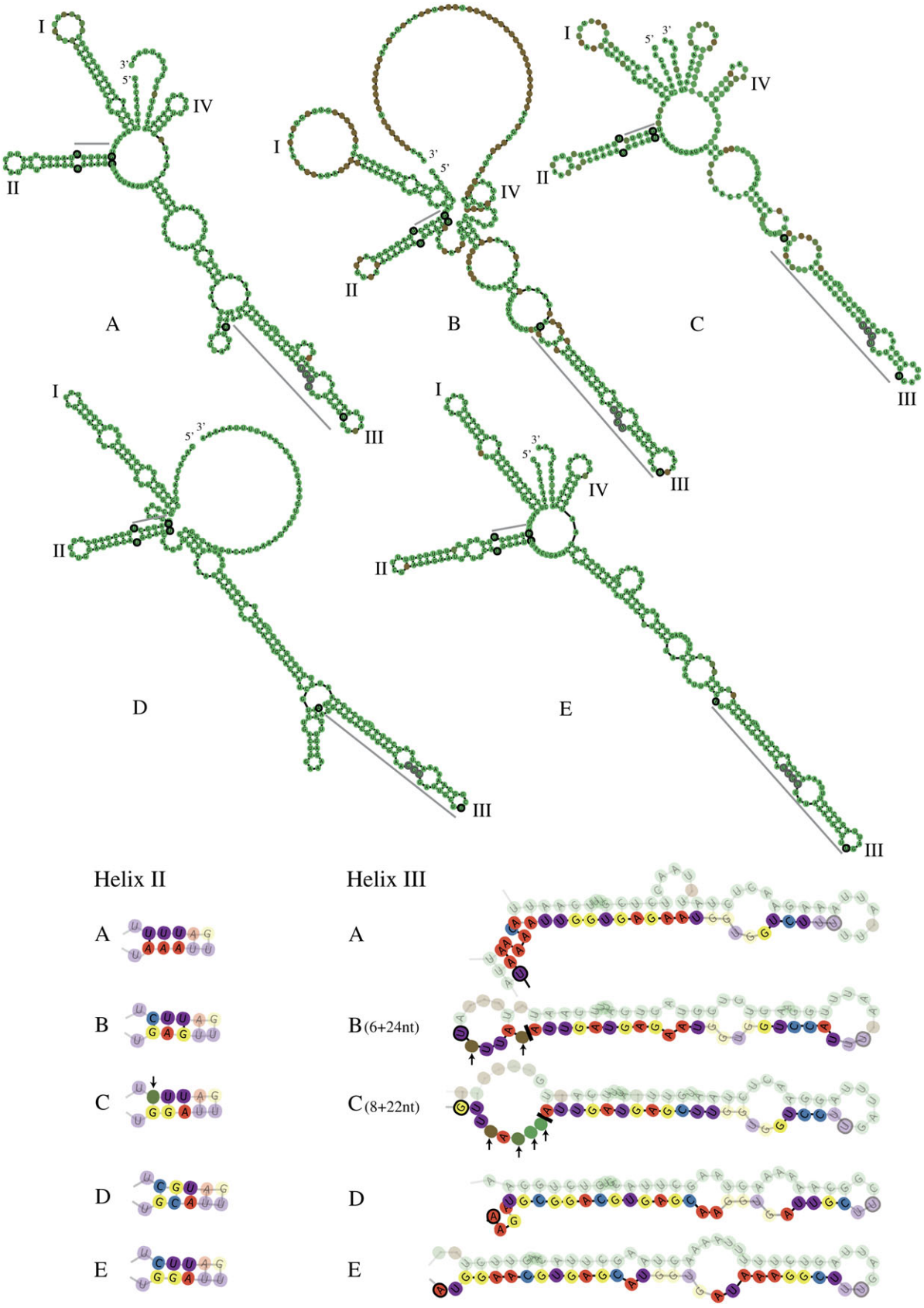

FIG. 5. Consensus ITS2 secondary structures for each of the five MAST-4 clades (*A-E*), showing three main helices (I–III). Base pairs highly conserved within each clade are shown in green; variable positions are shown in brown. Nucleotides with gray circles represent the UGGU motif. Details of helices II and III are represented at the bottom of the figure, highlighting in bright color the conserved nucleotides within each clade that differ among clades. Nucleotides conserved in all sequences appear in weak color. Black lines in helix III delimit the longest position (if this is smaller than 30 nt) of this conserved region until polymorphism appears. Arrows point to polymorphisms (in conserved regions) within a clade.

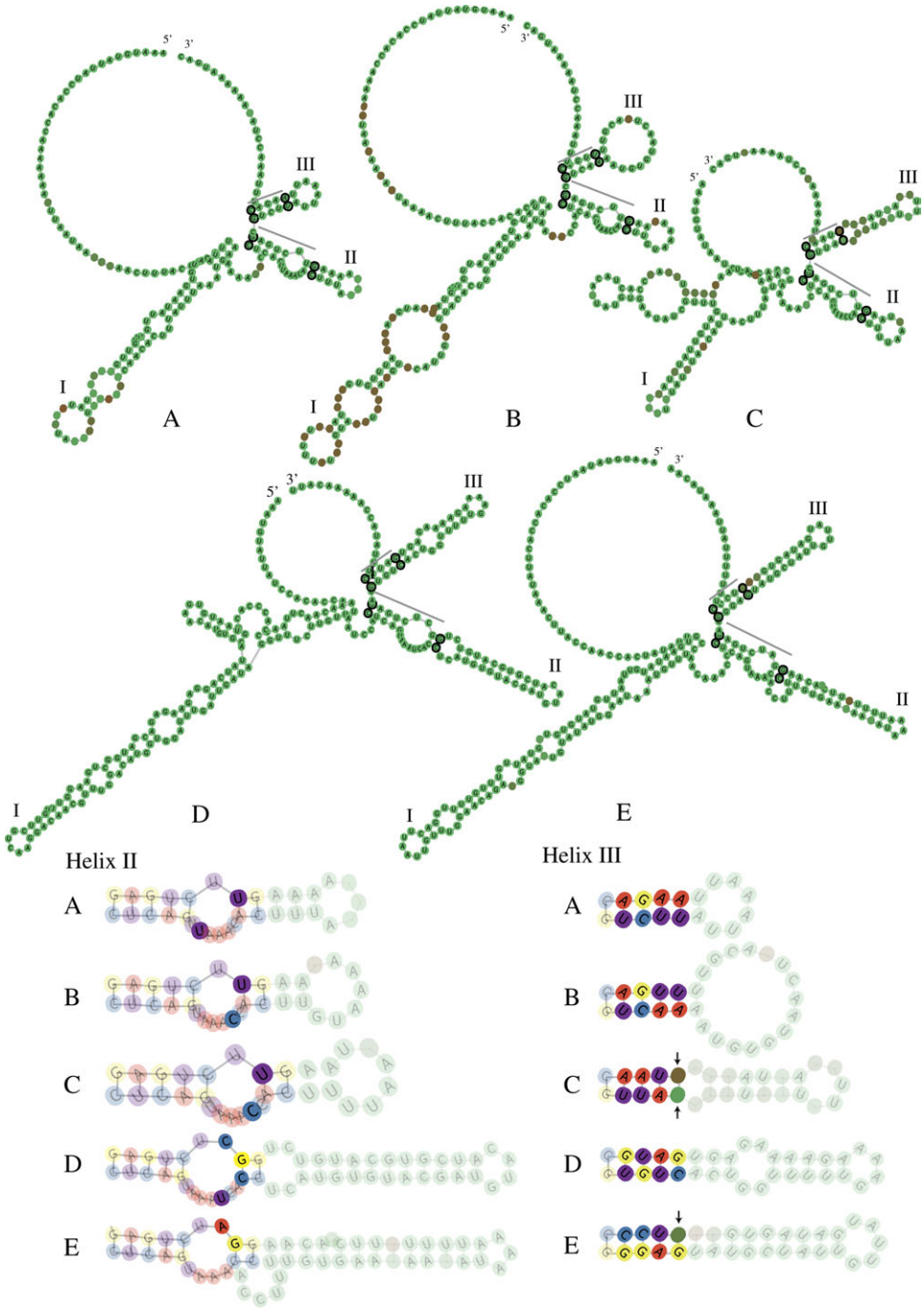

Fig. 6. Consensus ITS1 secondary structure for each of the five MAST-4 clades (*A-E*), showing three main helices (I–III). Base pairs highly conserved within each clade are shown in green; variable positions are shown in brown. Details of helices II and III are represented at the bottom of the figure, highlighting in bright color the conserved nucleotides within each clade that differ among clades. Nucleotides conserved in all sequences appear in weak color. Arrows point to polymorphisms (in conserved regions) within a clade.

in all cases except for clade D that had AGGT (fig. 5). As in other taxa studied, helices I and IV were very variable (in some cases, helix IV did not appear in the intraclade consensus structure). Overall, we identified helices II and III as regions in the ITS2 that were conserved within clades and differed among clades (with only a few position exceptions).

For ITS1 sequences, we found a common core secondary structure with three helices. Helix I was the most variable, helix II was the most conserved, even for the primary sequence, and the short helix III had a conserved secondary structure (fig. 6). In helix II, there was an initial stem with 5 bp common for all sequences, a loop with positions differing among clades, a stem conserved in clades A–C but very different in clades D and E and a final nonconserved loop. In Helix III, there was a basal stem of four conserved base pairs that started with CG in all clades and was followed by conserved base pairs within clades but differing

with CBCs between them. If we expanded it to the fifth base pair, polymorphisms appeared within clades C (three groups supported by CBCs and hemiCBCs) and E (two groups with a hemiCBC). As found in ITS2 secondary structures, we identified an ITS1 conserved region (in helix III) that was conserved within a clade and varied among clades.

Discussion

The group MAST-4 accounts for 9% of heterotrophic flagellates in all the oceans (except in polar waters), with an average abundance of 131 cells ml^{-1} (Massana, Terrado, et al. 2006). Its global population size is estimated to be about 10^{24} cells, 2,500 billion times the number of total birds in the world (Gaston and Blackburn 1997), which are divided in around 8,600 species (May and Beverton 1990). In contrast, the MAST-4 18S rDNA phylogeny reveals only five main clades, each one exhibiting low sequence divergence and conservation in specific parts of the ITS secondary structures. This lineage appears as a well-supported discrete group in 18S rDNA phylogenies, and the closest known outgroup sequences are only 91% similar. In addition, the maximal 18S rDNA sequence divergence within MAST-4 is 0.044 (table 1), a very low value as compared with other protist groups (Pernice M, personal communication). Pyrosequencing added more than one order of magnitude of sequences (with respect to currently available GenBank sequences) and confirmed the low MAST-4 diversity, as all found sequences affiliated to the five known clades. It is important to note that the MAST-4 pyrosequences were retrieved from a vast environmental protist survey. Only clone IND31.115 did not match any new pyrosequences, and we confirmed that this divergent sequence was not a chimera or a sequencing error. Instead, this sequence could be a pseudogene (Thornhill et al. 2007). Overall, it is remarkable that such a widespread and abundant protist group appears to have experienced so little evolutionary diversification. A similar scenario of low diversity and cosmopolitan distribution seems to exist in other picoeukaryotes, such as the prasinophyte *Micromonas* (Slapeta et al. 2006).

For a better interpretation of the detected genetic variability in this uncultured group, we sequenced for the first time its complete ITS region, since it has been observed that the ITS2 secondary structure can help in delimiting species (Coleman 2007, 2009). In particular, strains exhibiting at least one CBC in the conserved nucleotides of helices II and III were shown to belong to different biological species. Still the presence of a hemiCBC (one sided) could allow some weak degree of interbreeding (Coleman 2009). For instance, two *Pseudo-nitzschia* strains differing by three hemiCBCs produced zygotes but never gave viable offspring, thus being considered as separate species (Amato et al. 2007). We looked for these conserved regions among the consensus ITS2 secondary structures of each MAST-4 clade. The 5 bp in helix II are identical within each clade but differ by CBCs among clades, suggesting that each clade is a separate biological species. The only exception is clade C, which has a hemiCBC between several clones, so it could include at least two species. With respect to the conserved region in helix III, different size criteria have been invoked in the literature to correlate with sexual incompatibility, 18 positions (Coleman 2007), ~20 (Coleman 2003) or 30 (Coleman 2009). In clades A and E, we identified a region of 30 bp conserved within clades and differing among clades, whereas clades B and C exhibited a slightly shorter conserved fragment. So, using the most restrictive criteria of 30 bp, these two latter clades would include more than one species.

Whereas the ITS2 secondary structure has been widely investigated, the ITS1 still lacks of a universal core secondary structure model. In the eukaryotic taxa examined so far, it is typically represented by an open loop containing multiple double-stranded helices (Coleman et al. 1998; Gottschling et al. 2001; Goertzen 2003; Gottschling 2004; Hoshina 2010; Thornhill and Lord 2010). However, it seems that the generally accepted hypervariability of ITS1 was overestimated (Itskovich et al. 2008), and this region can be used, for example, to define species complex groups on the basis of the conserved helix II motif (Bridge et al. 2008). Similarly to the ITS2, the consensus ITS1 secondary structures for each MAST-4 clade suggest five different species (one per clade), with the presence of CBCs among clades in the conserved 4 bp stem in helix III. If this stem was elongated with one additional base pair, clades C and E would include more putative species.

In order to further investigate whether each MAST-4 clade is composed of one or several species, we contrasted their corresponding tree topologies recovered by the ITS1 and ITS2 regions. These are rapidly evolving spacers that can be used to explore questions related to the speciation process (Coleman 2007; Mullineux and Hausner 2009). We hypothesize that groups that have diversified enough to constitute different species will display congruent topologies in their ITS1 and ITS2 trees, since no recombination would exist among these markers. In contrast, groups that may still constitute one single species, or are in the process of speciation, might display incongruent topologies due to present or recent recombination events. Basically, this is the concordance–discordance principle used for the recognition of phylogenetic species (Taylor et al. 2000). The incongruent topologies displayed by clade A (fig. 4*B*) suggest that it could represent a single biological species. This is consistent with the low intraclade divergence in the ITS region (table 1), which is within the observed variation in other species (table 2). On the contrary, clade C displays the same topology in the two ITS trees, revealing three subclades that could qualify as separate species. In addition, clade C has an intraclade divergence similar to the minimum interclade distance (table 1) and also to the average minimum divergence between species (table 2). A similar scenario of high divergence is seen with the two sequences from clade B. Finally, clade E has some sequences intermixed in the trees and others with a consistent position, suggesting that it may include more than one species.

There are a few alternative explanations, besides speciation, for the ITS variability we observe. It is unlikely

Table 2. Different Examples of ITS Sequence Divergences (uncorrected p-distance; shown as minimum–maximum) among Related Strains or Species of Cultured Eukaryotes.

		Intraclonal			Intraspecies			Interspecies			
	Species	ITS	ITS1	ITS2	ITS	ITS1	ITS2	ITS	ITS1	ITS2	Ref.[a]
Diatoms	*Eunotia bilunaris*	0.000–0.052			0.000–0.123						1
	E. bilunaris	0.000–0.043			0.000–0.044						2
	Pseudo-nitzschia multistriata				0.006	0.010	0.006				3
	P. pungens	0.000–0.070			0.000–0.044	0.000–0.050	0.000–0.064				4
	P. seriata and *P. australis*								0.036	0.027	5
	P. decipiens and *P. dolorosa*				0.000–0.005			0.105–0.108			6
	P. delicatissima and *P. decipiens*				0.000–0.049			0.075–0.090			
	P. dolorosa and *P. delicatissima*				0.000–0.002			0.129–0.151			
	Several species (5.8S+ITS2)				0.000–0.070			0.110–0.260			7
Dinoflagellates	*Symbiodinium*	0.006–0.061	0.009–0.043	0.010–0.124							8
	Several species	0.000–0.017	0.000–0.034	0.000–0.026	0.000–0.021	0.000–0.040	0.000–0.021	0.042–0.577	0.038–0.734	0.020–0.732	9
	Several species							0.000–0.014			10
	Peridinium limbatum and *P. willei*					0.000–0.099	0.000–0.111		0.551–0.566	0.432–0.463	11
	Scrippsiella trochoidea				0.002–0.015						12
Ciliates	*Halteria grandinella*				0.001–0.082						13
Mollusca	*Haliotis*					0.000–0.049	0.000–0.044		0.380–0.590	0.380–0.480	14
Copepod	Several species						0.000–0.008			0.002–0.034	15
Magnoliophyta	Several species								0.000–0.480	0.000–0.440	16
Averages		0.001–0.049	0.005–0.039	0.005–0.075	0.000–0.042	0.002–0.050	0.001–0.042	0.077–0.200	0.201–0.481	0.144–0.363	

[a] References: 1) Vanormelingen et al. (2008), 2) Vanormelingen et al. (2007), 3) D'Alelio et al. (2008), 4) Casteleyn et al. (2008), 5) Fehling et al. (2004), 6) Lundholm et al. (2006), 7) Moniz and Kaczmarska (2009), 8) Thornhill et al. (2007), 9) Litaker et al. (2007), 10) Logares et al. (2008), 11) Kim et al. (2004), 12) Montresor et al. (2003), 13) Katz et al. (2005), 14) Coleman and Vacquier (2002), 15) Goetze (2003), and 16) Goertzen (2003).

that this variability is due to experimental artifacts, as we used very stringent sequencing and analysis methods, but it could be caused by intragenomic polymorphisms (Prokopowich et al. 2003). It is generally assumed that these polymorphisms are rapidly eliminated through a series of homogenizing mechanisms referred to as concerted evolution (Elder and Turner 1995; Ganley and Kobayashi 2007), but it is also known that some species have a fraction of the rDNA units that have escaped the process of concerted evolution (Keller et al. 2006; Simon and Weiss 2008). Intragenomic variation in species where this occurs is very low (lower than interspecific variation, Litaker et al. 2007) and typically only in extremely variable positions that are never paired in secondary structure (Behnke et al. 2004; Orsini et al. 2004; Casteleyn et al. 2008). Thus, the ITS can be treated as single copy region (Coleman 2003). Finally, we observed that MAST-4 cells have a relatively low rDNA copy number (around 30, Rodríguez-Martínez et al. 2009), which reduces the possibility of mutations.

In summary, despite the presence of a huge number of MAST-4 cells in the oceans, its diversity is structured into just five main clades, each representing at least one biological species. Clade A is particularly interesting because it seems to be composed of only one species, appearing in distant oceanic areas. Specifically, clade A showed no polymorphisms in the critical regions of the ITS2 and ITS1 secondary structures (figs. 5 and 6), the topologies of their ITS1 and ITS2 trees were incongruent (fig. 4), and the clade presented a very low sequence divergence (table 1). On the other hand, using the same three criteria, clade B would include at least two species, clade C three species, and clade E two species (clade D is still undersampled with only one ITS sequence). Altogether there is currently evidence of a maximum of only ten separate species within MAST-4. Within each of the clades, diversification appears to be very low, as indicated by the 18S and ITS rDNA markers. This low evolutionary diversification points to either a very recent evolutionary divergence and worldwide dispersal or to a very strong environmental filtering that penalizes any deviation from an optimal cell design.

Acknowledgments

Funding has been provided by FLAME (CGL2010-16304, MICINN, Spain) and BioMarKs (2008-6530, ERA-net Biodiversa, EU) projects to R.M., by an FPI fellowship from the Spanish Ministry of Education and Science to R.R.M., by a Marie Curie Intra-European Fellowship grant PIEF-GA-2009-235365 to R.L., and by NSF ATOL 0629521 to G.R. We thank M. Pernice for Mothur help, V. Balagué and C. Williams for molecular help, R.E. Collins for useful advice, and all people involved in the BioMarKs consortium, in particular, the coordinator Colomban de Vargas.

References

Álvarez I, Wendel J. 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Mol Phylogenet Evol.* 29:417–434.

Amato A, Kooistra W, Levialdi Ghiron J, Mann D, Pröschold T, Montresor M. 2007. Reproductive isolation among sympatric cryptic species in marine diatoms. *Protist* 158:193–207.

Behnke A, Friedl T, Chepurnov V, Mann D. 2004. Reproductive compatibility and rDNA sequence analyses in the Sellaphora pupula species complex (Bacillariophyta). *J Phycol.* 40:193–208.

Bridge PD, Schlitt T, Cannon PF, Buddie AG, Baker M, Borman AM. 2008. Domain II hairpin structure in ITS1 sequences as an aid in differentiating recently evolved animal and plant pathogenic fungi. *Mycopathologia* 166:1–16.

Casteleyn G, Chepurnov V, Leliaert F, Mann D, Bates S, Lundholm N, Rhodes L, Sabbe K, Vyverman W. 2008. Pseudo-nitzschia pungens (Bacillariophyceae): a cosmopolitan diatom species? *Harmful Algae* 7:241–257.

Coleman A. 2000. The significance of a coincidence between evolutionary landmarks found in mating affinity and a DNA sequence. *Protist* 151:1.

Coleman A. 2003. ITS2 is a double-edged tool for eukaryote evolutionary comparisons. *Trends Genet.* 19:370–375.

Coleman A. 2007. Pan-eukaryote ITS2 homologies revealed by RNA secondary structure. *Nucleic Acids Res.* 35:3322–3329.

Coleman AW. 2009. Is there a molecular key to the level of "biological species" in eukaryotes? A DNA guide. *Mol Phylogenet Evol.* 50:197–203.

Coleman AW, Maria Preparata R, Mehrotra B, Mai JC. 1998. Derivation of the secondary structure of the ITS-1 transcript in Volvocales and its taxonomic correlations. *Protist* 149:135–146.

Coleman AW, Vacquier VD. 2002. Exploring the phylogenetic utility of ITS sequences for animals: a test case for abalone (Haliotis). *J Mol Evol.* 54:246–257.

Côté CA, Greer CL, Peculis BA. 2002. Dynamic conformational model for the role of ITS2 in pre-rRNA processing in yeast. *RNA* 8:786–797.

D'Alelio D, Amato A, Kooistra W, Procaccini G, Casotti R, Montresor M. 2008. Internal transcribed spacer polymorphism in Pseudo-nitzschia multistriata (Bacillariophyceae) in the Gulf of Naples: recent divergence or intraspecific hybridization? *Protist* 160:9–20.

Díez B, Pedrós-Alió 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.

Drummond AJ, Ashton B, Buxton S, et al. (12 co-authors). 2010. Geneious v5.1 [Internet]. Available from: http://www.geneious.com.

Edvardsen B, Eikrem W, Green J, Andersen RA, Moon-van der Staay SY, Medlin LK. 2000. Phylogenetic reconstructions of the Haptophyta inferred from 18S ribosomal DNA sequences and available morphological data. *Phycologia* 39:19–35.

Elder JF, Turner BJ. 1995. Concerted evolution of repetitive DNA sequences in eukaryotes. *Q Rev Biol.* 70:297–320.

Evans KM, Wortley AH, Mann DG. 2007. An assessment of potential diatom "barcode" genes (cox1, rbcL, 18S and ITS rDNA) and their effectiveness in determining relationships in Sellaphora (Bacillariophyta). *Protist* 158:349–364.

Falkowski PG, Fenchel T, Delong EF. 2008. The microbial engines that drive earth's biogeochemical cycles. *Science* 320:1034–1039.

Fehling J, Green DH, Davidson K, Bolch CJ, Bates SS. 2004. Domoic acid production by Pseudo-nitzschia seriata (Bacillariophyceae) in Scottish waters. *J Phycol.* 40:622–630.

Ganley ARD, Kobayashi T. 2007. Highly efficient concerted evolution in the ribosomal DNA repeats: total rDNA repeat variation revealed by whole-genome shotgun sequence data. *Genome Res.* 17:184–191.

Gaston KJ, Blackburn TM. 1997. How many birds are there? *Biodivers Conserv.* 6:615–625.

Goertzen L. 2003. ITS secondary structure derived from comparative analysis: implications for sequence alignment and phylogeny of the Asteraceae. *Mol Phylogenet Evol.* 29:216–234.

Goetze E. 2003. Cryptic speciation on the high seas; global phylogenetics of the copepod family Eucalanidae. *Proc Biol Sci.* 270:2321–2331.

Gottschling M. 2004. Secondary structure models of the nuclear internal transcribed spacer regions and 5.8S rRNA in Calciodinelloideae (Peridiniaceae) and other dinoflagellates. *Nucleic Acids Res.* 32:307–315.

Gottschling M, Hilger H, Wolf M, Diane N. 2001. Secondary structure of the ITS1 transcript and its application in a reconstruction of the phylogeny of Boraginales. *Plant Biol.* 3: 629–636.

Gutell R, Larsen N, Woese C. 1994. Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. *Microbiol Mol Biol Rev.* 58:10.

Haas BJ, Gevers D, Earl AM, et al. (15 co-authors). 2011. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res.* 21:494–504.

Hoshina R. 2010. Secondary structural analyses of ITS1 in Paramecium. *Microbes Environ.* 25:313–316.

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755.

Itskovich V, Gontcharov A, Masuda Y, Nohno T, Belikov S, Efremova S, Meixner M, Janussen D. 2008. Ribosomal ITS sequences allow resolution of freshwater sponge phylogeny with alignments guided by secondary structure prediction. *J Mol Evol.* 67:608–620.

Jerome C, Lynn D. 1996. Identifying and distinguishing sibling species in the Tetrahymena pyriformis complex (Ciliophora, Oligohymenophorea) using PCR/RFLP analysis of nuclear ribosomal DNA. *J Eukaryot Microbiol.* 43:492–497.

Joseph N, Krauskopf E, Vera M, Michot B. 1999. Ribosomal internal transcribed spacer 2 (ITS2) exhibits a common core of secondary structure in vertebrates and yeast. *Nucleic Acids Res.* 27:4533.

Katoh K, Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform.* 9:286–298.

Katz LA, McManus GB, Snoeyenbos-West OLO, Griffin A, Pirog K, Costas B, Foissner W. 2005. Reframing the 'Everything is everywhere' debate: evidence for high gene flow and diversity in ciliate morphospecies. *Aquat Microb Ecol.* 41:55–65.

Keller I, Chintauan-Marquier IC, Veltsos P, Nichols RA. 2006. Ribosomal DNA in the grasshopper Podisma pedestris: escape from concerted evolution. *Genetics* 174:863–874.

Kim E, Wilcox L, Graham L, Graham J. 2004. Genetically distinct populations of the dinoflagellate Peridinium limbatum in neighboring Northern Wisconsin lakes. *Microb Ecol.* 48:521–527.

Lange M, Chen YQ, Medlin LK. 2002. Molecular genetic delineation of Phaeocystis species (Prymnesiophyceae) using coding and non-coding regions of nuclear and plastid genomes. *Eur J Phycol.* 37:77–92.

Letunic I, Bork P. 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23:127–128.

Litaker RW, Vandersea MW, Kibler SR, Reece KS, Stokes NA, Lutzoni FM, Yonish BA, West MA, Black MND, Tester PA. 2007. Recognizing dinoflagellate species using ITS rDNA sequences. *J Phycol.* 43:344–355.

Logares R, Daugbjerg N, Boltovskoy A, Kremp A, Laybourn-Parry J, Rengefors K. 2008. Recent evolutionary diversification of a protist lineage. *Environ Microbiol.* 10:1231–1243.

Logares R, Rengefors K, Kremp A, Shalchian-Tabrizi K, Boltovskoy A, Tengs T, Shurtleff A, Klaveness D. 2007. Phenotypically different microalgal morphospecies with identical ribosomal DNA: a case of rapid adaptive evolution? *Microb Ecol.* 53:549–561.

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.

Lowe CD, Montagnes DJS, Martin LE, Watts PC. 2010. Patterns of genetic diversity in the marine heterotrophic flagellate Oxyrrhis marina (Alveolata: Dinophyceae). *Protist* 161:212–221.

Lundholm N, Moestrup Ø, Kotaki Y, Hoef-Emden K, Scholin C, Miller P. 2006. Inter-and intraspecific variation of the Pseudo-nitzschia delicatissima complex (Bacillariophyceae) illustrated by rRNA probes, morphological data and phylogenetic analyses. *J Phycol.* 42:464–481.

Mai JC, Coleman AW. 1997. The internal transcribed spacer 2 exhibits a common secondary structure in green algae and flowering plants. *J Mol Evol.* 44:258–271.

Massana R. 2011. Eukaryotic picoplankton in surface oceans. *Annu Rev Microbiol.* 65:1–47.

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

Massana R, DeLong EF, Pedrós-Alió C. 2000. A few cosmopolitan phylotypes dominate planktonic archaeal assemblages in widely different oceanic provinces. *Appl Environ Microbiol.* 66:1777–1787.

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

Massana R, Guillou L, Terrado R, Forn I, Pedros-Alio C. 2006. Growth of uncultured heterotrophic flagellates in unamended seawater incubations. *Aquat Microb Ecol.* 45:171–180.

Massana R, Terrado R, Forn I, Lovejoy C, Pedrós-Alió C. 2006. Distribution and abundance of uncultured heterotrophic flagellates in the world oceans. *Environ Microbiol.* 8:1515–1522.

May RM, Beverton RJH. 1990. How many species? *Philos Trans R Soc Lond B Biol Sci.* 330:293.

Moniz MBJ, Kaczmarska I. 2009. Barcoding diatoms: is there a good marker? *Mol Ecol Resour.* 9:65–74.

Moniz MBJ, Kaczmarska I. 2010. Barcoding of diatoms: nuclear encoded ITS revisited. *Protist* 161:7–34.

Montresor M, Sgrosso S, Procaccini G, WHCF Kooistra. 2003. Intraspecific diversity in Scrippsiella trochoidea (Dinophyceae): evidence for cryptic species. *Phycologia* 42:56–70.

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

Müller T, Philippi N, Dandekar T, Schultz J, Wolf M. 2007. Distinguishing species. *RNA* 13:1469.

Mullineux T, Hausner G. 2009. Evolution of rDNA ITS1 and ITS2 sequences and RNA secondary structures within members of the fungal genera Grosmannia and Leptographium. *Fungal Genet Biol.* 46:855–867.

Not F, Gausling R, Azam F, Heidelberg JF, Worden AZ. 2007. Vertical distribution of picoeukaryotic diversity in the Sargasso Sea. *Environ Microbiol.* 9:1233–1252.

Orsini L, Procaccini G, Sarno D, Montresor M. 2004. Multiple rDNA ITS-types within the diatom Pseudo-nitzschia delicatissima (Bacillariophyceae) and their relative abundances across a spring bloom in the Gulf of Naples. *Mar Ecol Prog Ser.* 271:87–98.

Prokopowich CD, Gregory TR, Crease TJ. 2003. The correlation between rDNA copy number and genome size in eukaryotes. *Genome* 46:48–50.

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

Rambaut A. 2009. FigTree. Edinburgh (United Kingdom): Institute of Evolutionary Biology, University of Edinburgh. Available from: http://tree.bio.ed.ac.uk/software/figtree.

Rocap G, Distel D, Waterbury J, Chisholm S. 2002. Resolution of Prochlorococcus and Synechococcus ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. *Appl Environ Microbiol.* 68:1180.

Rodríguez-Martínez R, Labrenz M, Del Campo J, Forn I, Jürgens K, Massana R. 2009. Distribution of the uncultured protist MAST-4 in the Indian Ocean, Drake Passage and Mediterranean Sea assessed by real-time quantitative PCR. *Environ Microbiol.* 11: 397–408.

Rynearson TA, Lin EO, Armbrust EV. 2009. Metapopulation structure in the planktonic diatom Ditylum brightwellii (Bacillariophyceae). *Protist* 160:111–121.

Schloss PD, Westcott SL, Ryabin T, et al. (15 co-authors). 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol.* 75:7537–7541.

Schlötterer C, Hauser MT, von Haeseler A, Tautz D. 1994. Comparative evolutionary analysis of rDNA ITS regions in Drosophila. *Mol Biol Evol.* 11:513.

Schultz J, Muller T, Achtziger M, Seibel P, Dandekar T, Wolf M. 2006. The internal transcribed spacer 2 database—a web server for (not only) low level phylogenetic analyses. *Nucleic Acids Res.* 34:W704.

Seibel P, Müller T, Dandekar T, Schultz J, Wolf M. 2006. 4 SALE—a tool for synchronous RNA sequence and secondary structure alignment and editing. *BMC Bioinformatics* 7:498.

Selig C, Wolf M, Muller T, Dandekar T, Schultz J. 2007. The ITS2 Database II: homology modelling RNA structure for molecular systematics. *Nucleic Acids Res.* 36:D377–D380.

Simon UK, Weiss M. 2008. Intragenomic variation of fungal ribosomal genes is higher than previously thought. *Mol Biol Evol.* 25:2251–2254.

Slapeta J, López-García P, Moreira D. 2006. Global dispersal and ancient cryptic species in the smallest marine eukaryotes. *Mol Biol Evol.* 23:23–29.

Sorhannus U, Ortiz JD, Wolf M, Fox MG. 2010. Microevolution and speciation in Thalassiosira weissflogii (Bacillariophyta). *Protist* 161:237–249.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.

Stoeck T, Bass D, Nebel M, Christen R, Jones MDM, Breiner H-W, Richards TA. 2010. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol Ecol.* 19:21–31.

Taylor JW, Jacobson DJ, Kroken S, Kasuga T, Geiser DM, Hibbett DS, Fisher MC. 2000. Phylogenetic species recognition and species concepts in fungi. *Fungal Genet Biol.* 31:21–32.

Thompson J, Marcelino L, Polz M. 2002. Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by 'reconditioning PCR'. *Nucleic Acids Res.* 30:2083.

Thornhill DJ, Lajeunesse TC, Santos SR. 2007. Measuring rDNA diversity in eukaryotic microbial systems: how intragenomic variation, pseudogenes, and PCR artifacts confound biodiversity estimates. *Mol Ecol.* 16:5326–5340.

Thornhill DJ, Lord JB. 2010. Secondary structure models for the internal transcribed spacer (ITS) region 1 from symbiotic dinoflagellates. *Protist* 161:434–451.

Tippery N, Les D. 2008. Phylogenetic analysis of the internal transcribed spacer (ITS) region in Menyanthaceae using predicted secondary structure. *Mol Phylogenet Evol.* 49:526–537.

Vanormelingen P, Chepurnov V, Mann D, Cousin S, Vyverman W. 2007. Congruence of morphological, reproductive and ITS rDNA sequence data in some Australasian Eunotia bilunaris (Bacillariophyta). *Eur J Phycol.* 42:61–79.

Vanormelingen P, Chepurnov V, Mann D, Sabbe K, Vyverman W. 2008. Genetic divergence and reproductive barriers among morphologically heterogeneous sympatric clones of Eunotia bilunaris sensu lato (Bacillariophyta). *Protist* 159:73–90.

Wang S, Bao Z, Li N, Zhang L, Hu J. 2007. Analysis of the secondary structure of ITS1 in Pectinidae: implications for phylogenetic reconstruction and structural evolution. *Mar Biotechnol.* 9: 231–242.

Wiemers M, Keller A, Wolf M. 2009. ITS2 secondary structure improves phylogeny estimation in a radiation of blue butterflies of the subgenus Agrodiaetus (Lepidoptera: Lycaenidae: Polyommatus). *BMC Evol Biol.* 9:300.

Worden A. 2006. Picoeukaryote diversity in coastal waters of the Pacific Ocean. *Aquat Microb Ecol.* 43:165–175.

Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31:3406.