The Journal of Eukaryotic Microbiology

ISOP Published by the International Society of Protistologists

Journal of Eukaryotic Microbiology ISSN 1066-5234

ORIGINAL ARTICLE

Reference Tree and Environmental Sequence Diversity of Labyrinthulomycetes

Jingwen Pan, Javier del Campo & Patrick J. Keeling

Department of Botany, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

Keywords
18s rRNA; amphitremids; high-throughput sequencing; labyrinthulids; phylogeny; thraustochytrids.

Correspondence
P.J. Keeling, Department of Botany, University of British Columbia, 3529-6270 University Blvd, Vancouver, BC V6T 1Z4, Canada
Telephone number: +1-604-822-4906;
FAX number: +1-604-822-6089;
e-mail: pkeeling@mail.ubc.ca

Received: 22 January 2016; revised 27 April 2016; accepted May 17, 2016.

doi:10.1111/jeu.12342

ABSTRACT

Labyrinthulomycetes are heterotrophic stramenopiles that are ubiquitous in a wide range of both marine and freshwater habitats and play important roles in decomposition of organic matter. The diversity and taxonomy of Labyrinthulomycetes has been studied for many years, but we nevertheless lack both a comprehensive reference database and up-to-date phylogeny including all known diversity, which hinders many global insights into their ecological distribution and the relative importance of various subgroups in different environments. Here, we present a curated reference database and a phylogenetic tree of Labyrinthulomycetes small subunit ribosomal RNA (SSU or 18S rRNA) data. Based on this created reference database, we analyzed high-throughput environmental sequencing data, revealing many previously unknown environmental clades and exploring the ecological distribution of various subgroups. Particularly, a number of newly identified environmental clades that are widespread in the open ocean. Comparing the manually curated reference database to existing tools for identification of environmental sequences (e.g. PR2 or SILVA databases) suggests that the curated database provides a higher degree of specificity and a lower frequency of misidentification. The phylogenetic framework and database will be a useful tool for future ecological and evolutionary studies.

LABYRINTHULOMYCETES are a group of ubiquitous and diverse unicellular stramenopiles. They are characterized by the production of an ectoplasmic network, a membrane-bound, branched network secreted through a unique organelle called bothrosome (sagenogenetosome) that is involved in saprotrophic nutrient uptake. Since the first description of *Labyrinthula* in 1867 by Cienkowski, the classification within Labyrinthulomycetes has undergone several changes and rearrangements (Honda et al. 1999; Leander and Porter 2001; Leander et al. 2004; Olive 1975; Porter 1989; Yokoyama and Honda 2007; Yokoyama et al. 2007). Based on the most up-to-date taxonomical classification of this group, Labyrinthulomycetes are composed of three orders: Labyrinthulida, Thraustochytrida, and Amphitremida (Beakes et al. 2014; Gomaa et al. 2013; Takahashi et al. 2014).

Labyrinthula and *Aplanochytrium*, the two genera that crawl using their ectoplasmic network, are included in the order Labyrinthulida. Species of *Aplanochytrium* are distinguished by having their cells not embedded within the ectoplasmic network and by the production of nonflagellated "crawling spores" (Leander et al. 2004; Tsui et al. 2009).

Vegetative cells of species belonging to the Thraustochytrida are often spherical, unicellular, or colonial. They are immobile and the ectoplasmic network is only used to increase surface area for enzyme secretion and nutrient absorption. The family Thraustochytriidae represents most of the diversity within the order Thraustochytrida and is composed of *Aurantiochytrium*, *Botryochytrium*, *Parietichytrium*, the Quahog parasite, *Schizochytrium*, *Sicyoidochytrium*, *Thraustochytrium*, and *Ulkenia*. Ameboid cell stages have been observed in some genera including *Ulkenia*, *Sicyoidochytrium*, *Parietichytrium*, and *Botryochytrium* (Yokoyama et al. 2007). Species of *Althornia* (family Althorniidae) are free floating and do not have a bothrosome or produce an ectoplasmic network (Alderman and Jones 1971; Bower 1987; Moss 1985). *Amphifila* (family Amphifilidae) is newly named to accompany the rearrangement of *Diplophrys marina* into *Amphifila marina*

(Anderson and Cavalier-Smith 2012). Species of this genus differ from other Thraustochytrida in having pseudostomes instead of true bothrosomes, and ectoplasmic elements in the form of pseudopodia (Anderson and Cavalier-Smith 2012; Gomaa et al. 2013).

The original genus *Diplophrys* with its remaining species is now placed under family Diplophryidae. Diplophryidae and Amphitremidae (containing *Amphitrema* and *Archerella*) form the third-order Amphitremida of Labyrinthulomycetes (Gomaa et al. 2013; Takahashi et al. 2014). Cells of Amphitremida also possess pseudostomes and pseudopodia. Both *Amphifila* and *Diplophrys* bear refractive granules in their cytoplasm that are visible under light microscope (Anderson and Cavalier-Smith 2012; Gomaa et al. 2013). *Amphitrema* and *Archerella* both harbor photosynthetic zoochlorellae endosymbionts and are thus mixotrophic.

The family Oblongichytriidae contains only one genus, *Oblongichytrium*. While it has been classified under Thraustochytrida, phylogenetic studies often place *Oblongichytrium* sister to Labyrinthulida (Anderson and Cavalier-Smith 2012; Gomaa et al. 2013; Takahashi et al. 2014; Yokoyama et al. 2007) or sister to both Labyrinthulida and other Thraustochytrida (Collado-Mercado et al. 2010; Yokoyama and Honda 2007).

Labyrinthulomycetes can be found in a diverse range of habitats, including both marine and freshwater, from the epipelagic surface to the deep sea (Raghukumar 2002). They have also been isolated from various kinds of substrates, including but not limited to algae, mangrove leaves, seagrass, coral mucus, and mollusks (Raghukumar and Damare 2011). Most Labyrinthulomycetes are saprotrophic feeders through an osmotrophic or phagotrophic mode of nutrient uptake. In fact, they are often seen to be associated with detritus like fallen mangrove leaves, decomposing algae, and fecal pellets of marine invertebrates (Raghukumar and Raghukumar 1999; Tsui et al. 2009). The production of high level of omega-3 polyunsaturated fatty acids by *Aurantiochytrium*, including docosahexaenoic acid (DHA) and eicosapentaenoic acid (EPA), has also made Labyrinthulomycetes to be commercially valuable (Lee Chang et al. 2012).

Even though some aspects of the Labyrinthulomycetes have been studied in depth, their environmental diversity has not yet been fully explored. Collado-Mercado et al. (2010) and Ueda et al. (2015) both attempted to uncover the hidden diversity of this group in marine samples, but they are also known to be abundant in freshwater and soil environments, especially in the Amphifilidae and Amphitremida. In addition, we currently lack an up-to-date reference database for the identification of Labyrinthulomycetes in ecological studies. Large-scale data sets generated by high-throughput sequencing methods to access microbial community composition rely heavily on the accuracy of reference database to make taxonomic assignments, and the most commonly used of these, PR2 (Guillou et al. 2012) and SILVA (Quast et al. 2012) are not optimized for many groups of microbial eukaryotes.

Here, we develop a manually curated reference database for Labyrinthulomycetes SSU rRNA data. We use this to generate a reference tree and evaluate the diversity and taxonomy, allowing us to identify and name novel diversity within the group. We also applied this database to two high-throughput environmental sequence (HTES) databases, the VAMPS database (Huse et al. 2008; Sogin et al. 2006) and the Tara Ocean database (de Vargas et al. 2015) to examine the utility of the reference data and the diversity and distribution of Labyrinthulomycetes. Furthermore, the analysis of these HTES datasets will allow us to expand dramatically the described diversity of Labyrinthulomycetes in freshwater and marine environments.

MATERIALS AND METHODS

Reference phylogenetic tree construction

All GenBank SSU rDNA sequences taxonomically identified as Labyrinthulomycetes were retrieved using the corresponding taxid (35131). Mitochondrial sequences and complete genomes were excluded, as were sequences shorter than 500 bp. The remaining sequences were clustered at 97% identity using USEARCH v7.0.1090 (Edgar 2010). In order to build the tree, 44 other stramenopiles, one *Planomonas* sp. and one *Sabulodinium* sp. sequences were used as outgroups. All sequences were aligned and trimmed using MAFFT with default setting (Katoh and Standley 2013) and trimAl (Capella-Gutierrez et al. 2009), respectively. A maximum likelihood phylogenetic tree was constructed with RAxML 8.1.3 (Stamatakis 2014) using the rapid hill climbing algorithm and GTRCATI evolutionary model. Whether sequences were from Labyrinthulomycetes or not was determined based on the tree topology and literature (Leander and Porter 2001; Leander et al. 2004; Liu et al. 2013; Yokoyama and Honda 2007; Yokoyama et al. 2007). Verified sequences were then used to iteratively retrieve more sequences from GenBank using blastn (Camacho et al. 2009) (*E*-value = 10e-5) against NCBI nonredundant/nucleotide collection (nr/nt) as previously described (del Campo and Massana 2011; del Campo and Ruiz-Trillo 2013). The first 100 hits were kept for each query. After removing duplicated sequences, the new ones were added to our dataset and used to construct a phylogeny, as described above. New sequences belonging to Labyrinthulomycetes were blasted again against nr/nt in order to retrieve even more sequences. This process was repeated iteratively until no new sequences that branch with Labyrinthulomycetes were retrieved from GenBank. Sequences were then checked for chimeras using both the built-in function of Qiime v1.9.1 (identify_chimeric_seqs.py) (Caporaso et al. 2010b) against the SILVA database (v119, 97% identity) and USEARCH (uchime_denovo). Chimeric sequences were also manually examined.

The final phylogenetic tree was built using RAxML with the settings mentioned above. Statistical support for the consensus tree was calculated using nonparametric bootstrapping with 1,000 replicates. Support for Bayesian posterior probability was examined with MrBayes v3.2.2 (Altekar et al. 2004; Ronquist and Huelsenbeck 2003) using the GTR + Gamma model. The analysis was

performed using 64 MCMC chains with a sampling frequency of every 1,000th generation. A consensus tree was generated after discarding the first 50% of the total generations as "burn-in".

Reference database construction and annotation

Sequences were first identified for annotation based on previously published works on Labyrinthulomycetes phylogeny (Leander and Porter 2001; Leander et al. 2004; Liu et al. 2013; Yokoyama and Honda 2007; Yokoyama et al. 2007), excluding environmental clades. We tried to adopt the established taxonomy as our classification method as far as it was supported by our tree. New names based on the most representative cultured strain were added for those undescribed clusters when the bootstrap support for the clade was 70% or higher. If a group contained only environmental sequences, the group was then named LAB"0", where "0" is a number. Environmental singletons (OTUs represented by a single sequence) were unannotated. The classification for the monophyletic, nonenvironmental groups followed the most updated taxonomy available (Beakes et al. 2014; Gomaa et al. 2013). Metadata for the sequences in our dataset was downloaded from GenBank using custom scripts. For sequences still missing environmental data, their information was then collected manually from the literature. We were able to retrieve environmental metadata from Genbank for 924 sequences (out of 1,181). The 257 sequences that have no metadata available are all cultured strains.

Analysis of V9 HTES sequences

Sequences annotated as Labyrinthulomycetes or more generically as Stramenopiles were retrieved from two 18S rRNA V9 region databases (100–150 bp), VAMPS and Tara Oceans (Huse et al. 2014; de Vargas et al. 2015). The VAMPS (Visualization and Analysis of Microbial Population Structures, vamps.mbl.edu) database contains HTES data from different environments for both bacteria and eukarya. On the other hand, the Tara Oceans (taraoceans.sb-roscoff.fr/EukDiv/) database contains exclusively eukaryotic V9 reads from the sunlit ocean. The fasta file containing all reads was sorted by length using USEARCH and clustered into OTUs with 97% similarity using QIIME with default setting (UCLUST). OTUs were then aligned with the reference alignment using PyNAST (Caporaso et al. 2010a) embedded in QIIME (align_seqs.py). The reference alignment was the same alignment that was used to generate the reference phylogenetic tree. OTUs that the PyNAST algorithm failed to align were discarded. The PyNAST alignment output was merged with the reference alignment and filtered for gap positions using QIIME (filter_alignment.py) with gap filtering threshold set to 0.99 and entropy threshold set to 0.0001. Identification of Labyrinthulomycetes reads used a maximum likelihood phylogenetic approach by mapping the OTUs onto the Labyrinthulomycetes reference tree using the Evolutionary Placement Algorithm (EPA) of RAxML (Berger et al. 2011). OTUs that were not placed within the Labyrinthulomycetes were removed, together with their 97% clustered sequences. Trees using the remaining sequences were built consecutively until no more reads were placed outside the Labyrinthulomycetes. OTUs and their clustered sequences were then annotated according to their placement. OTUs that were not placed with any previously defined groups were assigned a new name as outlined above.

Abundance and richness distribution patterns

For abundance and richness analyses, comparisons between different groups were done at the order level, except for the order Thraustochytrida where the family level was used. Abundance represents the number of sequences found in each group, while richness represents the number of different OTUs with less than 97% similarity to each other. For data from VAMPS and Tara Oceans, abundance was calculated based on raw reads data through custom scripts that link the OTU table with their previous clustering frequency tables (available online for each database). The abundance of each group under each environmental category was then calculated using an Excel pivot table and a heatmap was generated to illustrate the distribution patterns. Richness was calculated by determining the number of discrete OTUs in each group.

RESULTS

Phylogeny of Labyrinthulomycetes

In total, 1,181 18S rDNA sequences longer than 500 bp were retrieved from GenBank by iteratively screening for sequences branching with taxonomically verified Labyrinthulomycetes. The phylogeny of these sequences was constructed with 332 OTU_{97}, after clustering the 1,181 sequences at the 97% level. Virtually, all of the previously described clusters of Labyrinthulomycetes were recovered with strong support (> 70/0.7). While most genera of Thraustochytriidae form a monophyletic group with support of 88/1, the genus *Oblongichytrium* branched basal to both Labyrinthulida and Thraustochytrida (Table S2 and Fig. S1). Considering the uncertain placement of *Oblongichytrium* among different studies, and to avoid non-monophyletic clades, we use the provisional order name "Oblongichytrida" conscious that it needs to be confirmed by further morphological investigations. Within Thraustochytriidae, species of Thraustochytrium branch at various locations, often interspersed with other genera, suggesting previous misidentification of some organisms. The phylogenetic tree also revealed 20 new environmental clades, most of which have over 70% bootstrap support and Bayesian posterior probability of 1 (Fig. 1). Among these new clades, 16 of them do not belong to any of the previously defined major groups. Former environmental groups defined by Collado-Mercado et al. (2010) have been renamed according to our name system for consistency. The correspondence between their naming and ours can be found in Table S1. The

ancestral node of LAB1, LAB6, and LAB8 is 50/0.99 supported and are placed into supergroup LAB1/6/8.

Estimating Labyrinthulomycetes relative abundance, richness, and host-associations from the reference database

Abundance and richness analyses were conducted by comparing different phylogenetic groups using a variety of parameters. The total abundance distribution between cultured and environmental GenBank sequences, as illustrated by the upper bars in Fig. 2, reveals that over half of the sequences in most of the groups are environmental, except for the Thraustochytriidae, which contains mostly sequences from cultured species. 18S rDNA sequences of Amphifilidae are highly variable, as illustrated by the over two-fold differences in abundance vs. richness. Based on the metadata collected, Thraustochytriidae and

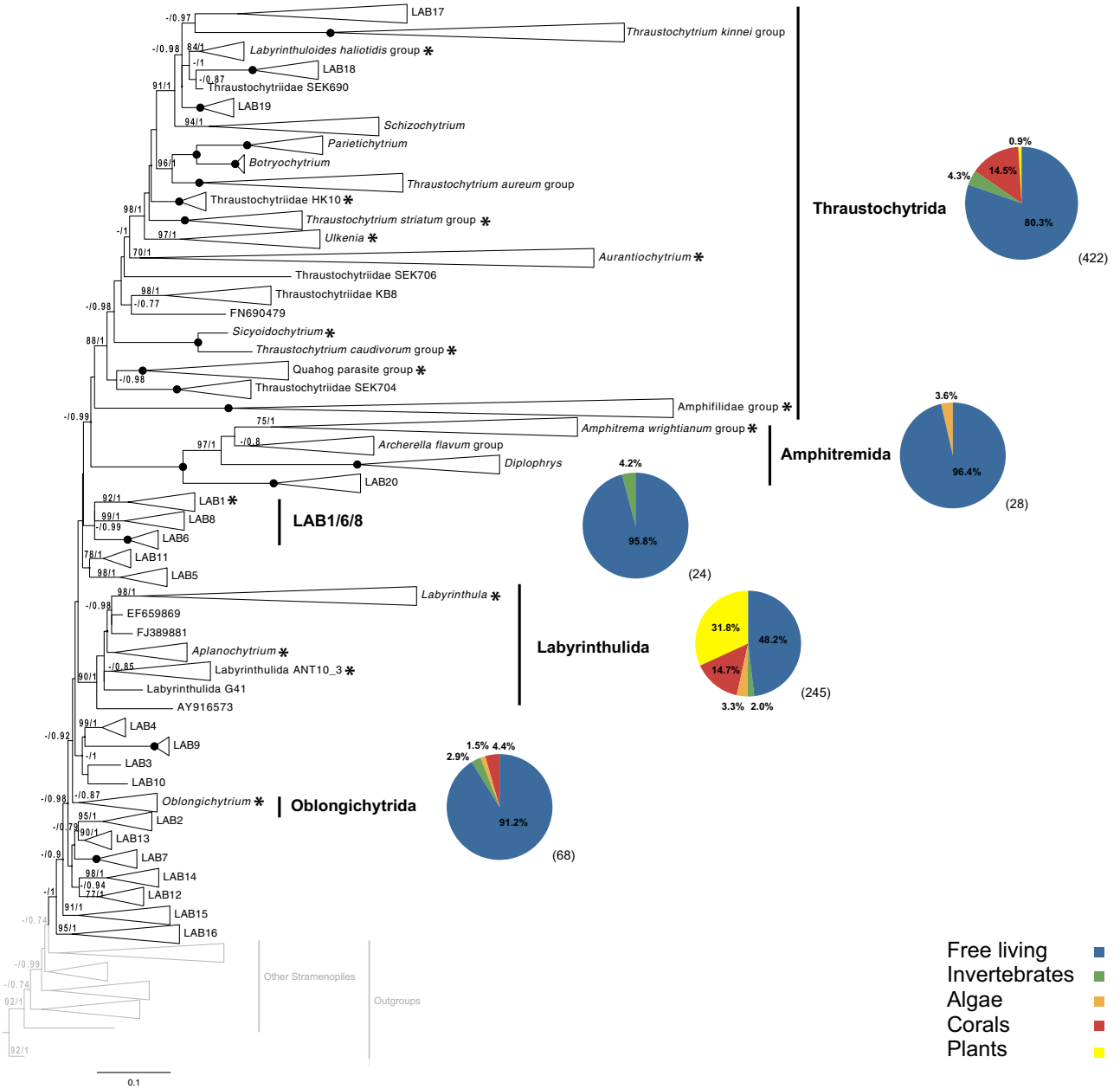


Figure 1 Diversity of Labyrinthulomycetes inferred from a maximum likelihood (RAxML) phylogenetic tree constructed using 18S rRNA sequences. Numbers at nodes represent bootstrap support/Bayesian posterior probability. Only values > 70% or 0.7 are shown. Nodes with support values of 100/1 are highlighted as black dot. Groups containing host-associated sequences are indicated by *. Host-associated sequence abundance in Labyrinthulomycetes major groups are shown as pie charts next to the order names. Numbers in brackets indicate the total number of sequences with metadata available.

Labyrinthulida are relatively common in marine environments. Mangrove forests, which are saltwater ecosystems found between terrestrial and marine environments, are also a common habitat for Labyrinthulomycetes. This is especially true for *Aurantiochytrium*, where 11% of the sequences in the database were collected from mangroves (Table S2). On the contrary, Amphifilidae, a basal subgroup of Thraustochytrida, are found primarily in freshwater and soil samples, with only three sequences from the marine environment. Similarly, Amphitremida also contain many freshwater sequences. The environmental clade AMP1 is the main marine representative of this lineage. Two sequences belonging to the *A. wrightianum* group were also retrieved from marine environments. All the environmental clades that are distinct from any defined orders were recovered from marine samples (Table S2).

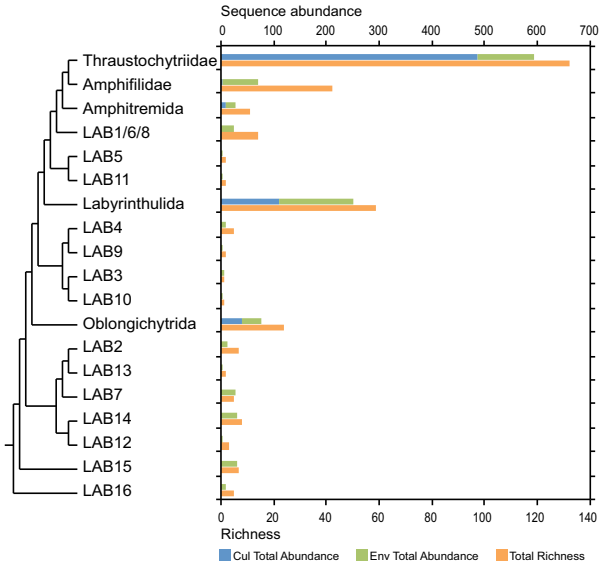


Figure 2 Abundance and richness of Labyrinthulomycetes based on retrieved GenBank sequences. On the left is a phylogenetic tree showing relationships among different groups (based on Fig. 1). On the right is a table of richness and abundance. For each group, the upper bar represents the total abundance while the lower bar indicates the richness. Upper *x*-axis: abundance, lower *x*-axis: richness. Cul Total Abundance, total abundance for cultured sequences; Env Total Abundance, total abundance for environmental sequences.

A total of 208 Labyrinthulomycetes sequences retrieved from GenBank (69 OTU_{97}), belonging to various phylogenetic groups, were isolated directly from biological substrates and they are denoted as being "host-associated" (Fig. 1). Over half of the Labyrinthulida sequences are host-associated, with two-third coming from plants (mainly *Labyrinthula* from seagrass). A second common association is between *Aplanochytrium* and coral mucus. Within *Aplanochytrium*, sequences from the subclade containing OTU_{97} representatives FJ389839, FJ389848, FJ389840, and FJ389872 are all associated with the massive coral *Favia* sp. and all come from the same study (Table S2) (Siboni et al. 2010). Another subclade of OTU_{97}, containing AF348521, AF348517, AF348518, and AF348516, is associated with a more diverse range of marine hosts, including coral, seagrass, and algae. In addition to Labyrinthulida, coral association can also be seen in Thraustochytriidae and Oblongichytrida (Table S2). The *T. striatum* group also contains a large number of sequences associated to *Favia* sp. (Siboni et al. 2010). Other groups of Thraustochytriidae that contain coral-associated sequences include *Sicyoidochytrium*, Thraustochytriidae HK10, and Ulkenia. Roughly 4% of Thraustochytriidae sequences are isolated from invertebrates, including the quahog parasite QPX, which parasitizes clams, the abalone parasite *Labyrinthuloides haliotidis* (*L. haliotidis* group), and all the sequences in *T. caudivorum*, which belong to flatworm parasitic species.

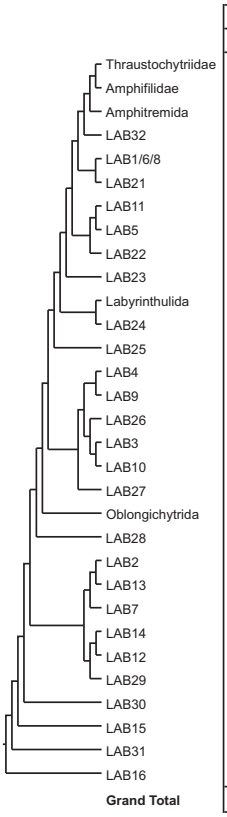
Estimating Labyrinthulomycetes relative abundance and richness from V9 data

While these analyses show that information about Labyrinthulomycetes exists in environmental survey data, HTES studies contain a great deal of more information that can be accessed by mapping their short reads onto the reference tree. We identified 520 OTUs corresponding to Labyrinthulomycetes and representing a total of 760,593 reads from VAMPS and Tara Oceans (Table S3). The original taxonomic assignments of these were based on PR2 and SILVA databases, but for both data sets, we found that the taxonomic identification based on the

Table 1. Comparison of taxonomic identification using the curated reference tree with automated identification

Tara Oceans (474 swarms)			VAMPS (2210 OTU_{99})		
Discarded 76	Others	25	Discarded 1,030	Others	862
	Labeled as Labyrinthulomycetes	51		Labeled Labyrinthulomycetes-Oomycetes	168
Included 398	Not labeled as Labyrinthulomycetes	7	Included 1,180	Not labeled as Labyrinthulomycetes-Oomycetes	3
	More accurate assignment	295		More accurate assignment	1,170
	Different subclades	62		Different subclades	5
	Same subclades	34		Same subclades	2

On the left are the Tara Oceans 'swarms' and on the right are VAMPS clusters identified as Labyrinthulomycetes. The upper row outlines numbers of sequences discarded as non-Labyrinthulomycetes, and numbers of Labyrinthulomycetes not identified. At the bottom, the specificity is compared, outlining how many Labyrinthulomycetes sequences were assigned to the wrong subgroup, or could be assigned to a narrower taxonomic group.



	GenBank		Tara Ocean		VAMPS	
	Abun.	Rich.	Abun.	Rich.	Abun.	Rich.
Thraustochytriidae	592	132	32737	52	78	25
Amphifilidae	69	42	132	7	96	28
Amphitremida	29	11	609	4	4	3
LAB32	0	0	5	1	0	0
LAB1/6/8	24	14	2128	20	110	13
LAB21	0	0	1990	2	3	1
LAB11	2	2	0	0	0	0
LAB5	3	2	307	2	2	1
LAB22	0	0	386	1	0	0
LAB23	0	0	3452	7	0	0
Labyrinthulida	252	59	137644	68	1115	64
LAB24	0	0	12107	4	3	1
LAB25	0	0	298	2	0	0
LAB4	8	5	752	7	0	0
LAB9	2	2	104	1	0	0
LAB26	0	0	162	2	0	0
LAB3	5	1	0	0	0	0
LAB10	2	1	0	0	0	0
LAB27	0	0	140	6	0	0
Oblongichytrida	78	24	3872	26	3014	58
LAB28	0	0	0	0	5	2
LAB2	12	8	908	5	0	0
LAB13	4	3	0	0	0	0
LAB7	27	5	106765	32	21	3
LAB14	31	7	386527	38	4	3
LAB12	2	2	5910	5	8	3
LAB29	0	0	0	0	17	1
LAB30	0	0	193	2	0	0
LAB15	31	7	58958	15	20	8
LAB31	0	0	7	1	0	0
LAB16	8	5	0	0	0	0
Grand Total	**1181**	**332**	**756093**	**310**	**4500**	**214**

Figure 3 Total abundance and richness for the three Labyrinthulomycetes databases. Numbers are calculated from raw data. Phylogenetic relationships among different groups are indicated by the dendrogram on the left. The topology of the cladogram differs from Fig. 1 tree because novel clades retrieved from the HTS have been added.

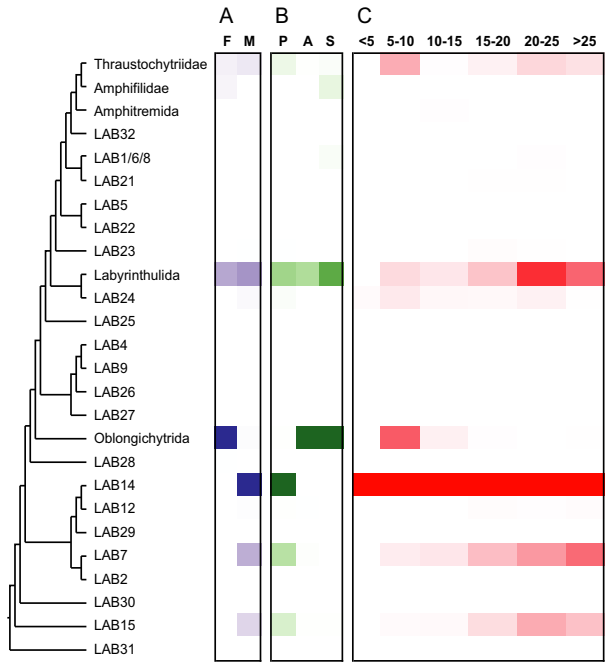


Figure 4 Heatmap showing the relative abundance distribution of major groups of Labyrinthulomycetes according to different environmental parameters (increasing color intensities indicate increasing relative abundances). (**A**) Freshwater and marine. (**B**) Photic water column, aphotic water column, and sediment. (**C**) Temperature range, in °C. F, freshwater; M, marine; P, photic zone; A, aphotic zone; S, sediment. For B & C, only marine samples were used since they were numerically dominant. The topology of the cladogram differs from Fig. 1 tree because novel clades retrieved from the HTS have been added.

curated reference data allowed for a significant improvement in the specificity of the assignment, and a reduction in the frequency of miss-assigned taxa (Table 1). For the Tara Ocean database alone, the improved annotation by our study has resulted in an over two-fold increase in the abundance of Labyrinthulomycetes identified in the collected samples (as compared with Database W6 in de Vargas et al. 2015).

Overall, 19 environmental clades were identified from the V9 dataset, seven of which were found to branch within previously defined lineages. Five of these belong to Thraustochytrida, one to Amphitremida, and one to Labyrinthulida (Table S3). The remainder were not found to belong to any previously identified lineages, and their phylogenetic positions within Labyrinthulomycetes can be seen in the dendrogram in Fig. 3 as well as in 4. From the HTES data, the abundance and richness of novel taxa is also evident, and some environmental clades surpass both the abundance and diversity of the better-studied groups. Analysis on the Tara Oceans database shows that LAB14, an environmental clade identified here for the first time, to be the most abundant of all the Labyrinthulomycetes clades, accounting for over 50% of the reads (Fig. 3). LAB7 and LAB15 also rank third and fourth in abundance, suggesting they too are ecologically significant but understudied. In the case of VAMPS, most of the reads belong to Oblongichytrida and Labyrinthulida (67% and 25%, respectively) (Fig. 3).

Based on the environmental information available for both the VAMPS and Tara Ocean databases, the abundance distribution was compared across the defined phylogenetic groups using different environmental parameters. Most of the 760,593 reads analyzed are derived from marine samples, with only 423 from freshwater. Oblongichytrida is the dominant group in freshwater, whereas LAB14 is the most abundant in marine data (Fig. 4A). Labyrinthulida are common in both freshwater and marine environments. Since the marine samples dominate the databases, we further analyzed marine metadata for depth (Fig. 4B) and temperature (Fig. 4C). Over 98% of Labyrinthulomycetes are recovered from the photic zone. Among the lineages compared, Labyrinthulida is the only one common in all three regions, even though it is not the most abundant (Fig. 4B). The most representative Labyrinthulomycetes subgroup in the photic zone is LAB14, whereas Oblongichytrida is the dominant taxa in the aphotic zone and the sediment. Although LAB14 dominates across all temperature ranges in the ocean (Fig. 4C), its dominance is less pronounced in warmer waters, and other taxa increase in abundance, such as Labyrinthulida,

LAB7, and LAB15. While Oblongichytrida is generally the minority under most temperature range, its abundance increases under 5–10 °C.

DISCUSSION

Phylogeny and classification

We have created a curated reference tree and database for Labyrinthulomycetes SSU rRNA, and used it to analyze the phylogeny, diversity, and distribution of the lineage. Overall, the topology of our reference phylogenetic tree is in agreement with several previous studies for the placement of most groups, like the sister placement of *Labyrinthula* and *Aplanochytrium* into Labyrinthulida, and the basal branching of Amphifilidae to Thraustochytriidae in the order Thraustochytrida. Our analyses also agree on the placement of Amphitremida deep within Labyrinthulomycetes and the placement of *Diplophrys* sister to *Amphitrema* and *Archerella* (Gomaa et al. 2013; Takahashi et al. 2014). The placement of *Oblongichytrium* outside of Thraustochytrida in this study is not very surprising, considering other studies have also shown similar results, even when different methods were used to obtain the phylogenetic tree (Anderson and Cavalier-Smith 2012; Collado-Mercado et al. 2010; Gomaa et al. 2013; Takahashi et al. 2014; Ueda et al. 2015; Yokoyama and Honda 2007; Yokoyama et al. 2007). Similar to these studies, members of the genus Thraustochytrium are also found scattered throughout Thraustochytriidae, probably due to previous misidentification.

Two genera of the Thraustochytriidae, *Althornia* and *Japonochytrium*, were not included in this study due to lack of publicly available sequences. The genus *Althornia* was named after isolation of *Althornia crouchii* Alderman and Jones (1971) from diseased oyster shells. According to Alderman and Jones, it is "a monocentric, biflagellate phycomycete with free-floating globose sporangia with a thick laminate wall" (Alderman and Jones 1971). It was placed in the Thraustochytriales based on a morphological study by Alderman et al. (Alderman et al. 1974). Since then, very little work has been done on this genus, and to date, no SSU rRNA gene sequence has been published. *Japonochytrium* was originally described by Kobayashi and Ookubo in 1953 for the species *Japonochytrium marinum*. Later, Harrison and Jones described the morphology and ultrastructure of a species that closely resembled *J. marinum*, denoted *Japonochytrium* sp. (Harrison and Jones 1974). The cultured strain ATCC28207 (sequence AB022104) was labeled as *J. marinum* by Tsui et al. (2009). However, this strain had previously been revised to *Ulkenia* sp. by Yokoyama et al. in 2007. AB022104 was clustered with AB022116 at 97%, and phylogenetic analyses also confirmed its placement within *Ulkenia* (97/1) (Fig. S1). It remains uncertain whether *Japonochytrium* is a real genus or misidentified *Ulkenia*. If future phylogenetic studies can confirm the placement of these two genera within Labyrinthulomycetes, it is then very possible that sequences belonging to them have already been included in our data.

In total, 39 new environmental clades were identified in our study. While most new clades branch well within Labyrinthulomycetes, the basal branching position of some of them in the majority of our analyses (e.g. LAB15 and LAB16), suggest that some of these lineages may represent sister lineages to the Labyrinthulomycetes sensu stricto; morphological observations would be required to tell if they match the descriptions applied to Labyrinthulomycetes or not. Additionally, 97% clustering of the hypervariable V9 data may have resulted in the identification of more novel clades than would be defined by full-length sequences. Nevertheless, the discovery of the new environmental clades, together with the fact that most of them are placed outside of Thraustochytrida, Amphitremida, and Labyrinthulida, shows the limitation of traditional culture-dependent approaches in uncovering the diversity of Labyrinthulomycetes and the benefit of large-scale environmental samplings.

Environmental distribution of Labyrinthulomycetes

Analyzing two HTES data sets revealed a number of interesting features of the abundance and distribution of various subgroups of Labyrinthulomycetes, as described above. Testing these will require identifying the target groups, in particular the more abundant groups currently comprised only of environmental sequences, but in the meantime, it is also useful to compare these results with what is known about Labyrinthulomycetes distribution in natural environments. Several studies have been carried out to investigate the presences, viability, and metabolic activity of Labyrinthulomycetes in deep-sea samples and under controlled deep-sea conditions. Using a combination of the AfDD staining technique and culturing methods, Raghukumar and collaborators were able to detect presence of thraustochytrids from water samples collected from the Arabian Sea up to 2,000 m in depth (Raghukumar et al. 2001). In another study, Raghukumar and Raghukumar (1999) demonstrated that during a 7-day incubation, thraustochytrids cultures were able to grow and maintain protease production and enzyme activity under 10 °C and 10 MPa . Riemann and Schaumann have also reported, using both AfDD staining and Nomarski microscopy, dense populations of thraustochytrids-like protists in a fast ice core drilled close to the southern shelf ice margin of the Weddell Sea. However, when these experiments were conducted, it was unclear which group of thraustochytrids (including *Oblongichytrium*) was being observed (Riemann and Schaumann 1993). Both of our reference and V9 databases contain sequences collected from the deep sea, and the most abundant are *Oblongichytrium* or Labyrinthulida (*Oblongichytrium* is the most abundant, followed by *Labyrinthula*).

Based on the environmental metadata, Labyrinthulomycetes are also relatively common in oxygen minimum zones (OMZs) and anoxic environments. Thraustochytrids have been reported from oxygen-limited environments, suggesting some members of Labyrinthulomycetes might be able to survive in anoxic habitats (Cathrine and Raghukumar 2009), perhaps by some facultative anaerobic metabolism.

Lastly, the association of Labyrinthulomycetes with other organisms is an obvious factor to consider (e.g. Raghukumar 2002; Raghukumar and Damare 2011), and our data shows that host-association likely evolved independently in many Labyrinthulomycetes lineages. Also, Labyrinthulomycetes from the same group can be associated with many, often very different, hosts. For example: *Labyrinthula* sp. have been isolated from the surface of seagrass and from the cytoplasm of single celled amoebozoan protists (Dyková et al. 2008).

A curated reference database increases the specificity of taxonomic assignments

One of the goals of a curated reference tree is that it improves the accuracy with which environmental sequences can be identified in two ways: sequences can be assigned to a lineage with lower levels of both false positive and negative identification, and sequences can be assigned to a lineage with a greater degree of specificity (i.e. they are more narrowly defined to a subgroup within that lineage). We compared taxonomic assignment of sequences from the two HTES data sets using our reference tree and the automated identification currently associated with the data sets (Table 1). In both cases using the reference tree, we identified putative false positive and negative identifications, as well as sequences assigned to the wrong subgroup of Labyrinthulomycetes, but all at relatively low frequencies. More commonly, however, the reference tree enabled sequences to be identified to a more narrowly defined taxonomic sub-group; in the case of Tara Oceans data, more than two-thirds of the swarms were assigned with greater specificity.

ACKNOWLEDGMENTS

This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (227301) and from a grant to the Centre for Microbial Diversity and Evolution from the Tula Foundation. PJK is a Senior Fellow of the Canadian Institute for Advanced Research. JdC was supported by a Marie Curie International Outgoing Fellowship grant (FP7-PEOPLE-2012-IOF – 331450 CAARL).

LITERATURE CITED

Alderman, D. J., Harrison, J. L., Bremer, G. B. & Jones, E. 1974. Taxonomic revisions in the marine biflagellate fungi: the ultrastructural evidence. *Mar. Biol.*, 25:345–357.

Alderman, D. J. & Jones, E. B. G. 1971. Physiological requirements of two marine phycomycetes, *Althornia crouchii* and *Ostracoblabe implexa*. *Trans. Br. Mycol. Soc.*, 57:213–225.

Altekar, G., Dwarkadas, S., Huelsenbeck, J. P. & Ronquist, F. 2004. Parallel metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics*, 20:407–415.

Anderson, O. R. & Cavalier-Smith, T. 2012. Ultrastructure of *Diplophrys parva*, a new small freshwater species, and a revised analysis of Labyrinthulea (Heterokonta). *Acta Protozool.*, 51:291–304.

Beakes, G. W., Honda, D. & Marco, T. 2014. 3 Systematics of the Straminipila: Labyrinthulomycota, Hyphochytriomycota, and Oomycota. Systematics and evolution, 2nd ed. Springer, Berlin Heidelberg. p. 39–97.

Berger, S. A., Krompass, D. & Stamatakis, A. 2011. Performance, accuracy, and web server for evolutionary placement of short sequence reads under maximum likelihood. *Syst. Biol.*, 60:291–302.

Bower, S. M. 1987. *Labyrinthuloides haliotidis* n. sp. (Protozoa: Labyrinthomorpha), a pathogenic parasite of small juvenile abalone in a British Columbia mariculture facility. *Can. J. Zool.*, 65:1996–2007.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. & Madden, T. L. 2009. BLAST+: architecture and applications. *BMC Bioinformatics*, 10:421.

del Campo, J. & Massana, R. 2011. Emerging diversity within Chrysophytes, Choanoflagellates and Bicosoecids based on molecular surveys. *Protist*, 162:435–448.

del Campo, J. & Ruiz-Trillo, I. 2013. Environmental survey meta-analysis reveals hidden diversity among unicellular Opisthokonts. *Mol. Biol. Evol.*, 30:802–805.

Capella-Gutierrez, S., Silla-Martinez, J. M. & Gabaldon, T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, 25:1972–1973.

Caporaso, J. G., Bittinger, K., Bushman, F. D., DeSantis, T. Z., Andersen, G. L. & Knight, R. 2010a. PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics*, 26:266–267.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Peña, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., Reeder, J., Sevinsky, J. R., Turnbaugh, P. J., Walters, W. A., Widmann, J., Yatsunenko, T., Zaneveld, J. & Knight, R. 2010b. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods*, 7:335–336.

Cathrine, S. J. & Raghukumar, C. 2009. Anaerobic denitrification in fungi from the coastal marine sediments off Goa, India. *Mycol. Res.*, 113:100–109.

Collado-Mercado, E., Radway, J. C. & Collier, J. L. 2010. Novel uncultivated labyrinthulomycetes revealed by 18S rDNA sequences from seawater and sediment samples. *Aquat. Microb. Ecol.*, 58:215–228.

Dyková, I., Fiala, I., Dvořáková, H. & Pecková, H. 2008. Living together: The marine amoeba *Thecamoeba hilla* Schaeffer, 1926 and its endosymbiont *Labyrinthula* sp. *Eur. J. Protistol.*, 44:308–316.

Edgar, R. C. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26:2460–2461.

Gomaa, F., Mitchell, E. A. D. & Lara, E. 2013. Amphitremida (Poche, 1913) is a new major, ubiquitous Labyrinthulomycete clade. *PLoS ONE*, 8:e53046.

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., de Vargas, C., Decelle, J., del Campo, J., Dolan, J. R., Dunthorn, M., Edvardsen, B., Holzmann, M., Kooistra, W. H. C. F., Lara, E., Le Bescot, N., Logares, R., Mahé , F., Massana, R., Montresor, M., Morard, R., Not, F., Pawlowski, J., Probert, I., Sauvadet, A.-L., Siano, R., Stoeck, T., Vaulot, D., Zimmermann, P. & Christen, R. 2012. The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Res.*, 41:D597–D604.

Harrison, J. L. & Jones, E. 1974. Ultrastructural aspects of the marine fungus *Japonochytrium* sp. *Arch. Microbiol.*, 96:305–317.
Honda, D., Yokochi, T., Nakahara, T., Raghukumar, S., Nakagiri, A., Schaumann, K. & Higashihara, T. 1999. Molecular phylogeny of labyrinthulids and thraustochytrids based on the sequencing of 18S ribosomal RNA gene. *J. Eukaryot. Microbiol.*, 46:637–647.
Huse, S. M., Dethlefsen, L., Huber, J. A., Welch, D. M., Relman, D. A. & Sogin, M. L. 2008. Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. *PLoS Genet.*, 4:e1000255.
Huse, S. M., Welch, D. & Voorhis, A. 2014. VAMPS: a website for visualization and analysis of microbial population structures. *BMC Bioinformatics*, 15:41.
Katoh, K. & Standley, D. M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.*, 30:772–780.
Leander, C. A. & Porter, D. 2001. The Labyrinthulomycota is comprised of three distinct lineages. *Mycologia*, 93:459–464.
Leander, C. A., Porter, D. & Leander, B. S. 2004. Comparative morphology and molecular phylogeny of aplanochytrids (Labyrinthulomycota). *Eur. J. Protistol.*, 40:317–328.
Lee Chang, K. J., Dunstan, G. A., Abell, G. C. J., Clementson, L. A., Blackburn, S. I., Nichols, P. D. & Koutoulis, A. 2012. Biodiscovery of new Australian thraustochytrids for production of biodiesel and long-chain omega-3 oils. *Appl. Microbiol. Biotechnol.*, 93:2215–2231.
Liu, Y., Singh, P., Sun, Y., Luan, S. & Wang, G. 2013. Culturable diversity and biochemical features of thraustochytrids from coastal waters of Southern China. *Appl. Microbiol. Biotechnol.*, 98:3241–3255.
Moss, S. T. 1985. An ultrastructural study of taxonomically significant characters of the Thraustochytriales and the Labyrinthulales. *Bot. J. Linn. Soc.*, 91:329–357.
Olive, L. S. 1975. The mycetozoans. Academic Press, New York. p. 215–241.
Porter, D. 1989. Phylum Labyrinthulomycota. Handbook of protoctista. Jones and Bartlett, Boston. p. 388–398.
Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. & Glockner, F. O. 2012. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.*, 41:D590–D596.
Raghukumar, S. 2002. Ecology of the marine protists, the Labyrinthulomycetes (Thraustochytrids and Labyrinthulids). *Eur. J. Protistol.*, 38:127–145.
Raghukumar, S. & Damare, V. S. 2011. Increasing evidence for the important role of Labyrinthulomycetes in marine ecosystems. *Bot. Mar.*, 54:3–11.
Raghukumar, S. & Raghukumar, C. 1999. Thraustochytrid fungoid protists in faecal pellets of the tunicate *Pegea confoederata*, their tolerance to deep-sea conditions and implication in degradation processes. *Mar. Ecol. Prog. Ser.*, 190:133–140.
Raghukumar, S., Ramaiah, N. & Raghukumar, C. 2001. Dynamics of thraustochytrid protists in the water column of the Arabian Sea. *Aquat. Microb. Ecol.*, 24:175–186.
Riemann, F. & Schaumann, K. 1993. Thraustochytrid protists in Antarctic fast ice? *Antarct. Sci.*, 5:279–280.
Ronquist, F. & Huelsenbeck, J. P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19:1572–1574.
Siboni, N., Rasoulouniriana, D., Ben-Dov, E., Kramarsky-Winter, E., Sivan, A., Loya, Y., Hoegh-Guldberg, O. & Kushmaro, A. 2010. Stramenopile microorganisms associated with the massive coral *Favia* sp. *J. Eukaryot. Microbiol.*, 57:236–244.
Sogin, M. L., Morrison, H. G., Huber, J. A., Mark Welch, D., Huse, S. M., Neal, P. R., Arrieta, J. M. & Herndl, G. J. 2006. Microbial diversity in the deep-sea and the underexplored "rare biosphere". *Proc. Natl Acad. Sci.*, 103:12115–12120.
Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30:1312–1313.
Takahashi, Y., Yoshida, M., Inouye, I. & Watanabe, M. M. 2014. *Diplophrys mutabilis* sp. nov., a new member of Labyrinthulomycetes from freshwater habitats. *Protist*, 165:50–65.
Tsui, C. K. M., Marshall, W., Yokoyama, R., Honda, D., Lippmeier, J. C., Craven, K. D., Peterson, P. D. & Berbee, M. L. 2009. Labyrinthulomycetes phylogeny and its implications for the evolutionary loss of chloroplasts and gain of ectoplasmic gliding. *Mol. Phylogenet. Evol.*, 50:129–140.
Ueda, M., Nomura, Y., Doi, K. & Nakajima, M. 2015. Seasonal dynamics of culturable thraustochytrids (Labyrinthulomycetes, stramenopiles) in estuarine and coastal waters. *Aquat. Microb. Ecol.*, 74:187–204.
de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Le Bescot, N., Probert, I., Carmichael, M., Poulain, J., Romac, S., Colin, S., Aury, J.-M., Bittner, L., Chaffron, S., Dunthorn, M., Engelen, S., Flegontova, O., Guidi, L., Horak, A., Jaillon, O., Lima-Mendez, G., Luke, J., Malviya, S., Morard, R., Mulot, M., Scalco, E., Siano, R., Vincent, F., Zingone, A., Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S., Acinas, S. G., Bork, P., Bowler, C., Gorsky, G., Grimsley, N., Hingamp, P., Iudicone, D., Not, F., Ogata, H., Pesant, S., Raes, J., Sieracki, M. E., Speich, S., Stemmann, L., Sunagawa, S., Weissenbach, J., Wincker, P. & Karsenti, E. 2015. Eukaryotic plankton diversity in the sunlit ocean. *Science*, 348:1261605.
Yokoyama, R. & Honda, D. 2007. Taxonomic rearrangement of the genus *Schizochytrium* sensu lato based on morphology, chemotaxonomic characteristics, and 18S rRNA gene phylogeny (Thraustochytriaceae, Labyrinthulomycetes): emendation for *Schizochytrium* and erection of *Aurantiochytrium* and *Oblongichytrium* gen. nov. *Mycoscience*, 48:199–211.
Yokoyama, R., Salleh, B. & Honda, D. 2007. Taxonomic rearrangement of the genus *Ulkenia* sensu lato based on morphology, chemotaxonomical characteristics, and 18S rRNA gene phylogeny (Thraustochytriaceae, Labyrinthulomycetes): emendation for *Ulkenia* and erection of *Botryochytrium*, *Parietichytrium*, and *Sicyoidochytrium* gen. nov. *Mycoscience*, 48:329–341.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article:

Figure S1. Diversity of Labyrinthulomycetes inferred from a maximum likelihood (RAxML) phylogenetic tree constructed using 18S rRNA sequences showing the position of all the represented OTUs.
Table S1. List of previously described environmental clades (Collado-Mercado et al. 2010) and the new clades they now belong to according to the present phylogenetic study.
Table S2. Labyrinthulomycetes GenBank sequences reference table.
Table S3. Labyrinthulomycetes V9 sequences reference table.