Molecular Phylogenetics and Evolution 190 (2024) 107964

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Phylogenomic position of genetically diverse phagotrophic stramenopile flagellates in the sediment-associated MAST-6 lineage and a potentially halotolerant placididean

Anna Cho [a,*], Denis V. Tikhonenkov [b], Gordon Lax [a], Kristina I. Prokina [b,c], Patrick J. Keeling [a]

[a] Department of Botany, University of British Columbia, Vancouver V6T 1Z4, British Columbia, Canada
[b] Papanin Institute for Biology of Inland Waters, Russian Academy of Science, Borok 152742, Russia
[c] Ecologie Systématique Evolution, CNRS, Université Paris-Saclay, AgroParisTech, Gif-sur-Yvette, France

ARTICLE INFO

Keywords:
MAST-6
Placididea
Bigyra
Stramenopile
Phylogenomics
Benthic protists

ABSTRACT

Unlike morphologically conspicuous ochrophytes, many flagellates belonging to basally branching stramenopiles are small and often overlooked. As a result, many of these lineages are known only through molecular surveys and identified as MArine STramenopiles (MAST), and remain largely uncharacterized at the cellular or genomic level. These likely phagotrophic flagellates are not only phylogenetically diverse, but also extremely abundant in some environments, making their characterization all the more important. MAST-6 is one example of a phylogenetically distinct group that has been known to be associated with sediments, but little else is known about it. Indeed, until the present study, only a single species from this group, *Pseudophyllomitus vesiculosus* (Pseudophyllomitidae), has been both formally described and associated with genomic information. Here, we describe four new species including two new genera of sediment-dwelling MAST-6, *Vomastramonas tehuelche* gen. et sp. nov., *Mastreximonas tlaamin* gen. et sp. nov., one undescribed *Pseudophyllomitus* sp., BSC2, and a new species belonging to Placididea, the potentially halotolerant *Haloplacidia sinai* sp. nov. We also provide two additional bikosian transcriptomes from a public culture collection, to allow for better phylogenetic reconstructions of deep-branching stramenopiles. With the SSU rRNA sequences of the new MAST-6 species, we investigate the phylogenetic diversity of the MAST-6 group and show a high relative abundance of MAST-6 related to *M. tlaamin* in samples across various depths and geographical locations. Using the new MAST-6 species, we also update the phylogenomic tree of stramenopiles, particularly focusing on the paraphyly of Bigyra.

1. Introduction

Stramenopiles are a diverse group of eukaryotes, including in their molecular sequence diversity, sizes, trophic modes, and morphologies. The best known are within a single subgroup, the Ochrophyta, which includes diatoms with diverse frustule shapes, microscopic phagotrophic flagellates that have lost photosynthesis (Dorrell et al., 2019; Kayama et al., 2020), and macroscopic multicellular brown algae like kelps. The diversity is less obvious at the morphological level in some basally branching groups of stramenopiles, but their molecular diversity is nonetheless significant. This is most obvious in the Bigyra Cavalier-Smith, 1998, which is a large assemblage composed of Sagenista Cavalier-Smith, 1995 and Opalozoa Cavalier-Smith, 1993. Other than saprotrophic Labyrinthulea (Sagenista), epiphytic *Solenicola setigera* (MAST-3), and symbiotic Opalinata (Opalozoa), the rest of the species of Bigyra are marine phagotrophic flagellates, generally in the size range of 2–10 μm (Gómez et al., 2011; Guillou et al., 1999; Lee, 2002; Moriya et al., 2002, 2000; Schoenle et al., 2022; Shiratori et al., 2017, 2015; Yubuki et al., 2015, 2010). These small and inconspicuous flagellates historically have been mistaken for cercozoans or discobans in light microscopy surveys (Larsen and Patterson, 1990; Lee, 2002; Patterson et al., 1993). Without detailed morphological examination and molecular surveys, it is difficult to discern among flagellated species of Bigyra, or indeed even between members of the major subdivisions, leading to both under-sampling and under-estimation of their diversity.

The diversity and abundance of Bigyra has accordingly been determined by molecular surveys, and these have revealed a number lineages without any morphological identity. They are simply referred to as

* Corresponding author.
E-mail address: acho@mail.ubc.ca (A. Cho).

https://doi.org/10.1016/j.ympev.2023.107964
Received 19 May 2023; Received in revised form 2 November 2023; Accepted 8 November 2023
Available online 10 November 2023

MArine Stramenopiles (MASTs), a term that was originally coined to include 18 uncharacterized lineages, most of which are still only known from small subunit (SSU) ribosomal RNA (rRNA) sequences identified from environmental sampling efforts (Logares et al., 2012; Massana et al., 2004, 2006, 2009, 2014). These surveys also showed significant differences in community composition among different environments, particularly between benthic and pelagic samples (Forster et al., 2016; Massana et al., 2015). Notably, MAST-6 (along with MAST-9, and -12) are common in sediments while rare in pelagic samples (Logares et al., 2012; Massana et al., 2015; Rodríguez-Martínez et al., 2020). Extensive ultrastructural and cellular examination of the first cultured MAST-6 lineage, *Pseudophyllomitus vesiculosus* (family Pseudophyllomitidae Shiratori et al., 2017), describe it as a relatively large algivore (dimensions up to 18.3 × 12.4 μm) with a flagellar apparatus ultrastructure characteristic of deep-branching stramenopiles (e.g., no x-fiber, R2 flagellar root with 13 microtubules) (Lee and Patterson, 2002; Moestrup and Thomsen, 1976; Yubuki et al., 2010; Shiratori et al., 2017). The genus *Pseudophyllomitus* (Lee, 2002) was erected to describe *Phyllomitus*-like taxa without two adhering flagella. This resulted in re-designation of four species (e.g., *Pseudophyllomitus apiculatus, P. granulatus, P. salinus*, and *P. vesiculosus*), however with no molecular and ultrastructural data available to indicate the monophyly of the genus. Later, a new family Pseudophyllomitidae (Shiratori et al., 2017) was erected, apparently corresponding to a MAST-6 clade however, with unknown phylogenetic position of the type species, *P. granulatus*. As a result, we refrain from using Pseudophyllomitidae in replacement of MAST-6 hereafter.

In multi-gene analyses, MAST-6 are closely related to many ecologically important groups such as MAST-4 (Cho et al., 2022; Thakur et al., 2019) which is one of the most common heterotrophic flagellate groups in coastal ecosystems, significantly affecting microbial food webs (Logares et al., 2012; Massana et al., 2006; Rodríguez-Martínez et al., 2009). Along with MAST-7, -8, -9, and MAST-11, MAST-4 and -6 form a clade known as Eogyrea (Cavalier-Smith and Scoble, 2013) which was formerly described as clade L that only contained environmental sequences (Cavalier-Smith and Chao, 2006). Previous phylogenomic analyses showed the monophyly of MAST-4 and MAST-6, forming a sister group to Labyrinthulea, a detrital decomposer that is also abundant in sediments (Collado-Mercado et al., 2010; Massana et al., 2015; Nakai and Naganuma, 2015; Rodríguez-Martínez et al., 2020). The diversity of MAST-6 has been demonstrated by SSU rRNA amplicon sequencing of various sediment studies from surface to deep-sea push-core sediments (Massana et al., 2015; Rodríguez-Martínez et al., 2020; Schoenle et al., 2021). These studies showed high relative abundance of MAST-6 in sediments and estimated up to 46 Operational Taxonomic Units (OTUs) assigned to MAST-6, with more than 1000 sequencing reads (Rodríguez-Martínez et al., 2020). Interestingly, two morphotypes of MAST-6 were observed in plankton samples that differ in sizes and seasonal abundances (Piwosz and Pernthaler, 2010), with the larger morphoptype (9.9–22 μm in length) showing rapid increase in abundance only for a week. This study showed MAST-6 is not only phylogenetically diverse, but also that community composition can respond quickly to fluctuating environments and food availability. Despite these advances, however, only a single transcriptome is available for MAST-6 (Shiratori et al., 2017; Thakur et al., 2019), limiting our understanding of its biology and its relationship to other stramenopiles, and character evolution of Sagenista.

Another major but under-sampled subgroup of Bigyra is the Placidida, a class within Opalozoa. Like Sagenista, the phylogenetic diversity of Placidida is largely represented by SSU rRNA genes, with -omic data available for only two species: *Wobblia lunata* (Moriya et al., 2000; Thakur et al., 2019) and *Placididea* sp. (the CaronLab strain formerly mis-labelled as '*Cafeteria*') (Keeling et al., 2014). Placidideans are closely related to MAST-3, another abundant and highly diverse heterotrophic flagellate group that plays an important role in marine food webs and, found in all coasts and open oceans around the world (Gómez et al., 2011; Logares et al., 2012; Massana et al., 2004). Unlike the MAST clades, SSU sequences of Placididea (Moriya et al., 2002) do not amplify well with V4-targeting primers (Lee et al., 2022). Consequently, the diversity of this group was reported either using V9-targeting SSU primers (Choi and Park, 2020; Lee et al., 2021) or by individually isolating placidideans (Moriya et al., 2002; Park and Simpson, 2010; Rybarski et al., 2021). Isolated placidideans are often from hypersaline environments (>40 ‰), however, many characterized halophilic placidideans can tolerate lower salinity (Park and Simpson, 2010; Rybarski et al., 2021), raising the possibility that they are also present in non-hypersaline environments.

In this study, we describe three new MAST-6 species and two new genera, and one new Placididea species, providing microscopic observations, transcriptomic data, and SSU rRNA sequence comparisons with previously generated environmental amplicon data. Strains PhM-7 (Placididea, *Haloplacidia sinai*) and Colp-33 (MAST-6, *Vomastramonas tehuelche*) were maintained in culture for a year, but subsequently lost. Two other MAST-6 species (*Mastreximonas tlaamin* and *Pseudophyllomitus* sp. BSC2) were obtained by single cell isolations. We also report transcriptomes from two cultured species, *Symbiomonas scintillans* RCC257 (Guillou et al., 1999) and *Caecitellus* sp. RCC1078 (O'Kelly and Nerad, 1998), to further fill out the diversity of deep-branching stramenopiles for phylogenomic analyses. We report the relative abundance and diversity of the four new species of MAST-6 and Placididea in publicly available environmental sequence surveys, and re-examine stramenopile phylogeny, particularly with the aim of resolving relationships within the Bigyra using a multi-gene approach.

2. Materials and methods

2.1. Sample collection and imaging

Strain PhM-7 (*Haloplacidia sinai* sp. nov.) was isolated from the Red Sea (average salinity 36–41 ‰), Sharm El Sheikh, Egypt (27°50′50.5″ N, 34°18′59.4″ E), scraped from coral at 75 m depth in April 2015. Strain Colp-33 (*Vomastramonas tehuelche* gen. et sp. nov.) was isolated from nearshore bottom sediments, Chile, Punta Arenas (53°37′49″ S, 70°56′58″ W, T = 9.4 °C, Salinity 24 ‰) in November 2015. These strains were propagated in a predator–prey culture with the bodonid *Procryptobia sorokini* as a steady food source but both perished after a year of cultivation. Light microscopy observations for PhM-7 and Colp-33 were made using a Zeiss AxioScopeA.1 equipped with phase contrast and DIC water immersion objectives (63x) and an AVT HORN MC-1009/S analog video camera. For scanning electron microscopy (SEM) imaging of PhM-7, cells from exponential growth phase were fixed at 22 °C for 10 min in a cocktail of 0.6 % glutaraldehyde and 2 % OsO_4 (final concentration) prepared using a 0.1 M cacodylate buffer (pH 7.2), and gently drawn onto a polycarbonate filter (diameter 24 mm, pores 0.8 μm). Following filtration, the specimen was taken through a graded ethanol dehydration and acetone, and critical-point dried. Then dry filters with fixed specimens were mounted on aluminum stubs, coated with gold–palladium, and observed with a JSM-6510LV scanning electron microscope (JEOL, Japan).

Two uncultured single cells, PRC5 (*Mastreximonas tlaamin* gen. et sp. nov.) and BSC2 (*Pseudophyllomitus* sp. BSC2), were isolated from oxic marine intertidal sediment. Sediment for PRC5 was collected from Powell River, British Columbia, Canada (49°50′42″ N, 124°31′60″ W) in August 2020; whereas the BSC2 sample was collected from Boka Santa Cruz, Curaçao (12°18′24″ N, 69°8′44″ W) in April 2022. Both cells were manually isolated using a drawn-out glass micropipette under a Leica DLIM inverted microscope and imaged with a Sony α7rIII camera. The cells were rinsed twice in filtered sea water and transferred into a 0.2 mL PCR tube containing lysis buffer (Picelli et al., 2014) and stored in −80 °C until cDNA synthesis.

Cultures of *Symbiomonas scintillans* strain RCC257 and *Caecitellus* sp. strain RCC1078 were obtained from the Roscoff culture collection

(France) in March 2022. The cultures were grown in 30 mL of 0.22 μm filtered f/2 medium (30 ‰) and autoclaved seawater (30 ‰), respectively, both with an autoclaved rice grain added. The cultures were kept in a 20 °C incubator with a 12 h:12 h light:dark cycle and sub-cultured every two weeks.

2.2. cDNA synthesis, library preparation and sequencing

Cells of PhM-7 (*H. sinai*) and Colp-33 (*V. tehuelche*) grown in clonal cultures were harvested when the cells had reached peak abundance and after most of the prey had been eaten. The cells were collected by centrifugation (2,000*g* for PhM-7 and 1,000*g* for Colp-33, both at room temperature) onto the 0.8 μm membrane of a Vivaclear mini column (Sartorium Stedim Biotech Gmng, Cat. No. VK01P042). Total RNA was then extracted using a RNAqueous-Micro Kit (Invitrogen, Cat. No. AM1931). In addition to the RNA extraction from the Colp-33 clonal cultures, 20 single cells were manually picked from its culture using a glass micropipette and transferred into a 0.2 mL PCR tube containing the cell lysis buffer for an additional Smart-Seq2 cDNA synthesis and library preparation (see below).

For cultures obtained from the Roscoff Culture Collection (RCC257 and RCC1078), TRIzol™ LS Reagent was used to extract total RNA, following the manufacturer's instructions with a modification at the aqueous-organic layer separation step. Briefly, 100 mL of each culture was centrifuged at 3220*g* for 20 min at 4 °C to pellet cells at the bottom of the centrifuge tubes. After carefully discarding the media, 1 mL of TRIzol™ LS was added to the pelleted cells. For an easier transfer of the aqueous phase containing the RNA without an interphase contaminant, the aqueous-organic layer separation by chloroform was done in Phasemaker™ (Invitrogen) tubes. The quality and quantity of the RNA yield was determined using a NanoDrop 1000 Spectophotometer v3.8.1 (Thermo Fisher Scientific). Additionally, using glass micropipettes, approximately 20 cells were manually isolated from each culture and processed in the same manner as the single-cell isolation method used for Colp-33 (*V. tehuelche*), PRC5, and BSC2.

For cDNA synthesis, the poly-A selection based Smart-Seq2 protocol was used (Picelli et al., 2014). For manually isolated single cells in the lysis buffer, 2–3 rounds of freeze–thaw steps were included prior to the cDNA synthesis (Onsbring et al., 2020). For RNA extracts, 4 μL of the extract was used for cDNA synthesis. The rest of the library preparation and sequencing steps (tagmentation, quality control, and adaptor ligation) for PRC5, BSC2, RCC257 and RCC1078 were carried out by the Sequencing and Bioinformatics Consortium (University of British Columbia, BC Canada), using the Illumina Nextera™ DNA Flex Library Preparation Kit. The sequencing was performed on a NextSeq (mid-output) platform with 150 bp paired-end library constructs. For PhM-7 and Colp-33, the libraries were prepared using Nextera™ XT DNA Library Preparation Kit (Illumina, Inc., Cat. # FC-131-1024) followed by Illumina Miseq 300 bp paired-end sequencing at GenoSeq, Sequencing & Genotyping Core (University of California Los Angeles, CA USA) for PhM-7, and Sequencing and Bioinformatics Consortium (University of British Columbia, BC Canada) for Colp-33. All the raw reads of the transcriptomes are deposited in the NCBI Short Read Archive (SRA) under the BioProject number PRJNA961826 (SRR24392492 to SRR24392501).

2.3. Transcriptome processing, assembly, and decontamination

Along with the six newly generated transcriptomes in this study, recently published transcriptomes of *Actinophrys sol* (Azuma et al., 2022) and its prey, *Chlorogonium capillatum*, were processed as follows. The quality of the raw sequencing reads was assessed using FastQC v0.11.9 (Andrews, 2010). To correct random sequencing errors of the short Illumina RNA-seq reads, *k*-mer based Rcorrector (version 3) was used on the raw reads (Song and Florea, 2015). The error-corrected reads were then trimmed using Trimmomatic v0.39 (Bolger et al., 2014) to remove remnant transposase-inserts from the library preparation, Nextera™ DNA Flex adaptors, low quality reads (-phred33), and Smart-Seq2 IS-primers with the leading and trailing cut-off at 3, SLIDINGWINDOW:4:15, and MINLEN:36. Processed forward, reverse, and unpaired transcripts were assembled using the *de novo* transcriptome assembler rnaSPAdes v3.15.1 (Bushmanova et al., 2019). Additionally, for species with two libraries prepared from both RNA extract and single cell isolations (i.e., Colp-33, RCC257, and RCC1078), the resulting transcripts were co-assembled. BlobTools v2.3.3 (Challis et al., 2020; Laetsch and Blaxter, 2017) was used to identify contaminants and visualize contig coverage. In short, megaBLAST was used to search assembled transcripts against the NCBI nucleotide database followed by a diamond BLASTX (Altschul et al., 1990, Buchfink et al., 2014) protein search against the UniProt reference database (Bateman et al., 2021; Buchfink et al., 2014). Both searches were performed with an e-value cut-off 1e-25. Bacterial, Viriplantae, metazoan, and archaeal reads were removed from all transcripts. To remove prey contaminants from PhM-7, Colp-33, and *A. sol*, the assembled transcripts were first searched against the transcriptome of the respective prey (*Procryptobia sorokini* for PhM-7 and Colp-33; and *C. capillatum* for *A. sol*) using BLASTn, followed by the removal of contigs with $\geq$ 95 % sequence identity. TransDecoder v5.5.0 (Haas, 2015) was used to predict open reading frames (ORFs) and the longest ORFs were annotated using a BLASTP search against UniProt database with the e-value cut-off 1e-5. BUSCO v5.2.2 (Simão et al., 2015) with 'stramenopiles_odb10′ database was used to assess the completeness of each transcriptome.

2.4. Small subunit sequences and amplicon processing using QIIME 2

Small subunit (SSU) rRNA sequences were extracted from PRC5 and BSC2 transcriptomes using barrnap v0.9 (Seemann, 2007). For *S. scintillans* and *Caecitellus* sp., SSU rRNA sequences were generated by polymerase chain reaction (PCR) amplification of cDNA using 18SFU and 18SRU eukaryotic primers (Tikhonenkov et al., 2016), followed by Sanger dideoxy sequencing. Although the SSU sequences for *S. scintillans* RCC257 and *Caecitellus* sp. and RCC1078 are available in GenBank, we did SSU PCR to confirm the species identity and to obtain longer sequences as the published *S. scintillans* RCC257 (accession KT861043) SSU is 760 bp. For all the downstream analyses, we used the newly obtained SSU sequences for these two cultured bikosia.

To obtain SSU rRNA sequences of Colp-33 and PhM-7, the cells were first harvested when the cultures had reached peak abundance and after the prey had been eaten (confirmed with light microscopy), followed by centrifugation (7000*g*, room temperature) onto an 0.8 μm membrane of a Vivaclear mini column (Sartorius Stedim Biotech Gmng, Cat. No. VK01P042). Total DNA was extracted from the filters using the MasterPure Complete DNA and RNA Purification Kit (Epicentre, Cat. No. MC85200). The SSU rRNA genes were PCR-amplified using the general eukaryotic primers EukA-EukB for strain Colp-33 (Medlin et al., 1988), and GGF-GGR for strain PhM-7 (Tikhonenkov et al., 2022). PCR products were subsequently cloned prior to sequencing (PhM-7) or sequenced directly (Colp-33), using Sanger dideoxy sequencing with two additional internal sequencing primers 18SintF and 18SintR (Tikhonenkov et al., 2022). All the SSU rRNA sequences from the four newly described species and two culture strains are deposited in GenBank with the accession OQ909082-OQ909087.

To compare SSU rRNA sequences of newly identified species to previously reported studies, five sediment datasets were obtained via The European Nucleotide Archive (ENA). The datasets are designated as follows: BioMarKs (Dunthorn et al., 2014; Massana et al., 2015), SouthChina (Wu and Huang, 2019), Norway (unpublished BioProjects PRJEB24876; PRJEB24158; PRJEB24888), Deepsea (Schoenle et al., 2021), and ISME2020 (Rodríguez-Martínez et al., 2020) (Table 1). For the sixth dataset (designated as ESBig), we obtained ten SSU rRNA sequences (ESBig130-139) assigned to Placididea, directly from the authors (Lee et al., 2022) (Table 1). These studies examined sediments

Table 1
List of selected amplicon sequencing datasets from the European Nucleotide Archive (ENA). Accession numbers are included only if selected samples of a given BioProject were analyzed. For example, out of 139 samples for the BioMarks dataset, only sediment samples (24) were processed to access the diversity of the newly identified MAST-6 species (*). Extracted length indicates the length of the amplicon sequence variants (ASVs) assigned to MAST-6 or Placididea lineages. The ESBig dataset was not processed in this study, but the sequences assigned to placidideans were directly obtained from the authors of the BioProject (**). All SouthChina ASVs and one Deepsea_MAST6 ASV were excluded from the main figures based on LWR-values and manual blastn searches (†).

Dataset designation	Sample environment	Sequencing technology	Targeted 18S rRNA region	Extracted length (bp)	Number of ASVs		Sample number	BioProject
					Placididea lineage	MAST-6 lineage		
BioMarKs*	Seafloor sediment	454 GS FLX Titanium	V4	380–384	0	12	24* (Run accessions: ERR861806-ERR861811, ERR861839, ERR861843, ERR861849, ERR861853, ERR861860, ERR861870, ERR861884, ERR861885, ERR861894, ERR861895, ERR861900, ERR861901, ERR861905, ERR861910, ERR861911, ERR861915-ERR861917)	PRJEB9133 (Dunthorn et al., 2014; Massana et al., 2015)
SouthChina	Seafloor sediment	454 GS FLX Titanium	V1-V2	396–429	0	9†	6	PRJNA341446 (Wu and Huang, 2019)
Norway	Marine and brackish sediment	Illumina MiSeq paired-end	V4	426–429	0	16	24	PRJEB24876, PRJEB24158, PRJEB24888
Deepsea	Abyssal seafloor sediment	Illumina Genome Analyzer II paired-end	V9	134–138	15	6†	20	PRJNA635512 (Schoenle et al., 2021)
ISME2020	Seafloor sediment	Illumina MiSeq paired-end	V4	182–425	0	61	49	PRJNA521526 (Rodríguez-Martínez et al., 2020)
ESBig**	Solar saltern	Illumina MiSeq paired-end	V9**	128–154	10**	0	**Accession number: MZ297173, MZ297191, MZ299824, MZ299825, MZ299969, MZ300048, MZ300314, MZ300350, MZ300439, MZ300768,	**PRJNA732544 (Lee et al., 2021)

from different bodies of water across the US, Europe, and Asia, including the South China Sea, North Atlantic Ocean, Mariana Basin, Philippine Basin, Bunnefjorden (Norway), Pacific Ocean, and the freshwater lake Pollevann (Norway). The depths of the sample sites vary from 20 m to 5497 m, and cover diverse marine, brackish and freshwater environments such as push-cores or surface sediments of seafloors, fjords, abyssal plains, and continental rises. Except for ESBig, all datasets were processed using QIIME 2 (q2cli v2020.11.1) (Bolyen et al., 2018). For the 454 pyrosequencing data (BioMarKs and SouthChina), the raw reads were imported and demultiplexed with '–type SampleData[SequencesWithQuality]' and '–input-format SingleEndFastqManifestPhred33′ options. After trimming the raw reads with respective primer-pair sequences (Martin, 2011), both 454 pyrosequencing and Illumina sequencing data were filtered with a DADA2 denoising step (Callahan et al., 2016). To remove chimeric sequences, denoised sequences were further processed with 'uchime-denovo' (Rognes et al., 2016). For taxonomic classification of amplicon sequence variants (ASVs), a QIIME 2 compatible PR2 v4.14.0 dataset was obtained (del Campo. et al., 2018; Guillou et al., 2013) and modified by manually adding the SSU rRNA sequences of the new species described from this study and relevant sequences from the recent PR2 database and GenBank (Guillou et al., 2013; Park and Simpson, 2010; Rybarski et al., 2021). The modified PR2 dataset was used to pre-train the QIIME 2 classifier using 'qiime feature-classifier fit-classifier-naïve-bayes' (Pedregosa et al., 2011). The trained classifier was then used to assign taxonomy to filtered representative amplicon sequence variants (ASVs). Amplicon sequence variants assigned to MAST-6 and Placididea were extracted and added to a stramenopile SSU rDNA alignment consisting of partial to nearly full-length sequences (Cho et al., 2022; Yubuki et al., 2015). Additionally, relevant environmental sequences from the PR2 database, GenBank, and 10 placididean-associated operational taxonomic units (OTUs) from ESBig (Lee et al., 2021) were added. The extracted feature sequences were further subjected to CD-HIT to remove duplicates (Li and Godzik, 2006).

To check presence and visualize relative abundance of newly acquired MAST-6 and Placididea species in the amplicon dataset (Table 1), feature tables from QIIME2 were exported and processed in RStudio (R v4.2.0) (R Team, 2016) with ggplot2 (Wickham, 2016).

2.5. Small-subunit (SSU) rRNA gene tree construction

Without adding the ASVs extracted from the sediment dataset, the compiled SSU rRNA sequences were aligned with MAFFT v7.481 (Katoh and Standley, 2013) and trimmed using trimAl 1.2rev59 (-gt 0.3, -st 0.001) (Capella-Gutiérrez et al., 2009). The resulting trimmed alignment was then combined with the extracted ASVs and realigned as described above (8,771 sites), followed by maximum likelihood inference using RAxML v8.2.12 (Stamatakis, 2014) under the GTRGAMMA model with 1000 replicates of ultrafast bootstrap (UFB). To further evaluate the phylogenetic placement of short amplicon sequences from the amplicon datasets (Table 1), additional phylogenetic supports were estimated using the Evolutionary Placement Algorithm (EPA) (Berger et al., 2011) with EPA-ng v0.3.8 (Barbera et al., 2019). This method used the reference ML tree constructed with the same conditions as above with partial to nearly full-length SSU rRNA sequences. The reference alignment file was generated using MAFFT followed by trimAl (-gt 0.3, -st 0.001), resulting in 8,871 sites. To determine the placement probability of each amplicon sequence variant (ASVs) assigned to MAST-6 or Placididea, a likelihood weight ratio (LWR) was determined using GAPPA (Czech et al., 2020). Likelihood weight ratios (LWRs) are values that represent the probability of an ASV placement in a given branch, across the tree. The ASVs with an LWR value higher than 95 % were

inspected for chimerism using BLASTn and passing sequences were considered high confidence (Dunthorn et al., 2014). The SSU rRNA tree with EPA analysis is hereafter referred as SSU-EPA tree.

To evaluate phylogenetic relationships of newly added MAST-6, placididean, and other species of Bigyra, another SSU rRNA phylogenetic tree was constructed without short amplicon sequences, hereafter referred as SSU-tree. A total of 224 SSU rRNA sequences ≥ 900 bp consisting of previously compiled datasets and new sequences (Aleoshin et al., 2016; Cho et al., 2022; Rybarski et al., 2021; Yubuki et al., 2015) were aligned using MAFFT v7.481 (Katoh and Standley, 2013), followed by trimming using trimAl 1.2rev59 (-gt 0.3, -st 0.001) (Capella-Gutiérrez et al., 2009). The phylogenetic tree was then constructed based on 1649 sites using IQ-TREE v2.1.0 (Minh et al., 2020) under TIM2 + F + R6, determined with ModelFinder (Kalyaanamoorthy et al., 2017) and 1000 UFB.

2.6. Phylogenomic matrix construction using PhyloFisher

The phylogenomic matrix including the predicted proteins of the newly produced transcriptomes was generated using PhyloFisher v1.1.2 (Tice et al., 2021). Briefly, annotated ORFs from the newly generated transcriptomes were searched against the 241 gene set embedded in PhyloFisher and the resulting homologs were then added to each of the gene alignments. For each of the updated 241 gene alignments, a single-gene tree was constructed using IQ-TREE v1.6.12 (Nguyen et al., 2015) under the L + G4 + X model and 1000 UFB. Each single-gene tree was manually screened using ParaSorter v1.0.4 to ensure correct orthologs were inferred from the newly added proteins. Predicted orthologs of recently published or relevant stramenopiles (Azuma et al., 2022; Keeling et al., 2014; Cho et al., 2022; Thakur et al., 2019; Richter et al., 2022) were kept. To generate a final concatenated phylogenomic matrix, 98 taxa (including 15 taxa for an outgroup) were selected, resulting in a 240 gene set with 76,516 amino acid (aa) sites. Beside the main concatenated matrix, two additional concatenated matrices were generated to evaluate the effects of ortholog completeness in determining the phylogeny: One that included only orthologs found in ≥ 39 % of taxa (233 orthologs with 74,531 aa sites; referred to as 39per-matrix), and another that included orthologs found in ≥ 59 % of taxa (215 orthologs with 67,630 aa sites; referred to as 59per-matrix). Additionally, we generated another matrix with the most recent genomic data of other MAST lineages (MAST-1, MAST-7, MAST-8, MAST-9, and MAST-11) (Labarre et al., 2021; Richter et al., 2022) with 74,898 aa sites (234 orthologs) composed of 104 taxa (hereafter, referred as MASTer-matrix).

2.7. Phylogenomic tree reconstruction, removal of fast-evolving sites, and recoding

The initial maximum likelihood (ML) tree of the main concatenated phylogenomic matrix was inferred using IQ-TREE v2.1.2 under the empirical profile mixture model LG + C60 + F + G4 (Quang et al., 2008) with 1000 UFB. The resulting ML tree was used as a guide to estimate posterior mean site frequencies (PMSF) (Wang et al., 2018), which was then used to re-estimate a final ML-PMSF tree with 100 non-parametric standard bootstraps under the same model. The construction of the ML-PMSF phylogenomic tree was repeated with the 39per- and 59per-matrices using the same model with 100 non-parametric standard bootstraps. To consider the effect of fast-evolving sites on tree topology, the main concatenated matrix was further subjected to a stepwise 10,000 aa site removal using PhyloFisher (fast_site_removal.py) followed by construction of ML-PMSF trees. For potential amino acid composition bias in the dataset, web-based Composition Profiler (Vacic et al., 2007) was used with default settings to compare relative abundances of GARP vs. FYMINK amino acids with “SwissProt 51” (Bairoch et al., 2005) as a background, in addition to examining a distance matrix tree output generated by ‘aa_comp_calculator.py’ in PhyloFisher. To remove potential amino acid composition bias, the main concatenated matrix was recoded with the Dayhoff 18 (Dayhoff et al., 1978; Hernandez and Ryan, 2021; Wang et al., 2008) option using PhyloFisher v1.2.4 (aa_recode.py) followed by a tree reconstruction under the MULTI18_GTR + FO model and 100 replicates of standard bootstrap with RAxML-NG v.1.1.0 (Kozlov et al., 2019).

To infer a phylogenomic tree using Bayesian estimation, the CAT-GTR mixture model was used with the -dgm 4 option in PhyloBayes-MPI v4.0.3 (Lartillot et al., 2009; Lartillot and Philippe, 2004). Four independent Markov Chain Monte Carlo (MCMC) chains were run in parallel for at least 10,000 generations. The consensus posterior probability and topology were estimated after discarding the first 20 % as burn-in and subsampling every second tree. Convergence of the four chains was tested with bpcomp.

3. Results

3.1. Phylogenomic tree of stramenopiles

The final phylogenomic matrix used for constructing the main phylogenomic tree was a concatenated alignment of 240 genes (76,517 sites) and 98 taxa (including 15 taxa belonging to an outgroup). The average percentage of genes present for each included transcriptome was 71.6 %, with 76.4 % of sites covered. These values were comparatively lower in the newly added transcriptomes: 8.5 % genes and 16.3 % sites for *Pseudophyllomitus* sp. BSC2; 21.1 % genes and 25 % sites for *Mastreximonas tlaamin*; 38.6 % genes and 52.5 % sites for *Vomastramonas tehuelche*; 38.2 % genes and 47.5 % sites for *Haloplacidia sinai*; 21.42 % genes and 35.42 % sites for *Caecitellus* sp.; 42.7 % genes and 58.3 % sites for *Symbiomonas scintillans* (Fig. 1). The BUSCO scores showed a similar pattern where *Pseudophyllomitus* sp. BSC2 and *M. tlaamin* had the lowest values (4 %:4% and 8 %:2% completed:fragmented) while *V. tehuelche* and *H. sinai* had 28 %:19 % and 26 %:16 % and, *S. scintillans* and *Caecitellus* sp. 43 %:10 % and 11 %:7%, respectively.

Based on the main phylogenomic tree inferred from ML analysis under LG + C60 + F + G4 + PMSF, Gyrista was monophyletic and Bigyra was paraphyletic (Fig. 1). Within Gyrista, Ochrophytes and Pseudofungi are monophyletic with strong support. In Ochrophyta, the Raphidophyceae, Phaeophyceae, and Xanthophyceae (RPX) clade formed a monophyletic group with the Chrysophyceae, Synurophyceae, and Synchromophyceae + Pinguiophyceae clade (CSS + Pi) with moderate bootstrap support (84 %). Bacillariophyceae + Bolidophyceae + Pelagophyceae and Dictyochophyceae (BBDPe) formed a fully supported clade. The monophyly of RPX, CCS + Pi, and BBDPe was moderately supported (84 %). However, phylogenetically unstable Eustigmatophyceae formed a weakly supported (71 %) clade with *Actinophrys sol*, a non-photosynthetic heliozoan stramenopile. The Eustigmatophyceae + *A. sol* clade branched sister to the rest of the Ochrophyta. In the ML-PMSF trees inferred for the MASTer, 39per- and 59per-matices, *A. sol* branched sister to CSS + Pi while Eustigmatophyceae branched with RPX with moderate support (81 % to 94 %, and 76 % to 89 %, respectively) (Fig. S1-2).

Within Bigyra, the three new MAST-6 species, *M. tlaamin*, *Pseudophyllomitus* sp. BSC2 and *V. tehuelche* formed a clade with *Pseudophyllomitus vesiculosus,* with *Pseudophyllomitus* sp. BSC2 being the immediate sister lineage to *P. vesiculosus*. MAST-6, MAST-4, and Labyrinthulea all formed a monophyletic group, Sagenista. In the tree inferred for the MASTer-matrix (Fig. S2), MAST-7 and MAST-11 formed a robust clade, which then branched sister to MAST-4. MAST-8 and MAST-9 formed close relationship with the grouping consisting of MAST-4, MAST-7, and MAST-11. MAST-6 in turn formed a robust monophyly with this grouping, composed of MAST-4, MAST-7, −8, −9, and −11. The new Placididea species, *H. sinai,* is closest to Placididea sp. (Caron Lab) and, together with *Wobblia lunata,* formed the monophyletic group Placididea. Placididea formed a sister lineage to the rest of the Placidozoa (Nanomonadea and Opalinata), however, the support value

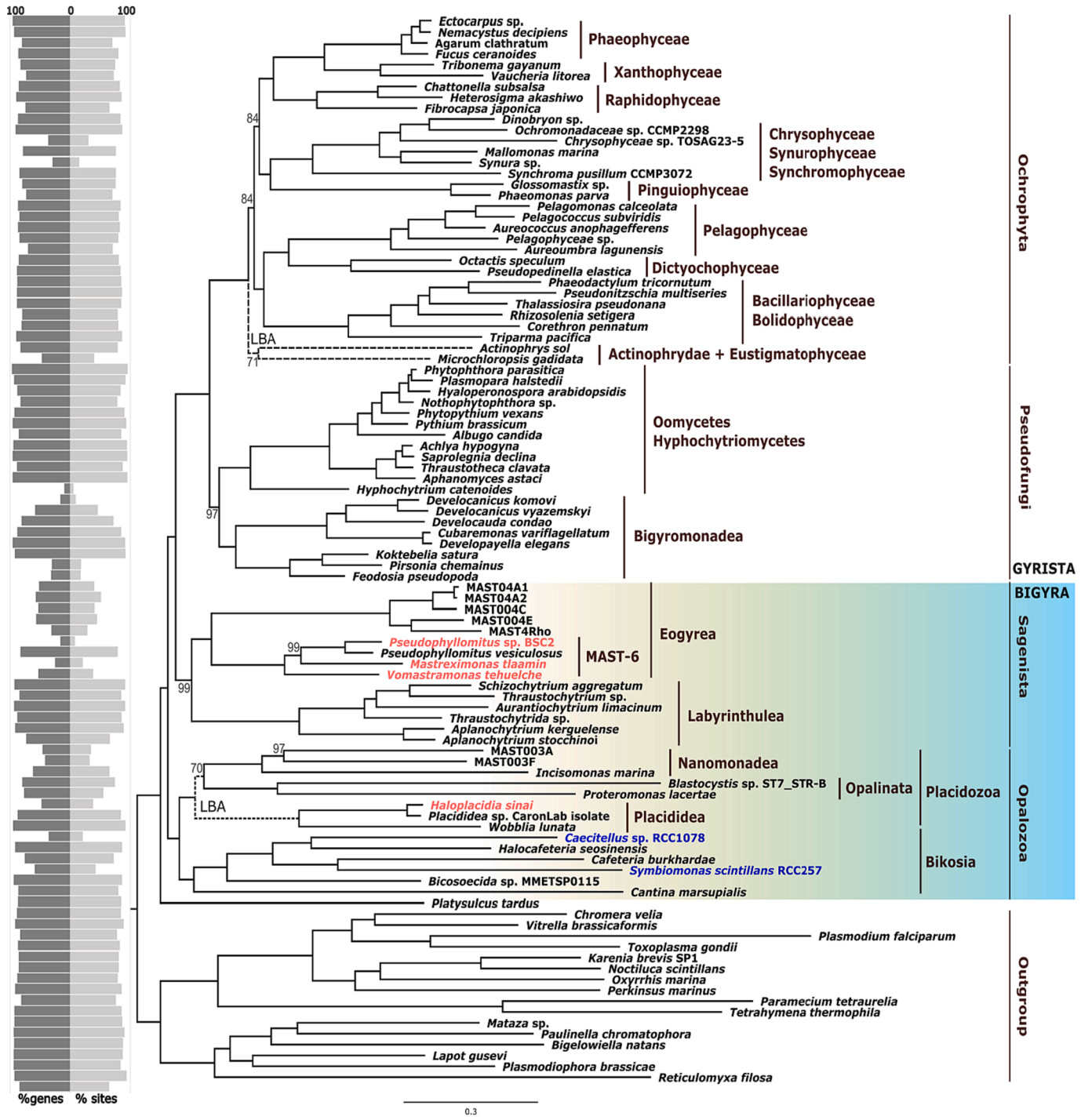

Fig. 1. Maximum-likelihood (ML) multi-gene tree of stramenopiles, including six new transcriptomes; four from newly described Bigyra in this study (light red), and two from culture collections (blue). The tree was constructed from concatenated alignments of 240 genes from 98 taxa (76,516 aa sites) under the site-heterogeneous model LG + C60 + F + G4 + PMSF with 100 standard bootstraps. Only nodes with $\leq$ 99 % support values are labelled, with unlabelled nodes indicating 100 % bootstrap support. Dashed branches indicate potential long branch attraction artefacts (LBA). The % genes (dark grey) and sites occupied (light grey) for each taxon are shown on the mirrored bar plot on the left. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

for the Nanomonadea (MAST-3) and Opalinata (*Blastocystis* sp.) clade was weak (70 % bootstrap). Placidozoa and Bikosia in turn formed a robust monophyletic group, the Opalozoa, which is the sister lineage to the rest of the stramenopiles, except for the most deep-branching *Platysulcus tardus*. *Symbiomonas scintillans* RCC257 branched sister to *Cafeteria burkhardae* (Fenchel and Patterson, 1988; Schoenle et al., 2020) and this clade formed a well-supported sister lineage with a clade composed of *Caecitellus* sp. RCC1078 and *Halocafeteria seosinensis* (Park et al., 2006) (Fig. 1).

When fast-evolving sites were removed from the concatenated matrix to assess the effects of long branch attraction, monophyly of the Ochrophyta, Gyrista, Sagenista, and Opalozoa were well supported up to 65 % site removal (50,000 aa; Fig. 2A). The monophyly of pseudofungi and the relationship between Gyrista and Sagenista were well supported

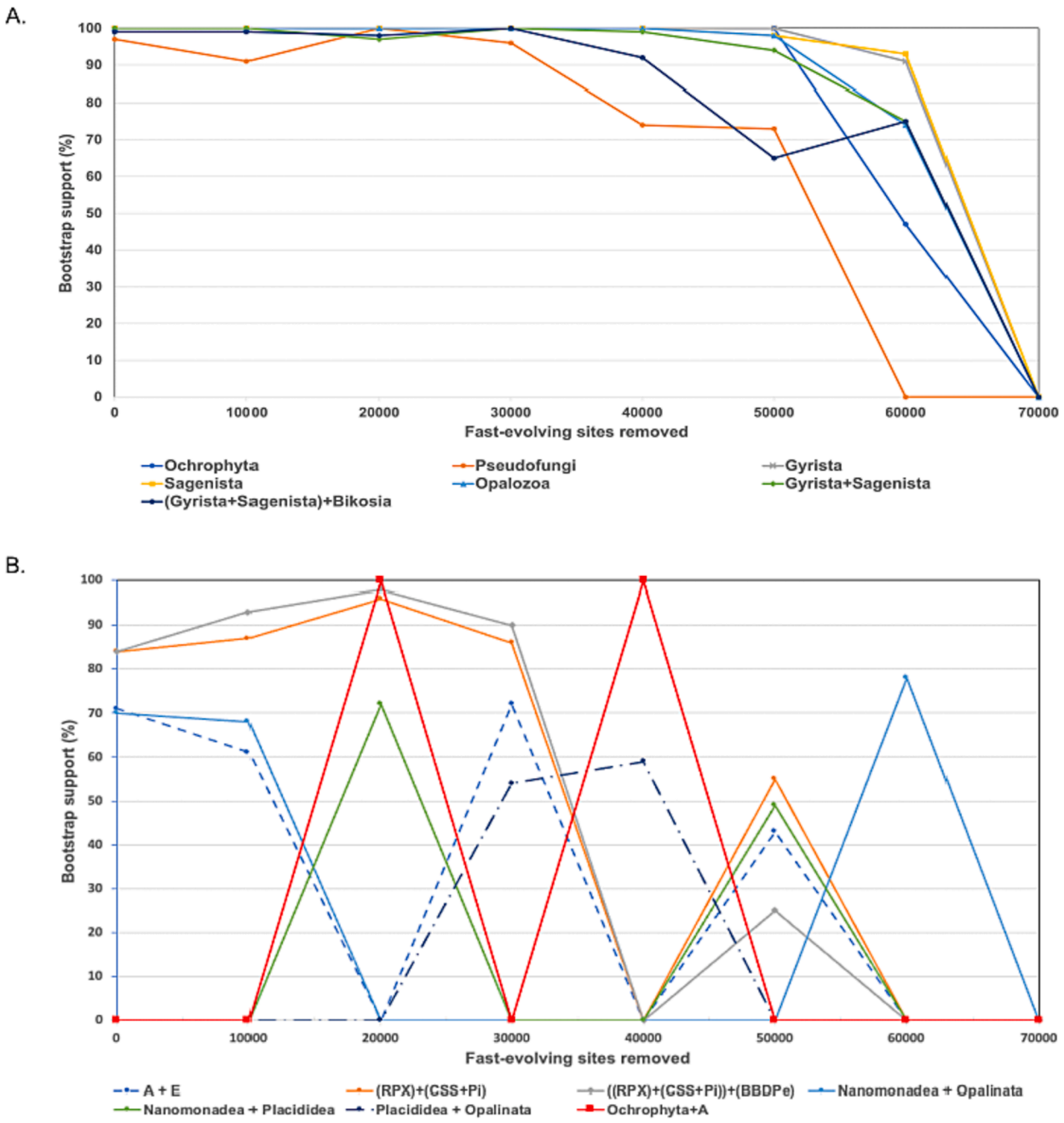

Fig. 2. Change in bootstrap support with the incremental removal of fast-evolving sites (10,000 sites removed at each step) for the monophyly of major stramenopile groups (A) and minor unstable groups (B). **A.** Monophyly of major stramenopile groups show strong bootstrap support up to 30,000 sites removed. Paraphyly of Bigyra, represented by "(Gyrista + Sagenista) + Bikosia" and "Gyrista + Sagenista". **B.** Monophyly of unstable groups showing fluctuation in bootstrap support. Bootstrap supports with zero values indicate alternative topology (not shown here) with weak support (22–55 %). Topologies within Opalozoa (Nanomonadea, Placididea and Oplinata) were unstable and weakly supported. Topologies within the Ochrophyta were also largely unstable, especially for Eustigmatophyceae ("E") and *Actinophrys sol* ("A"). "RPX" = Raphidophyceae + Phaeophyceae + Xanthophyceae; "CSS + Pi" = Chyrsophyceae + Synurophyceae + Synchromophyceae; "BBDPe" = Bacillariophyceae + Bolidophyceae + Dictyophyceae + Pelagophyceae.

up to 39 % sites removed (30,000 aa; Fig. 2A). However, the groups with weak to moderate supports (70–84 %) in Fig. 1 continued to show unstable relationships when fast-evolving sites were removed (Fig. 2B). Particularly, the placement of *A. sol*, *Microchloropsis gadidata* (Eustigmatophyceae) and subgroups of opalozoans changed. The alternative placement for *A. sol* was as a sister lineage to the rest of the ochrophytes when 20,000 aa sites (26 %) and 40,000 sites (52 %) were removed (Fig. 2B). For *M. gadidata*, it formed a sister lineage with Pinguiophyceae or the rest of the ochrophytes except *A. sol*. Although the paraphyly of Bigyra was always supported with the progression of fast-evolving site removal, the relationships among the subgroups kept changing with weak support (~70 %) (Fig. 2B).

To evaluate the effect of amino acid composition bias within subgroups of opalozoans (within Placidozoa), amino acid compositions of all the taxa within Placidozoa were compared upon inspecting the GC% of each transcriptome. All taxa belonging to Placididea were enriched in GARP amino acids compared to the background dataset, whereas all Opalinata were enriched in FYMINK, as was Nanomonadea with the exception of *Incisomonas marina*. Additionally, the amino acid composition of Placididea was more similar to Bikosia than the rest of the

Placidozoa (Fig. S3). However, when a phylogenomic tree was reconstructed using the recoded main matrix, the topology of the Placidozoa remained the same as Fig. 1, while the placement of *A. sol* and *M. gadidata* changed; *A. sol* being the sister lineage to Pinguiophyceae and *M. gadidata* being the sister lineage to RPX (Fig. S4).

For the Bayesian analysis, the chains did not converge (maxdiff = 1), with all chains conflicting with one another. When a consensus tree from each chain was compared, all the trees had the same topology of Sagenista and Bikosia that was also seen in the ML-PMSF inferred trees (Fig. 1; Fig. S5). For Placidozoa, all the consensus trees had Nanomonadea branching sister to a clade composed of Opalinata and Placididea, a different topology from the ML-PMSF analyses, except the one conducted with the MASTer-matrix (Fig. 1; Fig. S2 and S5). All chains had different Ochrophyta topologies, although the sub-clade relationship of BBPe was the same as the ML-PMSF inferred trees. The same was observed for the monophyletic CSS and the monophyletic RPX. The placement of *M. gadidata*, *A. sol*, and Pinguiophyceae were the most inconsistent. In all chains, Bigyromonadea formed a sister lineage to Ochrophyta and two out of four chains had monophyletic Bigyra (excluding *P. tardus*) (Fig. S5B-C).

3.2. New species represent phylogenetically diverse MAST-6 group in SSU rRNA analysis

To determine the genetic diversity of MAST-6 in publicly available sediment datasets, a SSU rRNA tree was constructed with the extracted amplicon sequence variants (ASVs) trained with the modified PR2 reference database, including the SSU rRNA sequences of the newly described species in this study. All the SSU rRNA sequences obtained from the newly described MAST-6 species (>1800 bp), *H. sinai* (>1800 bp) and two bikosia (>1600 bp) species are nearly full length. In total, 12 unique ASVs from BioMarKs were assigned to MAST-6 species; 9 for SouthChina; 16 for Norway; 6 for Deepsea; and 61 for the ISME2020 dataset. In general, studies that targeted the V4 region had sequence lengths between 183 and 460 bp; the V1-2 region 397 to 429 bp; and the V9 region 128–154 bp (Table 1). Shorter sequence lengths (~180 bp) from V4 targeted amplicon data are unpaired reads where low quality reverse reads were dropped.

The SSU rRNA analysis of environmental data revealed substantial diversity of the MAST-6 group, which were largely divided into four subgroups. The new MAST-6 species and previously cultured species were found within three sub-groups (Fig. 3); *M. tlaamin* in sub-group I, *Pseudophyllomitus* sp. BSC2 in sub-group II, and *V. tehuelche* in sub-group III. However, their phylogenetic relationship needs a further examination as the node supports were weak (Fig. S7). To evaluate the prevalence of these new MAST-6 species in the amplicon studies of sediment samples, relative abundance was plotted against other environmental MAST-6 ASVs (Fig. 4). All sediment datasets had relatively high abundance of MAST-6, particularly BioMarKs (65 % of all MASTs and 2.24 % of all ASVs) and ISME2020 (37 % of all MASTs and 13.9 % of all ASVs) studies. Amplicon sequence variants assigned to *M. tlaamin* (PRC5) were dominant MAST-6 groups in Deepsea (67 % of all MAST-6), ISME2020 (44 %) and Norway (20 %) datasets, while no sequences assigned to the new MAST-6 species were present in the BioMarKs study (Fig. 4B). It is important to note that not all ASVs assigned to *M. tlaamin* correspond exactly to the same species, but were assigned based on which most closely related MAST-6 species was available in the training dataset. This assignment may change as new MAST-6 transcriptomes representing each sub-group are added to the updated SSU reference data. Interestingly, we did not find shared MAST-6 ASVs across the four studies, which may be due to different sequencing technologies with different coverage or presence of biological sequence variants by different sampling sites and time. Within sub-group I (Fig. 3), the ASVs from the ISME2020 and Norway datasets are placed closest to *M. tlaamin*, while the ASVs from the Deepsea dataset are more distantly related. This indicates that MAST-6 species closely related to *M. tlaamin* are not only genetically diverse and abundant, but present in various sediment samples across different depths and geological locations. Additionally, within the MAST-6 sub-group I, *M. tlaamin* and the environmental sequence "SA2_3F7" are the only two with nearly full length SSU rRNA sequences, compared to sub-group II, which includes more close-to-full length SSU rRNA sequences. The addition of the *M. tlaamin* SSU rRNA sequence in the taxonomic assignment has markedly improved phylogenetic resolution among the MAST-6 lineages. A similar trend was observed in sub-group III where *V. tehuelche* was placed. Amplicon sequence variants from the BioMarKs dataset that were assigned to MAST-6, however, were mostly placed across the different sub-groups, except sub-group II. Along with many ASVs from ISME2020, six out of 12 unique MAST-6 ASVs of BioMarKs are placed within sub-group IV, which have no sequences from cell isolates with genomic data. When we visualized the abundance of different sub-groups across different datasets, Sub-group I was the dominant group in all cases. More sub-groups were present in ISME2020 and Norway and this is likely due to sequencing techniques (i.e., pyrosequencing in BioMarKs) and limited universality of the V9 primer used in Deepsea dataset (Fig. S6C) (Decelle et al., 2014). As QIIME2 generates ASVs, we interpreted the data without clustering, however, clustering the ASVs by ≥ 98 % sequence similarity resulted in 10 and 37 ASVs assigned to MAST-6 in BioMarKs and ISME2020, respectively. For other MAST lineages, ASVs assigned to MAST-1, −3, −9, and −12 were present in all studies. Depending on the dataset, the relative abundance these MAST lineages were high although values fluctuated depending on the sample within the study.

Two ASVs assigned to *V. tehuelche* were present in the SouthChina study, and no ASVs were assigned to *Pseudophyllomitus* sp. BSC2 (Fig. S6). However, based on initial phylogenetic evaluation of assigned MAST-6 sequences from the SouthChina study, blastn searches, and the EPA analysis (low LWR values with the equal likelihood of alternative placements), the sequences were excluded from main the SSU-EPA tree (Fig. S6-7). Additionally, one Deepsea ASVs assigned to MAST-6 was excluded from the downstream analysis based on the initial phylogenetic tree and blastn search placing it close to MAST-8 (Fig. S7). Aside from *M. tlaamin*, other MAST-6 sequences from cell isolates (*P. vesiculosus* and NY13S_181 clone) were found in Deepsea (1.5 %) and Norway (0.8 %), although in low relative abundance. The rest of the MAST-6 sequences were assigned to environmental "MAST-6_X" and "SA2_3F7" from the PR2 dataset and "MAST-6", a potentially new MAST-6 variant (Fig. 4B).

As an additional measure to quantify the confidence of the extracted SSU rRNA sequence placements, sequences with LWR values ≥ 95 % verified with blastn searches are highlighted in red in SSU-EPA tree (Fig. 3) and considered to be of high confidence (Berger et al., 2011; Dunthorn et al., 2014). No Deepsea_MAST6 ASVs had LWR values ≥ 95 %, with many of them having equally likely alternative placements (blue lines in nodes in Fig. 3).

3.3. The new Placididea may be rare in sediments

For the Deepsea study, the only ASVs with high LWR values were the ones assigned to Placididea species. Although there was a total of 15 ASVs assigned to Placididea, none were assigned to *H. sinai*. When the SSU sRNA tree was constructed including the 10 Placididean OTU sequences of ESBig, *H. sinai* formed a sister lineage with ESBig133 which were found in water samples with salinities of 78, 124 and 380 ‰ (Lee et al., 2021) and, *Placididea* sp. (Caron Lab), cultured in 36 ‰ (Caron, 2000; Keeling et al., 2014) (Fig. 3). This clade formed a sister-lineage to "Group-D" containing *Haloplacidia cosmopolita* (described in Park and Simpson, 2010; Rybarski et al., 2021), which can tolerate 15–175 ‰ salinity. Additionally, ESBig sequences and Deepsea placididean sequences were placed across the major subgroups of Placididea, despite being isolated from different geographical locations and a broad range of salinities (36 ‰ for Deepsea and 76–380 ‰ for ESBig) (Fig. 3). The confidence of the extracted SSU rRNA sequences placement within the

Fig. 3. A RAxML SSU rRNA phylogenetic tree (SSU-EPA tree) of stramenopiles. The tree was constructed under the GTR + GAMMA model with 1000 rapid bootstrap replicates, using an alignment of 527 stramenopile sequences and seven outgroup sequences (8,771 sites): 109 extracted ASVs assigned to MAST-6 or Placididea from the amplicon dataset, and 10 placididean OTU sequences from ESBig study. The four new Bigyra species are coloured in pink. The likelihood weight ratio (LWR) values calculated from our EPA analysis are coloured in red for high confidence (LWR $\geq$ 95 %), and in blue for low confidence (LWR $<$ 95 %), indicating equally likelihood of alternative placements. The label structure for the ASVs is "Dataset_MAST6/Placididea_count". Clades other than MAST-6 and Placididea are collapsed. For bootstrap supports, see Fig S6. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

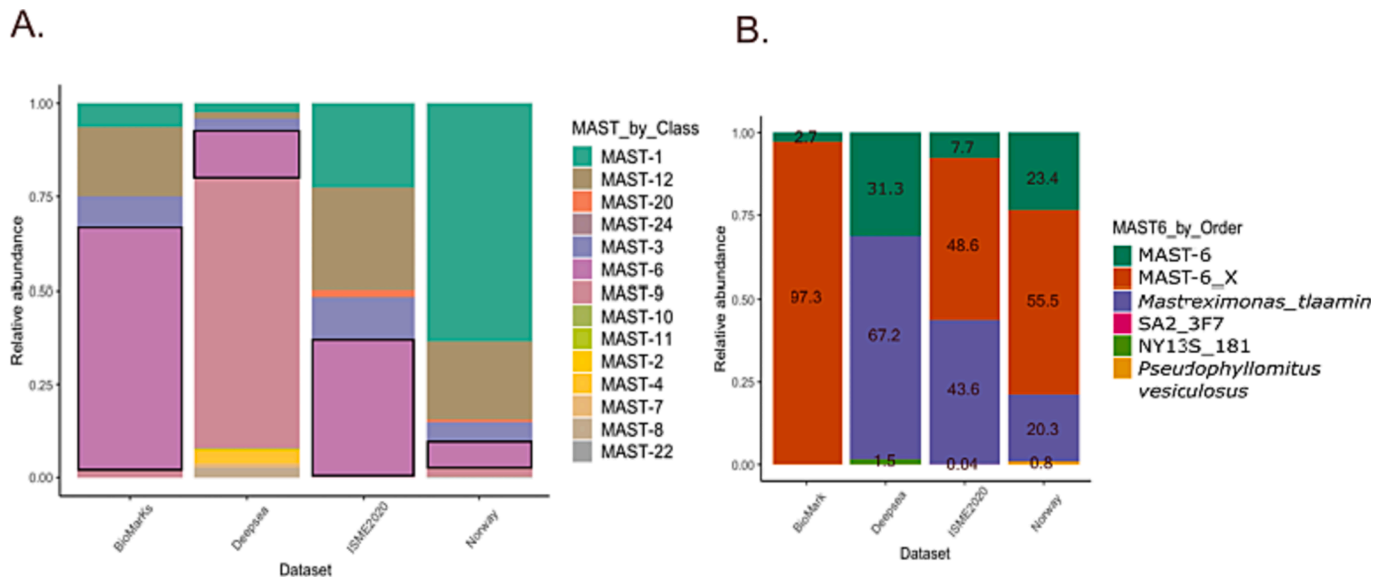

Fig. 4. Stacked bar plots of the relative abundance of unique ASVs assigned to main MAST groups (**A**) and MAST-6 (**B**) from four sediment datasets: BioMarKs, Deepsea, ISME2020, and Norway. Deepsea is the only study with a SSU rRNA gene primer targeting the V9 region. **A.** Composition of different MASTs from each dataset grouped by class level. Black frames indicate the relative abundance of MAST-6. **B.** Composition of MAST-6 lineages from each dataset grouped by order to further show higher taxonomic assignment. "MAST-6_X" represents an unknown MAST-6 lineages classified from the PR2 database, and "MAST-6" represents a potentially new MAST-6 lineage based on the updated taxonomic training database. "*Mastreximonas tlaamin*" is one of the new MAST-6 species descried in the current study. "*Pseudophyllomitus vesiculosus*" and "NY13S_181" are previously reported cultures and "SA2_3F7" is an environmental sequence.

partial-to-full length SSU sequences was inferred from LWR values ≥ 95 %. Seven out of 15 Deepsea and four out of 10 ESBig placididean ASVs showed high confidence (red nodes in SSU-EPA tree in Fig. 3).

3.4. Morphological description and new name designation

3.4.1. An undescribed Pseudophyllomitus sp. BSC2

The cell is a biflagellated, naked, and free-living single-celled protist. The outline of the cell is oblong and slightly concave at the middle, measuring 22 μm in length and 7 μm in width (Fig. 5A-E). Both flagella emerge subapically from a gullet which continues for two-thirds (approximately 5 μm) of the cell width. The anterior flagellum is ~ 1x cell length and directed forward. The posterior flagellum is 0.5x cell length and inserts to the left of the anterior flagellum. When the cell was stationary, the anterior flagellum beats rapidly in a sinusoidal wave, often sweeping to the right (Video S1). The posterior flagellum is anchored sideways, likely attached to the surface, and occasionally trailing behind when changing direction. The two flagella are clearly visible and do not adhere to each other, the morphological trait that separates *Pseudophyllomitus* from *Phyllomitus* species (Lee, 2002). Some refractile granules are visible at the cell surface. Although no feeding was observed at the time of sampling, the cell is likely be a phagotroph. The shape of the cell is comparable to *P. salinus* (Lackey, 1940) in its oblong shape however, it is distinguishable by the longer anterior flagellum and the shorter posterior flagellum, and the presence of refractile granules on the cell surface (Lee, 2002). The cell is also similar to *P. granulatus* (Larsen and Patterson, 1990; Lee and Patterson, 2002) in terms of length and movement of both flagella and presence of the vesicles on the cell surface. However, its oblong shape is distinguished from sac-shaped *P. granulatus*.

3.4.2. New genera and species designation

Mastreximonas gen. nov. Lax, Cho, and Keeling.

Taxonomy: Eukaryota; SAR Burki et al. 2008, emend. Sar Adl et al. 2012; Stramenopiles Patterson 1989, emend. Adl et al. 2005; Bigyra Cavalier-Smith, 1998 emend. 2006; Sagenista Cavalier-Smith, 1995; Eogyrea Cavalier-Smith 2013.

Diagnosis: Flagellated, naked, and single-celled protist. Cell outline is elongated sac-shape with a slightly flattened anterior end. Thick anterior flagellum emerging apically, posterior flagellum may be very short and trailing under the cell, or absent.

Etymology: Acronym for *mar*ine *stra*menopile, *éxi* (Greek έξι, number 6), and *monas* (Greek, fem.), commonly used for unicellular organisms.

Zoobank Registration. LSID for this publication: urn:lsid:zoobank.org:pub:583E6EDF-B1A2-4220-96D7-C4CF47DA9A6C. LSID for the new genus: urn:lsid:zoobank.org:act:960070EF-0259-4A31-936F-A372FED9B7FE.

Type species. *Mastreximonas tlaamin*

Mastreximonas tlaamin sp. nov. Lax, Cho, and Keeling.

Diagnosis: The cell measures 15.6 μm in length and 4.8–6.4 μm in width. The prominent anterior flagellum is markedly thicker than the posterior flagellum and roughly two-thirds of the cell length (13 μm), directed forward, and emerges apically from a gullet. The posterior flagellum was not observed. Many large vesicles (approximately 1.5–2.5 μm in diameter) are present in the cytoplasm and two similarly sized golden vacuoles (2.4 μm) are present at the posterior end. The cell swims in a circular motion with the anterior flagellum beating in a sine wave. The nucleus is located just below the base of the anterior flagellum and is 3.5–4.0 μm in diameter. Although no feeding was observed at the time of sampling, the cell is likely a phagotroph.

Type Figure: Fig. 5F.Video S2 of living cell *Mastreximonas tlaamin*.

Gene sequence: The SSU rRNA gene sequence has the GenBank Accession Number OQ909084.

Type material: The specimen shown in Fig. 5 F–J is the holotype. The actual specimen (single cell) was destroyed in the process of single-cell transcriptome sequencing by necessity (see International Code of Zoological Nomenclature, Art. 72.5.6, Declaration 45).

Type locality: Oxic marine intertidal sediment of the Powell River, British Columbia, Canada (49′50′'2'' N, 124′31′'0'' W).

Etymology: The species epithet 'tlaamin' is derived from the Tla'amin Nation, an indigenous First Nation in Powell River, BC. It means 'our people' in Tla'amin language.

Zoobank Registration: LSID for this publication: urn:lsid:zoobank.org:pub:583E6EDF-B1A2-4220-96D7-C4CF47DA9A6C. LSID for the new species:urn:lsid:zoobank.org:act:8B0835A7-679C-441A-AFFC-

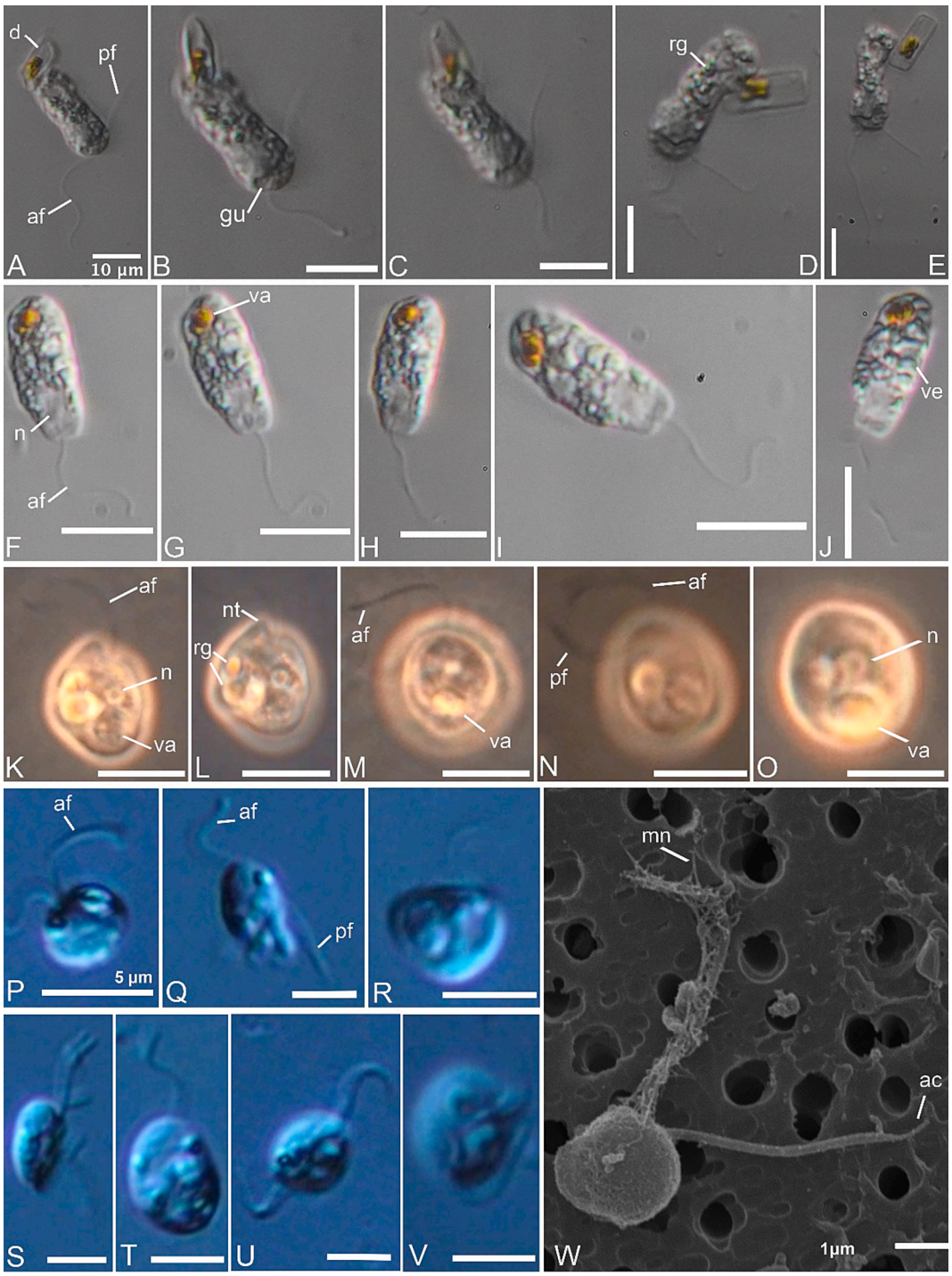

Fig. 5. Morphology of four new Bigyra. **A-E**. *Pseudophyllomitus* sp. BSC2. General view of the cell including anterior flagellum [af] and posterior flagellum [pf]. Both flagella emerge from a horizontal gullet [gu] and some refractile granules are visible on the cell surface [rg]. A diatom [d] is attached at the posterior end. **F-J**. *Mastreximonas tlaamin*, general view of the cell with an anterior flagellum. The nucleus [n] is visible just below the base of the anterior flagellum. Some vesicles [ve] and golden vacuoles [va] are present from the mid to posterior end of the cell. **K-O**. *Vomastramonas tehuelche*. General view of the cell with clearly visible anterior flagellum. A notch [nt] is present at the anterior end. Refractile granules and food vacuoles are present. **P-V.** *Haloplacidia sinai*, general view of the cell with two flagella. **W**. *H. sinai* in scanning electron micrograph, showing mastigonemes [mn] on the anterior flagellum and acroneme [ac] on the posterior flagellum. **Scale bars** are 10 μm for A-O, 5 μm for P-V, and 1 μm for W.

18D8596201BC.

Vomastramonas gen. nov. Tikhonenkov, Prokina, Cho, and Keeling.

Taxonomy: Eukaryota; SAR Burki et al. 2008, emend. Sar Adl et al. 2012; Stramenopiles Patterson 1989, emend. Adl et al. 2005; Bigyra Cavalier-Smith, 1998 emend. 2006; Sagenista Cavalier-Smith, 1995; Eogyrea Cavalier-Smith 2013.

Diagnosis: Biflagellate, naked, and solitary eukaryovorous protist. Cells are slightly flattened and ovoid, with a slightly narrowed posterior end and a notch at the anterior end. Both flagella are acronematic, emerging apically from a notch at the anterior end of the cell.

Etymology: Acronym for voracious, marine stramenopile, and monas (Greek, fem.) – commonly used for unicellular organisms.

Zoobank Registration. LSID for this publication: urn:lsid:zoobank.org:pub:583E6EDF-B1A2-4220-96D7-C4CF47DA9A6C. LSID for the new genus:urn:lsid:zoobank.org:act:610C621F-9C18-49E5-983A-6194BE4F97CB

Type species. Vomastramonas tehuelche

Vomastramonas tehuelche sp. nov. Tikhonenkov, Prokina, Cho, and Keeling

Diagnosis: cell body is 11.5–13 μm in length and 7.5–10 μm in width. Anterior flagellum is approximately equal to the cell length, posterior flagellum is 1.2–1.5 times longer than the cell. Anterior flagellum is markedly thicker than the posterior flagellum and clearly visible, directed forward and sideways, curved in form of an arc, vibrates very rapidly with a short wavelength but doesn't change its position during cell movement. Posterior flagellum is barely visible during cell movement, directed backwards. Cells swim close to the substrate in a circle, pushing off with the posterior flagellum, without rotation around its longitudinal axis and without changing the direction of movement. The anterior flagellum is directed towards the outer side of the circle when cell moves. When cell stops, posterior flagellum is directed sideways and curved in arc towards the anterior flagellum, so the flagella seem to stretch towards each other. Cells also can swim relatively straight, with small jerks. Numerous light-refracting granules and digestive vacuoles are present in the posterior half of the cell. No cysts.

Remarks: this species differs from the other member of MAST-6 clade, *Pseudophyllomitus vesiculosus* Shiratori et al., 2017 because the cells are not flexible and lack the rod or bar laid against the anterior side of the nucleus (Shiratori et al., 2017).

Type material: The specimen shown in Fig. 5K is the holotype (see International Code of Zoological Nomenclature, Art. 72.5.6, Declaration 45). The strain Colp-33 was propagated in a predator–prey culture with the bodonid *Procryptobia sorokini* as a steady food source but perished after a year of cultivation.

Type Figure: Fig. 5K. Video S3 illustrates a live cell of *Vomastramonas tehuelche* strain Colp-33.

Gene sequence: The SSU rRNA gene sequence has the GenBank Accession Number OQ909086.

Type locality: nearshore bottom sediments of the Strait of Magellan, Punta Arenas, Chile.

Etymology: Tehuelche is the collective name (in Araucanian) of the indigenous peoples of Patagonia.

Zoobank Registration: LSID for this publication: urn:lsid:zoobank.org:pub:583E6EDF-B1A2-4220-96D7-C4CF47DA9A6C. LSID for the new species:urn:lsid:zoobank.org:act:9C52101E-1B3E-45F2-9DB0-A6DFAB1349D1

Haloplacidia sinai sp. nov. Tikhonenkov, Cho, and Keeling

Taxonomy: Eukaryota; SAR Burki et al. 2008, emend. Sar Adl et al. 2012; Stramenopiles Patterson 1989, emend. Adl et al. 2005; Bigyra Cavalier-Smith, 1998 emend. 2006; Opalozoa Cavalier Smith 1991 emend. 2006; Placidozoa Cavalier-Smith 2013; Placididea Moriya, Nakayama & Inouye, 2002; *Haloplacidia* Rybarski, Nitsche & Arndt 2021.

Diagnosis: Cells are oval, roundish or irregularly ovoid, with the convex dorsal side and the flatter ventral side. Cell body is 5.4–8.3 μm in length and 3.4–6.6 μm in width. Anterior flagellum is approximately 1.5 times longer than the cell, posterior flagellum is approximately equal to the cell length. Posterior flagellum is acronematic and both flagella emerge from a shallow groove at the central part of the ventral side of the cell and oriented in the opposite directions. Anterior flagellum bears mastigonemes. Cells are often attached to the substrate with a posterior flagellum and produce very fast trembling movements. No cysts.

Remarks: this species differs from the other member of the genus, *H. cosmopolita* Rybarski, Nitsche & Arndt 2021, by having a slightly different shape of the cell without pronounced kidney-like morphology, and by the absence of cysts, even under starvation conditions (Rybarski et al., 2021).

Type material: the specimen shown in Fig. 5U is the holotype (see International Code of Zoological Nomenclature, Art. 72.5.6, Declaration 45).

Type Figure: Fig. 5U. Video S4 illustrates a live cell of *Haloplacidia sinai* strain PhM-7.

Gene sequence: The SSU rRNA gene sequence has the GenBank Accession Number

OQ909082.

Type locality: surface of corals in the Red Sea, Sharm El Sheikh, Egypt.

Etymology: named after the place it was found in the Mount Sinai region, where the Ten Commandments were given to Moses by God, according to the Book of Exodus in the Hebrew Bible. The English name Sinai came from Latin, ultimately from Hebrew סִינַי, pronounced /siˈnái/ .

Zoobank Registration: LSID for this publication: urn:lsid:zoobank.org:pub:583E6EDF-B1A2-4220-96D7-C4CF47DA9A6C. LSID for the new species:urn:lsid:zoobank.org:act:6EF3B31E-BAFB-4E5C-8010-DF3BBE3DCE43

4. Discussion

4.1. Updated taxon sampling and phylogeny of MAST-6

MAST-6 has been shown to be both abundant and diverse through various amplicon sequencing studies of sediment samples (Massana et al., 2015; Rodríguez-Martínez et al., 2020; Schoenle et al., 2021) (Table 1). Despite the known abundance and distribution across various sediment sites, inferring the diversity of MAST-6 species has been limited to a reference database composed of a handful of SSU rRNA sequences. Moreover, only a single taxon for which -omic-level data are available (i.e., *Pseudophyllomitus vesiculosus*) has represented the MAST-6 clade in phylogenomic analyses. In our present study, we generated transcriptomes of three new MAST-6 taxa: *Mastreximonas tlaamin*, *Vomastramonas tehuelche*, and *Pseudophyllomitus* sp. BSC2, and updated the deep phylogeny of stramenopiles. These three new MAST-6 species in turn reflect broader genetic diversity by representing different subgroups of the MAST-6 lineage.

Like previously described *P. vesiculosus*, all new MAST-6 species described here were found in sediments, and had relatively large and numerous vesicles or granules underlying the cell surface. The new *Pseudophyllomitus* sp. BSC2 was the most closely related to previously described *P. vesiculosus* and one of the longest *Pseudophyllomitus* species described so far (22 μm) (Lee and Patterson, 2002). The overall morphological characteristics were most similar to *P. granulatus* and—to a lesser extent—to *P. salinus*, however, due to not observing feeding behaviour, we refrained from establishing a new species for this cell. *Mastreximonas tlaamin* had a similar oblong shape to *Pseudophyllomitus* sp. BSC2 and branched sister to the two *Pseudophyllomitus* species. *Vomastramonas tehuelche*, on the other hand, had a more circular shape and formed a sister lineage to the rest of the MAST-6 species in the phylogenomic tree.

4.2. The new MAST-6 species broaden the genetic diversity

Our study showed that *M. tlaamin*-related ASVs were the most abundant MAST-6 across different sediment amplicon studies (Fig. 4), representing a largest MAST-6 sub-group consisting of ASVs from various sediment locations and depths (Fig. 3). Amplicon sequence variants assigned to *M. tlaamin* were absent in other studies (e.g., BioMarKs and SouthChina). This can be due to low abundance of *M. tlaamin* at the time of sampling and/or the combination of the technical pitfalls associated with pyrosequencing. Pyrosequencing has lower sequence coverage and is also prone to non-homopolymer errors (Luo et al., 2012), all of which may have contributed to the lack of *M. tlaamin*-related ASVs detections. None of the new MAST-6 species from this study were found within sub-group IV despite its high relative

abundance in the ISME2020 dataset (Fig. S6C). Future efforts in isolating and describing cells of the sub-group IV may not only confirm phylogenetic diversity but help us better understand the biology behind the sediment-associated MAST-6 species. Based on the absence of *Pseudophyllomitus* sp. BSC2-and *V. tehuelche*-assigned ASVs, and low relative abundance of ASVs assigned to *P. vesiculosus* in sediment studies, these MAST-6 species may be rare. Additionally, sample timing may have played a role as the cell abundance has been reported to be affected by seasonality and salinity (Piwosz and Pernthaler, 2010). For example, both small and large morphotypes of MAST-6 were observed to have short-lived peaks at mid-May to early-June in the Gulf of Gdansk shortly after freshwater inflow, followed by a substantial decline in relative abundance (Piwosz and Pernthaler, 2010). All datasets except BioMarKs were sampled mostly in September, with some sampled in August and July (sampling months for BioMarKs from February to October). These months were the time when the number of sub-group II associated MAST-6 were reported to be very low (Piwosz and Pernthaler, 2010). Although this work by Piwosz and Pernthaler was done on plankton samples, the rapid and short-lived seasonal fluctuation of MAST-6 abundance revealed that this group may respond quickly to changing environment, including the ones in sediments.

4.3. Rare and potentially halotolerant haploplacidia sinai and its implication in trait evolution

Haloplacidia sinai is the fourth new species reported here. *Haloplacidia sinai* belongs to Placididea, another significant clade of Bigyra that was represented by a couple transcriptomes before the present study. Similar to some of the previously described species of Placididea (Park and Simpson, 2010), *H. sinai* was found in a relatively high salinity environment. Although the present study did not detect any ASVs assigned to *H. sinai*, based on its relationship (Fig. 3) with other isolated cells cultured in broad range of salinity, *H. sinai* might also be found in non-hypersaline environments. Absence of ASVs assigned to *H. sinai* may be due to the choices of sampling habitats in the datasets examined, as we isolated the cell from coral scrapes. Interestingly, three "Deep-sea_Placididea" sequences formed a clade with the two *Suigetsumonas* spp. isolated from brackish lakes in Japan and Kenya (Okamura and Kondo, 2015; Rybarski et al., 2021) (Fig. 3), further demonstrating the broad range of salinity in which species of Placididea can be found.

The halophilic trait is not just limited to Placididea but can also be found in Bikosia. The extremely halophilic *Halocafeteria seosinensis* (Park et al., 2006; Park and Simpson, 2010) can survive between 75 and 363 ‰ (Park et al., 2006). Furthermore, several traits including differential gene expressions involved in anti-oxidization, membrane fluidity, O-linked glycosylation, and gene-duplication were linked to high salt adaptability of *H. seosinensis* (Harding et al., 2017). Exploring the evolution of halotolerancy in these deep-branching stramenopiles may lead to a better understanding of the ancestral state of the stramenopiles. For example, determining whether the trait evolved separately in Placididea and Bikosia or arose in the last common ancestor of the two groups, may help answering the transition between different salinity barrier in stramenopiles (Dunthorn et al., 2014; Jamy et al., 2022).

4.4. Phylogenomics of stramenopiles with a twist

In this study, *H. seosinensis* formed a sister lineage to *Caecitellus* sp., and this atypical mastigoneme-lacking group (O'Kelly and Nerad, 1998; Park et al., 2006) in turn formed a robust sister lineage to the clade composed of *S. scintillans* and *Cafeteria burkhardae* (Fig. 1; Fig. S1, 3). When we added the most recent genomic data of MAST-1, MAST-7, MAST-8, MAST-9, and MAST-11, the relationship remained the same (Fig. S2). The bikosian phylogenomic relationship in this study (Fig. 1) is consistent with previous SSU phylogenetic trees (Cavalier-Smith and Chao, 2006; Cavalier-Smith and Scoble, 2013; Guillou et al., 1999; Park et al., 2006; Shiratori et al., 2017, 2015). However, an alternative SSU rRNA phylogeny showed *H. seosinensis* forming a sister lineage to a clade composed of *Cafeteria* spp. and *Caecitellus* spp. (Yubuki et al., 2015), similar to the SSU-tree generated in our study (Fig. S8). This could be due to the fast-evolving nature of many bikosian SSU rRNA genes, as indicated by the long branch length of *S. scintillans* and *C. burkhardae* (Fig. S8). Additionally, the topology of Bikosia in the phylogenomic tree may change as there are far more bikosia that are not represented in transcriptomic or genomic datasets, such as diverse *Bicosoeca* spp. (Karpov, 1998), *Pseudobodo* spp. (Griessmann, 1913), and freshwater or soil bikosians, including *Siluania monomastiga* (Karpov et al., 1998), *Nerada mexicana* (Cavalier-Smith and Chao, 2006), *Adriamonas peritocrescens* (Verhagen et al., 1994), and *Paramonas globosa* (Cavalier-Smith and Chao, 2006; Saville-Kent, 1880) (Fig. S8).

The paraphyly of Bigyra has been repeatedly demonstrated in recent publications as more genomic data across different lineages of stramenopiles have become available (Azuma et al., 2022; Burki et al., 2016; Cho et al., 2022; Noguchi et al., 2016), including the ML-PMSF trees in our study (Fig. 1; Fig. S1-S2). However, the Bigyra are monophyletic in some other studies (Derelle et al., 2016; Thakur et al., 2019), as well as the consensus trees obtained from two of the four MCMC chains in this study (Fig. S5 B-C). As these studies all have differing numbers of taxa (as well as different taxa) and orthologs, and use different methods for data processing, it is difficult to pinpoint the steps that would have caused topological incongruencies across these analyses.

In contrast to previously published work (Azuma et al., 2022), *Actinophrys sol* is not placed sister to the rest of the Ochrophyta. Rather, it forms a weakly-supported clade with *Microchloropsis gadidata* (Eustigmatophyceae). As a single transcriptome is representing each of Eustigmatophyceae and Actinophrydae, we argue that this is the result of long branch attraction artefacts (LBA) caused by eroded phylogenetic signals (class II LBA), rather than parallel substitutions (class III) or saturation (Fig. 1) (Wägele and Mayer, 2007). This class II LBA was demonstrated in trees reconstructed from fast-evolving-site removal (Fig. 2B), Bayesian analysis (Fig. S5) and, 39per-, 59per-, and recoded matrix (Fig S1 and S4), where the placement is chaotic rather than showing a pattern. The Ochrophyta phylogeny was further complicated by other unstable relationships of Eustigmatophyceae, Pinguiophyceae, and among CCS, RPX, and BBPe. Phylogenomic discrepancies found in Ochrophyta nuclear datasets should be addressed by more taxon sampling to break the long branches (e.g., Marine OCHrophytes (MOCH) (Massana et al., 2014) and Olisthodiscophyceae (Barcytė et al., 2021)) and developing new phylogenomic models that can resolve short internal branches within early ochrophyte divergence (di Franco et al., 2021; Philippe et al., 2011; Ševcíková et al., 2015). A similar discrepancy between Maximum likelihood (ML) and Bayesian analyses was also observed in a recent study (Cho et al., 2022), when all four chains yielded different topologies compared to the ML tree. Interestingly, our consensus trees also differed from the ML analysis where all the consensus trees recovered Bigyromonadea forming sister lineage to Ochrophyta. This discrepancy between the ML and Bayesian analyses was also observed in said previous study (Cho et al., 2022). However, constrained AU tests (Nguyen et al., 2015; Shimodaira, 2002) failed to reject the monophyly of Bigyromonadea, together with Oomycetes (Winter 1897) and Hyphochytriomycetes (Dick 1983), forming a sister lineage to Ochrophyta in all four consensus trees (Cho et al., 2022).

Within the monophyletic Placidozoa (Placididea + Nanomonadea + *Blastocystis*), the relationship among the sub-groups was no longer strongly supported, in contrast to three other studies (Azuma et al., 2022; Cho et al., 2022; Thakur et al., 2019) and our Bayesian analysis (Fig. S5). Based on the amino acid composition of the placidozoan data used in this study, the topology appears to result from artefact due to enriched GARP amino acids in this group (Fig. 1; Fig. S3). However, repeating the ML-PMSF analysis without *H. sinai* (data not shown) recovered the same Placidozoa topology as previous studies (Azuma et al., 2022; Cho et al., 2022; Thakur et al., 2019) with weak support (76 %). Despite the suspected artefact due to amino acid composition bias,

the recoded tree analysis did not change the topology of Placidozoa, although the bootstrap support was weak. The present placidozoan topology is likely unstable in our dataset due to a combination of long branches leading to Placididea and Opalinata, and low taxon sampling in Bigyra. As hinted by the Opalinata + MAST-12 clade and diverse placidideans shown in the SSU-tree (Fig. S8), future efforts in increasing taxon sampling will likely help stabilize the placidozoan topology, in addition to deploying a phylogenetic model that can resolve LBA amongst stramenopiles.

5. Conclusion

The first impression of phagotrophic Bigyra to most observers may be a jumble of heterotrophic flagellates with few distinguishing features. It was only through SSU rRNA-amplicon sequencing that their identities and phylogenetic diversities were revealed. Even then the reference-dependent taxonomy assignment and usage of a single SSU primer-set often led to an under-detection of their diversity. Placididea on the other hand were initially discovered through cell isolates, but an assessment of their environmental distribution was limited due to its preferential amplification with a V9-targeting primer set (Lee et al., 2021; Rybarski et al., 2021). Despite the group's diversity and ability to survive in a broad range of salinity, only very limited transcriptome or genome data had been available prior to this study. After adding another transcriptome of a placididean (*H. sinai*), we observed a topology change in Placidozoa that conflicts with previous studies. Based on the long unbroken branch leading to Placididea and alternative tree construction methods, the topology from the current study may be an artefactual relationship caused by long branch attraction (Delsuc et al., 2005; Philippe et al., 2005). Combined with a lack of taxon sampling, the presence of highly divergent species, such as symbiotic Opalinata, *Incisomonas marina,* and their long-branching sister lineage, Bikosia, it is likely that currently available models cannot resolve the true relationship of Placidozoa. Although the phylogenomics of Ochrophyta are beyond the scope of the present study, we will note that it remained unresolved with conflicting ML and Bayesian analyses in this and previous studies (Azuma et al., 2022; Cho et al., 2022), all of which suggests more data will be required. Adding three new MAST-6 transcriptomes to our phylogenomic tree resulted in a robust monophyly of MAST-6 and MAST-4, a relationship only recently revealed in phylogenetic studies (Shiratori et al., 2017; Thakur et al., 2019; Cho et al., 2022). Along with the new MAST-6 species, we also showed phylogenomic relationship among Sagenista with recently published genomic data of MAST-7, −8, −9, and −11 for the first time. Newly described MAST-6 species improved the detection of considerable phylogenetic diversity of sediment-associated MAST-6 species from various sample sites, and demonstrated a higher diversity compared to that of the most abundant MAST-4 group (Logares et al., 2012; Rodríguez-Martínez et al., 2012). One of the abundantly detected MAST-6 is closely related to the newly described *M. tlaamin* (PRC5), while few or no ASVs were detected for *V. tehuelche* and *Pseudophyllomitus* sp. BSC2. This indicates different MAST-6 species may be rare and have different seasonal dynamics.

CRediT authorship contribution statement

Anna Cho: Conceptualization, Formal analysis, Writing – original draft, Writing – review & editing, Visualization. **Denis V. Tikhonenkov:** Writing – original draft, Writing – review & editing, Visualization, Investigation, Funding acquisition. **Gordon Lax:** Writing – review & editing, Visualization, Investigation. **Kristina I. Prokina:** Writing – review & editing, Investigation, Visualization. **Patrick J. Keeling:** Supervision, Funding acquisition, Conceptualization, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Dr. Noriko Okamoto (Hakai Institute) for assistance in sampling in Powell River, BC, Dr. Vladimir V. Aleoshin (Moscow State University) for providing a sample from Egypt, all members of Keeling Lab, especially Dr. Vittorio Boscaro, Dr. Elizabeth Cooney, Dr. Corey Holt, and Dr. Vojtěch Žárský. We thank Dr. Andrew Roger (Dalhousie) for advice in interpreting phylogenomic discrepancies, and Dr. Donald Wong (UBC) for advice in RNA extraction. We also thank the Tla'amin Nation for permission to designate a species after them. We would like to express our sincere gratitude to the editor and the reviewers who provided detailed and thorough suggestions, allowing us to significantly improve our manuscript. This work was supported by grants from GenomeBC (R02MSE) and NSERC Grant 2014-03994 to PJK and the Russian Science Foundation grant no. 23-14-00280, https://rscf.ru/project/23-14-00280/ to DVT (writing, microscopy analysis). GL was supported by GenomeBC (R02MSE) and AC was supported by NSERC-PGSD and UBC 4YF. The infrastructure of the laboratory of microbiology of the Papanin Institute for Biology of Inland Waters was supported by IBIW RAS program 121051100102-2.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2023.107964.

References

Aleoshin, V. v., Mylnikov, A.P., Mirzaeva, G.S., Mikhailov, K. v., Karpov, S.A., 2016. Heterokont predator Develorapax marinus gen. et sp. nov. – a model of the ochrophyte ancestor. Front Microbiol 7, 1–14. https://doi.org/10.3389/fmicb.2016.01194.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic Local Alignment Search Tool. J. Mol. Biol. 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Andrews, S., 2010. FastQC: A quality control tool for high throughput sequence data.

Azuma, T., Pánek, T., Tice, A.K., Kayama, M., Kobayashi, M., Miyashita, H., Suzaki, T., Yabuki, A., Brown, M.W., Kamikawa, R., 2022. An enigmatic stramenopile sheds light on early evolution in ochrophyta plastid organellogenesis. Mol. Biol. Evol. 39 https://doi.org/10.1093/molbev/msac065.

Bairoch, A., Apweiler, R., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M.J., Natale, D.A., O'Donovan, C., Redaschi, N., Yeh, L.S.L., 2005. The universal protein resource (UniProt). Nucleic Acids Res. 33 https://doi.org/10.1093/nar/gki070.

Barbera, P., Kozlov, A.M., Czech, L., Morel, B., Darriba, D., Flouri, T., Stamatakis, A., 2019. EPA-ng: massively parallel evolutionary placement of genetic sequences. Syst. Biol. https://doi.org/10.1093/sysbio/syy054.

Barcytė, D., Eikrem, W., Engesmo, A., Seoane, S., Wohlmann, J., Horák, A., Yurchenko, T., Eliáš, M., 2021. Olisthodiscus represents a new class of Ochrophyta. J. Phycol. 57, 1094–1118. https://doi.org/10.1111/jpy.13155.

Bateman, A., Martin, M.J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., Alpi, E., Bowler-Barnett, E.H., Britto, R., Bursteinas, B., Bye-A-Jee, H., Coetzee, R., Cukura, A., da Silva, A., Denny, P., Dogan, T., Ebenezer, T.G., Fan, J., Castro, L.G., Garmiri, P., Georghiou, G., Gonzales, L., Hatton-Ellis, E., Hussein, A., Ignatchenko, A., Insana, G., Ishtiaq, R., Jokinen, P., Joshi, V., Jyothi, D., Lock, A., Lopez, R., Luciani, A., Luo, J., Lussi, Y., MacDougall, A., Madeira, F., Mahmoudy, M., Menchi, M., Mishra, A., Moulang, K., Nightingale, A., Oliveira, C.S., Pundir, S., Qi, G., Raj, S., Rice, D., Lopez, M.R., Saidi, R., Sampson, J., Sawford, T., Speretta, E., Turner, E., Tyagi, N., Vasudev, P., Volynkin, V., Warner, K., Watkins, X., Zaru, R., Zellner, H., Bridge, A., Poux, S., Redaschi, N., Aimo, L., Argoud-Puy, G., Auchincloss, A., Axelsen, K., Bansal, P., Baratin, D., Blatter, M.C., Bolleman, J., Boutet, E., Breuza, L., Casals-Casas, C., de Castro, E., Echioukh, K.C., Coudert, E., Cuche, B., Doche, M., Dornevil, D., Estreicher, A., Famiglietti, M.L., Feuermann, M., Gasteiger, E., Gehant, S., Gerritsen, V., Gos, A., Gruaz-Gumowski, N., Hinz, U., Hulo, C., Hyka-Nouspikel, N., Jungo, F., Keller, G., Kerhornou, A., Lara, V., le Mercier, P., Lieberherr, D., Lombardot, T., Martin, X., Masson, P., Morgat, A., Neto, T.B., Paesano, S., Pedruzzi, I., Pilbout, S., Pourcel, L., Pozzato, M., Pruess, M., Rivoire, C., Sigrist, C., Sonesson, K., Stutz, A., Sundaram, S., Tognolli, M., Verbregue, L., Wu, C.H., Arighi, C.N., Arminski, L., Chen, C., Chen, Y., Garavelli, J.S., Huang, H., Laiho, K., McGarvey, P., Natale, D.A., Ross, K., Vinayaka, C.R., Wang, Q.,

Wang, Y., Yeh, L.S., Zhang, J., 2021. UniProt: The universal protein knowledgebase in 2021. D480–D489 Nucleic Acids Res. 49. https://doi.org/10.1093/nar/gkaa1100.
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. https://doi.org/10.1093/sysbio/syr010.
Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C., Al-Ghalith, G.A., Alexander, H., Alm, E.J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J.E., Bittinger, K., Brejnrod, A., Brislawn, C.J., Brown, C.T., Callahan, B.J., Caraballo-Rodríguez, A. M., Chase, J., Cope, E., Silva, R. da, Dorrestein, P.C., Douglas, G.M., Durall, D.M., Duvallet, C., Edwardson, C.F., Ernst, M., Estaki, M., Fouquier, J., Gauglitz, J.M., Gibson, D.L., Gonzalez, A., Gorlick, K., Guo, J., Hillmann, B., Holmes, S., Holste, H., Huttenhower, C., Huttley, G., Janssen, S., Jarmusch, A.K., Jiang, L., Kaehler, B., Kang, K. bin, Keefe, C.R., Keim, P., Kelley, S.T., Knights, D., Koester, I., Kosciolek, T., Kreps, J., Langille, M.G., Lee, J., Ley, R., Liu, Y.-X., Loftfield, E., Lozupone, C., Maher, M., Marotz, C., Martin, B., McDonald, D., McIver, L.J., Melnik, A. v, Metcalf, J.L., Morgan, S.C., Morton, J., Naimey, A.T., Navas-Molina, J.A., Nothias, L.F., Orchanian, S.B., Pearson, T., Peoples, S.L., Petras, D., Preuss, M.L., Pruesse, E., Rasmussen, L.B., Rivers, A., Michael S Robeson, I., Rosenthal, P., Segata, N., Shaffer, M., Shiffer, A., Sinha, R., Song, S.J., Spear, J.R., Swafford, A.D., Thompson, L.R., Torres, P.J., Trinh, P., Tripathi, A., Turnbaugh, P.J., Ul-Hasan, S., Hooft, J.J. van der, Vargas, F., Vázquez-Baeza, Y., Vogtmann, E., Hippel, M. von, Walters, W., Wan, Y., Wang, M., Warren, J., Weber, K.C., Williamson, C.H., Willis, A.D., Xu, Z.Z., Zaneveld, J.R., Zhang, Y., Knight, R., Caporaso, J.G., 2018. QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science. PeerJ Prepr. https://doi.org/10.7287/peerj.preprints.27295v1.
Buchfink, B., Xie, C., Huson, D.H., 2014. Fast and sensitive protein alignment using DIAMOND. Nat. Methods. https://doi.org/10.1038/nmeth.3176.
Burki, F., Kaplan, M., Tikhonenkov, D. v., Zlatogursky, V., Minh, B.Q., Radaykina, L. v., Smirnov, A., Mylnikov, A.P., Keeling, P.J., 2016. Untangling the early diversification of eukaryotes: A phylogenomic study of the evolutionary origins of centrohelida, haptophyta and cryptista. Proceedings of the Royal Society B: Biological Sciences 283. https://doi.org/10.1098/rspb.2015.2802.
Bushmanova, E., Antipov, D., Lapidus, A., Prjibelski, A.D., 2019. RnaSPAdes: A de novo transcriptome assembler and its application to RNA-Seq data. GigaScience 8. https://doi.org/10.1093/gigascience/giz100.
Callahan, B.J., Mcmurdie, P.J., Rosen, M.J., Han A, a.j.., A.W., 2016. DADA2: High resolution sample inference from Illumina amplicon data. Nat. Methods 13, 581–583. https://doi.org/10.1038/nmeth.3869.DADA2.
Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973. https://doi.org/10.1093/bioinformatics/btp348.
Caron, D.A., 2000. Symbiosis and mixotrophy among pelagic microorganisms. In: Kirchman, D.L. (Ed.), Microbial Ecology of the Oceans. John Wiley and Sons, New York, pp. 495–523.
Cavalier-Smith, T., 1993. The Protozoan Phylum Opalozoa. J. Eukaryot. Microbiol. 40, 609–615. https://doi.org/10.1111/j.1550-7408.1993.tb06117.x.
Cavalier-Smith, T., 1995. Zooflagellate phylogeny and classification. Tsitologiia 37, 1010–1029.
Cavalier-Smith, T., 1998. A revised six-kingdom system of life. Biol. Rev. Camb. Philos. Soc. 73, 203–266. https://doi.org/10.1017/s0006323198005167.
Cavalier-Smith, T., Chao, E.E.Y., 2006. Phylogeny and megasystematics of phagotrophic heterokonts (kingdom Chromista). J. Mol. Evol. 62, 388–420. https://doi.org/10.1007/s00239-004-0353-8.
Cavalier-Smith, T., Scoble, J.M., 2013. Phylogeny of Heterokonta: Incisomonas marina, a uniciliate gliding opalozoan related to Solenicola (Nanomonadea), and evidence that Actinophryida evolved from raphidophytes. Eur. J. Protistol. 49, 328–353. https://doi.org/10.1016/j.ejop.2012.09.002.
Challis, R., Richards, E., Rajan, J., Cochrane, G., Blaxter, M., 2020. BlobToolKit - interactive quality assessment of genome assemblies. G3: Genes. Genomes, Genetics 10, 1361–1374. https://doi.org/10.1534/g3.119.400908.
Cho, A., Tikhonenkov, D. v., Hehenberger, E., Karnkowska, A., Mylnikov, A.P., Keeling, P.J., 2022. Monophyly of diverse Bigyromonadea and their impact on phylogenomic relationships within stramenopiles. Mol Phylogenet Evol 171. https://doi.org/10.1016/j.ympev.2022.107468.
Choi, J., Park, J.S., 2020. Comparative analyses of the V4 and V9 regions of 18S rDNA for the extant eukaryotic community using the Illumina platform. Sci. Rep. 10 https://doi.org/10.1038/s41598-020-63561-z.
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. https://doi.org/10.3354/ame01361.
Czech, L., Barbera, P., Stamatakis, A., 2020. Genesis and Gappa: Processing, analyzing and visualizing phylogenetic (placement) data. Bioinformatics 36, 3263–3265. https://doi.org/10.1093/bioinformatics/btaa070.
Dayhoff, M.O., Schwartz, R.M., Orcutt, B.C., 1978. A model of evolutionary change in proteins. In: Atlas of Protein Sequence and Structure. National Biomedical Research Foundation, Silver Spring (MD), pp. 345–352.
Decelle, J., Romac, S., Sasaki, E., Not, F., Mahé, F., 2014. Intracellular diversity of the V4 and V9 regions of the 18S rRNA in marine protists (radiolarians) assessed by high-throughput sequencing. PLoS One 9. https://doi.org/10.1371/journal.pone.0104297.
del Campo, J., Kolisko, M., Boscaro, V., Santoferrara, L.F., Nenarokov, S., Massana, R., Guillou, L., Simpson, A., Berney, C., de Vargas, C., Brown, M.W., Keeling, P.J., Parfrey, L.W, 2018. EukRef: Phylogenetic curation of ribosomal RNA to enhance understanding of eukaryotic diversity and distribution. PLoS Biol 16, 1–14. https://doi.org/10.1371/journal.pbio.2005849.
Delsuc, F., Brinkmann, H., Philippe, H., 2005. Phylogenomics and the reconstruction of the tree of life. Nat. Rev. Genet. 6, 361–375. https://doi.org/10.1038/nrg1603.
Derelle, R., López-García, P., Timpano, H., Moreira, D., 2016. A Phylogenomic Framework to Study the Diversity and Evolution of Stramenopiles (=Heterokonts). Mol. Biol. Evol. 33, 2890–2898. https://doi.org/10.1093/molbev/msw168.
di Franco, A., Baurain, D., Glöckner, G., Melkonian, M., 2021. Lower statistical support with larger datasets : insights from the Ochrophyta radiation. Mol Biol Evol msab300. https://doi.org/https://doi.org/10.1093/molbev/msab300.
Dorrell, R.G., Azuma, T., Nomura, M., de Kerdrel, G.A., Paoli, L., Yang, S., Bowler, C., Ishii, K. ichiro, Miyashita, H., Gile, G.H., Kamikawa, R., 2019. Principles of plastid reductive evolution illuminated by nonphotosynthetic chrysophytes. Proc Natl Acad Sci U S A 116, 6914–6923. https://doi.org/10.1073/pnas.1819976116.
Dunthorn, M., Otto, J., Berger, S.A., Stamatakis, A., Mahé, F., Romac, S., de Vargas, C., Audic, S., Stock, A., Kauff, F., Stoeck, T., Edvardsen, B., Massana, R., Not, F., Simon, N., Zingone, A., 2014. Placing environmental next-generation sequencing amplicons from microbial eukaryotes into a phylogenetic context. Mol. Biol. Evol. 31, 993–1009. https://doi.org/10.1093/molbev/msu055.
Fenchel, T.M., Patterson, D.J., 1988. Cafeteria roenbergensis nov. gen., nov. sp., a heterotrophic microflagellate from marine plankton. Mar. Microb. Food Webs 3, 9–19.
Forster, D., Dunthorn, M., Mahé, F., Dolan, J.R., Audic, S., Bass, D., Bittner, L., Boutte, C., Christen, R., Claverie, J.M., Decelle, J., Edvardsen, B., Egge, E., Eikrem, W., Gobet, A., Kooistra, W.H.C.F., Logares, R., Massana, R., Montresor, M., Not, F., Ogata, H., Pawlowski, J., Pernice, M.C., Romac, S., Shalchian-Tabrizi, K., Simon, N., Richards, T.A., Santini, S., Sarno, D., Siano, R., Vaulot, D., Wincker, P., Zingone, A., de Vargas, C., Stoeck, T., 2016. Benthic protists: The under-charted majority. FEMS Microbiol. Ecol. 92 https://doi.org/10.1093/femsec/fiw120.
Gómez, F., Moreira, D., Benzerara, K., López-García, P., 2011. Solenicola setigera is the first characterized member of the abundant and cosmopolitan uncultured marine stramenopile group MAST-3. Environ. Microbiol. 13, 193–202. https://doi.org/10.1111/j.1462-2920.2010.02320.x.
Guillou, L., Chretiennot-dinet, M., Boulben, S., Moon-van der Staay, S.Y., Vaulot, D., 1999. Symbiomonas scintillans gen. et sp. nov. and Picophagus flagellatus gen. et sp. nov. (Heterokonta): Two New Heterotrophic Flagellates of Picoplanktonic Size. Protistology 150, 383–398. https://doi.org/10.1016/S1434-4610(99)70040-4.
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., 2013. The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. Nucleic Acids Res. 41, 597–604. https://doi.org/10.1093/nar/gks1160.
Haas, B., 2015. TransDecoder (Find Coding Regions Within Transcripts) [WWW Document]. GitHub. URL https://github.com/TransDecoder/TransDecoder.
Harding, T., Roger, A.J., Simpson, A.G.B., 2017. Adaptations to high salt in a halophilic protist: differential expression and gene acquisitions through duplications and gene transfers. Front. Microbiol. 8, 944. https://doi.org/10.3389/fmicb.2017.00944.
Hernandez, A.M., Ryan, J.F., 2021. Six-State Amino Acid Recoding is not an Effective Strategy to Offset Compositional Heterogeneity and Saturation in Phylogenetic Analyses. Syst. Biol. 70, 1200–1212. https://doi.org/10.1093/sysbio/syab027.
Jamy, M., Biwer, C., Vaulot, D., Obiol, A., Jing, H., Peura, S., Massana, R., Burki, F., 2022. Global patterns and rates of habitat transitions across the eukaryotic tree of life. Nat. Ecol. Evol. https://doi.org/10.1038/s41559-022-01838-4.
Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder : fast model selection for accurate phylogenetic estimates. Nat. Methods 14. https://doi.org/10.1038/nmeth.4285.
Karpov, S.A., Kersanach-, R., Williams, D.M., 1998. Ultrastructure and 18S rRNA gene sequence of a small heterotrophic flagellate Siluania monomastiga gen. et sp. nov. (bicosoecida). Eur. J. Protistol. 34, 415–425. https://doi.org/10.1016/S0932-4739(98)80010-2.
Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30, 772–780. https://doi.org/10.1093/molbev/mst010.
Kayama, M., Maciszewski, K., Yabuki, A., Miyashita, H., Karnkowska, A., Kamikawa, R., 2020. Highly Reduced Plastid Genomes of the Non-photosynthetic Dictyochophyceans Pteridomonas spp. (Ochrophyta, SAR) Are Retained for tRNA-Glu-Based Organellar Heme Biosynthesis. Front Plant Sci 11. https://doi.org/10.3389/fpls.2020.602455.
Keeling, P.J., Burki, F., Wilcox, H.M., Allam, B., Allen, E.E., Amaral-Zettler, L.A., Armbrust, E.V., Archibald, J.M., Bharti, A.K., Bell, C.J., Beszteri, B., Bidle, K.D., Cameron, C.T., Campbell, L., Caron, D.A., Cattolico, R.A., Collier, J.L., Coyne, K., Davy, S.K., Deschamps, P., Dyhrman, S.T., Edvardsen, B., Gates, R.D., Gobler, C.J., Greenwood, S.J., Guida, S.M., Jacobi, J.L., Jakobsen, K.S., James, E.R., Jenkins, B., John, U., Johnson, M.D., Juhl, A.R., Kamp, A., Katz, L.A., Kiene, R., Kudryavtsev, A., Leander, B.S., Lin, S., Lovejoy, C., Lynn, D., Marchetti, A., McManus, G., Nedelcu, A. M., Menden-Deuer, S., Miceli, C., Mock, T., Montresor, M., Moran, M.A., Murray, S., Nadathur, G., Nagai, S., Ngam, P.B., Palenik, B., Pawlowski, J., Petroni, G., Piganeau, G., Posewitz, M.C., Rengefors, K., Romano, G., Rumpho, M.E., Rynearson, T., Schilling, K.B., Schroeder, D.C., Simpson, A.G.B., Slamovits, C.H., Smith, D.R., Smith, G.J., Smith, S.R., Sosik, H.M., Stief, P., Theriot, E., Twary, S.N., Umale, P.E., Vaulot, D., Wawrik, B., Wheeler, G.L., Wilson, W.H., Xu, Y., Zingone, A., Worden, A.Z., 2014. The Marine Microbial Eukaryote Transcriptome Sequencing

Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing. PLoS Biol. 12, e1001889.
Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., Stamatakis, A., 2019. RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 35, 4453–4455. https://doi.org/10.1093/bioinformatics/btz305.
Labarre, A., Lopez-Escardo, D., Leonard, G., Bucchini, F., Obiol, A., Cruaud, C., Sieracki, M.E., Jaillon, O., Wincker, P., Vandepoele, K., Logares, R., Massana, R., n.d.. 2021. Comparative genomics reveals new functional insights in uncultured MAST species. ISME J. 15, 1767–1781. doi:10.1038/s41396-020-00885-8.
Laetsch, D.R., Blaxter, M.L., 2017. BlobTools : Interrogation of genome assemblies [version 1; peer review : 2 approved with reservations]. F1000Res 6, 1–16.
Larsen, J., Patterson, D.J., 1990. Some flagellates (Protista) from tropical marine sediments. J. Nat. Hist. 24, 801–937. https://doi.org/10.1080/00222939000770571.
Lartillot, N., Lepage, T., Blanquart, S., 2009. PhyloBayes 3: A Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics 25, 2286–2288. https://doi.org/10.1093/bioinformatics/btp368.
Lartillot, N., Philippe, H., 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol. Biol. Evol. 21, 1095–1109. https://doi.org/10.1093/molbev/msh112.
Lee, W.J., 2002. Redescription of the Rare Heterotrophic Flagellate (Protista) -Phyllomitus undulans Stein, 1878, and Erection of a New Genus -Pseudophyllomitus gen. n. Acta Protozool. 41, 375–381.
Lee, H.B., Jeong, D.H., Cho, B.C., Park, J.S., 2021. The diversity patterns of rare to abundant microbial eukaryotes across a broad range of salinities in a solar saltern. Microb. Ecol. https://doi.org/10.1007/s00248-021-01918-1.
Lee, W.J., Patterson, D.J., 2002. Abundance and biomass of heterotrophic flagellates, and factors controlling their abundance and distribution in sediments of Botany Bay. Microb. Ecol. 43, 467–481. https://doi.org/10.1007/s00248-002-2000-5.
Li, W., Godzik, A., 2006. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22, 1658–1659. https://doi.org/10.1093/bioinformatics/btl158.
Logares, R., Audic, S., Santini, S., Pernice, M.C., de Vargas, C., Massana, R., 2012. Diversity patterns and activity of uncultured marine heterotrophic flagellates unveiled with pyrosequencing. ISME J. 6, 1823–1833. https://doi.org/10.1038/ismej.2012.36.
Luo, C., Tsementzi, D., Kyrpides, N., Read, T., Konstantinidis, K.T., 2012. Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA Sample. PLoS One 7, e30087. https://doi.org/10.1371/journal.pone.0030087.
Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. Embnet J 17, 10–12. https://doi.org/10.14806/ej.17.1.200.
Massana, R., Castresana, J., Balague, V., Guillou, L., Romari, K., Groisillier, A., Valentin, K., Pedros-Alio, C., 2004. Phylogenetic and ecological analysis of novel marine stramenopiles. Appl. Environ. Microbiol. 70, 3528–3534. https://doi.org/10.1128/AEM.70.6.3528.
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. https://doi.org/10.1111/j.1462-2920.2006.01042.x.
Massana, R., Unrein, F., Rodríguez-Martínez, R., Forn, I., Lefort, T., Pinhassi, J., Not, F., 2009. Grazing rates and functional diversity of uncultured heterotrophic flagellates. ISME J. 3, 588–595. https://doi.org/10.1038/ismej.2008.130.
Massana, R., Campo, J., Sieracki, M.E., Audic, S., Logares, R., 2014. Exploring the uncultured microeukaryote majority in the oceans: reevaluation of ribogroups within stramenopiles. Int. Soc. Microbial Ecol. 8, 854–866. https://doi.org/10.1038/ismej.2013.204.
Massana, R., Gobet, A., Audic, S., Bass, D., Bittner, L., Boutte, C., Chambouvet, A., Christen, R., Claverie, J.M., Decelle, J., Dolan, J.R., Dunthorn, M., Edvardsen, B., Forn, I., Forster, D., Guillou, L., Jaillon, O., Kooistra, W.H.C.F., Logares, R., Mahé, F., Not, F., Ogata, H., Pawlowski, J., Pernice, M.C., Probert, I., Romac, S., Richards, T., Santini, S., Shalchian-Tabrizi, K., Siano, R., Simon, N., Stoeck, T., Vaulot, D., Zingone, A., de Vargas, C., 2015. Marine protist diversity in European coastal waters and sediments as revealed by high-throughput sequencing. Environ. Microbiol. 17, 4035–4049. https://doi.org/10.1111/1462-2920.12955.
Medlin, L., Elwood, H.J., Stickel, S., Sogin, M.L., 1988. The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. Gene 71, 491–499. https://doi.org/10.1016/0378-1119(88)90066-2.
Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear, R., Teeling, E., 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol. Biol. Evol. 37, 1530–1534. https://doi.org/10.1093/molbev/msaa015.
Moestrup, Ø., Thomsen, H.A., 1976. Fine structural studies on the flagellate genus Bicoeca I. - Bicoeca maris with particular emphasis on the flagellar apparatus. Protistologica 12, 101–120.
Moriya, M., Nakayama, T., Inouye, I., 2000. Ultrastructure and 18S rDNA Sequence Analysis of Wobblia lunata gen. et sp. nov., a New Heterotrophic Flagellate (Stramenopiles, Incertae Sedis). Protistology 151, 41–55.
Moriya, M., Nakayama, T., Inouye, I., 2002. A New Class of the Stramenopiles, Placididea Classis nova: Description of Placidia cafeteriopsis gen. et sp. nov. Protist 153, 143–156. https://doi.org/10.1078/1434-4610-00093.
Nakai, R., Naganuma, T., 2015. Diversity and ecology of thraustochytrid protists in the marine environment. In: Marine Protists: Diversity and Dynamics. Springer Japan, pp. 331–346. https://doi.org/10.1007/978-4-431-55130-0_13.
Nguyen, L.T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274. https://doi.org/10.1093/molbev/msu300.
Noguchi, F., Tanifuji, G., Brown, M.W., Fujikura, K., Takishita, K., 2016. Complex evolution of two types of cardiolipin synthase in the eukaryotic lineage stramenopiles. Mol. Phylogenet. Evol. 101, 133–141. https://doi.org/10.1016/j.ympev.2016.05.011.
O'Kelly, C.J., Nerad, T.A., 1998. Kinetid Architecture and Bicosoecid Affinities of the Marine Heterotrophic Nanoflagellate Caecitellus parvulus (Griessmann, 1913) Patterson et al., 1993. Eur. J. Protistol. 34, 369–375. https://doi.org/10.1016/S0932-4739(98)80006-0.
Okamura, T., Kondo, R., 2015. Suigetsumonas clinomigrationis gen. et sp. nov., a Novel Facultative Anaerobic Nanoflagellate Isolated from the Meromictic Lake Suigetsu. Japan. Protist 166, 409–421. https://doi.org/10.1016/j.protis.2015.06.003.
Onsbring, H., Tice, A.K., Barton, B.T., Brown, M.W., Ettema, T.J.G., 2020. An efficient single-cell transcriptomics workflow for microbial eukaryotes benchmarked on Giardia intestinalis cells. BMC Genomics 21, 1–9. https://doi.org/10.1186/s12864-020-06858-7.
Park, J.S., Cho, B.C., Simpson, A.G.B., 2006. Halocafeteria seosinensis gen. et sp. nov. (Bicosoecida), a halophilic bacterivorous nanoflagellate isolated from a solar saltern. Extremophiles 10, 493–504. https://doi.org/10.1007/s00792-006-0001-x.
Park, J.S., Simpson, A.G.B., 2010. Characterization of halotolerant Bicosoecida and Placididea (Stramenopila) that are distinct from marine forms, and the phylogenetic pattern of salinity preference in heterotrophic stramenopiles. Environ. Microbiol. 12, 1173–1184. https://doi.org/10.1111/j.1462-2920.2010.02158.x.
Patterson, D.J., Nygaard, K., Steinberg, G., Turley, C.M., 1993. Heterotrophic flagellates and other protists associated with oceanic detritus throughout the water column in the mid North Atlantic. J. Mar. Biol. Assoc. U. K. 73, 67–95. https://doi.org/10.1017/S0025315400032653.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 12, 2825–2830.
Philippe, H., Brinkmann, H., Lavrov, D. v., Littlewood, D.T.J., Manuel, M., Wörheide, G., Baurain, D., 2011. Resolving difficult phylogenetic questions: Why more sequences are not enough. PLoS Biol 9. https://doi.org/10.1371/journal.pbio.1000602.
Philippe, H., Zhou, Y., Brinkmann, H., Rodrigue, N., Delsuc, F., 2005. Heterotachy and long-branch attraction in phylogenetics. BMC Evol. Biol. 5 https://doi.org/10.1186/1471-2148-5-50.
Picelli, S., Faridani, O.R., Björklund, Å.K., Winberg, G., Sagasser, S., Sandberg, R., 2014. Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. 9, 171–181. https://doi.org/10.1038/nprot.2014.006.
Piwosz, K., Pernthaler, J., 2010. Seasonal population dynamics and trophic role of planktonic nanoflagellates in coastal surface waters of the Southern Baltic Sea. Environ. Microbiol. 12, 364–377. https://doi.org/10.1111/j.1462-2920.2009.02074.x.
Quang, L.S., Gascuel, O., Lartillot, N., 2008. Empirical profile mixture models for phylogenetic reconstruction. Bioinformatics 24, 2317–2323. https://doi.org/10.1093/bioinformatics/btn445.
Richter, D.J., Berney, C., Strassert, J.F.H., Poh, Y.-P., Herman, E.K., Muñoz-Gómez, S.A., Wideman, J.G., Burki, F., de Vargas, C., 2022. EukProt: A database of genome-scale predicted proteins across the diversity of eukaryotes. Peer Community Journal 2. https://doi.org/10.24072/pci.
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. https://doi.org/10.1111/j.1462-2920.2008.01779.x.
Rodríguez-Martínez, R., Rocap, G., Logares, R., Romac, S., Massana, R., 2012. Low evolutionary diversification in a widespread and abundant uncultured protist (MAST-4). Mol. Biol. Evol. 29, 1393–1406. https://doi.org/10.1093/molbev/msr303.
Rodríguez-Martínez, R., Leonard, G., Milner, D.S., Sudek, S., Conway, M., Moore, K., Hudson, T., Mahé, F., Keeling, P.J., Santoro, A.E., Worden, A.Z., Richards, T.A., 2020. Controlled sampling of ribosomally active protistan diversity in sediment-surface layers identifies putative players in the marine carbon sink. ISME J. 14, 984–998. https://doi.org/10.1038/s41396-019-0581-y.
Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahé, F., 2016. VSEARCH: A versatile open source tool for metagenomics. PeerJ 2016. https://doi.org/10.7717/peerj.2584.
Rybarski, A.E., Nitsche, F., Soo Park, J., Filz, P., Schmidt, P., Kondo, R., GB Simpson, A., Arndt, H., 2021. Revision of the phylogeny of Placididea (Stramenopiles): Molecular and morphological diversity of novel placidid protists from extreme aquatic environments. Eur J Protistol 81. https://doi.org/10.1016/j.ejop.2021.125809.
Saville-Kent, W., 1880. A manual of the Infusoria : including a description of all known flagellate, ciliate, and tentaculiferous protozoa, British and foreign, and an account of the organization and the affinities of the sponges by W. Saville Kent. D. Bogue, London : https://doi.org/10.5962/bhl.title.10143.
Schoenle, A., Hohlfeld, M., Rosse, M., Filz, P., Wylezich, C., Nitsche, F., Arndt, H., 2020. Global comparison of bicosoecid Cafeteria-like flagellates from the deep ocean and surface waters, with reorganization of the family Cafeteriaceae. Eur. J. Protistol. 73 https://doi.org/10.1016/j.ejop.2019.125665.
Schoenle, A., Hohlfeld, M., Hermanns, K., Mahé, F., de Vargas, C., Nitsche, F., Arndt, H., 2021. High and specific diversity of protists in the deep-sea basins dominated by diplonemids, kinetoplastids, ciliates and foraminiferans. Commun Biol 4. https://doi.org/10.1038/s42003-021-02012-5.
Schoenle, A., Hohlfeld, M., Rybarski, A., Sachs, M., Freches, E., Wiechmann, K., Nitsche, F., Arndt, H., 2022. Cafeteria in extreme environments: investigations on C. burkhardae and three new species from the atacama desert and the deep ocean. Eur. J. Protistol. 125905 https://doi.org/10.1016/j.ejop.2022.125905.

Seemann, T., 2007. Barrnap 0.9: BAsic Rapid Ribosomal RNA Predictor [WWW Document]. GitHub. URL https://github.com/tseemann/barrnap.
Ševcíková, T., Horák, A., Klimeš, V., Zbránková, V., Demir-Hilton, E., Sudek, S., Jenkins, J., Schmutz, J., Pribyl, P., Fousek, J., Vlcek, C., Lang, B.F., Oborník, M., Worden, A.Z., Eliáš, M., 2015. Updating algal evolutionary relationships through plastid genome sequencing: Did alveolate plastids emerge through endosymbiosis of an ochrophyte? Sci. Rep. 5 https://doi.org/10.1038/srep10134.
Shimodaira, H., 2002. An approximately unbiased test of phylogenetic tree selection. Syst. Biol. 51, 492–508. https://doi.org/10.1080/10635150290069913.
Shiratori, T., Nakayama, T., Ishida, K. ichiro, 2015. A New Deep-branching Stramenopile, Platysulcus tardus gen. nov., sp. nov. Protist 166, 337–348. https://doi.org/10.1016/j.protis.2015.05.001.
Shiratori, T., Thakur, R., Ishida, K. ichiro, 2017. Pseudophyllomitus vesiculosus (Larsen and Patterson 1990) Lee, 2002, a Poorly Studied Phagotrophic Biflagellate is the First Characterized Member of Stramenopile Environmental Clade MAST-6. Protist 168, 439–451. https://doi.org/10.1016/j.protis.2017.06.004.
Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E. v., Zdobnov, E.M., 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351.
Song, L., Florea, L., 2015. Rcorrector: Efficient and accurate error correction for Illumina RNA-seq reads. GigaScience 4. https://doi.org/10.1186/s13742-015-0089-y.
Stamatakis, A., 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033.
Team, Rs., 2016. RStudio: Integrated Development for R. https://doi.org/10.1002/ejoc.201200111.
Thakur, R., Shiratori, T., Ishida, K. ichiro, 2019. Taxon-rich Multigene Phylogenetic Analyses Resolve the Phylogenetic Relationship Among Deep-branching Stramenopiles. Protist 170, 125682. https://doi.org/10.1016/j.protis.2019.125682.
Tice, A.K., Žihala, D., Pánek, T., Jones, R.E., Salomaki, E.D., Nenarokov, S., Burki, F., Eliáš, M., Eme, L., Roger, A.J., Rokas, A., Shen, X.-X., Strassert, J.F.H., Kolísko, M., Brown, M.W., 2021. PhyloFisher: A phylogenomic package for resolving eukaryotic relationships. PLoS Biol. 19, e3001365.
Tikhonenkov, D. v., Janouškovec, J., Keeling, P.J., Mylnikov, A.P., 2016. The Morphology, Ultrastructure and SSU rRNA Gene Sequence of a New Freshwater Flagellate, Neobodo borokensis n. sp. (Kinetoplastea, Excavata). Journal of Eukaryotic Microbiology 63, 220–232. https://doi.org/10.1111/jeu.12271.
Tikhonenkov, D. v., Mikhailov, K. v., Gawryluk, R.M.R., Belyaev, A.O., Mathur, V., Karpov, S.A., Zagumyonnyi, D.G., Borodina, A.S., Prokina, K.I., Mylnikov, A.P., Aleoshin, V. v., Keeling, P.J., 2022. Microbial predators form a new supergroup of eukaryotes. Nature. https://doi.org/10.1038/s41586-022-05511-5.
Vacic, V., Uversky, V.N., Dunker, A.K., Lonardi, S., 2007. Composition Profiler: A tool for discovery and visualization of amino acid composition differences. BMC Bioinf. 8 https://doi.org/10.1186/1471-2105-8-211.
Verhagen, F.J.M., Zölffel, M., Brugerolle, G., Patterson, D.J., 1994. Adriamonas peritocrescens gen. nov., sp. nov., a new free-living soil flagellate (Protista, Pseudodendromonadidae Incertae Sedis). Eur. J. Protistol. 30, 295–308. https://doi.org/10.1016/S0932-4739(11)80076-3.
Wang, H.C., Susko, E., Spencer, M., Roger, A.J., 2008. Topological estimation biases with covarion evolution. J. Mol. Evol. 66, 50–60. https://doi.org/10.1007/s00239-007-9062-4.
Wägele, J., Mayer, C., 2007. Visualizing differences in phylogenetic information content of alignments and distinction of three classes of long-branch effects. BMC Evolutionary Biology 7, 147. https://doi.org/10.1186/1471-2148-7-147.
Wang, H.C., Minh, B.Q., Susko, E., Roger, A.J., 2018. Modeling Site Heterogeneity with Posterior Mean Site Frequency Profiles Accelerates Accurate Phylogenomic Estimation. Syst. Biol. 67, 216–235. https://doi.org/10.1093/sysbio/syx068.
Wickham, H., 2016. ggplot2: Elegant graphics for data analysis. Springer-Verlag, New York.
Wu, W., Huang, B., 2019. Protist diversity and community assembly in surface sediments of the South China Sea. Microbiologyopen 8. https://doi.org/10.1002/mbo3.891.
Yubuki, N., Leander, B.S., Silberman, J.D., 2010. Ultrastructure and Molecular Phylogenetic Position of a Novel Phagotrophic Stramenopile from Low Oxygen Environments: Rictus lutensis gen. et sp. nov. (Bicosoecida, incertae sedis). Protist 161, 264–278. https://doi.org/10.1016/j.protis.2009.10.004.
Yubuki, N., Pánek, T., Yabuki, A., Čepička, I., Takishita, K., Inagaki, Y., Leander, B.S., 2015. Morphological Identities of Two Different Marine Stramenopile Environmental Sequence Clades: Bicosoeca kenaiensis (Hilliard, 1971) and Cantina marsupialis (Larsen and Patterson, 1990) gen. nov., comb. nov. J. Eukaryot. Microbiol. 62, 532–542. https://doi.org/10.1111/jeu.12207.