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

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
Molecular Phylogenetics and Evolution 190 (2024) 107964

A. Cho et al.

2012; Massana et al., 2015; Rodríguez-Martínez et al., 2020). Extensive
are common in sediments while rare in pelagic samples (Logares et al.,
Massana et al., 2015). Notably, MAST-6 (along with MAST-9, and -12)
from environmental sampling efforts (Logares et al., 2012; Massana
lineage,

***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. sali-
nus*, 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 ecolog-
ically 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 Scole, 2013) which was
formerly described as clade I that only contained environmental se-
quences (Cavalier-Smith and Chao, 2006). Previous phylogenomic an-
alyses showed the monophyly of MAST-4 and MAST-6, forming a sister
group to Labyrinthulida, 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
(Rodriguez-Martínez et al., 2020). Interestingly, two morphotypes of MAST-6 were
observed in plankton samples that differ in sizes and seasonal abund-
ances (Piwoz and Pernthaler, 2010), with the larger morphotype
(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 Placi-
ididae, a class within Opalozoa. Like Sagenista, the phylogenetic di-
versity of Placiididae 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 *Placiididae* sp. (the CarsonLab strain
formerly mis-labelled as *Cafeteria*) (Keeling et al., 2014). Placiidaceae 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 Placiididae (Moriya et al., 2002)
do not amplify well with V4-targeting primers (Lee et al., 2022). Conse-
quently, 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 placiidaceans (Moriya et al., 2002; Park and
Simpson, 2010; Rybarski et al., 2021). Isolated placiidaceans are often
from hypersaline environments (>40 %≤), however, many characterized
halophilic placiidaceans 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 Placiididae species, providing microscopic obser-
vations, transcriptomic data, and SSU rRNA sequence comparisons with
previously generated environmental amplicon data. Strains PhM-7
(Placiididae, *Haloplacidia sinaii* and Colp-33 (MAST-6, *Vonaramonas
teuhelue*) were maintained in culture for a year, but subsequently lost.
Two other MAST-6 species (*Mastrreximus talaamin* and *Pseudophyllomi-
tus* sp. BSC2) were obtained by single cell isolations. We also report
transcriptomes from two cultured species, **Symbionomas scintillans**
RCC257 (Guillou et al., 1999) and *Cacicellus* sp. RCC1078 (O’Kelly
and Neral, 1998), to further fill out the diversity of deep-branching stra-
menopiles for phylogenomic analyses. We report the relative abundance
and diversity of the four new species of MAST-6 and Placiididae in
publicly available environmental sequence surveys, and re-examine
stramenopile phylogeny, particularly with the aim of resolving re-
lationships within the Bigyra using a multi-genie approach.

2. Materials and methods

2.1. Sample collection and imaging

Strain PhM-7 (*Haloplacidia sinaii* 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 (*Vonaramonas teuhelue* 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
*Procrystalidia 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) imag-
ing 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 % OsO4 (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
series without any artificial food. 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 Symbionomas scintillans** strain RCC257 and *Cacicellus* sp.
strain RCC1078 were obtained from the Roscoff culture collection
Cells of PhM-7 (H. sinai) and Colp-33 (V. tehuelsche) 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 (2,000g for PhM-7 and 1,000g for Colp-33, both at room temperature) onto the 0.8 µm membrane of a Vivaclear mini column (Sartorium Stedim Biotech Gmg, Cat. No. VK01P042). Total RNA was then extracted using a RNAQeous-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 3220g 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 Phasmaker™ (Invitrogen) tubes. The quality and quantity of the RNA yield was determined using a NanoDrop 1000 Spectrophotometer 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. tehuelsche), PRCS, 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) were carried out by the Sequencing and Bioinformatics Consortium (University of British Columbia, BC Canada), using the Illumina Nexterra™ 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 Nexterra™ 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, Chlorophonium 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 Trimomatic 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, SLINDINGWINDOW:4:15, and MINLEN:36. Processed forward, reverse, and unpaired transcripts were assembled using the de novo transcriptome assembler masSPAdes 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 (Procytophia sorokinii for PhM-7 and Colp-33, and C. capillatum for A. sol) using BLASTn, followed by the removal of contigs with ≥ 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 bmarp v0.9 (Seemann, 2007). For S. scintillans and Caecitellus sp., SSU rRNA sequences were generated by polymerase chain reaction (PCR) amplification of cDNA using 18SFP 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 (7000g, room temperature) onto an 0.8 µm membrane of a Vivaomni mini column (Sartorius Stedim Biotech Gmg, 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 DNA 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 (Schoene et al., 2021), and ISMe2020 (Rodriguez-Martinez et al., 2020) (Table 1). For the sixth dataset (designated as ESBig), we obtained ten SSU rRNA sequences (ESBig130-139) assigned to Placidae, directly from the authors (Lee et al., 2022) (Table 1). These studies examined sediments
from different bodies of water across the US, Europe, and Asia, including the South China Sea, North Atlantic Ocean, Mariana Basin, Philippine Basins, and the freshwater lake Polleven (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 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 2006). The trained classifier was then used to assign taxonomy to filtered representative classifier fit-classifier-naïve-bayes (Pedregosa et al., 2011). The trained classifier was further processed with ‘uchime-denovo′ and ‘–input-format SingleEndFastqManifestPhred33′ options. After trimming the raw reads with respective primer-pair sequences (Martin, 2011), both 454 pyrosequencing and Illumina sequences (Martin, 2011) were further processed with EPA-ng v0.3.8 (Barbera et al., 2019). This method used relevant environmental sequences from the PR2 database, GenBank, and modified SSU rRNA sequences were aligned with MAFFT v7.481 (Katoh et al., 2013) and trimmed using trimAl 1.2rev59 (-gt 0.3, -st 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-Gutierrez 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 GTR+Gamma model with 1000 replicates of ultrafast bootstrap (UBF). 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 identified as significant.

Table 1

<table>
<thead>
<tr>
<th>Dataset design</th>
<th>Sample environment</th>
<th>Sequencing technology</th>
<th>Targeted 18S rRNA region</th>
<th>Extracted length (bp)</th>
<th>Number of ASVs</th>
<th>Sample number</th>
<th>BioProject</th>
</tr>
</thead>
<tbody>
<tr>
<td>BioMarKs*</td>
<td>Seafloor sediment</td>
<td>454 GS FLX Titanium</td>
<td>V4</td>
<td>380-384</td>
<td>0</td>
<td>24*</td>
<td>PRJEB9133 (Dunthorn et al., 2014; Manas et al., 2015)</td>
</tr>
<tr>
<td>SouthChina</td>
<td>Seafloor sediment</td>
<td>454 GS FLX Titanium</td>
<td>V1-V2</td>
<td>396-429</td>
<td>0</td>
<td>6</td>
<td>PRJNA341446 (Wu and Huang, 2019)</td>
</tr>
<tr>
<td>Norwegian</td>
<td>Marine and brackish sediment</td>
<td>Illumina MiSeq paired-end</td>
<td>V4</td>
<td>426-429</td>
<td>0</td>
<td>24</td>
<td>PRJEB24876, PRJEB24158, PRJEB24888</td>
</tr>
<tr>
<td>Deepsea</td>
<td>Abyssal seafloor sediment</td>
<td>Illumina Genome Analyzer II paired-end</td>
<td>V9</td>
<td>134-138</td>
<td>15</td>
<td>20</td>
<td>PRJNA655512 (Schoenle et al., 2021)</td>
</tr>
<tr>
<td>ISME2020</td>
<td>Seafloor sediment</td>
<td>Illumina MiSeq paired-end</td>
<td>V4</td>
<td>182-425</td>
<td>0</td>
<td>49</td>
<td>PRJNA521526 (Rodríguez-Martínez et al., 2020)</td>
</tr>
<tr>
<td>ESBig**</td>
<td>Solar saltern</td>
<td>Illumina MiSeq paired-end</td>
<td>V9**</td>
<td>128-154</td>
<td>0</td>
<td><strong>Accession number: MZ299713, MZ299719, MZ299824, MZ299825, MZ299969, MZ300048, MZ300314, MZ300350, MZ300439, MZ300768, ERR861849, ERR861843, ERR861894, ERR861893, ERR861890, ERR861900, ERR861905, ERR861910, ERR861911, ERR861915-ERR861917</strong> (Lee et al., 2021)</td>
<td>**PRJNA732544 (Lee et al., 2021)</td>
</tr>
</tbody>
</table>
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, placideidan, and other species of Bigyra, another SSU rRNA phylogenetic tree was constructed without short ampiclon 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 1.2rev9 (-gt 0.3, -st 0.001) (Capella-Gutierrez 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 phylogenetic 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 belonging to an outgroup) were selected, resulting in a 240 gene set with 76,517 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., 2022; 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 + G60 + 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 (Vaic et al., 2007) was used with default settings to compare relative abundances of GARP vs. FYINK amino acids with “SwissProt 51” (Raierch 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 + F0 model and 100 replicates of standard bootstrap with RAxML-NG v1.1.0 (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 bcmp.

3. Results

3.1. Phylogenomic tree of stramenopiles

The final phylogenomic matrix used for constructing the main phylogenetic 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 Pseudophyllum 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 Symbimonomas scintillans (Fig. 1). The BUSCO scores showed a similar pattern where Pseudophyllum 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 + G60 + 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 Sagenista. The monophyly of RPX, CCS, and BBDPe was moderately supported (84 %). Bacillariophyceae, Phaeophyceae, 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-matrices, 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, Pseudophyllum sp. BSC2 and V. tehuelche formed a clade with Pseudophyllum vesticulosus, with Pseudophyllum sp. BSC2 being the immediate sister lineage to V. vesticulosus. MAST-6, MAST-4, and Labyrinthulae all formed a monophyletic group, Sagenista. In the tree inferred for the MASTer-matrix (Fig. S3), 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 Placidooza (Nanomonadea and Opalinata), however, the support value
for the Nanomonadea (MAST-3) and Opalinata (Blastocystis sp.) clade was weak (70 % bootstrap). Placidooza 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 Platyctiscus 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 Halocaceteria 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 are well supported.
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 Pinguioophyceae 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 Placidooza), amino acid compositions of all the taxa within Placididea 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 phylogenetic 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 Pinguiophiaceae 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 Pinguiophiaceae were the most inconsistent. In all chains, Bigynomorpha formed a sister lineage to Ochrophyta and two out of four chains had monophyletic Bigyra (excluding P. tardo) (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 BioMarkks 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 diversity of the MAST-6 group, which were largely divided into four sub-groups (Fig. 3). Interestingly, we did not find shared 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 geographical 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 BioMarkks 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 BioMarkks) and limited universality of the V9 primer used in Deepsea dataset (Fig. S6C) (Oxcell et al., 2014). As QIME2 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 BioMarkks and ISME2020, respectively. For other MAST lineages, ASVs assigned to MAST-1, –9, –3, –6, 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 rRNA tree was constructed including the 10 Placidean 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 placidean sequences were placed across the major subgroups of Placideidae, 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 ≥ 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.)
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 sinuousoidal 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 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.


**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 marine stramnepolice, εξι (Greek ξι, 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-


**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.
A. Cho et al.  


Type species. Vomastramonas tehuechelle

Vomastramonas tehuechelle 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 tehuechelle strain Colp-33. 

Gene sequence: The SSU rRNA gene sequence has the GenBank Accession Number Q9999986. 

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

Etymology: Tehuechelle 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-A6DFAB139D1

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


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 acronemotic 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. cosmopólitá 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 sinaí strain PhM-7. 

Gene sequence: The SSU rRNA gene sequence has the GenBank Accession Number QQ999982. 

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 יָהוּד=yohuḏ pronounced /ʃiːm/.


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 tehuechele, 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 sub-groups 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 tehuechelle, 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 *Pseudophyllumus* sp. BSC2-and *V. tehueluca*-assigned ASVs, and low relative abundance of ASVs assigned to *P. veliculasius* 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 Gdańsk shortly after freshwater inflow, followed by a substantial decline in relative abundance (Piwosz and Pernthaler, 2010). All datasets except BioMarkKs were sampled mostly in September, with some sampled in August and July (sampling months for BioMarkKs 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 sinaí and its implication in trait evolution

*Haploplacidia sinaí* is the fourth new species reported here. *Haploplacidia sinaí* 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. sinaí* was found in a relatively high salinity environment. Although the present study did not detect any ASVs assigned to *H. sinaí*, based on its relationship (Fig. 3) with other isolated cells cultured in broad range of salinity, *H. sinaí* might also be found in non-hypersaline environments. Absence of ASVs assigned to *H. sinaí* 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 halotolerance in these deep-branched 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 *Caeccitellus* sp., and this atypical mastigoneme-lacking group (*O. scintillans* and *Caeccitellus burkhardiae* (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 genes (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 18S rRNA phylogeny showed *H. seosinensis* forming a sister lineage to a clade composed of *Caecefteria* spp. and *Caeccitellus* 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 bikonian SSU 18S rRNA genes, as indicated by the long branch length of *S. scintillans* and *C. burkhardiae* (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), *Pseudobooldo* spp. (Griessmann, 1913), and freshwater or soil bikosia, including *Siluania monomastigia* (Karpov et al., 1998), *Nerada mexicana* (Cavalier-Smith and Chao, 2006), *Adriaonas peri-tocrescens* (Verhagen et al., 1994), and *Parramosa 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; Brezulianu 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 (Dorrelle 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 orthology, 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), *Actinophys sol* is not placed sister to the rest of the Ochrophyta. Rather, it forms a weakly-supported clade with *Microchloropsis gudidates* (Eustigmatophyceae). As a single transcriptome is representing each of Eustigmatophyceae and Actinophryidae, 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, Pinguio phyceae, 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 (Barcet 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; Sevcikova 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 Hypochytriomycetes (Dick 1983), forming a sister lineage to Ochrophyta in all four consensus trees (Cho et al., 2022).

Within the monophyletic Placididea (*Placididea + Nanomemone + Blastosyctis*), the relationship among the sub-groups was no longer strongly supported, in the most recent three of our 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 placidoozoan 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 the most recent genomic data of MAST-6 species may be rare. Additionally, sample timing affected by seasonality and salinity (Piwosz and Pernthaler, 2010). 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).
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 TRNA-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. Placidea 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 placideid (H. sinui), we observed a topology change in Placidae that conflicts with previous studies. Based on the long unbroken branch leading to Placidea 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-sibling sister lineage, Bikosia, is likely that currently available models cannot resolve the true relationship of Placidae. 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 phylogenetic 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; Rodriguez-Martinez et al., 2012). One of the abundantly detected MAST-6 is closely related to the newly described M. tlaunin (PRCS), while few or no ASVs were detected for V. tehuelche and Pseudophyllumus 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. Tikho- nenkov: Writing – original draft, Writing – review & editing, Visualization. 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. Vojtech Zarsky. We thank Dr. Andrew Roger (Dalhouse) 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://rsfc.ru/project/23-14-00280/ to DVT (writing, microscopy analysis). GL was supported by GenomeBC (R02MSE) and AC was supported by NSERC-PSGD 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 12105110102-2.

Appendix A. Supplementary material

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

References


