

Hemiselmis aquamarina sp. nov. (Cryptomonadales, Cryptophyceae), a
cryptophyte with a novel phycobiliprotein type (Cr-PC 564)

3 Karoline Magalhães ^{a,1}

4 (ORCID- 0000-0001-5007-8912)

5 Adriana Lopes Santos ^b

6 (ORCID-0000-0002-0736-4937)

7 Daniel Vaulot ^{c, b}

8 (ORCID-0000-0002-0717-5685)

9 Mariana Cabral Oliveira ^a

10 (ORCID: 0000-0001-8495-2962)

11

¹² ^aLaboratório de Algas Marinhas Édson José de Paula, Departamento de Botânica,
¹³ Instituto de Biociências, Universidade de São Paulo, Rua do Matão, 277, 05508-090,
¹⁴ Butantã, São Paulo, SP, Brazil.

^b Asian School of the Environment, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798

¹⁷ ^cSorbonne Université, CNRS, UMR 7144, ECOMAP team Station Biologique de
¹⁸ Roscoff, Roscoff, France.

19

20 Submitted to Protist: revised version Tuesday, July 06, 2021

21 **Abstract:** Cryptophytes are a small group of photosynthetic biflagellated organisms
22 distributed worldwide in fresh, brackish and marine waters. Although members of this
23 class are easily distinguished from other groups, species identification is difficult and
24 studies concerning their diversity are scarce. Two strains of an undescribed *Hemiselmis*
25 species were isolated from the marine waters off Brazil and Japan. Analyses of
26 morphology, phycobiliprotein spectral characterization, molecular phylogeny and ITS2
27 secondary structure comparisons were performed to assist the identification. The
28 morphological features of *Hemiselmis aquamarina* sp. nov. matches that of other
29 species from the same genus, but it has a new type of phycocyanin. Molecular
30 phylogeny and ITS2 secondary structure support *H. aquamarina* as a distinct species.
31 Furthermore, phylogenetic inferences indicate *H. aquamarina* [as closely related to](#) *H.*
32 *tepida*, *H. andersenii* and *H. rufescens*. Currently, all *Hemiselmis* species have been
33 described from the North Hemisphere and most from the subtropical region. *H.*
34 *aquamarina* is the first species of this genus described from the South Atlantic.
35 **Key index words:** morphology; molecular phylogeny; phytoplankton; South Atlantic
36 Ocean; Pacific Ocean.

37 **Abbreviations:** BA, Bayesian analysis; bp, base pairs; [BS, bootstrap](#); CBC,
38 compensatory base change; Cr-PC, cryptophyte phycocyanin; Cr-PE, cryptophyte
39 phycoerythrin; [DHBV, 15,16-dihydrobiliverdin](#); ITS2, internally transcribed spacer 2;
40 h-CBC, hemi- compensatory base change; [ML, maximum likelihood](#); nSSU, nuclear
41 small ribosomal subunit; nmSSU, nucleomorph small ribosomal subunit; PBPs,
42 phycobiliproteins; [PCB, phycocyanobilin](#); PP, posterior probability; RCC, Roscoff
43 Culture Collection; SPC, surface periplast component.

44 **Introduction**

45 Cryptophytes are unicellular organisms, mostly photosynthetic, widespread in the
46 pelagic zones of brackish, marine, and freshwater environments (Klaveness 1985).
47 Their cells possess a furrow gullet system surrounded by ejectosomes (extrusive
48 organelles) and with a pair of flagella inserted. This system has a strong impact on cell
49 morphology, resulting in an asymmetrical shape and a unique way of swimming, which
50 makes the group easily recognizable by light microscopy.

51 The plastid [of cryptophytes is](#) surrounded by four membranes, originating from
52 secondary endosymbiosis with a red alga ancestor and it still possesses a relict of the
53 endosymbiont nucleus, the nucleomorph (Douglas and Penny 1999, Douglas et al. 2001,
54 Keeling 2010).

55 The photosynthetic pigments of cryptophytes include chlorophylls *a*, *c*, carotenoids and
56 phycobiliproteins-PBPs (Cunningham et al. 2019, Spear-Bernstein and Miller 1989).

57 The PBPs of cryptophytes are a heterodimer composed of two subunits, α and β , with
58 four bilins linked at the positions α -Cys 18 [\(19\)](#), β -DiCys 50, 61, β -Cys 82, and β -Cys
59 158 (Glazer and Wedemayer 1995, Wedemayer et al. 1996, Wemmer et al. 1993).

60 While four isomeric bilins are found in cyanobacteria and red algae PBPs, cryptophytes
61 PBPs display six bilins: [15,16-dihydrobiliverdin \(DHBV\)](#), [phycocyanobilin \(PCB\)](#),
62 [phycoerythrobilin, mesobiliverdin, and the two acryloyl bilins: bilin 584 and bilin 618](#).
63 [The last two bilins](#) are [only](#) known within this group (Glazer and Wedemayer 1995,
64 Wedemayer et al. 1991).

65 A given strain of cryptophyte possesses a single type of spectroscopically distinct PBP,
66 cryptophyte phycocyanin (Cr-PC), or phycoerythrin (Cr-PE), and its classification is
67 determined by its bluish or reddish appearance, respectively (Overkamp et al. 2014,

68 Wedemayer et al. 1996, Wemmer et al. 1993). Moreover, all Cr-PE investigated up to
69 date possess a single absorption peak (Hoef-Emden 2008). Currently, eight different
70 types of PBPs have been recognized according to their maximum visible absorption
71 spectrum: Cr-PC 569, Cr-PC 577, Cr-PC 615, Cr-PC 630, Cr-PC645, Cr-PE 545, Cr-PE
72 555 and Cr-PE566 (Cunningham et al. 2019, Greenwold et al. 2019).

73 *Hemiselmis* was first described in 1949 from the Isle of Man, UK, by the typification of
74 *Hemiselmis rufescens* Parke (1949), which has a reddish plastid. The next species
75 described was *Hemiselmis virescens* Droop (1955), a blue-green member, from
76 Cumbrae, Scotland. In 1967, when Butcher revised the classification of cryptophytes, he
77 created the family Hemiselmidaceae (Butcher 1967) based on the position of the gullet
78 across the short axis of the cell. Moreover, he described two subgenera, *Hemiselmis* and
79 *Plagiomonas*, distinguished by their reddish and bluish color, respectively. Assisted
80 mostly by light microscopy investigations, he assigned eleven species to *Hemiselmis*,
81 including seven new descriptions and two freshwater species with problematic
82 classification history (*Nephroselmis olivacea* sensu Pascher and *Sennia parvula* Skuja).
83 Unfortunately, he attributed many type localities for the majority of his descriptions,
84 except for *H. amyliifera*, *H. oculata*, and *H. rotunda*. Therefore, due to the uncertainty
85 concerning the morphological traits elected by Butcher, the inaccuracy in localities
86 typification, the absence of cultures from which the types were obtained and,
87 consequently, the inability to validate them, his descriptions of *Hemiselmis* species have
88 been treated as illegitimate (Lane and Archibald 2008).

89 In the last decades, five new species of *Hemiselmis* have been described with the help of
90 the rRNA operon sequences. *Hemiselmis amylosa* Clay & Kugrens was the first
91 freshwater species described for the genus, from Colorado Lake, USA (Clay and

92 Kugrens 1999). *Hemiselmis andersenii* Lane & Archibald, *Hemiselmis*
93 *cryptochromatica* Lane & Archibald, *Hemiselmis pacifica* Lane & Archibald and
94 *Hemiselmis tepida* Lane & Archibald are from marine environments. Except for *H.*
95 *pacifica*, which is from the North Pacific Ocean, all others were recorded from the
96 North Atlantic Ocean (Lane and Archibald 2008).

97 We investigated in detail two *Hemiselmis* strains, BMAK265 and RCC4102, collected
98 off the coasts of Brazil and Japan, respectively. Different microscopy techniques, PBP
99 visible absorption spectra, and sequences of the rRNA operon were used to assist in
100 species identification. Furthermore, we sequenced seven other strains of *Hemiselmis*
101 and *Chroomonas* available in the Roscoff Culture Collection (RCC). The two strains
102 cited above correspond to a yet undescribed species, designed herein as *Hemiselmis*
103 *aquamarina*, with a unique type of PBP, named Cr-PC 564.

104 **Taxonomy section**

105 *Hemiselmis aquamarina* K. Magalhães & M. C. Oliveira, sp. nov.

106 *Description:* free-swimming cells, reniform in lateral view, with rounded ends. Length
107 from 4.5 to 7.5 µm, width from 2.5 to 4.5 µm. Cells ovate in ventral/dorsal and circular
108 apical/antapical view. Two sub-equal flagella. Superficial periplast component with
109 hexagonal plates. Single dorsal plastid, parietal, blue-green. Single subapical pyrenoid,
110 starch coated, with single thylakoids penetrating the core. Conspicuous refractive body.
111 Accessory pigment Cr-PC 564. The SEM stub is available at the Herbarium of the
112 Botanical Institute of São Paulo (SP), voucher #SP469.780.

113 *Holotype* (here designated): Frozen pellet of strain BMAK265 (in a metabolically
114 inactive state) has been deposited as type at the Roscoff Culture Collection under DNA
115 record # 2161.

116 *Molecular diagnosis:* nSSU (MT605165, MT605166), ITS2 (MT628030- MT628033)
117 and nmSSU (MT605187- MT605190) rRNA.

118 *Type strain:* BMAK265 (synonymous RCC5634)

119 *Other strain:* RCC4102

120 *Type locality:* 23.59745 S, 45.02833W, coastal area of Ubatuba, São Paulo, Brazil.

121 *Etymology:* the epithet refers to the color of the cells in culture; light blue-green,
122 aquamarine.

123 **Results**

124 **Morphological characterization**

125 The cells of *Hemiselmis aquamarina* are asymmetrical and reniform (bean-shaped) in
126 lateral view, sizing from 4.5 to 7.5 μm in length. The form of the cells is variable in the
127 culture, from elliptical to rounded (Figure 1). One conspicuous refractive body, also
128 called Corps de Maupas, is located above the nucleus, near the cell center (Figure 1C-
129 D).

130 One parietal lobed plastid, boat-shaped and light blue-green ([Supplementary Figure](#)
131 [2A](#)), occupies the dorsal part of the cell extending towards the lateral sides (Figure 1).
132 [Dense cultures are hunter-green \(Supplementary Figure 2B\)](#). Four membranes enclose
133 the plastid and a prominent subapical starch-coated pyrenoid can be seen (Figure 1A-B,
134 [Supplementary Figure 3](#)). The thylakoids are densely packed within the plastid with
135 little free stroma; sometimes a parallel arrangement pattern is observed (Supplementary
136 Figure 3A). Single thylakoids penetrate the pyrenoid core. No stigma was detected.

137 The nucleomorph is situated in the ventral part of the plastid, directly below the
138 subapical pyrenoid, almost in the center of the cell. A double-membrane envelope
139 surrounds the nucleomorph, which has a granular matrix ([Supplementary Figure 3A-B](#)).

140 The main nucleus of the cell occupies the antapical pole and its last membrane is
141 continuous with the plastid complex ([Supplementary Figure 3A, C, E](#)). A small Golgi
142 body, with many vesicles, is located behind the flagellar region ([Supplementary Figure](#)
143 [3B-E](#)). A tubular mitochondrion extends mainly through the longitudinal direction of
144 the cell, near the flagellar apparatus, nucleus and between the periplast and the plastid.

145 Two unequal flagella are inserted ventrally, almost in the center of the *H. aquamarina*
146 cell (Figures 1-2). One flagellum is slightly larger than the other (Figure 1F), and
147 mastigonemes were found on both (Figure 2B, D-E). Close to the flagellar insertion, a
148 small gullet extends obliquely from the vestibulum towards the antapical pole of the
149 cell, and it is surrounded by large ejectosomes (Figures 1C-D, [Supplementary Figure](#)
150 [3B-E](#)). A furrow [is absent and a](#) contractile vacuole [was not observed](#). Cells are free-
151 swimming and very active. Frequently, when resting, they quickly start to rotate on their
152 axis and suddenly go away (see Supplementary Data 1 and 2).

153 The surface periplast component (SPC) of *H. aquamarina* consists of large hexagonal
154 plates and seems quite granular (Figure 2). In *Hemiselmis*, the periplast of the cells is
155 more delicate than in other cryptophytes and cells can collapse after critical point drying
156 (Figure 2B). Gaps in the periplast occur, which are occupied by ejectosomes (Figure
157 2D-E). We observed a mid-basal line at the antapical pole of the cell (Figure 2F).

158 **Phycobiliprotein spectral characterization**

159 The PBP extracts of both *H. aquamarina* strains have very close spectral characteristics.
160 Both pigment extracts are light purple ([Supplementary Figure 2C](#)). Small variations of
161 spectral signatures between the two strains are observed. The curve of BMAK265
162 shows the maximum absorption at 564 nm and a second peak at 616-620 nm. RCC4102
163 has the highest peak of absorption at 557-566 nm, and another peak at 616-619 nm
164 (Figure 3).

165 **Phylogenetic analysis**

166 Molecular phylogeny inferences performed using Bayesian (BA) [and maximum](#)
167 [likelihood \(ML\) analysis](#) indicate with maximum posterior probability (PP) [and](#)
168 [bootstrap \(BS\) support](#) that the genus *Hemiselmis* is a distinct and monophyletic

169 lineage, sister of the *Komma* / “*Chroomonas*” clade. Within Hemiselmidaceae, *H.*
170 *cryptochromatica* is the early diverting branch, fully supported in concatenated
171 (nucleomorph and nuclear SSU) and nuclear small ribosomal subunit (nSSU)
172 phylogenies. *H. amylosa* follows *H. cryptochromatica* as the sister clade of the
173 remaining species. Furthermore, all analyses support a sister relationship between *H.*
174 *pacifica* and *H. virescens* (Figure 4, Supplementary Figure 1).

175 The close relationship between *H. aquamarina*, *H. tepida*, *H. rufescens*, and *H.*
176 *andersenii* is recovered in concatenated and nucleomorph small ribosomal subunit
177 (nmSSU) phylogenies, although with unreliable support (below 0.75 PP and 70 BS). *H.*
178 *rufescens* and *H. andersenii* are indicated as sister species, supported by 0.95 PP in BA,
179 but not by ML (below 70 BS) in the concatenated SSU rRNA inferences. This
180 relationship is not recovered from the single genes phylogenies, which show unresolved
181 relationships between these species.

182 *H. tepida* and *H. aquamarina* are pointed out as sister taxa in concatenated (0.86 PP)
183 and nmSSU (0.94 PP, 70 BS) rRNA inferences. However, nSSU phylogeny suggests a
184 close relationship between *H. andersenii* and *H. aquamarina*, although the support is
185 too low to make any conclusion (below 0.75 PP and 70 BS). The *H. aquamarina* clade
186 is fully supported by PP in all BA performed and with high BS in ML inferences
187 (Figure 4, Supplementary Figure 1). In all trees, the UTEX 2000 strain clusters into the
188 *H. aquamarina* clade and, therefore, it is classified as *H. cf. aquamarina*.

189 Secondary structures of nuclear ITS2

190 The lengths of the ITS2 region are quite similar for all *Hemiselmis* strains and clones
191 analyzed, with a mean length of 335 nt (SD ± 9 nt). The shortest ITS2 sequence is found
192 in one clone of *H. aquamarina* (BMAK265, 327 nt), and the longest one in *H. cf.*

193 *virescens* (RCC3575, 360 nt). Clones of the same strain display different lengths of the
194 ITS2 sequence.

195 The predicted ITS2 secondary structures of *Hemiselmis* have four helices as reported for
196 most eukaryotes. ITS2 displays highly conserved single-stranded sequences between
197 5.8S and helix I, helices I-II and helices II-III (see Supplementary Figure 4). However,
198 other regions could not be well aligned interspecifically due to the high divergence of
199 nucleotides. These regions are located at helices I (middle and apex portion), II (apex
200 portion), III (middle portion) and IV.

201 Helix I of *H. aquamarina* and *H. cf. virescens* is branched in all acceptable predictions
202 generated. The apex portion of helix I have variations of nucleotides between clones and
203 strains of *H. aquamarina*. Helix II is conserved at the first eight nucleotides and has a
204 uracil- uracil mismatch at the sixth position in all *Hemiselmis* species (Figure 5). The
205 apex of helix II show deletions of three base pairs (pb) between clones of *H.*
206 *aquamarina* in both strains (Figure 5, positions 15-17). Helix III, the longest one, shows
207 high conserved nucleotides between positions 31-48 (Figure 6, Supplementary Figure
208 4). Helix IV is the most variable one and could not be accurately aligned
209 interspecifically. The single-stranded sequences between helices III-IV and helix IV-
210 LSU are quite dissimilar. Therefore, the consensus motif indicating the termini of helix
211 IV could not be predicted. For more details, see Supplementary Data 4.

212 Structural comparison of ITS2 helices at conserved base pairs (bp) between *Hemiselmis*
213 species show many compensatory base changes (CBCs) and hemi-CBCs (h-CBCs) at
214 the most conserved helices (Figures 5-6). *H. aquamarina* has a unique molecular
215 signature of ITS2 secondary structure between species of the genus. Its strains and

- 216 clones have no CBC across helices I, II and III. Therefore, *H. aquamarina* can also be
217 distinguished from the other *Hemiselmis* species by CBCs.

218 **Discussion**

219 Many features observed by light and electron microscopy unveiled synapomorphies in
220 Hemiselmidaceae, which are congruent with molecular phylogeny results, indicating
221 that this family is indeed a natural grouping from evolutionary processes. All currently
222 described *Hemiselmis* species have a lateral insertion of the flagella and hexagonal
223 plates of SPC (Butcher 1967, Clay and Kugrens 1999, Lane and Archibald 2008,
224 Wetherbee et al. 1986). Before this work, four species of *Hemiselmis* (*H. amylosa*, *H.*
225 *rufescens/brunnescens*, *H. simplex*, and *H. virescens*) have been investigated by
226 transmission electron microscopy (Butcher 1967, Clay and Kugrens 1999, Lucas 1970,
227 Santore and Greenwood 1977, Santore 1982). The arrangement and form of the
228 organelles, such as the starch-coated pyrenoid, the nucleomorph, the Golgi body, the
229 nucleus and the mitochondrion, are congruent between *Hemiselmis* species. Santore
230 (1982) had difficulty separating *H. rufescens/H. brunnescens* and *H. virescens* using
231 ultrastructural data. Conclusively, *Hemiselmis* is easily distinguishable by morphology
232 from other genera of cryptophytes, but species identification is not clear (see
233 Supplementary Table 2). Accordingly, *Hemiselmis aquamarina* identification must rely
234 on molecular tools as reported for other species of cryptophytes (Lane and Archibald
235 2008, Hoef-Emden 2007, 2018).

236 The PBP type has been correlated with phylogenetic analyses (Deane et al. 2002, Hoef-
237 Emden 2008, Marin et al. 1998). While Cr-PE 545 has been suggested as a
238 plesiomorphic state (Cunningham et al. 2019) and it is found in many genera, Cr-PCs
239 are more diversified and some types are restricted to last divergent lineages.

240 Cr-PC 564 of *H. aquamarina* seems to close to Cr-PC 569, Cr-PE 566 and Cr-PC 615.
241 Cr-PC 569, Cr-PE 566 and Cr-PC564 maximum absorption peaks are notable closes
242 (569, 566 and 564, respectively). Cr-PE 566 has a single maximum absorption peak.
243 The second peak of Cr-PC 569 and Cr-PC 564 do not overlap. The second peak of Cr-
244 PC 564 is close to the major absorption peak of Cr-PC 615, but the main peak of Cr-PC
245 564 is not related to the second one of Cr-PC 615.

246 Therefore, none of the previously described PBPs in cryptophytes has the same spectral
247 signature of Cr-PC 564 from *H. aquamarina*, which has features common to Cr-PCs
248 and Cr-PEs. These differences of absorption peaks in Cr-PC 564 are indicative of
249 different bilin composition and/ or linkage sites. Accordingly, its bilin composition must
250 differ from other Cr-PCs at the linkage sites β -Cys 158 and -Cys 18, as the sites β -
251 DiCys 50, 61 and β -Cys 82 are frequently linked to DHBV and PCB bilins, respectively
252 (Wedemayer et al. 1996, Overkamp et al. 2014). Possibly, the β -Cys 158 position could
253 be linked to bilin 584, as reported in Cr-PC 569 and Cr-PE 566, and α -Cys 18 position
254 to PCB or bilin 618, as in Cr-PC 615 and Cr-PC 569, respectively. However, bilin
255 composition alone does not determine the absorption spectra of a given PBP: the native
256 protein environment of the bilins also contributes significantly to these properties

257 (Glazer and Wedemayer 1995, Wemmer et al. 1993).

258 *H. aquamarina* is closely related to *H. tepida*, *H. andersenii* and *H. rufescens*, as
259 previously indicated in literature by the placement of UTEX2000 in phylogenetic trees
260 (Cunningham et al. 2019, Hoef-Emden 2008, 2018). The accurate relationship between
261 these species is not well solved by phylogenetic inferences due to the medium/ low
262 supports of branches in all methods applied. This could be attributed to the low
263 evolutionary rate of SSU rRNA. However, it seems that *H. tepida* and *H. andersenii* are

264 closer to *H. aquamarina* than *H. rufescens*. All phylogenetic inferences indicate with
265 high statistical support that *Hemiselmis aquamarina* forms a new branch and can be
266 regarded as a new species. For species identification propose, nmSSU and ITS2 rRNA
267 sequences are more divergent between *Hemiselmis* and, accordingly, more accurate for
268 diagnosis.

269 The predicted ITS2 secondary structures of *Hemiselmis* are similar to the others
270 previously published for the group (see Hoef-Emden 2007, 2018, Majaneva et al 2014).
271 Intraspecific variations of ITS2 length have been found in species of *Chroomonas*
272 (Hoef-Emden 2018) as pointed out for *H. aquamarina*. Homopolymeric stretch of uracil
273 in helix II observed in *H. aquamarina* was also reported in *Chroomonas nordstedtii*
274 Hansgirg (Hoef-Emden 2018). Substitutions and indels can occur in the apex of the
275 helices with high frequency (Coleman 2000). Accordingly, the differences found within
276 strains and clones of *H. aquamarina* are likely the result of intragenomic copy variation
277 due to the several numbers of copies of rRNA operon in the genome (Coleman 2007).

278 CBCs in the ITS2 secondary structure are indicative of species separation due to the
279 conserved pairing structure needed for rRNA processing (Coleman 2007, 2009, Muller
280 2007). For example, in Volvocales, the absence of CBCs in ITS2 helices II and III is in
281 agreement with sexual compatibility and, therefore, can predict mating affinity (i.e.
282 biological species concept, Coleman 2000). Moreover, most of CBCs in these regions
283 are non-homoplasious changes and can present molecular signatures, which detect
284 unambiguously taxa and clades (Caisová et al. 2011). Therefore, the presence of CBCs
285 in these conserved regions between *Hemiselmis aquamarina* and other species of the
286 genus indicates it as distinct species.

287 **Conclusion**

288 The evidence presented here suggests the existence of a new species, named *Hemiselmis*
289 *aquamarina*, containing a new type of PBP (Cr-PC 564) and represented by two strains
290 from Brazil (BMAK265) and Japan (RCC4102). Both strains are cultured at 20°C,
291 suggesting that this species prefers relatively warm water conditions. This corresponds
292 to the prevailing conditions at the locations where the strains were collected which are
293 subjected to dominant currents coming from tropical regions (Brazil and Kuroshio
294 currents, respectively).

295 **Material and Methods**

296 **Sampling, strain isolation and culturing conditions**

297 The strain from the coast of Brazil was isolated in August 2011, close to the Anchieta
298 Island, Ubatuba, São Paulo ($23^{\circ} 35.85'S$, $45^{\circ} 01.70'W$). Water was collected at a depth
299 of 40 m with a Nansen bottle. The water sample was enriched with Erd-Schreiber
300 medium diluted 10 times. After a few days, a single cell was selected by micro-pipetting
301 and carefully washed in a sterilized medium. The culture is maintained in Erd-Schreiber
302 medium (Throndsen 1997), 32-35 salinity, $20^{\circ}C$ temperature, a photoperiod of 12:12
303 L:D cycle at $50 \mu\text{mol m}^{-2}\text{s}^{-1}$. This strain is deposited in the Banco de Microrganismos
304 Marinhos Aydar & Kutner ([http://www.io.usp.br/index.php/infraestrutura/banco-de-](http://www.io.usp.br/index.php/infraestrutura/banco-de-microorganismos.html)
305 [microorganismos.html](http://www.io.usp.br/index.php/infraestrutura/banco-de-microorganismos.html)), as BMAK265, and in the RCC as RCC5634.

306 The strain from Japan was isolated in August of 2013, during the Oshoro-Maru Cruise.
307 Four liters of surface water were collected from station S3 near Kuroasaki, Iwate ($39^{\circ} 59'N$,
308 $142^{\circ} 15'E$) and concentrated by tangential flow filtration to 100 mL. Single cells
309 were isolated in K medium at $20^{\circ}C$ by micro-pipetting and then maintained under these
310 conditions. This strain is deposited in the Roscoff Culture Collection as RCC4102.

311 Although we wished to perform similar analyses on strain UTEX 2000, which SSU
312 rRNA sequences are very close to those of BMAK265 and RCC4102, we could not
313 obtain this strain from The Culture Collection of Algae at the University of Texas at
314 Austin (UTEX) where it is cryopreserved only and not distributed
315 (<https://utex.org/products/utex-lb-2000>).

316 **Phycobiliprotein extraction**

317 Cultures were grown in 50 ml polycarbonate culture flasks for 2-3 weeks, following the
318 conditions described above, and harvested by centrifugation (300 g, 8 min). The pellet
319 was frozen and kept at -80°C until processing. PBP extractions were performed for
320 strains BMAK265, RCC4102, RCC4116, RCC4216 and RCC659 (Supplementary
321 Table 1) following Hill and Rowan (1989). The absorption spectra of the pigment
322 extract were determined using an Epoch 2 microplate spectrophotometer (BioTek
323 Instruments, Inc., Winooski, VT, USA).

324 **Morphological observations**

325 Dense BMAK265 (2 – 3 weeks old) cultures were used for morphological
326 investigations. For differential interference contrast (DIC) and phase contrast, living
327 cells on glass slides sealed with coverslips were observed with a Leica DM 4000 B
328 (Leica Microsystems, Wentzler, Germany). Morphometric values from 40 live
329 individuals were obtained from calibrated pictures. Natural fluorescence of the plastid
330 was observed from fixed (2% glutaraldehyde) cells with a confocal microscope, Zeiss
331 LSM 440 Axiovent 100 (Carl Zeiss, Jena, Germany) equipped with 543 nm laser and a
332 570 nm long-pass filter.

333 For electron microscopy observations, cells were harvested by gentle centrifugation (3
334 min, 100-150 g) and then immediately fixed for 90 min with a solution containing
335 glutaraldehyde (2%), sodium cacodylate trihydrate (0.1 M) and sucrose (0.8M), as
336 described in Majaneva et al. (2014). The cells were washed using the latter solution
337 (without glutaraldehyde) and post-fixed with osmium tetroxide (1 %) buffered in
338 cacodylate trihydrate (0.1 M) for 60 min. Cells were then washed twice in cacodylate
339 buffer (0.1M). For scanning Electron microscopy (SEM), a sample was dehydrated in a

340 series of increasing ethanol concentrations (70, 90, 95 and 100 %). It was subsequently
341 critical-point dried (Balzers CPD 030, Bal-Tec, Vaduz, Liechtenstein), gold-coated
342 (Balzers SCD 050) and examined in a Zeiss Sigma VP. For Transmission Electron
343 microscopy (TEM), cells were dehydrated in an acetone series (50, 70, 90, 95 and
344 100%), embedded in Spurr resin, thin sectioned and examined in a Philips CM120
345 TEM.

346 **DNA extraction, PCR, cloning and sequencing**

347 Material for molecular analyses was obtained as specified in the pigment extraction
348 section. Genomic DNA was extracted using *NucleoSpin® Plant* II kit following the
349 manufacturer's instructions. We performed PCRs of nSSU, nmSSU and ITS2 rRNA
350 with Platinum® *Taq* DNA polymerase kit (Invitrogen™, Carlsbad, USA). Primers and
351 cycling conditions are available in Supplementary Table 3. We analyzed the strains
352 BMAK265 (RCC5634), RCC659, RCC1504, RCC2614, RCC3436, RCC3575,
353 RCC4102, RCC4116 and RCC4216 (Supplementary Table 1). Since there might be
354 multiple copies of the rRNA operon in the genome (Prokopowich et al. 2003, Thornhill
355 et al. 2007), we cloned all sequences of BMAK265 and RCC4102 and the ones that
356 intragenomic variation was observed using TOPO® TA Cloning® kit (Invitrogen,
357 Carlsbad, CA, USA). To avoid polymerase errors in cloning, we used Phusion High-
358 Fidelity PCR Master Mix (Thermo Fisher Scientific).

359 The PCRs products yielding a single band of the expected size on an agarose gel (1%)
360 were purified using the GFX Illustra kit (GE Healthcare Life Sciences, Little Chalfont,
361 Buckinghamshire, UK) following the manufacturer's instructions. Sequencing reactions
362 were performed with the Terminator Cycle Sequencing Ready Reaction kit (Applied

363 Biosystems™, Hammonton, NJ, USA) and samples were sequenced with 3730 Applied
364 Biosystems.

365 Generated contigs were searched by BLAST (<https://blast.ncbi.nlm.nih.gov/>) against
366 NCBI sequences to ensure that each contig corresponded to the organisms we were
367 investigating. We discarded contigs with high levels of ambiguity and noise. The
368 consensus sequences were assembled using *Geneious* 9.7 (Biomatters, Auckland, New
369 Zealand, <https://www.geneious.com/>) by comparison to a reference sequence obtained
370 from BLAST. All sequences generated were deposited to Genbank (see Supplementary
371 Table 1 for accession numbers).

372 **Phylogenetic analyses**

373 We built the datasets for phylogenetic analyses using the sequences obtained and from
374 the NCBI database. Alignments were performed in AliView (Larsson 2014) with the
375 Muscle algorithm (Edgar 2004) and refined by visual inspection. The appropriate
376 evolution model was chosen using JModelTest 2.1.7 (Darriba et al. 2012). Aligned
377 sequences datasets were subjected to likelihood mapping tests with varying degrees of
378 indel regions in Tree-Puzzle 5.3 (Schmidt et al. 2003), to determine whether the
379 phylogenetic signal was increased with or without missing data. For these analyses, we
380 used the specific molecular evolution model for nucleotide substitution recommended
381 for each rRNA region. For the following analyses, we used the alignments with best-
382 solved quartets.

383 We concatenated nSSU (1495 pb, 56 sequences) and nmSSU (1347 pb, 51 sequences)
384 sequences using SeaView (Gouy et al. 2010), resulting in a dataset of 56 sequences with
385 2,842 bp (Supplementary Data 3). We applied the molecular evolution modelss
386 HKY+G+I and GTR+G+I for the nSSU and nmSSU, respectively, in a partitioned

387 [Bayesian analysis](#) in the concatenated alignment. Moreover, to determine if the
388 topologies were congruent between nSSU and nmSSU, we performed separated
389 phylogenetic inferences by [Bayesian analysis](#) as described below (Supplementary
390 Figure 1).

391 [Bayesian analyses](#) were performed with MrBayes 3.2 (Ronquist et al. 2012) with two
392 consecutive runs of 1×10^7 generations, four Markov chains, and a sampling frequency
393 of 100 generations. Runs convergence and likelihood were checked in Tracer V1.6
394 (Rambaut et al. 2014). The split frequency of the runs was below the guidance
395 recommendation. We applied a relative burn-in of 25%.

396 [Maximum likelihood trees were inferred using raxmlGUI 2.0 \(Edler et al. 2021\) by ML](#)
397 [+ rapid bootstrap analysis and two threads for comparisons between topologies and](#)
398 [nodes support with Bayesian analysis. 1000 replicates were applied for bootstrap](#)
399 [estimation. The GTR+CAT evolutionary model was applied for single region inferences](#)
400 [and in the concatenated alignment in a partitioned analysis.](#)

401 We decided to include strains that had just one marker in the concatenated dataset, such
402 as *Hemiselmis amylosa*. Incomplete taxa can be accurately placed in phylogenies and
403 improve results in cases of misleading long branches (Wiens 2006). The sequences of
404 *Teleaulax*, *Plagioselmis*, *Hamisia*, *Guillardia*, *Proteomonas*, *Rhodomonas*, *Rhinomonas*
405 and *Storeatula* were used to root the tree due to their distant phylogenetic relationship to
406 the *Hemiselmis* clade (Deane et al. 2002, Hoef-Emden et al. 2002, Hoef-Emden 2008,
407 2018).

408 Secondary structure prediction of the nuclear ITS2 region

409 All strains sequenced for nuclear ITS2 rRNA region, except RCC659 and RCC2614,
410 were cloned due to intragenomic variation, which was observed in a first sequencing.

411 Fourteen secondary structures of ITS2 rRNA were predicted from *H. aquamarina*
412 (BMAK265 and RCC4102, seven clones), *H. cf. andersenii* (RCC2614), *H. rufescens*
413 (RCC659), *H. cf. virescens* (RCC3575, one clone) and *Chroomonas cf. debatzensis*

414 Hoef-Emden (RCC1504 and RCC3436, four clones). The complementary regions of
415 5.8S and LSU rRNA, ITS2 boundaries, were annotated using Hidden Markov Models
416 with the default parameters of the ITS2 database annotation tool

417 (<http://its2.bioapps.biozentrum.uni-wuerzburg.de/>, Ankenbrand et al. 2015). Automated
418 secondary structures of ITS2 rRNA predictions of the entire sequence and, when
419 necessary, of single helices, were acquired by online web services using the default
420 folding options of the Mfold (Zuker, 2003) and RNAsstructure (Reuter and Mathews,
421 2010). Several alternative secondary structures were predicted per sequence. Structure
422 choices were based on the conserved hallmarks of ITS2 secondary structures (see
423 Coleman 2000, 2007, 2009) and comparisons with previously published structures of
424 cryptophytes. The ITS2 sequences were aligned with MAFFT and G-INS-I algorithm
425 (Katoh et al. 2002). For each sequence in the alignment, a preliminary secondary
426 structure was annotated in Vienna file format, imported to 4SALE (Seibel et al. 2008)
427 and manually edited by a comparative analysis of each nucleotide position
428 (Supplementary Data 4). The consensus secondary structure of *H. aquamarina* was
429 generated in 4SALE using the default option (Supplementary Figure 4). Due to the
430 variability of the paring positions in helices I and IV of ITS2 between Hemiselmis
431 sequences, we decided to adopt the CBC clade-based concept sensu Coleman (2000).

432 Accordingly, unambiguously aligned positions of helices II and III were used for
433 numbering the positions common between species

434 **Data availability**

435 Alignments and ITS2 secondary structure data are available at
436 <https://figshare.com/s/be3127e6bc7edfb267ce>. Strains are available from the Roscoff
437 Culture Collection (<http://www.roscoff-culture-collection.org/>).

438 **Acknowledgments**

439 We wish to thank Willian da Silva Oliveira, André Nakasato and Rosario Petti from
440 Marine Algae Laboratory of IB USP, Waldir Caldeira and Sheila Schuindt, from the
441 Electron Microscopy Laboratory of IB USP, for the substantial support with confocal
442 electron microscopies, Michael Guiry, for the revision of the taxonomic section and
443 Priscillia Gourvil for isolation of the RCC4102 strain, and Catherine Gérikas Ribeiro for
444 help with DNA work in the laboratory.

445 **Funding**

446 Financial support was obtained from the CNPq to KM (163070/2013-0) and MCO
447 (305687/2018-2; 406351/2016-3), from the FAPESP research project within the Biota-
448 Program (2013/11833-3) and the CNRS GDRI (International Scientific Network)
449 "Diversity, Evolution and Biotechnology of Marine Algae"

450 **CRediT author statement**

451 KM: conceptualization, methodology, validation, formal analysis, investigation, data
452 curation, writing (original draft). ALS: methodology, data curation and validation,
453 writing (review and editing). DV: validation, resources, writing (review and editing)

454 and funding acquisition. MCO: resources, writing (review and editing), supervision,
455 project administration and funding acquisition.

456 Declaration of Competing Interest

457 The authors declare that they have no conflict of interest.

458 **References**

- 459 **Ankenbrand MJ, Keller A, Wolf M, Schultz J, Förster F** (2015) ITS2 database V:
460 Twice as much. Mol Biol Evol **32**:3030–3032.
- 461 **Butcher RW** (1967) An introductory account of the smaller algae of British coastal
462 waters: Cryptophyceae. London.
- 463 **Caisová L, Marin B, Melkonian M** (2011) A close-up view on ITS2 evolution and
464 speciation - a case study in the Ulvophyceae (Chlorophyta, Viridiplantae). BMC Evol
465 Biol **11**:262.
- 466 **Clay BL, Kugrens P** (1999) Characterization of *Hemiselmis amylosa* sp. nov. and
467 phylogenetic placement of the blue-green cryptomonads *H. amylosa* and *Falcomonas*
468 *daucooides*. Protist **150**:297–310.
- 469 **Coleman AW** (2000) The significance of a coincidence between evolutionary
470 landmarks found in mating affinity and a DNA sequence. Protist **151**:1–9.
- 471 **Coleman AW** (2007) Pan-eukaryote ITS2 homologies revealed by RNA secondary
472 structure. Nucleic Acids Res **35**:3322–3329.
- 473 **Coleman AW** (2009) Is there a molecular key to the level of “biological species” in
474 eukaryotes? A DNA guide. Mol Phylogenet Evol **50**:197–203.
- 475 **Cunningham BR, Greenwold MJ, Lachenmyer EM, Heidenreich KM, Davis AC,**
476 **Dudycha JL, Richardson TL** (2019) Light capture and pigment diversity in marine
477 and freshwater cryptophytes Collier, J (ed). J Phycol **55**:552–564.
- 478 **Darriba D, Taboada GL, Doallo R, Posada D** (2012) jModelTest 2: more models,
479 new heuristics and parallel computing. Nat Methods **9**:772–772.

- 480 **Deane JA, Strachan IM, Saunders GW, Hill DRA, McFadden GI** (2002)
- 481 Cryptomonad evolution: nuclear 18S rDNA phylogeny versus cell morphology and
- 482 pigmentation. *J Phycol* **38**:1236–1244.
- 483 **Douglas SE, Penny SL** (1999) The plastid genome of the cryptophyte alga, *Guillardia*
- 484 *theta*: complete sequence and conserved synteny groups confirm its common ancestry
- 485 with red algae. *J Mol Evol* **48**:236–244.
- 486 **Douglas S, Zauner S, Fraunholz M, Beaton M, Penny S, Deng LT, Wu X, Reith M,**
- 487 **Cavalier-Smith T, Maier UG** (2001) The highly reduced genome of an enslaved algal
- 488 nucleus. *Nature* **410**:1091–6.
- 489 **Droop MR** (1955) Some new supra-littoral Protista. *J Mar Biol Ass UK* **34**:233–245.
- 490 **Edgar RC** (2004) MUSCLE: multiple sequence alignment with high accuracy and high
- 491 throughput. *Nucleic Acids Res* **32**:1792–7.
- 492 [**Edler D, Klein J, Antonelli A, Silvestro D** \(2021\) raxmlGUI 2.0: A graphical interface](#)
- 493 [and toolkit for phylogenetic analyses using RAxML](#) Matschiner, M (ed). *Methods Ecol*
- 494 [Evol](#) **12**:373–377.
- 495 **Glazer AN, Wedemayer GJ** (1995) Cryptomonad biliproteins- an evolutionary
- 496 perspective. *Photosynth Res* **46**:93–105.
- 497 **Gouy M, Guindon S, Gascuel O** (2010) SeaView version 4: A multiplatform graphical
- 498 user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol*
- 499 **27**:221–224.
- 500 [**Greenwald MJ, Cunningham BR, Lachenmyer EM, Pullman JM, Richardson TL,**](#)
- 501 [**Dudycha JL** \(2019\) Diversification of light capture ability was accompanied by the](#)

- 502 [evolution of phycobiliproteins in cryptophyte algae. Proc R Soc B Biol Sci](#)
503 [286:20190655.](#)
- 504 **Hill DRA, Rowan KS** (1989) The biliproteins of the Cryptophyceae. *Phycologia*
505 **28**:455–463.
- 506 [Hoef-Emden K \(2007\) Revision of the genus *Cryptomonas* \(Cryptophyceae\) II:](#)
507 [Incongruences between the classical morphospecies concept and molecular phylogeny](#)
508 [in smaller pyrenoid-less cells. Phycologia 46:402–428.](#)
- 509 **Hoef-Emden K** (2008) Molecular phylogeny of phycocyanin-containing cryptophytes:
510 evolution of biliproteins and geographical distribution. *J Phycol* **44**:985–993.
- 511 **Hoef-Emden K** (2018) Revision of the Genus *Chroomonas* HANSGIRG: the benefits
512 of DNA-containing specimens. *Protist* **169**:662–681.
- 513 **Hoef-Emden K, Marin B, Melkonian M** (2002) Nuclear and nucleomorph SSU rDNA
514 phylogeny in the Cryptophyta and the evolution of cryptophyte diversity. *J Mol Evol*
515 **55**:161–79.
- 516 **Katoh K, Misawa K, Kuma KI, Miyata T** (2002) MAFFT: A novel method for rapid
517 multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*
518 **30**:3059–3066.
- 519 **Keeling PJ** (2010) The endosymbiotic origin, diversification and fate of plastids. *Philos*
520 *Trans R Soc Lond B Biol Sci* **365**:729–748.
- 521 **Klaveness D** (1985). Classical and modern criteria for determining species of
522 Cryptophyceae. *Bull Plankt Soc Japan* **32**:111–123.
- 523 **Lane CE, Archibald JM** (2008) New marine members of the genus *Hemiselmis*

- 524 (Cryptomonadales, Cryptophyceae). J Phycol **44**:439–450.
- 525 **Larsson A** (2014) AliView: a fast and lightweight alignment viewer and editor for large
526 datasets. Bioinformatics **30**:3276–3278.
- 527 **Lucas I** (1970) Observations on the ultrastructure of representatives of the genera
528 *Hemiselmis* and *Chroomonas* (Cryptophyceae). Br Phycol J **5**:29–37.
- 529 **Majaneva M, Remonen I, Rintala J-M, Belevich I, Kremp A, Setälä O, Jokitalo E,**
530 **Blomster J** (2014) *Rhinomonas nottbecki* n. sp. (Cryptomonadales) and molecular
531 phylogeny of the family Pyrenomonadaceae. J Eukaryot Microbiol **61**:480–492.
- 532 **Marin B, Klingberg M, Melkonian M** (1998) Phylogenetic relationships among the
533 Cryptophyta: analyses of nuclear-encoded SSU rRNA sequences support the
534 monophyly of extant plastid-containing lineages. Protist **149**:265–76.
- 535 **Müller T, Philippi N, Dandekar T, Schultz J, Wolf M** (2007) Distinguishing species.
536 RNA **13**:1469–1472.
- 537 **Overkamp KE, Langklotz S, Aras M, Helling S, Marcus K, Bandow JE, Hoef-**
538 **Emden K, Frankenberg-Dinkel N** (2014) Chromophore composition of the
539 phycobiliprotein Cr-PC577 from the cryptophyte *Hemiselmis pacifica*. Photosynth Res
540 **122**:293–304.
- 541 **Parke M** (1949) Studies on marine flagellates. J Mar Biol Assoc United Kingdom
542 **28**:255–286.
- 543 **Prokopowich CD, Gregory TR, Crease TJ** (2003) The correlation between rDNA
544 copy number and genome size in eukaryotes. Genome **46**:48–50.
- 545 **Rambaut A, Suchard M, Xie D, Drummond A** (2014) Tracer v1.6.

- 546 [**Reuter JS, Mathews DH**](#) (2010) RNAstructure: software for RNA secondary structure
547 prediction and analysis. *BMC Bioinformatics* **11**:129.
- 548 **Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget
549 B, Liu L, Suchard M a., Huelsenbeck JP** (2012) Mrbayes 3.2: Efficient bayesian
550 phylogenetic inference and model choice across a large model space. *Syst Biol* **61**:539–
551 542.
- 552 **Santore UJ** (1982) The ultrastructure of *Hemiselmis brunnescens* and *Hemiselmis
553 virescens* with additional observations on *Hemiselmis rufescens* and comments on the
554 Hemiselmidaceae as a natural group of the Cryptophyceae. *Br Phycol J* **17**:81–99.
- 555 **Santore UJ, Greenwood AD** (1977) The mitochondrial complex in cryptophyceae.
556 *Arch Microbiol* **112**:207–218.
- 557 **Schmidt HA, Petzold E, Vingron M, Von Haeseler A** (2003) Molecular
558 phylogenetics: parallelized parameter estimation and quartet puzzling. *J Parallel Distrib
559 Comput* **63**:719–727.
- 560 **Seibel PN, Müller T, Dandekar T, Wolf M** (2008) Synchronous visual analysis and
561 editing of RNA sequence and secondary structure alignments using 4SALE. *BMC Res
562 Notes* **1**:91.
- 563 **Spear-Bernstein L, Miller KR** (1989) Unique location of the phycobiliprotein light-
564 harvesting pigment in the Cryptophyceae. *J Phycol* **25**:412–419.
- 565 **Thornhill DJ, Lajeunesse TC, Santos SR** (2007) Measuring rDNA diversity in
566 eukaryotic microbial systems: how intragenomic variation, pseudogenes, and PCR
567 artifacts confound biodiversity estimates. *Mol Ecol* **16**:5326–5340.

- 568 **Thronsdon J** (1997) The Planktonic Marine Flagellates. In: Identifying Marine
569 Phytoplankton, Thomas, C (ed), Academic Press, pp. 591–730.
- 570 **Wedemayer GJ, Wemmer DE, Glazer AN** (1991) Phycobilins of cryptophycean
571 algae. Structures of novel bilins with acryloyl substituents from phycoerythrin 566. J
572 Biol Chem **266**:4731–41.
- 573 **Wedemayer GJ, Kidd DG, Glazer AN** (1996) Cryptomonad biliproteins: Bilin types
574 and locations. *Photosynth Res* **48**:163–170.
- 575 **Wemmer DE, Wedemayer GJ, Glazer AN** (1993) Phycobilins of cryptophycean
576 algae. Novel linkage of dihydrobiliverdin in a phycoerythrin 555 and a phycocyanin
577 645. *J Biol Chem* **268**:1658–1669.
- 578 **Wetherbee R, Hill DRA, Mcfadden GI** (1986) Periplast Structure of the Cryptomonad
579 Flagellate *Hemiselmis brunnescens*. *Protoplasma* **131**:11–22.
- 580 **Wiens JJ** (2006) Missing data and the design of phylogenetic analyses. *J Biomed*
581 Inform **39**:34–42.
- 582 **Zuker M** (2003) Mfold web server for nucleic acid folding and hybridization
583 prediction. *Nucleic Acids Res* **31**:3406–3415.

584 **Figures**

585 **Figure 1:** Morphology of *H. aquamarina* (BMAK265). Scale bars 5 µm. (A) Lateral
586 view of the cell in the differential interface contrast. Cell reniform, slightly acute in
587 apical/ antapical ends. (B) Dorso-lateral view in the bright field with polarized light. (C)
588 Lateral view in phase contrast image displaying rows of ejectosomes and the bright
589 refractive body. (D) Dorsal view in phase-contrast showing two refractive bodies. (E)
590 Lateral view obtained with confocal microscopy showing the natural fluorescence of the
591 plastid and its boat-shaped form. (F) Dorso-lateral view of the cell by confocal
592 microscopy showing natural fluorescence of plastid and the cell delimitation. Plastid
593 (Pl), large ejectosomes (Le), large flagellum (Lf), pyrenoid (P), refractive body (R),
594 starch (S) and small flagellum (Sf).

595 **Figure 2:** Morphology of *H. aquamarina* (BMAK265) by scanning electron
596 microscopy. (A) Apical view showing the vestibular region and the flagella. (B) Lateral
597 view showing the SPC hexagonal plates, flagella and large ejectosome discharged. (C)
598 Ventro-lateral view of the vestibular region showing differences in SPC plates. (D)
599 Dorso-lateral view of the cell displaying the hexagonal SPC plates, flagellar insertion
600 and small ejectosomes discharged. (E) Dorso-lateral view close to the flagellar
601 insertion. (F) Antapical view of the cell showing the SPC mid-basal line. All scale bars
602 1µm, except in panel (E) where it is 500nm. Flagellum (F), large ejectosomes (Le),
603 large flagellum (LF), middle-basal line (MI), mastigonemes (Ms), small ejectosome
604 (Se), small flagellum (Sf) and vestibulum (V).

605 **Figure 3:** Visible absorption spectra of *H. aquamarina* strains phycobiliprotein extract
606 (Cr-PC564).

607 **Figure 4:** Bayesian tree of the concatenated nSSU and nmSSU rRNA. Phycobiliprotein
608 type (PC- phycocyanin and PE- phycoerythrin) is indicated for each clade. (-) Absence
609 of phycobiliprotein information. Bold names correspond to sequences obtained in this
610 study. Bold circles represent cultures from which the species holotype was obtained
611 according to literature information. Bayesian inference was performed applying the
612 substitution models HKY+G+I and GTR+G+I for the nSSU and nmSSU, respectively.

613 Maximum likelihood inferences were performed using GTR+CAT in a partitioned
614 analysis. Nodes with * are fully supported by posterior probability or bootstrap.
615 Supports below 0.75 posterior probability and 70% bootstrap are omitted. Scale bar
616 indicates the rate of nucleotide substitution per site.

617 **Figure 5:** Helix II of the predicted secondary structure of ITS2. The conserved base
618 pairs among different strains are numbered. CBCs and h-CBCs are in blue, emphasized
619 by solid and dotted arrows, respectively. h-CBCs are indicated just in positions where
620 CBCs occur. The pyrimidine- pyrimidine mismatches are in bold. Nucleotides with *
621 represent an indel region in some clones. Nucleotides that differ between clones are
622 labeled by rectangular boxes.

623 **Figure 6:** Helix III of the predicted secondary structure of ITS2. The conserved base
624 pairs among different strains are numbered. CBCs and h-CBCs are in blue, emphasized
625 by solid and dotted arrows, respectively. h-CBCs are indicated just in positions where
626 CBCs occur. Nucleotides that differ between clones are labeled by rectangular boxes.

627 **Supplementary Tables**

628 **Supplementary Table 1:** List of strains, associated sequences and metadata used in this
629 study. For the molecular markers (nSSU, nmSSU and ITS2) GenBank accession
630 numbers (#) are listed (sequences in bold were obtained in this work). Phycobiliprotein
631 (PBP) type, PC for phycocyanin and PE for phycoerythrin. Temperature, country,
632 habitat and locality.

633 **Supplementary Table 2:** Description of *Hemiselmis* species from the literature.

634 **Supplementary Table 3:** PCR cycling conditions and primers used for PCRs and
635 sequencing reactions.

636 **Supplementary Figures**

637 **Supplementary Figure 1:** Molecular phylogeny trees of *Hemiselmis* estimated by

638 Bayesian inferences. Nodes support represent posterior probability and bootstrap.

639 Strains in bold were sequenced in the present study. (A) Phylogeny inference of nSSU

640 rRNA gene using HKY+G+I as nucleotide substitution model. (B) Phylogeny inference

641 based on sequences of nmSSU rRNA gene, using the nucleotide substitution model

642 GTR+G+I. Maximum likelihood trees were obtained with GTR+CAT for A and B.

643 Nodes with * are fully supported by posterior probability or bootstrap. Supports below

644 0.75 PP or 70% of BS are omitted. Scale bar indicates the rate of nucleotide substitution

645 per site.

646 **Supplementary Figure 2:** *Hemiselmis aquamarina* color images. A) Color picture of

647 cells in light microscopy. B) Dense cultures aspect. C) Cr-PC 564 pigment after

648 extraction.

649 **Supplementary Figure 3:** Ultrastructure of *H. aquamarina* (BMAK265) in
650 transmission electron microscopy. (A) Longitudinal section showing the single plastid,
651 nucleomorph, the mitochondrion, antapical starch grain surrounded by thylakoids and
652 the nucleus. (B) Transversal section showing the plastid, nucleomorph, Golgi body and
653 ventral large ejectosome. (C) Longitudinal section of the cell showing the ventral gullet,
654 large ejectosome, nucleus, dorsal plastid, nucleomorph and Golgi body. (D)
655 Longitudinal section showing the large ejectosomes, mitochondrion, plastid and starch
656 coated pyrenoid. (E) Longitudinal section of a cell displaying the flagellar insertion and
657 the Golgi body, pyrenoid, plastid, and nucleus. (F) Longitudinal section of the flagellar
658 region. Scale bars of panels represent 1 μ m (A, C), 500 nm (B, D, E) and 200 nm (F).
659 Basal body (Bb), plastid (Pl), central pair of microtubules (Cp), double microtubules
660 (Dm), flagellum (F), Golgi body (G), gullet (Gu), large ejectosome (LE), mitochondrion
661 (M), nucleus (Nu), nucleomorph (Nm), starch (S).

662 **Supplementary Figure 4:** *Hemiselmis aquamarina* (BMAK265 and RCC4102)
663 consensus secondary structure of nuclear ITS2. Most conserved sites between
664 *Hemiselmis* species are highlighted in grey. Consensus motifs of the helices termini
665 evidenced by blue squares.

666

667 **Supplementary Data**

668 **Supplementary Data 1:** RCC4102- *Hemiselmis aquamarina* movement at 400X.

669 <https://www.youtube.com/watch?v=07q36edleew>

670 **Supplementary Data 2:** BMAK265 (RCC5634) - *Hemiselmis aquamarina* movement

671 at 400X. <https://www.youtube.com/watch?v=rgsID2wDj4g>

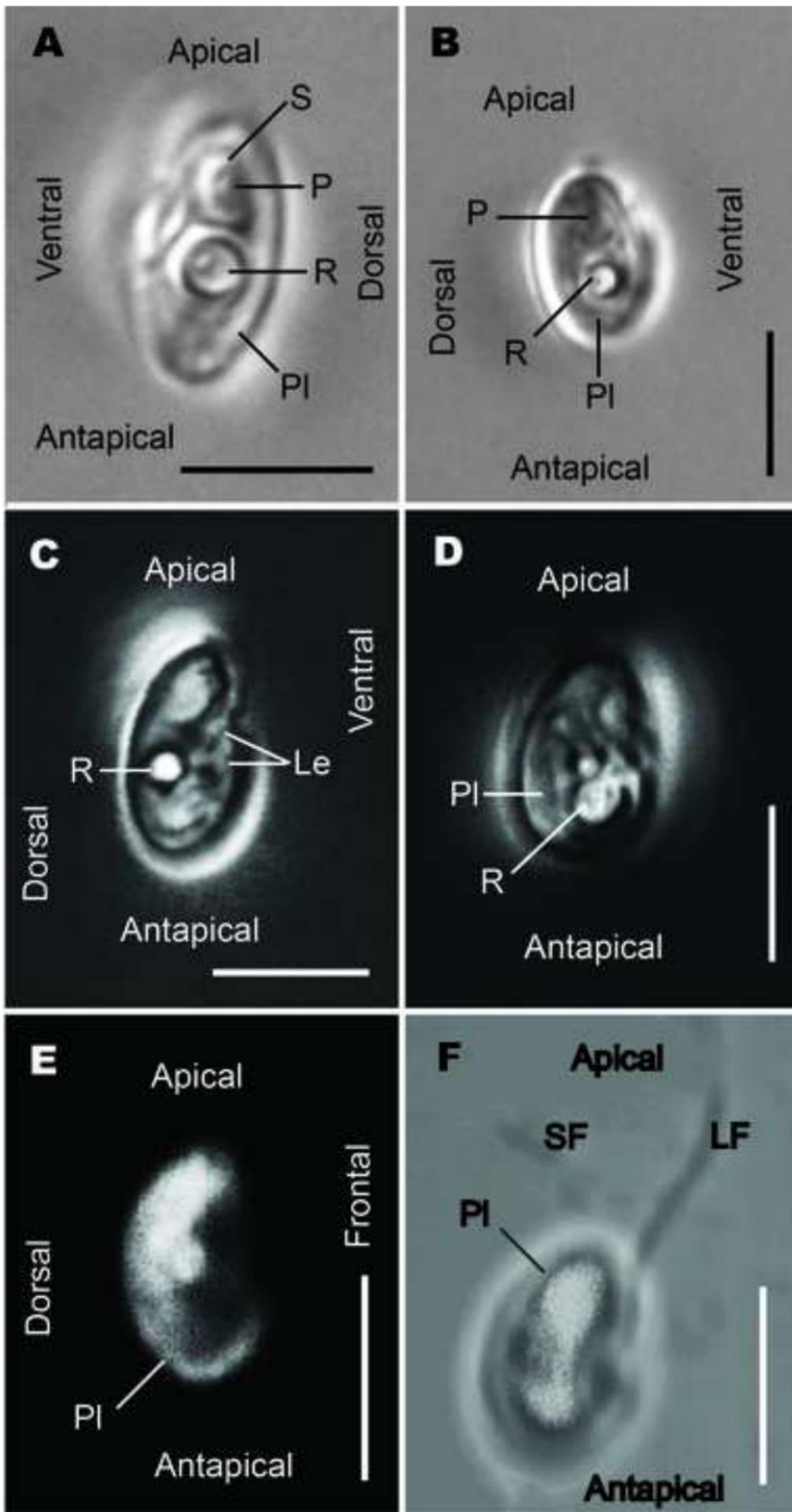
672 **Supplementary Data 3:** Alignment of concatenated SSU rRNA sequences from

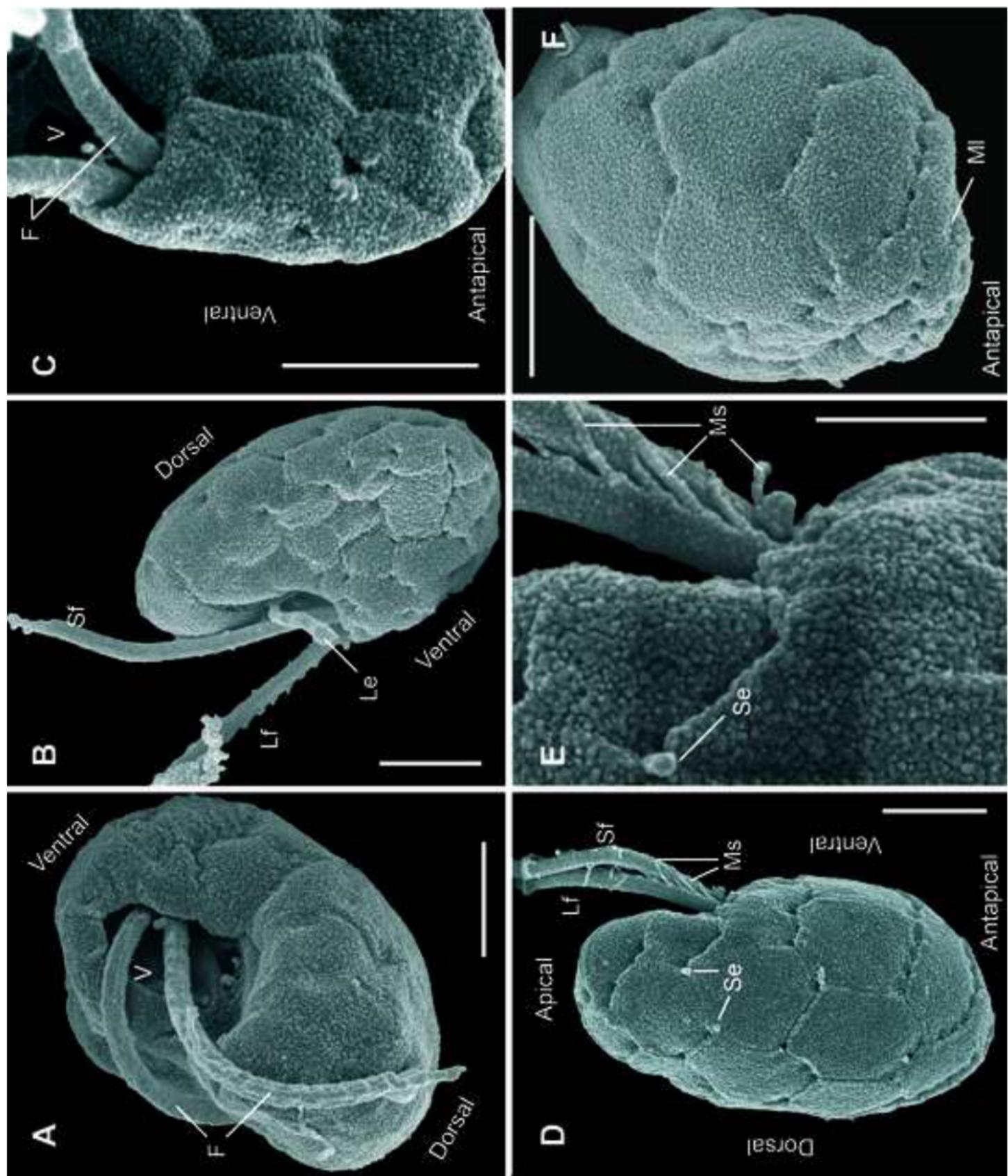
673 nucleus and nucleomorph used to construct Figure 4. Available at

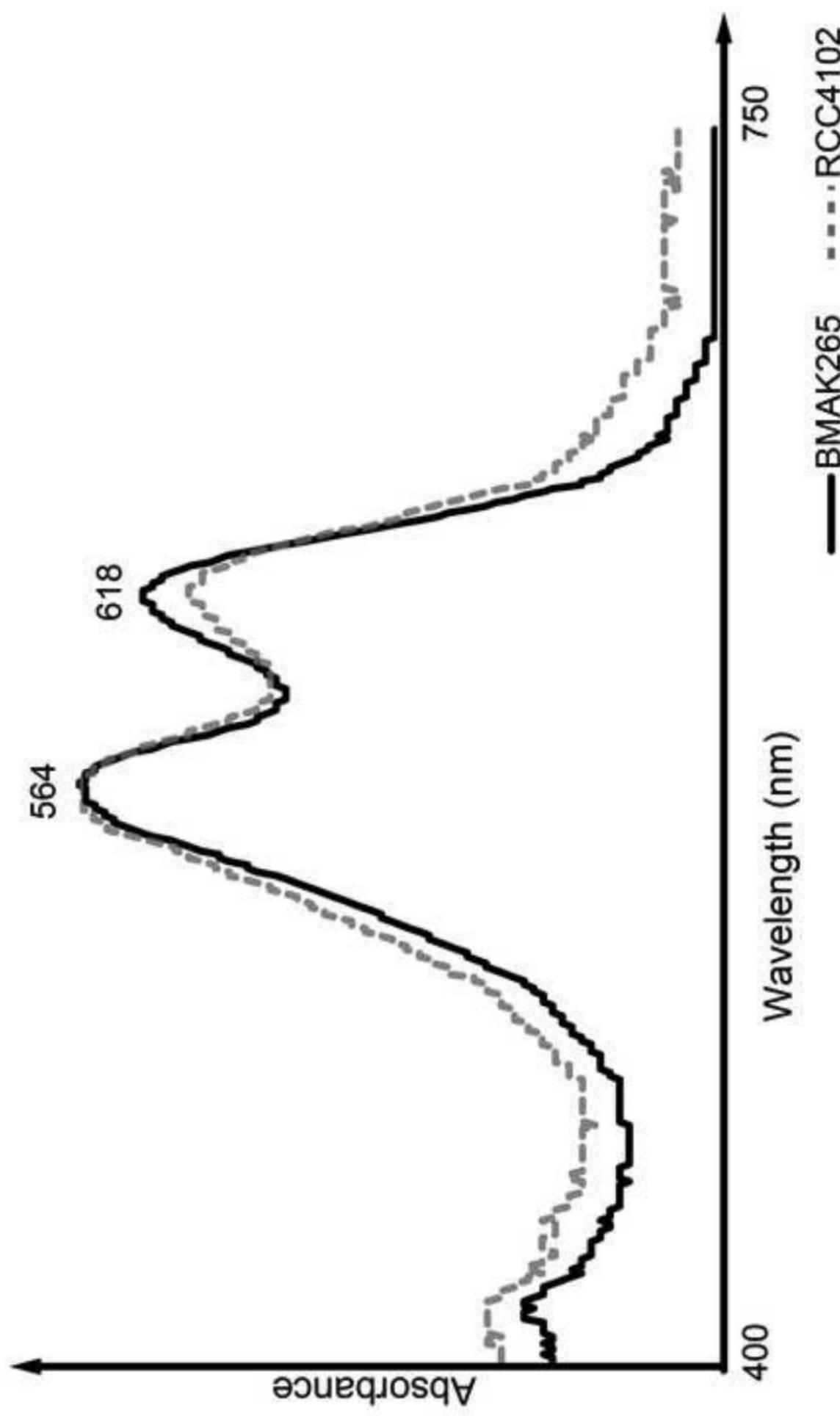
674 <https://figshare.com/s/be3127e6bc7edfb267ce>

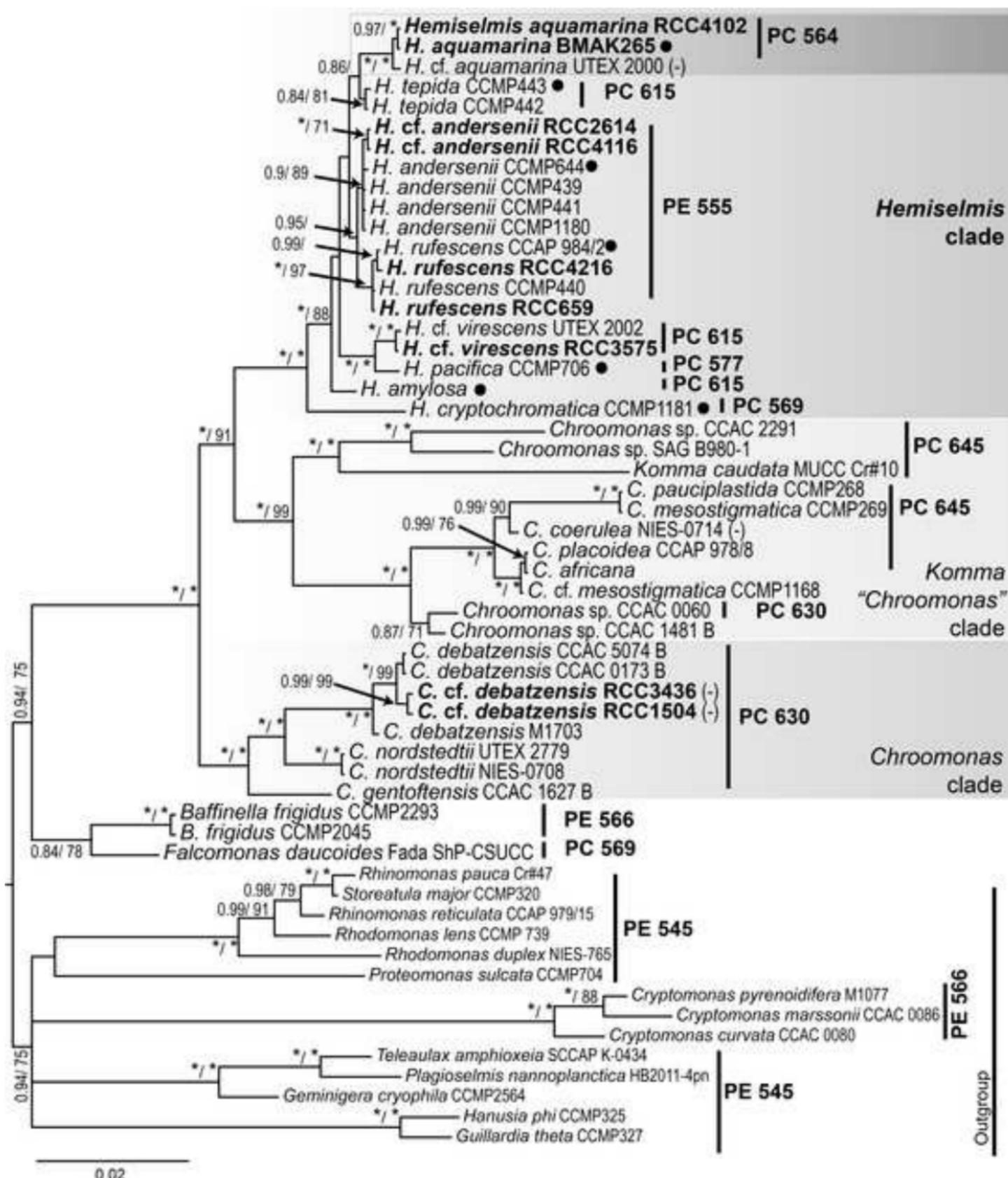
675 **Supplementary Data 4:** Alignment of predicted ITS2 rRNA secondary structures.

676 Available at <https://figshare.com/s/be3127e6bc7edfb267ce>









RCC659 *Hemiselmis rufescens*

RCC2614 *Hemiselmis cf. anderseni*

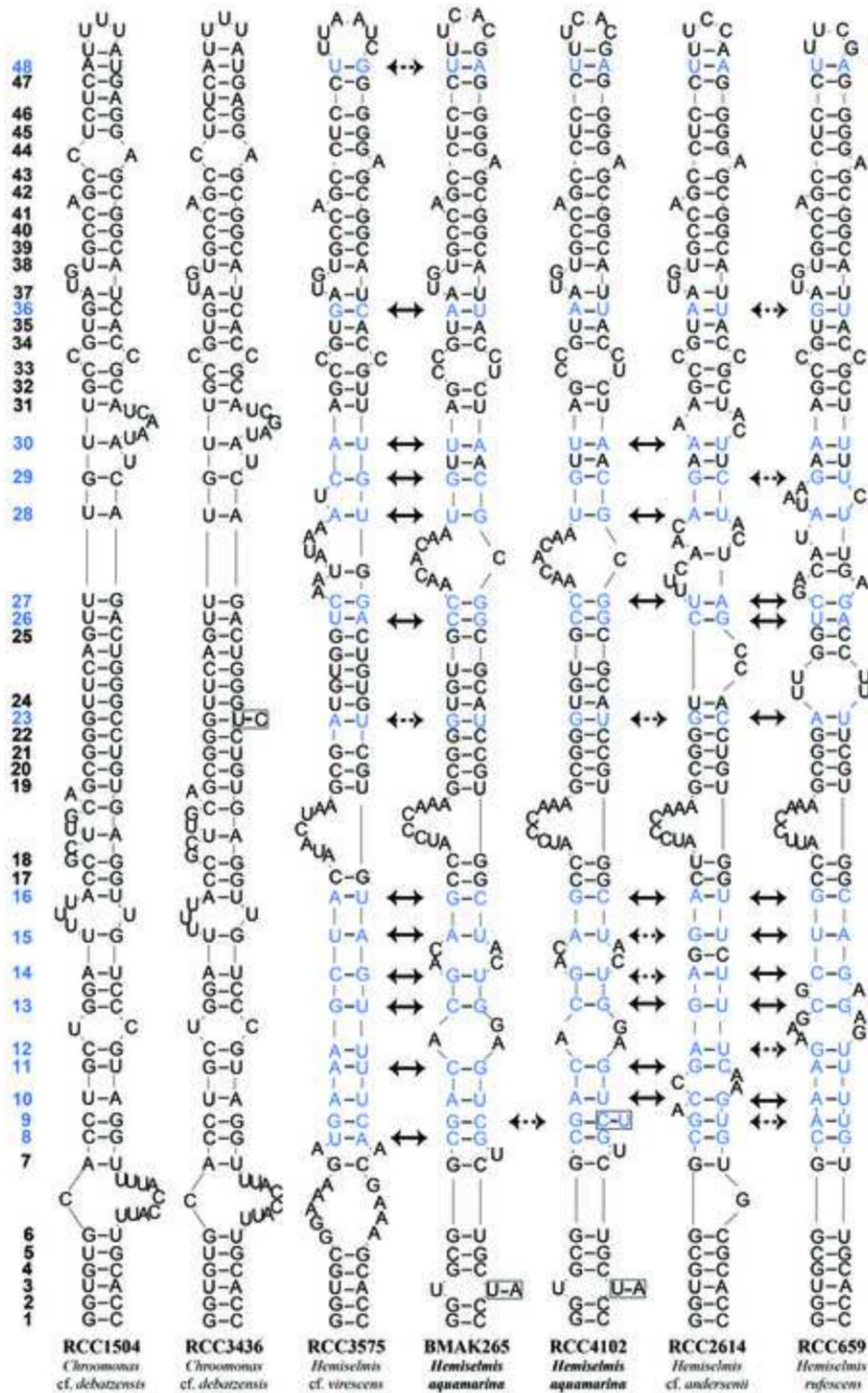
RCC4102 *Hemiselmis aquamarina*

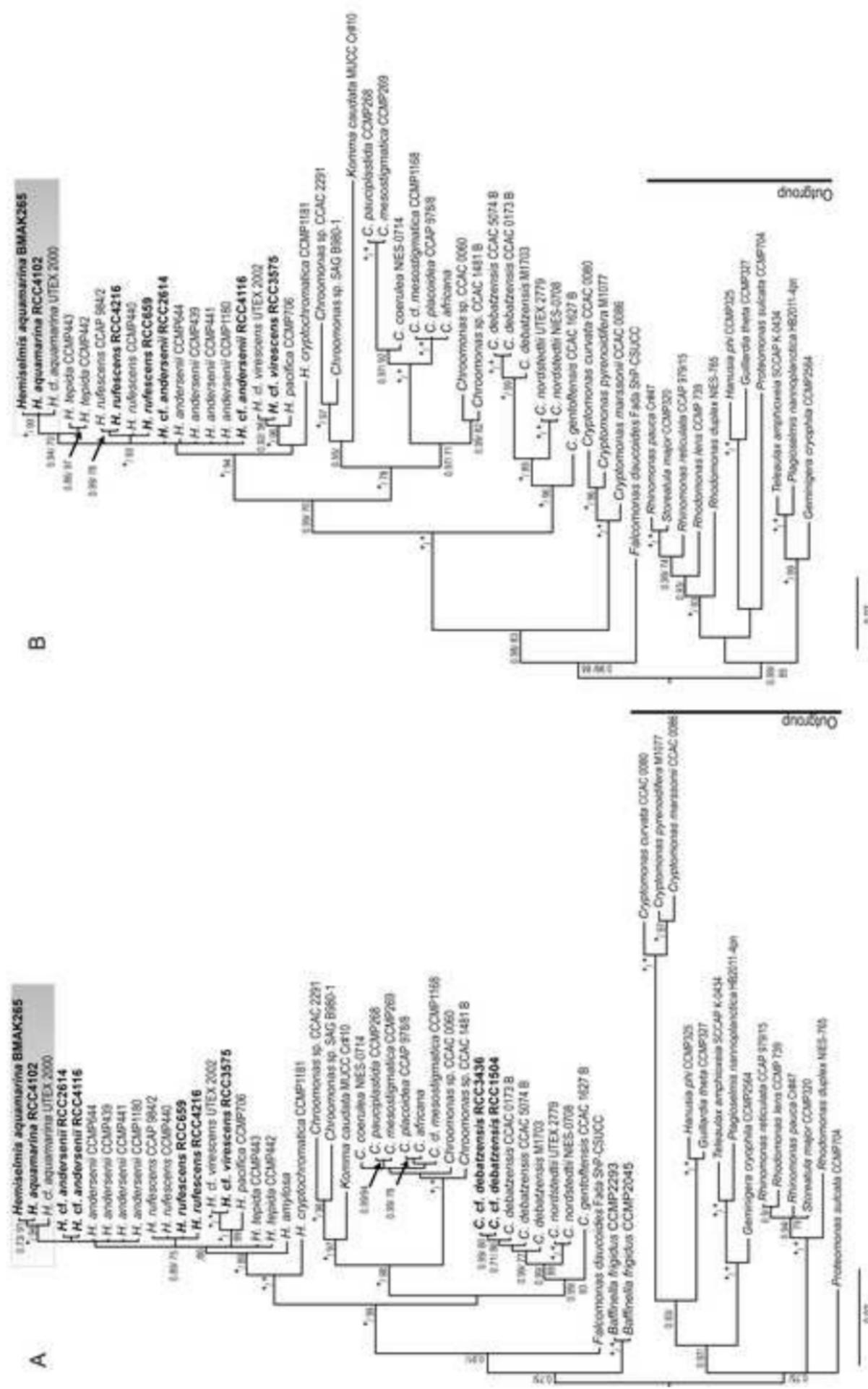
BMAK265 *Hemiselmis aquamarina*

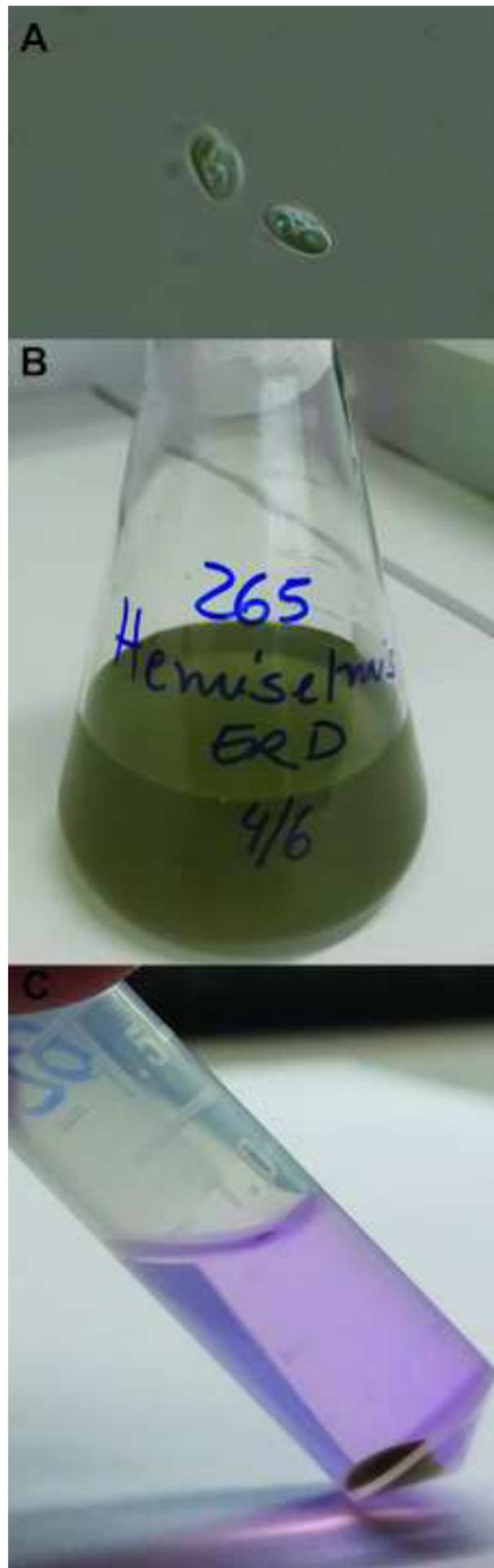
RCC3575 *Hemiselmis cf. virescens*

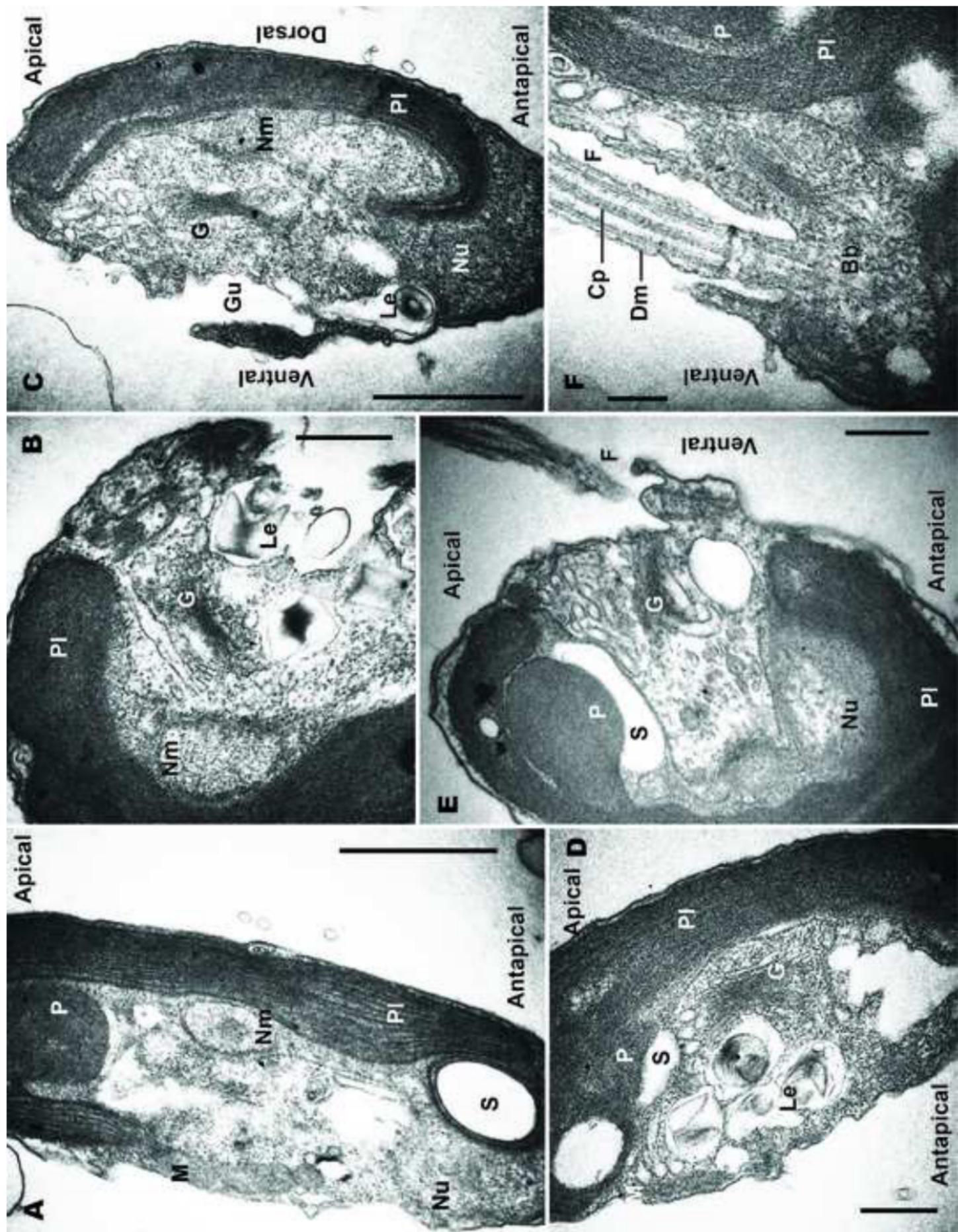
RCC3436 *Chromonas cf. debatzensis*

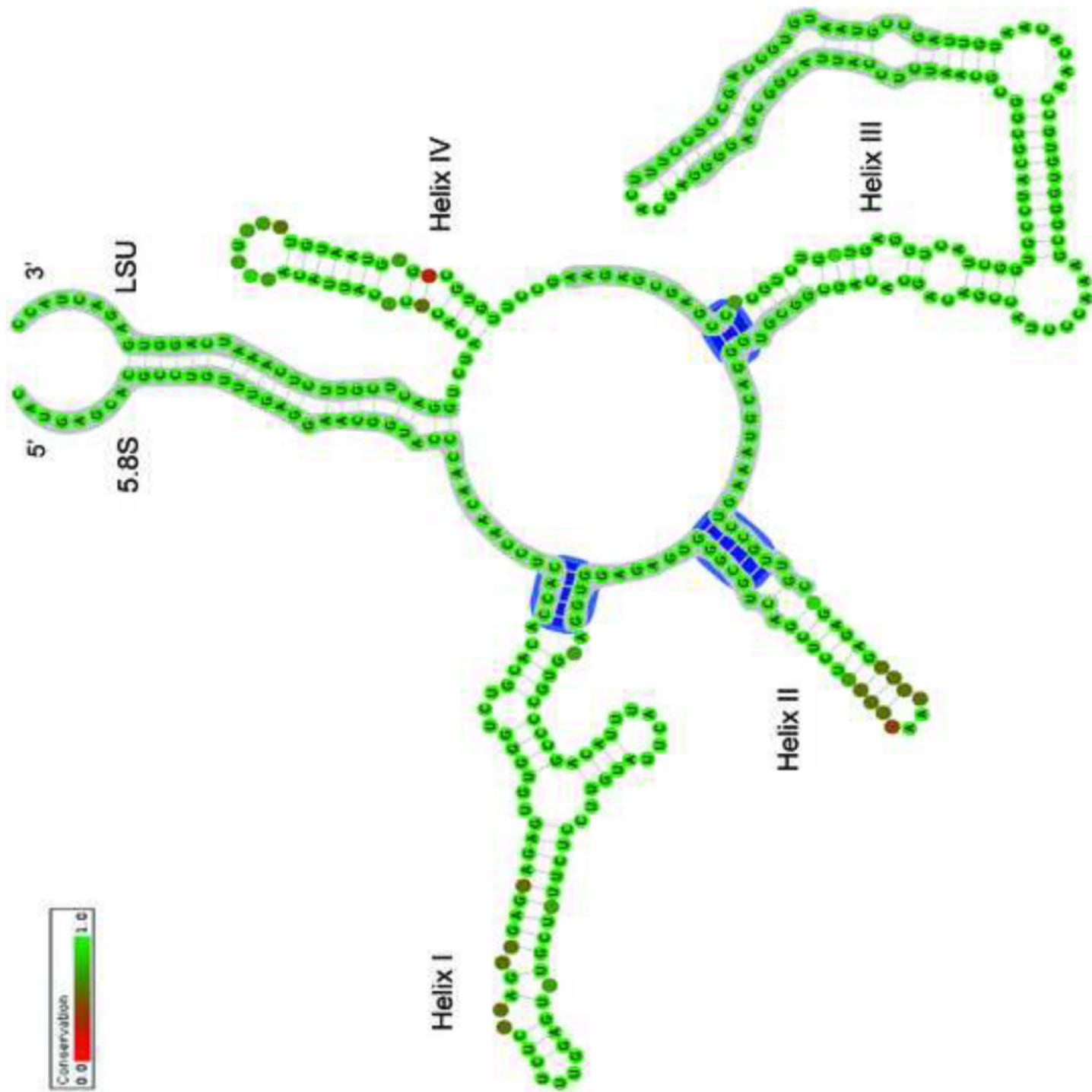
RCC1504 *Chromonas cf. debatzensis*











Supplementary Table 1: List of strains, associated sequences and metadata used in this study. For the molecular markers (nSSU, nmSSU and ITS2) GenBank accession numbers (#) are listed (sequences in bold were obtained in this work). Phycobiliprotein (PBP) type, PC for phycocyanin and PE for phycoerythrin. Temperature, country, habitat and locality.

Name	Strain	nSSU #	nmSSU #	ITS2 #	PBP	Temp.	Country	Habitat	Locality
<i>Hemiselmis amylosa</i> ⁽⁴⁾	Cowl- CSU	AF143944		PC 615 ⁽⁵⁾	USA	freshwater			North America, Cowdry Lake/ Colorado
<i>Hemiselmis andersenii*</i>	CCMP1180	AM901353	AM901022	PE 545/ 555 ⁽²⁾	20°C	MEX	marine		Gulf of Mexico
<i>Hemiselmis andersenii*</i>	CCMP441	AM901350	AM901019	PE 555 ⁽²⁾	20°C	MEX	marine		Gulf of Mexico, Gulf Stream
<i>Hemiselmis andersenii*</i>	CCMP644	AM901351	DQ519365	PE 555 ⁽²⁾	20°C	USA	marine		Gulf of Mexico, Gulf Stream
<i>Hemiselmis andersenii*</i>	CCMP439	AJ007283	AJ420690	PE 555 ⁽²⁾	24°C	USA	marine		Gulf of Mexico, Cape San Blas
<i>Hemiselmis cf. andersenii</i>	RCC4116	MT628036	MT628035	MT628029	PE 551/ 553	20°C	JPN	marine	North Pacific Ocean
<i>Hemiselmis cf. andersenii</i>	RCC2614	MF179473	MF179479	MT628034	15°C	GBR	marine		North Atlantic Ocean, North Sea
<i>Hemiselmis aquamarina</i>	RCC4102	MF179476	MT605191- MT605193	MT628026- MT628028	PC 564	20°C	JPN	marine	North Pacific Ocean
<i>Hemiselmis aquamarina</i>	BMAK265	MT605165	MT605187- MT605166	MT628030- MT628033	PC 564	20°C	BRA	marine	South Atlantic Ocean, Ubatuba, SP
<i>Hemiselmis cf. aquamarina</i>	UTEX2000 ⁽⁵⁾	AM901367	AM901034		USA		marine		North Atlantic Ocean, Virginia, York River
<i>Hemiselmis cryptochromatica</i> *	CCMP1181	AM901354	AM901023	PC 659 ⁽⁴⁾	14°C	USA	marine		North Atlantic Ocean, Boothbay Harbor, Maine
<i>Hemiselmis pacifica</i> *	CCMP706	AM901352	AM901020	PC 576 ⁽¹⁾	14°C	USA	marine		North Atlantic Ocean, Washington, San Juan Island

<i>Hemiselmis rufescens</i>	RCC659	MF179475	MF179481	MT628025	PE 554/ 555	15°C	NOR	marine	North Atlantic Ocean, North Sea
<i>Hemiselmis rufescens</i>	RCC4216	MF179474	MF179480		PE 554/ 555	15°C	FRA	marine	North Atlantic Ocean, English Channel
<i>Hemiselmis rufescens*</i>	CCAP 984/2	AJ007282	AM901016		PE 554 ⁽²⁾		GBR	marine	North Atlantic Ocean, English Channel
<i>Hemiselmis rufescens*</i>	CCMP440	AM901349	AM901018		PE 555 ⁽³⁾	20°C	USA	marine	North Atlantic Ocean, Maine, West Boothbay Harbor
<i>Hemiselmis tepida*</i>	CCMP443	AJ007284	AJ420691		PC 612 ⁽³⁾	20°C	USA	marine	North Atlantic Ocean, Galveston Channel, Texas
<i>Hemiselmis tepida*</i>	CCMP442	HM126533	EF594307		PC 612 ⁽³⁾	20°C	USA	marine	North Atlantic Ocean, Galveston Channel, Texas
RCC3575, (M1635, CACC1635		MF179477	MF179484	MT628024	PC 612⁽³⁾	17°C	SWE	marine	Kristineborg, Baltic Sea
B)									
<i>Hemiselmis virescens*</i>	UTEX 2002	AM901368	AM901035		PC 614 ⁽³⁾		USA	marine	North Atlantic Ocean, Virginia, York River
<i>Chroomonas africana</i>		HG328376	HG328384		PC 645 ⁽³⁾		ZAF	marine	South Atlantic Ocean, Cape Province, Yzerfontein region
<i>Chroomonas coerulea</i>	NIES-0714	HG328381	HG328389				JPN	freshwater	Asia, Honshu, Nagano, Sugadaira
<i>Chroomonas debatzensis*</i>	CCAC 5074 B	MG387973	MG387972				FRA	marine	North Atlantic Ocean, English Channel, Roscoff
<i>Chroomonas debatzensis*</i>	CCAC 0173 B (M1318)	AJ007279	AJ420679				FRA	marine	North Atlantic Ocean, English Channel, île de Batz

<i>Chroomonas debatzensis</i> *	M1703	AJ420699	AJ420681	PC 630 ⁽³⁾	DNK	Jutland, Hjerting
<i>Chroomonas cf. debatzensis</i>	RCC3436	MF589232		MT628020, MT628021	Unknown	marine Unknown
<i>Chroomonas cf. debatzensis</i>	RCC1504	MT628016		MT628022, MT628023	17°C	FRA marine
<i>Chroomonas gentoftensis</i> *	B (M1627)	AM901360	AM901029	PC 630 ⁽³⁾	DNK	Baltic Sea, Sjaelland, Bellevue Strandbad
<i>Chroomonas mesostigmatica</i>	CCMP1168	AF508268	AM901021	PC 645 ⁽³⁾	Unknown	Unknown
<i>Chroomonas mesostigmatica</i>	CCMP0269	AM901347	AM901017	PC 645 ⁽³⁾	USA	North Atlantic Ocean, Maryland, Assateague Island
<i>Chroomonas nordstedtii</i> *	NIES-0708	HG328378	HG328386		JPN	Asia, Hokkaido, Sapporo, Hokkaido University
<i>Chroomonas nordstedtii</i> *	UTEX2779	AM901369	AM901036	PC 630 ⁽³⁾	USA	North America, Colorado, Wellington Reservoir
<i>Chroomonas pauciplastida</i>	CCMP0268	AM901346	DQ519363	PC 645 ⁽³⁾	USA	North Atlantic, Nantucket Sound
<i>Chroomonas placoidea</i>	CCAP 978/08	AM901345	AM901015	PC 645 ⁽³⁾	GBR	North Atlantic, Irish Sea, Yorkshire
<i>Chroomonas sp.</i>	CCAC 0060 (M0874)	AM901357	AM901026	PC 630 ⁽³⁾	DEU	Europe, Griether Ort
<i>Chroomonas sp.</i>	SAG 980-1	AJ420698	AJ420677	PC 645 ⁽³⁾	GBR	Europe, Wales
<i>Chroomonas sp.</i>	CCAC 2291 (M2291/1)	AM901366	AM901033	PC 645 ⁽³⁾	DEU	Europe, Cologne, Wahner Heide
<i>Chroomonas sp.</i>	CCAC 1481 B (M1481)	AJ007278	AJ420680	PC 645 ⁽³⁾	DEU	Europe, Spessart, Biebergemünd

<i>Komma caudata</i>	MUCC Cr#10	U53122	U53121	PC 645 ⁽³⁾	AUS	freshwater	Oceania, Wimmera river
<i>Cryptomonas curvata</i>	CCAC 0080	AM051189	AJ715462	DEU	freshwater	Europe, Muenster	
<i>Cryptomonas marssonii</i>	CCAC 0086	AM051191	AJ566173	DEU	freshwater	Muenster	
<i>Cryptomonas pyrenoidifera</i>	B (M1077)	AM051197	AJ566180	DEU	freshwater	Europe, Cologne	
<i>Falcomonas daucoides</i>	Fada ShP- CSUCC	AF143943	AJ420689	PC 569 ⁽³⁾	USA	marine	North Pacific Ocean, Shannon Point, Washington
<i>Geminigera cryophila</i>	CCMP2564	DQ452091	DQ452092	ATA	marine	Southern Ocean, McMurdo Sound	
<i>Guillardia theta</i>	CCMP327	X57162	AJ010592	PE 545 ⁽³⁾	USA	marine	North Atlantic, Long Island Sound
<i>Hanusia phi</i>	CCMP325	U53126	U53125	USA	marine	North Atlantic, Milford, Connecticut USA, Long Island Sound	
<i>Plagioselmis nannoplantica</i>	HB2011-4pn	KC928320	KC928321	Unknown	Unknown	Unknown	
<i>Proteomonas sulcata</i>	CCMP704	AJ007285	AJ420692	Unknown	Unknown	Unknown	
<i>Rhinomonas pauca</i>	MUCC Cr#47	U53132	U53131	PE546 ⁽³⁾	AUS	marine	Bass Strait, Hobsons Bay
<i>Rhinomonas reticulata</i> var. <i>reticulata</i>	CCAP 979/15	HF952562	HF952608	GBR	marine	English Channel, Plymouth Sound	
<i>Rhodomonas duplex</i>	NIES-765	HF952604	HF952620	JPN	marine	East China Sea, Okinawa, Yaga	
<i>Rhodomonas lens</i>	CCMP739	HF952574	HF952611	marine	Gulf of Mexico	Unknown	
<i>Storeatula major</i>	CCMP320	U53130	U53129	Unknown	Unknown	Unknown	
<i>Teleaulax amphioxiae</i>	SCCAP K- 0434	AJ007287	AJ421146	Unknown	Unknown	Unknown	

Names in bold represents the strains sequenced in this study.

- * Names revised in the study of Lane and Archibald (2008) and Hoef-Emden (2018).
- ⁽¹⁾ Value of pigments according to Lane and Archibald (2008).
- ⁽²⁾ Value of biliprotein maximum absorption according to Hill and Rowan (1989) and references therein.
- ⁽³⁾ Value of biliprotein maximum absorption according to Hoef-Emden (2008) and references therein.
- ⁽⁴⁾ Value of biliprotein maximum absorption according to Cunningham et al. (2019).
- ⁽⁵⁾ Species name and phycobiliprotein characterization are according to Clay and Kugrens (1999).
- ⁽⁶⁾ Strain is deposited in UTEX culture collection as *Chroomonas* sp. (see <https://utex.org/products/utex-lb-2000>)

Supplementary Table 2: Description of *Hemiselmis* species from the literature.

<i>Hemiselmis</i> species	Cell shape			Cell measurements			Biliprotein strain	Authentic strain	Habitat	Type locality
	lateral view	dorsal/ ventral view	cell end	length (μ)	width (μ)	Chromatophore description				
<i>H. amylosa</i> Clay & Kugrens, 1999	bean-shaped	cylindrical	rounded	5.-6	3-3.5	parietal	Cr-PC 615	Cowl CSU	freshwater	Cowdry Lake, Colorado (USA)
<i>H. aquamarina</i> Magalhães & Oliveira, 2020	bean-shaped	ovoid	rounded	4.5-7.5	2.7-4.4	parietal, lobate, light green	Cr-PC 564	BMAK265 (syn. RCC5634)	marine, coastal	Ubatuba, SP (BRA)
<i>H. andersenii</i> Lane & Archibald, 2008	reniform	ovoid	acute	5.5-8.5	3.-5	parietal, orange to dark red	Cr-PE 555	CCMP644	marine, oceanic	Gulf Stream
<i>H. cryptochromatica</i> Lane & Archibald, 2008	reniform	obvoid to pyriform	rounded	4.5-6.5	3-4.5	faint gray	Cr PC 630	CCMP1181	marine, coastal	Boothbay Harbor, Maine (USA)
<i>H. pacifica</i> Lane & Archibald, 2008	reniform	ovate	acute	7-8.5	4-6.0	parietal, grass- green to grey- green	Cr PC 615	CCMP706	marine, coastal	Friday Harbor, San Juan Island, Washington (USA)
<i>H. rifescens</i> Parke, 1949	bean-shaped		acute	4-8.5	3.5-5	parietal, lobated, French Rose	Cr-PE 555	PCCS63	marine, coastal	Port Erin, Isle of Man, (GBR)
<i>H. tepla</i> Lane & Archibald, 2008	reniform	ovate	rounded	5.5-7	3.5-4.5	parietal, absinthe-green to emerald	Cr-PC 612	CCMP443	marine, coastal	Galveston Channel, Texas (USA)
<i>H. virescens</i> Droop, 1955	bean-shaped/ falcate	cylindrical		5-7	2.5-3	brilliant green, turquoise to bottle-green	Cr-PC 612	no. 64	marine, coastal	Cumbrae, Scotland (GBR)
<i>H. amyliifera</i> Butcher nom. inval. 1967	oblong, ovoid	slightly compressed	rounded	7.5	2.5-3	single, olive- yellow (CC 235)	unknown	unavailable	marine	Lowestoft, Suffolk (GBR)
<i>H. anomala</i> Butcher nom. inval. 1967	bean-shaped	ovoid	rounded	5-6.5	3-3.5	parietal, Paris green or Neptune green	unknown	unavailable	marine	Carmathenshire, Caernarvonshire (GBR)

<i>H. brunnescens</i> Butcher, nom. inval. 1967	ovate, bean-shaped	slightly compressed/ elliptical	obtuse- rounded	5-5.5	3	salmon-pink/ tangerine- orange	Cr-PE 555	PCC14 (syn. CCAP984/2)	marine	English Channel
<i>H. cyclorea</i> Butcher nom. inval. 1967	ovate, bean-shaped	slightly compressed/ elliptical	obtuse	4.5- 5.5	4.5- 5	poppy red, parietal	unknown	unavailable	marine	Fishguard, Lowestoft and Conway (GBR)
<i>H. oculata</i> Butcher nom. inval. 1967	ovate, bean-shaped	elliptical, uncompressed	rounded	5-8	4-4.5	two, poppy red parietal	unknown	unavailable	marine	Knap Buoy, English Channel
<i>H. simplex</i> Butcher, nom. inval. 1967	bean- shaped	ovate, lanceolate or cylindrical	rounded	5-6.5	3.5-4	Paris green	unknown	unavailable	marine, coastal	Yorkshire, Lowestoft, Southend, and Conway (GBR)
<i>H. rotunda</i> Butcher nom. inval. 1967	ovate	uncompressed	very rounded	4-5	2.5-3	parietal, lobated, Neptune green, absinthe	unknown	unavailable	brackish, marine	Bembridge, Isle of Wight (GBR)

Supplementary Table 3

PCRs cycles for nmSSU of initial denaturation at 94°C during 5min. Followed by 35 cycles of: (i) 94°C for 30sec; (ii) 60°C for 1min and (iii) 72°C for 2min. Final extension step at 72°C for 7 minutes. PCRs for nSSU were performed as indicated in (Majaneva et al. 2014) and ITS2 as described in Hoef-Emden & Melkonian 2003.

Primers used for PCRs and sequencing reactions

Primers for nmSSU PCRs		
Sequences of oligonucleotides (5'- 3')		
CrNM1F	CAG TAG TCA TAT GCT TGT CTT AAG	(Hoef-Emden and Melkonian 2003)
SSUBR	TTG ATC CTT CTG CAG GTT CAC CTA C	(Hoef-Emden and Melkonian 2003)
18S 5'	CCA CCT GGT TGA TCC TGC CAG T	(Sogin 1990)
18S 3'	GAT CCT TCT GCA GGT TCA CCT ACG GAA	(Sogin 1990)
Primers for nSSU PCRs		
18SNF2	TGA TGG TCC CTT ACT ACA	(Majaneva et al. 2014)
SSUR	CTT GTT ACG ACT TCT CCT	(Majaneva et al. 2012)
Primers for nSSU and nmSSU sequencing reactions		
18S 5'	CCA CCT GGT TGA TCC TGC CAG T	(Sogin 1990)
CrNM1F	CAG TAG TCA TAT GCT TGT CTT AAG	(Hoef-Emden and Melkonian 2003)
528F	CGG TAA TTC CAG CTC C	(Sogin 1990)
1055F	GGT GGT GCA TGG CCG	(Bellorin et al. 2002)
18S 3'	GAT CCT TCT GCA GGT TCA CCT ACG GAA	(Sogin 1990)
536R	GAA TTA CCG CGG CTG CTG	(Bird et al. 1992)
1055R	CGG CCA TGC ACC ACC	(Bird et al. 1992)
18SNF2	TGATGGTCCCTTACTACA	(Majaneva et al. 2014)
SSUBR	TTG ATC CTT CTG CAG GTT CAC CTA C	(Hoef-Emden and Melkonian 2003)
SSUR	CTT GTT ACG ACT TCT CCT	(Majaneva et al. 2012)
Primers for ITS PCR and sequencing reaction		
ITS03F- 800	CGA TGA AGA ACG YAG CGA	(Hoef-Emden and Melkonian 2003)
ITS05R- 700	TAC TTG TTC GCT ATC GGT CTC T	(Hoef-Emden and Melkonian 2003)

References

- Bellorin, A.M., Oliveira, M.C. & Oliveira, E.C. 2002. Phylogeny and systematics of the marine algal family Gracilariaeae (Gracilariales, Rhodophyta) based on small subunit rDNA and its sequences of Atlantic and Pacific species. *J Phycol.* 38:551–63.
- Bird, C.J., Rice, E.L., Murphy, C. a. & Ragan, M. a. 1992. Phylogenetic relationships in the Gracilariales (Rhodophyta) as determined by 18S rDNA sequences*. *Phycologia*. 31:510–22.
- Hoef-Emden, K. & Melkonian, M. 2003. Revision of the Genus Cryptomonas (Cryptophyceae): a Combination of Molecular Phylogeny and Morphology Provides Insights into a Long-Hidden Dimorphism. *Protist*. 154:371–409.
- Majaneva, M., Remonen, I., Rintala, J.-M., Belevich, I., Kremp, A., Setälä, O., Jokitalo, E. et

Supplementary Table 3

- al. 2014. Rhinomonas nottbecki n. sp. (Cryptomonadales) and Molecular Phylogeny of the Family Pyrenomonadaceae. *J Eukaryot Microbiol.*
- Majaneva, M., Rintala, J.-M., Piisilä, M., Fewer, D.P. & Blomster, J. 2012. Comparison of wintertime eukaryotic community from sea ice and open water in the Baltic Sea, based on sequencing of the 18S rRNA gene. *Polar Biol.* 35:875–89.
- Sogin, M.L. 1990. Amplification of ribosomal RNA genes for molecular evolution studies. In Innes, M. A., Gelfand, D. H., Snisnky, J. J. & White, T. J. [Eds.] *PCR Protocols. A Guide to Methods and Applications*. M.A. Innes. Academic Press, San Diego, pp. 307–14.

Declaration of Competing Interest

The authors declare that they have no conflict of interest

CRediT author statement

KM: conceptualization, methodology, validation, formal analysis, investigation, data curation, writing (original draft). ALS: methodology, data curation and validation, writing (review and editing). DV: validation, resources, writing (review and editing) and funding acquisition. MCO: resources, writing (review and editing), supervision, project administration and funding acquisition.