

ORIGINAL PAPER

Hemiselmis aquamarina sp. nov. (Cryptomonadales, Cryptophyceae), A Cryptophyte with A Novel Phycobiliprotein Type (Cr-PC 564)



Karoline Magalhães^{a,1}, Adriana Lopes Santos^b, Daniel Vaultot^{c,b}, and Mariana Cabral Oliveira^a

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

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

^cSorbonne Université, CNRS, UMR 7144, ECOMAP Team Station Biologique de Roscoff, Roscoff, France

Submitted March 24, 2021; Accepted July 30, 2021

Monitoring Editor: Michael Melkonian

Cryptophytes are a small group of photosynthetic biflagellate organisms distributed worldwide in fresh, brackish and marine waters. Although members of this class are easily distinguished from other groups, species identification is difficult and studies concerning their diversity are scarce. Two strains of an undescribed *Hemiselmis* species were isolated from the marine waters off Brazil and Japan. Analyses of morphology, phycobiliprotein spectral characterization, molecular phylogeny and ITS2 secondary structure comparisons were performed to assist the identification. The morphological features of *Hemiselmis aquamarina* sp. nov. matches that of other species from the same genus, but it has a new type of phycocyanin. Molecular phylogeny and ITS2 secondary structure support *H. aquamarina* as a distinct species. Furthermore, phylogenetic inferences indicate *H. aquamarina* as closely related to *H. tepida*, *H. andersenii* and *H. rufescens*. Currently, all *Hemiselmis* species have been described from the Northern Hemisphere and most from the subtropical region. *H. aquamarina* is the first species of this genus described from the South Atlantic.

© 2021 Elsevier GmbH. All rights reserved.

Key words: Morphology; molecular phylogeny; phytoplankton; South Atlantic Ocean; Pacific Ocean.

¹ Corresponding author;
e-mail karolinemfl@alumni.usp.br (K. Magalhães).

Introduction

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

The plastid of cryptophytes is surrounded by four membranes, originating from secondary endosymbiosis with a red algal ancestor and it still possesses a relict of the endosymbiont nucleus, the nucleomorph (Douglas and Penny 1999; Douglas et al. 2001; Keeling 2010).

The photosynthetic pigments of cryptophytes include chlorophylls *a*, *c*, carotenoids and phycobiliproteins-PBPs (Cunningham et al. 2019; Spear-Bernstein and Miller 1989). The PBPs of cryptophytes are a heterodimer composed of two subunits, α and β , with four bilins linked at the positions α -Cys 18 (19), β -DiCys 50, 61, β -Cys 82, and β -Cys 158 (Glazer and Wedemayer 1995; Wedemayer et al. 1996; Wemmer et al. 1993). While four isomeric bilins are found in cyanobacteria and red algal PBPs, cryptophyte PBPs display six bilins: 15,16-dihydrobiliverdin (DHBV), phycocyanobilin (PCB), phycoerythrobilin, mesobiliverdin, and the two acryloyl bilins: bilin 584 and bilin 618. The last two bilins are only known within this group (Glazer and Wedemayer 1995; Wedemayer et al. 1991).

A given strain of cryptophyte possesses a single type of spectroscopically distinct PBP, cryptophyte phycocyanin (Cr-PC), or phycoerythrin (Cr-PE), and its classification is determined by its bluish or reddish appearance, respectively (Overkamp et al. 2014; Wedemayer et al. 1996; Wemmer et al. 1993). Moreover, all Cr-PE investigated up to date possess a single absorption peak (Hoef-Emden 2008). Currently, eight different types of PBPs have been recognized according to their maximum visible absorption spectrum: Cr-PC 569, Cr-PC 577, Cr-PC 615, Cr-PC 630, Cr-PC645, Cr-PE 545, Cr-PE 555 and Cr-PE566 (Cunningham et al. 2019; Greenwold et al. 2019).

Hemiselmis was first described in 1949 from the Isle of Man, UK, by the typification of *Hemiselmis rufescens* Parke (1949), which has a reddish plastid. The next species described was *Hemiselmis virescens* Droop (1955), a blue-green member, from

Cumbrae, Scotland. In 1967, when Butcher revised the classification of cryptophytes, he created the family Hemiselmidaceae (Butcher 1967) based on the position of the gullet across the short axis of the cell. Moreover, he described two subgenera, *Hemiselmis* and *Plagiomonas*, distinguished by their reddish and bluish color, respectively. Assisted mostly by light microscopy investigations, he assigned eleven species to *Hemiselmis*, including seven new descriptions and two freshwater species with problematic classification history (*Nephroselmis olivacea* sensu Pascher and *Sennia parvula* Skuja). Unfortunately, he attributed many type localities for the majority of his descriptions, except for *H. amyliifera*, *H. oculata*, and *H. rotunda*. Therefore, due to the uncertainty concerning the morphological traits chosen by Butcher, the inaccuracy in localities for typification, the absence of cultures from which the types were obtained and, consequently, the inability to validate them, his descriptions of *Hemiselmis* species have been treated as illegitimate (Lane and Archibald 2008).

In the last decades, five new species of *Hemiselmis* have been described with the help of the rRNA operon sequences. *Hemiselmis amylosa* Clay & Kugrens was the first freshwater species described for the genus, from Colorado Lake, USA (Clay and Kugrens 1999). *Hemiselmis andersenii* Lane & Archibald, *Hemiselmis cryptochromatica* Lane & Archibald, *Hemiselmis pacifica* Lane & Archibald and *Hemiselmis tepida* Lane & Archibald are from marine environments. Except for *H. pacifica*, which is from the North Pacific Ocean, all others were recorded from the North Atlantic Ocean (Lane and Archibald 2008).

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

Taxonomy Section

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

Description: free-swimming cells, reniform in lateral view, with rounded ends. Length from 4.5 to

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

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

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

Type strain: BMAK265 (synonymous RCC5634)

Other strain: RCC4102

Type locality: 23.59745 S, 45.02833 W, Ubatuba, São Paulo, Brazil.

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

Results

Morphological Characterization

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

One parietal lobed plastid, boat-shaped and light blue-green (Supplementary Material Fig. S2A), occupies the dorsal part of the cell extending towards the lateral sides (Fig. 1). Dense cultures are hunter-green (Supplementary Material Fig. S2B). Four membranes enclose the plastid and a prominent subapical starch-coated pyrenoid can be seen (Fig. 1A-B, Supplementary Material Fig. S3). The thylakoids are densely packed within the plastid with little free stroma; sometimes a parallel arrangement pattern is observed (Supplementary Material Fig. S3A). Single thylakoids penetrate the pyrenoid core. No stigma was detected.

The nucleomorph is situated in the ventral part of the plastid, directly below the subapical pyrenoid, almost in the center of the cell. A double-membrane envelope surrounds the nucleomorph,

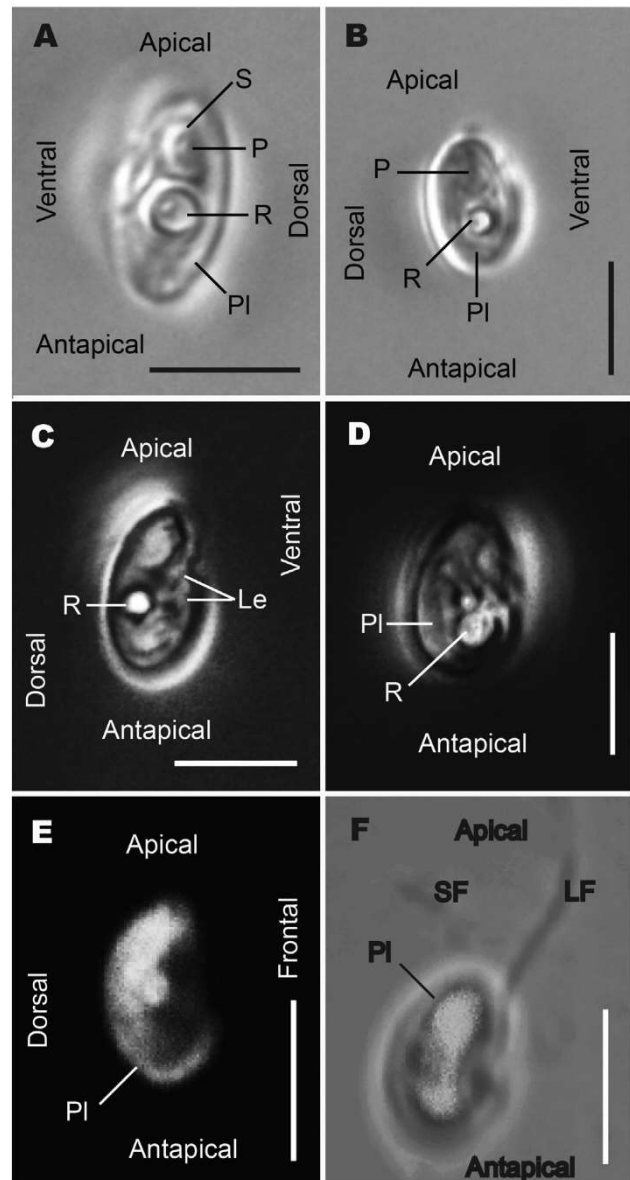


Figure 1. Morphology of *H. aquamarina* (BMAK265). Scale bars 5 μm . (A) Lateral view of the cell in the differential interface contrast. Cell reniform, slightly acute in apical/ antapical ends. (B) Dorso-lateral view in the bright field with polarized light. (C) Lateral view in phase contrast image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

which has a granular matrix ([Supplementary Material Fig. S3A-B](#)). The main nucleus of the cell occupies the antapical pole and its last membrane is continuous with the plastid complex ([Supplementary Material Fig. S3A, C, E](#)). A small Golgi body, with many vesicles, is located behind the flagellar region ([Supplementary Material Fig. S3B-E](#)). A tubular mitochondrion extends mainly through the longitudinal direction of the cell, near the flagellar apparatus, nucleus and between the periplast and the plastid.

Two unequal flagella are inserted ventrally, almost in the center of the *H. aquamarina* cell ([Figs 1–2](#)). One flagellum is slightly larger than the other ([Fig. 1F](#)), and mastigonemes were found on both ([Fig. 2B, D-E](#)). Close to the flagellar insertion, a small gullet extends obliquely from the vestibulum

towards the antapical pole of the cell, and it is surrounded by large ejectosomes ([Fig. 1C-D, Supplementary Material Fig. S3B-E](#)). A furrow is absent and a contractile vacuole was not observed. Cells are free-swimming and very active. Frequently, when resting, they quickly start to rotate on their axis and suddenly go away (see [Supplementary Material Data 1 and 2](#)).

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

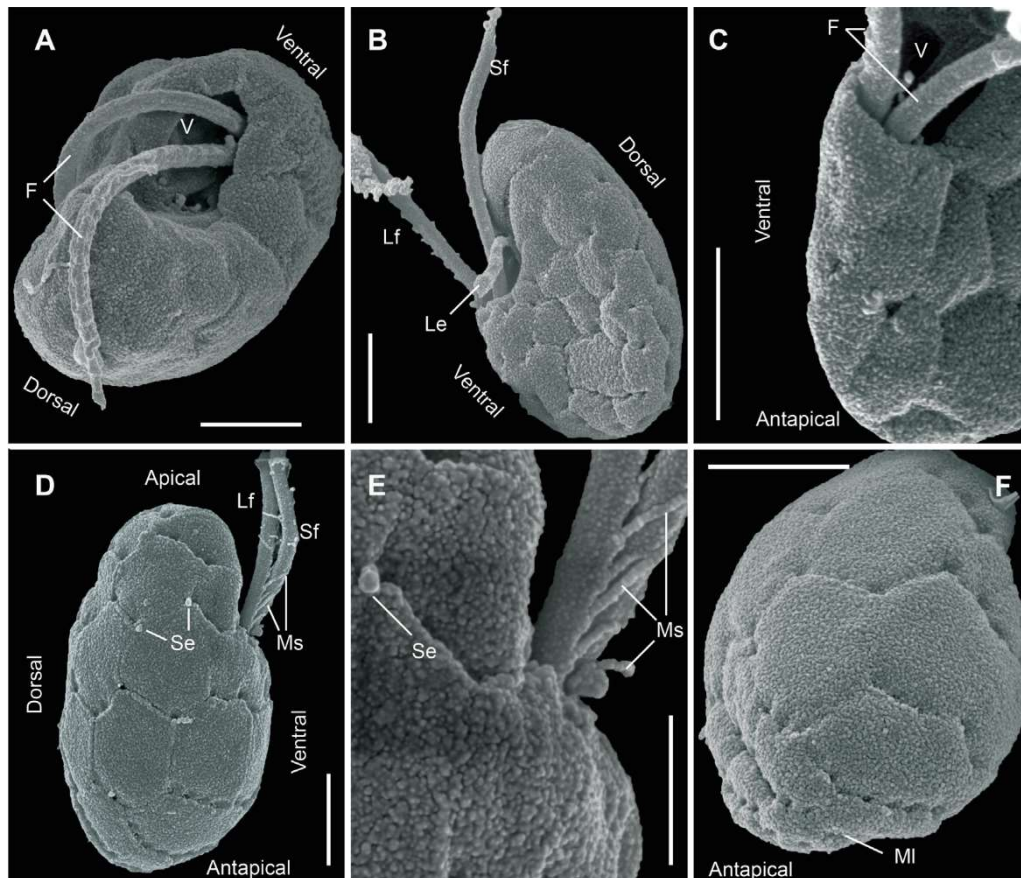


Figure 2. Morphology of *H. aquamarina* (BMAK265) by scanning electron microscopy. **(A)** Apical view showing the vestibular region and the flagella. **(B)** Lateral view showing the SPC hexagonal plates, flagella and large ejectosome discharged. **(C)** Ventro-lateral view of the vestibular region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

Phycobiliprotein Spectral Characterization

The PBP extracts of both *H. aquamarina* strains have very close spectral characteristics. Both pigment extracts are light purple ([Supplementary Material Fig. S2C](#)). Small variations of spectral signatures between the two strains are observed. The curve of BMAK265 shows the maximum absorption at 564 nm and a second peak at 616–620 nm. RCC4102 has the highest peak of absorption at 557–566 nm, and another peak at 616–619 nm ([Fig. 3](#)).

Phylogenetic Analysis

Molecular phylogeny inferences performed using Bayesian (BA) and maximum likelihood (ML) analysis indicate with maximum posterior probability (PP) and bootstrap (BS) support that the genus *Hemiselmis* is a distinct and monophyletic lineage, sister of the *Komma* / “*Chroomonas*” clade. Within Hemiselmidaceae, *H. cryptochromatica* is the early diverting branch, fully supported in concatenated (nucleomorph and nuclear SSU) and nuclear small ribosomal subunit (nSSU) phylogenies. *H. amylosa* follows *H. cryptochromatica* as the sister clade of the remaining species. Furthermore, all analyses support a sister relationship between *H. pacifica* and *H. virescens* ([Fig. 4](#), [Supplementary Material Fig. S1](#)).

The close relationship between *H. aquamarina*, *H. tepida*, *H. rufescens*, and *H. andersenii* is recovered in concatenated and nucleomorph small

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

H. tepida and *H. aquamarina* are pointed out as sister taxa in concatenated (0.86 PP) and nmSSU (0.94 PP, 70 BS) rRNA inferences. However, nSSU phylogeny suggests a close relationship between *H. andersenii* and *H. aquamarina*, although the support is too low to make any conclusion (below 0.75 PP and 70 BS). The *H. aquamarina* clade is fully supported by PP in all BA performed and with high BS in ML inferences ([Fig. 4](#), [Supplementary Material Fig. S1](#)). In all trees, the UTEX 2000 strain clusters into the *H. aquamarina* clade and, therefore, it is classified as *H. cf. aquamarina*.

Secondary Structures of Nuclear ITS2

The lengths of the ITS2 region are quite similar for all *Hemiselmis* strains and clones analyzed, with a mean length of 335 nt (SD $\pm$ 9 nt). The shortest ITS2 sequence is found in one clone of *H. aquamarina* (BMAK265, 327 nt), and the longest one in *H. cf. virescens* (RCC3575, 360 nt). Clones of the same strain display different lengths of the ITS2 sequence.

The predicted ITS2 secondary structures of *Hemiselmis* have four helices as reported for most

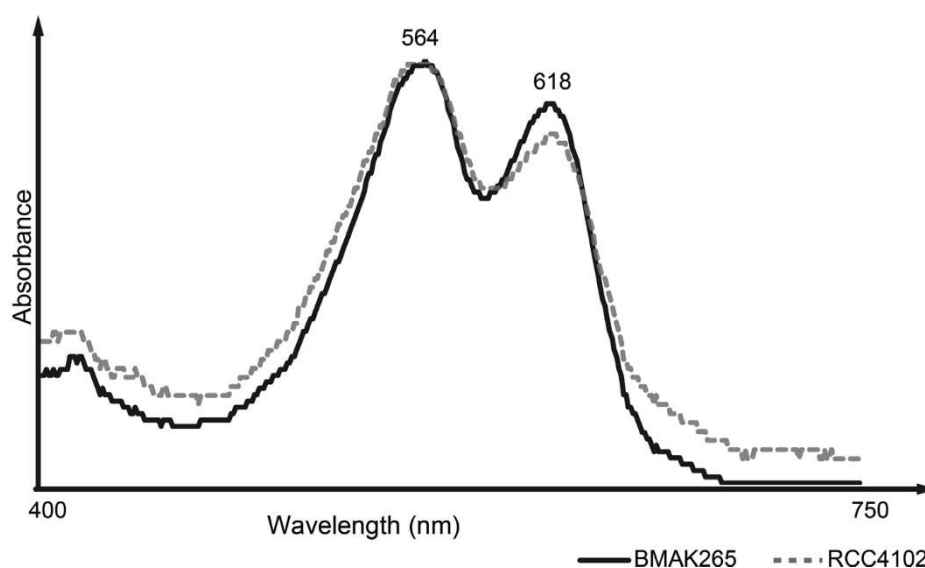


Figure 3. Visible absorption spectra of *H. aquamarina* strains phycobiliprotein extract (Cr-PC564).

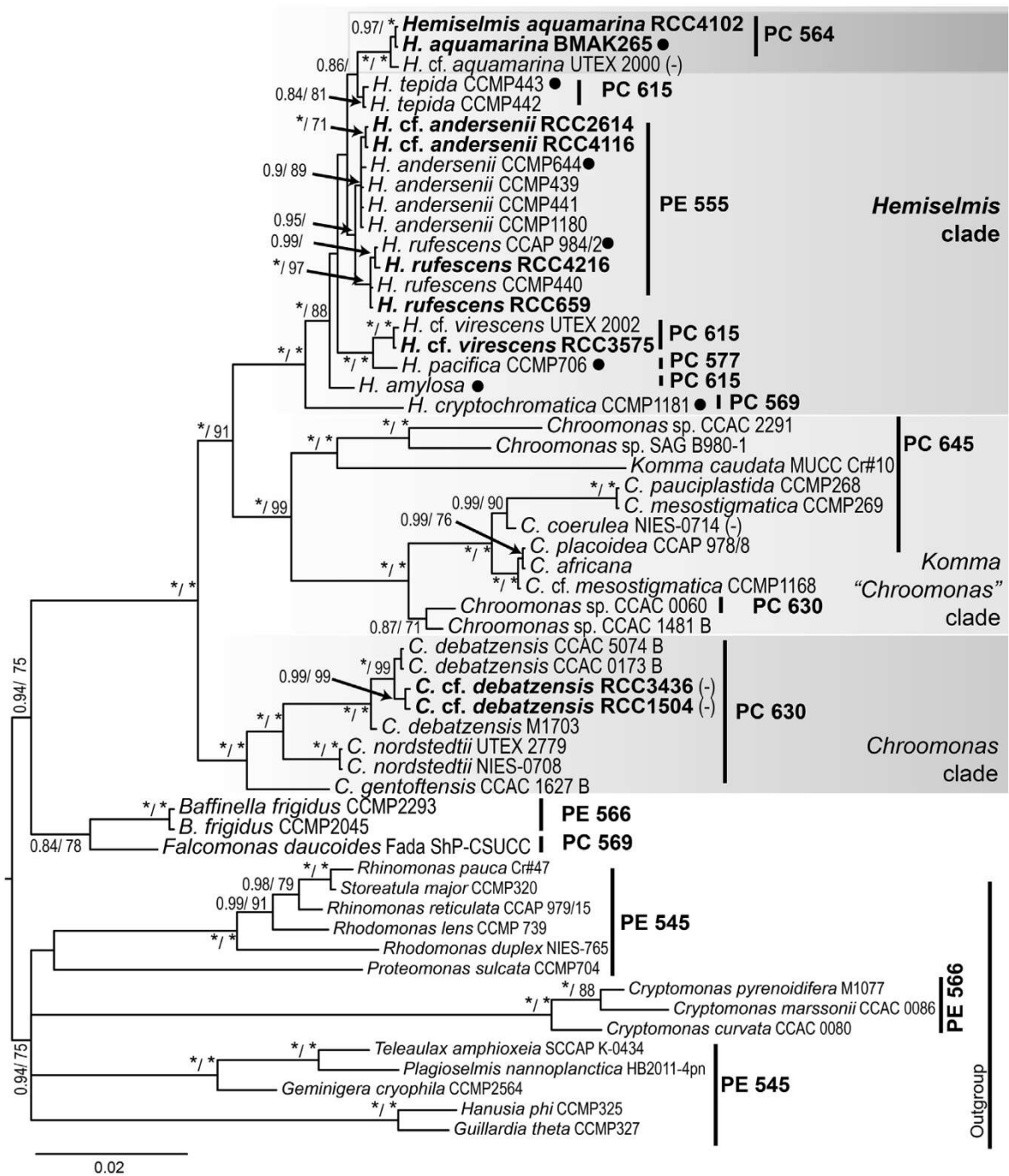


Figure 4. Bayesian tree of the concatenated nSSU and nmSSU rRNA. Phycobiliprotein type (PC- phycocyanin and PE- phycoerythrin) is indicated for each clade. (-) Absence of phycobiliprotein information. Bold names correspond to sequences obtained in this study. Bold circles represent cultures from which the species holotype was obtained according to literature information. Bayesian inference was performed applying the substitution models HKY + G + I and GTR + G + I for the nSSU and nmSSU, respectively. Maximum likelihood inferences were performed using GTR + CAT in a partitioned analysis. Nodes with * are fully supported by posterior probability or bootstrap. Supports below 0.75 posterior probability and 70% bootstrap are omitted. Scale bar indicates the rate of nucleotide substitution per site.

eukaryotes. ITS2 displays highly conserved single-stranded sequences between 5.8S and helix I, helices I-II and helices II-III (see [Supplementary Material Fig. S4](#)). However, other regions could not be well aligned interspecifically due to the high divergence of nucleotides. These regions are located at helices I (middle and apex portion), II (apex portion), III (middle portion) and IV.

Helix I of *H. aquamarina* and *H. cf. virescens* is branched in all acceptable predictions generated. The apex portion of helix I have variations of nucleotides between clones and strains of *H. aquamarina*. Helix II is conserved at the first eight nucleotides and has a uracil- uracil mismatch at the sixth position in all *Hemiselmis* species ([Fig. 5](#)). The apex of helix II shows deletions of three base pairs (pb) between clones of *H. aquamarina* in both strains ([Fig. 5](#), positions 15–17). Helix III, the longest one, shows highly conserved nucleotides between positions 31–48 ([Fig. 6](#), [Supplementary Material Fig. S4](#)). Helix IV is the most variable one and could not be accurately aligned interspecifically. The single-stranded sequences between helices III-IV and helix IV-LSU are quite dissimilar. Therefore, the consensus

motif indicating the termini of helix IV could not be predicted. For more details, see [Supplementary Material Data 4](#).

Structural comparison of ITS2 helices at conserved base pairs (bp) between *Hemiselmis* species shows many compensatory base changes (CBCs) and hemi-CBCs (h-CBCs) at the most conserved helices ([Figs 5–6](#)). *H. aquamarina* has a unique molecular signature of ITS2 secondary structure between species of the genus. Its strains and clones have no CBC across helices I, II and III. Therefore, *H. aquamarina* can also be distinguished from the other *Hemiselmis* species by CBCs.

Discussion

Many features observed by light and electron microscopy unveiled synapomorphies in Hemiselmidae, which are congruent with molecular phylogeny results, indicating that this family is indeed a natural grouping from evolutionary processes. All currently described *Hemiselmis* species have a lateral insertion of the flagella and hexagonal plates of SPC ([Butcher 1967](#); [Clay and Kugrens](#)

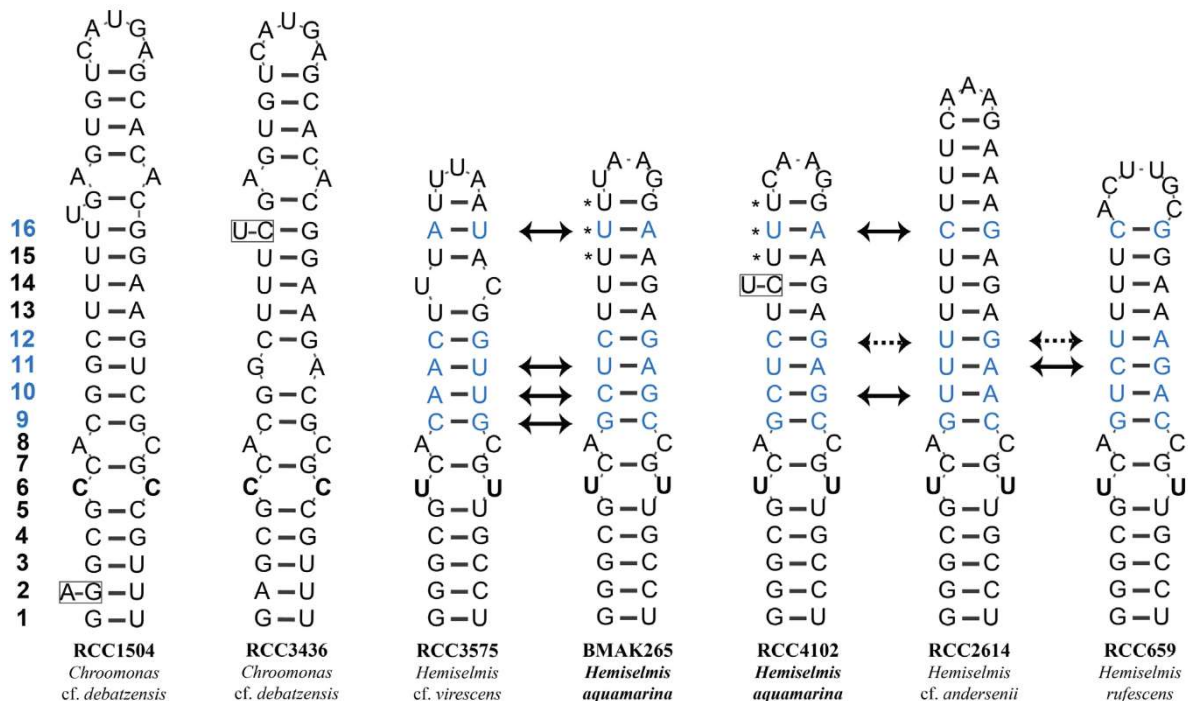


Figure 5. Helix II of the predicted secondary structure of ITS2. The conserved base pairs among different strains are numbered. CBCs and h-CBCs are in blue, emphasized by solid and dotted arrows, respectively. h-CBCs are indicated just in positions where CBCs occur. The pyrimidine- pyrimidine mismatches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

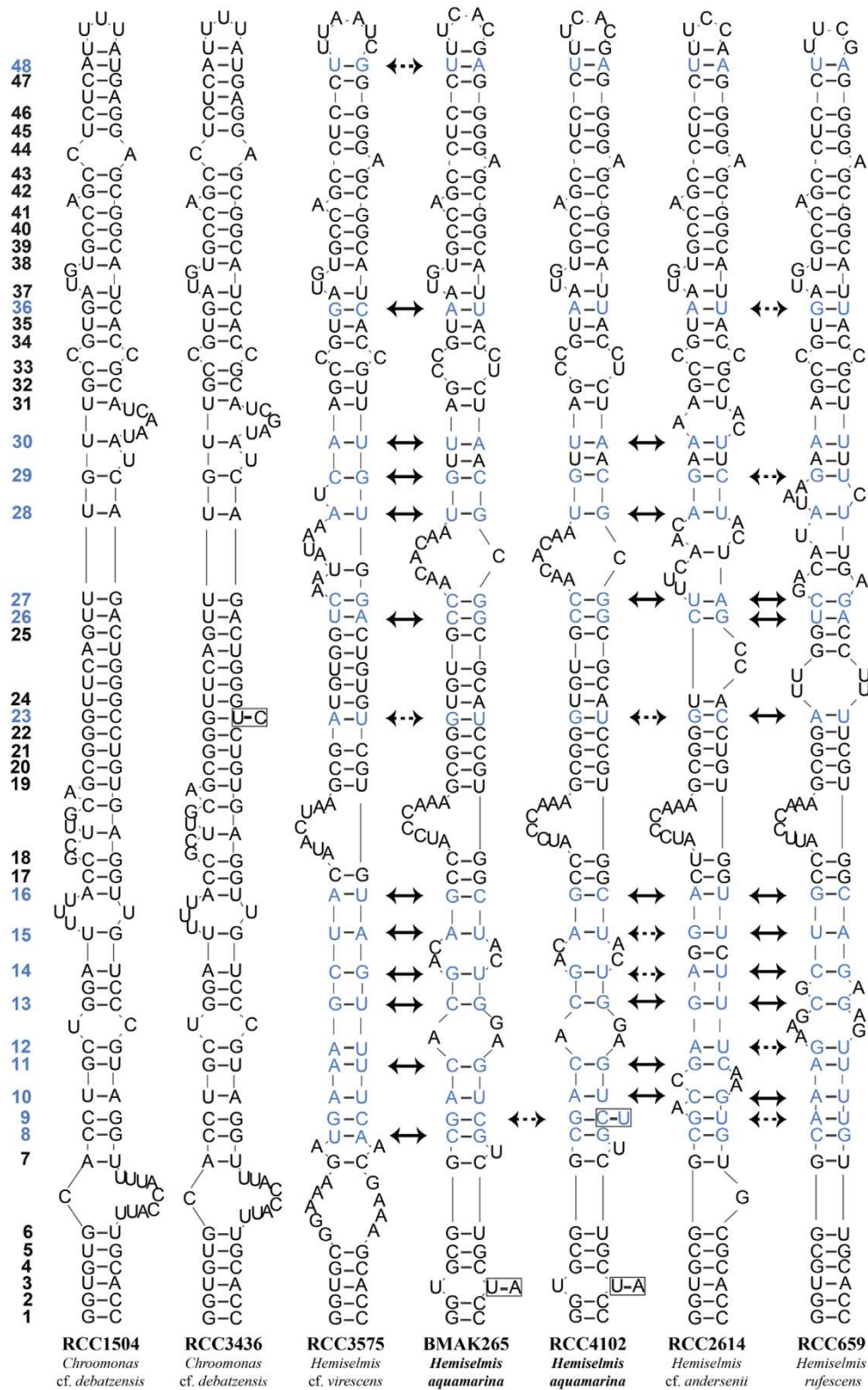


Figure 6. Helix III of the predicted secondary structure of ITS2. The conserved base pairs among different strains are numbered. CBCs and h-CBCs are in blue, emphasized by solid and dotted arrows, respectively. h-CBCs are indicated just in positions where CBCs occur. Nucleotides that differ between clones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

1999; Lane and Archibald 2008; Wetherbee et al. 1986). Before this work, four species of *Hemiselmis* (*H. amylosa*, *H. rufescens/brunnescens*, *H. simplex*, and *H. virescens*) have been investigated by transmission electron microscopy (Butcher 1967; Clay and Kugrens 1999; Lucas 1970; Santore and Greenwood 1977; Santore 1982). The arrangement and form of the organelles, such as the starch-coated pyrenoid, the nucleomorph, the Golgi body, the nucleus and the mitochondrion, are congruent between *Hemiselmis* species. Santore (1982) had difficulty separating *H. rufescens/H. brunnescens* and *H. virescens* using ultrastructural data. Conclusively, *Hemiselmis* is easily distinguishable by morphology from other genera of cryptophytes, but species identification is not clear (see Supplementary Material Table S2). Accordingly, *Hemiselmis aquamarina* identification must rely on molecular tools as reported for other species of cryptophytes (Lane and Archibald 2008; Hoef-Emden 2007, 2018).

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

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

Therefore, none of the previously described PBPs in cryptophytes has the same spectral signature of Cr-PC 564 from *H. aquamarina*, which has features common to Cr-PCs and Cr-PEs. These differences of absorption peaks in Cr-PC 564 are indicative of different bilin composition and/or linkage sites. Its bilin composition must differ from other Cr-PCs at the linkage sites β -Cys 158 and α -Cys 18, as the sites β -DiCys 50, 61 and β -Cys 82 are frequently linked to DHBV and PCB bilins, respectively (Overkamp et al. 2014; Wedemayer et al. 1996). Possibly, the β -Cys 158 position could be linked to bilin 584, as reported in Cr-PC 569 and Cr-PE

566, and α -Cys 18 position to PCB or bilin 618, as in Cr-PC 615 and Cr-PC 569, respectively. However, bilin composition alone does not determine the absorption spectra of a given PBP: the native protein environment of the bilins also contributes significantly to these properties (Glazer and Wedemayer 1995; Wemmer et al. 1993).

H. aquamarina is closely related to *H. tepida*, *H. andersenii* and *H. rufescens*, as previously indicated in literature by the placement of UTEX 2000 in phylogenetic trees (Cunningham et al. 2019; Hoef-Emden 2008, 2018). The accurate relationship between these species is not well resolved by phylogenetic inferences due to the medium/low support values of branches in all methods applied. This could be attributed to the low evolutionary rate of SSU rRNA. However, it seems that *H. tepida* and *H. andersenii* are closer to *H. aquamarina* than *H. rufescens*. All phylogenetic inferences indicate with high support that *Hemiselmis aquamarina* forms a new branch and can be regarded as a new species. For species identification proposes, nmSSU and ITS2 rRNA sequences are more divergent between *Hemiselmis* and, accordingly, more accurate for diagnosis.

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

CBCs in the ITS2 secondary structure are indicative of species separation due to the conserved pairing structure needed for rRNA processing (Coleman 2007, 2009; Müller et al. 2007). For example, in Volvocales, the absence of CBCs in ITS2 helices II and III is in agreement with sexual compatibility and, therefore, can predict mating affinity (i.e. biological species concept, Coleman 2000). Moreover, most of CBCs in these regions are non-homoplasious changes and can present molecular signatures, which detect unambiguously taxa and

clades (Caisová et al. 2011). Therefore, the presence of CBCs in these conserved regions between *Hemiselmis aquamarina* and other species of the genus indicates it as a distinct species.

Conclusion

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

Methods

Sampling, strain isolation and culturing conditions: The strain from the coast of Brazil was isolated in August 2011, close to the Anchieta Island, Ubatuba, São Paulo (23° 35.85'S, 45° 01.70' W). Water was collected at a depth of 40 m with a Nansen bottle. The water sample was enriched with Erd-Schreiber medium diluted 10 times. After a few days, a single cell was selected by micro-pipetting and carefully washed in a sterilized medium. The culture is maintained in Erd-Schreiber medium (Thronsdon 1997), 32–35 salinity, 20 °C temperature, a photoperiod of 12:12 L:D cycle at 50 µmol photons m⁻².s⁻¹. This strain is deposited in the Banco de Microrganismos Marinhos Aydar & Kutner (<http://www.io.usp.br/index.php/infraestrutura/banco-de-microrganismos.html>), as BMAK265, and in the Roscoff Culture Collection as RCC5634.

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

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

Phycobiliprotein extraction: Cultures were grown in 50 ml polycarbonate culture flasks for 2–3 weeks, following the conditions described above, and harvested by centrifugation (300 g, 8 min). The pellet was frozen and kept at –80 °C until processing. PBP extractions were performed for strains BMAK265, RCC4102, RCC4116, RCC4216 and RCC659 (Supplementary Material Table S1) following Hill and Rowan (1989). The absorption spectra of the pigment extract were determined using an Epoch 2 microplate spectrophotometer (BioTek Instruments, Inc., Winooski, VT, USA).

Morphological observations: Dense BMAK265 (2–3 weeks old) cultures were used for morphological investigations. For differential interference contrast (DIC) and phase contrast, living cells on glass slides sealed with coverslips were observed with a Leica DM

4000B (Leica Microsystems, Wentzler, Germany). Morphometric values from 40 live individuals were obtained from calibrated pictures. Natural fluorescence of the plastid was observed from fixed (2% glutaraldehyde) cells with a confocal microscope, Zeiss LSM 440 Axiovert 100 (Carl Zeiss, Jena, Germany) equipped with 543 nm laser and a 570 nm long-pass filter.

For electron microscopy observations, cells were harvested by gentle centrifugation (3 min, 100–150 g) and then immediately fixed for 90 min with a solution containing glutaraldehyde (2%), sodium cacodylate trihydrate (0.1 M) and sucrose (0.8 M), as described in Majaneva et al. (2014). The cells were washed using the latter solution (without glutaraldehyde) and post-fixed with osmium tetroxide (1 %) buffered in cacodylate trihydrate (0.1 M) for 60 min. Cells were then washed twice in cacodylate buffer (0.1 M). For Scanning Electron microscopy (SEM), a sample was dehydrated in a series of increasing ethanol concentrations (70, 90, 95 and 100 %). It was subsequently critical-point dried (Balzers CPD 030, Bal-Tec, Vaduz, Liechtenstein), gold-coated (Balzers SCD 050) and examined in a Zeiss Sigma VP. For Transmission Electron microscopy (TEM), cells were dehydrated in an acetone series (50, 70, 90, 95 and 100%), embedded in Spurr resin, thin sectioned and examined in a Philips CM120 TEM.

DNA extraction, PCR, cloning and sequencing: Material for molecular analyses was obtained as specified in the pigment extraction section. Genomic DNA was extracted using NucleoSpin® Plant II kit following the manufacturer's instructions. We performed PCRs of nSSU, nmSSU and ITS2 rRNA with Platinum® Taq DNA polymerase kit (Invitrogen™, Carlsbad, USA). Primers and cycling conditions are available in Supplementary Material Table S3. We analyzed the strains BMAK265 (RCC5634), RCC659, RCC1504, RCC2614, RCC3436, RCC3575, RCC4102, RCC4116 and RCC4216 (Supplementary Material Table S1). Since there might be multiple copies of the rRNA operon in the genome (Prokopowich et al. 2003; Thornhill et al. 2007), we cloned all sequences of BMAK265 and RCC4102 and the ones that intragenomic variation was observed using TOPO® TA Cloning® kit (Invitrogen, Carlsbad, CA, USA). To avoid polymerase errors in cloning, we used Phusion High-Fidelity PCR Master Mix (Thermo Fisher Scientific).

The PCR products yielding a single band of the expected size on an agarose gel (1%) were purified using the GFX Illustra kit (GE Healthcare Life Sciences, Little Chalfont, Buckinghamshire, UK) following the manufacturer's instructions. Sequencing reactions were performed with the Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems™, Hammondon, NJ, USA) and samples were sequenced with 3730 Applied Biosystems.

Generated contigs were searched by BLAST (<https://blast.ncbi.nlm.nih.gov/>) against NCBI sequences to ensure that each contig corresponded to the organisms we were investigating. We discarded contigs with high levels of ambiguity and noise. The consensus sequences were assembled using Geneious 9.7 (Biomatters, Auckland, New Zealand, <https://www.geneious.com/>) by comparison to a reference sequence obtained from BLAST. All sequences generated were deposited to Genbank (see Supplementary Material Table S1 for accession numbers).

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

the following analyses, we used the alignments with best-solved quartets.

We concatenated nSSU (1495 pb, 56 sequences) and nmSSU (1347 pb, 51 sequences) sequences using SeaView (Gouy et al. 2010), resulting in a dataset of 56 sequences with 2,842 bp (Supplementary Material Data 3). We applied the molecular evolution models HKY + G + I and GTR + G + I for the nSSU and nmSSU, respectively, in a partitioned Bayesian analysis in the concatenated alignment. Moreover, to determine if the topologies were congruent between nSSU and nmSSU, we performed separated phylogenetic inferences by Bayesian analysis as described below (Supplementary Material Fig. S1).

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

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

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

Secondary structure prediction of the nuclear ITS2 region:

All strains sequenced for nuclear ITS2 rRNA region, except RCC659 and RCC2614, were cloned due to intragenomic variation, which was observed in a first sequencing. Fourteen secondary structures of ITS2 rRNA were predicted from *H. aquamarina* (BMAK265 and RCC4102, seven clones), *H. cf. andersenii* (RCC2614), *H. rufescens* (RCC659), *H. cf. virescens* (RCC3575, one clone) and *Chroomonas cf. debatzensis* Hoef-Emden (RCC1504 and RCC3436, four clones). The complementary regions of 5.8S and LSU rRNA, ITS2 boundaries, were annotated using Hidden Markov Models with the default parameters of the ITS2 database annotation tool (<http://its2.bioapps.biozentrum.uni-wuerzburg.de/>, Ankenbrand et al. 2015). Automated secondary structures of ITS2 rRNA predictions of the entire sequence and, when necessary, of single helices, were acquired by online web services using the default folding options of the Mfold (Zuker 2003) and RNAstructure (Reuter and Mathews 2010). Several alternative secondary structures were predicted per sequence. Structure choices were based on the conserved hallmarks of ITS2 secondary structures (see Coleman 2000, 2007, 2009) and comparisons with previously published structures of cryptophytes. The ITS2 sequences were aligned with MAFFT and G-INS-I algorithm (Katoh et al. 2002). For each sequence in the alignment, a preliminary secondary structure was annotated in Vienna file format, imported to 4SALE (Seibel et al. 2008) and manually edited by a comparative analysis of each nucleotide position (Supplementary Material Data 4). The consensus secondary structure of *H. aquamarina* was generated in 4SALE using the default option (Supplementary Material Fig. S4). Due to the variability of the pairing positions in helices I and IV of ITS2 between *Hemiselmis* sequences, we decided to adopt the CBC clade-based concept sensu Coleman (2000). Accordingly,

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

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

CRedit authorship contribution statement

Karoline Magalhães: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data curation, Writing - original draft. **Adriana Lopes Santos:** Methodology, Data curation, Validation, Writing - review & editing. **Daniel Vaultot:** Validation, Resources, Writing - review & editing, Funding acquisition. **Mariana Cabral Oliveira:** Resources, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Acknowledgements

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

Funding

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

Declaration of Competing Interest

The authors declare that they have no conflict of interest.

Appendix A. Supplementary Material

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.

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

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

References

- Ankenbrand MJ, Keller A, Wolf M, Schultz J, Förster F** (2015) ITS2 database V: Twice as much. *Mol Biol Evol* **32**:3030–3032
- Butcher RW** (1967) An Introductory Account of the Smaller Algae of British Coastal Waters: Cryptophyceae. Fish Invest Ser IV, Her Majesty's Stationary Office; London, 54 p
- Caisová L, Marin B, Melkonian M** (2011) A close-up view on ITS2 evolution and speciation - a case study in the Ulvophyceae (Chlorophyta, Viridiplantae). *BMC Evol Biol* **11**:262
- Clay BL, Kugrens P** (1999) Characterization of *Hemiselmis amylosa* sp. nov. and phylogenetic placement of the blue-green cryptomonads *H. amylosa* and *Falcomonas daucooides*. *Protist* **150**:297–310
- Coleman AW** (2000) The significance of a coincidence between evolutionary landmarks found in mating affinity and a DNA sequence. *Protist* **151**:1–9
- Coleman AW** (2007) Pan-eukaryote ITS2 homologies revealed by RNA secondary structure. *Nucleic Acids Res* **35**:3322–3329
- Coleman AW** (2009) Is there a molecular key to the level of “biological species” in eukaryotes? A DNA guide. *Mol Phylogenet Evol* **50**:197–203
- Cunningham BR, Greenwold MJ, Lachenmyer EM, Heidenreich KM, Davis AC, Dudycha JL, Richardson TL** (2019) Light capture and pigment diversity in marine and freshwater cryptophytes. *J Phycol* **55**:552–564
- Darriba D, Taboada GL, Doallo R, Posada D** (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* **9**:772–772.
- Deane JA, Strachan IM, Saunders GW, Hill DRA, McFadden GI** (2002) Cryptomonad evolution: nuclear 18S rDNA phylogeny versus cell morphology and pigmentation. *J Phycol* **38**:1236–1244
- Douglas SE, Penny SL** (1999) The plastid genome of the cryptophyte alga, *Guillardia theta*: complete sequence and conserved syntenic groups confirm its common ancestry with red algae. *J Mol Evol* **48**:236–244
- Douglas S, Zauner S, Fraunholz M, Beaton M, Penny S, Deng LT, Wu X, Reith M, Cavalier-Smith T, Maier UG** (2001) The highly reduced genome of an enslaved algal nucleus. *Nature* **410**:1091–1096
- Droop MR** (1955) Some new supra-littoral Protista. *J Mar Biol Ass UK* **34**:233–245
- Edgar RC** (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**:1792–1797
- Edler D, Klein J, Antonelli A, Silvestro D** (2021) raxmlGUI 2.0: A graphical interface and toolkit for phylogenetic analyses using RAxML. *Matschiner M* (ed) *Methods Ecol Evol* **12**:373–377
- Glazer AN, Wedemayer GJ** (1995) Cryptomonad biliproteins- an evolutionary perspective. *Photosynth Res* **46**:93–105
- Gouy M, Guindon S, Gascuel O** (2010) SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* **27**:221–224
- Greenwold MJ, Cunningham BR, Lachenmyer EM, Pullman JM, Richardson TL, Dudycha JL** (2019) Diversification of light capture ability was accompanied by the evolution of phycobiliproteins in cryptophyte algae. *Proc R Soc B Biol Sci* **286**:20190655
- Hill DRA, Rowan KS** (1989) The biliproteins of the Cryptophyceae. *Phycologia* **28**:455–463
- Hoef-Emden K** (2007) Revision of the genus *Cryptomonas* (Cryptophyceae) II: Incongruences between the classical morphospecies concept and molecular phylogeny in smaller pyrenoid-less cells. *Phycologia* **46**:402–428
- Hoef-Emden K** (2008) Molecular phylogeny of phycocyanin-containing cryptophytes: evolution of biliproteins and geographical distribution. *J Phycol* **44**:985–993
- Hoef-Emden K** (2018) Revision of the Genus *Chroomonas* HANSGIRG: the benefits of DNA-containing specimens. *Protist* **169**:662–681
- Hoef-Emden K, Marin B, Melkonian M** (2002) Nuclear and nucleomorph SSU rDNA phylogeny in the Cryptophyta and the evolution of cryptophyte diversity. *J Mol Evol* **55**:161–179
- Katoh K, Misawa K, Kuma KI, Miyata T** (2002) MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**:3059–3066
- Keeling PJ** (2010) The endosymbiotic origin, diversification and fate of plastids. *Philos Trans R Soc Lond B Biol Sci* **365**:729–748
- Klaveness D** (1985) Classical and modern criteria for determining species of Cryptophyceae. *Bull Plankt Soc Japan* **32**:111–123
- Lane CE, Archibald JM** (2008) New marine members of the genus *Hemiselmis* (Cryptomonadales, Cryptophyceae). *J Phycol* **44**:439–450
- Larsson A** (2014) AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30**:3276–3278

- Lucas I** (1970) Observations on the ultrastructure of representatives of the genera *Hemiselmis* and *Chroomonas* (Cryptophyceae). *Br Phycol J* **5**:29–37
- Majaneva M, Remonen I, Rintala JM, Belevich I, Kremp A, Setälä O, Jokitalo E, Blomster J** (2014) *Rhinomonas nottbecki* n. sp. (Cryptomonadales) and molecular phylogeny of the family Pyrenomonadaceae. *J Eukaryot Microbiol* **61**:480–492
- Marin B, Klingberg M, Melkonian M** (1998) Phylogenetic relationships among the Cryptophyta: analyses of nuclear-encoded SSU rRNA sequences support the monophyly of extant plastid-containing lineages. *Protist* **149**:265–276
- Müller T, Philippi N, Dandekar T, Schultz J, Wolf M** (2007) Distinguishing species. *RNA* **13**:1469–1472
- Overkamp KE, Langklotz S, Aras M, Helling S, Marcus K, Bandow JE, Hoef-Emden K, Frankenberg-Dinkel N** (2014) Chromophore composition of the phycobiliprotein Cr-PC577 from the cryptophyte *Hemiselmis pacifica*. *Photosynth Res* **122**:293–304
- Parke M** (1949) Studies on marine flagellates. *J Mar Biol Ass UK* **28**:255–286
- Prokopowich CD, Gregory TR, Crease TJ** (2003) The correlation between rDNA copy number and genome size in eukaryotes. *Genome* **46**:48–50
- Rambaut A, Suchard M, Xie D, Drummond A** (2014) Tracer v1.6.
- Reuter JS, Mathews DH** (2010) RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* **11**:129
- Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard M, Huelsenbeck JP** (2012) Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* **61**:539–542
- Santore UJ** (1982) The ultrastructure of *Hemiselmis brunnescens* and *Hemiselmis virescens* with additional observations on *Hemiselmis rufescens* and comments on the Hemiselmidaceae as a natural group of the Cryptophyceae. *Br Phycol J* **17**:81–99
- Santore UJ, Greenwood AD** (1977) The mitochondrial complex in Cryptophyceae. *Arch Microbiol* **112**:207–218
- Schmidt HA, Petzold E, Vingron M, Von Haeseler A** (2003) Molecular phylogenetics: parallelized parameter estimation and quartet puzzling. *J Parallel Distrib Comput* **63**:719–727
- Seibel PN, Müller T, Dandekar T, Wolf M** (2008) Synchronous visual analysis and editing of RNA sequence and secondary structure alignments using 4SALE. *BMC Res Notes* **1**:91
- Spear-Bernstein L, Miller KR** (1989) Unique location of the phycobiliprotein light-harvesting pigment in the Cryptophyceae. *J Phycol* **25**:412–419
- Thornhill DJ, Lajeunesse TC, Santos SR** (2007) Measuring rDNA diversity in eukaryotic microbial systems: how intragenomic variation, pseudogenes, and PCR artifacts confound biodiversity estimates. *Mol Ecol* **16**:5326–5340
- Throndsen J** (1997) The Planktonic Marine Flagellates. In Thomas C (ed) *Identifying Marine Phytoplankton*. Academic Press: San Diego, pp 591–730
- Wedemayer GJ, Kidd DG, Glazer AN** (1996) Cryptomonad biliproteins: Bilin types and locations. *Photosynth Res* **48**:163–170
- Wedemayer GJ, Wemmer DE, Glazer AN** (1991) Phycobilins of cryptophycean algae. Structures of novel bilins with acryloyl substituents from phycoerythrin 566. *J Biol Chem* **266**:4731–4741
- Wemmer DE, Wedemayer GJ, Glazer AN** (1993) Phycobilins of cryptophycean algae. Novel linkage of dihydrobiliverdin in a phycoerythrin 555 and a phycocyanin 645. *J Biol Chem* **268**:1658–1669
- Wetherbee R, Hill DRA, McFadden GI** (1986) Periplast structure of the cryptomonad flagellate *Hemiselmis brunnescens*. *Protoplasma* **131**:11–22
- Wiens JJ** (2006) Missing data and the design of phylogenetic analyses. *J Biomed Inform* **39**:34–42
- Zuker M** (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**:3406–3415