



HAL
open science

Revision of the Genus *Micromonas* Manton et Parke (Chlorophyta, Mamiellophyceae), of the Type Species *M. pusilla* (Butcher) Manton & Parke and of the Species *M. commoda* van Baren, Bachy and Worden and Description of Two New Species Based on the Genetic and Phenotypic Characterization of Cultured Isolates

Nathalie Simon, Elodie Foulon, Daphné Grulois, Christophe Six, Yves Desdevises, Marie Latimier, Florence Le Gall, Margot Tragin, Aude Houdan, Evelyne Derelle, et al.

► **To cite this version:**

Nathalie Simon, Elodie Foulon, Daphné Grulois, Christophe Six, Yves Desdevises, et al.. Revision of the Genus *Micromonas* Manton et Parke (Chlorophyta, Mamiellophyceae), of the Type Species *M. pusilla* (Butcher) Manton & Parke and of the Species *M. commoda* van Baren, Bachy and Worden and Description of Two New Species Based on the Genetic and Phenotypic Characterization of Cultured Isolates. *Protist*, 2017, 168 (5), pp.612-635. 10.1016/j.protis.2017.09.002 . hal-01614739

HAL Id: hal-01614739

<https://hal.sorbonne-universite.fr/hal-01614739>

Submitted on 11 Oct 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted Manuscript



Title: Revision of the Genus *Micromonas* Manton et Parke (Chlorophyta, Mamiellophyceae), of the Type Species *M. pusilla* (Butcher) Manton and Parke and of the Species *M. commoda* van Baren, Bachy and Worden and Description of Two New Species Based on the Genetic and Phenotypic Characterization of Cultured Isolates<!--<RunningTitle>The Genus *Micromonas* (Chlorophyta, Mamiellophyceae) Revisited</RunningTitle>-->

Authors: Nathalie Simon, Elodie Foulon, Daphné Grulois, Christophe Six, Yves Desdevises, Marie Latimier, Florence Le Gall, Margot Tragin, Aude Houdan, Evelyne Derelle, Fabien Jouenne, Dominique Marie, Sophie Le Panse, Daniel Vaultot, Birger Marin

PII: S1434-4610(17)30078-0
DOI: <http://dx.doi.org/10.1016/j.protis.2017.09.002>
Reference: PROTIS 25593

To appear in:

Received date: 1-9-2016
Accepted date: 6-9-2017

Please cite this article as: Simon, Nathalie, Foulon, Elodie, Grulois, Daphné, Six, Christophe, Desdevises, Yves, Latimier, Marie, Le Gall, Florence, Tragin, Margot, Houdan, Aude, Derelle, Evelyne, Jouenne, Fabien, Marie, Dominique, Panse, Sophie Le, Vaultot, Daniel, Marin, Birger, Revision of the Genus *Micromonas* Manton et Parke (Chlorophyta, Mamiellophyceae), of the Type Species *M. pusilla* (Butcher) Manton and Parke and of the Species *M. commoda* van Baren, Bachy and Worden and Description of Two New Species Based on the Genetic and Phenotypic Characterization of Cultured Isolates. *Protist* <http://dx.doi.org/10.1016/j.protis.2017.09.002>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ORIGINAL PAPER

Revision of the Genus *Micromonas* Manton et Parke (Chlorophyta, Mamiellophyceae), of the Type Species *M. pusilla* (Butcher) Manton & Parke and of the Species *M. commoda* van Baren, Bachy and Worden and Description of Two New Species Based on the Genetic and Phenotypic Characterization of Cultured Isolates

Running title: The Genus *Micromonas* (Chlorophyta, Mamiellophyceae) Revisited

Nathalie Simon^{a,1}, Elodie Foulon^a, Daphné Grulois^a, Christophe Six^a, Yves Desdevises^b, Marie Latimier^a, Florence Le Gall^a, Margot Tragin^a, Aude Houdan^a, Evelyne Derelle^b, Fabien Jouenne^a, Dominique Marie^a, Sophie Le Panse^c, Daniel Vaultot^a, and Birger Marin^d

^aSorbonne Universités, Université Pierre et Marie Curie - Paris 06 and Centre National de la recherche Scientifique (CNRS), UMR 7144, Laboratoire Adaptation et Diversité en Milieu Marin, Station Biologique de Roscoff, Place Georges Teissier, 29680 Roscoff, France

^bSorbonne Universités, Université Pierre et Marie Curie - Paris 06 and Centre National de la recherche Scientifique (CNRS), UMR 7232, BIOM, Observatoire Océanologique, 66650 Banyuls/Mer, France

^cSorbonne Universités, Université Pierre et Marie Curie - Paris 06 and Centre National de la recherche Scientifique (CNRS), FR2424, Imaging Core Facility, Station Biologique de Roscoff, Place Georges Teissier, 29680 Roscoff, France

^dBotanisches Institut, Biozentrum Köln, Universität zu Köln, Zülpicher Str. 47b, 50674 Köln, Germany

Submitted September 1, 2016; Accepted September 6, 2017

Monitoring Editor: B. S. C. Leadbeater

Corresponding author; fax +33 2 98 29 23 24

e-mail simon@sb-roscoff.fr (N. Simon).

The green picoalgal genus *Micromonas* is broadly distributed in estuaries, coastal marine habitats and open oceans, from the equator to the poles. Phylogenetic, ecological and genomic analyses of culture strains and natural populations have suggested that this cosmopolitan genus is composed of several cryptic species corresponding to genetic lineages. We performed a detailed analysis of variations in morphology, pigment content, and sequences of the nuclear-encoded small-subunit rRNA gene and the second internal transcribed spacer (ITS2) from strains isolated worldwide. A new morphological feature of the genus, the presence of tip hairs at the extremity of the hair point, was discovered and subtle differences in hair point length were detected between clades. Clear non-homoplasious synapomorphies were identified in the small-subunit rRNA gene and ITS2 spacer sequences of five genetic lineages. These findings lead us to provide emended descriptions of the genus *Micromonas*, of the type species *M. pusilla*, and of the recently described species *M. commoda*, as well as to describe 2 new species, *M. bravo* and *M. polaris*. By clarifying the status of the genetic lineages identified within *Micromonas*, these formal descriptions will facilitate further interpretations of large-scale analyses investigating ecological trends in time and space for this widespread picoplankter.

Keywords: *Micromonas*; Mamiellophyceae; Chlorophyta; green algae; ITS2; molecular signature.

Introduction

Micromonas pusilla (Butcher) Manton & Parke, a motile marine microalga of very small size (1-3 μm), was first described by Butcher (1952) as *Chromulina pusilla*, based on material from the Conway estuary (North Wales) and initially classified, using light microscopy, as a member of the Chrysophyceae. This species was also identified as a dominant member of the ultraplankton and probably the most abundant organism on the British Islands list by Knight-Jones and Walne (1951). Ultrastructural and biochemical characteristics of the original isolate as well as of other strains originating from the English Channel, led Manton (1959) and Manton and Parke (1960) to classify *M. pusilla* within the green algae (Chlorophyceae). This species was further classified within the Prasinophyceae Christensen based on analogies between its light harvesting complexes and those of *Mamiella* Moestrup and *Mantoniella* Desikachary (Fawley et al. 1990). Phylogenetic analyses confirmed the affiliation of *Micromonas* within the order Mamiellales, sometimes termed prasinophyte clade II (Nakayama et al. 1998; Fawley et al. 2000; Guillou et al. 2004), that was raised to class status (Mamiellophyceae) by Marin and Melkonian (2010). *Micromonas* is also the 'type' of a previously described class, the Micromonadophyceae (Mattox and Stewart 1984),

introduced to replace the name Prasinophyceae by excluding *Tetraselmis*. The class Micromonadophyceae was declared invalid by Marin and Melkonian (2010).

In the diagnosis by Manton and Parke (1960), based on a neo-type culture isolated off Plymouth in the English Channel, *M. pusilla* is described as a pear-shaped naked cell 1-3 μm long and 0.7-1 μm broad, with a single mitochondrion, nucleus, Golgi body and chloroplast. The single flagellum is laterally attached and includes a 1 μm long basal part (the flagellum proper) and a slender hair-point (ca 3 μm long according to Manton and Parke 1960). In addition to these characteristics, a distinctive swimming behaviour (Manton and Parke 1960) allows identification using light microscopy. The pigment suite of *Micromonas* is typical of members of the Mamiellales (Mamiellophyceae, see above) (Latasa et al. 2004). A pigment named Chl $\text{C}_{\text{CS-170}}$, first detected in the tropical *Micromonas* strain CS-170 by Jeffrey (1989), has been reported to occur in other *Micromonas* strains as well as in strains of other green algal genera (such as *Ostreococcus* and *Prasinococcus* respectively members of the Mamiellophyceae and Palmophyllaceae) isolated from the deep sea (Latasa et al. 2004). The life cycle of *Micromonas* has not yet been elucidated, but a palmelloid phase with cells 2.5-5 μm long was reported in the original descriptions (Butcher 1952; Manton and Parke 1960) but apparently not observed since. The presence in the genome sequence of *Micromonas* isolates of meiosis-related genes, low GC regions with features of sex chromosomes, and genes coding for cell wall components suggest that sexual differentiation and formation of a resistant life-cycle stage may occur (Worden et al. 2009) as in other Chlorophyta (e.g. some members of the Pyramimonadales, Nephroselmidophyceae and Chlorophyceae (Graham et al. 2009; Leliaert et al. 2012).

Micromonas has a worldwide distribution (Thomsen and Buck 1998) and is of major ecological importance in temperate coastal waters (Not et al. 2004, 2005; Thronsen and Kristiansen 1988) as well as polar oceanic waters (Balzano et al. 2012b; Lovejoy et al. 2007; Thronsen and Kristiansen 1991). Evidence of phagotrophy has been recently reported for an arctic strain of *Micromonas*, suggesting that in addition to contributing significantly to primary production, this genus might have an impact on prokaryotic populations (McKie-Krisberg and Sanders 2014).

Several studies based on phylogenetic analyses of different genetic markers from culture isolates collected worldwide have distinguished three to seven genetic clades and suggested the existence of cryptic species within *Micromonas* (Guillou et al. 2004; Lovejoy et al. 2007; Slapeta et al. 2006; van Baren et al. 2016; Worden 2006; Worden et al. 2009; Wu et al. 2014, Table 1). Marin and Melkonian (2010) even suggested that some of these clades should be raised to genus status as they are genetically as different from the neotype culture of *Micromonas pusilla* as they are from *Mantoniella*. Studies of clade distributions using culture approaches or phylogenetic probes also suggested that genetic lineages within *Micromonas* occupy different ecological niches (Foulon et al. 2008; Lovejoy et al. 2007) and interact with specific viral populations (Baudoux et al. 2015). Comparison of the genome sequences of CCMP1545 which derives from the neo-type culture of *M. pusilla*, and RCC299 that belongs to a different clade, also suggests ecological differentiation through selection and acquisition processes that lead to different repertoires of genes in these two strains (van Baren et al. 2016; Worden et al. 2009). These differences, associated with extensive genomic divergence and rearrangements, led van Baren et al. (2016) to propose the description of a new *Micromonas* species, *M. commoda*.

In order to further clarify the status of the main genetic lineages identified within the last 15 years within the genus *Micromonas*, we conducted a detailed analysis of the morphology, pigment content, as well as small subunit rRNA (SSU rRNA) gene and second internal transcribed spacer (ITS2) sequences of individual strains isolated worldwide. Our findings lead us to provide a revised description of the genus *Micromonas*, of the type species *M. pusilla*, and of the species *M. commoda*, as well as to describe 2 new species.

Results and Discussion

Clear Molecular Signatures Distinguish Deeply Diverging Clades as Well as Sub-clades in the Genus *Micromonas*

All previously published phylogenies (among which the multigene analyses by Slapeta et al. 2006) and genomic analyses of *Micromonas* strongly suggest that this genus comprises a genetically diverse complex of cryptic species or clades that have been attributed different codes (Guillou et al. 2004; Lovejoy et al. 2007; Slapeta et al. 2006; van Baren et al. 2016; Worden et al. 2009; Wu et al. 2014; Table 1).

In order to better assess the genetic divergences within and between clades, as well as to identify synapomorphic signatures for the most highly supported clades and further characterize species within this genus, we produced near full length SSU rRNA gene and ITS2 sequences for 13 new isolates and retrieved published sequences (from isolates or environmental clone libraries) that were assigned to *Micromonas* (Table 2). Signatures in these two markers are indeed now commonly used as diagnostic characters of the Mamiellophyceae (Marin and Melkonian 2010; Subirana et al. 2013).

The phylogenetic analyses of the SSU-rDNA sequences (corresponding to 42 unique isolates of *Micromonas* and 26 environmental sequences) allowed us to recover the major deeply diverging lineages A.ABC.12, B.E.3, B._.4 and C.D.5 distinguished in previous studies and labelled using names that combine identifiers used by Guillou et al. (2004), Slapeta et al. (2006) and Worden (2006) (Guillou Clade.Slapeta Clade(s).Worden Clade(s)) as in Worden (2006)(Table 1, Fig. 1). None of our new isolates fell into clade B._.4 identified by Worden (2006) and composed solely of environmental sequences. An additional rather deeply diverging clade, already distinguished as an “unknown clade” in Wu et al. (2014), included environmental sequences retrieved from coastal surface waters (<10m) of the South China Sea (Wu et al. 2014) and Red Sea (Acosta et al. 2013), as well as a sequence retrieved from strain RCC1109, isolated from a Mediterranean lagoon.

Within some of the lineages described above, sub-clades identified in previously published phylogenies were recovered in our SSU rDNA and ITS2 phylogenies (Figs 1, 2). Within clade B.E.3, the arctic *Micromonas* sequences (obtained exclusively from polar environmental sequences and arctic isolates) clustered apart within an arctic sub-clade (clade Ea) as established by Lovejoy et al. (2007) (Figs 1, 2) while other sequences grouped together in a non-arctic clade as in Simmons et al. (2015). Sub-clades A.A.2, A.B.1 and A.C.1 (corresponding to clades A, B and C of Slapeta et al. 2006), although rather weakly supported, were recovered when ITS2 sequences were included in analyses (Fig. 2) but A.B.1 and A.C.1 were paraphyletic in SSU rRNA gene phylogenies (Fig. 1). It is also important to note that a recombination event in a highly conserved protein-coding marker (β -tubulin gene) was suspected between sub-clades A.A.2 and A.B.1 (Slapeta et al. 2006).

Within these clades, genetic divergences calculated for the highly conserved SSU rDNA, did not exceed 1.1 % while between-clade divergence was as high as 3.5 % (between the non-arctic B.E.3 and A.ABC.12 clades, Table 3). Between-clade divergences were in the same range as those calculated between *Mantoniella* and *Micromonas* species (1.5 to 3.2 %, Table 3).

A careful synapomorphy search using the SSU rDNA alignment that included all sequences of *Micromonas* from Table 2, sequences of other Mamiellophyceae taxa, as well as a large number (> 2000) of Viridiplantae sequences, allowed us to identify unique molecular signatures for the entire genus *Micromonas* and for three *Micromonas* clades, i.e. the complete clade B sensu Guillou et al. (2004) (B.E.3, Ea unknown clade and B._.4), sub-clade B._.4, and C.D.5 (Fig. 3, Table 4), while within these clades, only SNPs were encountered. These synapomorphies, which were unique (no homoplasies) within the class Mamiellophyceae, were designated here as clade-specific signatures. However these signatures showed parallel changes (homoplasies) for various distantly related green algae (Table 4). SSU rDNA signatures were mapped upon the secondary structure of the SSU rRNA molecule, and were identified as compensating base pair changes (CBCs) in intramolecular rRNA helices, as illustrated in Figure 3.

To substantiate the remaining *Micromonas* clades with molecular signatures it was necessary to investigate the more variable ITS2 marker. Several studies have shown the utility of ITS2 sequences (second internal transcribed spacer, separating the 5.8S and 28S rRNA genes) in addressing species level phylogenies (Kawasaki et al. 2015; Nakada et al. 2010; Subirana et al. 2013), and the presence of CBCs in conserved helices has been correlated with the inability of the respective organisms to sexually mate (Coleman 2000, 2007, 2009; Müller et al. 2007; but see Caisovà et al. 2011). ITS2 RNA transcripts of *Micromonas* strains displayed a highly conserved intramolecular folding pattern (secondary structure) with four universal helices separated by single-stranded linkers, as already known for the sister genus *Mantoniella* and other Mamiellophyceae (Marin and Melkonian 2010; Subirana et al. 2013 and Fig 4). Comparisons of each helix at the secondary structure level among *Micromonas* strains revealed homologous base pair positions across taxa, and revealed their evolution via CBCs and single-sided hemi-CBCs in full detail. All evolutionary steps were then precisely mapped upon branches of the phylogenetic tree of *Micromonas* clades, which distinguished between unique synapomorphies (= non-homoplasious within *Micromonas*) as well as homoplasious changes (parallelisms, reversals and convergences; Fig. 4). As a result, four *Micromonas* clades, i.e. arctic E_a, non-arctic B.E.3, A.ABC.12 and C.D.5, gained support by unique synapomorphic signatures in ITS2 helices (Fig. 4 and Table 4). No signature, neither in the SSU rDNA nor in the ITS2 sequences, was recovered for the sub-clades, A.A.2, A.B.1 and A.C.1 and for the unknown clade distinguished in Wu et al. (2014) (for sub-clade B._4, no ITS2 sequence is available). It should be noted that *M. commoda*, which was recently (van Baren et al. 2016) erected upon subclades A.A.1 plus A.B.2, was only supported by a single hemi-CBC in ITS2 (bp 10 of Helix 2; Fig. 4B).

The presence of molecular signatures in the ITS2 and/or SSU rDNA of clades A.ABC.12, non-arctic B.E.3, arctic E_a, B._4 and C.D.5 strongly supports the hypothesis that these clades represent distinct species. Some of these synapomorphies, which showed no homoplasies within *Micromonas*, have been included in the taxonomic diagnoses, in order to provide an unambiguous characterization of the whole genus *Micromonas* and of four species (18S rDNA and/or ITS2).

New Morphological Features for the Genus *Micromonas* and Infrageneric Morphometric Variations

The high genetic divergence recorded between lineages within the genus *Micromonas* probably corresponds to diversifications that occurred millions of years ago, in the Late Cretaceous for the deepest divergence according to Slapeta et al. (2006). This genetic divergence, associated to ecological diversification (Baudoux et al. 2015; Foulon et al. 2008; Lovejoy et al. 2007), could be expected to be associated with morphological variations. We thus carefully examined cells from different *Micromonas* strains belonging to the main genetic clusters using light and/or electron microscopy in order to detect potential distinctive morphological characters. We discovered two 1 µm long flagellar hairs at the tip of the hair point (tip hairs, Marin and Melkonian 1994) for *Micromonas* strains belonging to the 3 main genetic lineages (RCC372, RCC449, RCC472, RCC746, RCC804 and RCC834, Fig. 5). All other described flagellate genera within the class Mamiellophyceae (*Mamiella*, *Mantoniella*, *Dolichomastix*, *Crustomastix*, *Monomastix*) possess various types of flagellar hairs, and tip hairs have been reported in for example, *Mamiella* and *Mantoniella* (Marin and Melkonian 1994). Tip hairs seem to be easily lost and were probably overlooked in previous electron microscopical studies of *Micromonas* cells. This new morphological feature is quoted in the emendation of the genus *Micromonas* (see below).

No distinctive character in cell body size, shape, flagellar insertion (LM, SEM and TEM, Fig. 6) was detected among strains belonging to different genetic clades, but flagellar length, measured in exponentially growing cells, varied among strains and clades (Fig. 7). Differences were due to variations in hair point length, whereas the proximal part of the flagellum was similar in length for all strains (approx. 1 µm, data not shown). For a given strain, longer lengths were obtained when TEM whole-mount preparations were used for measurements, but flagellar length as estimated with LM or TEM did not vary significantly among growth stages and cell cycle (Fig. 7, A to D).

Micromonas pusilla (clade C.D.5) strains possessed a significantly longer flagellum (3.92 ± 0.13 µm) than both clades A.ABC.12 and non-arctic B.E.3 (Mann-Whitney pairwise comparisons,

$p < 10^{-3}$). Flagellar length of clade A.ABC.12 strains ($2.34 \pm 0.09 \mu\text{m}$) and non-arctic B.E.3 strains ($1.8 \pm 0.10 \mu\text{m}$) was also significantly different (Mann-Whitney pairwise comparison, $p < 10^{-2}$), but flagellar length of strains belonging to sub-clades A.A.2, A.B.1 and A.C.1 was not significantly different (Kruskal-Wallis test, $p = 0.22$).

Differences in hair point length cannot be used alone to assign cells to a specific lineage since measured lengths on individual cells overlapped between the different clades (Fig. 7E). Within prasinophytes, as well as within the Mamiellophyceae, the flagellum of *Micromonas* is extremely unusual in that it is the only one to possess a long hair point (Sym and Pienaar 1993). This hair point contains the prolongation of the central pair of microtubules present in the flagellum (Manton 1959). Central pairs of microtubules are known to regulate motility (Mitchell 2004). Studies have shown that the hair point in *Micromonas* is motile by rotation (Omoto and Witman 1981; Omoto et al. 1999). The consequences of a reduction in hair point size on the swimming ability of a cell are difficult to predict. The swimming behaviour was estimated to be similar by Guillou et al. (2004) for several strains belonging to the three genetic clades. It would be interesting to investigate this aspect in more detail in order to formulate and test hypotheses concerning the role of the hair points (such as escaping predators or moving to nutrient spots or prey) for this pelagic genus.

Variability in Pigment Content Within the Genus *Micromonas*

Pigment content is thought to possess a critical selective value for marine phytoplanktonic organisms and to be connected to niche adaptation (Six et al. 2004, 2008; Stomp et al. 2004). *Micromonas* displays the classical pigment suite of prasinoxanthin-containing green algae, and more specifically of Mamiellophyceae (Latasa et al. 2004; Marin and Melkonian, 2010). In addition, a chlorophyll pigment (Chl c_{CS-170}) first detected in the tropical *Micromonas* strain CS-170 by (Jeffrey 1989) has been reported to occur in *Micromonas* strain RCC372, but to be absent from strains RCC418 and CCMP490 (Latasa et al. 2004). Chl c_{CS-170} has recently been identified as a [7-Methoxycarbonyl-8-vinyl] protochlorophyllide (Alvarez et al. 2013). Because it was detected in *Ostreococcus* and *Micromonas* strains isolated mainly in deep waters (Jeffrey 1989; Latasa et al. 2004; Rodríguez et al. 2005), as well as in a *Prasinococcus* strain also isolated near the bottom of the photic zone (Latasa et al. 2004), this pigment has been hypothesized to be a potential biomarker to identify low light ecotypes. In order to identify potential pigment signatures for individual lineages within *Micromonas*, we analyzed the pigment content of 37 *Micromonas* isolates. All strains displayed the classical pigment suite of Mamiellophyceae and 13 strains possessed Chl c_{CS-170} (Fig. 1). Of the 16 strains analyzed within clade A.ABC.12, Chl c_{CS-170} was present in 11 strains and occurred in each of the 3 sub-clades distinguished by phylogenetic analysis. Strains of clade A.ABC.12 that possessed this pigment were isolated at different depths: surface (RCC299), 5 m (RCC836), 25 m (RCC448, RCC451 and RCC808), and 120 m (RCC450) or unknown depths. Strains that lacked this pigment were isolated from surface waters (RCC570, RCC676) or unknown depths. Chl c_{CS-170} was not detected in strain RCC1109 from the “unknown clade” of Wu et al. (2014) and arctic strain CCMP2099 (Lovejoy et al. 2007). It was detected in only 1 isolate (RCC806) of clade B.E.3 and 1 isolate (RCC833) of clade C.D.5 (out of 10 isolates analysed for each of these clades). While RCC806 was isolated from surface waters in the bay of Naples, RCC833 was isolated from the Gulf of Mexico at a depth of 275 m. Other strains of clades B.E.3 and C.D.5 for which isolation depth information is available were isolated from surface waters or from 1800 m (RCC497; but this strain probably originate from cells attached to larger particles and transported to depth through sedimentation). Hence, Chl c_{CS-170} cannot serve as a biomarker for any of the genetic clades distinguished. Its higher occurrence in isolates retrieved from deeper environments provides some evidence for a link to physiological adaptation to low light.

Genetic Clades of *Micromonas pusilla* Correspond to Distinct Species Rather than Distinct

Genera

The genus *Micromonas* was described by Manton and Parke (1960) and originally included both *M. pusilla* and *M. squamata*. *Micromonas squamata* Manton & Parke was transferred to the genus *Mantoniella* by Desikachary (Desikachary 1972) because this species has both body and flagellar scales (that are absent in *Micromonas pusilla*), and because the flagellar insertion is different in *Mantoniella squamata*. The description of the genus *Micromonas* was not revised by Desikachary to take into account this modification. Given the high genetic divergences observed between the deeply diverging *Micromonas* clades (values were similar to that observed between *Micromonas* clades and e.g. *Mantoniella* species, Table 3) Marin and Melkonian (2010) suggested that the corresponding clades should be raised to genus status. The monophyly of *Micromonas* although moderately supported in some phylogenies of the Mamiellophyceae (e.g. in the SSU rRNA gene Viridiplantae phylogeny of Marin and Melkonian 2010) was strongly supported in the multigene phylogeny reported by Slapeta et al. (2006). *Mantoniella* and *Micromonas* also exhibit several important morphological differences. *Mantoniella* has two very unequal flagella and is covered by an outer layer of large, flattened, spider web-like scales (Desikachary 1972; Moestrup 1990) while cells of all clades of *Micromonas* are scale-less and possess a single peculiar true flagellum with a long hair point. In addition, all *Micromonas* lineages have a commonly shared SSU rDNA synapomorphy (C-G in bp 1 of Helix 11) to the exclusion of *Mantoniella* and the remaining Mamiellophyceae (U-A in this position; Table 4). We thus kept all lineages within the same genus and provided an emendation of the genus *Micromonas* in order to include, besides the absence of scale covering as a morphological character, the distinctive molecular SSU rDNA signature, and our discovery of new morphological features (the presence of tip hairs at the extremity of its hair point, and variability in hair point length, see above). This genus that has colonised most ocean surface waters, is probably best adapted to coastal habitats as suggested by the distribution of stations from which isolates or environmental sequences were obtained (Fig 8a, b).

Delineating species is a highly challenging task because not only the definition of what constitutes a species but also the criteria to consider for delineation are controversial. As a consequence, delimitations that are congruent across methods are recommended (De De Queiroz 2007; Leliaert and De Clerck 2017). Within the genus *Micromonas*, the elevation of clades A.ABC.12, non-arctic B.E.3, arctic E_a and C.D.5 to species status appears fully justified. Besides the high sequence divergences between clades, the distinctive morphological features and/or molecular signatures that we discovered for each of these clades are congruent with ecological (Baudoux et al. 2015; Foulon et al. 2008; Lovejoy et al. 2007) and genomic specificities (Simmons et al. 2015; Worden et al. 2009) identified in previous studies. These distinctive features are detailed below.

Clade C.D.5 comprises a group of strains with remarkable similarities in the sequences of different nuclear, mitochondrial and plastid genes (Slapeta et al. 2006; Worden et al. 2009 and this study). Clear synapomorphic signatures both in the SSU rRNA and ITS2 including CBCs and hemi-CBCs separate this clade from all other clades in the genus *Micromonas*. *Micromonas* C.D.5 clade strains were also recently found to possess specific introner elements (IE) not found in other *Micromonas* lineages (D-IEs, Simmons et al. 2015). If the nuclear genome of strain CCMP1545 consists of 19 chromosomes (van Baren et al. 2016; Worden et al. 2009), PFGE genome sizes analyses suggest that strains of clade C.D.5 possess 19 or 20 chromosomes (Supplementary Material Table S1, Fig. S1). Variations in chromosome numbers but also chromosome rearrangements have already been observed within the mamiellophycean species *Ostreococcus lucimarinus* (Palenik et al. 2007; Rodríguez et al. 2005). How these variations impact meiosis and fecundation has not been studied in this group of algae but interfertility between cultured strains presenting important genomic variations has been observed for some fungi and green algal species (Delneri et al. 2003; Flowers et al. 2015). In addition, strains belonging to clade C.D.5 possess a distinctively long hair point. Studies by Foulon et al. (2008) and Baudoux et al. (2015) suggest that this clade, which appears less prominent than other clades in the environment, has a distinct ecological niche and interacts with specific virus populations. This clade includes strains derived from the original isolate PL27 (Table 2), the neo-type culture upon which the original

description of *M. pusilla* was based. We therefore restricted the name *Micromonas pusilla* (sensu stricto) to those cells that have a slender hair point (as quoted in the original description by Manton and Parke, 1960) ca 1.5 to 7 μm long, as well as the distinctive synapomorphic signatures highlighted in Table 4 and Figures 3 and 4. *Micromonas pusilla* seems to exhibit a broad biogeographical distribution. Isolates or sequences of this species have been obtained mostly from coastal zones across a wide range of latitudes, in all major oceanic provinces (Table 2, Fig 8). However *M. pusilla* seems to be a minor component of the genus *Micromonas* in these coastal waters and while it has been shown to become the dominating *Micromonas* species in oceanic waters, it was detected at low absolute concentrations of the order of 100 cells mL^{-1} (Foulon et al. 2008).

Clade A.ABC.12 has been detected worldwide in a wide range of biogeographic regions (Table 2, Fig. 8A). It was found to be the most abundant and ubiquitous of the *Micromonas* lineages distinguished by Foulon et al. (2008). The genomes of strain RCC299, that belongs to clade A.A.2, and strain CCMP1545, a strain derived from the neotype strain of *M. pusilla*, are highly divergent since they each harbour at least 19% of unique genes (van Baren et al. 2016; Worden et al. 2009). For these reasons, a subset of strains belonging to the closely related subclades A.A.2 and A.B.1 have recently been assigned to the new species *Micromonas commoda* by van Baren et al. (2016). However, our synapomorphy search revealed only a single unique signature for subclades A.A.2 and A.B.1 together, i.e. a hemi-CBC in Helix 2 of ITS2 (bp 10). In contrast, the entire clade A.ABC.12, including subclade A.C.1, was distinguished from other *Micromonas* lineages by several prominent ITS2 signatures, including a CBC in Helix 2 (bp. 14; Table 4, Fig. 4B). Slight variations of whole-genome and chromosomes sizes were recorded between strains analysed in clade A.ABC.12 but all strains possessed 17 chromosomes (Supplementary Material Table S1, Fig. S1), as was shown by whole genome sequencing for strain RCC299 (van Baren et al. 2016). Likewise, the flagellar lengths of clade A.ABC.12 strains proved to be significantly different from those of other lineages, but similar among sub-clades distinguished by Slapeta et al. (2006). A majority of strains of clade A.ABC.12 possess the pigments Chl CCS-170 , potentially linked to a physiological adaptation to low light, while this pigment was not detected in most strains from other clades. Peculiarities of introner elements were also detected in the genomes of strains from this clade which possess a distinct IE family (ABC-IE, Simmons et al. 2015)). For all of these reasons, we elevated the entire deeply diverging A.ABC.12 clade to species level, under the name *M. commoda*. The extent of genetic variation within species is highly variable and the relative contribution of its determinants (effective population size, mutation rates, life-history traits) still largely unknown (Leffler et al. 2012; Romiguier et al. 2014). The rather high genetic diversity and structure within this clade (this study; Slapeta et al. 2006; Worden et al. 2009) may be the result of ongoing speciation events. It may also correspond to natural intra-specific polymorphism and could reflect peculiarities associated to key species traits such as life-history strategies and responses to short-term environmental perturbations (Leffler et al. 2012; von Dassow et al. 2015).

The non-arctic B.E.3 sub-clade (clade E1 in Simmons et al. 2015), although not always highly supported in SSU rRNA gene phylogenies, was highly supported when ITS2 sequences were added to the dataset and clear synapomorphic ITS2 signatures were identified for this clade (Table 4, Fig. 4B). The strains examined within this sub-clade possessed 21 chromosomes (Supplementary Material Table S1, Fig. S1). The non-arctic B.E.3 sub-clade appears to be ubiquitous (Table 2, Fig. 8). Foulon et al. (2008) suggested that this clade, that outcompeted other *Micromonas* lineages in coastal environments in summer, was well adapted to warm (presumably well illuminated) inshore habitats. Genetic polymorphism within this clade was comparable to that estimated for clade A.ABC.12. Isolates of this lineage had a significantly shorter hair point than all other isolates. This clade is thus described as a new species, *Micromonas bravo*.

The "arctic" *Micromonas* clade (Ea) discovered by Lovejoy et al. (2007) appears widespread in the Arctic Ocean (Balzano et al. 2012b; Lovejoy et al. 2007) and has recently been reported in the Southern Ocean and the deep currents that transport arctic water to the Southern Ocean (Simmons et al. 2015). Isolates that fall in this clade and environmental sequences assigned to it were all retrieved from arctic waters, but not from the Antarctic (Table 2, Fig. 8a). These strains are clearly psychrophilic and possess highly similar SSU rDNA and ITS2 sequences (Balzano et al. 2012a; Lovejoy et al. 2007). McKie-Krisberg et al. (2014) provided evidence that the arctic isolate CCMP 2099 was capable of phagotrophy. This trophic strategy may be specific to

the arctic lineage (ingestion of fluorescently labeled bacteria was not observed for *Micromonas* in the Mediterranean, Unrein et al. 2014), but may also be environmentally determined. In any case, this highly supported clade, that shows clear synapomorphic ITS2 signatures (Table 4, Fig. 4B) and is strictly associated with polar waters, corresponds to a distinct biological unit, which is also described here as a new species, *Micromonas polaris*.

Compared to other *Micromonas* clades, information available for clades B._4 and for the “unknown clade” discovered by Wu et al. (2014) is scarce. Both clades exhibit high SSU rDNA genetic divergences with the *Micromonas* species described. Environmental sequences of clade B._4 have been detected in different regions of the Mediterranean Sea, Red Sea and Pacific Ocean (Fig 8B) but no culture is available to date. Environmental sequences from the “unknown clade” (Wu et al. 2014) have been found in the Red and South China Seas (Fig 8B), and a single culture strain has been isolated in a Mediterranean lagoon (Fig. 8A). Interestingly, temperatures at the 3 corresponding sampling stations were particularly high (> 28°C, Wu et al. 2014, Acosta et al. 2013). We propose the designation of *Micromonas* candidate species 1 for clade B._4 and the designation of *Micromonas* candidate species 2 for the “unknown clade” reported by Wu et al. (2014).

Our results and the analysis of existing data allowed us to clarify the status of genetic lineages identified earlier within *Micromonas*, to write or revise formal descriptions for four species and suggest the existence of two additional species. This work will facilitate further interpretations of large-scale analyses investigating ecological trends in time and space for this widespread microalga. In the future population metagenomics analyses such as that conducted by Vannier et al. (2016), but also eco-physiological studies using cultured strains should help decipher further the extent of species diversity and evolutionary history of the genus *Micromonas*.

Taxonomic Revisions

Revision of the Genus *Micromonas*

Micromonas I. Manton & M. Parke, 1960, J. Mar. Biol. Assoc. U.K. 39: pp. 292, 298, emend. Simon, Foulon and Marin

Emended diagnosis: Motile cells ellipsoid to pyriform, slightly compressed, naked (cell wall absent, no organic body scales), 1-3 µm long, 0.7-1 µm broad; one flagellum attached laterally, less than 1 µm long, with a 1 to 7 µm long hair point; tip hairs about 1 µm long at the extremity of the hair point; cells without body and flagellar scales, chloroplast single appearing crescent in side view with a large pyrenoid filling the concavity; starch shell around the pyrenoid visible under electron microscopy; stigma absent; one mitochondrion lying on inner face of the chloroplast; no contractile vacuole; nucleus sub-spherical, situated near the flagellar base; fission in motile or palmelloid phase. First base pair of Helix 11 in the nuclear encoded small subunit rRNA is C-G instead of U-A. Broadly distributed in estuaries, coastal habitats and open oceans from the poles to the equator.

Type species: *Micromonas pusilla* (Butcher) Manton & Parke

Emendation of the Type Species *Micromonas pusilla*

Micromonas pusilla (R.W. Butcher) I. Manton & M. Parke, 1960, J. Mar. Biol. Assoc. U.K. 39: pp. 292, 298, emend. Simon, Foulon and Marin

Emended diagnosis: Characters of the genus. Flagellum including hair point longer than 3 µm. Nuclear genome comprising 19 or 20 chromosomes. In the nuclear-encoded 18S rRNA, base pair 10 of Helix 25 is A-U instead of U-A. In ITS2, base pair 13 of Helix 3 is U-A instead of C-G. Broadly distributed, mostly from coastal zones.

Authentic strain: RCC834 derived from strain Plymouth no.27 (PL27), isolated by M. Parke from the surface waters at position 50°15' N, 04° 13' W (13 April 1950).

Emendation of the species *Micromonas commoda*

Micromonas commoda J. van Baren, C. Bachy and A. Worden, 2016, in: van Baren et al. 2016, BMC Genomics 17:267, p. 6, emend. Simon, Foulon and Marin

Emended diagnosis: Characters of the genus. Flagellum including hair point approximately 2.5 µm in length. Nuclear genome comprising 17 chromosomes. In ITS2 of the nuclear-encoded rRNA operon, base pair 14 of Helix 2 is A-U instead of G-C, and base pair 4 of Helix 4 is U-G instead of C-G. Worldwide distribution, often with high abundance.

Authentic strain: clonal strain CCMP2709 derived from the isolate RCC299 (=NOUM17), isolated by S. Boulben from the Equatorial Pacific at 22°20'S, 166°20'W (10 February 1998).

Micromonas bravo Simon, Foulon and Marin, sp. nov.

Diagnosis: Characters of the genus. Short flagellum of approximately 1.8 µm length (including hair point). Nuclear genome comprising 21 chromosomes. In ITS2 of the nuclear-encoded rRNA operon, base pair 14 of Helix 2 is G-U instead of G-C. Broadly distributed, especially in warm coastal environments.

Holotype: Cells of *M. bravo* strain RCC434 preserved in a metabolically inactive state (cells embedded in resin for electron microscopy) have been deposited at the RCC.

Type locality: Strain RCC434 was isolated from Mediterranean Sea surface waters off Spain (41° 40' N, 2° 48' E) by L. Guillou (20 March 2001).

Etymology: The species name refers to the international code word for the letter B in the NATO phonetic alphabet and to clade B, originally chosen by Guillou et al. (2004) and referred to as “non-polar B.E.3” in this article.

Micromonas polaris Simon, Foulon and Marin, sp. nov.

Diagnosis: Characters of the genus. In ITS2 of the nuclear-encoded rRNA operon, base pair 16 of Helix 2 is U-A instead of C-G. Psychrophilic microalgae, restricted to polar waters.

Holotype: Cells of *M. polaris* strain RCC2306 preserved in a metabolically inactive state (cells embedded in resin for electron microscopy) have been deposited at the RCC.

Type locality: RCC 2306 was isolated from 70 m in the Beaufort Sea (71° 24' N, 132° 40' W) by D. Vaultot (15 August 2009).

Etymology: The species name refers to its distribution in polar marine waters.

Methods

Cultures: Thirty-eight culture strains (Table 2) of picoeukaryotes assigned to the genus *Micromonas* based on morphological characters or analysis of the SSU rRNA gene sequences were obtained from the Roscoff Culture Collection (RCC, Roscoff, France, <http://roscoff-culture-collection.org>). Some of these strains corresponded to isolates acquired by the RCC from the Bigelow National Center for Marine Algae and Microbiota (NCMA, <https://ncma.bigelow.org/>), the National Institute of Technology and Evaluation Biological Resource Center (NBRC, <http://www.nite.go.jp/en/nbrc/cultures/>), the North East Pacific Culture Collection (NEPCC, <http://www3.botany.ubc.ca/cccm/>) or the CSIRO Australian National Algae Culture Collection (<https://www.csiro.au/en/Research/Collections/ANACC>). All strains were maintained at either 20 °C or 4 °C under a 12:12 h LD (light : dark) regime in K medium (Keller et al. 1987). Light was provided by Sylvania Daylight fluorescent bulbs with an intensity of 100 $\mu\text{mol photon}\cdot\text{m}^{-2}$.

DNA amplification and sequencing: Sequences of the SSU rRNA gene and ITS 1 and 2 were obtained for 27 of the strains listed in Table 2. DNA was extracted by a modified cetyltrimethylammonium bromide (CTAB) protocol (Winnepenninckx et al. 1993). Cells (200 mL of culture) were harvested by centrifugation. The pellet was resuspended in 0.8 mL of CTAB buffer and incubated for 30 min at 60 °C with 0.1 mg/ml proteinase K. DNA was extracted by the addition of 0.8 mL of chloroform:isoamyl alcohol (24:1). After gentle agitation for 2 min, the organic phase was removed by a 10 min centrifugation step at 4 °C. The aqueous phase was recovered and incubated with 0.6 mL of isopropanol for 30 min at room temperature to precipitate the DNA. DNA was further washed by the addition of 1 mL of EtOH 76 % / ammonium acetate 10mM, dried, resuspended in sterile water and stored at -20 °C.

Extracted DNA was used as a template to amplify the nuclear small subunit ribosomal (SSU rRNA gene) and internal transcribed spacers (ITS rRNA). The eukaryotic primers Euk328f (5'-ACC TGG TTG ATC CTG CCA G-3') and Euk329r (5'-TGA TCC TTC YGC AGG TTC AC-3') were used to amplify the SSU rDNA as described in Romari and Vaultot (2004) with the following conditions:

an initial incubation step at 95 °C for 5 min, followed by 34 cycles with a denaturing step at 95 °C for 1 min, an annealing step at 62 °C for 2 min and an extension step at 72 °C for 3 min; these cycles were followed by a final extension step at 72 °C for 7 min. The primers D1 (5'-GTA GGT GAA CCT GCG GAA GGA-3' and R1 (5'-CCT TGG TCC GTG TTT CTA GAC-3') and D2 (5'-ACC CGC CGA ATT TAA GCA TA-3') and R2 (5'-AGG GGA ATC CTT GTT AGT TTC-3'), which are complementary to regions respectively upstream from the large subunit 28S rDNA and near the 3' end of the 18S rDNA gene, were used to amplify the ITS1, 2 and 5.8S rDNA as described in Guillou et al. (2004), with the following conditions: an initial incubation step at 94 °C for 12 min, followed by 30 cycles with a denaturing step at 94 °C for 1 min, an annealing step at 58 °C for 2 min and an extension step at 72 °C for 3 min; these cycles were followed by a final extension step at 72 °C for 10 min. Polymerase chain reactions were carried out in an automated thermocycler (iCycler, Bio-Rad, Marne-la-Coquette, France). The PCR mixture (25 µl final volume) contained 2.5 µl of Mg free buffer 10X (1X final concentration, Promega, Madison, Wisconsin), 2.5 µl of MgCl₂ solution (2.5 mM final concentration), 2 µl of deoxynucleoside triphosphate (dNTP, 400 µM final concentration each, Eurogentec), 0.5 µl of each primer (1 µM final contraction each), 0.125 µl of Taq Polymerase (5 units per µl, Promega, Madison, Wisconsin), sterile water and 1 µl of extracted DNA. PCR products were cloned using the TOPO TA cloning kit (Invitrogen, Carlsbad, CA, USA) following the protocol provided by the manufacturer.

Genetic polymorphism was then assessed by analysing several clones by RLFP. Clone inserts were amplified by PCR and then digested with the restriction enzyme *HaeIII* (0.25 units per µl, BioLabs, NewEngland) for 3 h at 37 °C. The digested products were separated by electrophoresis at 70 V for 2 h on a 1% agarose gel. When a strain presented several RFLP patterns, all were recovered for sequencing. PCR products were purified using the "QIAprep Miniprep" (Qiagen, Courtaboeuf, France) following the manufacturer's recommendations and then directly sequenced in both directions using the M13f and M13r primers from the TOPO TA cloning kit and fluorescent nucleotides (Big Dye Terminator) from the 'DNA Sequencing' kit (Applied Biosystems, Norwalk, Connecticut). Sequencing reaction conditions were as follows: an initial incubation step at 94 °C for 5 min, followed by 50 cycles with a denaturing step at 96 °C for 30 s, an annealing step at 55 °C for 30 s and an extension step at 60 °C for 4 min followed by cooling at 4 °C. The sequencing was performed using an ABI 3100 xl (Applied Biosystems).

Phylogenetic and genetic distances analyses: Additional relevant sequences were included for phylogenetic analyses in addition to the sequences obtained from the cultures listed in Table 2. These sequences were retrieved from GenBank and the National Center for Biotechnology Information (NCBI, <http://www.ncbi.nlm.nih.gov/>) and included SSU rRNA gene sequence and ITS rRNA sequence of strains which have been previously analysed (Bellec et al. 2014; Guillou et al. 2004; Marin and Melkonian 2010; Nakayama et al. 1998; Slapeta et al. 2006; Worden 2006;) (Table 2), a selection of environmental sequences grouping with *Micromonas*, and sequences of *Mantoniella* used as outgroups. Environmental SSU rRNA gene sequences were retrieved from the Protist Ribosomal Reference (PR²) database (Guillou et al. 2013) and selected based on their length (> 1695 bp). Sequences were then selected from each of the major deep branching clades based upon a preliminary phylogenetic analysis. The SSU rDNA and ITS rDNA sequences from RCC299 were retrieved from the whole chromosome 8 sequence (accession number: NC_013045) from the isolate which contained three identical copies of these genes. Sequences were aligned automatically using MUSCLE (SSU rRNA gene) or manually, taking into account the secondary structure (analyzed by MFold; <http://mfold.bioinfo.rpi.edu/>). Alignments (1865 bp for the concatenated markers, including 1606 bp for SSU) were analysed by two phylogenetic methods: Bayesian inference (BI) and maximum likelihood (ML). The homogeneity of SSU and ITS datasets was first assessed using a partition homogeneity test (Farris et al. 1994) on the pooled dataset (SSU + ITS) with PAUP*4.0b10 (Swofford 2002). Bayesian inference was performed with MrBayes 3.2.2 (Ronquist et al. 2012) using a HKY85 (Hasegawa-Kishino-Yano) evolutionary model accounting for substitution rate heterogeneity and a proportion of invariable sites, chosen with jModelTest 2 (Darriba et al. 2012) using the Akaike Information Criterion. The reconstruction used 2 runs of 4 chains of 10⁶ generations, with trees sampled every 100 generations and burn-in value set to 20 % of the sampled trees. Majority-rule consensus was kept as conservative estimates. Maximum likelihood reconstruction was carried out with PhyML (Guindon and Gascuel 2003; Guindon et al. 2005) using the same evolutionary model as BI and validated with a bootstrap procedure using 100 replicates. Uncorrected pairwise genetic distances

between clades were computed with PAUP* 4.0b10 (Swofford 2002). New sequences have been deposited to GenBank under the accession numbers KU244630 to KU244682 (Table 2).

Search for unique molecular signatures of clades: To screen the SSU rDNA for synapomorphies of *Micromonas* and sub-clades, a taxon-rich alignment containing more than 2000 viridiplants (green algae/embryophytes) was used in order to reveal all existing homoplasies (especially parallelisms) immediately, by application of the 'list of apomorphies' function of PAUP* 4.0b10, followed by several manual steps as described previously (Marin and Melkonian 2010; Marin et al. 2005). Similarly, the secondary structural ITS2 alignment of the class Mamiellophyceae, which was used by Marin and Melkonian (2010), was extended by novel *Micromonas* sequences after full reconstruction of their ITS2 secondary structures using MFold (<http://mfold.bioinfo.rpi.edu/>). The ITS2 synapomorphy search was confined to those positions, which formed universal base pairs in all members of *Micromonas* (bold base pair numbers in Fig. 4).

Morphological analyses and flagella measurements and statistical tests:

Morphological analyses were conducted under LM, TEM and SEM. For LM cells were fixed with Lugol solution. Fourteen μL of exponentially growing culture was mixed with 1 μL of acidic Lugol solution (Thronsdén 1978) on a microscopic slide. Cells were immediately examined with an Olympus BX51 microscope, at objective x100 under Nomarski interference contrast. Images were obtained with a SPOT camera (LM, G-Spot, Diagnostic Instruments Inc, USA, SPOT software, version 4.0.9, Diagnostic Instruments Inc, USA). For the examination of cells under TEM, whole-mounts were prepared by placing a drop of exponentially growing culture fixed with glutaraldehyde (1 % final) on a formvar (chloroform-formvar 0.8 %) coated copper grid (diameter 3.05 mm, type G200, TAAB). After 15 min the grid was rinsed in distilled water and air dried. Whole-mounts were contrasted for 15 min in 0.2 μm filtered uranyl acetate (20 %), rinsed in distilled water and air dried again. Grids were observed using a JEOL-JEM 1400 electron microscope operating at 80 kV and images were obtained with a Gatan Orius Ultrascan camera (TEM). For SEM examination, cultures were sampled during late exponential phase. Cells were fixed in 1-2% glutaraldehyde and 5-10 ml were filtered through Nucleopore filters (13 mm diameter, 2 μm pores) by gravity (volume depending on cell density and filter clogging). The filter was rinsed with growth medium then with 0.1 M cacodylic acid buffer, 10 min both. 0.5 ml of 1% osmium tetroxide in 0.1 M cacodylic acid buffer was added in the syringe for 30 min. Three rinses in 0.1 M cacodylic acid buffer were applied (5 min each) and then dehydration was achieved by serial transfers through progressive aqueous-ethanol series (70%, 90%, 96%, once and finally 100%, three times, 10 min each). All the filter-holders were placed in Critical Point Dryer and filters were subsequently placed on stubs with carbon tabs. Gold-Palladium sputtering has been finally applied to the cells before observation on field-emission scanning electron microscope HITACHI S-4800 at the University of Oslo.

The flagellum of *M. pusilla* includes a proximal wider portion (the true flagellum) and the distal hair point (slender distal portion) (Fig. 6, A and B). We used the term flagellum to designate the structure including both the true flagellum and the hair point. Flagellar lengths were estimated from cell images obtained using light microscopy (LM) or transmission electron microscopy (TEM). Flagella lengths were measured using the Image J software (Schneider et al. 2012) on images. Between 57 to 67, and 18 to 31 flagella were measured for LM and TEM respectively for each strain. Kruskal-Wallis and Mann-Whitney tests were performed to compare mean flagella lengths (as measured using LM) among genetic lineages. Statistical tests were performed using the PAST software (Hammer et al. 2001).

Pigment analyses: The pigment content of 40 isolates was determined in 50 mL of exponentially growing cells, growing under identical light conditions (100 $\mu\text{mol photon.m}^{-2}$ with a 12:12 light:dark cycle). Analytical and semi-preparative high-performance liquid chromatograph (HPLC) separations followed a protocol adapted from Zapata et al. (Zapata et al. 2000). Chlorophylls and carotenoids were detected by absorbance at 440 nm and identified by diode array spectroscopy (Jeffrey et al. 1997). *Micromonas* pigments were identified by co-chromatography with authentic standards (Sigma-Aldrich). The pigment analyses were replicated on larger volumes of cultures when needed, especially when the presence/absence of Chl c_{CS-170} could not be clearly established because of a poor resolution of the targeted absorbance peaks.

Mapping of *Micromonas* isolates and environmental sequences: In order to depict the distribution of *Micromonas* species and clades, we retrieved quality controlled and annotated

eukaryotic 18S rRNA gene sequences obtained from cultured strains and environmental samples from the PR² database (Guillou et al. 2013). All *Micromonas* sequences were extracted, yielding a final dataset of 516 environmental sequences and 44 isolates (corresponding to isolates listed in Table 2 plus RCC913 and RCC966). Chimeric sequences were filtered out by assigning the first 300 and last 300 base pairs of the sequences with the software Mothur v1.35.1 (Schloss et al. 2009). If a conflict of assignment between the beginning and the end of the sequences was detected, sequences were BLASTed against GenBank to confirm whether they were chimeras, and if this proved to be the case they were removed from any further analysis. Assignment of sequences to species or clades was achieved by aligning them using MAFFT v1.3.3 (Kato et al. 2002) and constructing phylogenetic trees using FastTree v1.0 (Price et al. 2009) run within the Geneious software v7.1.7 (Kearse et al. 2012). Phylogenetic trees were compared to that of Figure 1. Sequences that fell into one of the clades defined in Figure 1 were assigned to that clade. For each sequence, we extracted metadata from GenBank (such as sampling coordinates, date and publication details) or from culture collections databases, when available. Other metadata were obtained from the literature. This information was used to map the distribution of isolates and sequences using Tableau Desktop 9.2 (<http://www.tableau.com/>).

Acknowledgements

We thank Anne-Claire Baudoux, David Demory, John Dolan and Ian Probert for helpful discussions and English proofreading. We thank Fabienne Rigaut-Jalabert and Priscilla Gourvil for culture maintenance and Francisco Rodríguez for his help with HLPC analyses. Adriana Zingone kindly provided several *Micromonas* cultures. Bertrand Ytournal produced some of the sequences. This work was supported by the following programs: PICOVIR (N° BLAN07-1_200218) and REVIREC (N°ANR-12-BSV7-0006) from ANR (Agence Nationale pour la Recherche), “Souchothèque de Bretagne” (Contrat de Projet Etat-Région with funds from Région Bretagne, Département du Finistère and EU FEDER), ASSEMBLE EU FP7 research infrastructure initiative (EU-RI-227799), MaCuMBA (FP7-KBBE-2012-6-311975). EF and MT benefited from doctoral fellowships from the Region Bretagne and the Université Pierre and Marie Curie.

References

- Acosta F, Ngugi D, Stingl U** (2013) Diversity of picoeukaryotes at an oligotrophic site off the Northeastern Red Sea Coast. *Aquat Biosyst* **9**:16
- Alvarez S, Rodriguez F, Riobo P, Garrido JL, Vaz B** (2013) Chlorophyll cCS-170 isolated from *Ostreococcus* sp. is [7-methoxycarbonyl-8-vinyl]protochlorophyllide *a*. *Org Lett* **15**:4430–4433
- Balzano S, Marie D, Gourvil P, Vaultot D** (2012b) Composition of the summer photosynthetic pico and nanoplankton communities in the Beaufort Sea assessed by T-RFLP and sequences of the 18S rRNA gene from flow cytometry sorted samples. *ISME J* **6**:1480–1498
- Balzano S, Gourvil P, Siano R, Chanoine M, Marie D, Lessard S, Sarno D, Vaultot D** (2012a) Diversity of cultured photosynthetic flagellates in the North East Pacific and Arctic Oceans in summer. *Biogeosciences* **9**:4553–4571
- van Baren MJ, Bachy C, Reistetter EN, Purvine SO, Grimwood J, Sudek S, Yu H, Poirier C, Deerinck TJ, Kuo A, Grigoriev IV, Wong C-H, Smith RD, Callister J, Wei C-L, Schmutz J, Worden AZ** (2016) Evidence-based green algal genomics reveals marine diversity and ancestral characteristics of land plants. *BMC Genomics* **17**:267
- Baudoux A-C, Lebretonchel H, Dehmer H, Latimier M, Edern R, Rigaut-Jalabert F, Ge P, Guillou L, Foulon E, Bozec Y, Cariou T, Desdevises Y, Derelle E, Grimsley N, Moreau H, Simon N** (2015) Interplay between the genetic clades of *Micromonas* and their viruses in the Western English Channel. *Environ Microbiol Rep* **7**:765–773
- Bellec L, Clerissi C, Edern R, Foulon E, Simon N, Grimsley N, Desdevises Y** (2014) Cophylogenetic interactions between marine viruses and eukaryotic picophytoplankton. *BMC Evol Biol* **14**:59
- Butcher RW** (1952) Contribution to our knowledge of the smallest marine algae. *J Mar Biol Assoc UK* **31**:175–191
- 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

- 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
- Darriba D, Taboada GL, Doallo R, Posada D** (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* **9**:772–772
- von Dassow P, John U, Ogata H, Probert I, Bendif EM, Kegel JU, Audic S, Winckler P, Da Silva C, Claverie J-M, Doney S, Glover DM, Mella Flores D, Herrera Y, Lescot M, Garet-Delmas M-J, de Vargas C** (2015) Life-cycle modification in open oceans accounts for genome variability in a cosmopolitan phytoplankton. *ISME J* **9**:1365–1377
- Delneri D, Colson I, Grammenoudi S, Roberts IN, Louis EJ, Oliver SG** (2003) Engineering evolution to study speciation in yeasts. *Nature* **422**:68–72
- Desikachary TV** (1972) Notes on Volvocales. *Ind Curr Sci* **41**:445–447
- Farris JS, Källersjö M, Kluge AG, Bult C** (1994) Testing significance of incongruence. *Cladistics* **10**:315–319
- Fawley MW, Osterbauer N, Lee CM, Jiao S** (1990) The light-harvesting complex of *Mamiella gilva*: a character linking scaly and naked members of the Micromonadophyceae (Chlorophyta). *Phycologia* **29**:511–514
- Fawley MW, Yun Y, Qin M** (2000) Phylogenetic analyses of 18S rDNA sequences reveal a new coccoid lineage of the Prasinophyceae (Chlorophyta). *J Phycol* **36**:387–393
- Flowers JM, Hazzouri KM, Pham GM, Rosas U, Bahmani T, Khraiweh B, Nelson DR, Jikakli K, Abdrabu R, Harris EH, Lefebvre PA, Hom EFY, Salehi-Ashtiani K, Purugganan MD** (2015) Whole-genome resequencing reveals extensive natural variation in the model green alga *Chlamydomonas reinhardtii*. *Plant Cell* **27**:2353–2369

- Foulon E, Not F, Jalabert F, Cariou T, Massana R, Simon N** (2008) Ecological niche partitioning in the picoplanktonic green alga *Micromonas pusilla*: evidence from environmental surveys using phylogenetic probes. *Environ Microbiol* **10**:2433–2443
- Graham LE, Graham JM, Wilcox LW** (2009) *Algae* (2nd Edition). Benjamin Cummings (Pearson): San Francisco, 720 p
- Guillou L, Eikrem W, Chrétiennot-Dinet M-J, Le Gall F, Massana R, Romari K, Pedrós-Alió C, Vaultot D** (2004) Diversity of picoplanktonic prasinophytes assessed by direct nuclear SSU rDNA sequencing of environmental samples and novel isolates retrieved from oceanic and coastal marine ecosystems. *Protist* **155**:193–214
- Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner L, Boutte C, Burgaud G., de Vargas C, Decelle J, del Campo J, Dolan J, Dunthorn M, Edvardsen B, Holzmann M, Kooistra WHCF, Lara E, Le Bescot N, Logares R, Mahé F, Massana R, Montresor M, Morard R, Not F, Pawlowski J, Probert I, Sauvadet A-L, Siano R, Stoeck T, Vaultot D, Zimmermann P, Christen R** (2013) The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small SubUnit rRNA sequences with curated taxonomy. *Nucleic Acids Res* **41**:D597–D604
- Guindon S, Gascuel O** (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by Maximum Likelihood. *Syst Biol* **52**:696–704
- Guindon S, Lethiec F, Duroux P, Gascuel O** (2005) PHYML Online - A web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res* **33**:557–559
- Hammer Ø, Harper DAT, Ryan PD** (2001) PAST: Palaeontological statistics software package for education and data analysis. *Palaeontol Electron* **4**:1–9
- Jeffrey SW** (1989) Chlorophyll *c* Pigments and their Distribution in the Chromophyte Algae. In Green, JC, Leadbeater, BSC, Diver, WL (eds) *The Chromophyte Algae: Problems and Perspectives*. Clarendon Press: Oxford, pp 13–36
- Jeffrey SW, Mantoura RFC, Bjornland T** (1997) Data for the Identification of 47 Key Phytoplankton Pigments. In Jeffrey, SW, Mantoura, RFC, Wright, SW (eds) *Phytoplankton Pigments in Oceanography*. UNESCO Publishing: Paris, pp 449–559

- Katoh K, Misawa K, Kuma K, Miyata T** (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**:3059–3066
- Kawasaki Y, Nakada T, Tomita M** (2015) Taxonomic revision of oil-producing green algae, *Chlorococcum oleofaciens* (Volvocales, Chlorophyceae), and its relatives. *J Phycol* **51**:1000–1016
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A** (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**:1647–1649
- Keller MD, Selvin RC, Claus W, Guillard RRL** (1987) Media for the culture of oceanic ultraphytoplankton. *J Phycol* **23**:633–638
- Knight-Jones EW, Walne PR** (1951) *Chromulina pusilla* Butcher, a dominant member of the ultraplankton. *Nature* **167**:445–447
- Latasa M, Scharek R, Gall F Le, Guillou L, Le Gall F** (2004) Pigment suites and taxonomic groups in Prasinophyceae. *J Phycol* **40**:1149–1155
- Leffler EM, Bullaughey K, Matute DR, Meyer WK, Ségurel L, Venkat A, Andolfatto P, Przeworski M** (2012) Revisiting an old riddle: what determines genetic diversity levels within species? *PLoS Biol* **10**:e1001388
- Leliaert F, De Clerck O** (2017) Refining species boundaries in algae. *J Phycol* **53**:12–16
- Leliaert F, Smith DR, Moreau H, Herron MD, Verbruggen H, Delwiche CF, de Clerck O** (2012) Phylogeny and molecular evolution of the green algae. *CRC Crit Rev Plant Sci* **31**:1–46
- Lovejoy C, Vincent WF, Bonilla S, Roy S, Martineau MJ, Terrado R, Potvin M, Massana R, Pedrós-Alió C** (2007) Distribution, phylogeny, and growth of cold-adapted picoprasinophytes in arctic seas. *J Phycol* **43**:78–89
- Manton I** (1959) Electron microscopical observations on a very small flagellate: the problem of *Chromulina pusilla* Butcher. *J Mar Biol Assoc UK* **38**:319–333
- Manton I, Parke M** (1960) Further observations on small green flagellates with special reference to

possible relatives of *Chromulina pusilla* Butcher. J Mar Biol Assoc UK **39**:275–298

Marin B, Melkonian M (1994) Flagellar hairs in Prasinophytes (Chlorophyta): Ultrastructure and distribution on the flagellar surface. J Phycol **30**:659–678

Marin B, Melkonian M (2010) Molecular phylogeny and classification of the Mamiellophyceae class. nov. (Chlorophyta) based on sequence comparisons of the nuclear- and plastid-encoded rRNA operons. Protist **161**:304–336

Marin B, Nowack ECM, Melkonian M (2005) A plastid in the making: Evidence for a second primary endosymbiosis. Protist **156**:425–432

Mattox KR, Stewart KD (1984) Classification of the Green Algae: A Concept Based on Comparative Cytology. In Irvine, DEG, John, DM (eds) The Systematics of Green Algae. Academic Press: London, UK, pp 29–72

McKie-Krisberg ZM, Sanders RW (2014) Phagotrophy by the picoeukaryotic green alga *Micromonas*: implications for Arctic Oceans. ISME J **10**:1953–1961

Mitchell DR (2004) Speculations on the evolution of 9+2 organelles and the role of central pair microtubules. Biol Cell **96**:691–696

Moestrup Ø (1990) Scale structure in *Mantoniella squamata*, with some comments on the phylogeny of the Prasinophyceae (Chlorophyta). Phycologia **29**:437–442

Müller T, Philippi N, Dandekar T, Schultz J, Wolf M (2007) Distinguishing species. RNA **13**:1469–1472

Nakada T, Shinkawa H, Ito T, Tomita M (2010) Recharacterization of *Chlamydomonas reinhardtii* and its relatives with new isolates from Japan. J Plant Res **123**:67–78

Nakayama T, Marin B, Kranz HD, Surek B, Huss VAR, Inouye I, Melkonian M (1998) The basal position of scaly green flagellates among the green algae (Chlorophyta) is revealed by analyses of nuclear-encoded SSU rRNA sequences. Protist **149**:367–380

Not F, Latasa M, Marie D, Cariou T, Vaultot D, Simon N (2004) A single species *Micromonas pusilla* (Prasinophyceae) dominates the eukaryotic picoplankton in the western English Channel.

Appl Environ Microbiol **70**:4064–4072

Not F, Massana R, Latasa M, Marie D, Colson C, Eikrem W, Pedrós-Alió C, Vaultot D, Simon N (2005) Late summer community composition and abundance of photosynthetic picoeukaryotes in Norwegian and Barents Seas. *Limnol Oceanogr* **50**:1677–1686

Omoto CK, Witman GB (1981) Functionally significant central-pair rotation in a primitive eukaryotic flagellum. *Nature* **290**:708–710

Omoto CK, Gibbons I., Kamiya R, Shingyoji C, Takahashi K, Witman GB (1999) Rotation of the central pair microtubules in eukaryotic flagella. *Mol Biol Cell* **10**:1–4

Palenik B, Grimwood J, Aerts A, Rouzé P, Salamov A, Putnam N, Dupont C, Jorgensen R, Derelle E, Rombauts S, Zhou K, Otilar R, Merchant SS, Podell S, Gaasterland T, Napoli C, Gendler K, Manuell A, Tai V, Vallon O, Piganeau G, Jancek S, Heijde M, Jabbari K, Bowler C, Lohr M, Robbens S, Werner G, Dubchak I, Pazour GJ, Ren Q, Paulsen I, Delwiche C, Schmutz J, Rokhsar D, Van de Peer Y, Moreau H, Grigoriev IV (2007) The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *Proc Natl Acad Sci USA* **104**:7705–10

Price MN, Dehal PS, Arkin AP (2009) FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* **26**:1641–1650

Queiroz K De (2007) Species concepts and species delimitation. *Syst Bot* **56**:879–886

Rodríguez F, Derelle E, Guillou L, Le Gall F, Vaultot D, Moreau H (2005) Ecotype diversity in the marine picoeukaryote *Ostreococcus* (Chlorophyta, Prasinophyceae). *Environ Microbiol* **7**:853–859

Romari K, Vaultot D (2004) Composition and temporal variability of picoeukaryote communities at a coastal site of the English Channel from 18S rDNA sequences. *Limnol Oceanogr* **49**:784–798

Romiguier J, Gayral P, Ballenghien M, Bernard A, Cahais V, Chenuil A., Chiari Y, Dernet R, Duret L, Faivre N, Loire E, Lourenco JM, Nabholz B, Roux C, Tsagkogeorga G, Weber AA-T, Weinert LA, Belkhir K, Bierne N, Glémin S, Galtier N (2014) Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature* **515**:261–3

Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L,

- Suchard MA, Huelsen-beck JP** (2012) MrBayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* **61**:539–542
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF** (2009) Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**:7537–7541
- Schneider CA, Rasband WS, Eliceiri KW** (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**:671–675
- Simmons MP, Bachy C, Sudek S, van Baren MJ, Sudek L, Ares M, Worden AZ** (2015) Intron invasions trace algal speciation and reveal nearly identical arctic and antarctic *Micromonas* populations. *Mol Biol Evol* **32**:2219–2235
- Six C, Thomas JC, Brahmsha B, Lemoine Y, Partensky F** (2004) Photophysiology of the marine cyanobacterium *Synechococcus* sp. WH8102, a new model organism. *Aquat Microb Ecol* **35**:17–29
- Six C, Finkel Z V, Rodriguez F, Marie D, Partensky F, Campbell DA** (2008) Contrasting photoacclimation strategies in ecotypes of the eukaryotic picoplankton *Ostreococcus*. *Limnol Oceanogr* **53**:255–265
- Slapeta J, López-García P, Moreira D** (2006) Global dispersal and ancient cryptic species in the smallest marine eukaryotes. *Mol Biol Evol* **23**:23–29
- Stomp M, Huisman J, de JF, Veraart AJ, Gerla D, Rijkeboer M, Ibelings BW, Wollenzien IA, Stal L** (2004) Adaptive divergence in pigment composition promotes phytoplankton biodiversity. *Nature* **432**:104–107
- Subirana L, Péquin B, Michely S, Escande M-L, Meilland J, Derelle E, Marin B, Piganeau G, Desdevises Y, Moreau H, Grimsley N** (2013) Morphology, genome plasticity, and phylogeny in the genus *Ostreococcus* reveal a cryptic species, *O. mediterraneus* sp. nov. (Mamiellales, Mamiellophyceae). *Protist* **164**:643–659
- Swofford DL** (2002) PAUP*. Phylogenetic Analysis Using Parsimony (*and others methods).

- Sym SD, Pienaar RN** (1993) Further observations on *Trichocystis*, a subgenus of *Pyramimonas* (Prasinophyceae, Chlorophyta). *Phycologia* **32**:338–350
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S** (2013) MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* **30**:2725–2729
- Thomsen HA, Buck KR** (1998) Nanoflagellates of the central California waters: taxonomy, biogeography and abundance of primitive, green flagellates (Pedinophyceae, Prasinophyceae). *Deep - Sea Res Part II - Top Stud Oceanogr* **45**:1687–1707
- Thronsen J** (1978) Productivity and abundance of ultra- and nanoplankton in Oslofjorden. *Sarsia* **63**:273–284
- Thronsen J, Kristiansen S** (1988) Nanoplankton communities in Haltenbanken waters (Norwegian Sea) during an oil spill experiment, July—August 1982. *Sarsia* **73**:71–74
- Thronsen J, Kristiansen S** (1991) *Micromonas pusilla* (Prasinophyceae) as part of pico- and nanoplankton communities of the Barents Sea. *Polar Res* **10**:201–207
- Unrein F, Gasol JM, Not F, Forn I, Massana R** (2014) Mixotrophic haptophytes are key bacterial grazers in oligotrophic coastal waters. *ISME J* **8**:164–176
- Vannier T, Leconte J, Seeleuthner Y, Mondy S, Pelletier E, Aury J-M, de Vargas C, Sieracki M, Ludicone D, Vaultot D, Winckler P, Jaillon O** (2016) Survey of the green picoalga *Bathycoccus* genomes in the global ocean. *Sci Rep* **6**:37900
- Winnepenninckx B, Backeljau T, De Wachter R** (1993) Extraction of high molecular weight DNA from molluscs. *Trends Genet* **9**:407
- Worden AZ** (2006) Picoeukaryote diversity in coastal waters of the Pacific Ocean. *Aquat Microb Ecol* **43**:165–175
- Worden AZ, Lee J-H, Mock T, Rouzé P, Simmons MP, Aerts AL, Allen AE, Cuvelier M, Derelle E, Everett MV, Foulon E, Grimwood J, Gundlach H, Henrissat B, Napoli C, McDonald SM, Parker MS, Rombauts S, Salamov A, Von Dassow P, Badger JH, Coutinho PM, Demir E, Dubchak I, Gentemann C, Eikrem W, Gready JE, John U, Lanier W, Lindquist EA, Lucas S, Mayer KFX, Moreau H, Not F, Otilar R, Panaud O, Pangilinan J, Paulsen I, Piegu B, Poliakov**

A, Robbens S, Schlutz J, Toulza E, Wyss T, Zelensky A, Zhou K, Armbrust EV, Bhattacharya D, Goodenough UW, Van de Peer Y, Grigoriev IV (2009) Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. *Science* **324**:268–272

Wu W, Huang B, Zhong C (2014) Photosynthetic picoeukaryote assemblages in the South China Sea from the Pearl River estuary to the SEATS station. *Aquat Microb Ecol* **71**:271–284

Zapata M, Rodriguez F, Garrido JL (2000) Separation of chlorophylls and carotenoids from marine phytoplankton: a new HPLC method using a reversed phase C-8 column and pyridine-containing mobile phases. *Mar Ecol - Prog Ser* **195**:29–45

Figure Legends

Figure 1 : Phylogenetic reconstruction based on near full-length SSU rRNA gene sequences from *Micromonas* strains and a selection of environmental sequences (in blue). The tree was built via Bayesian inference (BI) and maximum likelihood (ML). Numbers are posterior probabilities (BI) and bootstrap values in % (ML) indicating clade support. *Mantoniella* sequences were used as outgroup taxa. Clades distinguished in Guillou et al. (2004), Slapeta et al. (2006), Worden (2006), Lovejoy et al. (2007) and Wu et al. (2014) are indicated (see also Table 1 for a comparison between clades labelling). The SSU rDNA clade to which *Micromonas commoda* van Baren, Bachy and Worden belongs is also indicated. Strain PL27 is the strain upon which the original description of *Micromonas pusilla* was based. Strains originating from the same original isolate are indicated. Black and white squares indicate the presence and absence of Chl C_{CS-170} . Pigments of CCMP2099 were not analysed in this study but the absence of Chl C_{CS-170} in this strain was reported by Lovejoy et al. (2007).

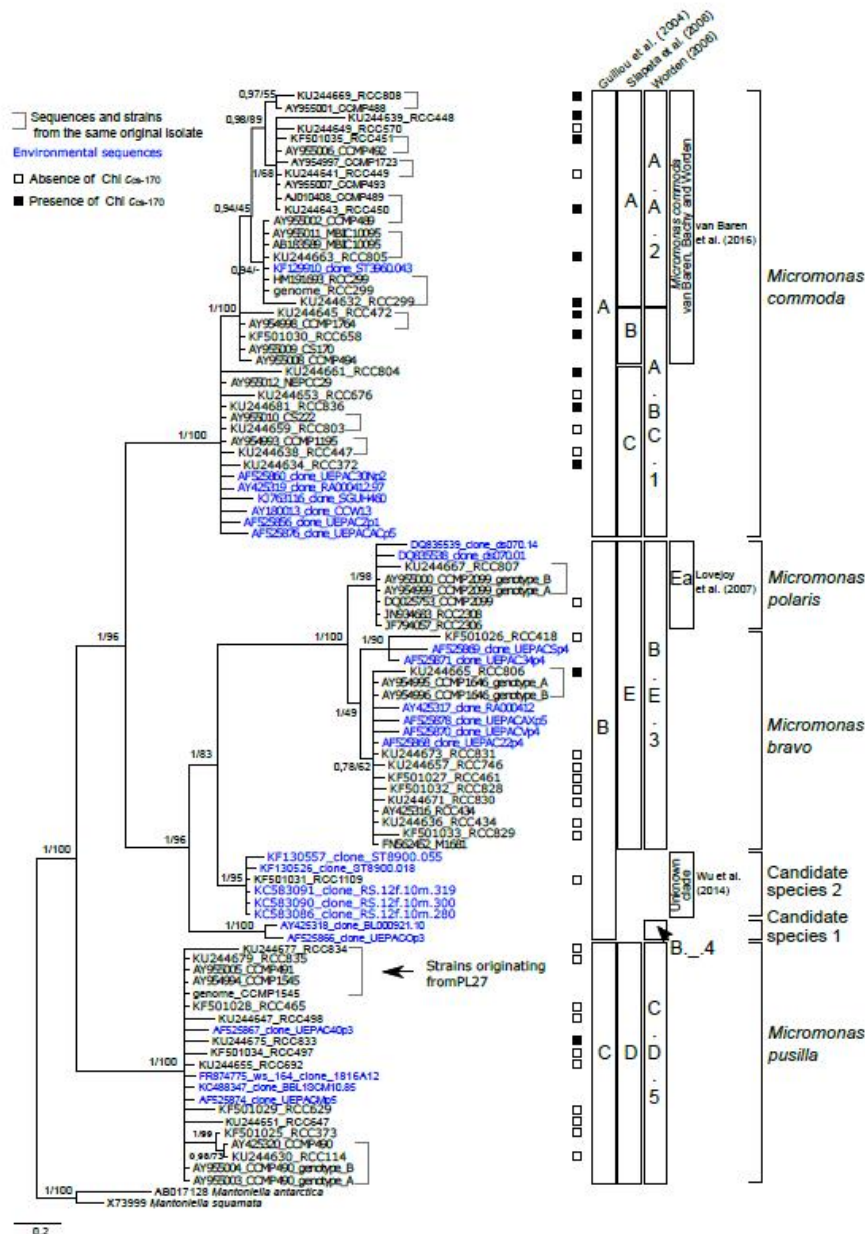


Figure 2. Phylogenetic reconstruction based on combined SSU rRNA gene and ITS2 sequences from *Micromonas* strains. Clades labelling is identical to that used in Figure 1 and Table 1. Note that clade B_4 (candidate species 1) of Figure 1 is based on the analysis of SSU rRNA gene environmental sequences. ITS2 sequences corresponding to this clade are not available. Rest of legend as in Figure 1.

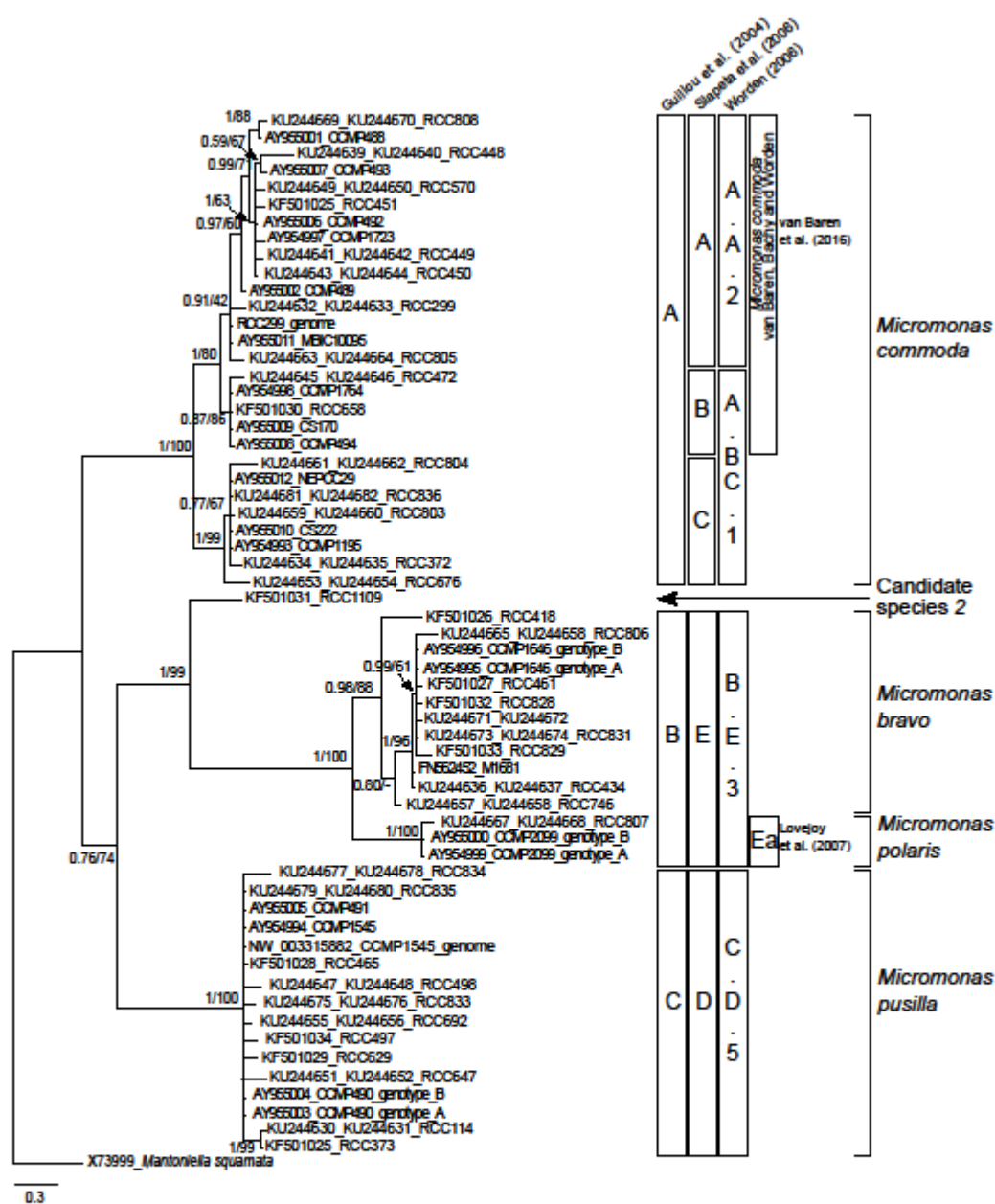
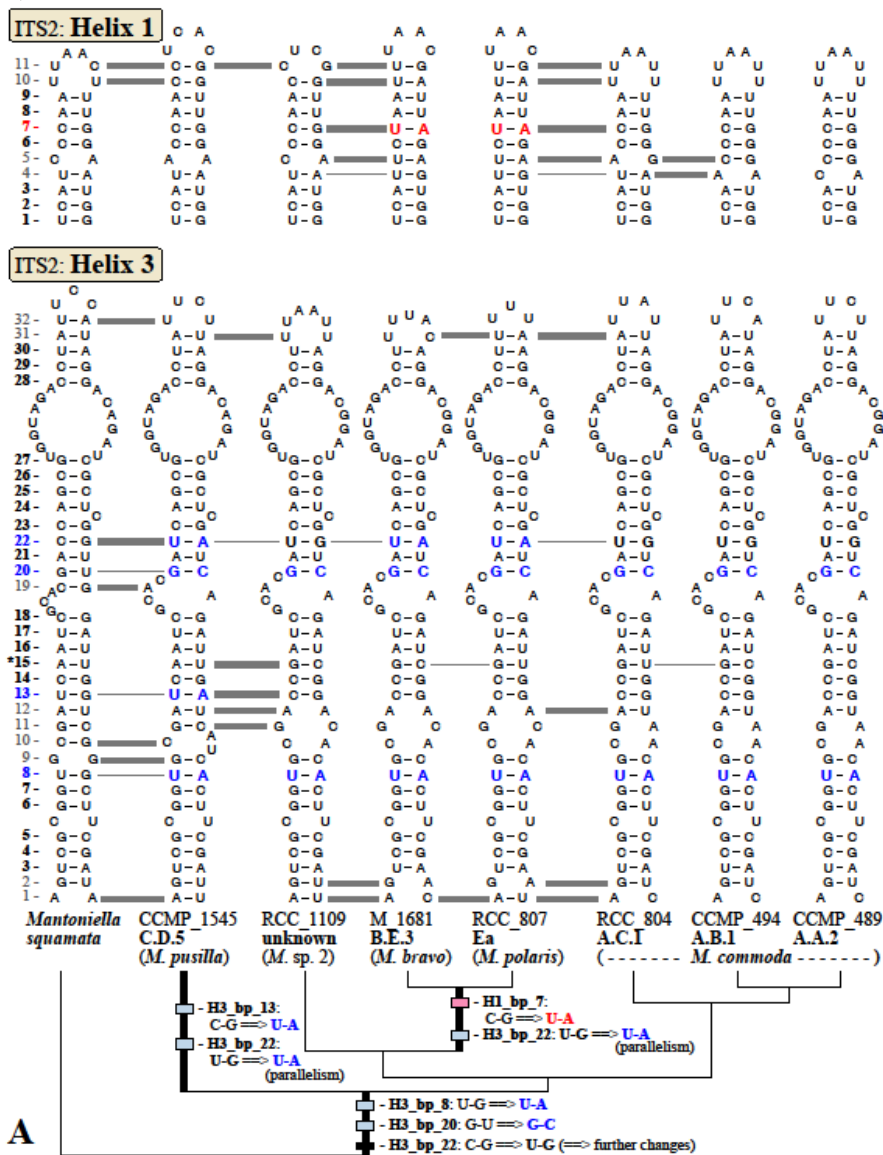
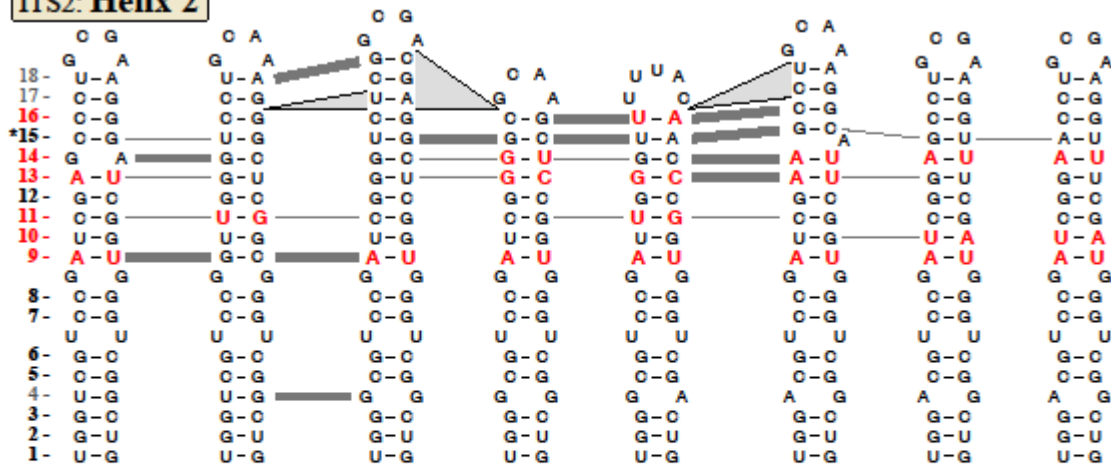


Figure 4. Molecular signatures of *Micromonas* species revealed by comparison of ITS2 secondary structures. Helices 1 and 3 (A) and 2 and 4 (B) of *Mantoniella* and *Micromonas* are shown. All base pairs are numbered, with numbers of universal base pairs (= paired in all members of *Micromonas*) in **bold**, and non-universal pair numbers in grey. Double-sided CBCs (compensatory base changes) vs. hemi-CBCs are highlighted by thick vs. thin grey lines, whereas base pairing/dissociation events are indicated by dotted lines. In Helix 2, length differences may be explained by double-sided indel events (grey triangles). Synapomorphic signatures of *Micromonas* clades were identified with PAUP using an alignment of all Mamiellophyceae (Marin and Melkonian 2010), and mapped on those branches of the phylogenetic tree where they occurred. Base pairs with clear synapomorphies are highlighted with colors (pink and blue respectively for helices 1 and 2 in A and 3 and 4 in B) while all branches with synapomorphy support are highlighted in bold. Several universal base pairs showed too many changes (CBCs and hemi-CBCs), suggesting alternative, equally likely explanations for their evolution, and therefore could not be unambiguously mapped upon the tree; these hypervariable pairs were flagged by an asterisk (*), and were not used as clade signatures. The same holds for two base pairs in Helix 4 (indicated by two asterisks**), where the precise secondary structure remained ambiguous due to presence of adjacent unpaired nucleotides. Tracing base pair evolution in the stem regions (= helices) by CBCs and hemi-CBCs revealed clear molecular signatures for clades (unique synapomorphies within *Micromonas*) as well as several homoplasious changes (e.g. parallelisms, convergences and reversals) in helices 2, 3 and 4.



ITS2: Helix 2



ITS2: Helix 4

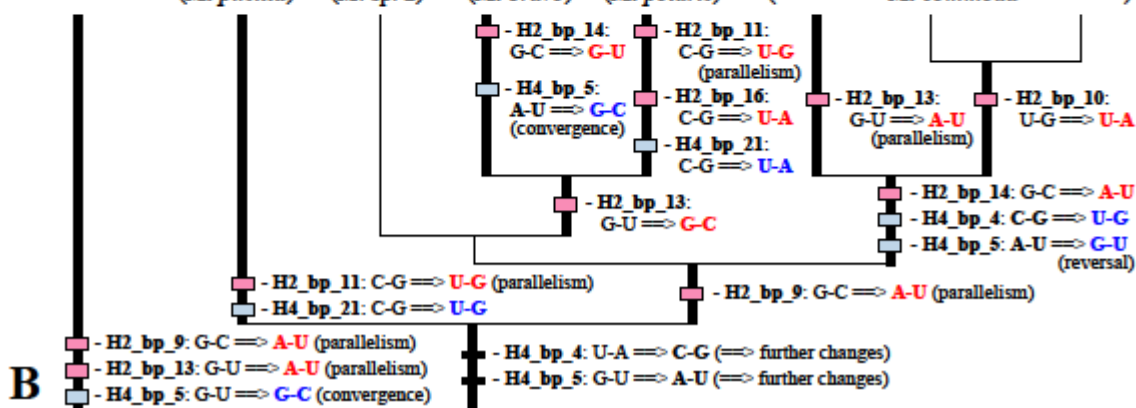
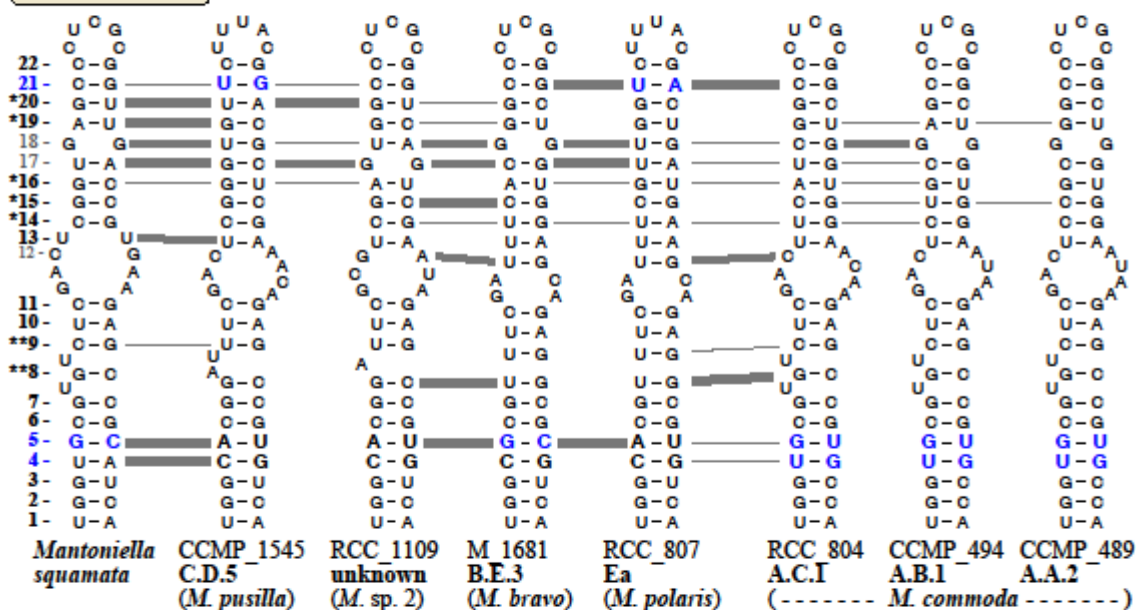


Figure 5. Transmission electron microscopy pictures of flagellar extremity of *Micromonas* strains showing the tip hairs. (A) RCC449. Arrows indicate the double tip hairs. (B, C) RCC 372. Details of the tip hairs in negative stained samples. Scale bar (A) = 200 nm, (B) = 50 nm, (C) = 20 nm.

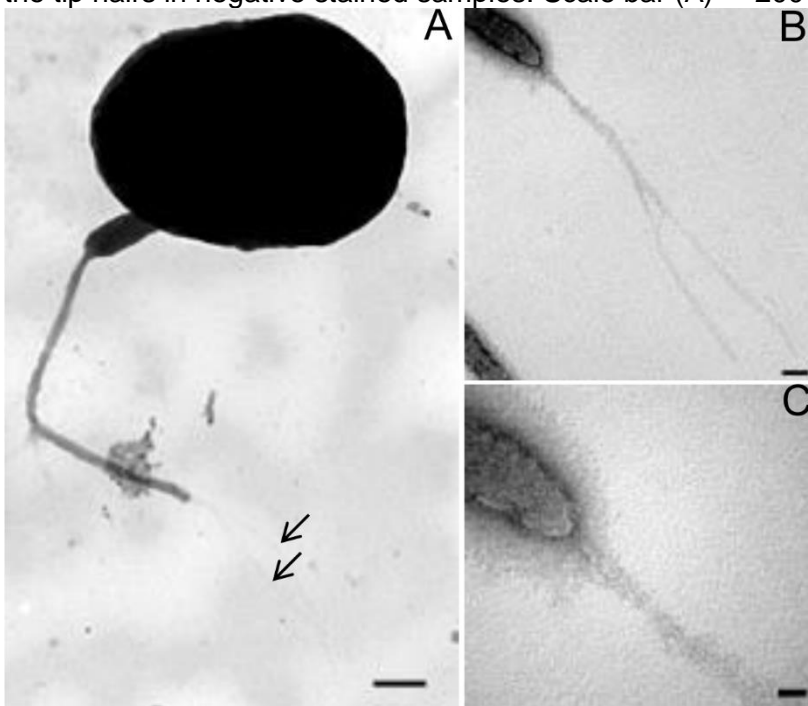


Figure 6. Scanning (A-D) and transmission (E-H) electron microscopy pictures of *Micromonas* spp. cells. (A) *Micromonas commoda*, RCC299. (B, E) *Micromonas bravo*, RCC434. (C, G) *Micromonas* candidate species 1, RCC1109. (D, H) *Micromonas pusilla*, RCC834. (F) *Micromonas polaris*, RCC2306.

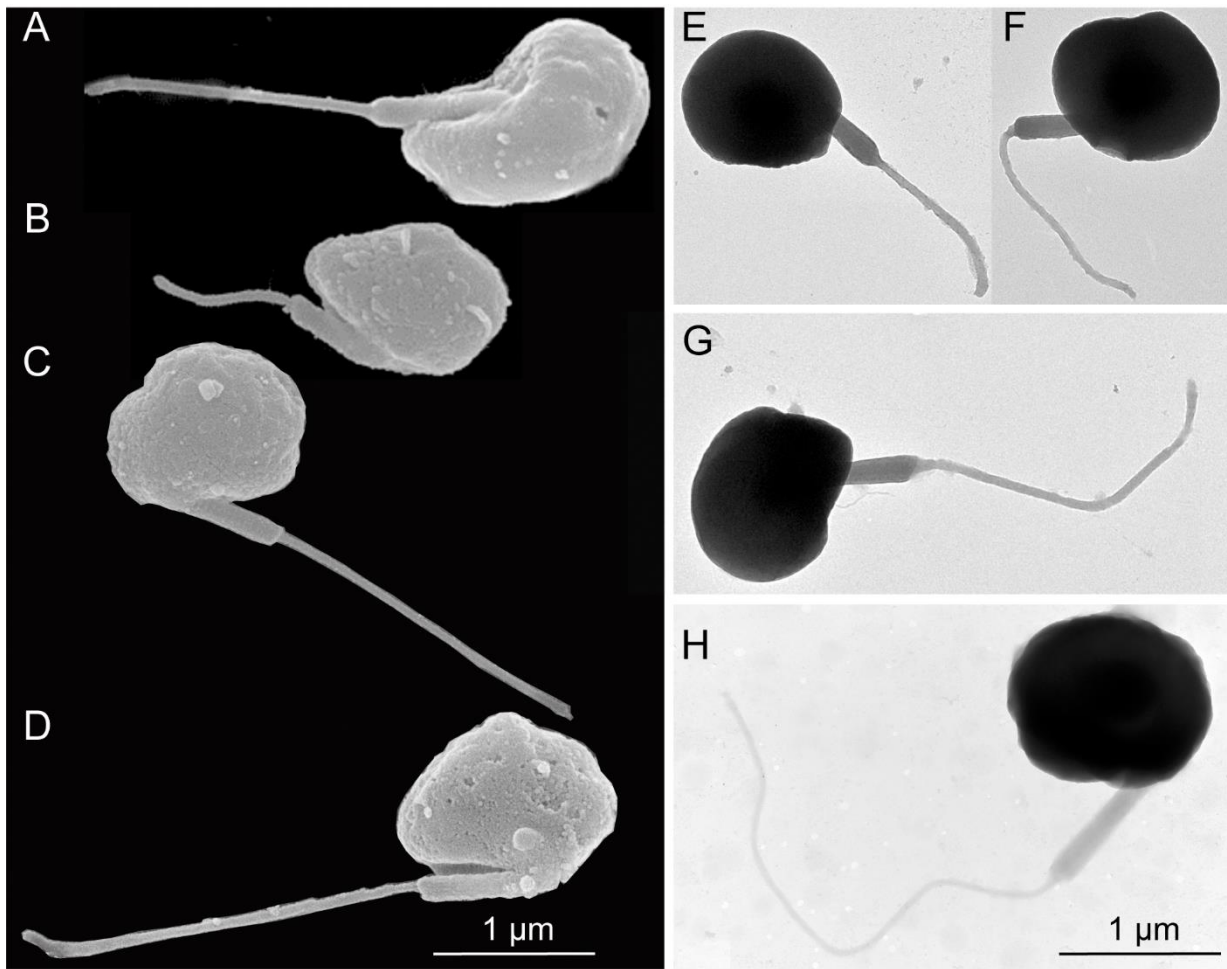


Figure 7. Flagellar length variations in *Micromonas* strains. (A) and (B) TEM and LM pictures of *Micromonas pusilla* strain RCC834. Arrows indicate the extremities of the structure measured (hair point and true flagellum). Scale bars (A) = 200 nm and (B) = 2 μ m. (C) and (D) Flagellar length variations in *Micromonas* strain RCC834 as estimated with TEM and LM along (C) the growth curve and (D) a cell division cycle. (E) Box plots showing flagellar length variations among cells, strains and clades. The median, first and last quartiles, as well as minimum and maximum lengths are shown (number of cells measured per strain is between 56 and 66).

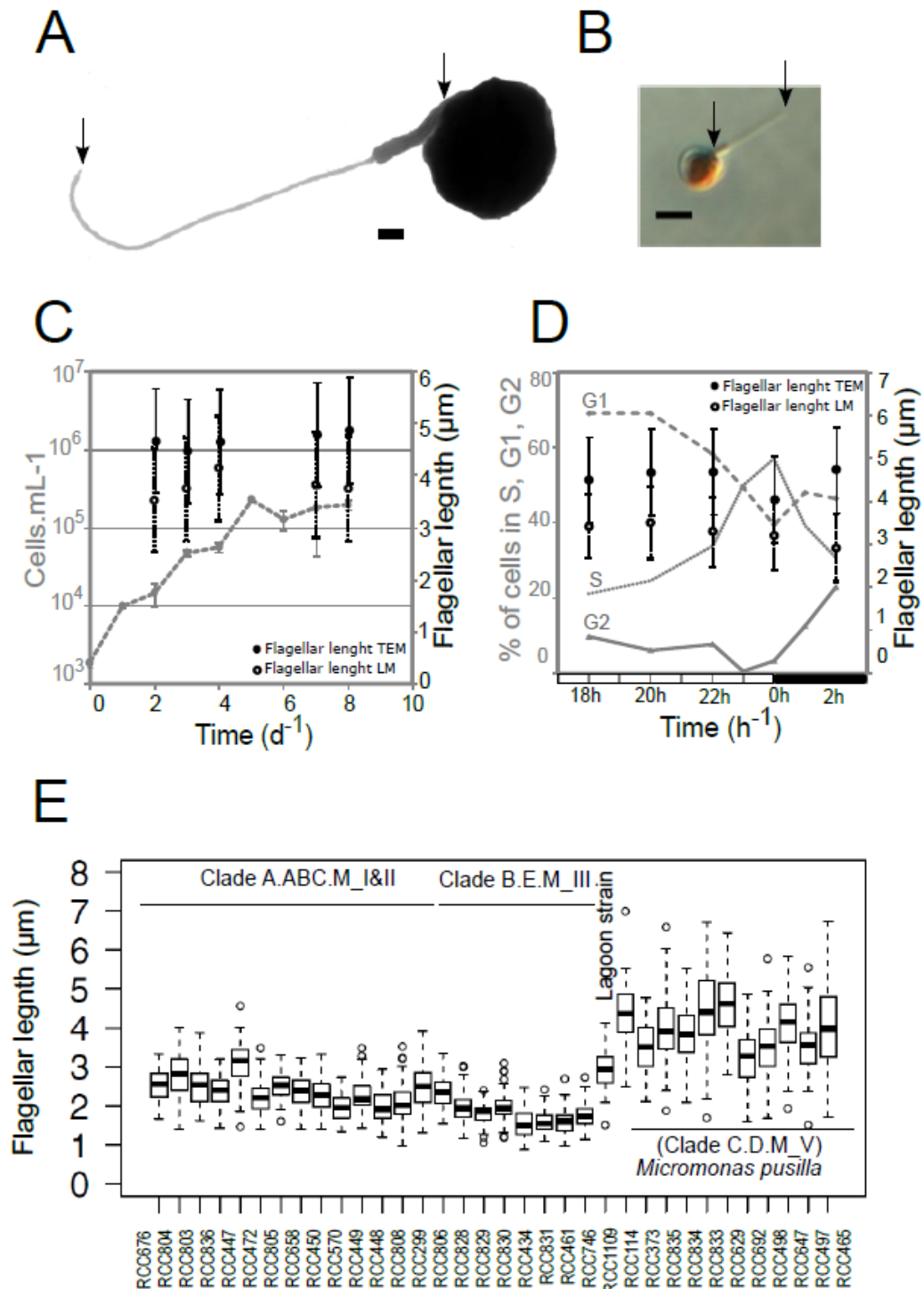


Figure 8. Distribution of stations from which *Micromonas* spp. isolates (**A**) or environmental sequences (**B**) have been reported. Environmental sequences were retrieved from the PR² database (Guillou et al. 2013). Environmental sequences that did not fall into clades identified as species or candidate species were assigned to the category *Micromonas* sp. These figures are also available as interactive maps at <https://tinyurl.com/krvumys>.

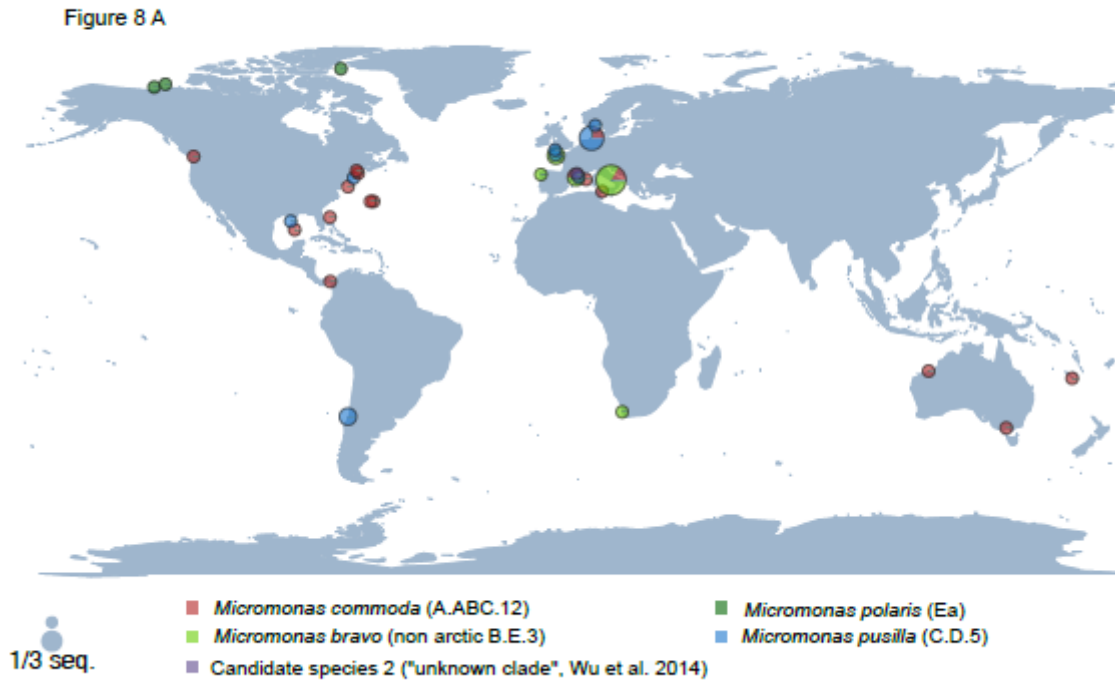
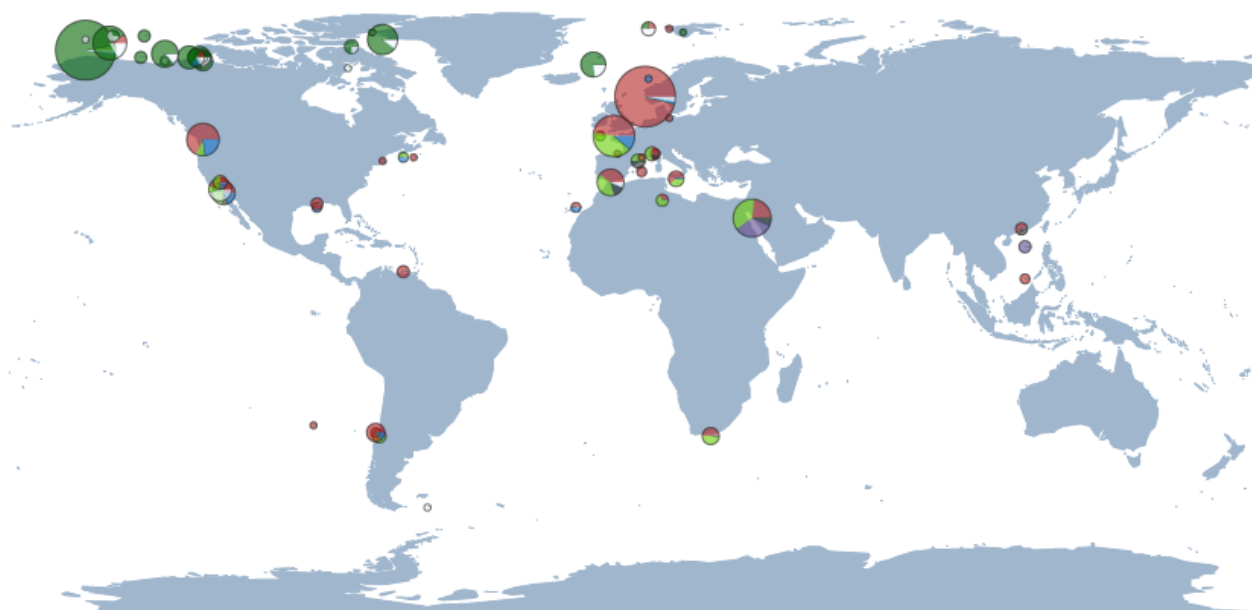


Figure 8 B




 1/10/50
 sequences

- *Micromonas commoda* (A.ABC.12)
 - *Micromonas bravo* (non arctic B.E.3)
 - Candidate species 1 (B._4)
 - Candidate species 2 ("Unknown clade", Wu et al., 2014)
- *Micromonas polaris* (arctic B.E.3)
 - *Micromonas pusilla*
 - Micromonas* sp.

Tables

Table 1: Names or codes used in the literature to designate the infrageneric entities distinguished within the genus *Micromonas* since 2004. The codes created by Worden (2006) combine identifiers used by Guillou et al. (2004), Slapeta et al. (2006) and their own study (Guillou Clade.Slapeta Clade(s).Worden Clade).

This study	Guillou et al. (2004)	Slapeta et al. (2006)	Lovejoy et al. (2007)	Worden (2006)	Worden et al. (2009)	Wu et al. (2014)	Simmons et al. (2015)	Van Baren et al. (2016)
<i>Micromonas commoda</i>	A	A		A.A. 2	M_II			<i>M. commoda</i>
		B		A.BC. 1	M_I			
		C			M_I			
<i>Micromonas bravo</i>	B	E		B.E. 3	M_III		E1	
<i>Micromonas polaris</i>			Ea				E2	
<i>Micromonas</i> candidate species 1				B._. 4	M_IV			
<i>Micromonas</i> candidate species 2						Unknown clade		
<i>Micromonas pusilla</i>	C	D		C.D. 5	M_V			

Table 2: *Micromonas* strains included in this work. Each line of the table corresponds to a single isolate, except for RCC834 and 835 which were obtained from the same original isolate. Strain names in the Roscoff Culture Collection (RCC) and original culture collections are provided. Alternative names given to strains derived from original isolates are also provided. Culture collections include: the National Center for Marine Algae and Microbiota (NCMA) formerly Provasoli-Guillard National Center for Culture of Marine Phytoplankton (CCMP), the National Institute of Technology and Evaluation Biological Resource Center (NBRC = MBIC), the North East Pacific Culture Collection (NEPCC), the CSIRO Collection of living Microalgae (CS) and the Culture Collection of Algae at the University of Cologne (CCAC). GenBank accession numbers are indicated when available. Clades are named using a three letter code after Guillou et al. (2004), Slapeta et al. (2006) and Worden (2006). Strains for which a new sequence was obtained in the frame of this study are in bold. Accession numbers in bold and underlined were obtained respectively from strains in bold and underlined. Data concerning isolation conditions were retrieved from culture collections except for depth of RCC806 (Zingone, pers. com.). - = data not available. Strains with a grey background correspond to type or neotype strains of species. See Fig. 7A for a map of all strains.

Species	Clade	RCC N°	Other strain name	Isolation region	Geographic coordinates	Isolation depth	Isolation date	SSU rDNA Accession number	ITS, 5.8S Accession number	SSU rDNA – ITS Accession number
<i>M. pusilla</i>	C.D.5	114	DW8, CCMP490	North Atlantic Ocean	70°40' W - 41°31' N	-	18-06-1964	AY425320, KU244630	KU244631	AY955003, AY955004
<i>M. pusilla</i>	C.D.5	373		Skagerrak, Baltic Sea	9°6' E - 58°11' N	-	-			KF501025
<i>M. pusilla</i>	C.D.5	465	RA010613-65-8	English Channel	3°57' W - 48°45' N	surface	13-06-2001			KF501028
<i>M. pusilla</i>	C.D.5	497	BL_105-7	Mediterranean Sea	3°33' E - 41°43' N	1800m	-			KF501034
<i>M. pusilla</i>	C.D.5	498	BL_74-8	Mediterranean Sea		surface	28-02-2001	AY665979 (partial) KU244647	KU244648	
<i>M. pusilla</i>	C.D.5	629	He010117-D1-D5	North Sea	7°54' E - 54°11' N	surface	17-01-2001			KF501029
<i>M. pusilla</i>	C.D.5	647	He010418-I1-B6	North Sea	7°54' E - 54°11' N	surface	18-04-2001	KU244651	KU244652	
<i>M. pusilla</i>	C.D.5	692	He010619-D2-C6	North Sea	7°54' E - 54°11' N	surface	19-06-2001	KU244655	KU244656	
<i>M. pusilla</i>	C.D.5	833	M1 E21	Gulf of Mexico	91°18' W - 27°42' N	275m	20-04-2004	KU244675	KU244676	
<i>M. pusilla</i>	C.D.5	834	PLY27, CCMP 1545, UTEX LB991, CCAP LB 1965/4, MpUK	English Channel	4°13' W - 50°15' N	surface	13-04-1950	KU244677	KU244678	AY954994 From genome (http://genome.jgi.doe.gov/)
<i>M. pusilla</i>	C.D.5	835	PLY27, CCMP 491, UTEX	English Channel	4°13' W - 50°15' N	surface	13-04-1950	KU244679	KU244680	AY955005

			LB99, CCAP LB 1965, CS-98								
<i>M. commoda</i>	A.A.2	299	NOUM17, CCMP2709	South Pacific Ocean	166°20' E - 22°20' S	surface	10-02-1998	HM191693, KU244632	KU244633	From genome (http://genome.jgi.doe.gov/	
<i>M. commoda</i>	A.A.2	448	CCMP487, IB4	Sargasso Sea	64°0' W - 34°0' N	25m	07-07-1980	KU244639	KU244640		
<i>M. commoda</i>	A.A.2	449	CCMP1723, Mp7/1	Mediterranean Sea	11°24' E - 37°12' N	-	15-05-1992	KU244641	KU244642	AY954997	
<i>M. commoda</i>	A.A.2	450	CCMP489, D120M	Sargasso Sea	64°22' W - 28°59' N	120m	30-05-1988	AJ010408 KU244643	KU244644	AY955002	
<i>M. commoda</i>	A.A.2	451	CCMP492, IIE1	North Atlantic	72°22' W - 38°42' N	25m	11-07-1980			AY955006, KF501035	
<i>M. commoda</i>	A.A.2	570	BL_151	Mediterranean Sea	2°48' E - 41°40' N	surface	25-06-2001	KU244649	KU244650		
<i>M. commoda</i>	A.A.2	805	MBIC 10095, H02-10, NBRC 102743	Pacific Ocean	-	-	-	AB183589 KU244663	KU244664	AY955011	
<i>M. commoda</i>	A.A.2	-	CCMP 493, IV03	Gulf of Mexico	90° W - 25° N	-	15-02-1981			AY955007	
<i>M. commoda</i>	A.A.2	808	CCMP 488, IE2	Sargasso Sea	65°0' W - 34°0' N	25m	07-07-1980	KU244669	KU244670	AY955001	
<i>M. commoda</i>	A.B.1	472	CCMP1764, M97-105	Pacific Ocean	78°13' W - 8°30' N	-	01-03-1997	KU244645	KU244646	AY954998	
<i>M. commoda</i>	A.B.1	658	CS-170	Pacific Ocean	-	-	01-01-1982			AY955009, KF501030	
<i>M. commoda</i>	A.B.1	-	CCMP494	North Atlantic	69° W-43° N		01-03-1982			AY955008	
<i>M. commoda</i>	A.C.1	372	Naples	Mediterranean Sea	-	-	01-01-1986	KU244634	KU244635		
<i>M. commoda</i>	A.C.1	447	CCMP1195, CCMP9	North Atlantic Ocean	69°38' W - 43°50' N	-	19-11-1986	KU244638		AY954993	
<i>M. commoda</i>	A.C.1	676	H0400U-D4	North Sea	7°54' E - 54°11' N	surface	2000	KU244653	KU244654		
<i>M. commoda</i>	A.C.1	803	CS-222	Southern Ocean	-	-	01-01-1988	KU244659	KU244660	AY955010	
<i>M. commoda</i>	A.C.1	804	NEPCC 29	Pacific Ocean	-	-	-	KU244661	KU244662	AY955012	
<i>M. commoda</i>	A.C.1	836	Mnorbal	Mediterranean Sea	6°5' E - 41°0' N	5m	09-03-2003	KU244681	KU244682		
<i>M. bravo</i>	B.E.3	418	RD010614-71-1	English Channel	3°51' W - 48°37' N	surface	14-06-2001			KF501026	
<i>M. bravo</i>	B.E.3	434	BL_122	Mediterranean Sea	2°48' E - 41°40' N	surface	20-03-2001	AY425316 KU244636	KU244637		
<i>M. bravo</i>	B.E.3	461	RD010614-71-4	English Channel	3°51' W - 48°37' N	surface	14-06-2001			KF501027	
<i>M. bravo</i>	B.E.3	746	A1_Arousa 90, Arousa 2	Atlantic Ocean	8°49' W - 42°35' N	10m	10-10-1990	KU244657	KU244658		
<i>M. bravo</i>	B.E.3	806	CCMP 1646	Mediterranean Sea	14°20' E - 40°45' N	surface *	08-04-1993	KU244665	KU244666	AY954995, AY954996	
<i>M. bravo</i>	B.E.3	828	Mp1	Mediterranean Sea	-	-	01-06-1993			KF501032	
<i>M. bravo</i>	B.E.3	829	MpPart	Mediterranean Sea	-	-	08-01-1997			KF501033	
<i>M. bravo</i>	B.E.3	830	MpCO	Mediterranean Sea	-	-	21-11-1996	KU244671	KU244672		
<i>M. bravo</i>	B.E.3	831	Mp2	Mediterranean Sea	-	-	24-11-1993	KU244673	KU244674		

<i>M. bravo</i>	B.E.3	-	CCAC 1681 B (M 1681)	W. Cape, South Africa	18°0' E - 33°2' S	surface	06-1996			FN562452
<i>M. polaris</i>	Ea	807	CCMP 2099, MicroWC7343	Arctic Ocean	74°45' W - 76°17' N	55m	04-07-1998	DQ025753 KU244667	KU244668	AY954999, AY955000
<i>M. polaris</i>	Ea	2308	MALINA S623	Arctic Ocean	70°41' W -136°3' N	56m	08-19-2009	JN934683	KT860583	
<i>M. polaris</i>	Ea	2306	MALINA S522	Arctic Ocean	71°24' W -132°40' N	70m	08-15-2009	JF794057	KT860582	

Table 3: Minimum and maximum % of differences (p-distance) between SSU rDNA sequences of *Micromonas* clades and subclades and *Mantoniella* species. The analysis included the 97 sequences from the phylogenetic tree presented in Figure 1. Analysis was conducted using MEGA v6 (Tamura et al. 2013). All positions containing gaps and missing data were removed. The final dataset contained 1559 positions. Clades were distinguished based upon phylogenetic analyses based on SSU rRNA gene and ITS2 (see text) and are named as in Figure 1. *Micromonas* sp. 1 and 2 are candidate species (see text). The % of differences obtained between and within species or candidate species described in this study are respectively in bold and grey. Values obtained between sequences of *Micromonas* and *Mantoniella* are in bold and underlined. * Wu et al. (2014)

Genus	Species	Clade	<i>Micromonas</i>								<i>Mantoniella</i>	
			<i>M. commoda</i> (clade A.ABC.12)	<i>M. polaris</i>	<i>M. bravo</i>	<i>M. sp. 2</i>	<i>M. sp. 1</i>	<i>M. pusilla</i>	<i>M. squamata</i>	<i>M. antarctica</i>		
			A.A.2	A.B.1	A.C.1	Ea	Non-arctic B.E.3	Unknown clade*	B._.4	C.D.5		
<i>Micromonas</i>	<i>M. commoda</i> (A.ABC.12)	A.A.2	0.0-0.9	0.1-0.8	0.3-1.1	2.4-3.3	2.5-3.5	1.9-2.6	1.9-2.5	1.9-2.8	2.1-2.6	2.3-2.8
		A.B.1		0.0-0.3	0.1-0.8	2.4-3.1	2.6-3.2	1.8-2.2	1.8-2.1	1.9-2.6	2.0-2.2	2.3-2.6
		A.C.1			0.0-0.7	2.5-3.3	2.5-3.4	1.7-2.2	1.8-2.3	1.8-2.7	1.9-2.3	2.2-2.6
	<i>M. polaris</i>	Ea				0.0-0.4	0.4-1.3	1.6-2.1	1.9-2.2	2.6-3.4	2.7-3.0	2.9-3.2
	<i>M. bravo</i>	Non arctic B.E.3					0.0-1.1	1.5-2.0	2.1-2.6	2.4-3.2	2.6-3.1	2.6-3.2

	<i>M. sp. 2</i>	Unknwon clade*						0.0-0.2	1.2-1.3	2.1-2.6	1.9-2.0	2.3-2.4
	<i>M. sp. 1</i>	B._.4							0.2	2.3-3.0	2.1	2.6
	<i>M. pusilla</i>	C.D.5								0.0-0.7	1.5-1.9	1.7-2.1
<i>Mantoniella</i>	<i>M. squamata</i>										-	0.6
	<i>M. antarctica</i>											-

Table 4. Synapomorphy support for the genus *Micromonas* and its sub-clades in the 18S rRNA molecule and the second internal transcribed spacer (ITS2). For 18S rRNA synapomorphies, all homoplasious changes (parallelisms) of unrelated Viridiplantae are listed. Numbering of 18S rRNA helices after the European ribosomal RNA database (<http://web.archive.org/web/20110208210644/http://bioinformatics.psb.ugent.be/webtools/rRNA>); for ITS2 helices see Figure 4. The high sequence diversity of ITS2 sequences precluded alignments beyond members of the class Mamiellophyceae, and the analysis of ITS2 base pairs was therefore confined to this class.

Taxon/character	Synapomorphy	Homoplasies
<i>Micromonas</i>		
18S rRNA - Helix 11: bp 1	U-A ==> C-G	C-G parallel in <i>Acrosiphonia</i> , <i>Nautococcus</i>
<i>Micromonas pusilla</i> (C.D.5)		
18S rRNA – Helix 25: bp 10	U-A ==> A-U	A-U parallel in <i>Parachlorella</i> spp., <i>Heterotetracystis akinetos</i> , prasinophyte CCMP 1205
ITS2 – Helix 3: bp 13	C-G ==> U-A	U-A unique within <i>Micromonas</i>
ITS2 – Helix 4: bp 21	C-G ==> U-G	U-G unique within <i>Micromonas</i>
<i>Micromonas commoda</i> (A.ABC.12)		
ITS2 – Helix 2: bp 14	G-C ==> A-U	A-U unique within <i>Micromonas</i>
ITS2 – Helix 4: bp 5	A-U ==> G-U	G-U unique within <i>Micromonas</i>
Subclades A.A.2 A.B.1		
ITS2 – Helix 2: bp 10	U-G ==> U-A	A unique within <i>Micromonas</i>
<i>M. bravo</i>, <i>M. polaris</i> and candidate species 1 and 2		

(B.E.3, unknown clade and B._.4)		
18S rRNA – Helix 10: bp 3	C-G ==> U-A	U-A parallel in <i>Pseudoscourfieldia</i>
18S rRNA – Helix 29: bp 7	G-C ==> C-G	C-G parallel in <i>Leptosira</i> , Sphaeropleaceae
<i>M. bravo</i> and <i>M. polaris</i> (B.E.3)		
ITS2 – Helix 1: bp 7	C-G ==> U-A	U-A unique within <i>Micromonas</i>
ITS2 – Helix 2: bp 13	G-U ==> G-C	G-C unique within <i>Micromonas</i>
<i>Micromonas bravo</i> (non-arctic B.E.3)		
ITS2 – Helix 2: bp 14	G-C ==> G-U	G-U unique within <i>Micromonas</i>
ITS2 – Helix 4: bp 5	A-U ==> G-C	G-C unique within <i>Micromonas</i>
<i>Micromonas polaris</i> (Ea)		
ITS2 – Helix 2: bp 16	C-G ==> U-A	U-A unique within <i>Micromonas</i>
ITS2 – Helix 4: bp 21	C-G ==> U-A	U-A unique within <i>Micromonas</i>
Candidate species 1 (B._.4)		
18S rRNA - Helix 11: bp 4	C-G ==> U-G	U-G parallel in <i>Ostreococcus</i>
18S rRNA - Helix E23_1: bp 6	A-U ==> C-G	C-G parallel in e.g. <i>Tetracystis/Chlorococcum</i> , <i>Chlorosarcinopsis</i> , <i>Chlamydomonas</i> spp. (e.g. <i>C. moewusii</i>)