Protist, Vol. 164, 643–659, xx 2013
http://www.elsevier.de/protis
Published online date xxx

Protist

ORIGINAL PAPER

Morphology, Genome Plasticity, and Phylogeny in the Genus *Ostreococcus* Reveal a Cryptic Species, *O. mediterraneus* sp. nov. (Mamiellales, Mamiellophyceae)

Lucie Subirana[a,b], Bérangère Péquin[c], Stéphanie Michely[d], Marie-Line Escande[a,b], Julie Meilland[a,b,2], Evelyne Derelle[a,b], Birger Marin[e], Gwenaël Piganeau[a,b], Yves Desdevises[a,b], Hervé Moreau[a,b], and Nigel H. Grimsley[a,b,1]

[a]CNRS, UMR7232 Laboratoire de Biologie Intégrative des Organisms Marins (BIOM), Observatoire Océanologique, Banyuls-sur-Mer, France
[b]UPMC Univ Paris 06, Observatoire Océanologique, Banyuls-sur-Mer, France
[c]Québec-Océan, Département de Biologie and Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec (QC) G1 V 0A6, Canada
[d]INRA UMR1319, AgroParisTech, Micalis, Bimlip, Bât. CBAI, BP 01, 78850 Thiverval-Grignon, France
[e]Biozentrum Köln, Botanisches Institut, Universität zu Köln, Zülpicher Str. 47b, 50674 Köln, Germany

Submitted December 4, 2011; Accepted June 18, 2013
Monitoring Editor: David Moreira

Coastal marine waters in many regions worldwide support abundant populations of extremely small (1-3 μm diameter) unicellular eukaryotic green algae, dominant taxa including several species in the class Mamiellophyceae. Their diminutive size conceals surprising levels of genetic diversity and defies classical species' descriptions. We present a detailed analysis within the genus *Ostreococcus* and show that morphological characteristics cannot be used to describe diversity within this group. Karyotypic analyses of the best-characterized species *O. tauri* show it to carry two chromosomes that vary in size between individual clonal lines, probably an evolutionarily ancient feature that emerged before species' divergences within the Mamiellales. By using a culturing technique specifically adapted to members of the genus *Ostreococcus*, we purified >30 clonal lines of a new species, *Ostreococcus mediterraneus sp. nov.*, previously known as *Ostreococcus* clade D, that has been overlooked in several studies based on PCR-amplification of genetic markers from environment-extracted DNA. Phylogenetic analyses of the S-adenosylmethionine synthetase gene, and of the complete small subunit ribosomal RNA gene, including detailed comparisons of predicted ITS2 (internal transcribed spacer 2) secondary structures, clearly support that this is a separate species. In addition, karyotypic analyses reveal that the chromosomal location of its ribosomal RNA gene cluster differs from other *Ostreococcus* clades.

Key words: Chromosome; karyotype; culture; ribosomal gene; barcode; picoeukaryote; ITS2.

[1]Corresponding author;
[2]Current address: CNRS, UMR6112, LPG-BIAF, Bio-Indicateurs Actuels et Fossiles, Université Angers 2 Boulevard Lavoisier, 49045 Angers CEDEX 01, France
e-mail nigel.grimsley@obs-banyuls.fr (N.H. Grimsley).

http://dx.doi.org/10.1016/j.protis.2013.06.002

Introduction

Although bacteria represent the largest biomass of marine life, photosynthetic eukaryotic protists nevertheless account for a large proportion of the oceans' primary production, largely due to their rapid turnover time (Li et al. 1992; Stockner 1988). Within the extremely diverse protistan marine community, picoeukaryotes of the Chlorophyta are present worldwide, and can represent a high proportion of eukaryotic plankton, particularly in coastal regions (Gross 1937; Knight-Jones and Walne 1951; Lovejoy 2007; Manton and Parke 1960; Worden et al. 2004; see Massana 2011 for a recent review). The abundance of particular species or ecotypes may be governed by their adaptations to the local environment, and cryptic species may exist (Cardol et al. 2008; Foulon et al. 2008; Jancek et al. 2008; Lovejoy and Potvin 2010; Rodriguez et al. 2005; Worden et al. 2009), but little information exists about their population structures and why such sympatric communities of cryptic species of phytoplankton exist. Grimsley et al. (2010) showed that genetic recombination between strains, a hallmark of sexual reproduction in the large sense, must be inferred in the genealogy of wild-type *O. tauri* cultures to explain the distribution and sequences of neutral genetic markers. We are particularly interested in the Mamiellophyceae (Marin and Melkonian 2010) because (i) they are distributed worldwide, as documented by sequence data from analyses of environmental DNA extractions, especially for the genera *Bathycoccus*, *Micromonas* and *Ostreococcus* (Massana 2011; Slapeta et al. 2006; Vaulot et al. 2008; Viprey et al. 2008) in the order Mamiellales, (ii) several complete genome sequences of species from this group are available (reviewed in Piganeau et al. 2011a) permitting detailed phylogenetic and evolutionary comparisons to be made, (iii) they can be grown easily as clonal cultures, facilitating genetic and physiological analyses. Numerous examples of these three genera are maintained in culture collections (Vaulot et al. 2004).

The genus *Ostreococcus* was initially described using one Mediterranean strain, *Ostreococcus tauri* (Chrétiennot-Dinet et al. 1995; Courties et al. 1994). Subsequently, several more strains presenting a similar morphology were isolated, but DNA sequence analysis of their small subunit ribosomal RNA gene including the more variable ITS sequences (two internal transcribed spacer regions, separating the three rRNA genes) revealed that the genus *Ostreococcus* should be divided in four clades A, B, C and D (Guillou et al. 2004; Rodriguez et al. 2005). This analysis strongly suggested a complex of cryptic species, the initial species *O. tauri* belonging to the clade C, although this hypothesis could not be tested experimentally since the sexual cycle is unknown. The complete genome of *O. tauri* was then analysed (Derelle et al. 2006). When a second strain isolated from Pacific was described (now known as *O. lucimarinus*, type member of clade A), its karyotype and its genome sequence were clearly too divergent for it to interbreed with *O. tauri*, leading to its designation as *O. lucimarinus* (Palenik et al. 2007; without valid taxonomic description, this name is currently a 'nomen nudum'). Many other strains were then isolated from various oceanic areas, and their ribosomal gene sequences confirmed their classification in four clades. These clades may represent groups with differing environmental adaptations, clade B strains being "low light" adapted strains, whereas the others may be "high light" or "polyvalent light" strains (Rodriguez et al. 2005). This association between clades and adaptations was confirmed recently by Demir-Hilton et al. (2011) who also showed that co-occurrence of both ecotypes (which we now know are probably different species) at the same geographical location is rare, and that factors explaining clade distribution were more complex than irradiance alone. They proposed that these two "low light" and "high light" "ecotypes" might better be described as oceanic and coastal clades/species, respectively. Until now, the two clades C and D have only been found in the Mediterranean whereas strains of the two other clades A and B have been isolated from various oceanic origins.

In order to describe the population structure of a species, numerous wild-type individuals or clones should be isolated to assess the level of intraspecific polymorphism. Different, independently isolated, wild-type clones of *Ostreococcus spp.* may belong to any one of four phylogenetically distinct clades that certainly represent different species, because the extent of the differences seen in their genome characteristics such as the level of sequence homology (Jancek et al. 2008; Palenik et al. 2007) or the number and size of chromosomes, preclude the possibility of conventional meiosis occurring. Since they may show no clear phenotypical differences, even when examined by electron microscopy, molecular markers must be used to describe individuals and to determine whether intraspecific genetic exchanges, the hallmark of a biologically defined species, occur. We have already isolated several clonal lines of *O. tauri* and inferred that sexual recombination occurs

by genetic analyses of the segregation of DNA sequence polymorphisms between strains in the genealogy of 19 strains from a natural population (Grimsley et al. 2010). Sexual cycles have not yet been described in the Mamiellophyceae, but the frequency of meiosis is probably as low as it is in wild yeasts (Tsai et al. 2008). This would not be surprising because the genetic systems of these species are rather similar, involving haploid unicellular organisms, most likely with defined mating types, dispersed over geographically large regions, often at relatively low cell densities, that grow mainly by clonal division.

Here, we present the isolation and culture of more than 30 new clade D strains of *Ostreococcus*, previously thought to be rare, using techniques that specifically favour the growth of very small autotrophic picoeukaryotes. We make detailed electron microscopical examinations of individual strains from different clades of *Ostreococcus*, characterize their karyotypes by pulsed field electrophoresis (PFGE), and present a detailed comparison of ITS2 secondary structures. These morphological and molecular analyses led us to propose a species definition within the genus *Ostreococcus*, and to suggest some guidelines for discriminating between species in very small unicellular eukaryotes for which morphological criteria are scarce.

Results and Discussion

Ostreococcus sp. Clades C and D are Common in the NW Mediterranean Sea

We isolated new strains of picoeukaryotic Mamiellophyceae from the N.W. Mediterranean Sea. This region provides an interesting variety of productive seawater habitats that favour the growth of a diverse range of protists. Its coastline is punctuated by innumerable shallow seawater lagoons of varying sizes (Fig. 1) that are alimented by small rivers and are connected to the sea by narrow channels called “grau”. Some of the lagoons thus have variable levels of salinity and the largest, such as Thau and Leucate, are used for culture of oysters. Our strain isolation protocol was designed to maximise the probability of finding picoeukaryotic green algae by filtration of each seawater sample through 1.2 μm filters before addition of appropriate culture media (see Methods). Only cultures becoming visibly green after 1-3 weeks were retained for further analysis. To assure clonality, single colonies were re-isolated from plates (Grimsley et al. 2010).

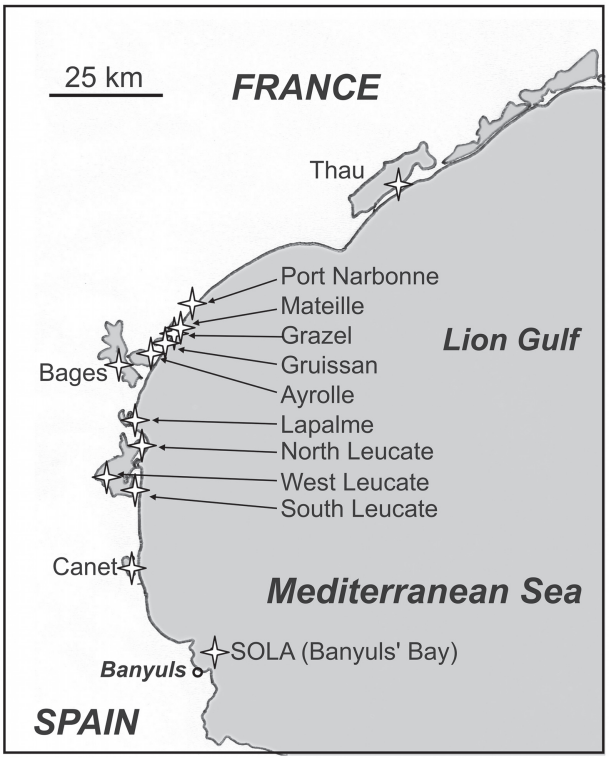


Figure 1. Seawater sample collection sites in the N. W. Mediterranean Sea. Map of the South of France showing the Eastern coastal region adjoining Spain, with the names and places of the marine lagoons where the samples were collected (open stars). Barcelona Harbour (site where the first strain was isolated, Spain, not shown) lies 148 km South West of Banyuls’ Bay along the same coastline.

Over a period of 5 years, a total of 45 clonal lines of *Ostreococcus spp.* were produced from over 200 independent samplings at different N. W. Mediterranean coastal sites (Grimsley et al. 2010 and Supplementary Table S1). The identity of individual clones was determined by amplifying and sequencing the nuclear small subunit ribosomal RNA (18S rRNA) gene, including ITS1, 5.8S rRNA, and ITS2 sequences, and sequence alignment with reference members of the clade in question using BioEdit (Hall 1999). Most of them were either *Ostreococcus tauri* (clade C) (13 clones), or *Ostreococcus* clade D (35 clones), whereas only one clade A strain was isolated. As almost only clades C and D were isolated, one other clade A and one clade B strain from the Roscoff Culture Collection (RCC) and originating from other locations than the Mediterranean were also analysed to compare the genetic distances based on the 18S sequence found inside and between the *Ostreococcus*

Table 1. Number of differences in the 18S gene sequence between the four *Ostreococcus* clades.

Clade Comparisons	Substitutions	Deletions	Difference, %
A-B	15	0	0.86%
A-C	3	0	0.17%
A-D	19	2	1.21%
B-C	14	0	0.81%
B-D	30	2	1.84%
C-D	16	2	1.04%

clades (Table 1 and see below). Several other species of picoeukaryotes were also found, mostly identified provisionally by searching databases for best BLAST hits, including *Bathycoccus prasinos* (three clones), *Nannochloris* sp. (four clones), *Pycnococcus* sp. (one clone) and *Micromonas* sp. (one clone). Only one isolate of *Micromonas* and three strains of *Bathycoccus prasinos* were found despite the relative abundance of these two genera reported in previous PCR-based analyses (Viprey et al. 2008; Zhu et al. 2005). This is not surprising, given the slightly larger size of these cells (about 2 μm) they would have been excluded by the pore size of our filters in the majority of isolations. The complete genome sequence of one of the isolates of *B. prasinos* has now been completely analysed (Moreau et al. 2012).

Among the *Ostreococcus spp* 18S rDNA sequences analysed, the percentage of identity between the clades for the 18S gene varied from 0.17% (three substitutions in 1767 bases) between clades A and C to 1.8% (30 substitutions and 2 deletions) between clades B and D (Table 1), suggesting that the four clades represent at least four different species. In contrast, within-clade polymorphisms were limited to a single base pair substitution in the ITS2 region in both clades C and D, whereas the 18S rRNA region remained strictly identical within a clade.

Subtle Differences in Morphology Distinguish Strains but not Clades of *Ostreococcus*

We grew representatives of the four clades of *Ostreococcus* and made detailed examinations of several hundreds of transmission electron microscope images, to look for small but reproducible differences in their phenotypes (Fig. 2). One clade A (RCC356), one clade B (RCC809), three clade C (RCC1561, RCC1558 and RCC1117, but only one clade C example is shown in Figure 2 to save space) and one clade D (RCC789). *Ostreococcus* strains were grown in parallel under identical culture conditions and reduced light intensity (see Methods, clade B strains cannot be grown at high light intensities). Although the cells of all of these species were about the same size (on average, from 100 randomly chosen electron micrographs, clade A: 1.14±0.33 μm (mean ± SD); clade B: 1.05±0.23 μm; clade C: 1.22±0.31 μm; clade D: 1.27±0.32 μm), and no clear differences in the sizes of their plastid, nuclei and cytoplasmic volumes were visible, small differences were observed when many cells of the same clonal line were examined (Fig. 2). However, such differences were never sufficiently consistent between individual cells of a single strain for us to consider them as reliable criteria to use for species' descriptions. For example, all *Ostreococcus* strains possessed one starch granule inside a single chloroplast in each clade. Both the size of these granules and the proportion of cells where the granule is visible showed significant differences between individual strains (granule size Kruskal–Wallis test, $p = 0.014$, granule presence Chi-square $p = 0.0006$), but less so between the three different clades (granule size Kruskal–Wallis test, $p = 0.13$, granule presence Chi-square $p = 0.03$). Another difference observed between the strains was the presence of secretion granules (see Henderson et al. 2007, for a detailed description of these features), which were abundant in some strains and almost absent in others. However, again, these differences were not clade-dependent and were highly variable between the three strains of clade C (Chi-square $p = 0.001$). The only morphological difference, which has been observed and which seems to be linked to a clade, is an external membrane-like structure outside the cells, usually apparently detached from the plasmalemma of the clade B "low light" strain RCC809 (Fig. 2B), and which was never observed in other clades. In conclusion, the morphological differences observed were unreliable for distinguishing clades because of the levels of variations observed between different clonal lines within clades. Despite identical culture conditions, we could not show that these differences are due to taxonomic affiliation,

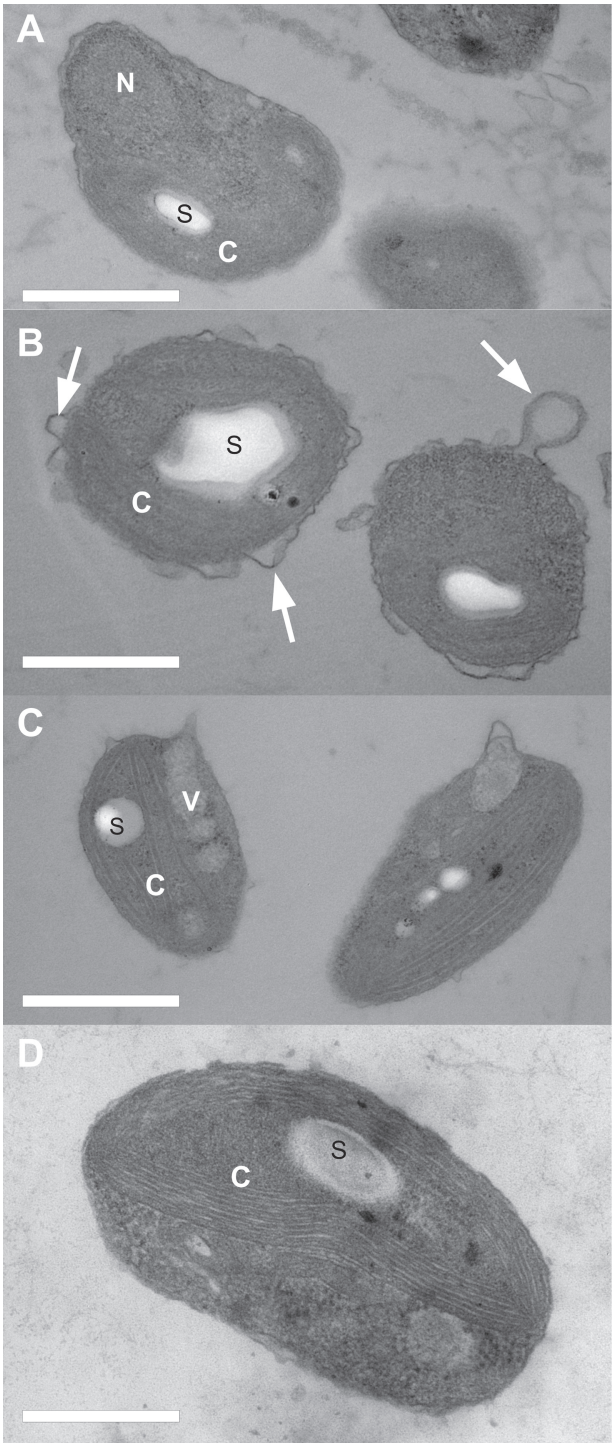
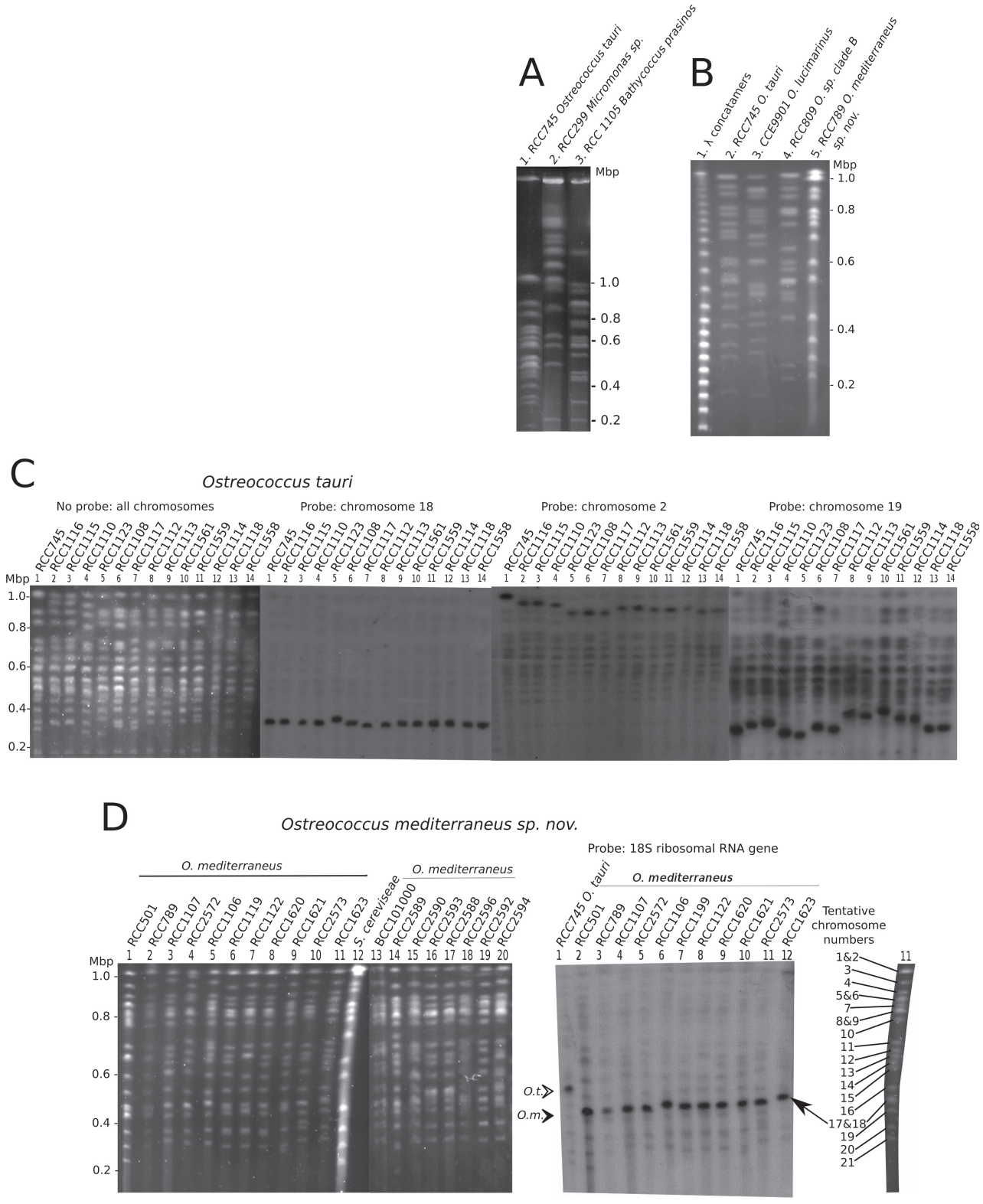


Figure 2. Cell morphologies of different clades of *Ostreococcus spp*. Transmission electron micrographs of four *Ostreococcus spp.* strains: **A** – clade A *O. lucimarinus*; **B** – clade B RCC809 "low light"; **C** – clade C *O. tauri* RCC745; **D** – clade D *Ostreococcus sp.* RCC789 (a clonal line from the original RCC501) *Ostreococcus mediterraneus* sp. nov. (clade D) (Chlorophyta), showing examples of the morphological characteristics that were visible in some strains when many cells of the same clonal line were examined. Although these characteristics were strain-dependent, they were not clade-dependent. N: nucleus, C: chloroplast, S: starch granule, V: secretion granule (vesicle), arrows: protruding external membrane.

but instead propose that they could reflect within-clade diversity in the way each clonal strain in the population of one species responds to the culture conditions imposed in our incubator (individual phenotypes of different members of the same population).

Ostreococcus Strains Show Large Variations in Size of Two Unusual Chromosomes

Pulsed field gel electrophoresis (PFGE) was used to compare the karyotypes of newly isolated strains with reference strains (Fig. 3, panels A to D). Figure 3 panel A shows the large variations in karyotypes observed between different species of the order Mamiellales and panel B compares different reference species within the genus *Ostreococcus* (see also Rodriguez et al. 2005). Chromosome sizes of individual clonal lines of *O. tauri* (panel C) and *O.* sp. clade D (panel D) are then compared. Eighteen of the 20 chromosomes migrated mainly in a similar way within the different strains of clade C, but surprisingly chromosomes 2 and 19 showed remarkably large variations in size (*O. tauri* Fig. 3). Comparing the gel photograph on the left in Figure 3 panels with the autoradiographs on the right, a chromosome 2-specific probe shows that in lane 1 chromosome 2 (usually the 2nd largest chromosome) migrates close to chromosome 1, the largest chromosome, whereas in subsequent lanes it remains the 2nd largest chromosome in all of the subsequent clade C strains, but varies in size between strains (overall, between about 0.97 Mb and 1.1 Mb). Similarly, the right panel shows that chromosome 19, the 2nd smallest, shows enormous size variations between strains, from about 280 to 436 kb. Other chromosomes show mainly much smaller variations in size, for example the smallest chromosome seen close to the bottom edge of Figure 3 panel A, or chromosome 18 migrating around 323 kb seen using a specific radioactive probe (Fig. 3, 2nd panel); these chromosomes should be considered as being about the same size, as small variations might also arise due


A
1. RCC745 Ostreococcus tauri
2. RCC299 Micromonas sp.
3. RCC 1105 Bathycoccus prasinos
Mbp
1.0
0.8
0.6
0.4
0.2
B
1. λ concatamers
2. RCC745 O. tauri
3. CCE9901 O. lucimarinus
4. RCC809 O. sp. clade B
5. RCC789 O. mediterraneus sp. nov.
Mbp
1.0
0.8
0.6
0.4
0.2
C
Ostreococcus tauri
No probe: all chromosomes
Probe: chromosome 18
Probe: chromosome 2
Probe: chromosome 19
RCC745 RCC1116 RCC1115 RCC1110 RCC1123 RCC1108 RCC1117 RCC1112 RCC1113 RCC1561 RCC1559 RCC1114 RCC1118 RCC1558
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Mbp
1.0
0.8
0.6
0.4
0.2
D
Ostreococcus mediterraneus sp. nov.
O. mediterraneus
O. mediterraneus
RCC501 RCC789 RCC1107 RCC2572 RCC1106 RCC1119 RCC1122 RCC1620 RCC1621 RCC2573 RCC1623 S. cerevisieae BCC101000 RCC2589 RCC2590 RCC2593 RCC2588 RCC2596 RCC2592 RCC2594
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Mbp
1.0
0.8
0.6
0.4
0.2
Probe: 18S ribosomal RNA gene
RCC745 O. tauri
O. mediterraneus
RCC501 RCC789 RCC1107 RCC2572 RCC1106 RCC1199 RCC1122 RCC1620 RCC1621 RCC2573 RCC1623
1 2 3 4 5 6 7 8 9 10 11 12
O.t.
O.m.
Tentative chromosome numbers
1&2
3
4
5&6
7
8&9
10
11
12
13
14
15
16
17&18
19
20
21
11

to differences in the amount of DNA in individual gel tracks.

The large and small variable-sized chromosomes are known from genomic analyses to be atypical in terms of lowered GC content, predicted coding sequences, higher densities of transposable and repetitive DNA elements, so henceforth we refer to them as BOC (big outlier chromosome) and SOC (small outlier chromosome). We thus propose that the variations in size of 2 and 19 in *O. tauri* are associated with the content of their genomes in these regions. Such variations might be due, for example, to a reduced level of recombination (Jancek et al. 2008) that could lead to accumulation of differences between strains by mutation or chromosomal rearrangements. Suppressed recombination is a phenomenon often associated with sex-linked chromosomal regions (Bergero and Charlesworth 2009), but we have not been able to find putative mating-type loci by searching for homologies to related green algae such as *Chlamydomonas reinhardtii* (Lee et al. 2008; Pröschold et al. 2005), although a conserved set of genes necessary for sexual recombination is present (Derelle et al. 2006). This is not surprising. Even if these chromosomes were involved in mating type differentiation, mating type systems have evolved independently in many species (Bergero and Charlesworth 2009; Lee et al. 2010). Two atypical chromosomes with a similar structure also exist in four other species of Mamiellales whose genomes have been characterized, namely *Micromonas* (two clades) (Worden et al. 2009), *Ostreococcus lucimarinus* (Palenik et al. 2007) and in the genome of *Bathycoccus prasinos* (Moreau et al. 2012). We thus propose that this may be a feature common to all members of the order Mamiellales that predates the evolutionary separation of species in this group. Clearly it is not a general feature of the Chlorophyta, since this kind of large-scale genomic structure is not seen in other green algae. In *Volvox carteri* (Prochnik et al. 2010), such islands were not reported but in *Chlamydomonas reinhardtii* (Merchant et al. 2007), a few AT-rich islands were found and in *Chlorella variabilis* small lower GC-rich islands are seen distributed throughout the genome (Blanc et al. 2010). More detailed comparisons of the BOC and SOC between different members of the order Mamiellales can be found in Moreau et al. (2012). Further investigations are required to investigate the generality of this feature in the class Mamiellophyceae.

Clade D Strains Represent a New Species

Surprisingly, we find that one group (clade D) of *Ostreococcus* spp. is very frequently obtained in our cultures from this geographic region. The only representative in culture before this study as far as we know is RCC789, a clonal cell lineage derived from the original RCC501 (Massana et al. 2004). Phylogenetic analyses, (including 18S rDNA, ITS1, 5.8S rDNA and ITS2 sequences, Figure 4A) strongly suggest that the clade D group is a separate species, although the highly conserved 18S rDNA alone is known to underestimate species diversity in protists in many cases, where identical 18S rDNA sequences may hide a species complex (Piganeau et al. 2011b; Slapeta et al. 2006).

The karyotypes of clade D strains were clearly quite different from those of other clades of *Ostreococcus* (Fig. 3, panel D), also consistent with the hypothesis that these individuals represent a different biological species. The ribosomal RNA gene cluster lies on a different chromosome to that found in *O. tauri* (panel D). We were unable to find primers that would PCR-amplify sequences from the genome of *Ostreococcus* clade D (except for one PCR-amplified sequence) because of sequence divergence with other clades despite trying 12 primer pairs that lie on these chromosomes in *O. tauri* (see Methods). We propose that clade D should have 21 chromosomes based on the available PFGE data. Further investigations are required to determine which chromosomes represent the BOC and the SOC in clade D.

Figure 3. Genome structure in the Mamiellales as seen by pulsed field gel electrophoresis.
Pulse field gel electrophoresis (PFGE) of DNA extracted from different clonal lines of Mamiellales' strains showing sizes of chromosomes. Panels **A** and **B** show ethidium bromide stained chromosomes of strains from different genera of the Mamiellophyceae, panel **C** shows 14 strains of the species *O. tauri* and **D** shows 19 strains of *O. mediterraneus* sp. nov. Numbers on the left of the panels indicate expected DNA sizes (Mbp) of the observed bands. In panels C and D specific features of clonal strains are illuminated by autoradiographs to the right of the corresponding gels on the left. **C** – autoradiographs of *O. tauri* (*O.t.*) using a typical chromosome (no. 18), the big outlier chromosome (BOC, chromosome 2) and the small outlier chromosome (SOC, no. 19 in *O. tauri*). **D** – the chromosomal location of the 18S ribosomal gene locus (or loci) in *O. mediterraneus* sp. nov (*O.m.*) is typical of this species.

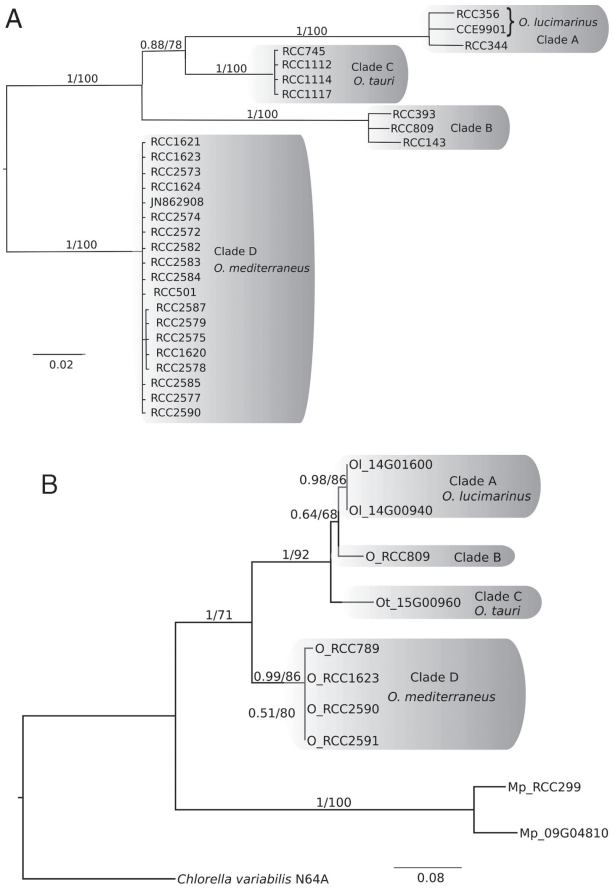


Figure 4. Phylogenetic relationships between *Ostreococcus spp*. clades in the Order Mamiellales. Phylogenies of several *Ostreococcus* strains including *mediterraneus* sp. nov. (clade D). Phylogenetic trees based on **A**: ribosomal RNA gene sequences 18S rDNA+ITS1+5.8S rDNA+ITS2 (GenBank accession numbers are shown for *O. tauri*, and **B**: partial S-adenosylmethionine synthetase DNA sequences, reconstructed using Bayesian inference (BI) and maximum likelihood (ML). Numbers on branches indicate posterior probabilities (BI)/bootstrap support values in % (ML). *Micromonas pusilla* [Mp] and *Ostreococcus* [Ol and Ot] numbers refer to the coding sequences of genes from complete genomes recovered from the "pico-Plaza" comparative genomics website (http://bioinformatics.psb.ugent.be/pico-plaza/, Proost et al. 2009). The scale bars represent the number of substitutions per site. RCC – Roscoff Culture Collection accessions.
(see SupplementaryTable S1 for further details).

In clade D strains karyotypic analyses and a Southern blot using the 18S rDNA as a probe clearly show that 18S rDNA lies on the same chromosome in all of the clade D strains, migrating more quickly than the control hybridisation to chromosome 12 in *O. tauri* (Fig. 3, panel D). In *O. lucimarinus* the 18S rDNA copies are present on chromosomes 8 and 12 (Palenik et al. 2007). In order to further verify the phylogenetic position of clade D strains, 12 primer pairs, chosen using highly conserved coding sequences from *O. tauri* and *O. lucimarinus* (see Methods), were tested by PCR. However, only one of these gave amplified product carrying a clear ORF. The translated DNA sequence of this sequence showed similarity to the coding sequence for S-adenosylmethionine synthetase, and presented sufficiently numerous substitutions between clades of the Mamiellales to perform a phylogeny (Fig. 4B). This analysis, and the rRNA/ITS phylogeny showed that the phylogenetic distance between clade D and the other clades of *Ostreococcus* is greater than the distance between *O. lucimarinus* and *O. tauri* (Fig. 4).

In contrast to the conserved 18S rRNA gene, the internal transcribed spacers (ITS1 and 2, located between 18S, 5.8S, and 28S rRNA genes) showed high sequence diversity among *Ostreococcus* clades, accompanied by high conservation of the intramolecular folding pattern (secondary structure) of the ITS2 RNA transcript. In the family Bathycoccaceae (= *Ostreococcus* and *Bathycoccus*) ITS2 consists of five double-stranded stem regions (the 'universal' helices 1-4, and an additional helix) separated by single-stranded linkers (Marin and Melkonian 2010). In stem regions of ITS2 and rRNA molecules, base pairs are structurally conserved because the majority of observed nucleotide substitutions are either compensatory base changes (CBCs = double-sided nucleotide exchanges that retain base pairing) or hemi-CBCs (exchange of one RNA-nucleotide in a base pair; Caisová et al. 2011; Müller et al. 2007; Zuker 2003). In several taxonomic studies, CBCs in ITS2 helices have been investigated for the prediction of species boundaries (the CBC species concept; Coleman 2000, 2009) and molecular delineation of species. The CBC species concept was based upon the observation that presence of already one distinguishing CBC in the 'conserved parts' of the 'hallmark' helices (sensu Coleman 2007), i.e. Helix 2 (base pairs 1 to about 11) and/or Helix 3, was correlated with the inability of the respective pair of organisms to sexually mate (Coleman 2000).

Therefore, we performed a detailed comparison of predicted ITS2 secondary structures in the Bathycoccaceae, with a focus on molecular signatures (i.e. distinguishing characters) of clade D of *Ostreococcus*, including a record of all CBCs and hemi-CBCs that evolved between *Ostreococcus* clades (Fig. 5A). The 'conserved parts' of

Helix 2 as well as Helix 3 revealed CBCs when the sister genera *Bathycoccus* and *Ostreococcus* were compared, but these parts did not reveal any CBC that evolved among clades A-D of *Ostreococcus* (except one CBC of clade C, i.e. of *O. tauri* [Helix 3, bp 13]; Fig. 5A). Thus, clade D (as well as clades A and B) is not distinguished from other *Ostreococcus* clades in the strict meaning of the CBC concept (Coleman 2000). A similar scenario has been found for other organisms where CBCs in the 'conserved parts' of ITS2 evolved much slower than morphological and/ or molecular (e.g. mating barriers) differences, which may determine species boundaries (Assunção et al. 2012; Caisová et al. 2011, 2013; Wiemers et al. 2009). Thus, *Ostreococcus* appears as an additional case where the CBC concept cannot be applied to predict species boundaries. In contrast to 'conserved parts', the 'variable' stem regions of ITS2 revealed many differences (CBCs and hemi-CBCs) between *Ostreococcus* clades, most of which evolved in a homoplasious way, i.e. with parallelisms and reversals (all observed differences in Helices 1, 2, 3, and the additional ITS2 helix; Fig. 5A). However, Helix 4 contained five base pairs (bp) with unique character states for *Ostreococcus* clade D, i.e. without homoplasies among other *Ostreococcus* clades, which represented molecular signatures for this clade (hemi-CBC in bp 7; CBCs in bp 8, 13, 14 and 21; Fig. 5A). Although Helix 4 is well alignable among the four *Ostreococcus* clades, it should be noted that primary and secondary structures of Helix 4 in *Bathycoccus* and other Mamiellophyceae are considerably different, which prevents any comparison at higher taxonomic levels.

The relatively conserved linker regions between ITS2 helices provided additional signatures for clade D (Fig. 5A). One of these signatures (C as nt 2 in the 3-nt-linker between helices 1 and 2) is unique (without homoplasies) within the whole class Mamiellophyceae (not shown).

One part of the conserved 18S rRNA gene provided additional signatures for *Ostreococcus* clades, i.e. Helix E23_7 (Fig. 5B). In *Ostreococcus*, this helix evolved faster than other parts of the 18S rRNA molecule, which likely is related to the process of length reduction of Helix E23_7 by loss of apical base pairs, compared to other Mamiellophyceae (including *Bathycoccus*) and green algae (Fig. 5B). Similar length reduction occurred only in the pedinophyte flagellate *Marsupiomonas* (see alignment in Fig. 5B). Among the well alignable basal portion of Helix E23_7 (bp 1-7), two base pairs differed between *Ostreococcus mediterraneus* (clade D) and the remaining Bathycoccaceae (but not from other green algae), i.e. bp 3 and 7 (Fig. 5B). The identification of 'unique' signatures in the ITS2 and 18S rRNA molecules allow an unambiguous definition of clade D, despite the similar morphology among *Ostreococcus* clades.

A New Marine Picoeucaryote: *Ostreococcus mediterraneus* sp. nov. (Chlorophyta, Mamiellophyceae, Mamiellales)

The original strain isolated from Barcelona harbour, RCC501, was initially described on the basis of its ribosomal gene DNA sequence (NCBI AY425313, Massana et al. 2004) and placed phylogenetically close to *Bathycoccus prasinos* (Eikrem and Throndsen 1990). Physiological analyses of growth rate in differing light irradiance regimes showed it to behave similarly to *O. tauri* (clade C), and its pigment composition was typical of the Mamiellales, but its karyotype by PFGE was clearly different from that of other clades of *Ostreococcus* (Rodriguez et al. 2005). Here, we confirm by electron microscopic analyses, that we could not distinguish clade D from other clades of *Ostreococcus*, and its morphological description is thus identical to that of other members of the genus. For our taxonomic description of clade D of *Ostreococcus* as a new species, we used molecular signatures in the ITS2 and 18S rRNA markers to avoid any ambiguity in the taxonomic diagnosis.

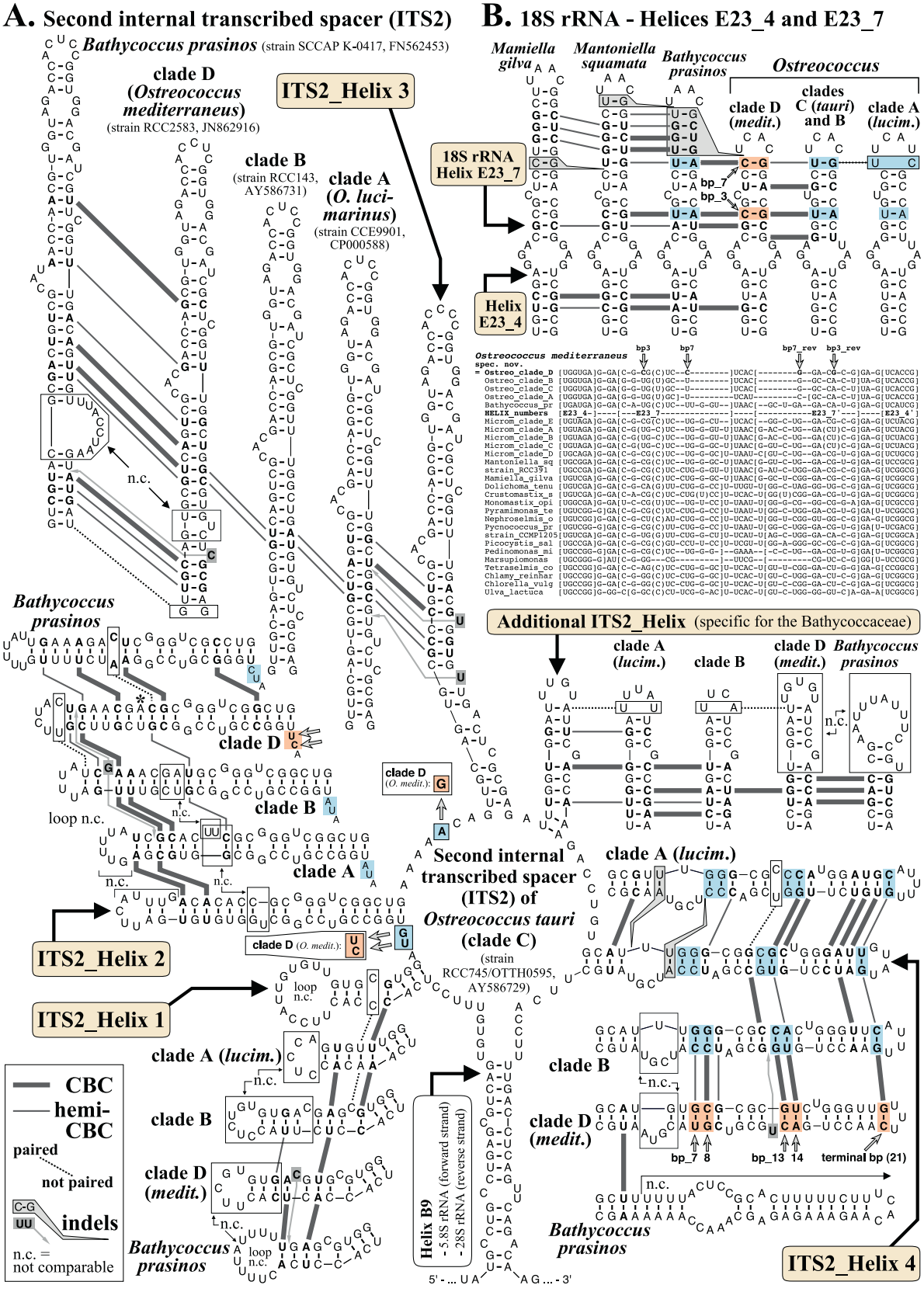
Ostreococcus mediterraneus Marin et Grimsley sp. nov.

Diagnosis: Cells small, about $1.27 \pm 0.5\,\mu m$, spherical or ovoid, without flagella, devoid of cell wall and not covered by scales, with a single semilunar chloroplast containing a central starch granule, without pyrenoid. Marine habitats. Base pair 3 in Helix E23_7 of the nuclear-encoded SSU rRNA is C-G. In the second internal transcribed spacer (ITS2) of the nuclear rRNA operon, nucleotides 2/3 of the linker between helices 1 and 2 are C/U, and base pair 8 of Helix 4 is C-G.

Type locality: Spain, Mediterranean Sea, Barcelona harbour.

Holotype: Cells of *O. mediterraneus* strain RCC789 permanently preserved in resin for electron microscopy (Department of Life Sciences, Botanical Collections, The Natural History Museum London, Cromwell Road, London, UK, SW7 5BD).

Culture of the species: RCC789 (re-isolated from strain RCC501).


A. Second internal transcribed spacer (ITS2)
Bathycoccus prasinos (strain SCCAP K-0417, FN562453)
clade D (Ostreococcus mediterraneus) (strain RCC2583, JN862916)
clade B (strain RCC143, AY586731)
clade A (O. lucimarinus) (strain CCE9901, CP000588)
ITS2_Helix 3
18S rRNA Helix E23_7
Helix E23_4
n.c.
Bathycoccus prasinos
clade D
clade B
clade A
loop n.c.
clade D (O. medit.):
ITS2_Helix 2
ITS2_Helix 1
Second internal transcribed spacer (ITS2) of Ostreococcus tauri (clade C) (strain RCC745/OTTH0595, AY586729)
clade A (lucim.)
clade B
clade D (medit.)
Bathycoccus prasinos
Helix B9
- 5.8S rRNA (forward strand)
- 28S rRNA (reverse strand)
5' - ... U A
A G ... - 3'
CBC
hemi-CBC
paired
not paired
indels
n.c. = not comparable
B. 18S rRNA - Helices E23_4 and E23_7
Mamiella gilva
Mantoniella squamata
Bathycoccus prasinos
Ostreococcus
clades C (tauri) and B
clade D (medit.)
clade A (lucim.)
bp_7
bp_3
Ostreococcus mediterraneus spec. nov.
bp3
bp7
bp7_rev
bp3_rev
HELIX numbers
Additional ITS2_Helix (specific for the Bathycoccaceae)
clade A (lucim.)
clade B
clade D (medit.)
Bathycoccus prasinos
n.c.
Second internal transcribed spacer (ITS2) of Ostreococcus tauri (clade C) (strain RCC745/OTTH0595, AY586729)
clade A (lucim.)
clade B
clade D (medit.)
bp_7 8
bp_13 14
terminal bp (21)
Bathycoccus prasinos
ITS2_Helix 4

Etymology: The species name refers to the distribution in the Mediterranean Sea (mediterraneus [lat.] = midland).

Additional observations support the above description: (i) no name yet exists for this clade, which is taxonomically equivalent to *O. lucimarinus* (nomen nudum, clade A) and *O. tauri* (clade C) (ii) its 18S ribosomal DNA sequence resides on a chromosome that is clearly different in size (≈450 kb, probably chromosome 17 and/or 18, see Fig. 3) to those harbouring the 18S clusters in other species (*O. lucimarinus* — on both chromosomes 8 and12 [539 kb and 702 kb, respectively, Palenik et al. 2007], *O. tauri* — chromosome 12 only, [540 kb, Derelle et al. 2006] and RCC809 clade B – putative chromosomes 9 and 10 [609 kb and 650 kb, respectively, unpublished data, appearing as scaffolds 9 and 10 on the publicly available assembly].

Since there are no clear morphological distinguishing features, we chose "*mediterraneus*" (latin for "midland") as its species name, (in a similar way to the original species of this genus, *O. tauri*, coming from the Thau lagoon, Chrétiennot-Dinet et al. 1995). To our knowledge all members of this species in collections have been so far isolated from the NW Mediterranean Sea or from saltwater lagoons in this region. While this article was in preparation two other ribosomal gene sequences nearly identical to *O. mediterraneus* were identified in samples from the North Sea (GenBank accession numbers FR874724, FR874725), again from a "midland" location in a Norwegian saltwater Fjord close to Bergen, about 20 km away from the open sea, suggesting that this species may also be a habitant of coastal waters in other regions. Type material was derived from strain RCC789, which is a clone of the first cultured isolate of *O. mediterraneus*.

Can the Genus *Ostreococcus* be Used as a Paradigm for the Definition of Picoeukaryotic Species?

One major distinction between the prokaryotic and unicellular eukaryotic organisms that co-exist in a marine environment is in the genetic structure of their populations. Exchange of genetic information by conjugation and/or transduction is well-documented in prokaryotes and the definition of a species is thus blurred by the mobility of their plasmids, genomes, and bacteriophages. A rather empirical division of prokaryotes into different species by the relatedness of their "core" genomes (Lerat et al. 2005) may thus be formulated by using the growing body of available complete bacterial genomes. In contrast, much less is known

Figure 5. Molecular signatures of *Ostreococcus mediterraneus* revealed by comparison of ITS2 and 18S rRNA secondary structures in the family Bathycoccaceae. **A.** The second internal transcribed spacer (ITS2) of *Ostreococcus tauri* (= *Ostreococcus* clade C) with the 'universal' helices 1-4 and an additional helix (between helices 3 and 4) characteristic for the family Bathycoccaceae. Helices of *O. tauri* were compared with homologous ITS2 helices from other *Ostreococcus* clades (A, B, D) and the sister genus *Bathycoccus*, by highlighting CBCs (compensatory base changes), hemi-CBCs, base pairing/ dissociation events (dotted lines), as well as hypothetical insertions/ deletions (indels). A few regions, which were highly divergent and showed no intermediate evolutionary stages, were flagged as 'not comparable' (n.c.). Among 19 clade D isolates (= *O. mediterraneus*), rRNA/ ITS sequences were completely identical except for one nucleotide (ITS2 Helix 2, bp 12 reverse), which was either A (marked by *; 14 isolates incl. RCC501) or T (e.g. in strain RCC1620; the resulting U U mismatch is not shown). Tracing base pair evolution in the stem regions (= helices) by CBCs and hemi-CBCs revealed several homoplasious changes (e.g. parallelisms and reversals) in helices 1, 2, and in the additional helix, whereas five base pairs in Helix 4 represented molecular signatures for *O. mediterraneus*, i.e. without homoplasies in the Bathycoccaceae, most of which evolved as CBCs. All signatures for *O. mediterraneus* were pointed by open arrows and orange shading (web version only; homologous nucleotides in other Bathycoccaceae were shaded in blue color). Base pair 8 of Helix 4 (corresponding to bp 7 in other *Ostreococcus* clades) was selected for the taxonomic diagnosis. Three single-stranded nucleotide positions in the conserved linkers between helices 1, 2 and 3 revealed further distinguishing characters of *O. mediterraneus* (open arrows).
B. Alignment and secondary structure comparison of Helix E23_7 of the 18S rRNA molecule. Helix E23_7 is unusually short in *Ostreococcus*, consisting of only seven base pairs (six in clade A), compared to most other Viridiplantae except *Marsupiomonas* (see alignment). In the alignment, secondary structure information was integrated by [brackets] to indicate stem regions, and normal (brackets) for single-stranded nucleotides within helices. Base pairs 3 and 7 of *O. mediterraneus* (both: C-G) differed from all other members of the Bathycoccaceae (marked as in Fig. 5A). Accession numbers are HE610132 (*Pedinomonas minor*), HE610136 (*Marsupiomonas pelliculata*) and AB425960 (*Ulva lactuca*); for other 18S rDNA accession numbers, see Marin and Melkonian (2010).

about population structures within the diverse community of eukaryotic picoplankton, where gene flow is probably more strictly regulated at a molecular level by cell partner recognition in sexual reproduction. Sexual exchanges govern both the stability of ecological niche specializations, by preserving the genetic integrity of a species, and the levels of gene flow and evolution within the species. In picoeukaryotes, definition of a biological species often remains a challenging task for diverse reasons, such as inability to culture a species for laboratory analyses, unknown sexual cycles and unsuspected diversity within phenotypically similar groups.

Here, we show that these strains cannot be differentiated by using morphological criteria and that molecular criteria are necessary. Pulsed field gel electrophoresis (PFGE) showed that although at least two chromosomes can vary in size between individual lines of clade D, the global pattern of chromosomes is rather characteristic for each species, and that the ribosomal gene clusters are present at species-specific positions on different chromosomes. Molecular data supported the notion that clade C and clade D strains represent biologically distinct species. Their karyotypes are quite different, and certainly cannot represent members of the same population. A striking feature of their karyotypes is the presence, in both clades, of certain chromosomes whose sizes vary widely between individuals. Derelle et al. (2006) first reported the presence of two chromosomes that had different DNA sequence properties to other chromosomes, including (i) lower GC content (ii) more repetitive DNA elements and transposons and (iii) a higher proportion of genes that may be derived from prokaryotes. Given that all of the characteristic BOC and SOC chromosomes have been observed in all of the six fully sequenced genomes of Mamiellales (reviewed in Piganeau et al. 2011a), we suggest that this feature is most probably also found in clade D strains. Although we cannot currently be sure of which chromosomes they are in this species, the 2nd largest and smallest chromosomes are clear candidates (Fig. 3 panel D). Derelle et al. (2006) speculated that the BOC might be involved in sex determination, because mating type loci are often associated with regions of suppressed recombination and consequently evolve differently to the rest of the genome.

Although there is no formal demonstration of a sexual life-cycle, the genomes of Mamiellales encode numerous genes that are probably important for meiosis (Worden et al. 2009) and there is genetic evidence for recombination in natural populations of *O. tauri* (Grimsley et al. 2010). Variations in chromosome size might also be maintained in a region of suppressed recombination, but this is not by itself sufficient evidence to conclude that this chromosome carries a mating type locus. In contrast, in plant pathogenic fungi smaller variable sized-chromosomes can show high levels of polymorphism or "plasticity" as seen in their adaptive race to remain virulent on host plants that express diverse resistance genes (Goodwin et al. 2011; Ma et al. 2010; Rep and Kistler 2010; Stukenbrock et al. 2010). Given the abundance of prasinoviruses attacking *Ostreococcus* (Bellec et al. 2010), and the observation that variable genomic islands contain a significant excess of genes involved in virus-host interactions in cyanobacteria (Avrani et al. 2011), we might speculate that at certain loci on the small chromosome size variability could be involved in resistance/susceptibility to viruses in the Mamiellales (Thomas et al. 2011). Indeed, in *O. tauri*, we have preliminary evidence that the SOC may be associated with resistance or susceptibility to prasinoviruses. Clerissi et al. (2012) showed that variations in chromosome size may be linked to viral specificity, but their studies gave little indication about which genes or molecular mechanisms might be involved. Alternatively, this might reflect genomic plasticity in which part of the SOC is translocated to the BOC, since smaller SOC are sometimes observed with slightly larger BOC in clade C (Fig. 3)

Despite a century and a half of work since Darwin's book "The Origin of the Species" (1859), the biological definition of a species remains difficult for many organisms, and this is particularly true for microbial species in general. There are differing levels of karyotypic variation within certain groups, as exemplified in the fungi, where exceptional plasticity is observed in certain species (Whittle et al. 2011; Zolan 1995), whereas in unicellular yeast "sensu stricto" identical morphologies and ribosomal gene sequences hide a species complex (Liti et al. 2006). Currently, at least three kinds of criteria are in use for differentiating species (see Coyne and Orr 2004, and references therein for a thorough treatment of this subject). Firstly, the biological species concept (BSC), defined as group of organisms capable of interbreeding and producing fertile offspring, but very few of the extremely diverse and innumerable eukaryotic protist lineages have known sexual cycles. Secondly, a phylogenetic species definition (PSD) is often the only pragmatic solution available for differentiating between these organisms, but this is often based on the comparison of one or a few DNA sequences, and may give an underestimation of true diversity (Piganeau et al. 2011b). Thirdly, when more genomic data are available

for closely related species, such as in yeast, and now as more and more data are becoming available from new generation sequencing techniques, the genotypic cluster species concept (GCSC) can be used (Mallet 1995). This concept compares the level of polymorphism between orthologous regions of the two genomes; if this exceeds ~5% hybrids cannot produce viable descendants even if the gene order is colinear, mainly because recombination becomes mechanistically difficult in face of the induced turmoil of mismatch repair (see for example Liti et al. 2006). The ribosomal 18S rRNA gene sequence only differs by 0.17% between *O. tauri* and *O. lucimarinus*, predicting that they are the same species by current classification criteria when operational taxonomic units (OTU) similar to those used for environmental rDNA sequence data (Caron et al. 2009) are applied, but their complete genomes share only ~75% of DNA sequence identity (between orthologous genes, amino acid sequence being only 70% identical, Jancek et al. 2008) and the sequential order of their genes (colinearity of chromosomes) is highly shuffled, so chromosome pairing is quite impossible. We thus propose that, as the 18S rDNA barcode marker underestimates microbial biological species diversity, when a sufficient number of reads is available any difference in this sequence identified with certainty between two strains of microbial eukaryotes represents two different species, even if they cannot be discriminated morphologically. In contrast, the reverse is not true, and strains having a complete identity of their 18S rRNA genes do not necessarily represent the same species but can hide cryptic species. In the future, we thus expect that single amplified genome analyses of the unknown and phylogenetically diverse protists found in nature will increase the depth of known eukaryotic diversity and expand our knowledge about their population structures.

Methods

Isolation culture, and identification of strains: The techniques used for isolation and growth of strains was done essentially as reported recently (Grimsley et al. 2010), except that the pre-filtration step on 3 μm filter was omitted, and smaller volumes were filtered. Briefly, 15 ml lots of collected sea water sample were passed individually through 1.2 μm disposable filter units using a syringe and mixed with 15 ml of Keller's medium for the initial culture period of about 3 weeks before plating out for individual clones on gel-solidified Keller's medium or on L1 medium (Guillard and Hargraves 1993). Colonies were picked off for further growth after 3 weeks. The same primer pairs (Grimsley et al. 2010) were used for PCR amplification of the 18S ribosomal gene for DNA sequence analyses. To amplify the S-adenosylmethionine synthetase (SAM) gene fragment (709 bp) from clade D strains the following primer pairs were used: chr14-4 fwd: 5′-aagctcgcggatcaaatctc-3′, chr14-4 rev: 5′-atcttacgaccggtcaaacc-3′. DNA sequences have been deposited in NCBI GenBank under the accession numbers JN862902-JN862919 (ribosomal RNA gene sequences) and JQ009201-JQ009203 (SAM sequences).

Phylogenetic analyses: 18S rDNA and ITS, as well as S-Adenosylmethionine synthetase (SAM) sequence alignments were performed with MUSCLE (Edgar 2004) and ambiguous parts were removed using GBlocks (Castresana 2000) when necessary (not for SAM sequences). The lengths of the resulting alignments were 2170 bp for 18S-ITS sequences and 609 bp (203 AA) for SAM sequences. The homogeneity of 18S rDNA and ITS alignments was established using a partition homogeneity test (Farris et al. 1994), and sequences were concatenated. Phylogenetic reconstructions were based on DNA sequences as well as amino acid (AA) for SAM, using Bayesian inference (BI) and maximum likelihood (ML). Evolutionary models were selected via Akaike Information Criterion using ProtTest (Abascal et al. 2005) for AA sequences (a WAG + I [proportion of invariant sites] model was selected) and jModelTest (Guindon and Gascuel 2003; Posada 2008) for non coding DNA sequences (leading to a GTR + I model). Bayesian analysis were carried out with MrBayes 3.2.1 (Ronquist et al. 2012), with 4 chains of 10^6 generations, trees sampled every 400 generations, and burnin value set to 20% of the sampled trees. In BI, coding DNA sequences were considered with an evolutionary model taking the genetic code into account (Goldman and Yang 1994; Muse and Gaut 1994), and AA sequences were analysed with a mixed model. We checked that standard deviation of the split frequencies fell below 0.01 to ensure convergence in tree search. Maximum likelihood reconstructions were carried out using PAUP*4.0b10 (Swofford 2003) for DNA and PhyML (Guindon and Gascuel 2003) for AA, and validated with 100 bootstrap replicates.

RNA secondary structure predictions were performed using the Mfold web interface (Zuker 2003; http://mfold.bioinfo.rpi.edu/), using entire ITS2 sequences and/or individual helices as input, together with a comparative search for covariations (CBCs) among taxa.

Preparation of genomic DNA and Pulsed Field Electrophoresis: For each *Ostreococcus* strain (clade C and D), cells were harvested from 2 x 200 ml of culture (1.6 x 10^7 cells/ml) by centrifugation at 8000 g for 20 min. Genomic DNA was extracted from one pellet using a CTAB protocol (Winnepenninckx et al. 1993). Cells of the second pellet were resuspended and embedded in 1% low melting point agarose strings (2.6 x 10^9 cells/ml) and then lysed with proteinase K at 37 °C. As previously described (Rodriguez et al. 2005), the PFGE parameters used were 0.8% agarose, 3 V/cm, 0.5x TBE (Tris-Borate 89 mM, EDTA 2 mM buffer), 120° switching angle, 14 °C, 90 s switch time for 45 h followed by 140 s switch time for 27 h. After denaturation and drying, the gel was directly hybridized with radioactive probes (Mead et al. 1988). For the clade C (chromosomes 2, 18 and 19) and clade D (18S RNA gene) PFGE hybridizations, *Ostreococcus tauri* RCC 745 genomic DNA was used as template to amplify by PCR several DNA fragments of about 600 bp. Universal primers were used for the amplification of the 18S, ITS1, 5.8S, ITS2 and partial 28S RNA (see Grimsley et al. 2010 for primer sequences, PCR conditions and source references) and the following primers were used for i) chromosome 2: γ-tubulin_Fw 5′- cgacggtgtagcaagctatg-3′, γ-tubulin_Rv 5′-gaaaagtgccatccattcgt-3′; cdc-25_Fw 5′-tggcgacacgtgaggaactt-3′, cdc-25_Rv 5′-gttgaacctctcgtttcgaa-3′; Ot02g0410_Fw: 5′-tcgaagatatctcccggatggct-3′, Ot02g0410_Rv

5′- cgccatcttctctagagcgg-3′, ii) chromosome 19: 19a_Fw 5′-gatgtcgacgaagtcttccgat-3′, 19a_Rv 5′-tcacgcgggataatgacgcagat-3′; 19b_ Fw 5′-ttacatgatcaagcaccctctc-3′, 19b_Rv 5′-cggtgtcgtcggagcgacaaagc-3′; 19c_ Fw 5′-gacgtgatcgatagaagcagacc-3′, 19c_Rv 5′-cgcgctgagaatttactcg-3′; 19d_ Fw 5′-aggacgctcgtggtgagaacac-3′, 19d_Rv 5′-atcgtcctcaattgtcaaggc -3′; 19e_ Fw 5′-gcgatgacggtgctctacc -3′, 19e_Rv 5′-gcgcgtggagttatccccgaacc-3′, iii) chromosome 18: 18a, b, c and d (Grimsley et al. 2010). After purification, these amplicons were pooled for each chromosome, labelled with [α-^{32}P]-dCTP by random priming (Prime-a gene kit, Promega) and used as probes on PFGE dried gels.

Electron microscopy: For transmission electron microscopy (TEM), the cells were prepared according to Chrétiennot-Dinet et al. (1995). Briefly, they were fixed in 1% glutaraldehyde in their culture medium before harvesting by centrifugation (15′, 3000 g) and embedding in molten agarose (37 °C). After solidification the plug was fixed for 2 h at 4 °C in 2.5% glutaraldehyde with one volume of 0.4 M cacodylate buffer and two volumes of culture medium. The plug was then washed 3 times for 30′ in 1: 1: 1 x 0.4 M cacodylate buffer: 1 x culture medium. For postfixation, 1% OsO4 in 0.2 M cacodylate at 4 °C for 1 h was used. After two further washes in 0.2 M cacodylate, the plug was cut into small pieces before serial dehydration in ethyl alcohol and embedded in Epon 812. Thin sections were stained with uranyl acetate and lead citrate before examination on a 7500 Hitachi transmission electron microscope.

Acknowledgements

We would like to thank Laetitia Bariat and the cytometry/imaging platform of the Banyuls' Oceanographic Observatory, Patrick Le Gall of the IFREMER laboratory at Sète for help with collecting water from the Thau lagoon, Eve Toulza for stimulating discussions, and the captain and crew of the Néréis II for collection of water from Banyuls' Bay (SOLA). We are additionally grateful to Lionel Fibla, Richard Cooke and Michèle Laudié for help with primer design, PCR amplifications and DNA sequencing, and to the reviewers for their helpful detailed comments. Laure Guillou and Ramon Massana collected and maintained the original isolate (now RCC789), and kindly gave permission for us to use this strain for the description of a new species. We are depositing our strains for distribution in the Roscoff Culture Collection. The work was supported by the following French National Agency grants: PICOVIR ANR-07-BLAN-0210 (coordinator N.H.G.), PHYTADAPT ANR NT09_567009 (coordinator H.M.), and AP-FRB2009-PICOPOP (to G.P.)

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.protis.2013.06.002.

References

Abascal F, Zardoya R, Posada D (2005) ProtTest: selection of best-fit models of protein evolution. Bioinformatics **21**:2104–2105

Assunção P, Jaén-Molina R, Caujapé-Castells J, de la Jara A, Carmona L, Freijanes K, Mendoza H (2012) Molecular taxonomy of *Dunaliella* (Chlorophyceae), with a special focus on *D. salina*: ITS2 sequences revisited with an extensive geographical sampling. Aquat Biosystems **8**:2

Avrani S, Wurtzel O, Sharon I, Sorek R, Lindell D (2011) Genomic island variability facilitates *Prochlorococcus*-virus coexistence. Nature **474**:604–608

Bellec L, Grimsley N, Derelle E, Moreau H, Desdevises Y (2010) Abundance, spatial distribution and genetic diversity of *Ostreococcus tauri* viruses in two different environments. Environ Microbiol Reports **2**:313–321

Bergero R, Charlesworth D (2009) The evolution of restricted recombination in sex chromosomes. Trends Ecol Evol **24**:94–102

Blanc G, Duncan G, Agarkova I, Borodovsky M, Gurnon J, Kuo A, Lindquist E, Lucas S, Pangilinan J, Polle J, Salamov A, Terry A, Yamada T, Dunigan DD, Grigoriev IV, Claverie J-M, Van Etten JL (2010) The *Chlorella variabilis* NC64A genome reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex. Plant Cell **22**:2943–2955

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

Caisová L, Marin B, Melkonian M (2013) A consensus secondary structure of ITS2 in the Chlorophyta identified by phylogenetic reconstruction. Protist **164**:482–496

Cardol P, Bailleul B, Rappaport F, Derelle E, Beal D, Breyton C, Bailey S, Wollman FA, Grossman A, Moreau H, Finazzi G (2008) An original adaptation of photosynthesis in the marine green alga *Ostreococcus*. Proc Natl Acad Sci USA **105**:7881–7886

Caron DA, Countway PD, Savai P, Gast RJ, Schnetzer A, Moorthi SD, Dennett M, Moran D, Jones A (2009) Defining DNA-based operational taxonomic units for microbial-eukaryote ecology. Appl Environ Microbiol **75**:5797–5808

Darwin C (1859) On the Origin of the Species by Natural Selection. Murray

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol **17**:540–552

Chrétiennot-Dinet M-J, Courties C, Vaquer A, Neveux J, Claustre H, Lautier J, Machado MC (1995) A new marine picoeucaryote - *Ostreococcus tauri* (Chlorophyta. Prasinophyceae) Phycologia **34**:285–292

Clerissi C, Desdevises Y, Grimsley NH (2012) Prasinoviruses of the marine green alga *Ostreococcus tauri* are mainly species-specific. J Virol **86**:4611–4619

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

Courties C, Vaquer A, Troussellier M, Lautier J, Chrétiennot-Dinet M-J, Neveux J, Machado C, Claustre H (1994) Smallest eukaryotic organism. Nature **370**, 255–255

Coyne JAJ, Orr HA (2004) Speciation. Sinauer Associates

Demir-Hilton E, Sudek S, Cuvelier ML, Gentemann CL, Zehr JP, Worden AZ (2011) Global distribution patterns of distinct clades of the photosynthetic picoeukaryote *Ostreococcus*. ISME J **5**:1095–1107

Derelle E, Ferraz C, Rombauts S, Rouze P, Worden AZ, Robbens S, Partensky F, Degroeve S, Echeynie S, Cooke R, Saeys Y, Wuyts J, Jabbari K, Bowler C, Panaud O, Piegu B, Ball SG, Ral JP, Bouget FY, Piganeau G, De Baets B, Picard A, Delseny M, Demaille J, Van de Peer Y, Moreau H (2006) Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. Proc Natl Acad Sci USA **103**:11647–11652

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res **32**:1792–1797

Eikrem W, Throndsen J (1990) The ultrasructure of *Bathycoccus* gen-nov and *Bathycoccus prasinos*, a nonmotile planktonic alga (Chlorophyta, Prasinophyceae) from the Mediterranean and Atlantic. Phycologia **29**:344–350

Farris JS, Kallersjo M, Kluge AG, Bult C (1994) Testing significance of incongruence. Cladistics **10**:315–320

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

Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol **11**:725–736

Goodwin SB, M'barek SB, Dhillon B, Wittenberg AHJ, Crane CF, Hane JK, Foster AJ, Van der Lee TAJ, Grimwood J, Aerts A, Antoniw J, Bailey A, Bluhm B, Bowler J, Bristow J, van der Burgt A, Canto-Canché B, Churchill ACL, Conde-Ferràez L, Cools HJ, Coutinho PM, Csukai M, Dehal P, De Wit P, Donzelli B, van de Geest HC, van Ham RCHJ, Hammond-Kosack KE, Henrissat B, Kilian A, Kobayashi AK, Koopmann E, Kourmpetis Y, Kuzniar A, Lindquist E, Lombard V, Maliepaard C, Martins N, Mehrabi R, Nap JPH, Ponomarenko A, Rudd JJ, Salamov A, Schmutz J, Schouten HJ, Shapiro H, Stergiopoulos I, Torriani SFF, Tu H, de Vries RP, Waalwijk C, Ware SB, Wiebenga A, Zwiers L-H, Oliver RP, Grigoriev IV, Kema GHJ (2011) Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensable structure, chromosome plasticity, and stealth pathogenesis. PLoS Genet **7**, e1002070–e1002070

Grimsley N, Pequin B, Bachy C, Moreau H, Piganeau G (2010) Cryptic sex in the smallest eukaryotic marine green alga. Mol Biol Evol **27**:47–54

Gross F (1937) Notes on the culture of some marine plankton organisms. J Mar Biol Assoc UK **21**:753–768

Guillard RRL, Hargraves PE (1993) *Stichochrysis immobilis* is a diatom, not a chrysophyte. Phycologia **32**:234–236

Guillou L, Eikrem W, Chrétiennot-Dinet M-J, Le Gall F, Massana R, Romari K, Pedros-Alio C, Vaulot 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

Guindon SP, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol **52**:696–704

Hall T (1999) BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser **41**:95–98

Henderson G, Gan L, Jensen G (2007) 3-D ultrastructure of *O. tauri*: electron cryotomography of an entire eukaryotic cell. PLoS ONE **2**:e749

Jancek S, Gourbiere S, Moreau H, Piganeau G (2008) Clues about the genetic basis of adaptation emerge from comparing the proteomes of two *Ostreococcus* ecotypes (Chlorophyta, Prasinophyceae). Mol Biol Evol **25**:2293–2300

Knight-Jones EW, Walne PR (1951) *Chromulina pusilla* Butcher, a dominant member of the ultraplankton. Nature **167**:445–446

Lee JH, Lin HW, Joo S, Goodenough U (2008) Early sexual origins of homeoprotein heterodimerization and evolution of the plant KNOX/BELL family. Cell **133**:829–840

Lee SC, Ni M, Li W, Shertz C, Heitman J (2010) The evolution of sex: a perspective from the fungal kingdom. Microbiol Mol Biol Rev **74**:298–340

Lerat E, Daubin V, Ochman H, Moran NA (2005) Evolutionary origins of genomic repertoires in bacteria. PLoS Biol **3**:e130

Li WKW, Dickie PM, Irwin BD, Wood AM (1992) Biomass of bacteria, cyanobacteria, prochlorophytes and photosynthetic eukaryotes in the Sargasso Sea. Deep Sea Res Part A Oceanogr Res Pap **39**:501–519

Liti G, Barton DBH, Louis EJ (2006) Sequence diversity, reproductive isolation and species concepts in *Saccharomyces*. Genetics **174**:839–850

Lovejoy C (2007) Distribution, phylogeny, and growth of cold-adapted picoprasinophytes in arctic seas. J Phycol **43**:78–89

Lovejoy C, Potvin M (2010) Microbial eukaryotic distribution in a dynamic Beaufort Sea and the Arctic Ocean. J Plankton Res **33**:431–444

Ma L-J, van der Does HC, Borkovich KA, Coleman JJ, Daboussi M-J, Di Pietro A, Dufresne M, Freitag M, Grabherr M, Henrissat B, Houterman PM, Kang S, Shim W-B, Woloshuk C, Xie X, Xu J-R, Antoniw J, Baker SE, Bluhm BH, Breakspear A, Brown DW, Butchko RAE, Chapman S, Coulson R, Coutinho PM, Danchin EGJ, Diener A, Gale LR, Gardiner DM, Goff S, Hammond-Kosack KE, Hilburn K, Hua-Van A, Jonkers W, Kazan K, Kodira CD, Koehrsen M, Kumar L, Lee Y-H, Li L, Manners JM, Miranda-Saavedra D, Mukherjee M, Park G, Park J, Park S-Y, Proctor RH, Regev A, Ruiz-Roldan MC, Sain D, Sakthikumar S, Sykes S, Schwartz

DC, Turgeon BG, Wapinski I, Yoder O, Young S, Zeng Q, Zhou S, Galagan J, Cuomo CA, Kistler HC, Rep M (2010) Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. Nature **464**:367–373

Mallet JA (1995) Species' definition for the modern synthesis. Trends Ecol Evol **10**:294–299

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 (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

Massana R, Balague V, Guillou L, Pedros-Alio C (2004) Picoeukaryotic diversity in an oligotrophic coastal site studied by molecular and culturing approaches. FEMS Microbiol Ecol **50**:231–243

Massana R (2011) Eukaryotic picoplankton in surface oceans. Annu Rev Microbiol **65**:91–110

Mead JR, Arrowood MJ, Current WL, Sterling CR (1988) Field inversion gel electrophoretic separation of *Cryptosporidium spp*. chromosome-sized DNA. J Parasitol **74**: 366–369

Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, Witman GB, Terry A, Salamov A, Fritz-Laylin LK, Marechal-Drouard L, Marshall WF, Qu L-H, Nelson DR, Sanderfoot AA, Spalding MH, Kapitonov VV, Ren Q, Ferris P, Lindquist E, Shapiro H, Lucas SM, Grimwood J, Schmutz J, Cardol P, Cerutti H, Chanfreau G, Chen C-L, Cognat V, Croft MT, Dent R, Dutcher S, Fernandez E, Fukuzawa H, Gonzalez-Ballester D, Gonzalez-Halphen D, Hallmann A, Hanikenne M, Hippler M, Inwood W, Jabbari K, Kalanon M, Kuras R, Lefebvre PA, Lemaire SD, Lobanov AV, Lohr M, Manuell A, Meier I, Mets L, Mittag M, Mittelmeier T, Moroney JV, Moseley J, Napoli C, Nedelcu AM, Niyogi K, Novoselov SV, Paulsen IT, Pazour G, Purton S, Ral J-P, Riano-Pachon DM, Riekhof W, Rymarquis L, Schroda M, Stern D, Umen J, Willows R, Wilson N, Zimmer SL, Allmer J, Balk J, Bisova K, Chen C-J, Elias M, Gendler K, Hauser C, Lamb MR, Ledford H, Long JC, Minagawa J, Page MD, Pan J, Pootakham W, Roje S, Rose A, Stahlberg E, Terauchi AM, Yang P, Ball S, Bowler C, Dieckmann CL, Gladyshev VN, Green P, Jorgensen R, Mayfield S, Mueller-Roeber B, Rajamani S, Sayre RT, Brokstein P, Dubchak I, Goodstein D, Hornick L, Huang YW, Jhaveri J, Luo Y, Martinez D, Ngau WCA, Otillar B, Poliakov A, Porter A, Szajkowski L, Werner G, Zhou K, Grigoriev IV, Rokhsar DS, Grossman AR (2007) The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. Science **318**:245–250

Moreau H, Verhelst BA, Couloux, Derelle E, Rombauts S, Grimsley N, Van Bel M, Poulain J, Katinka M, Hohmann-Marriott M, Piganeau Gwenael, Rouzé P, Da Silva C, Wincker P, Van de Peer Y, Vandepoele K (2012) Gene functionalities and genome structure in *Bathycoccus prasinos* reflect cellular specializations at the base of the green lineage. Genome Biol **13**:R24

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

Muse SV, Gaut BS (1994) A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. Mol Biol Evol **11**:715–724

Palenik B, Grimwood J, Aerts A, Rouze P, Salamov A, Putnam N, Dupont C, Jorgensen R, Derelle E, Rombauts S, Zhou KM, Otillar 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 QH, 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**:7705–7710

Piganeau G, Grimsley N, Moreau H (2011a) Genome diversity in the smallest marine photosynthetic eukaryotes. Res Microbiol **162**:570–577

Piganeau G, Eyre-Walker A, Grimsley N, Moreau H (2011b) How and why DNA barcodes underestimate the diversity of microbial eukaryotes. PLoS ONE **6**, e16342–e16342

Posada D (2008) jModelTest: Phylogenetic Model Averaging. Mol Biol Evol **25**:1253–1256

Prochnik SE, Umen J, Nedelcu AM, Hallmann A, Miller SM, Nishii I, Ferris P, Kuo A, Mitros T, Fritz-Laylin LK, Hellsten U, Chapman J, Simakov O, Rensing SA, Terry A, Pangilinan J, Kapitonov V, Jurka J, Salamov A, Shapiro H, Schmutz J, Grimwood J, Lindquist E, Lucas S, Grigoriev IV, Schmitt R, Kirk D, Rokhsar DS (2010) Genomic analysis of organismal complexity in the multicellular green alga *Volvox carteri*. Science **329**:223–226

Proost S, Van Bel M, Sterck L, Billiau K, Van Parys T, Van de Peer Y, Vandepoele K (2009) PLAZA: A comparative genomics resource to study gene and genome evolution in plants. Plant Cell **21**:3718–3731

Pröschold T, Harris EH, Coleman AW (2005) Portrait of a species: *Chlamydomonas reinhardtii*. Genetics **170**:1601–1610

Rep M, Kistler HC (2010) The genomic organization of plant pathogenicity in *Fusarium* species. Curr Opin Plant Biol **13**:420–426

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

Ronquist FF, Teslenko MM, van der Mark PP, Ayres DLD, Darling AA, Höhna SS, Larget BB, Liu LL, Suchard MA, Huelsenbeck JP (2012) MrBayes 3. 2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol **61**:539–542

Slapeta J, Lopez-Garcia P, Moreira D (2006) Global dispersal and ancient cryptic species in the smallest marine eukaryotes. Mol Biol Evol **23**:23–29

Stockner JG (1988) Phototrophic picoplankton: An overview from marine and freshwater ecosystems. Limnol Oceanogr **33**:765–775

Stukenbrock EH, Jørgensen FG, Zala M, Hansen TT, McDonald BA, Schierup MH (2010) Whole-genome and chromosome evolution associated with host adaptation and speciation of the wheat pathogen *Mycosphaerella graminicola*. PLoS Genet **6**:e1001189

Swofford DL (2003) PAUP*. Phylogenetic analysis using parsimony (*and other methods) Version 4. Sinauer Associates, Sunderland, Massachusetts

Thomas R, Grimsley N, Escande M, Subirana L, Derelle E, Moreau H (2011) Acquisition and maintenance of resistance to viruses in eukaryotic phytoplankton populations. Environ Microbiol **13**:1412–1420

Tsai IJ, Bensasson D, Burt A, Koufopanou V (2008) Population genomics of the wild yeast *Saccharomyces paradoxus*: Quantifying the life cycle. Proc Natl Acad Sci USA **105**:4957–4962

Vaulot D, Eikrem W, Viprey M, Moreau H (2008) The diversity of small eukaryotic phytoplankton (<3 μm) in marine ecosystems. FEMS Microbiol Rev **32**:795–820

Vaulot D, Le Gall F, Marie D, Guillou L, Partensky F (2004) The Roscoff Culture Collection (RCC): a collection dedicated to marine picoplankton. Nova Hedwigia **79**:49–70

Viprey M, Guillou L, Ferreol M, Vaulot D (2008) Wide genetic diversity of picoplanktonic green algae (Chloroplastida) in the Mediterranean Sea uncovered by a phylum-biased PCR approach. Environ Microbiol **10**:1804–1822

Whittle CA, Nygren K, Johannesson H (2011) Consequences of reproductive mode on genome evolution in fungi. Fung Genet Biol **48**:661–667

Winnepenninckx B, Backeljau T, De Wachter R (1993) Extraction of high molecular weight DNA from molluscs. Trends Genet **9**:407

Wiemers M, Keller A, Wolf M (2009) ITS2 secondary structure improves phylogeny estimation in a radiation of blue butterflies of the subgenus *Agrodiaetus* (Lepidoptera: Lycaenidae: *Polyommatus*). BMC Evol Biol **9**:300

Worden AZ, Nolan JK, Palenik B (2004) Assessing the dynamics and ecology of marine picophytoplankton: The importance of the eukaryotic component. Limnol Oceanogr **49**:168–179

Worden A, Lee J-H, Mock T, Rouzé P, Simmons MP, Aerts AL, Allen AE, Cuvelier ML, 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, Otillar R, Panaud O, Pangilinan J, Paulsen I, Piegu B, Poliakov A, Robbens S, Schmutz 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

Zhu F, Massana R, Not F, Marie D, Vaulot D (2005) Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. FEMS Microbiol Ecol **52**:79–92

Zolan ME (1995) Chromosome-length polymorphism in fungi. Microbiol Rev **59**:686–698

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res **31**:3406–3415