Cryptic Sex in the Smallest Eukaryotic Marine Green Alga

Nigel Grimsley,[1,2] Bérangère Péquin,[1,2] Charles Bachy,[1,2] Hervé Moreau,[1,2] and Gwenaël Piganeau*[,1,2]

[1]Université Pierre et Marie Curie, UPMC Univ Paris 06, UMR 7621, LOBB, Observatoire Océanologique, Banyuls/mer, France

[2]Centre National de la recherche Scientifique, CNRS, UMR 7621, LOBB, Observatoire Océanologique, Banyuls/mer, France

***Corresponding author:** E-mail: gwenael.piganeau@obs-banyuls.fr.

Associate editor: Peter Lockhart

Abstract

Ostreococcus spp. are common worldwide oceanic picoeukaryotic pelagic algae. The complete genomes of three strains from different ecological niches revealed them to represent biologically distinct species despite their identical cellular morphologies (cryptic species). Their tiny genomes (13 Mb), with ~20 chromosomes, are colinear and densely packed with coding sequences, but no sexual life cycle has been described. Seventeen new strains of one of these species, *Ostreococcus tauri*, were isolated from 98 seawater samplings from the NW Mediterranean by filtering, culturing, cloning, and plating for single colonies and identification by sequencing their ribosomal 18S gene. In order to find the genetic markers for detection of polymorphisms and sexual recombination, we used an *in silico* approach to screen available genomic data. Intergenic regions of DNA likely to evolve neutrally were analyzed following polymerase chain reaction amplification of sequences using flanking primers from adjacent conserved coding sequences that were present as syntenic pairs in two different species of *Ostreococcus*. Analyses of such DNA regions from eight marker loci on two chromosomes from each strain revealed that the isolated *O. tauri* clones were haploid and that the overall level of polymorphism was ~0.01. Four different genetic tests for recombination showed that sexual exchanges must be inferred to account for the between-locus and between-chromosome marker combinations observed. However, our data suggest that sexual encounters are infrequent because we estimate the frequency of meioses/mitoses among the sampled strains to be 10^{-6}. *Ostreococcus tauri* and related species encode and express core genes for mitosis and meiosis, but their mechanisms of cell division and recombination, nevertheless, remain enigmatic because a classical eukaryotic spindle with 40 canonical microtubules would be much too large for the available ~0.9-μm^3 cellular volume.

Key words: picoplankton, population genetics, *Ostreococcus*, recombination detection, polymorphism.

Research article

Introduction

Although the smallest eukaryotic green algae account for only a modest fraction of the oceanic biomass, they nevertheless contribute significantly to primary production in many oceanic waters (Stockner 1988; Li 1994; Worden 2006). *Ostreococcus* spp. are worldwide distributed unicellular autotrophic ~1-μm diameter cells, each containing one chloroplast, one mitochondrion, and a small (~13 Mb) nuclear genome. Sequence analysis of the genomes of three different strains of *Ostreococcus* spp. isolated from different ecological niches (coastal, oceanic, and deepwater) has been completed (Derelle et al. 2006; Palenik et al. 2007; Jancek et al. 2008) and yields information about their niche specializations.

All ecotypes harbor coding sequences with high similarity to several core meiotic genes (Derelle et al. 2006) and similarities to homeodomain proteins known to be important in early zygote development after fusion of plus and minus gametes in other species (Lee et al. 2008), but only a haploid phase of *Ostreococcus* spp. has so far been observed. In fact, no sexual phase has been described for any of the related species in this order (Mamiellales), including *Bathycoccus* sp. and *Micromonas* sp., despite rapid progress in analysis of their complete genomes (Worden et al. 2009). All these species have well-structured compact genomes, about 20 linear chromosomes, with a high level of between-strain synteny, despite their high levels of nucleotide sequence divergence (Palenik et al. 2007). Thus, circumstantial evidence about these genomes and their predicted meiotic functionalities intimate the likely existence of an unobserved diploid phase in this group of species that are fast becoming models at the cutting edge of genomics and ecology.

Although numerous important studies describe marine plankton diversity (Lopez-Garcia et al. 2001; Massana and Pedros-Alio 2008; Vaulot et al. 2008) and much more work is still required to catalogue marine microbial diversity, population genetic studies of planktonic species are scarce and absent for picoplanktonic eukaryotes. The main reason for this lack of knowledge of intraspecific diversity is that only a minority of marine protists can be isolated and cultured in the laboratory (for a review, see Massana and Pedros-Alio 2008), a prerequisite for their physiological characterization, for the extraction of sufficient DNA and for testing the induction of mating between strains.

Not surprisingly, our knowledge about the life cycles and the occurrence of sexual reproduction in free-living planktonic eukaryotes is mainly limited to certain diploid groups such as diatoms, where meiosis may be induced by the size of their mineral extracellular envelope (McMullin et al. 2005; Telford et al. 2006; Chepurnov et al. 2008) and freshwater ciliates (Coleman 2005; Snoke et al. 2006; Catania et al. 2009). In the green algae, better defined as

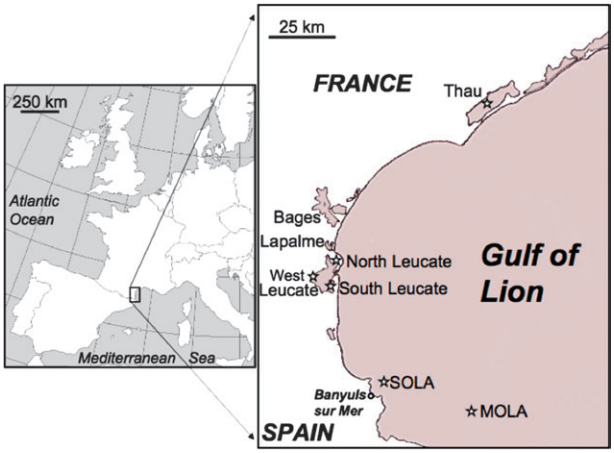

FIG. 1. The locations of sampling sites in coastal lagoons and in the open sea are indicated by open stars.

Chlorophyta, life cycles have been described in only a few species. Many of the multicellular Chlorophyceae are not attractive as models because of their multinucleate cells (Waaland et al. 2004). In freshwater systems, colonial *Volvox* spp. and unicellular *Chlamydomonas* are used as models (Hallmann 2003; Proschold et al. 2005). The latter genera are haplontic, the diploid phase being reduced to a short part of their life cycles. No sexual stages have yet been reported in the taxonomically diverse order Mamiellales that includes *Ostreococcus tauri*, the subject of this work, although genes predicted to encode and express meiosis-specific functions have been found in the genome (Derelle et al. 2006). The phylogenetically closest species with an observed sexual stage within the class Prasinophyceae is the freshwater *Nephroselmis* spp., where microscopy revealed cell fusion (syngamy) followed by division of one cell into four daughter cells, suggesting that sexual reproduction occurs (Suda et al. 1989, 2004).

In this study, we take advantage of the available genome sequence data of *O. tauri* and *O. lucimarinus* to investigate the population genetic diversity of this species from 17 strains isolated from five locations in the Gulf of Lion, NW Mediterranean Sea. We analyze this polymorphism data to assess whether this organism is haploid and whether it reproduces sexually.

Materials and Methods

Isolation Culture and Identification of *O. tauri* Strains

Samples of surface water (0.5–1 l) were collected from different saltwater lagoons or coastal waters in the NW Mediterranean (Fig. 1) and maintained at about 8 °C before returning to the laboratory. In all, 98 samplings were performed among these different sites over a 2-year period (January 2006 to January 2008). Each sample was treated separately, and many samples were subdivided into separate subcultures. To remove larger plankton, the samples were filtered successively by passage through a 3-μm pore size filter (Millipore) and then through a 1.2-μm filter (Sartorius). Aliquots (50 ml) of each filtrate were then mixed at a ratio of 1:1 with Keller's medium (Keller and Guillard 1985) to favor growth of phytoplankton. Antibiotics were added (kanamycin 20 μg ml^{-1}, penicillin 25 μg ml^{-1}, and neomycin 20 μg ml^{-1} final concentrations) to reduce bacterial growth. After 2–3 weeks in culture (20 °C, 100 μm^{-2}s^{-1} continuous white light), many of the cultures appeared pale green and most of the latter were found to contain picoplankton with flow cytometric characteristics similar to our characterized control strain OTTH0595. These lines were retained for further study (supplementary table 1, Supplementary Material Online). In order to obtain clonal cell lines, serial dilutions of such cultures were plated into K medium solidified with 0.15% agarose. Individual colonies usually appeared after 3–4 weeks and were picked off and resuspended in liquid K medium for further growth and analysis. In some cases, two or more such clones were retained for further analyses. DNA was prepared after harvesting cells from 100 ml of each clonal culture (cell density $> 2 \times 10^7$ ml^{-1}), and the identification of the species present was determined by polymerase chain reaction (PCR)–mediated amplification of the 18S ribosomal gene region using the specific primers: 18S forward (5′-ACCTGGTTGATCCTGCCAG-3′), 18S reverse (5′- TGATCCTTCCGCAGGTTCAC-3′), euk528f (5′-CCGCGGTAATTCCAGCTC-3′), eits2dir (5′-GTAGGTGAACCTGCGGAAGGA-3′), and etts3rev (5′-GGGGAATCCTTGTTAGTTTC-3′) (Moon-van der Staay et al. 2001; Guillou et al. 2004; Rodriguez et al. 2005). PCR amplifications were done using Promega GoTaq in conditions recommended by the manufacturer, with 1.5 mM $MgCl_2$, then cycling: 1 cycle of 2 min, 95 °C; 35 cycles (1 min, 95 °C; 1 min, 55 °C; 2 min, 72 °C); 1 cycle (4 min, 72 °C); then held at 4 °C until further use. Sixteen of the clonal cultures obtained in this way (16% of the total number of prasinophyte cell lines isolated) were found to be *O. tauri*, which is 100% identical over the 18S rRNA sequence, and the remaining strains belong to *Ostreococcus* sp. clade D, *Bathycoccus* spp., or *Micromonas* sp. The previously characterized OTTH95 (Courties et al. 1994; Derelle et al. 2006), sequenced in 2001, was then included in the analysis, together with the present clonal descendant of this strain, which has been maintained in the laboratory under continuous illumination for 8 years (~3000 cell divisions), giving 18 strains. One of the strains isolated from Leucate lagoon died during the study so that Ch1b marker analysis (see below) including this marker have been performed only for 17 strains. The sequences reported in this study have been deposited in the GenBank database (accession numbers GQ426331-482).

Genetic Markers

In order to optimize the detection of neutral polymorphic markers, we developed a specific search procedure. Because codon usage is extremely biased in these species (Palenik et al. 2007) and this correlates positively with the level of mRNA expression (G.P., unpublished data), synonymous polymorphism is likely to be under selection and would

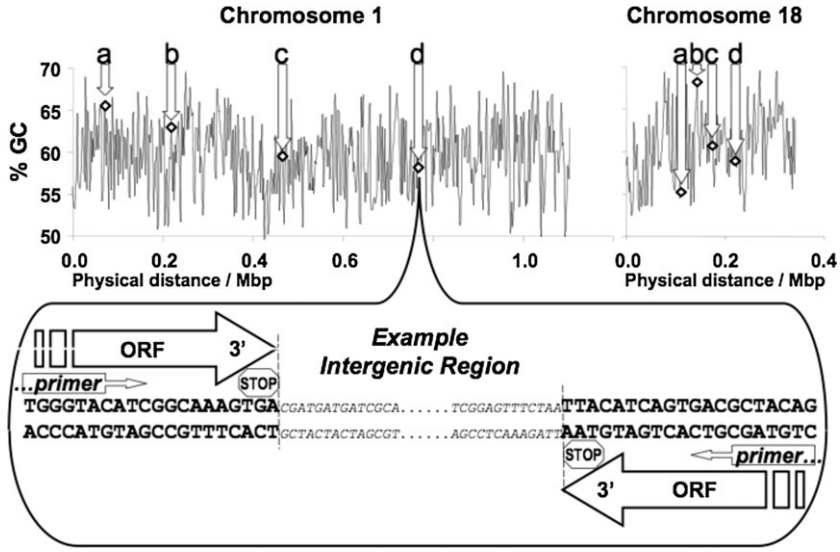

Fig. 2. The chromosomal positions of the eight markers with varying %GC compositions (ordinates) on chromosomes 1 and 18 are indicated by open diamonds. The strategy for PCR amplification of polymorphic regions from conserved flanking sequences is exemplified for one marker (Ch1d).

thus lead to an underestimation of the neutral level of polymorphism. Selection on codon usage is known to occur in several bacterial and eukaryotic species (for a review, see Hershberg and Petrov 2008). We therefore avoided using coding sequences for this study. Because intronic sequences are more constrained than synonymous sites in *Ostreococcus* (Piganeau and Moreau 2007), intergenic regions are the best candidates for investigation of neutral polymorphism. Our choice of eight intergenic markers relies on the analysis of the distribution of constraint in intergenic regions based on the whole-genome comparison of two *Ostreococcus* genomes from different species of *O. tauri* and *O. lucimarinus* (Piganeau et al. 2009). In the latter study, we showed that tail-to-tail intergenic regions, that is, intergenic regions bordered by two stop codons, are less constrained than head-to-head (intergenic regions bordered by two start codons) or head-to-tail (intergenic regions bordered by one stop and one start codon) gene orientations. We wrote a script to screen chromosomes 1 and 18 for four approximately evenly spaced tail-to-tail intergenic regions between 200 and 400 bp long. To avoid spurious intergenic regions due to annotation errors, we considered only orthologous intergenic regions, that is, intergenic regions bordered by two genes having an orthologous gene in the other species. We thus additionally ensured that PCR primers flanking the regions bearing neutral polymorphisms to be amplified fell into conserved protein-coding sequences, thereby maximizing the probability these would target sufficiently similar orthologous sequences among all the strains being studied. Less than a dozen such candidates on chromosomes 1 and 18 remained, and we picked four intergenic regions on each chromosome with a comparable distance between each marker. The position of these markers on *O. tauri*'s genome sequence and the nature of the bordering coding sequences are given in figure 2 and table 1.

Data Analysis

Sequences were aligned manually using Se–Al (Rambaut 2002) and compared with the genome sequence of the original isolate OTTH0595 to identify possible polymorphisms. All polymorphic sites observed were checked on the chromatogram of both the reverse and the forward strands. Special attention was given to the possible occurrence of double peaks, the hallmark of heterozygosity and thus diploidy, but no cases of double peaks were observed at these loci. The nature of polymorphism (nucleotide or indel) and Watterson's infinite-sites estimator of the

Table 1. Marker Choice on *Ostreococcus tauri*'s Chromosomes.

Name	Length	GenBank Accession of Flanking Genes	Positions on Chromosome	Primers (5′–3′)	
				Forward	Reverse
Ch1a	222	CAL49960.1–CAL49961.1	72 478–72 699	TCCAAACTTACGAAGAGAAGG	CTTCGTGTACCAGAACTTGC
Ch1b	290	CAL50060.1–CAL50061.1	219 590–219 879	CAGCCTCATCAACGAAGG	ATTGTCGGCGGTAGTACG
Ch1c	205	CAL50210.1–CAL50211.1	466 415–466 619	GAGCTAAATTTAAGAGGGAACG	GTGAAACCACCGAAGAAACC
Ch1d	212	CAL50326.1–CAL50327.1	677 385–677 596	CCTCTTCGTCTCCCTCGTTC	ACCCGTCGTTTTTGTATTCG
Ch18a	275	CAL58515.1–CAL58516.1	111 494–111 768	GGTTCAATTCGTTCATGGAG	CAGTCGACGCTATGGAATC
Ch18b	243	CAL58540.1–CAL58541.1	143 948–144 190	TAGACGGGACCGAGTTGG	GCTTCCGAGATCGATTACC
Ch18c	359	CAL58558.1–CAL58559.1	174 659–175 018	CGAAAGAAGAAGGAATCTGC	GGAGAAGATGCAGGAAGG
Ch18d	370	CAL58590.1–CAL58591.1	220 231–220 601	GTGAACGACAAGTACGAAGG	GAACGTATCCATGCAGATCC

Table 2. Polymorphisms Found for Each Marker.

Locus	No. of Sequences	nt pol[a]	Indel[b]	Length	θ[c]	π[c]	Haplotypes
Ch1a	18	14	1	222	0.0168	0.0204	8
Ch1b	18	9	2	280	0.0092	0.0086	7
Ch1c	18	4	0	204	0.0056	0.0042	4
Ch1d	17	7	3	203	0.0100	0.0067	4
Ch1abcd	17	37	6	912	0.0118	0.0107	14
Ch18a	18	9	1	291	0.0088	0.0127	5
Ch18b	18	9	4	506	0.0051	0.0047	7
Ch18c	18	13	0	359	0.0103	0.0038	5
Ch18d	18	11	0	370	0.0079	0.0077	7
Ch18abcd	18	42	5	1526	0.0086	0.0066	14

[a] nt pol: Nucleotide polymorphisms.
[b] indel: Insertion or deletion polymorphisms.
[c] Nucleotide polymorphism estimates per site.

population-scaled mutation rate θ (Watterson 1975), $\theta = 2N_e\mu$ for haploids (where N_e is the effective population size and μ is the per-site mutation rate), were estimated for each locus and for the concatenated sequences for each chromosome.

$\theta = P/L\sum_{i=1}^{n-1}\frac{1}{i}$, where P is the number of polymorphisms, L is the number of sites, and n is the number of sequences sampled. We also estimated θ by the average number of pairwise differences per site between individuals in the sample (Tajima 1983), π, or average heterozygosity (Hartl and Clark 1997).

$\pi = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} d_{ij}/Ln(n-1)/2$, where d_{ij} is the number of differences between sequence i and sequence j. Both polymorphism estimates should be equal under the infinite site, steady state, neutral evolution model of sequence evolution (Tajima 1989).

We used the SplitsTree4 package (Huson and Bryant 2006) with default parameters, available at http://www.splitstree.org, to infer the network (Bryant and Moulton 2002) of our data.

Empirical and simulation studies of several recombination detection methods suggest that different methods detect different signatures of recombination so that the definitive conclusions about the presence of recombination should not be derived on the basis of a single method (Posada and Crandall 2001; Posada 2002). We used four recombination detection methods to gain evidence of recombination from sequence data (Piganeau and Eyre-Walker 2004; Posada and Crandall 2001), chosen for their power to detect rare recombination events from polymorphism data (Posada and Crandall 2001), as previously described (Piganeau and Eyre-Walker 2004). They are as follows:

1. The "maxchi" test of Maynard Smith (1992) and the "geneconv" test of Sawyer (1989). We considered global P values, which automatically correct for multiple comparisons, for internal fragments. We used the default parameters because there has been no systematic analysis of the behavior of the geneconv test.
2. The relationship between the measure of linkage disequilibrium, r^2 (Hill 1974), and distance between sites, with the significance assessed by the Mantel test (the LDr^2 test). The relationship between the measure of linkage disequilibrium, $|D'|$ (Lewontin 1964), and distance between sites, with the significance assessed by the Mantel test (the LD$|D'|$ test).

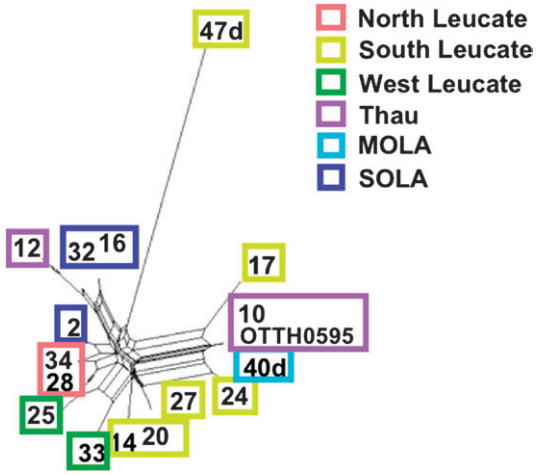

Fig. 3. Haplotype network of the 17 strains based on the polymorphisms of the eight marker sequences obtained using NeighborNet. For clarity, only two digits of the Banyuls Culture Collection numbers are shown (see supplementary table S1 for details of strains, Supplementary Material Online).

One thousand randomizations were performed for each test to assess significance. However, if none of the randomized data sets exceeded the observed value, we repeated the analysis with 10,000 and then 100,000 randomizations. We adapted the RecombiTEST codes available at http://www.lifesci.sussex.ac.uk/CSE/test/index.php (Piganeau et al. 2004) to take indel polymorphisms into account.

We also used the LDhat package (McVean et al. 2002), available at http://www.stats.ox.ac.uk/~mcvean/LDhat/LDhat1.0/LDhat1.0.html, to estimate the population-scaled recombination rate r for each region, $r = 2N_e r$ for haploids (Stumpf and McVean 2003).

Results

We obtained sequences for all markers for each strain, a total of 912 and 1,526 bp of intergenic sequences, corresponding to four markers on the largest chromosome (Ch1), and on one of the smallest chromosomes (Ch18), respectively (table 2). The 18 strains sequenced represented a total of 16 haplotypes if indel and nucleotide polymorphisms are considered and 13 haplotypes if only nucleotide polymorphisms are considered.

Interestingly, different haplotypes were found in the same water sample four times (SOLA, South Leucate, West Leucate, North Leucate), suggesting that *O. tauri*'s populations are polymorphic in the wild at the scale of the sample (i.e., 1 liter of seawater). We found a pair of identical genotypes from the same sample once and once from the same location at different times. Both of these pairs are thus likely to be clones, suggesting in turn that clonal reproduction occurs in the wild, as in laboratory conditions. A network representation of our data set with 13 haplotypes (excluding indel polymorphisms) is presented in figure 3.

Table 3. Output of Recombination Detection Methods.

Locus	maxchi P[a]	genconv P	LD|D'| ρ	P	LD|r^2| ρ	P	LDhat R[b]
Ch1a	0.330	0.220	−0.161	0.109	−0.116	0.124	
Ch1b	ND[c]	0.300	ND	ND	−0.257	0.073	
Ch1c	ND	0.834	ND	ND	ND	ND	
Ch1d	ND	0.250	ND	ND	−0.13	0.037	
Ch1abcd	0.004	0.037	0.085	0.8	−0.128	0.016	22.4
Ch18a	ND	0.589	ND	ND	0.137	0.664	
Ch18b	ND	0.367	0.062	0.631	−0.047	0.382	
Ch18c	0.820	0.798	ND	ND	−0.065	0.275	
Ch18d	0.170	0.336	ND	ND	−0.210	0.080	
Ch18abcd	0.000	0.000	−0.131	0.002	−0.211	0.001	2.9

[a] *P*: *P* value of the recombination detection test.
[b] *R*: Number of recombination events estimated by LDhat pairwise program.
[c] ND: No (or insufficient) data to perform the test.

Haplotypes collected from our six sampling sites are highly interconnected, suggesting that there is no geographical population structure at the scale of our sampling area. We found no correlation between pairwise genetic distance and geographical distance (Spearman's rho = 0.05, *P* value = 0.522), as expected if there is no spatial structure at the scale of our sample (the maximum distance between samples being 108 km).

The average nuclear polymorphism level is 0.0041 (0.0098 if insertion–deletion polymorphisms are included), consistent with the recent 0.0038 estimate of the level of polymorphism found in intergenic regions in the wild yeast *Saccharomyces paradoxus* (Tsai et al. 2008). Interestingly, the laboratory culture of the sequenced strain has one single indel difference with the genome sequence sequenced in 2001.

We screened the alignments for signatures of recombination (table 3). The number of polymorphic sites for each intergenic region was less than 10 for five out of the eight intergenic regions tested so that they were below the level of recombination detection for one of the methods (maxchi). Also, the four possible haplotypes for each pair of polymorphic sites were not present for most intergenic regions taken separately so that the linkage disequilibrium versus distance test measured by D' could not be performed ($|D'| = 1$ for all pairs of sites if one haplotype is lacking). None of the intergenic regions individually provides evidence for recombination (one test is significant for region Ch1d, but this is not significant after correction for multiple tests). However, when all intergenic regions are assembled for each strain, we find significant evidence for recombination on both chromosomes. On chromosome 1, three out of the four recombination detection methods are significant, whereas all four tests of recombination are significant for chromosome 18. Consistent with these results, the estimates of the number of recombination events in the genealogy of the samples are positive with LDhat and equal to 22.4 on chromosome 1 and 2.9 on chromosome 18 (table 3).

We then assessed whether we could find some evidence of chromosomal segregation, another hallmark of meiosis. To assess this, we checked that the recombination point predicted by the maxchi test was between the two chromosomes. Looking at the haplotypes, the hallmark of chromosomal segregation can be observed in figure 4: haplotypes 4 and 5 bearing chromosome 1 of type 1–3 and chromosome 18 of haplotype 9.

Polymorphism data can be used to estimate the number of mutations and recombination events since the last common ancestor of all of our strains, $M = \theta^* L$ and $R = r^* L$, with L the number of sites of the region analyzed (Stumpf and McVean 2003). We can thus estimate the meiosis to mitosis ratio or frequency of sexual reproduction, hereafter f_{SR}. Two additional assumptions about 1) the number of observed recombination events over the L sites considered per meiosis, r_O, and 2) the number of mutations over L per mitosis, m, are needed to go further. Thus, f_{SR} can be expressed as a function of R, M, m, and r_O as:

$$f_{SR} = \frac{\text{number of meioses}}{\text{number of mitoses}} = \frac{R/r_O}{M/m} = \frac{mR}{Mr_O}.$$

Let us assume that each meiosis leads to one crossover event per *Ostreococcus* chromosome, as observed in many organisms (van Veen and Hawley 2003; Jensen-Seaman et al. 2004), that is, $r_O = L/L_{\text{chrom}}$, with L_{chrom} the chromosome length. Let us take the yeast mutation rate estimate as a proxy because there is no available estimate of the rate of mutation per generation in *Ostreococcus* yet. The rate of mutation per

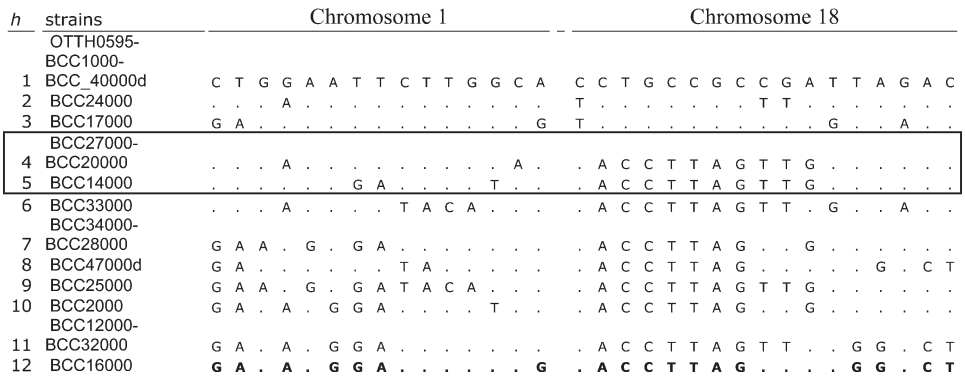

Fig. 4. Evidence of chromosome segregation. *h*: Haplotypes defined by informative sites only (polymorphism with single occurrence excluded). The boxed haplotypes show evidence for interchromosomal recombination.

site per mitosis is 0.33×10^{-9} in yeast (Lynch et al. 2008) so that $m = L^{*}0.33 \times 10^{-9}$.

We have three recombination events, R, on the sequences of chromosome 18 for $M = 13.8$ mutation events (table 3). This thus estimates $f_{SR} = 3 \times 10^{-7}$ for chromosome 18 and $f_{SR} = 1 \times 10^{-6}$ for chromosome 1 and suggests a very low proportion of meiosis to mitosis in the history of our sample. However, this is a lower limit because all methods that estimate recombination rates, including the one we used, just detect the fraction of recombination events that change the genealogy (Hudson and Kaplan 1985), which depends in turn on the frequency of crossing between different haplotypes.

Discussion

We characterized the level of nuclear genetic diversity in the unicellular picophytoplanktonic species *O. tauri* using eight intergenic marker sequences from 17 isolated strains and one complete genome sequence. The polymorphism data we collected gave indirect evidence of crossing-over and chromosomal segregation, revealing that *O. tauri* populations reproduce sexually. Given the homogeneity and low values of polymorphism levels across intergenic regions, we can rule out the possibility that mutational hotspots may have produced the pattern observed (for a review, see Innan and Nordborg 2002). However, we found no evidence that the genetic distance between our strains was in any way related to their geographical position, whether in a seawater lagoon, in Banyuls' Bay, or further out to sea at the MOLA sampling station. Perhaps this result is not so surprising as the water in this coastal region is mixed by currents created by the wind, the general flow in the Gulf of Lion, and by more limited exchanges of seawater through the narrow channels connecting lagoons such as Leucate to the coastal sea [the "Grau de Leucate" and the "Grau de St Ange," see Arnaud and Raimbault (1969)]. Additionally, the South East of France is subject to frequent strong winds, and marine microorganisms are known to be dispersed in aerosols (Aller et al. 2005). Furthermore, several studies on protist diversity aimed at assessing whether "morphospecies," which is indistinguishable through morphology, were worldwide distributed or endemic, using molecular markers (Foissner 2008). Evidence for both ubiquitous (same sequence different locations) and cryptic species have been identified in Micromonas morphospecies (Slapeta et al. 2006) on worldwide distributed collection sites. Whether population structure can been assessed over a worldwide scale would require collection of *O. tauri* strains from various worldwide locations, but this has not yet been done.

Taken together, our sequence data also show clearly that the cultured cell lines are haploid because careful inspection of sequencing profiles showed no ambiguities of bases that would be caused by amplification of heterozygous allele pairs; there was thus no evidence for within-clone polymorphism. No such polymorphisms were observed when the OT95 genome was completely sequenced either (Derelle et al. 2006). The expected proportion of heterozygous sites that would be present if individuals were diploid can be estimated from the level of polymorphism estimated from the average pairwise difference between sequences (table 2). Because π is around 0.009 over 2,470 sequenced base pairs, we would expect at least 21 ambiguous sites per sequenced strain if the strains were diploid. As no such ambiguities were observed among the 17 strains studied, we are confident that this species is haploid in the laboratory. However, we cannot completely exclude the possibility that *O. tauri* is diploid in the wild, the haploid state being induced by the laboratory culture process, because there are many experimental manipulation steps between sampling water to isolating a clonal culture, including several dilutions. Nevertheless, most Chlorophyta so far investigated are haploids for most of their life cycle (Davies and Grosman 1998; Lee et al. 2008; Togashi and Cox 2008), so haploidy in the wild seems likely.

The high proportion of *Ostreococcus* spp. among isolated phytoplankton clones is due in part to the small filter exclusion limit chosen for sample preparation (1.2 μm); relatively few eukaryotic species are so small. Baker's yeast cells, for example, are typically 3–4 μm in diameter. Wild-type *O. tauri* first observed in nature from the Thau lagoon, when observed by transmission electron microscopy in sections, were 0.97 ± 0.28 μm long and 0.7 ± 0.17 μm wide (Courties et al. 1994) and thus could easily pass through; even diploid *O. tauri* cells are likely to pass through (assuming that they have double the cell volume). Although we have no information about the rigidity of the cell surface, no cell wall or envelope has been observed, and the cell membrane can be deformed by apposed organelles (Henderson et al. 2007), hinting that its flexibility might facilitate passage through even smaller pores.

Because most picoplanktonic species cannot currently be maintained in culture in laboratory conditions, direct sequencing of PCR-amplified–cloned 18S rRNA gene fragments from filtered seawater has been successfully used to investigate their astounding diversity (Lopez-Garcia et al. 2001; Moon-van der Staay et al. 2001; Moreira and Lopez-Garcia 2002), and the description of many planktonic species resides in their sole 18S rRNA sequence. Understandably, the frontiers of the species concept is still a matter of debate in single-cell organisms (Schlegel and Meisterfeld 2003), and the "pragmatic" definition of species can either rely on morphological observation "morphospecies" or on molecular markers "operational taxonomic units" (Weisse 2008). Morphospecies have been shown to contain several cryptic species in many lineages (Weisse 2008), and 100% identity on a marker gene as the 18S rRNA may also underestimate the number of species in some lineages. Genetic diversity studies routinely use a 1–2% identity on 18S rRNA to delineate operational taxonomic units in planktonic species. Both the range of intergenic polymorphism, close to previously described polymorphism in unicellular organisms (Lynch et al. 2008; Tsai et al. 2008), and the indirect evidence of recombination from samples identified from their 18S rRNA a posteriori give some credit to the use of the 100% sequence identity of the 18S rRNA gene

as an appropriate marker for *O. tauri* species' identification, species sensu biological species concept as defined by Mayr (1942).

Although we conclude that sexual exchange of genetic information in *O. tauri* occurs in nature, its mechanism remains enigmatic. In fact, the general mechanism of cell division in the Mamiellales is unclear because the cellular volume of these species is too small to accommodate a classical eukaryotic spindle, an apparatus which is additionally involved in mechanical separation of chromosomes and resolution of chiasmata at meiosis. Separation of ~20 chromosomes would require 40 canonical microtubules, whereas only one has so far been observed (Henderson et al. 2007). The cells of budding yeast have about 60 times the volume of *O. tauri* cells; its microtubules are about 40 nm wide, and the spindle stretches out from 0.8 to 8 μm long during cell division (Westermann et al. 2007; Bouck et al. 2008). In the absence of experimental evidence, we might speculate that division might occur in an unconventional way, with one single pair of microtubules, relying on a yet undiscovered chromosomal segregation process. Could such a mechanistic reduction have been driven by selection for small cell size in a nutrient-limited environment and thus be the key step contributing to the global success of eukaryotic picoplankton?

Supplementary Material

Supplemetary table S1 is available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We would like to thank Pascal Romans and the team of the Néréis II boat for providing water samples. We are grateful to the Genomics of phytoplankton team, Laure Bellec, Yves Desdevises, Rozenn Thomas for stimulating discussions, and Lucie Subirana for her kind and efficient technical assistance. This work was funded by GDR1928 to G.P. and ANR-07-BLAN-0210-01 "PICOVIR" to N.G.

References

Aller JY, Kuznetsova MR, Jahns CJ, Kemp PF. 2005. The sea surface microlayer as a source of viral and bacterial enrichment in marine aerosols. *J Aerosol Sci*. 36:801–812.

Arnaud P, Raimbault R. 1969. L'etang de Salses-Leucate ses principaux caracteres physico-chimiques et leurs variations (en 1955–1956 et de 1960 à 1968). *Rev Trav Inst Pêches Maritimes (ISTPM)*. 33:335–443.

Bouck DC, Joglekar AP, Bloom KS. 2008. Design features of a mitotic spindle: balancing tension and compression at a single microtubule kinetochore interface in budding yeast. *Annu Rev Genet*. 42:335–359.

Bryant D, Moulton V. 2002. NeighborNet: an agglomerative method for the construction of planar phylogenetic networks. In: R Guigo, and D Gusfield, editors. Berlin: Springer. p. 375–1291.

Catania F, Wurmser F, Potekhin AA, Przybos E, Lynch M. 2009. Genetic diversity in the *Paramecium aurelia* species complex. *Mol Biol Evol*. 26:421–431.

Chepurnov VA, Mann DG, von Dassow P, Vanormelingen P, Gillard J, Inze D, Sabbe K, Vyverman W. 2008. In search of new tractable diatoms for experimental biology. *Bioessays*. 30:692–702.

Coleman AW. 2005. Paramecium aurelia revisited. *J Eukaryot Microbiol*. 52:68–77.

Courties C, Vaquer A, Troussellier M, Lautier J, Chretiennot-Dinet MJ, Neveux J, Machado C, Claustre H. 1994. Smallest eukaryotic organism. *Nature*. 370:255.

Davies JP, Grosman AR. 1998. The use of Chlamydomonas (Chlorophyta: Volvocales) as a model algal system for genome studies and the elucidation of photosynthetic processes. *J Phycol*. 34:907–917.

Derelle E, Ferraz C, Rombauts S, et al. (26 co-authors). 2006. Genome analysis of the smallest free-living eukaryote Ostreococcus tauri unveils many unique features. *Proc Natl Acad Sci USA*. 103:11647–11652.

Foissner W. 2008. Protist diversity and distribution: some basic considerations. *Biodivers Conserv*. 17:235–242.

Guillou L, Eikrem W, Chretiennot-Dinet MJ, 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.

Hallmann A. 2003. Extracellular matrix and sex-inducing pheromone in Volvox. *Int Rev Cytol*. 227:131–182.

Hartl DL, Clark AG. 1997. Principles of population genetics. Sinauer Associates, Sunderland (MA).

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

Hershberg R, Petrov DA. 2008. Selection on codon bias. *Annu Rev Genet*. 42:287–299.

Hill WG. 1974. Estimation of linkage disequilibrium in randomly mating populations. *Heredity*. 33:229–239.

Hudson RR, Kaplan NL. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics*. 111:147–164.

Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol*. 23:254–267.

Innan H, Nordborg M. 2002. Recombination or mutational hot spots in human mtDNA? *Mol Biol Evol*. 19:1122–1127.

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.

Jensen-Seaman MI, Furey TS, Payseur BA, Lu Y, Roskin KM, Chen CF, Thomas MA, Haussler D, Jacob HJ. 2004. Comparative recombination rates in the rat, mouse, and human genomes. *Genome Res*. 14:528–538.

Keller M, Guillard R. 1985. Factors significant to marine dinoflagellate culture. In: Anderson DM, White AW, Baden DG, editors. Toxic dinoflagellates. New York: Elsevier. p. 113–116.

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

Lewontin RC. 1964. The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics*. 49:49–67.

Li W. 1994. Primary production of prochlorophytes, cyanobacteria, and eukaryotic ultraphytoplankton: measurements from flow cytometric sorting. *Limnol. Oceanogr*. 39:169–175.

Lopez-Garcia P, Rodriguez-Valera F, Pedros-Alio C, Moreira D. 2001. Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature*. 409:603–607.

Lynch M, Sung W, Morris K, et al. (11 co-authors). 2008. A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc Natl Acad Sci USA*. 105:9272–9277.

Massana R, Pedros-Alio C. 2008. Unveiling new microbial eukaryotes in the surface ocean. *Curr Opin Microbiol*. 11:213–218.
Maynard Smith J. 1992. Analyzing the mosaic structure of genes. *J Mol Evol*. 34:126–129.
Mayr E. 1942. Systematics and the origin of species. New York: Columbia University Press.
McMullin ST, Lonergan TP, Rynearson CS. 2005. Twelve-month drug cost savings related to use of an electronic prescribing system with integrated decision support in primary care. *J Manag Care Pharm*. 11:322–332.
McVean G, Awadalla P, Fearnhead P. 2002. A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics*. 160:1231–1241.
Moon-van der Staay SY, De Wachter R, Vaulot D. 2001. Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature*. 409:607–610.
Moreira D, Lopez-Garcia P. 2002. The molecular ecology of microbial eukaryotes unveils a hidden world. *Trends Microbiol*. 10: 31–38.
Palenik B, Grimwood J, Aerts A, et al. (38 co-authors). 2007. The tiny eukaryote Ostreococcus provides genomic insights into the paradox of plankton speciation. *Proc Natl Acad Sci USA*. 104:7705–7710 .
Piganeau G, Eyre-Walker A. 2004. A reanalysis of the indirect evidence for recombination in human mitochondrial DNA. *Heredity*. 92:282–288.
Piganeau G, Gardner M, Eyre-Walker A. 2004. A broad survey of recombination in animal mitochondria. *Mol Biol Evol*. 21: 2319–2325.
Piganeau G, Moreau H. 2007. Screening the Sargasso Sea metagenome for data to investigate genome evolution in Ostreococcus (Prasinophyceae, Chlorophyta). *Gene*. 406: 184–190.
Piganeau G, Vandepoele K, Gourbiere S, Van de Peer Y, Moreau H. 2009. Unravelling cis-regulatory elements in the genome of the smallest photosynthetic eukaryote: phylogenetic footprinting in Ostreococcus. *J Mol Evol*.
Posada D. 2002. Evaluation of methods for detecting recombination from DNA sequences: empirical data. *Mol Biol Evol*. 19:708–717.
Posada D, Crandall KA. 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc Natl Acad Sci USA*. 98:13757–13762.
Proschold T, Harris EH, Coleman AW. 2005. Portrait of a species: Chlamydomonas reinhardtii. *Genetics*. 170:1601–1610.
Rambaut A. 2002. Se-Al is an application for creating multiple sequence alignments from nucleotide and amino acid sequences. [cited 2009 September 14]; Available from: http://tree.bio.ed.ac.uk/software/seal/.
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.
Sawyer S. 1989. Statistical tests for detecting gene conversion. *Mol Biol Evol*. 6:526–538.
Schlegel M, Meisterfeld R. 2003. The species problem in protozoa revisited. *Eur J Protistol*. 39:349–355.
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.
Snoke MS, Berendonk TU, Barth D, Lynch M. 2006. Large global effective population sizes in Paramecium. *Mol Biol Evol*. 23:2474–2479.
Stockner JG. 1988. Phototrophic picoplankton—an overview from marine and fresh-water ecosystems. *Limnol Oceanogr*. 33: 765–775.
Stumpf MP, McVean GA. 2003. Estimating recombination rates from population-genetic data. *Nat Rev Genet*. 4:959–968.
Suda S, Watanabe M, Inouye I. 1989. Evidence of sexual reproduction in the primitive green alga Nephroselmis Olivacea (Prasinophyceae). *J Phycol*. 25:596–600.
Suda S, Watanabe M, Inouye I. 2004. Electron microscopy of sexual reproduction in Nephroselmis olivacea (Prasinophyceae, Chlorophyta). *J Phycol*. 52:273–283.
Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics*. 164:1667–1676.
Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 123:185–195.
Telford RJ, Vandvik V, Birks HJ. 2006. Dispersal limitations matter for microbial morphospecies. *Science*. 312:1015.
Togashi T, Cox PA. 2008. Equal sex ratios of a marine green alga, Bryopsis plumosa. *J Integr Plant Biol*. 50:648–652.
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.
van Veen JE, Hawley RS. 2003. Meiosis: when even two is a crowd. *Curr Biol*. 13:R831–R833.
Vaulot D, Eikrem W, Viprey M, Moreau H. 2008. The diversity of small eukaryotic phytoplankton (< or =3 microm) in marine ecosystems. *FEMS Microbiol Rev*. 32:795–820.
Waaland J, Stiller J, Cheney D. 2004. Macroalgal candidates for genomics. *Journal of Phycology*. 40:26–33.
Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol*. 7: 256–276.
Weisse T. 2008. Distribution and diversity of aquatic protists: an evolutionary and ecological perspective. *Biodivers Conserv*. 17: 243–259.
Westermann S, Drubin DG, Barnes G. 2007. Structures and functions of yeast kinetochore complexes. *Annu Rev Biochem*. 76:563–591.
Worden AZ. 2006. Picoeukaryote diversity in coastal waters of the Pacific Ocean. *Aquatic Microbial Ecol*. 43:165–175.
Worden AZ, Lee JH, Mock T, et al. 2009. Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes Micromonas. *Science*. 324:268–272.