References and Notes

1. M. Tibayrenc, S. Ben Abderrazak, F. Guerrini, A. Banuls, *Arch. Inst. Pasteur Tunis* **70**, 375 (1993).
2. J. M. Kelly, J. M. Law, C. J. Chapman, G. J. Van Eys, D. A. Evans, *Mol. Biochem. Parasitol.* **46**, 253 (1991).
3. D. Nolder, N. Roncal, C. R. Davies, A. Llanos-Cuentas, M. A. Miles, *Am. J. Trop. Med. Hyg.* **76**, 573 (2007).
4. C. Ravel *et al.*, *Int. J. Parasitol.* **36**, 1383 (2006).
5. L. Jenni *et al.*, *Nature* **322**, 173 (1986).
6. M. W. Gaunt *et al.*, *Nature* **421**, 936 (2003).
7. A. A. Capul, T. Barron, D. E. Dobson, S. J. Turco, S. M. Beverley, *J. Biol. Chem.* **282**, 14006 (2007).
8. S. Martinez-Calvillo *et al.*, *Mol. Biochem. Parasitol.* **116**, 147 (2001).
9. P. B. Joshi, J. R. Webb, J. E. Davies, W. R. McMaster, *Gene* **156**, 145 (1995).
10. A. A. Capul, S. Hickerson, T. Barron, S. J. Turco, S. M. Beverley, *Infect. Immun.* **75**, 4629 (2007).
11. Materials and methods are available as supporting material on *Science* Online.
12. D. E. Dobson, L. D. Scholtes, P. J. Myler, S. J. Turco, S. M. Beverley, *Mol. Biochem. Parasitol.* **146**, 231 (2006).
13. J. Lukes *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 9375 (2007).
14. A. C. Ivens *et al.*, *Science* **309**, 436 (2005).
15. T. Thiel, R. Kota, I. Grosse, N. Stein, A. Graner, *Nucleic Acids Res.* **32**, e5 (2004).
16. A. K. Cruz, R. Titus, S. M. Beverley, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 1599 (1993).
17. C. Ravel, P. Dubessay, P. Bastien, J. M. Blackwell, A. C. Ivens, *Parasitol. Today* **14**, 301 (1998).
18. M. E. Moody, L. D. Mueller, D. E. Soltis, *Genetics* **134**, 649 (1993).
19. D. E. Dobson *et al.*, *J. Biol. Chem.* **278**, 28840 (2003).
20. M. D. Rose, *Annu. Rev. Cell Dev. Biol.* **12**, 663 (1996).
21. W. Gibson, L. Garside, M. Bailey, *Mol. Biochem. Parasitol.* **51**, 189 (1992).
22. J. Heitman, *Curr. Biol.* **16**, R711 (2006).
23. M. Hope *et al.*, *Mol. Biochem. Parasitol.* **104**, 1 (1999).
24. C. M. Turner, G. Hide, N. Buchanan, A. Tait, *Exp. Parasitol.* **80**, 234 (1995).
25. J. C. Dujardin *et al.*, *Acta Trop.* **59**, 293 (1995).
26. J. M. Schwenkenbecher *et al.*, *Int. J. Parasitol.* **36**, 237 (2006).
27. I. L. Mauricio, M. K. Howard, J. R. Stothard, M. A. Miles, *Parasitology* **119**, 237 (1999).
28. M. E. Grigg, S. Bonnefoy, A. B. Hehl, Y. Suzuki, J. C. Boothroyd, *Science* **294**, 161 (2001).

29. This research was supported in part by the Intramural Research Program of the NIH, National Institute of Allergy and Infectious Diseases (NIAID), and in part by NIAID grant support (S.M.B., D.E.D., and N.S.A. A1029646 and A1020941). We thank K. Owens for inventorying and shipping parasite strains; K. Beacht for mouse infection studies; and M. Grigg, L. Miller, and A. Sher for critical review of the manuscript. Maxicircle sequences have been submitted to GenBank and had been assigned accession numbers FJ349262 to FJ349263 (12*S* rRNA) and FJ349264 to FJ349265 (Divergent region).

Supporting Online Material

www.sciencemag.org/cgi/content/full/324/5924/265/DC1
Materials and Methods
Figs S1 to S5
Tables S1 to S3
References

8 December 2008; accepted 18 February 2009
10.1126/science.1169464

Green Evolution and Dynamic Adaptations Revealed by Genomes of the Marine Picoeukaryotes *Micromonas*

Alexandra Z. Worden,[1]* Jae-Hyeok Lee,[2]† Thomas Mock,[3]†‡ Pierre Rouzé,[4]† Melinda P. Simmons,[1]† Andrea L. Aerts,[5] Andrew E. Allen,[6] Marie L. Cuvelier,[1,7] Evelyne Derelle,[8] Meredith V. Everett,[7] Elodie Foulon,[9] Jane Grimwood,[5,10] Heidrun Gundlach,[11] Bernard Henrissat,[12] Carolyn Napoli,[13] Sarah M. McDonald,[1] Micaela S. Parker,[3] Stephane Rombauts,[4] Aasf Salamov,[5] Peter Von Dassow,[9] Jonathan H. Badger,[6] Pedro M. Coutinho,[11] Elif Demir,[1] Inna Dubchak,[5] Chelle Gentemann,[14] Wenche Eikrem,[15] Jill E. Gready,[16] Uwe John,[17] William Lanier,[18] Erika A. Lindquist,[5] Susan Lucas,[5] Klaus F. X. Mayer,[10] Herve Moreau,[8] Fabrice Not,[9] Robert Otillar,[5] Olivier Panaud,[19] Jasmyn Pangilinan,[5] Ian Paulsen,[20] Benoit Piegu,[19] Aaron Poliakov,[5] Steven Robbens,[4] Jeremy Schmutz,[5,10] Eve Toulza,[21] Tania Wyss,[22] Alexander Zelensky,[23] Kemin Zhou,[5] E. Virginia Armbrust,[3] Debashish Bhattacharya,[18] Ursula W. Goodenough,[2] Yves Van de Peer,[4] Igor V. Grigoriev[5]

Picoeukaryotes are a taxonomically diverse group of organisms less than 2 micrometers in diameter. Photosynthetic marine picoeukaryotes in the genus *Micromonas* thrive in ecosystems ranging from tropical to polar and could serve as sentinel organisms for biogeochemical fluxes of modern oceans during climate change. These broadly distributed primary producers belong to an anciently diverged sister clade to land plants. Although *Micromonas* isolates have high 18*S* ribosomal RNA gene identity, we found that genomes from two isolates shared only 90% of their predicted genes. Their independent evolutionary paths were emphasized by distinct riboswitch arrangements as well as the discovery of intronic repeat elements in one isolate, and in metagenomic data, but not in other genomes. Divergence appears to have been facilitated by selection and acquisition processes that actively shape the repertoire of genes that are mutually exclusive between the two isolates differently than the core genes. Analyses of the *Micromonas* genomes offer valuable insights into ecological differentiation and the dynamic nature of early plant evolution.

Ancestral green algae were of fundamental importance to the eukaryotic greening that shaped the geochemistry of our planet. This process began over a billion years ago when a cyanobacterium was captured by a heterotrophic protist and incorporated as an endosymbiont, giving rise to the first eukaryotic alga (*1*). The extant Prasinophytae retain characteristics that are believed to have been present in the last common ancestor of green algae (chlorophytes) and land plants (streptophytes, including charophyte algae) (*2*). Most prasinophytes within the monophyletic marine order Mamiellales (Fig. 1A and fig. S1), such as *Micromonas*, are tiny (≤2 μm in diameter) and known as picoeukaryotes. *Micromonas* is a motile unicell, with a single chloroplast and mitochondrion (Fig. 1A, inset), first reported as a dominant phytoplankter in the 1950s (*3*) and now recognized as having a global distribution (Fig. 1B) (*4*).

Today's oceans contain a polyphyletic diversity of algae, some with plastids that share ancestry with land plants (green algae) and others (chromalveolates) that are derived from red algae through secondary or tertiary (eukaryotic-eukaryotic) endosymbioses (*5*, *6*). Unlike most episodic chromalveolate bloomers and the freshwater green alga *Chlamydomonas* (*7*), the Mamiellales have reduced genomes, as first shown in *Ostreococcus* (*8*, *9*). *Ostreococcus* has a narrower environmental distribution than *Micromonas* (Fig. 1B) and a smaller genome (12 to 13 Mb containing only ~8000 genes). Open-ocean bacteria, including SAR11 and *Prochlorococcus* (*10*, *11*), show similar patterns of cell size and genome minimization. Conditions favoring picophytoplankton growth, such as increased stratification, less mixing, and reduced nutrient concentrations in ocean surface waters, are predicted climate change outcomes, and thus picoeukaryote dynamics may be useful ecosystem indicators.

We sequenced the nuclear genomes of *Micromonas* isolates RCC299 and CCMP1545 (Table 1 and figs. S2 and S3) (*12*). These isolates are from distant ocean provinces and fall into distinct phylogenetic clades that can co-occur (Fig. 1) (*12*, *13*) but are generally considered a single species (*Micromonas pusilla*). Transmission electron microscopy revealed no morphological differences (*12*), and 18*S* ribosomal DNA (rDNA) identity was high (97%). Surprisingly, only 90% of their 10,056 (RCC299) and 10,575 (CCMP1545) predicted genes (table S1) were shared (Fig. 2A). In contrast, *Ostreococcus lucimarinus* and *O. tauri* share 97% of cataloged genes (*12*), and yeast genera can share ~95% of homologs (*14*). The divergence we observed between the *Micromonas* isolates supports their classification as distinct species.

Synteny, GC content, and codon usage pointed to a shared evolutionary history for RCC299 and

CCMP1545 but underscored their genomic divergence [supporting online material (SOM) text S1]. Each genome contained a region that had 14% lower than average GC content, composing 7% (RCC299) and 8% (CCMP1545) of the genome (figs. S3 and S4), which also had higher transcriptional activity (SOM text S1). Similar regions in *Ostreococcus* (*8*, *9*) form smaller genome proportions. DNA alignment between RCC299 and CCMP1545 low-GC regions was poor, protein colinearity was absent, and codon usage was different, in contrast to normal GC chromosomes (figs. S4 to S6).

Two major evolutionary themes emerged from our analyses. First, the common ancestor of the Mamiellales had already undergone genomic reduction, highlighted by their organellar genomes (SOM text S2, fig. S7, and tables S2 to S4). Second, *Micromonas* appeared to be less derived than *Ostreococcus*, rendering insights into the genetic composition of the proto-prasinophyte (the common ancestor of plants and prasinophytes) and specialization in extant species. Most "core" nucleus-encoded genes (genes common to the four Mamiellales genomes) were found to have known functions (Fig. 2, A and B) in key pathways (SOM text S3 to S6, tables S5 to S9, and fig. S8), such as photosynthesis, and included selenoproteins (SOM text S3 and table S10). A significant proportion of genes grouped with land plants (Fig. 2C). Core genes branching with chromalveolates (mostly diatoms and brown algae) (Fig. 2C) presumably reflected losses (or extensive divergence) in other green lineage organisms and red algae or perhaps horizontal gene transfer (HGT).

The proto-prasinophyte features we discovered in *Micromonas* included transcription factors that probably belong to the "basal green toolkit" (SOM text S7, figs. S9 to S11, and table

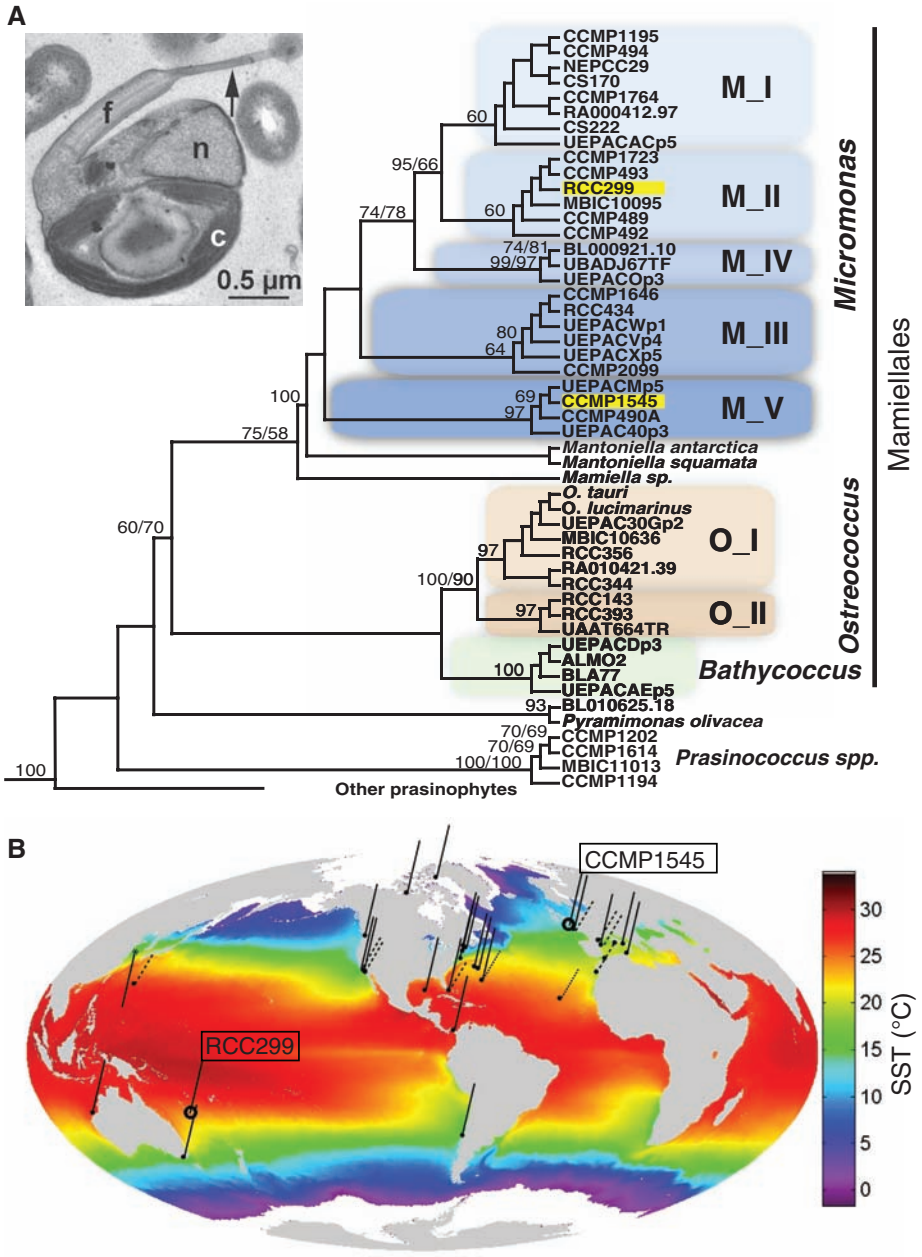


Fig. 1. *Micromonas* phylogeny and distribution. (**A**) A consensus neighbor-joining (NJ) distance 18*S* rRNA gene tree illustrating the distinct *Micromonas* clades (*12*). Bootstrap values represent a percent of 1000 replicates (NJ), and where provided the second value represents the maximum-likelihood bootstrap percentages. The genome isolates sequenced in this work are highlighted (yellow). The previously sequenced *Ostreococcus tauri* and *O. lucimarinus* neighbor each other in clade O_I. The relationship to plants and other photosynthetic lineages is shown in fig S1. (Inset) *Micromonas* thin section showing the nucleus (n), chloroplast (c), flagellum (f), and mucronate extension (the thin tip at the end of the flagellum, indicated by the arrow). (**B**) Mean sea surface temperature (SST) for 2006 measured with global high-resolution SST (GHRSST) blended infrared and microwave SSTs, and locations where *Micromonas* (solid lines and circles around the isolates used in this work) and *Ostreococcus* (dashed lines) 18*S* rDNA sequences have been recovered. *Micromonas* appeared in all temperature regimes.

[1]Monterey Bay Aquarium Research Institute, Moss Landing, CA 95039 USA. [2]Department of Biology, Washington University at St. Louis, St. Louis, MO 63130, USA. [3]School of Oceanography, University of Washington, Seattle, WA 98195, USA. [4]Department of Plant Systems Biology, Flanders Institute for Biotechnology (VIB) and Department of Molecular Genetics, Ghent University, 9052 Gent, Belgium. [5]U.S. Department of Energy (DOE) Joint Genome Institute (JGI), Walnut Creek, CA 94598, USA. [6]J. Craig Venter Institute, San Diego, CA 92121, USA. [7]Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, FL 33149, USA. [8]Observatoire Océanologique, CNRS–Université Pierre et Marie Curie, 66651 Banyuls sur Mer, France. [9]Station Biologique de Roscoff, CNRS–Université Pierre et Marie Curie, Roscoff Cedex, France. [10]Stanford Human Genome Center, Stanford University School of Medicine, Palo Alto, CA 94304, USA. [11]Institute of Bioinformatics and System Biology, German Research Center for Environmental Health, 85764 Neuherberg, Germany. [12]Architecture et Fonction des Macromolécules Biologiques, Universities of Aix-Marseille I and II, Marseille 13288, France. [13]Biology Institute, University of Arizona, Tucson, AZ 85719, USA. [14]Remote Sensing Systems, Santa Rosa, CA 95401, USA. [15]Avdeling for Marinbiologi og Limnologi, University of Oslo, Oslo N-0316, Norway. [16]Division of Molecular Bioscience, College of Medicine, Biology and the Environment, Australian National University, Canberra ACT 2601, Australia. [17]Alfred Wegener Institute for Polar and Marine Research, Am Handelshafen, Bremerhaven 27570, Germany. [18]Department of Biology, University of Iowa, Iowa City, IA 52242, USA. [19]Laboratoire Genome et Development des Plantes Université de Perpignan, 66860 Perpignan, France. [20]Department of Chemistry and Biomolecular Sciences, Macquarie University, New South Wales 2109, Australia. [21]Ecosystèmes Lagunaires, Université Montpellier II, F-34095 Montpellier Cedex 05, France. [22]Department of Biology, University of Miami, Miami, FL 33149, USA. [23]Department of Genetics, Erasmus Medical Center, Rotterdam 3015 CE, Netherlands.

*To whom correspondence should be addressed. E-mail: azworden@mbari.org

†These authors contributed equally to this work.

‡Present address: School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK.

S11). For example, early-branching land plants encode most higher-plant transcription factor families except for the YABBY family (*15*), which was therefore posited to be evolutionarily associated with the development of leaves. However, we found YABBY in *Micromonas*, although it is absent from *Chlamydomonas* and *Ostreococcus*, indicating that it was part of the basal toolkit (fig. S11). We also found diversified homeodomains (fig. S12 and table S12) that are relevant to the evolution of green regulatory networks.

Although prasinophytes are often considered asexual, our observations indicated that the proto-prasinophyte was sexual. First, meiotic-specific and non-meiotic representatives of the *RECA-RAD51*, *TOP6A/SPO11*, and *MUTS* gene families were found (SOM text S5 and table S13). Second, the low-GC regions showed features of sex chromosomes, including RWP-RK transcription factor family genes (SOM text S7 and table S14). Third, numerous Mamiellales genes encoded hydroxyproline-rich glycoproteins (HRGP) (SOM text S6, table S15, and fig. S13), which are cell-wall components in *Chlamydomonas* and plants (*16*). Like the many carbohydrate-active enzymes (SOM text S6 and table S17), this was unexpected because cell walls have not been observed in *Micromonas* or *Ostreococcus* (Fig. 1A, inset) (*4*). In *Chlamydomonas*, one HRGP gene set is expressed only after sexual fusion to produce a thick adhesive zygote wall (*17*). *Micromonas* may behave similarly. Collectively, these data indicate the occurrence of sexual differentiation and the formation of a resistant life-cycle stage.

Fourteen percent of genes were shared between RCC299 and CCMP1545 but not with *Ostreococcus* (Fig. 2, SOM text S3 and S8, table S18, and fig. S14). Shared enzymes for the synthesis and remodelling of peptidoglycan in the plastid provided new insight into the evolutionary history of the ancestral cyanobacterial endosymbiont (SOM text S6) (*18*, *19*). The shared genes also showed more rapid evolutionary rates than core genes (fig. S15), indicating that they escaped constraints acting on the Mamiellales core but still probably play important roles, given their presence in both isolates. Moreover, a larger proportion of "unique" genes (used here to mean genes mutually exclusive between RCC299 and CCMP1545) branched with opisthokont or bac-

Table 1. Characteristics of the *Micromonas* genomes.

Characteristic	CCMP1545	RCC299
Genome size (Mb)	21.9	20.9
G+C (%)	65	64
Number of genes	10,575	10,056
Gene size (bp)	1,557	1,587
Multiexon genes (%)	50	37
Introns (per gene)	0.90	0.57
Intron length (bp)	187	163

Table 2. Genes with associated TPP riboswitches in RCC299, CCMP1545, and *Ostreococcus* (both *O. tauri* and *O. lucimarinus*). The position of the riboswitch relative to the gene is indicated in the columns headed "Riboswitch." DC, domain containing; NF, not found with protein-protein basic local alignment search tool (BLASTP) or protein-nucleotide six-frame translation (TBLASTN). See SOM text S15 for gene descriptions. Protein IDs refer to JGI genome browser protein IDs.

Gene name	RCC299 Protein ID	RCC299 Riboswitch 5′	RCC299 Riboswitch 3′	CCMP1545 Protein ID	CCMP1545 Riboswitch 5′	CCMP1545 Riboswitch 3′	*Ostreococcus* Presence	*Ostreococcus* Riboswitch 5′	*Ostreococcus* Riboswitch 3′
NMT1	102273	no	yes	58387	no	no	NF	-	-
FOLR-like	106264	no	yes	NF	-	-	NF	-	-
EFG-DC	56895	no	yes	NF	-	-	NF	-	-
SSSF-F	NF	-	-	48760	yes	yes	yes	yes	no
SSSF-P	NF	-	-	60112	yes	yes	yes	yes	no

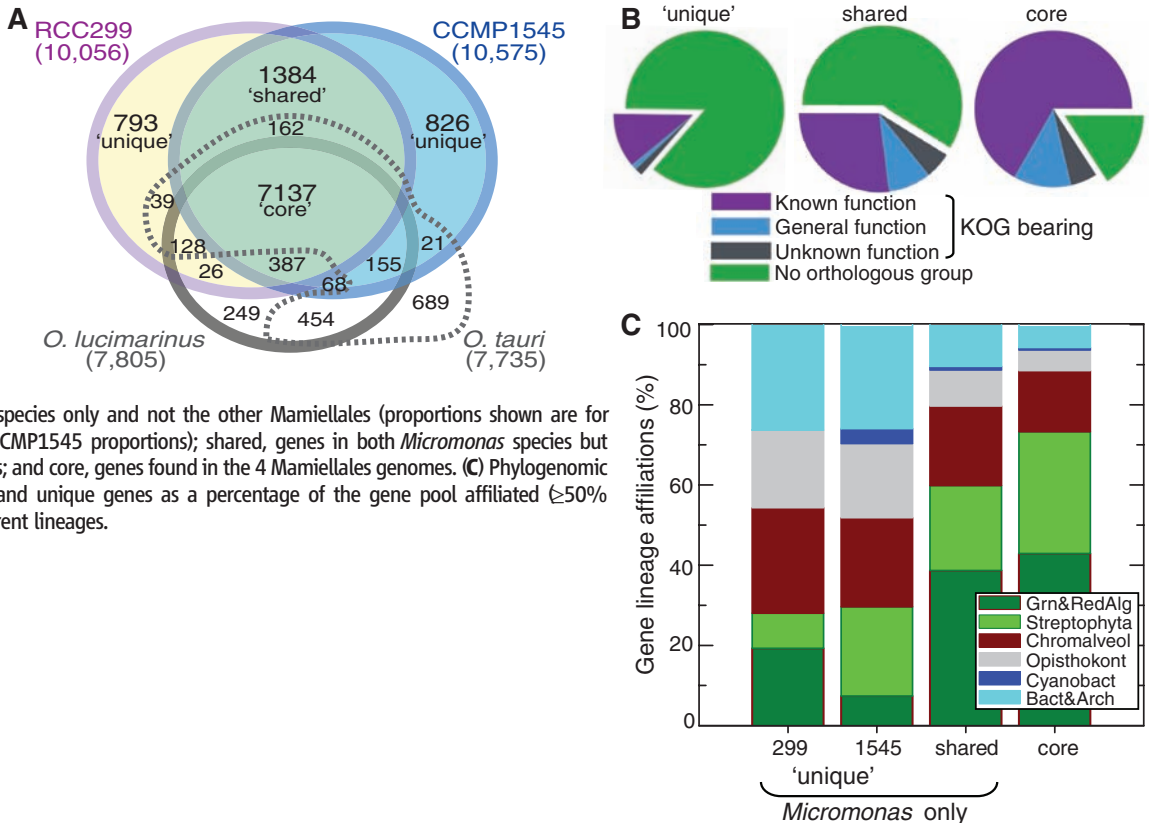


Fig. 2. Comparison of Mamiellales gene complements. (**A**) Venn diagram comparing RCC299 and CCMP1545, *O. tauri* and *O. lucimarinus* gene complements. Circle sizes roughly represent relative numbers of genes in each genome. (**B**) Proportions of genes within eukaryotic orthologous groups (KOGs) and without KOG placement for the gene pools: unique, genes in one *Micromonas* species only and not the other Mamiellales (proportions shown are for RCC299; see fig. S14 for CCMP1545 proportions); shared, genes in both *Micromonas* species but neither *Ostreococcus* species; and core, genes found in the 4 Mamiellales genomes. (**C**) Phylogenomic profiling for core, shared, and unique genes as a percentage of the gene pool affiliated (≥50% bootstrap values) with different lineages.

terial lineages (Fig. 2C), which is consistent with acquisition by means of HGT. Many were of unknown function (Fig. 2B) but may provide useful indicator information. Following early genome reduction, fundamentally different selection/acquisition processes acting on the unique genes appear to have promoted differentiation.

Marked differences in nutrient transport were seen as compared with that in other green-lineage organisms. Between the *Micromonas* species, 52 of the 59 transporter gene families common to land plants were present as well as several transporter gene families found in marine chromalveolates but not in other green-lineage members (SOM text S9 and table S19). Both *Micromonas* spp. had more transporter families represented and higher numbers of transporters than *Ostreococcus*, although CCMP1545 was missing specific transporter gene families, including some related to nitrogen uptake (SOM text S9 and table S19). These differences possibly reflected environmental parameters; for instance, RCC299 is from highly oligotrophic waters, in which nutrient scavenging is essential.

We explored other genomic features related to competition and mortality that influence community structure (SOM text S10 to S13 and figs. S16 to S18). Two types of carbon-concentrating mechanisms (CCM) were identified (SOM text S12 and figs. S17 and S18) that can alleviate CO_2 limitation during blooms. The more unusual *Micromonas* CCM, a C_4-like carbon fixation pathway, includes a nicotinamide adenine dinucleotide phosphate–dependent malic-enzyme (NADP-ME) that is targeted to the plastid lumen, an atypical localization that probably reduces CO_2 leakage (SOM text S12). Because C_4-like pathways have now been identified in the four Mamiellales genomes and in diatoms (SOM text S12), they may represent a fairly basic necessity rather than a rare form of optimization in a few taxa. Both *Micromonas* species appeared to have more robust defenses against heavy-metal toxicity and reactive-oxygen species (SOM text S13 and table S20) than *Ostreococcus*. The larger *Micromonas* genome sizes may thus facilitate broader physiological response capabilities than the *Ostreococcus* genomes.

We found few (CCMP1545) (table S21) or no (RCC299) recognizably functional transposable elements (TEs). Most eukaryotes, including *Ostreococcus* (*9*), contain many TEs, and TE content is positively correlated with genome size above an ~10 Mb threshold (*20*, *21*). Any relic or degenerate TEs in *Micromonas* had low similarity to known TEs, and structural features of class II elements were not found. GC bias was thought responsible for the high proportion of TEs in the low-GC regions of *Ostreococcus* and for loss of synteny in these regions (*9*). However, the low-GC regions of *Micromonas*, although rearranged (fig. S5), had few simple repeats, con-

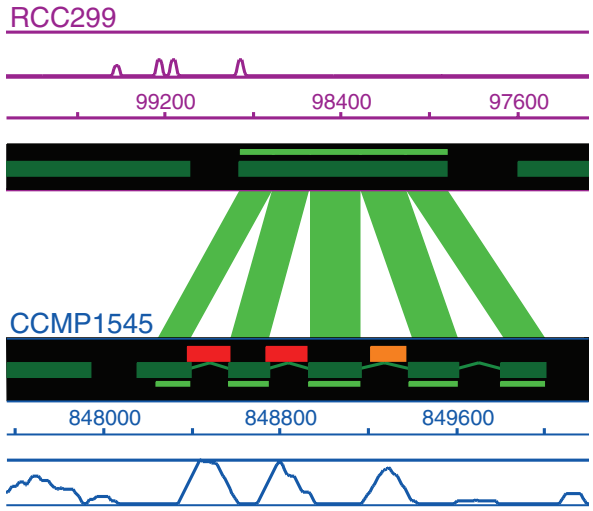
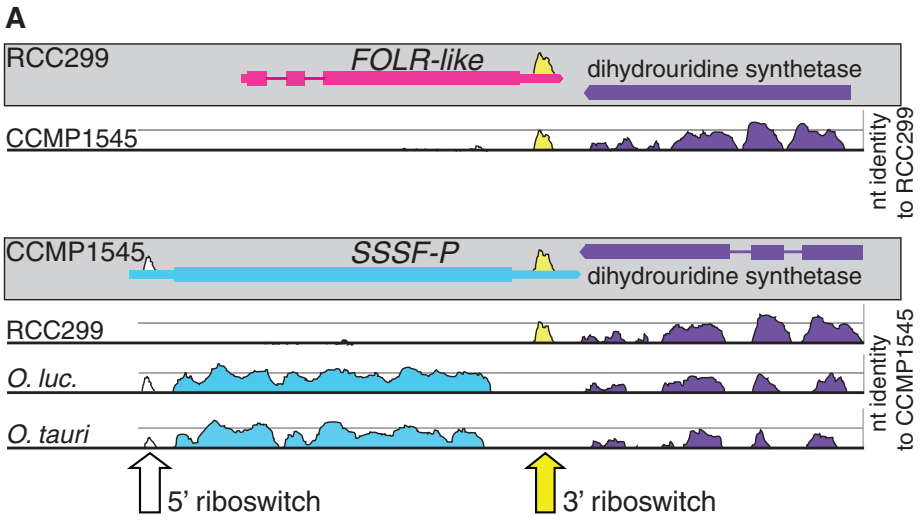


Fig. 3. Depiction of *Micromonas* orthologs with and without IEs. Single-exon (horizontal dark green bars represent exons) RCC299 [JGI protein identification (ID) 84234, chromosome 8] corresponds to a multi-exon gene in CCMP1545 (JGI protein ID 70142, scaffold 11). Different IE elements are shown (red and orange) within introns (horizontal light green lines). Diagonally oriented green lines show syntenic relationships by connecting exons with >70% nucleotide identity [minimum 100 base pairs (bp)]. Purple (RCC299, reversed orientation) and blue (CCMP1545) curves and peaks represent 16-nucleotide oligomer frequencies.



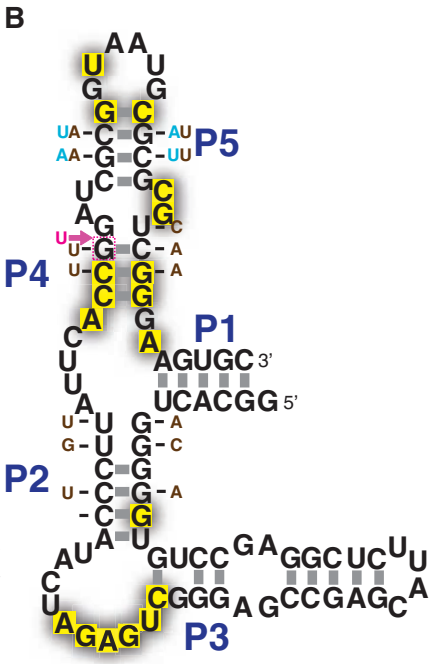


Fig. 4. TPP riboswitch arrangements. (**A**) High nucleotide identity of 3′ riboswitch sequences (yellow profiles) associated with *FOLR*-like (pink; RCC299 only) and *SSSF-P* (blue; CCMP1545) and identity between CCMP1545 and *Ostreococcus* 5′ riboswitches (white profiles) associated with *SSSF-P* homologs (blue). Plant riboswitches are often located in 3′UTRs (*25*), whereas bacterial and fungal riboswitches are often located in 5′UTRs. CCMP1545 has them in both positions. The downstream gene (purple) is a putative dihydrouridine synthase conserved in the four Mamiellales genomes. (**B**) Predicted secondary structure of *FOLR*-like–associated riboswitch showing the positions that are conserved among a range of organisms, particularly plants (yellow background), and a conserved position in all known plant riboswitches but not conserved in *Micromonas* (pink boxed U). Nucleotides adjacent to P2, P4, and P5 regions reflect differences in the CCMP1545 *SSSF-P* 3′ riboswitch (blue) and CCMP1545 *SSSF-F* 5′ riboswitch (brown). Differences in the more variable P1 and P3 are not marked in order to maintain the figure's simplicity.

tained only potential relic TEs, and showed high transcriptional activity (theoretically facilitating TE insertion) (SOM text S1), which suggests TE activity/propagation is actively hindered.

We discovered intronic repeat sequences in CCMP1545 that were absent from RCC299 and other published genomes (SOM text S14, tables S22 and S23, and figs. S19 to S22). These abundant introner elements (IEs) were located within introns, extended nearly to donor and acceptor sites (Fig. 3 and fig. S21), and lacked known TE characteristics (*22*). RCC299 genes generally had fewer introns than IE-bearing CCMP1545 homologs (Fig. 3), and CCMP1545 had a higher overall intron frequency (Table 1). The 9904 IEs fell into four heterogeneously distributed subfamilies (fig. S22 and table S22), making up 9% of the genome. We also found IEs in Sargasso Sea metagenome data (*23*) that have flanking coding domains with a high similarity to CCMP1545 but lower similarity to RCC299. *Micromonas* 18*S* rDNA sequences in the same metagenome data belong to uncultured clade M_IV (Fig. 1A) (*13*). Given the extent of genome reduction, the abundance of IE suggests that they are functional or resistant to purging.

Putative RNA interference (RNAi) components also differed between the *Micromonas* species (SOM text S4 and table S6). Only RCC299 had an argonaute-encoding gene. A version of argonaute is also found in *Chlamydomonas* and plants but not *Ostreococcus*. DEAD box and SDE3 gene analyses provided circumstantial evidence for a diverged RCC299 RNA helicase. Argonaute can act to combat TE invasion (*24*), which is notable given that RCC299 had no recognizable TEs or IEs.

Both *Micromonas* spp. had putative thiamine pyrophosphate (TPP) riboswitches, untranslated mRNAs that regulate gene expression by means of metabolite binding (*25*, *26*). These were not associated with homologous genes nor with known thiamine-biosynthesis–related genes, except for *N-myristoyltransferase 1* (*NMT1*) (Table 2 and SOM text S15). CCMP1545 riboswitches were located at both gene ends (Fig. 4A), an arrangement never before seen, and formed two divergent groups: 5′ riboswitches shared with *Ostreococcus* and 3′ riboswitches shared with RCC299 (Fig. 4B). A conserved 3′ riboswitch was shared between *folate receptor* (*FOLR*)–like (RCC299) and *SSSF-P* (CCMP1545), even though these genes were not held in common; yet *Ostreococcus* also had *SSSF-P* and a 5′ riboswitch (Fig. 4A). Only one of the seven *Micromonas* riboswitches was associated with a multi-exon gene (*FOLR*-like). Thus, it appears that the putative riboswitches in *Micromonas* act akin to bacterial riboswitches and lack the spliceosomal functions that evolved in other eukaryotes (*26*).

Deficiencies in the thiamine-biosynthesis pathway (*27*, *28*) were notable (SOM text S15). However, comparison with other lineages indicated the *Micromonas* riboswitch-containing genes represent ancient thiamine-pathway components. We identified TPP riboswitches associated with *SSSF-P* in SAR11 bacteria, which also lack classical thiamine-biosynthesis genes (*10*), and with *SSSF-F* in *Chlamydomonas* and *Volvox*. The functional importance of the gene-riboswitch associations is supported by the same gene-riboswitch pairings being found in these disparate lineages (SOM text S15).

The *Micromonas* genomes reveal features of the ancestral algae that initiated the billion-year trajectory of the green lineage and the greening of Earth. Their divergence, combined with acquisition strategies that are consistent with HGT, highlight the dynamic nature of marine protistan evolution and provide a springboard for unraveling functional aspects of phytoplankton populations. The challenge now is to identify biogeochemically important features within this natural diversity and apply them in assessing ecological transformations caused by environmental change.

References and Notes

1. J. D. Hackett *et al.*, *Mol. Biol. Evol.* **24**, 1702 (2007).
2. L. A. Lewis, R. M. McCourt, *Am. J. Bot.* **91**, 1535 (2004).
3. E. W. Knight-Jones, P. R. Walne, *Nature* **167**, 445 (1951).
4. A. Z. Worden, F. Not, in *Microbial Ecology of the Oceans,* D. L. Kirchman, Ed. (Wiley, Hoboken, NJ, 2008), pp. 159–205.
5. A. Reyes-Prieto, A. P. Weber, D. Bhattacharya, *Annu. Rev. Genet.* **41**, 147 (2007).
6. C. E. Lane, J. M. Archibald, *Trends Ecol. Evol.* **23**, 268 (2008).
7. S. S. Merchant *et al.*, *Science* **318**, 245 (2007).
8. E. Derelle *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 11647 (2006).
9. B. Palenik *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7705 (2007).
10. S. J. Giovannoni *et al.*, *Science* **309**, 1242 (2005).
11. G. Rocap *et al.*, *Nature* **424**, 1042 (2003).
12. Materials and methods are available as supporting material on *Science* Online.
13. A. Z. Worden, *Aquat. Microb. Ecol.* **43**, 165 (2006).
14. C. Hall, S. Brachat, F. S. Dietrich, *Eukaryot. Cell* **4**, 1102 (2005).
15. S. K. Floyd, J. L. Bowman, *Int. J. Plant Sci.* **168**, 1 (2007).
16. J. H. Lee, S. Waffenschmidt, L. Small, U. Goodenough, *Plant Physiol.* **144**, 1813 (2007).
17. U. Goodenough, H. Lin, J. H. Lee, *Semin. Cell Dev. Biol.* **18**, 350 (2007).
18. T. Cavalier-Smith, *Trends Plant Sci.* **5**, 174 (2000).
19. G. I. McFadden, *Curr. Opin. Plant Biol.* **2**, 513 (1999).
20. B. Gaut, J. Ross-Ibarra, *Science* **320**, 484 (2008).
21. M. Lynch, *The Origins of Genome Architecture* (Sinauer, Sunderland, MA 2007).
22. T. Wicker *et al.*, *Nat. Rev. Genet.* **8**, 973 (2007).
23. J. Venter *et al.*, *Science* **304**, 66 (2004).
24. A. A. Aravin, G. J. Hannon, J. Brennecke, *Science* **318**, 761 (2007).
25. A. Wachter *et al.*, *Plant Cell* **19**, 3437 (2007).
26. M. T. Cheah, A. Wachter, N. Sudarsan, R. R. Breaker, *Nature* **447**, 497 (2007).
27. M. T. Croft, M. Moulin, M. E. Webb, A. G. Smith, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 20770 (2007).
28. D. DellaPenna, R. L. Last, *Science* **320**, 479 (2008).
29. P. Deschamps *et al.*, *Genetics* **178**, 2373 (2008).

30. We thank the Culture Collection for Marine Phytoplankton and Roscoff Culture Collection for providing isolates, in particular F. LeGall, A. Houdan, and D. Vaulot. We also thank R. Gausling, C. Perle, Q. Ren, D. Root, L. Stal, J. Van Wye, T. Weissman, R.M. Welsh, and U. Wollenzien. F. Partensky, N. Simon, P. Deschamps, and S. Ball facilitated chloroplast and starch (*29*) annotations; C. Rancurel and B. Cantarel assisted with carbohydrate-active enzymes (with CNRS funding). We are grateful to S. Giovannoni for thoughtful criticism of the manuscript and overall enthusiasm. Genome sequencing was performed under the DOE Biological and Environmental Research Program contracts DE-AC02-05CH11231, DE-AC52-07NA27344, DE-AC02-06NA25396, and DEFC02-99ER62873. U.W.G. and J.-H.L. were funded by NSF Molecular and Cellular Biosciences (MCB) grant 0326829. Funding carrying the project from inception to completion was provided by a Young Investigator in Marine Microbiology award to A.Z.W. from the Gordon and Betty Moore Foundation with additional funds from NSF MCB grant 0429359 and the Lucille and David Packard Foundation. A.Z.W. coordinated the project and annotation; A.Z.W. and U.W.G. wrote the manuscript with input and sections from J.-H.L., T.M., P.R., and M.P.S. (joint second authors are listed in alphabetical order), and Y.V.P. and D.B. performed intellectually based editing to which S. Rombauts and M.S.P. contributed; I.V.G. coordinated the sequencing and analysis at JGI. A.L.A., A.E.A., M.L.C., E. Derelle, M.V.E., E.F., J.G., H.G., B.H., C.N., S.M.M., M.S.P., S. Rombauts, A.S., and P.V.D. also made substantial contributions (listed in alphabetical order). J.H.B. and A.E.A. constructed the phylogenomic analysis tool. A.Z.W., E.V.A., K.F.X.M., U.W.G., and Y.V.P. supervised analyses; A.Z.W. conceived the study with input from D.B., H.M., and E.V.A. All others contributed as members of the *Micromonas* genome consortium or JGI sequencing and are listed in alphabetical order. RCC299 and CCMP1545 assemblies and annotations are available at www.jgi.doe.gov/MicromonasRCC299 and www.jgi.doe.gov/MicromonasCCMP1545, respectively. Genome assemblies together with predicted gene models and annotations were deposited at DNA Data Bank of Japan/European Molecular Biology Laboratory/GenBank under the project accession numbers ACCO00000000 and ACCP00000000 for RCC299 and CCMP1545, respectively.

Supporting Online Material

www.sciencemag.org/cgi/content/full/324/5924/268/DC1
Materials and Methods
SOM Text
Figs. S1 to S22
Tables S1 to S25
References

14 October 2008; accepted 23 February 2009
10.1126/science.1167222