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

Alexandra Z. Worden,†* Jae-Hyeok Lee,†† Thomas Mock,†‡ Pierre Rouzé,†‡ Melinda P. Simmons,†‡ Andrea L. Aerts,§ Andrew E. Allen,¶ Marie L. Cuvelier,‖ Evelyne Derelle,¶ Meredith V. Everett,†‡ Steven Robbens,¶ Jeremy Schmutz,§,10 Eve Toulza,‖ Tania Wyss,‖ Alexander Zelensky,‖ Olivier Panaud,‖‡ Jasmyn Pangilinan,† Ian Paulsen,‖² Benoit Piegu,‖² Aaron Poliakov,† Ian Paulsen,‖² Benoit Piegu,‖² Aaron Poliakov,†

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 18S 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 ribosomal arrangements as well as the discovery of intron 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.

Ancient 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 (strepotyphyes, including charophyte algae) (2). Most prasinophytes within the monophyletic marine order Mamielales (Fig. 1A and fig. S1), such as *Micromonas*, are tiny (<2 mm 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 (eukaryot-eukaryotic) endosymbioses (5, 6). Unlike most episcopic 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 pico-phytoplankton 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 18S 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 catalyzed 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). Sec- ond, 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 seleno-

proteins (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
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 second GC regions (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 MTS 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, table S15, and fig. S13), which are cell-wall components in Chlamydomonas and plants (16). Like the many carbohydrate-active enzymes (SOM text S6, table S15, and fig. S13), 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.

Table 1. Characteristics of the Micromonas genomes.

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

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.

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Protein ID</th>
<th>Riboswitch 5'</th>
<th>Riboswitch 3'</th>
<th>Presence</th>
<th>Riboswitch 5'</th>
<th>Riboswitch 3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>NMT1</td>
<td>102273</td>
<td>no</td>
<td>yes</td>
<td>NF</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>FOLR-like</td>
<td>106264</td>
<td>no</td>
<td>yes</td>
<td>NF</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>EFG-DC</td>
<td>56895</td>
<td>no</td>
<td>yes</td>
<td>NF</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SSSF-F</td>
<td>NF</td>
<td>-</td>
<td>-</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>SSSF-P</td>
<td>NF</td>
<td>-</td>
<td>-</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>60112</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>

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 S9 transporter gene families common to land plants were present as well as several transporters 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 CO2 limitation during blooms. The more unusual Micromonas CCM, a C4-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 CO2 leakage (SOM text S12). Because C4-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) recognizable 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) or multi-exon (red and orange) gene families are depicted in 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 identity) or 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 viral 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 S21, S22, and figs. S19 to S22). These abundant intron elements (IEs) were located within introns, extended nearly to donor and acceptor sites (Fig. 3 and fig. S21), and lacked known TE characteristics within introns, extended nearly to donor and acceptor sites (Fig. 3, and CCMP1545 that were absent from RCC299 and S22 and S23, and figs. S19 to S22). These abundant intron 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* 18S DNA sequences in the same metagenome data belong to uncultured clade M4 (Fig. 1A) (43). 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 pyrophosphatase (TPP) riboswitches, untranslated miRNAs 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) (27, 28) and 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

12. Materials and methods are available as supporting material on Science Online.

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. Gaulting, C. Perle, Q. Ren, D. Root, L. Stal, J. Van Wye, T. Weissman, R.M. Welsh, and U. Wollenstein. 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 DEF02-99ER62873. U.W.G. and J.-H.L. were funded by NSF Molecular and Cellular Bionics (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 Lucile 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; U.V.P. coordinated the sequencing and analysis at JGI. A.L.A., A.E.A., M.L.C., E. Derelle, M.V.E., E.J., 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. conducted the phylogenetic 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 ACD0000000000 and ACQ0000000000 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
ERRATUM

Reports: “Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*” by A. Z. Worden et al. (10 April, p. 268). The affiliation listed for authors Panaud and Piegu was incorrect. They are at the Laboratoire Genome et Développement des Plantes, UMR CNRS/Institut pour la Recherche et le Développement/University of Perpignan Via Domitia, Université de Perpignan, 66860 Perpignan, France.