Newly identified and diverse plastid-bearing branch on the eukaryotic tree of life

Eunsoo Kim1,4, James W. Harrison1,5,6, Sebastien Sudet1,4, Meredith D. M. Jones1,4,7, Heather M. Wilcox1,4, Thomas A. Richards1,2,7,8,9,10, Alexandra Z. Worden1,2, and John M. Archibald1,2

*Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, NS, Canada B3J 1X5; 2School of Biosciences, University of Exeter, Exeter EX4 4QD, United Kingdom; 3Monterey Bay Aquarium Research Institute, Moss Landing, CA 95039; and 4Department of Zoology, Natural History Museum London, London SW7 5BD, United Kingdom

Edited by Jeffrey D. Palmer, Indiana University, Bloomington, IN, and approved November 22, 2010 (received for review September 6, 2010)

The use of molecular methods is altering our understanding of the microbial biosphere and the complexity of the tree of life. Here, we report a newly discovered uncultured plastid-bearing eukaryotic lineage named the rappemonads. Phylogenies using near-complete plastid ribosomal DNA (rDNA) operons demonstrate that this group represents an evolutionarily distinct lineage branching with haptophyte and cryptophyte algae. Environmental DNA sequencing revealed extensive diversity at North Atlantic, North Pacific, and European freshwater sites, suggesting a broad ecophysiology and wide habitat distribution. Quantitative PCR analyses demonstrate that the rappemonads are often rare but can form transient blooms in the Sargasso Sea, where high 16S rRNA gene copies mL–1 were detected in late winter. This pattern is consistent with these microbes being a member of the rare biosphere, whose constituents have been proposed to play important roles under ecosystem change. Fluorescence in situ hybridization revealed that cells from this unique lineage were 6.6 ± 1.2 × 5.7 ± 1.0 μm, larger than numerically dominant open-ocean phytoplankton, and appear to contain two to four plastids. The rappemonads are unique, widespread, putatively photosynthetic algae that are absent from present-day ecosystem models and current versions of the tree of life.

Photosynthetic marine organisms perform roughly half of the primary production on earth (1). In addition to cyanobacteria, phytoplankton taxa distributed across the eukaryotic tree are responsible for photosynthetic CO2 uptake (2, 3). Eukaryotic phytoplankton taxa differ significantly in relative abundances depending on environmental conditions, as do their contributions to primary production, which are poorly quantified (4–6). Analyses of environmental ribosomal RNA (rRNA) genes have transformed our understanding of such microbial eukaryotes. Eukaryotic algae possess both nuclear (18S) and plastid (16S) rRNA genes, and environmental sequence surveys using both markers have revealed that lineages such as the stramenopiles (heterokonts), haptophytes, and cryptophytes, which have ornate calcium carbonate plates (20), the overall level of plastid evolution is notable, and it presumably results in differing genomic repertoires and physiologies. Contrast, for example, the diatoms Thalassiosira pseudonana and Phaeodactylum tricornutum, which encompasses considerable diversity. Pairwise 16S rDNA identities range down to 94% identity (across a 492-nt fragment) within the group, whereas cryptophytes and haptophytes range down to 96% and 87% identity, respectively. The latter is considered a phylum and includes organisms as disparate as the pavlovalae, the soft-bodied picophytoflagellates that form a major portion of oceanic picoplankton biomass (6), and coccolithophores, which have ornate calcium carbonate plates (20). The overall level of 16S rDNA divergence within the newly discovered group identified herein is notable, and it presumably results in differing genomic repertoires and physiologies. Contrast, for example, the diatom Thalassiosira pseudonana and Phaeodactylum tricornutum.
which share 98% identity over the 492-nt fragment and 96% identity across the entire plastid 16S rRNA gene but whose genomes share only 30% identity across the entire plastid 16S rRNA gene but whose genomes share only 30% identity. Environmental diversity and sample locations. Maximum likelihood phylogenetic tree of environmental 16S rDNA sequences obtained herein (bold) as well as OM270. (Inset) Map shows the approximate positions of the sites sampled. Bootstrap support values (≥ 95%) are from RaxML and Log-Det distance analyses, respectively. (Inset) Map shows the approximate positions of the sites sampled. (Inset) Map shows the approximate positions of the sites sampled.

Environmental Distribution and Cellular Morphology of the Rappemonads. To investigate distributions in the environment, a rappemonad-specific qPCR assay was developed (SI Materials and Methods). Sequencing of environmental qPCR products and tests against nontarget and plasmid controls verified rappemonad specificity (SI Materials and Methods). The unique lineage was detected in 23 of 48 marine euphotic-zone samples ranging from 15 ± 4–4,318 ± 38 16S rRNA gene copies mL⁻¹ (Table S4). Sixteen samples in which cells were not detected showed high inhibition, an issue frequently seen with environmental DNA extractions and generally thought to be caused by the presence of unidentified inhibitory substances (23). This required dilution of these samples to levels at which detection limits were poor (in one case, minimum detection was 647 copies mL⁻¹, although detection was typically much better than this; Table S4); these data were not included in additional statistical analyses comparing abundance with environmental parameters. High 16S rRNA gene copies mL⁻¹ were detected in what appeared to be a late-winter bloom in sub-surface waters at the Bermuda Atlantic Time-series site (BATS) (Fig. 3). The water column at this time showed several fluorescence maxima, and rappemonads were concentrated at the shallowest of these, indicating that deeper maxima were composed of other taxa. Very few or no rappemonads were detected in stratified summertime conditions, when there was a pronounced deep chlorophyll maximum (Fig. 3). In addition, 11 of 12 samples from a North Pacific anticyclonic eddy, in which colder and more nutrient-rich waters, akin to the higher nutrient availability in late-winter BATS samples, were brought to the surface, resulting in a shallower mixed layer, had measureable gene copy numbers (averaging 186 ± 78 gene copies mL⁻¹). Other samples from the 500-mile transect in the North Pacific had fewer (Table S4).

Although depth, temperature, salinity, phosphate, chlorophyll a, and nitrate plus nitrite were measured, no statistically significant differences were identified between samples where rappemonads were detected and those samples where none were detected (t tests, Mann–Whitney rank sum), considering only

<table>
<thead>
<tr>
<th>Table S1</th>
<th>Palmaria palmata (Z18289)</th>
<th>Porphryya yeozenensis (AP006715)</th>
<th>Bangia atropurpurea (AF545616)</th>
<th>Bangia fusco purpurea (AF170716)</th>
</tr>
</thead>
<tbody>
<tr>
<td>99% Cluster 0001 (e.g. OM270)</td>
<td>99% Cluster 0002 (e.g. NP67-155D3B0A5_6Oct07_86m)</td>
<td>99% Cluster 0006 (e.g. NP67-155D3B026_6Oct07_86m)</td>
<td>99% Cluster 0018 (e.g. FS01F37_1Aug05_65m)</td>
<td></td>
</tr>
<tr>
<td>Palmaria palmata</td>
<td>Porphryya yeozenensis</td>
<td>Bangia atropurpurea</td>
<td>Bangia fusco purpurea</td>
<td></td>
</tr>
</tbody>
</table>

**Fig. 1.** Environmental diversity and sample locations. Maximum likelihood phylogenetic tree of environmental 16S rDNA sequences obtained herein (bold) as well as OM270. (Inset) Map shows the approximate positions of the sites sampled. Bootstrap support values (≥ 95%) are from RaxML and Log-Det distance analyses, respectively. (Inset) Map shows the approximate positions of the sites sampled.
those in which nutrient concentrations were above detection limits. The upper range of phosphate concentrations in rappemonad-positive samples was 0.69 μM, lower than for all samples (1.19 μM). In addition, chlorophyll a ranged from 0.07–0.69 μg L⁻¹ for rappemonad-containing samples and from 0.03–2.71 μg L⁻¹ for all samples. Average temperatures of the samples investigated and those that contained rappemonads were identical (17 ± 4 °C). Rappemonad sequences were detected in waters ranging from 11 °C to 24 °C, with sequences also recovered from 26 °C waters (Table S2), although this sample was not screened by qPCR because the DNA was not extracted in a quantitative manner. The temperature range of rappemonad-containing samples again indicates that this lineage may have a broad ecophylogical range.

We also characterized rappemonads morphologically. Oligonucleotide probes targeting two different 16S rRNA regions of distinct rappemonad subgroups within clade I (Fig. 1) were designed for use with tyramide signal amplification FISH and verified for specificity on a series of nontarget controls. The probes were applied to two samples for which FISH filters were available and notable gene copies mL⁻¹ were detected by qPCR. Rappemonads (n = 88) measured 5.7 ± 1.0 (SD) μm in width (shortest dimension) and 6.6 ± 1.2 (SD) μm in length (longest dimension). Each cell appeared to contain two, three, or four plastids (Fig. 4 and Fig. S3), with four being the most common (46 (52%) of 88 cells); it is conceivable that instances of three or four plastids when only one is present, and cell orientation can bias imaging. By comparison, the plastids of haptophytes mostly occur singly or in pairs and are often bilobed (2). The microcopy analyses also revealed a faint reddish fluorescence colocalized with the hybridized plastid compartments using a DAPI filter set (excitation, G365; emission, LP420), presumably derived from residual chlorophyll pigments.

The average cell biovolume of rappemonads was 112 μm³, and the average carbon content was 27 pg of carbon cell⁻¹, based on cellular dimensions and an established carbon conversion factor (24). This is a significantly greater cellular carbon content than in the picophytoplankton (<2–3 μm in diameter) that dominate such regions, for example, small haptophytes that range from 1–3 pg of carbon cell⁻¹ (6). Large cell size may also be responsible for the rarity of reported rappemonad sequences. The majority of environmental 16S (and 18S) rDNA clone libraries, especially those using primer sets targeting plastid-16S rRNA genes, are constructed from water prefiltered through 3-μm pore-sized filters (25, 26), which would select against these cells. In addition, a high abundance of heterotrophic bacteria (10⁴–10⁶ mL⁻¹) could effectively swamp 16S rDNA libraries constructed using universal primers, such that few plastid-derived sequences are attained.
What Are Rappemonads? We explored the possibility that rappemonad plastid 16S rDNA sequences are derived from other lineages represented in 18S rDNA phylogenies. One recently reported uncultured group is the biliphytes (or picobiliphytes). This putatively plastid-bearing eukaryotic lineage was previously detected in marine 18S rDNA clone libraries (27), and subsequent phylogenetic analyses indicated that they represented an uncultured lineage possibly related to cryptophytes, although no analysis provided bootstrap support for this relationship in excess of 50% (28, 29). Several lines of evidence suggest that rappemonads do not correspond to a putative plastid associated with biliphytes. First, the latter appear to be smaller than rappemonads, with two biliphyte clades being 3.5 ± 0.9 × 3.0 ± 0.9 and 4.1 ± 1.0 × 3.5 ± 0.8, respectively (29). Biliphytes detected here in the North Pacific, using FISH probes (28, 29), were similar in size to those reported by Cuvelier et al. (29). The highly punctate phycobilin-like (orange) fluorescence reported previously (28, 29) was not seen colocalized with the hybridized (North Pacific) cells, however, and biliphytes were not present at levels significantly above background counts for negative controls. Given that punctate orange fluorescence is detected in some (28, 29) but not all instances, biliphytes are likely not obligate photoautotrophs but rather facultative mixotrophs or phagotrophs, whereby transient detection of orange fluorescence could represent ingested prey items (e.g., the cyanobacterium Synechococcus).

We also applied a two-step group-specific nested 18S rDNA PCR protocol targeting the majority of known biliphyte diversity (SI Materials and Methods) but did not recover any biliphyte sequences from the same freshwater samples for which we had previously recovered rappemonad plastid 16S rDNA sequences. Although reaction-specific PCR biases cannot be completely ruled out, this is consistent with the idea that the biliphyte 18S rDNA and rappemonad plastid 16S rDNA do not share the same environmental distribution, and thus correspond to two separate groups. In addition, although rappemonads and biliphytes were recovered in marine 16S and 18S rDNA clone libraries, respectively, from the same depth and site (Table S2, station 67-155) for which 768 clones were sequenced per size fraction and primer pair, rappemonads were only found in the 3- to 20-μm size fraction, whereas biliphyte 18S rDNA sequences were in the 0.8- to 3-μm size fraction. Finally, we used a combination of nested forward primers and general reverse primers to amplify most of the biliphyte nuclear 18S-ITS1-5.8S-ITS2-28S rRNA gene cluster. A multigene phylogeny was then constructed, and the biliphyte branching position within eukaryotes was retested, using more phylogenetic information than previously published (28, 29). Like previous studies, the resulting phylogeny lacked bootstrap support above 50% for placement of biliphytes with respect to known eukaryotic groups. Furthermore, the phylogeny demonstrated that biliphytes are separated from cryptophyte and haptophyte algae by multiple branches resolved with weak bootstrap support, although the maximum likelihood bootstrap analyses demonstrated that glaucophytes formed a moderately supported clade (71%) with cryptophytes and katablepharids, to the exclusion of biliphytes (Fig. S4). This suggests a different branching relationship than that from 18S rDNA analyses alone (28, 29).

Alternative phylogenetic topology tests of the nuclear biliphyte and plastid rappemonad alignments were then used to investigate whether we could reject the hypothesis that the two groups represented equivalent branching positions on the plastid and nuclear rDNA trees. These data support the placement of rappemonad plastid sequences with the wider haptophyte/cryptophyte plastid radiation, whereas placement of the biliphyte lineage remains ambiguous (Fig. S2). The data suggest that the rappemonad plastid and biliphyte nucleus have incongruent ancestries, although this may result from secondary or tertiary endosymbioses or from potentially methodological artifacts. Taken in sum, multiple lines of indirect evidence do not support rappemonads and biliphytes being one and the same cellular entity. One possibility is that rappemonads harbor a plastid derived from a common ancestor it shared with haptophytes, and thus represent a lineage that diverged before the diversification of Pavlovophyceae and Cocolithophyceae (or Prymnesiophyceae), the only two classes of haptophytes known to date. Detection of recently obtained environmental 18S rDNA sequences showing affinity to the host component of haptophytes is consistent with this hypothesis (6, 30, 31).

Conclusion

We have reported the discovery of a putatively photosynthetic diverse plastid-bearing microbial lineage in marine and freshwater environments. Together with definitive proof that rappemonads contain plastids and perform photosynthesis, elucidation of their evolution and functional ecology will help to address how these organisms thrive. For example, should the presence of rappemonads in both marine and freshwater systems indicate relatively flexible halotolerance, this could be advantageous under conditions in which oceans are freshening as a result of ice melt, such as the Arctic (32). Microbes within the rare biosphere are thought to be essential for system stability and important under climate change scenarios (33). Furthermore, future initiatives to identify the specific eukaryotic host lineage corresponding to these plastid 16S rDNA sequences and its specific relationship to known haptophytes will enhance understanding of the origin, spread, and diversification of red algal-derived secondary plastids.

The discovery of unique plastid-bearing lineages, such as the rappemonads described herein, demonstrates that our current understanding of aquatic microbial community structure is far from complete. This has important repercussions for the current inability to model global biogeochemistry under perturbation scenarios, which relies not only on knowledge of “who is there” but on their resilience to change and capacity for both acclimation and adaptation.

Materials and Methods

Detailed descriptions of environmental sampling procedures and locations, DNA extraction protocols, clone library construction, and DNA sequencing are provided in SI Materials and Methods. Identification of chimeric sequences, molecular phylogenetic analyses and alternative topology tests, qPCR procedures and controls, and tyramide signal amplification FISH experiments are described in SI Materials and Methods. Tables S5–S8 present a list of PCR primers for biliphyte 18S rDNA gene amplification (Table S5), primers for biliphyte nucleus-encoded rRNA gene cluster amplification (Table S6), 99% sequence clusters for the phylogeny shown in Fig. 1 (Table S7), and oligonucleotide probes for TSA-FISH experiments (Table S8).
ACKNOWLEDGMENTS. We thank the captains and crews of the research vessels Oceanus, Waton Smith, and Western Flyer as well as the participants, particularly M. Cuvelier, A. Engman, and E. Demir. Some DNA samples were kindly provided by R. Paerl, J. Zehr, S. Giovannoni, and C. Carlson. Rebecca salina materials were provided by R. Kamikawa, T. Matsumoto, and T. Nakayama. We thank G. Weinstock, the Washington University St. Louis Genome sequencing center, J. Vassar, and R. Gausling for clone library assistance. N. Onodera assisted with sequencing haplotype DNA operons. J.W.H. was supported by a University of Exeter Studentship. M.D.M.J. was supported by National Environment Research Council (London) Grant NE/F011709/1, and T.A. R. thanks the Leverhulme Trust for fellowship support and the Natural Environment Research Council (London) and Biotechnology and Biological Sciences Research Council for funding. At-sea research, qPCR development, and some library construction and sequencing as well as A.Z.W., H.M.W., and S.S. were supported by National Science Foundation Grant OCE-0836721, Moore Foundation Grant MM-I-1668, and the David and Lucille Packard Foundation. E.K. and the Archibald Laboratory were supported by Dalhousie University’s Centre for Comparative Genomics and Evolutionary Bioinformatics and the Tula Foundation. A.Z.W. and J.M.A. acknowledge support from the Canadian Institute for Advanced Research’s Integrated Microbial Biodiversity Program.