# Global analysis of plastid diversity reveals apicomplexan-related lineages in coral reefs

Jan Janouškovec[1],
Aleš Horák[1,2], Katie L. Barott[3],
Forest L. Rohwer[3],
and Patrick J. Keeling[1,*]

The presence of relict non-photosynthetic plastids in obligate intracellular apicomplexan parasites (e.g. *Plasmodium*) has proved puzzling in many ways, but the recent discovery of their photosynthetic relative, *Chromera velia*, has begun to shed much-needed light on the origin and evolution of these plastids [1,2]. The intense interest that this single species has generated demonstrates how surprisingly little we know about photosynthetic relatives of apicomplexans as a whole. Here, we investigate global plastid diversity and distribution by comprehensively searching existing prokaryotic sequence surveys for eukaryotic plastids. From more than 1.6 million bacterial sequences, we identified 9,799 plastid-derived sequences, most of which were previously mis-labeled as 'novel bacteria' sequences. 98.8% of these plastid-derived sequences could be assigned to well-defined algal lineages, most often green algae, diatoms, and haptophytes. The exceptions were 121 sequences, all of which were related to apicomplexan parasites, and nearly all of which were derived from coral reef environments. Close relatives of *C. velia* were rare, but two other clusters were more common and globally distributed, one of which was tightly associated with corals. Overall, all of the major new lineages of algae we discovered were related to apicomplexans, suggesting that apicomplexans represent a large pool of unexplored algal diversity.

Comparing the recently-completed plastid genomes of two deep-branching photosynthetic relatives of apicomplexans [2] to public databases revealed that these lineages were present in microbial surveys, but were mistakenly identified as novel bacteria [3]. To examine simultaneously the extent to which plastids infiltrate bacterial surveys and the diversity and ecological distribution of plastid-bearing apicomplexan relatives, we developed a protocol for characterizing plastid sequences from environmental data (Supplemental experimental procedures; see Supplemental information). Over 1.6 million environmental sequences were pre-filtered, resulting in 42,989 candidate plastid genes, which were individually aligned to representative plastids and cyanobacteria. Phylogenies showed 9,799 sequences unambiguously originated from plastids, with 98.8% branching within well-defined algal lineages (predominantly within green algae, diatoms, and haptophytes; Figure 1A). Most plastid sequences were mis-identified as novel bacterial lineages, but some 'novel plastid lineages' were also mis-identified; for example, the prasinophyte clade 16S-IX [4] branched with stramenopiles, and not with green algae (data not shown). Many other sequences added to the diversity of existing uncultured clades, for example the prasinophyte clade 16S-VIII, deep-branching haptophytes, and rappemonads [5] and a novel lineage, S25_1200, which branched within haptophytes in our analysis (Figure 1A). Overall, sequences derived from numerous independent studies using different amplification primers successfully captured
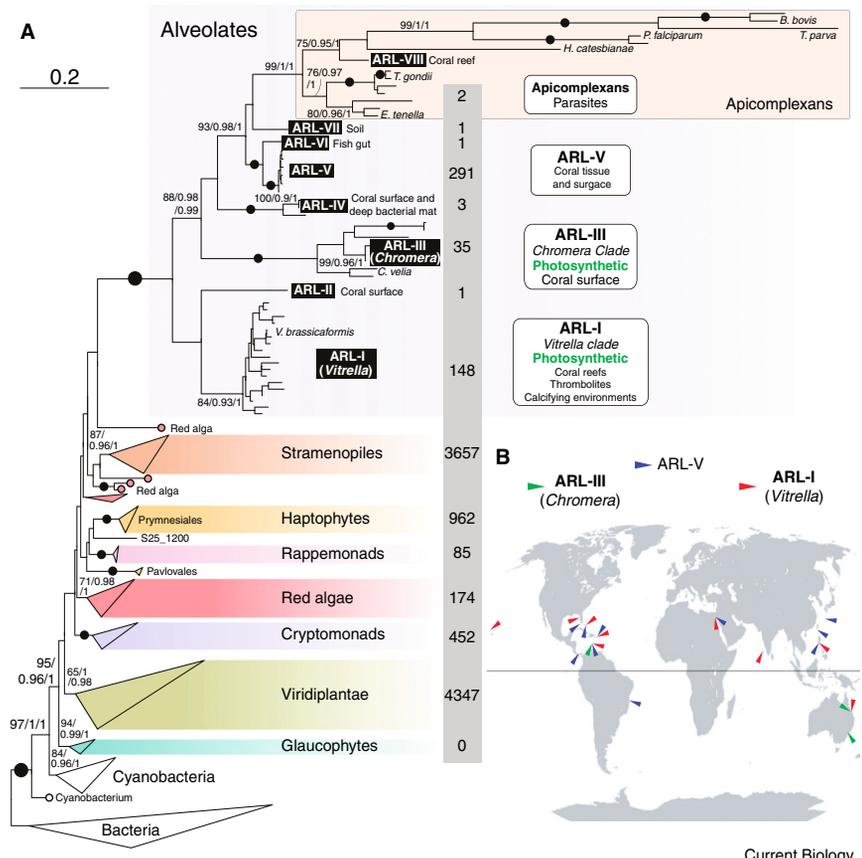


Figure 1. Phylogeny of novel ARL plastids and their environmental distribution.
(A) Rooted phylogeny of plastid 16S rRNA showing the positions of all new apicomplexan-related lineages (ARL, white type on black). For each plastid lineage, the number of environmental sequences detected is shown in the grey box (for alveolates this includes all sequences identified, for other groups only sequences detected in the automated search). Rappemonads and unidentified clone S25_1200 fell within the haptophytes in these analyses, and red algae were paraphyletic. The tree was constructed with RAxML and numbers at nodes correspond to support from rapid bootstrap/PhyML SH-like aLRT/MrBayes posterior probability when greater than 50/0.90/0.95 (supports inside ARL-I and ARL-V clades were omitted for clarity). Black dots indicate complete support. (B) Global distribution of ARL-I (*Vitrella*), ARL-III (*Chromera*) and ARL-V from environmental 16S and 18S rRNA sequences.

representatives of all known major plastid lineages, except glaucophytes, which are not abundant even in the best-sampled environments, and dinoflagellates, which are notoriously divergent in sequence [6] and do not amplify with universal primers. We therefore conclude that the systematic exclusion of new algal lineages is possible, but that this exclusion is unlikely to be widespread, and the recovered environmental sequences generally represent existing algal diversity.

The only exceptions that were not readily identifiable were 121 plastid sequences that did not fall within a major algal lineage, and all of these sequences branched as sisters to apicomplexans. Inclusion of shorter sequences that were not retrieved in automated searches resulted in 483 sequences comprising 8 distinct apicomplexan-related lineages (ARLs; Figure 1A), six of which were composed entirely of unidentified sequences. The phylogenetic relationship between ARLs and apicomplexans was confirmed by increasing plastid and bacterial sampling, excluding short and quickly evolving sequences, and including divergent dinoflagellate plastid sequences, including *Symbiodinium* (Figure S1), none of which changed the overall topology of Figure 1A, confirming that none of the ARLs represent the well-known zooxanthellae coral symbionts. Only three ARLs were phylogenetically diverse, and interestingly, two of these lineages include the only known photosynthetic plastids: *Chromera* [1] falls in ARL-III (35 sequences) and the recently described *Vitrella* [7] falls in ARL-I (148 sequences). However, the most abundant clade of all was ARL-V, which is composed entirely of unidentified sequences that were, like *Vitrella* plastids, previously misidentified as novel bacteria related to cyanobacteria [3,8].

The environmental distribution of the ARL sequences showed a strong association with coral reef environments, predominantly off-shore tropical and warm subtropical waters (Figure 1B; Table S1). The large ARL-V clade was strictly associated with coral reefs and its members were identified from tissue and surfaces of at least 16 coral species, but whether this clade represents a photosynthetic

symbiont or parasite remains unknown. It is possible that ARL-V corresponds to the plastids of Genotype N, an apicomplexan parasite of coral [9], but the phylogeny is more consistent with ARL-VIII corresponding to Genotype N. Unlike ARL-V, *Vitrella*-related ARL-I sequences were not strictly associated with coral reefs, but were also found in calcifying thrombolite mats [10]. Searching eukaryotic environmental data for sequences related to *Vitrella* nuclear rRNA confirmed its association with coral, reef sediment, and thrombolites, but also extended its range to a stromatolite mat and a coastal calcium carbonate sink — environments united by the dominance of calcium carbonate precipitates (Table S1 in Supplemental Information).

Overall, these analyses lead to a number of conclusions. On a practical level, plastid sequences significantly infiltrate environmental sequence surveys of bacteria, and recognizing this should prevent further mis-identification of eukaryotic organelles as novel bacterial phyla. More broadly, the major hot-spots of unknown algal diversity identified here are under-explored haptophytes, prasinophytes, and most surprisingly, several apicomplexan-related lineages. The multiple new apicomplexan-related plastid lineages identified here predominantly associate with coral reef environments. The two known photosynthetic groups, represented by *Chromera* and *Vitrella*, are both more diverse and abundant than previously thought, but the often-overlooked *Vitrella* emerges as the more abundant, widespread, and diverse. The newly recognized ARL-V clade is the most abundant of all apicomplexan-related plastid clades — its intermediate position between parasitic apicomplexans and the lineages known to include free-living photosynthetic species, together with its tight association with corals, focuses a great deal of interest on identifying these organisms in nature, and determining whether they are intracellular endosymbionts or parasites. Indeed, these organisms, more so than the cultured phototrophs, may be critical for elucidating the origin

and evolution of parasitism in apicomplexans.

**References**
1. Moore, R.B., Oborník, M., Janouskovec, J., Chrudimský, T., Vancová, M., Green, D.H., Wright, S.W., Davies, N.W., Bolch, C.J.S., Heimann, K., *et al*. (2008). A photosynthetic alveolate closely related to apicomplexan parasites. Nature *451*, 959–963.
2. Janouskovec, J., Horák, A., Oborník, M., Lukes, J., and Keeling, P.J. (2010). A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. Proc. Natl. Acad. Sci. USA *107*, 10949–10954.
3. Sørensen, K., Glazer, B., Hannides, A., and Gaidos, E. (2007). Spatial structure of the microbial community in sandy carbonate sediment. Mar. Ecol. Prog. Series *346*, 61–74.
4. Shi, X.L., Lepère, C., Scanlan, D.J., and Vaulot, D. (2011). Plastid 16S rRNA gene diversity among eukaryotic picophytoplankton sorted by flow cytometry from the South Pacific Ocean. PLoS ONE *6*, e18979.
5. Kim, E., Harrison, J.W., Sudek, S., Jones, M.D.M., Wilcox, H.M., Richards, T.A, Worden, A.Z., and Archibald, J.M. (2011). Newly identified and diverse plastid-bearing branch on the eukaryotic tree of life. Proc. Natl. Acad. Sci. USA *108*, 1496–1500.
6. Zhang, Z., Green, B.R., and Cavalier-Smith, T. (2000). Phylogeny of ultra-rapidly evolving dinoflagellate chloroplast genes: a possible common origin for sporozoan and dinoflagellate plastids. J. Mol. Evo. *51*, 26–40.
7. Oborník, M., Modrý, D., Lukeš, M., Cernotíková-Stříbrná, E., Cihlář, J., Tesařová, M., Kotabová, E., Vancová, M., Prášil, O., and Lukeš, J. (2012). Morphology, ultrastructure and life cycle of *Vitrella brassicaformis* n. sp., n. gen., a novel chromerid from the Great Barrier Reef. Protist *163*, 306–323.
8. Klaus, J.S., Janse, I., Heikoop, J.M., Sanford, R. a, and Fouke, B.W. (2007). Coral microbial communities, zooxanthellae and mucus along gradients of seawater depth and coastal pollution. Environ. Micro. *9*, 1291–1305.
9. Toller, W., Rowan, R., and Knowlton, N. (2002). Genetic evidence for a protozoan (phylum Apicomplexa) associated with corals of the Montastraea annularis species complex. Coral Reefs *21*, 143–146.
10. Myshrall, K.L., Mobberley, J.M., Green, S.J., Visscher, P.T., Havemann, S.A, Reid, R.P., and Foster, J.S. (2010). Biogeochemical cycling and microbial diversity in the thrombolitic microbialites of Highborne Cay, Bahamas. Geobiology *8*, 337–354.

[1]Canadian Institute for Advanced Research, Department of Botany, University of British Columbia, Vancouver, BC, V6T1Z4 Canada. [2]Institute of Parasitology, Biology Centre of the Academy of Sciences of the Czech Republic and Department of Molecular Biology, Faculty of Sciences, University of South Bohemia, Branišovská 31, 370 05 České Budejovice, Czech Republic. [3]Biology Department, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182, USA.
*E-mail: pkeeling@mail.ubc.ca