LETTER

https://doi.org/10.1038/s41586-019-1072-z

A widespread coral-infecting apicomplexan with chlorophyll biosynthesis genes

Waldan K. Kwong[1]*, Javier del Campo[1], Varsha Mathur[1], Mark J. A. Vermeij[2,3] & Patrick J. Keeling[1]

Apicomplexa is a group of obligate intracellular parasites that includes the causative agents of human diseases such as malaria and toxoplasmosis. Apicomplexans evolved from free-living phototrophic ancestors, but how this transition to parasitism occurred remains unknown. One potential clue lies in coral reefs, of which environmental DNA surveys have uncovered several lineages of uncharacterized basally branching apicomplexans[1,2]. Reef-building corals have a well-studied symbiotic relationship with photosynthetic Symbiodiniaceae dinoflagellates (for example, *Symbiodinium*[3]), but the identification of other key microbial symbionts of corals has proven to be challenging[4,5]. Here we use community surveys, genomics and microscopy analyses to identify an apicomplexan lineage—which we informally name 'corallicolids'—that was found at a high prevalence (over 80% of samples, 70% of genera) across all major groups of corals. Corallicolids were the second most abundant coral-associated microeukaryotes after the Symbiodiniaceae, and are therefore core members of the coral microbiome. In situ fluorescence and electron microscopy confirmed that corallicolids live intracellularly within the tissues of the coral gastric cavity, and that they possess apicomplexan ultrastructural features. We sequenced the genome of the corallicolid plastid, which lacked all genes for photosystem proteins; this indicates that corallicolids probably contain a non-photosynthetic plastid (an apicoplast[6]). However, the corallicolid plastid differs from all other known apicoplasts because it retains the four ancestral genes that are involved in chlorophyll biosynthesis. Corallicolids thus share characteristics with both their parasitic and their free-living relatives, which suggests that they are evolutionary intermediates and implies the existence of a unique biochemistry during the transition from phototrophy to parasitism.

Apicomplexan parasites rely on highly specialized systems to infect animal cells, live within those cells and evade host defences. Recently, it has come to light that these parasites evolved from phototrophic ancestors. Most apicomplexans have been found to retain relict plastids[6], and two photosynthetic 'chromerids' (*Chromera velia* and *Vitrella brassicaformis*) isolated from coral reef environments have been found to be the closest free-living relatives to the parasitic Apicomplexa[7–9]. The finding that the photosynthetic relatives of apicomplexans are somehow linked to coral reefs has prompted a major re-evaluation of the ecological conditions and symbiotic associations that drove the evolution of parasitism in this clade[2,10–12]. Corals (class Anthozoa) have not traditionally been considered a common host for apicomplexans: sporadic reports over the last 30 years include the morphological description of a single coccidian (*Gemmocystis cylindrus*) from histological sections of eight Caribbean coral species[13], and the detection of 18S rRNA gene sequences of apicomplexans (known as the 'type-N' apicomplexan) in Caribbean, Australian and Red Sea corals[14–16]. Plastid 16S rRNA gene surveys have also revealed that a number of uncharacterized apicomplexan-related lineages (notably, the 'ARL-V' lineage) are closely associated with reefs worldwide[1,2]. These lineages appear to occupy a phylogenetic position that is intermediate between the obligate parasitic Apicomplexa and the free-living chromerids, which makes them promising candidates for studying the transition between these different lifestyles.

To address evolutionary questions surrounding the transitional steps to parasitism, and to reconcile the currently incomparable data on the extent of apicomplexan diversity in corals, we sampled diverse anthozoan species from around the island of Curaçao in the southern Caribbean and surveyed the composition of their eukaryotic and prokaryotic microbial communities (Supplementary Table 1). From a total of 43 samples that represent 38 coral species, we recovered apicomplexan type-N 18S rRNA genes (putatively encoded by the nucleus) and ARL-V 16S rRNA genes (putatively encoded by the plastid) from 62% and 84% of samples, respectively (Fig. 1a, Supplementary Table 2). The type-N genes were only detected in corals that were also positive for ARL-V, which suggests that they come from the same organism; the high abundance of Symbiodiniaceae probably depressed our detection of the type-N apicomplexan. Excluding Symbiodiniaceae, type-N apicomplexans were the most common microbial eukaryote detected in corals, comprising 2.1% of the total sequence reads (56% of all non-Symbiodiniaceae reads). No other apicomplexan-related lineage was present, except for six reads of *Vitrella* 16S rRNA in a single sample. We also searched 31 publically available coral metagenomic and metatranscriptomic datasets that collectively amount to 15.8 Gb of assembled sequence (Supplementary Table 3). Sequences that correspond to rRNAs from type-N and ARL-V were present in 27 and 12 datasets, respectively (Fig. 1b); the discrepancy probably reflects the lower copy number of plastid rRNA genes. We further identified a suite of organelle-derived protein-coding genes (see below), and for each gene we found only a single apicomplexan sequence type to be predominant (Fig. 1b, Extended Data Fig. 1). All of these data are consistent with the presence of a single dominant apicomplexan lineage in corals, for which we propose the name corallicolids (meaning 'coral-dwellers', from Latin *corallium* combined with the suffix *–cola*, derived from the Latin *incola*). Our results indicate that this is the second most abundant microeukaryote that lives in association with coral, after the Symbiondiniaceae.

The high prevalence of corallicolids in wild corals is suggestive of a tight symbiosis (defined as two organisms that engage in long-term interactions), across a broad diversity of coral species. To test the host range of this symbiosis, we analysed 102 commercial aquarium samples that represented at least 61 species from across the major clades of Anthozoa. We detected corallicolid 18S rRNA genes in 53% of aquarium samples, including in soft-bodied octocorals, zoanthids, anemones and corallimorphs (Fig. 1c). Combined with existing data from wild corals (Supplementary Table 4), corallicolids were found in 1,271 of 1,546 samples (82% prevalence) and in 43 out of 62 host genera (70%), from all parts of the anthozoan phylogeny that we have examined thus far. Ecologically, the distribution of corallicolids is highly restricted: we searched large-scale 18S rRNA datasets from various terrestrial and marine ecosystems (1,014 samples, 837 million sequence reads), and found that type-N was almost

[1]Department of Botany, University of British Columbia, Vancouver, British Columbia, Canada. [2]Aquatic Microbiology, Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, The Netherlands. [3]CARMABI Foundation, Willemstad, Curaçao, The Netherlands. *e-mail: waldankwong@gmail.com

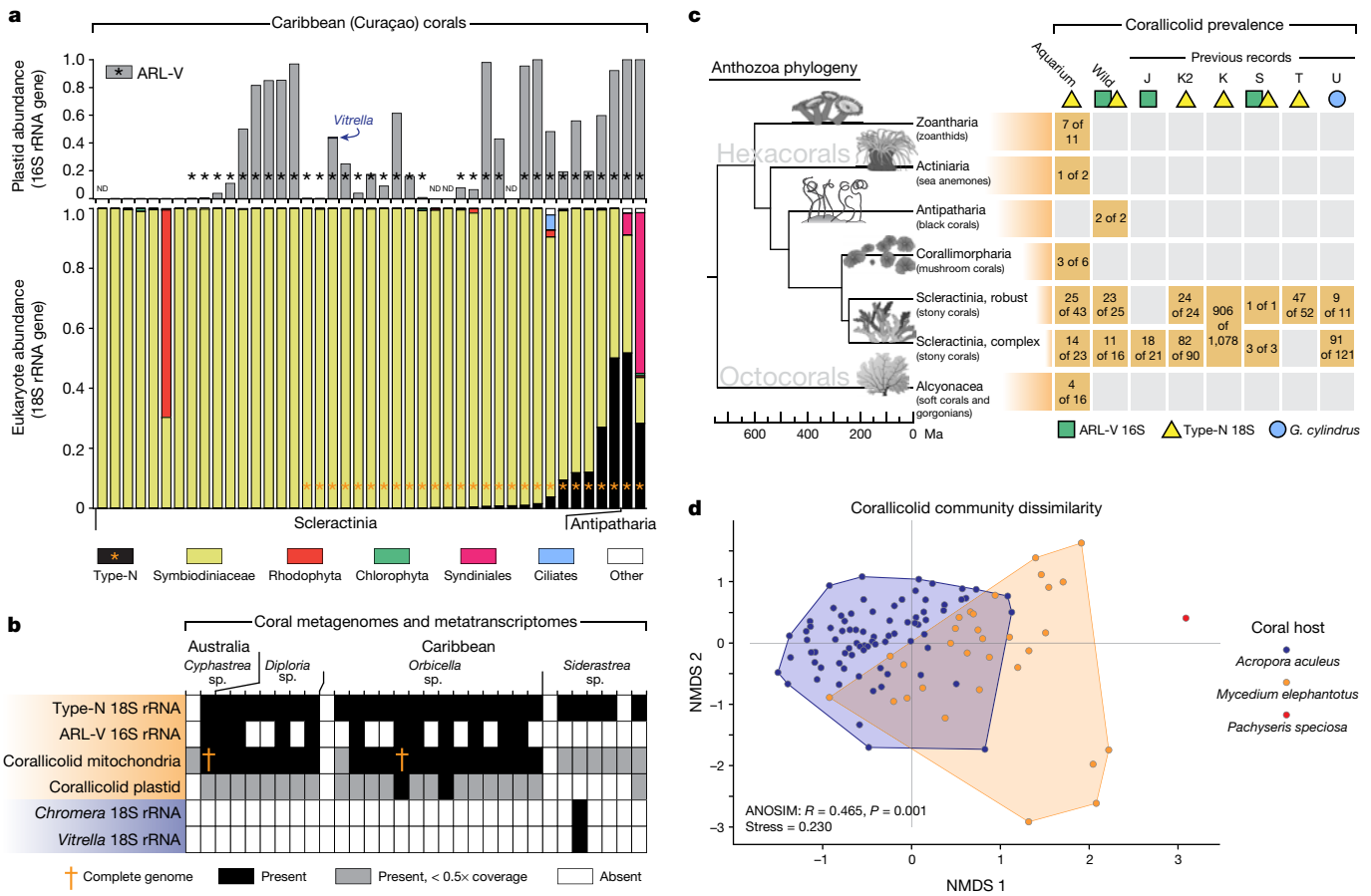


Fig. 1 | A single apicomplexan symbiont is present in diverse corals. **a**, Relative abundances of plastids (top) and microbial eukaryote communities (bottom), based on 16S and 18S rRNA gene sequencing. Each column represents data from a single coral sample. The presence of ARL-V and type-N is also denoted by asterisks (in addition to bars) to aid in visualization for samples in which they are present at low abundances. The 16S rRNA gene primers that we used excluded the detection of Symbiodiniaceae plastids. ND, not determined, owing to failed sequencing. **b**, Presence of Apicomplexa-derived sequences in public metagenomic and metatranscriptomic datasets. Host species indicated at the top. Shading indicates presence or absence, or coverage of contigs compared to the complete organellar genomes. **c**, Presence of corallicolids across the anthozoan phylogeny. Data include aquarium and wild-collected samples from this study, and wild samples from previous studies (labelled as J (ref. [11]), K (ref. [23]), K2 (ref. [29]), S (ref. [15]), T (ref. [14]) and U (ref. [13])) that used various methods to detect ARL-V, type-N and/or *G. cylindrus* (details in Supplementary Table 4). The coral phylogeny is from published data (see Methods). Ma, million years ago. **d**, Non-metric multidimensional scaling (NMDS) plot of ARL-V community diversity in three coral species, showing correlation with host identity (analysis of similarities (ANOSIM) using 999 permutations). Sequence data[17] clustered at 99.5% similarity.

exclusively associated with coral reef environments (Extended Data Fig. 2, Supplementary Table 5). These results agree with a previous survey of ARL-V based on 16S rRNA datasets[2]. The association of particular lineages of corallicolids with certain coral species (that is, host specificity) could provide further evidence for a close symbiosis that spans long evolutionary timescales. To test this, we analysed a 16S rRNA gene amplicon dataset of three scleractinian coral species that were sampled from the Great Barrier Reef and the Coral Sea[17]. There was strong support for host-specific communities of ARL-V (Fig. 1d), but no association with geographical location or sampling depth (Extended Data Fig. 3).

To further characterize corallicolid biology, we chose an aquacultured green mushroom coral as a model (*Rhodactis* sp.) (Fig. 2a). Corallicolid cells were first visualized by hybridizing coral tissues with fluorescent probes specific to type-N 18S rRNA and ARL-V 16S rRNA. An overlapping fluorescence signal from both probes was observed within the cnidoglandular lobes of the mesenterial filaments (Fig. 2b, c). This tissue region is dense with nematocysts and secretory cells: mesenterial filaments help to digest food within the gastric cavity, and can be expelled from the polyp body for defence[18]. The only formally described apicomplexan from coral, *G. cylindrus*, was also found in mesenterial filaments[13]. No genetic sequence data for *G. cylindrus* are available, but its similarities in localization, cell size (about 10 μm) and Coccidia-like morphology to the cells that we identified indicate that *G. cylindrus* is probably a corallicolid.

Transmission electron microscopy of infected tissue showed cells with classical apicomplexan features (for example, a conical cortex of microtubules and inner-membrane complex) that reside within a parasitophorous vacuole located inside host cells (Fig. 2d). Corallicolid cells were often closely clustered together, which is consistent with reproductive stages in other apicomplexans (for example, schizogony or oocyst development). The cells contained numerous large (up to 1.6 μm), darkly stained elliptical organelles that possessed striated internal features (Extended Data Fig. 4). These distinctive structures could be homologous to known apicomplexan compartments such as rhoptries or even plastids, but identification will require localization of functionally relevant marker proteins.

The most fundamental question about the relationship between corallicolids and their host—and, by extension, how they affect our views of apicomplexan origins—is whether they are photosynthetic or parasitic. To address this, we sequenced the genome of the corallicolid plastid, and assessed the phylogenetic position of corallicolids using plastid, mitochondrial and nuclear data (Fig. 3a). We first retrieved all possible corallicolid plastid and mitochondrial sequences from metagenomic datasets (Supplementary Table 3) by using homologues from *C. velia*, *V. brassicaformis*, and from

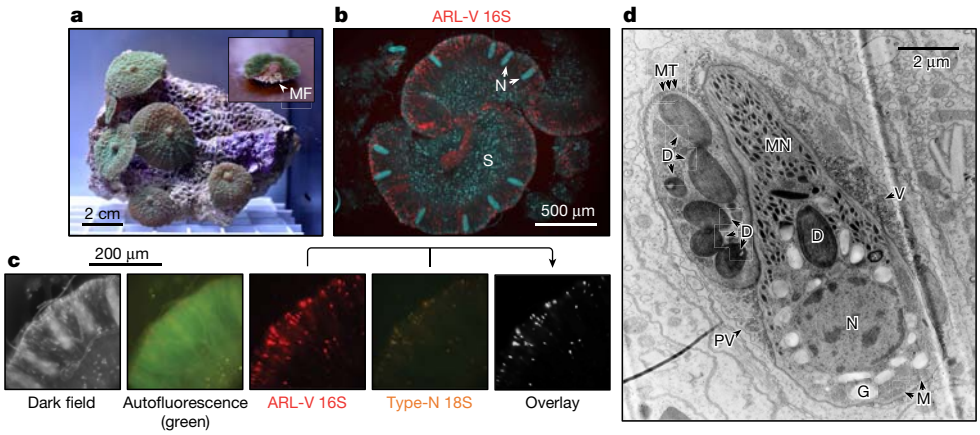


Fig. 2 | Corallicolids are located intracellularly within the mesenterial filaments, and possess cellular features of apicomplexans. a, The *Rhodactis* sp. coral from which the corallicolid wkC1 was imaged and sequenced. Inset shows cross-section, with mesenterial filaments (MF) lining the gastric cavity. **b**, Fluorescence in situ hybridization imaging localizes corallicolids (red, plastid rRNA) to cnidoglandular bands. N, nematocysts; S, Symbiodiniaceae cells. **c**, Co-localization of ARL-V (red) and type-N (orange) signals. Unlike Symbiodiniaceae, corallicolids do not exhibit autofluorescence. **d**, Transmission electron micrograph showing corallicolid ultrastructure. Two adjacent cells are visible, oriented perpendicularly. D, dark-staining organelles; G, polysaccharide granules; M, mitochondria; MN, micronemes; MT, microtubules; N, nucleus; PV, parasitophorous vacuole; V, extracellular vesicles. Imaging was conducted in triplicate; representative results are shown.

the parasitic apicomplexans *Toxoplasma gondii* and *Plasmodium falciparum* as search queries. Close matches were retrieved from 29 out of 31 datasets and manually inspected to confirm their apicomplexan origin. Two complete mitochondrial genomes were assembled, as well as fragments of plastid genomes (Fig. 1b). We then used a combination of metagenomic sequencing and primer walking to obtain a complete corallicolid plastid genome from the *Rhodactis* sp. host (sample number wkC1).

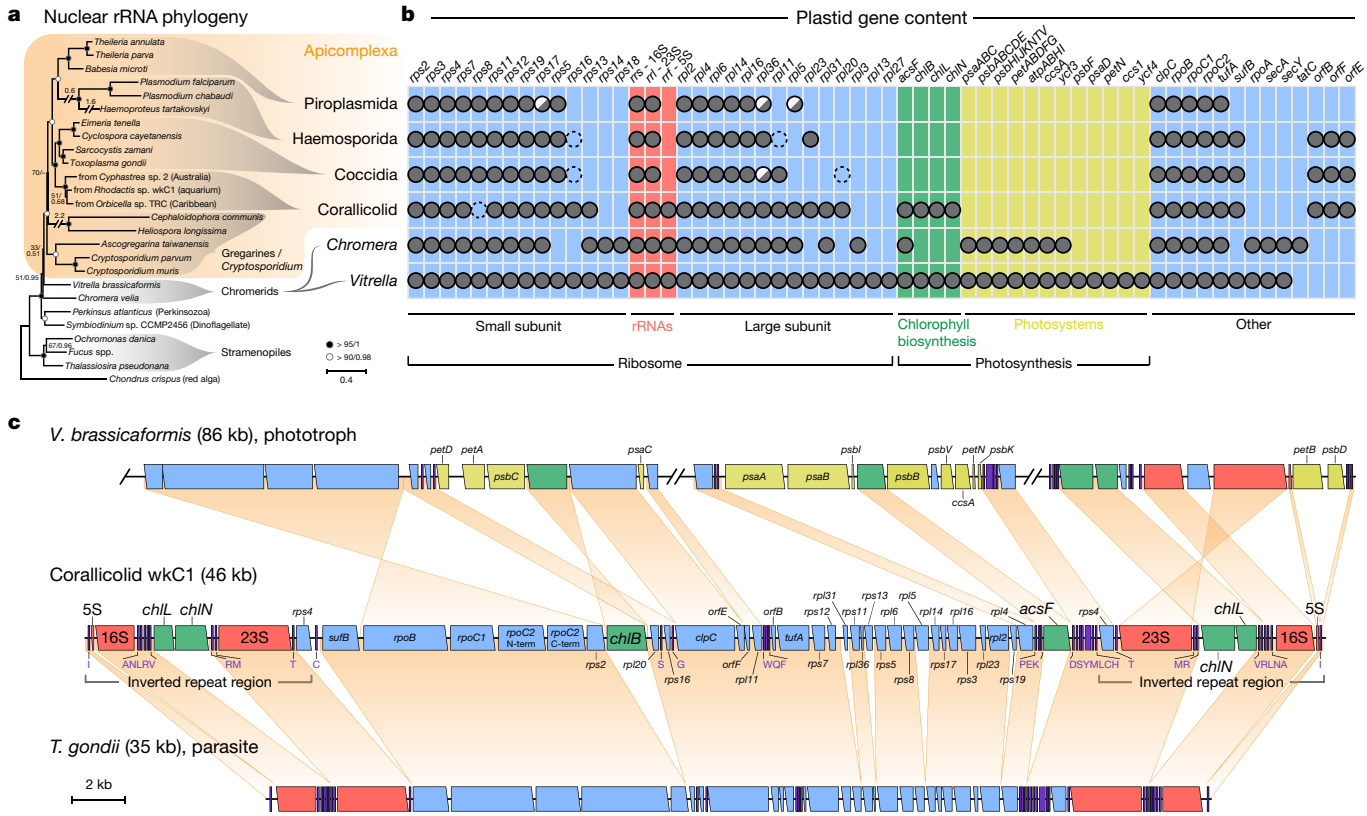


Fig. 3 | Corallicolids share traits with both the parasitic Apicomplexa and free-living phototrophs. a, Maximum likelihood phylogeny based on concatenation of nucleus-encoded 18S, 5.8S and 28S rRNA genes. Corallicolids branch within the parasitic Apicomplexa, as sister to the Coccidia. Values before and after the forward slash at nodes denote the maximum likelihood bootstrap values and Bayesian posterior probabilities, respectively. **b**, Plastid gene content of the Apicomplexa and free-living relatives. Presence is denoted by filled circles; variable presence between species within a clade is denoted by half-filled circles; dotted circles represent possible gene remnants (pseudogenes or highly divergent sequences). *Chromera* lacks plastid-encoded *chlB*, *chlL* and *chlN* genes, and instead uses nuclear-encoded genes for light-dependent protochlorophyllide reduction. For photosystem genes, *psaABC* indicates *psaA*, *psaB* and *psaC*; the equivalent abbreviated notation is also used for *psbABCDE*, *psbHIJKNTV*, *petABDFG* and *atpABHI*. **c**, The corallicolid plastid genome compared to those of *V. brassicaformis* (free-living and photosynthetic) and *T. gondii* (obligate parasite). Gene colours correspond to the gene categories in **b**. Purple bars denote tRNAs. Orange shading links regions of genomic synteny. Genomes are shown at the same scale; only select portions of the *V. brassicaformis* plastid genome are shown.

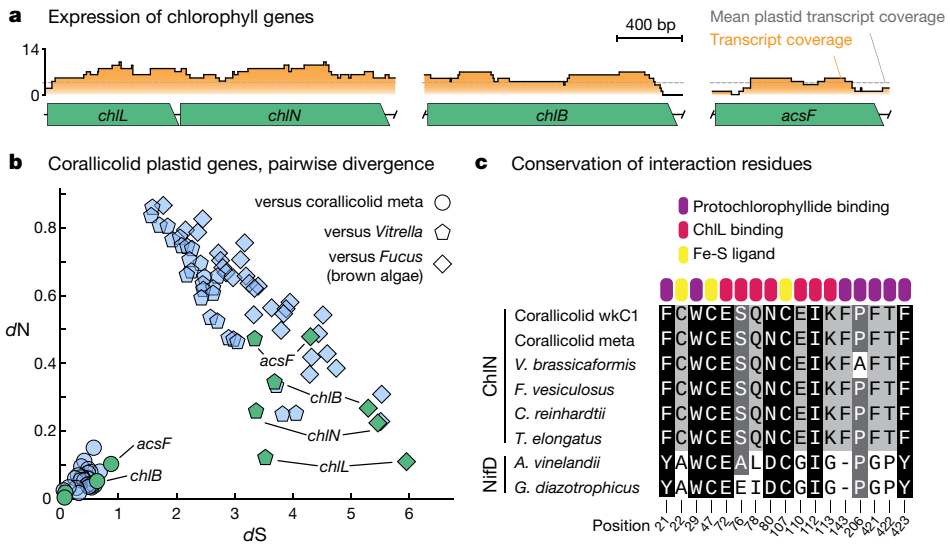


Fig. 4 | Corallicolid chlorophyll biosynthesis genes are probably functional. a, Transcript coverage from metatranscriptomes, showing that the expression levels of the four chlorophyll biosynthesis genes are similar to those of the rest of the plastid. **b**, Plot of *d*N and *d*S (*d*S, synonymous substitutions per synonymous site) for corallicolid wkC1 plastid genes compared to homologues. Variants of corallicolid plastid genes ('corallicolid meta') were obtained from metagenomes and metatranscriptomes. Chlorophyll biosynthesis genes (in green) have low *d*N compared to other plastid genes (in blue), which indicates that they are slow-evolving and under purifying selection. **c**, Protein sequence alignment of ChlN shows the conservation of key amino acid residues, which suggests functional conservation (full alignments in Extended Data Fig. 7). Sequences are from *V. brassicaformis* (chromerid), *Fucus vesiculosus* (brown alga), *Chlamydomonas reinhardtii* (green alga), *Thermosynechococcus elongatus* (cyanobacterium), and *Azotobacter vinelandii* and *Gluconacetobacter diazotrophicus* (bacteria). NifD, nitrogenase alpha chain, from a closely related protein family. Position numbers correspond to the *T. elongatus* sequence[30].

The genome of the corallicolid plastid shares a combination of similarities with the apicoplast and its photosynthetic relatives. At 46 kb, it is intermediate in size; it has lost a substantial number of genes, including those that encode photosystems (Fig. 3b). Photosystems are essential for photosynthesis and are still largely encoded by the plastid in chromerids[19] and other photosynthetic organisms. Their absence in the corallicolid plastid strongly suggests that corallicolids are non-photosynthetic. However, corallicolids also retain a number of other genes that have been lost in previously known apicoplasts. These include the gene for 5S rRNA and two genes that encode proteins that interact with it, *rpl5* and *rps13*[20]. Most notable was the presence of *chlL*, *chlN*, *chlB* and *acsF*, which are involved in chlorophyll biosynthesis (Extended Data Fig. 5). These are the only four genes in this pathway that the ancestral plastid would have encoded, and they grouped phylogenetically with homologues from *V. brassicaformis*, which indicates that they are derived from the photosynthetic apicomplexan ancestor and are not the result of horizontal acquisition (Extended Data Fig. 6). These findings provide a window into the evolutionary process that led to the reduction in size of the apicoplast genome, which probably occurred in a stepwise manner in which all photosystem genes were first lost from the common ancestor with chromerids, followed by the loss of chlorophyll biosynthesis genes in the parasitic apicomplexans (Fig. 3c).

The retention of *chlL*, *chlN*, *chlB* and *acsF* in the face of otherwise-severe gene loss indicates that these genes remain under strong selection. Not only were they transcribed at appreciable levels, they were also among the slowest evolving genes (that is, genes with a low rate of substitutions at non-silent (nonsynonymous) sites (*d*N)) in the genome of the corallicolid plastid and were conserved at key residues, which implies purifying selection (Fig. 4, Extended Data Fig. 7). The function that underlies this selection is less obvious: corallicolid cells were unpigmented and were not autofluorescent (at >550 nm, using 450–500-nm excitation) and—considering the absence of photosystem genes in the plastid genome—are probably not photosynthetic. Although it is conceivable that the photosystem genes have moved to the nucleus, such an arrangement would be unprecedented. Even in dinoflagellates, which contain the most reduced photosynthetic plastid genomes, some photosystem components remain encoded by the plastid[21]. By contrast, all other known non-photosynthetic organisms with tiny plastid genomes have lost their photosystems[22]. We did not detect any corallicolid photosystem homologues in our metagenomes (Supplementary Table 6), but a complete nuclear genome sequence will be needed to definitively rule out the presence of photosystems.

Chlorophyll itself has no natural biological function outside of photosynthesis, so if photosystems are indeed absent, corallicolids must have evolved a novel use for either chlorophyll or its closely related precursors or derivatives. However, these molecules generally function in light harvesting, which would be destructive to cellular integrity without the coupling of the resulting high-energy compounds to photosynthesis. Other possibilities are functions in light sensing, photo-quenching or the regulation of haem synthesis, but these too leave open the question of what the cell would do with the high-energy end products. Moreover, we detected corallicolids in sun coral (*Tubastrea* sp.) and black coral (order Antipatharia), both of which are considered to be non-photosynthetic corals, which further suggests that corallicolids deviate from classical modes of light harvesting.

Whatever the function of these genes may be, phylogenetic analyses suggest that corallicolids may not be the only apicomplexan lineage to retain them. Whole-plastid genome phylogenies placed corallicolids at the base of the Apicomplexa (Extended Data Fig. 6b) (consistent with 16S rRNA analyses[1,2]), which was parsimonious with respect to gene content (Fig. 3b). However, nuclear rRNA and protein phylogenies placed corallicolids deep within the parasitic apicomplexans, as sister to the Coccidia[14] (Fig. 3a, Extended Data Fig. 8). The mitochondrial genome phylogeny was consistent with this (Extended Data Fig. 6a), as was plastid gene synteny, amino acid identity, the use of UGA as a tryptophan codon in the plastid and single-gene phylogenies of several plastid genes (Extended Data Fig. 9). There is no straightforward explanation for this incongruence, but the dearth of data for deep-branching apicomplexan plastids makes it more likely that the plastid phylogeny is misleading, and that continued sampling will yield other lineages in which chlorophyll biosynthesis is retained. The validity of the nuclear phylogenies would also imply that plastid-encoded chlorophyll biosynthesis (and the concomitant cellular biochemistry) was separately and repeatedly lost in the lineages that led to Haemosporidia, Coccidia and *Cryptosporidium*, and that some similarities among Apicomplexa

parasites—including *Toxoplasma* and *Plasmodium*—may be the result of convergent evolution.

Most reef-dwelling corals are photosynthetic by virtue of symbiosis with the Symbiodiniaceae. Although the nature of the interaction between corallicolids and corals almost certainly differs from this, there is also no evidence that corallicolids are pathogenic. Corallicolids and the Symbiodiniaceae had non-overlapping localizations within host tissue (Fig. 2). Sampled corals were almost exclusively in good health (Supplementary Table 1), and the presence of corallicolids did not appear to be detrimental or to correlate with host disease. We speculate that if corallicolids do induce pathology, their negative effects are probably minor, strain-specific or arise opportunistically. Elucidation of the corallicolid life cycle may help to assess their effect on coral health and their role in reef ecosystems. *G. cylindrus* sporozoites and oocysts have previously been described in coral tissues[13], but the existence of life stages outside corals—perhaps in another host—remains a possibility. The planulae (larvae) of brooding coral species have previously been found[23] to contain type-N (which indicates vertical transmission of the symbiont), whereas the gametes of broadcast-spawning corals did not, which suggests horizontal or environmental acquisition in these species. The potential for mixed modes of transmission that are contingent on the traits of hosts has also been reported for Symbiodiniaceae[24,25].

Corals are found across temperate and tropical oceans, and are fundamental to the building of coral-reef ecosystems. In recent years, there have been alarming losses of healthy reefs worldwide, owing to stressors such as climate change and pollution. Changes in the microbiome are associated with coral stress and disease[26,27], and rising temperatures can result in coral bleaching due to the expulsion of Symbiodiniaceae[28]. Understanding the intricate symbiotic relationships between corals and their microorganisms is crucial in the effort to decipher the processes that lie behind reef degradation. The identity and nature of coral–microorganism associations, outside of those with the Symbiodiniaceae, remain poorly characterized[4,5]. Here we show that diverse anthozoans are colonized by members of a single apicomplexan lineage with unusual characteristics. Corallicolids are core coral symbionts, the discovery of which has implications for our understanding of coral biology and the evolution of parasitism.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-019-1072-z.

Received: 20 August 2018; Accepted: 6 March 2019;
Published online 3 April 2019.

1. Janouškovec, J., Horák, A., Barott, K. L., Rohwer, F. L. & Keeling, P. J. Global analysis of plastid diversity reveals apicomplexan-related lineages in coral reefs. *Curr. Biol.* **22**, R518–R519 (2012).
2. Mathur, V., del Campo, J., Kolisko, M. & Keeling, P. J. Global diversity and distribution of close relatives of apicomplexan parasites. *Environ. Microbiol.* **20**, 2824–2833 (2018).
3. LaJeunesse, T. C. et al. Systematic revision of Symbiodiniaceae highlights the antiquity and diversity of coral endosymbionts. *Curr. Biol.* **28**, 2570–2580 (2018).
4. Hernandez-Agreda, A., Gates, R. D. & Ainsworth, T. D. Defining the core microbiome in corals' microbial soup. *Trends Microbiol.* **25**, 125–140 (2017).
5. Ainsworth, T. D., Fordyce, A. J. & Camp, E. F. The other microeukaryotes of the coral reef microbiome. *Trends Microbiol.* **25**, 980–991 (2017).
6. Lim, L. & McFadden, G. I. The evolution, metabolism and functions of the apicoplast. *Phil. Trans. R. Soc. Lond. B* **365**, 749–763 (2010).
7. Moore, R. B. et al. A photosynthetic alveolate closely related to apicomplexan parasites. *Nature* **451**, 959–963 (2008).
8. Oborník, M. et al. Morphology, ultrastructure and life cycle of *Vitrella brassicaformis* n. sp., n. gen., a novel chromerid from the Great Barrier Reef. *Protist* **163**, 306–323 (2012).
9. Woo, Y. H. et al. Chromerid genomes reveal the evolutionary path from photosynthetic algae to obligate intracellular parasites. *eLife* **4**, e06974 (2015).
10. Cumbo, V. R. et al. *Chromera velia* is endosymbiotic in larvae of the reef corals *Acropora digitifera* and *A. tenuis*. *Protist* **164**, 237–244 (2013).
11. Janouškovec, J., Horák, A., Barott, K. L., Rohwer, F. L. & Keeling, P. J. Environmental distribution of coral-associated relatives of apicomplexan parasites. *ISME J.* **7**, 444–447 (2013).
12. Mohamed, A. R. et al. Deciphering the nature of the coral–*Chromera* association. *ISME J.* **12**, 776–790 (2018).
13. Upton, S. & Peters, E. A new and unusual species of coccidium (Apicomplexa: Agamococcidiorida) from Caribbean scleractinian corals. *J. Invertebr. Pathol.* **47**, 184–193 (1986).
14. Toller, W. W., Rowan, R. & Knowlton, N. Genetic evidence for a protozoan (phylum Apicomplexa) associated with corals of the *Montastraea annularis* species complex. *Coral Reefs* **21**, 143–146 (2002).
15. Šlapeta, J. & Linares, M. C. Combined amplicon pyrosequencing assays reveal presence of the apicomplexan "type-N" (cf. *Gemmocystis cylindrus*) and *Chromera velia* on the Great Barrier Reef, Australia. *PLoS ONE* **8**, e76095 (2013).
16. Clerissi, C. et al. Protists within corals: the hidden diversity. *Front. Microbiol.* **9**, 2043 (2018).
17. Hernandez-Agreda, A., Leggat, W., Bongaerts, P., Herrera, C. & Ainsworth, T. D. Rethinking the coral microbiome: simplicity exists within a diverse microbial biosphere. *Mbio* **9**, e00812–e00818 (2018).
18. Lang, J. Interspecific aggression by scleractinian corals: 2. Why the race is not only to the swift. *Bull. Mar. Sci.* **23**, 260–279 (1973).
19. Janouškovec, J., Horák, A., Oborník, M., Lukes, J. & Keeling, P. J. A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. *Proc. Natl Acad. Sci. USA* **107**, 10949–10954 (2010).
20. Dontsova, O. A. & Dinman, J. D. 5S rRNA: structure and function from head to toe. *Int. J. Biomed. Sci.* **1**, 1–7 (2005).
21. Barbrook, A. C., Voolstra, C. R. & Howe, C. J. The chloroplast genome of a *Symbiodinium* sp. clade C3 isolate. *Protist* **165**, 1–13 (2014).
22. Smith, D. R. Plastid genomes hit the big time. *New Phytol.* **219**, 491–495 (2018).
23. Kirk, N. L. et al. Tracking transmission of apicomplexan symbionts in diverse Caribbean corals. *PLoS ONE* **8**, e80618 (2013).
24. Baird, A. H., Guest, J. R. & Willis, B. L. Systematic and biogeographical patterns in the reproductive biology of scleractinian corals. *Annu. Rev. Ecol. Evol. Syst.* **40**, 551–571 (2009).
25. Quigley, K. M., Willis, B. L. & Bay, L. K. Heritability of the *Symbiodinium* community in vertically- and horizontally-transmitting broadcast spawning corals. *Sci. Rep.* **7**, 8219 (2017).
26. Meyer, J. L., Paul, V. J. & Teplitski, M. Community shifts in the surface microbiomes of the coral *Porites astreoides* with unusual lesions. *PLoS ONE* **9**, e100316 (2014).
27. Glasl, B., Herndl, G. J. & Frade, P. R. The microbiome of coral surface mucus has a key role in mediating holobiont health and survival upon disturbance. *ISME J.* **10**, 2280–2292 (2016).
28. Davy, S. K., Allemand, D. & Weis, V. M. Cell biology of cnidarian–dinoflagellate symbiosis. *Microbiol. Mol. Biol. Rev.* **76**, 229–261 (2012).
29. Kirk, N. L., Thornhill, D. J., Kemp, D. W., Fitt, W. K. & Santos, S. R. Ubiquitous associations and a peak fall prevalence between apicomplexan symbionts and reef corals in Florida and the Bahamas. *Coral Reefs* **32**, 847–858 (2013).
30. Bröcker, M. J. et al. Crystal structure of the nitrogenase-like dark operative protochlorophyllide oxidoreductase catalytic complex $(ChlN/ChlB)_2$. *J. Biol. Chem.* **285**, 27336–27345 (2010).

Acknowledgements We thank C. Zwimpfer and B. Ross for assistance with sample processing and electron microscopy. This work was funded by the Canadian Institutes for Health Research grant MOP-42517 (to P.J.K.), the Natural Sciences and Engineering Research Council of Canada Fellowship PDF-502457-2017 and a Killam Postdoctoral Research Fellowship (to W.K.K.) and the Marie Curie International Outgoing Fellowship FP7-PEOPLE-2012-IOF - 331450 CAARL and a Tula Foundation grant to the Centre for Microbial Biodiversity and Evolution (to J.d.C.).

Reviewer information *Nature* thanks Christopher Howe, Patrick Wincker and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions W.K.K., J.d.C. and P.J.K. designed the study. W.K.K., J.d.C., M.J.A.V. and P.J.K. obtained samples. J.d.C., V.M. and W.K.K. conducted microbial community analyses. W.K.K. performed all other analyses. W.K.K. and P.J.K. wrote the manuscript with input from all authors.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-019-1072-z.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1072-z.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to W.K.K.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Microbial community survey of wild corals. Corals were collected from several locations in Curaçao in April 2015, under the collecting permits of the Dutch Antillean Government (government reference: 2012/48584) provided to the CARMABI Foundation (CITES Institution code AN001) (Supplementary Table 1). Whole samples—including skeleton and tissue—were homogenized using a mortar and pestle, and DNA was extracted with the RNA PowerSoil Total RNA Isolation Kit coupled with the DNA Elution Accessory Kit (MO BIO Laboratories). DNA concentration was quantified on a Qubit 2.0 Fluorometer (Thermo Fisher Scientific). To avoid inclusion of host DNA, genes were amplified with primer sets designed to exclude metazoans.

Prokaryotic microbiome amplicon preparation and sequencing was performed by the Integrated Microbiome Resource facility at the Centre for Comparative Genomics and Evolutionary Bioinformatics at Dalhousie University. PCR amplification from template DNA was performed in duplicate using high-fidelity Phusion polymerase. A single round of PCR was done using 'fusion primers' (Illumina adaptors + indices + specific regions) targeting the V6–V8 region of the bacterial and archaeal 16S rRNA gene (primer set B969F + BA1406R (~440–450-bp fragment))[31] with multiplexing. PCR products were verified visually by running a high-throughput Invitrogen 96-well E-gel. The duplicate amplicons from the same samples were pooled in one plate, then cleaned up and normalized using the high-throughput Invitrogen SequalPrep 96-well Plate Kit. The samples were then pooled to make one library, which was quantified fluorometrically before sequencing on an Illumina MiSeq using a 300-bp paired-end read design.

Eukaryotic microbiome amplicons were prepared using PCR with high-fidelity Phusion polymerase (Thermo Fisher Scientific), using metazoan-excluding primers that target the V4 region of the 18S rRNA gene (UnonMetaF 5′-GTGCCAGCAGCCGCG-3′, UnonMetaR 5′-TTTAAGTTTCAGCC TTGCG-3′)[32,33]. PCR was performed using the following protocol: 30 s at 98 °C, followed by 35 cycles each consisting of 10 s at 98 °C, 30 s at 51.1 °C and 1 min at 72 °C, ending with 5 min at 72 °C. PCR products were visually inspected for successful amplification using gel electrophoresis with 1% agarose gels. PCR products were then cleaned using the QIAquick PCR Purification Kit (Qiagen) and quantified on a Qubit 2.0 Fluorometer. Amplicon sequencing was performed by the Integrated Microbiome Resource facility at the Centre for Comparative Genomics and Evolutionary Bioinformatics at Dalhousie University, as above, but using the eukaryote-specific primer set E572F + E1009R (~440-bp fragment)[31].

Amplicon reads were processed (dereplication, chimaera detection and singleton removal) using VSEARCH[34]. Operational taxonomic units (OTUs) were clustered at 97% similarity using VSEARCH and analysed using QIIME 1.9.1[35]. The taxonomic identity of each OTU was assigned on the basis of the SILVA 128.1 database[36], modified to include the small subunit rRNA of the coral-skeleton-boring algae *Ostreobium queketii*[37], using the assign_taxonomy function in QIIME. OTUs that were unassigned were inspected using BLAST against the GenBank nr database, and manually reassigned to the closest hit if possible. OTUs represented by fewer than four reads were removed, as were OTUs that were identified as metazoan 18S rRNA or mitochondria. Samples with fewer than 1,500 reads were excluded from the initial analysis. In total, 863,280 reads (average 20,554 per sample) were obtained in the eukaryotic 18S rRNA dataset after filtering. For the prokaryotic 16S rRNA dataset, a total of 254,611 reads (average 8,780 per sample) were obtained after filtering. For samples with fewer than 1,500 final reads, the pre-filtered OTU tables were inspected manually to determine relative abundances of ARL-V and type-N. Owing to the specificity of the prokaryotic primer set used, Symbiodiniaceae plastid 16S rRNA gene sequences were generally not amplified (Supplementary Table 2). Differences in detecting ARL-V and type-N from the same samples may be due to differential read depth between the eukaryotic and prokaryotic datasets.

Survey of corallicolid distribution and diversity. Sequences annotated as Alveolates were retrieved from three publicly available 18S rRNA datasets (VAMPS[38], BioMarKs[39] and Tara Oceans[40]), covering a wide range of environments from soils and freshwater to the sunlit ocean and sediments. Additionally, sequences from 18 other studies that focus on eukaryotic microbiomes of various marine hosts (including corals, sponges and eelgrass) were retrieved by BLAST search, using an 80% identity threshold against type-N 18S rRNA. In summary, the analysed data represent 1,014 samples and 837 million sequence reads, containing both V4- and V9-region reads and several size fractions. Reads were sorted by length using USEARCH, and clustered into OTUs with 97% similarity using QIIME 1.9.1 with default settings (UCLUST). OTUs were then aligned with the reference alignment using PyNAST (the align_seqs.py function in QIIME). The reference alignment was the same alignment that was used to generate the reference phylogenetic tree. OTUs that the PyNAST algorithm failed to align were discarded. The PyNAST alignment output was merged with the reference alignment, and filtered for gap positions using filter_alignment.py in QIIME with the gap filtering threshold set to 0.99 and the entropy threshold set to 0.0001. Identification of apicomplexan and chrompodellid reads used a maximum likelihood phylogenetic approach by mapping the OTUs onto our reference tree using the evolutionary placement algorithm of RAxML 8.2.10[41]. The reference tree was constructed in RAxML 8.2.10 (GTR + GAMMAI model, 1,000 bootstraps). OTUs that were not placed within the Apicomplexa and chrompodellids were removed. Trees using the remaining sequences were built consecutively until no more reads were placed outside our two groups of interest.

To determine whether corallicolids exhibit host specificity, we reanalysed a recent 16S rRNA gene amplicon dataset of three deeply sampled scleractinian coral species ($n = 309$ individuals) from the Great Barrier Reef and Coral Sea[17]. ARL-V sequences were retrieved using BLASTn against the corallicolid wkC1 16S rRNA sequence. Sequences with a length match of <320 bp and identity of $<85\%$ were discarded. Out of 17.2 million sequences in the dataset, 381,504 matched these criteria. These were subsequently clustered at 99.5% similarity (reflecting potential species- and strain-level variants) in QIIME 1.9.1 using UCLUST. Singleton OTUs were removed, which resulted in 5,470 final OTUs. Beta-diversity was calculated in QIIME (binary Jaccard metric, subsampled at 100 reads per sample), and the resulting distance matrix was transformed into two-dimensional NMDS space for visualization. Demultiplexed reads from *Pachyseris speciosa* ($n = 123$) were not available, hence the data point shown for *P. speciosa* represents the averaged community composition for this coral across all 123 samples.

Metagenome database mining. Thirty-one coral-derived metagenomic and metatranscriptomic assemblies were retrieved from the Joint Genome Institute Integrated Microbial Genomes and Microbiomes database (Supplementary Table 3). These were screened for the presence of the ARL-V 16S rRNA gene (DQ200412) and the 18S rRNA genes of type-N (AF238264), *C. velia* (NC_029806) and *V. brassicaformis* (HM245049) by BLASTn searches. To retrieve corallicolid plastid and mitochondrial sequences, protein-encoding genes of *V. brassicaformis* (HM222968 and ref. [42]), *C. velia* (HM222967), *T. gondii* (U87145) and *P. falciparum* (LN999985, AY282930) plastids and mitochondria were translated and searched against the datasets with tBLASTn. All hits were manually inspected by BLAST against the GenBank nr database to verify that they corresponded to apicomplexan-related organisms. Hits that most closely matched sequences from *Symbiodinium* spp. and *Ostreobium* spp. were discarded. Hits were assembled into longer contigs using Geneious R9 (Biomatters). Assembled sequences were deposited in GenBank (see 'Data availability').

Searches were conducted for the conserved, nucleus-encoded genes *EIF5B*, *HSP75* (also known as *TRAP1*), *HSP90* (also known as *HSP90AA1*), *RPL27A*, *RPL3*, *RPL5*, *RPL8*, *RPS19*, *RPS21*, *RPS27*, *RPS8*, *RUVBL1*, *TCP1-beta* (also known as *CCT2*) and *VPC*, as well as for the plastid-encoded photosystem genes *psaA*, *psaB*, *petB*, *psbA*, *psbB*, *psbC*, *psbD*, *psbE* and *petD*, using a BLASTp cut-off of 1×10^{-10} and query length cut-off of 0.40. Hits were phylogenetically placed using FastTree 2.1.5[43] in Geneious R9. Sequences that fell within the Apicomplexa or were sister to chromerids and chrompodellids were considered to be candidate corallicolid genes. We cannot rule out the possibility that the photosystems are encoded in the nucleus: they may have been present at low abundance, and therefore undetected (metagenome and metatranscriptome coverage was $<1\times$ whole genome), or horizontally acquired from distantly related organisms and thus missed using this phylogenetic identification approach. We built a concatenated protein phylogeny of putative nucleus-encoded genes (see 'Phylogenetic analyses').

Aquarium coral survey. Coral samples were purchased from Aquariums West, or online from Canada Corals and Fragbox Corals (Supplementary Table 1). Identification was based on morphology and/or vendor labels. Corals were thoroughly rinsed with salt water (Instant Ocean Reef Crystals Salt Mix) and cut into smaller pieces containing at least 1 polyp or—for larger specimens—a portion containing skeleton, tissue and part of the oral disc. Samples were homogenized using a mortar and pestle, and DNA was extracted with the DNeasy PowerBiofilm Kit (Qiagen) according to the manufacturer's instructions. Samples were screened by PCR for the presence of the type-N 18S rRNA gene using primers 18N-F2 (5′-TAGGAATCTAAACCTCTTCCA-3′) and 18N-R1 (5′-CAGGAACAAGGGTTCCCGACC-3′)[14]. PCR was performed with Phusion DNA polymerase (Thermo Fisher Scientific) using 32 cycles of amplification (98 °C for 8 s, 60.5 °C for 15 s and 72 °C for 30 s) after an initial incubation for 30 s at 98 °C. Selected amplicons were Sanger-sequenced to verify that the type-N 18S rRNA gene was amplified.

Fluorescence microscopy. Corallicolids were visualized in the green mushroom coral (Corallimorpharia, Discosomatidae, *Rhodactis* sp.), which lacks a calcium carbonate skeleton, has a large polyp structure and is amenable to tissue fixation with the following method. Dissected tissues were placed directly in Carnoy's solution (6:3:1 ethanol:chloroform:acetic acid) and soaked overnight. Tissues were then washed with 80% (v/v) ethanol for 3×10 min, and in bleaching solution (80% v/v ethanol, 6% v/v H_2O_2) for 2×10 min. Samples were left in bleaching solution for 7 days, with replacement of solution every other day. After bleaching, tissues were washed with 100% ethanol for 2×10 min, PBSTx (0.3% v/v Triton X-100 in

phosphate-buffered saline, pH 7.4) for 3 × 10 min, and with hybridization buffer (0.9 M NaCl, 20 mM Tris-HCl, 30% v/v formamide, 0.01% w/v sodium dodecyl sulphate) for 3 × 10 min. Fluorescence in situ hybridization was carried out by incubating tissues in hybridization buffer with DAPI DNA stain (0.01 mg/ml) and fluorescent-labelled DNA probes (0.1 μM), overnight in the dark with agitation. Samples were then washed with PBSTx for 3 × 10 min, placed on glass slides with ProLong Diamond Antifade Mountant (Thermo Fisher Scientific), and visualized on a Zeiss Axioplan 2 microscope. The following probes were used: wk16P (5′- CTGCGCATATAAGGAATTAC-3′) with 5′ Texas Red-X label, targeting type-N 18S rRNA; wk17P (5′-TCAGAAGAAAGTCAAAAACG-3′) with 5′ Alexa Fluor 532 label, targeting ARL-V 16S rRNA; wk18P (5′-GCCTTCCCACAT CGTTT-3′) with 5′ Texas Red-X label, targeting Gammaproteobacteria as a control. Ectoderm, endoderm and the mesenterial filament tissues from at least three individuals were viewed; the presence of corallicolids was unambiguously detected in only the mesenterial filaments. Sample sizes were not predetermined. Sample randomization was not applicable. Blinding was not used during data acquisition and analysis.

Electron microscopy. Mesenterial filaments from *Rhodactis* sp. were fixed in 2.5% glutaraldehyde in 35‰ salt water (Instant Ocean Reef Crystals Salt Mix), then rinsed 3× in 0.1 M sodium cacodylate buffer and postfixed with 1% osmium tetroxide in 0.1 M sodium cacodylate buffer. Samples were then rinsed 3 times in distilled water, dehydrated with successive washes in 30%, 50%, 70%, 90%, 95% and 3 × 100% ethanol, and embedded in Spurr's epoxy resin following infiltration. The resulting blocks were cut into 70-nm sections, and stained using 2% aqueous uranyl acetate (12 min) and 2% aqueous lead citrate (6 min). Sections were viewed under a Hitachi H7600 transmission electron microscope (UBC Bioimaging Facility).

Metagenomic sequencing and plastid genome closing. Insufficient read coverage and the presence of sequence variants meant that an unambiguous plastid assembly was not possible from publically available metagenomes. Thus, we used a combination of metagenomic sequencing and primer walking to obtain a complete corallicolid plastid genome from the *Rhodactis* sp. host. To enrich for corallicolids, the cnidoglandular lobes of *Rhodactis* sp. were removed from mesenterial filaments after soaking in 100% ethanol. DNA was extracted using the DNeasy PowerBiofilm Kit (Qiagen). Two Nextera XT libraries were generated using 1 ng and 10 ng of template DNA, according to the manufacturer's instructions (Illumina). Libraries were sequenced on a single lane of Illumina HiSeq 2500 (The Centre for Applied Genomics, The Hospital for Sick Children), generating 260.8 million paired-end 125-bp reads. Reads were assembled with MEGAHIT 1.1.2[44]. The mitochondrial genomes of the coral and the 18S, 5.8S and 28S rRNA genes of corallicolids were retrieved from the assembly. Reads were mapped using Bowtie 2[45] against mitochondrial and plastid contigs assembled from the 31 previous metagenomic datasets (see above). Read coverage was <1×; therefore, to close the corallicolid organelle genomes, gap-spanning regions were PCR-amplified with sequence-specific primers and Phusion DNA polymerase (Thermo Fisher Scientific). Amplicons were sequenced using the Sanger-sequencing method, and the resulting reads were assembled in Geneious R9. A predicted 68-bp hairpin region in the plastid genome was unable to be bridged by PCR. The sequence for this region was filled in using reads from the *Cyphastrea* sp. 2 (GOLD Analysis Project ID Ga0126343) metagenome that spanned the gap.

Phylogenetic analyses. *Nuclear rRNA gene phylogeny*. Related to Fig. 3a. 18S and 28S rRNA sequences were aligned with SINA 1.2.11[46], and 5.8S rRNA sequences were aligned with MUSCLE and edited manually in Geneious R9. Alignments of the three genes were concatenated. A maximum likelihood phylogeny was built with the GTR + GAMMA model (1,000 bootstrap replicates) in RAxML 8.2.10. A phylogeny based on Bayesian inference was constructed using the GTR + GAMMA model (2×10^5 generation run-time with tree sampling every 200 generations, 0.25 fraction burn-in) in MrBayes 3.2[47]).

Nuclear protein phylogeny. Related to Extended Data Fig. 8. Seven putative nucleus-encoded corallicolid proteins (Supplementary Table 6) were aligned against orthologues from Apicomplexa, Dinoflagellata and Ciliophora taxa using MUSCLE in Geneious R9, and concatenated. Phylogenies were built using RAxML 8.2.10 (GTR + GAMMA model, 1,000 bootstraps) and MrBayes 3.2 (GTR + GAMMA model, 10^5 generation run-time with tree sampling every 200 generations, 0.25 fraction burn-in).

Mitochondrial protein phylogeny. Related to Extended Data Fig. 6a. The three mitochondria-encoded genes, *cox1*, *cox3* and *cob*, were translated and aligned with MUSCLE in Geneious R9. The genes from *V. brassicaformis* were excluded owing to their extreme divergence and the resulting long branch in the final tree. Phylogenies were constructed in RAxML 8.2.10 (MtZoa + GAMMA model, 1,000 bootstraps) and MrBayes 3.2 (mixed + GAMMA model, 10^5 generation run-time with tree sampling every 200 generations, 0.25 fraction burn-in).

Plastid protein phylogeny. Related to Extended Data Fig. 6b. Nineteen plastid-encoded genes common to apicomplexans, chromerids and stramenopiles (*rpl2*, *rpl4*, *rpl6*, *rpl14*, *rpl16*, *rps2*, *rps3*, *rps4*, *rps7*, *rps8*, *rps11*, *rps12*, *rps17*, *rps19*, *clpC*, *rpoB*, *rpoC1*, *rpoC2* and *tufA*) were translated and aligned with MUSCLE in Geneious R9. Split genes were concatenated, and the most-conserved copy of duplicated (paralogous) genes were used for this analysis. Genes were translated with Genetic Code 4, as the TGA stop codon codes for tryptophan in chromerid and apicomplexan plastids (including in that of corallicolids). Protein sequences were concatenated and phylogenies were built in RAxML 8.2.10 (cpREV + GAMMA + F model, 1,000 bootstraps), in MrBayes 3.2 (cpREV + GAMMA model, 3×10^5 generation run-time with tree sampling every 500 generations, 0.25 fraction burn-in), and with the neighbour-joining algorithm (Jukes–Cantor model, 1,000 bootstraps). Phylogenies for concatenated ChlN, ChlB and ChlL proteins and for AcsF (Extended Data Fig. 6c) were also generated as above with RAxML 8.2.10 and with MrBayes 3.2 (cpREV + GAMMA model, 10^5 generation run-time with tree sampling every 5,200 generations, 0.25 fraction burn-in).

Plastid single-gene analysis. Related to Extended Data Fig. 9b. Phylogenetic trees were constructed for each of the above 19 proteins, as well as for SufB, Rps5, Rpl11, Rpl36 and the 16S and 23S rRNA genes (*rrs* and *rrl*, respectively). MUSCLE and SINA were used to align proteins and rRNAs, respectively. Trees were built using RAxML 8.2.10 (cpREV + GAMMA + F and GTR + GAMMA models, 500 bootstraps), MrBayes 3.2 (cpREV + GAMMA model and GTR + GAMMA and Poisson + GAMMA models, tree sampling every 200 generations for 10^5 generations, 0.25 fraction burn-in), neighbour joining (Jukes–Cantor model, 1,000 bootstraps) and maximum parsimony (all sites, 100 bootstraps) in MEGA 7[48].

Accession numbers of the sequences used in this study are listed in Supplementary Table 7.

Other methods. Coral phylogeny (Fig. 1c) was based on the established topology of the clades[49–51] and on published analyses that use molecular clocks and fossils to date clade divergences[52–55].

Pairwise *d*N values were calculated from translation-aligned nucleotide sequences using codeml from the PAML package[56]. The following settings were used: Seqtype = 1: codons, alpha = 0 (fixed), Small_Diff = 5e-07, model = 0: one w, runmode = -2, clock = 0, Mgene = 0, CodonFreq = 2: F3X4, estFreq = 0, fix_blength = 0, optimization method = 0, icode = 3: mould mt. Additional pairwise comparisons were made using KaKs_Calculator[57], with the GY-HKY method.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

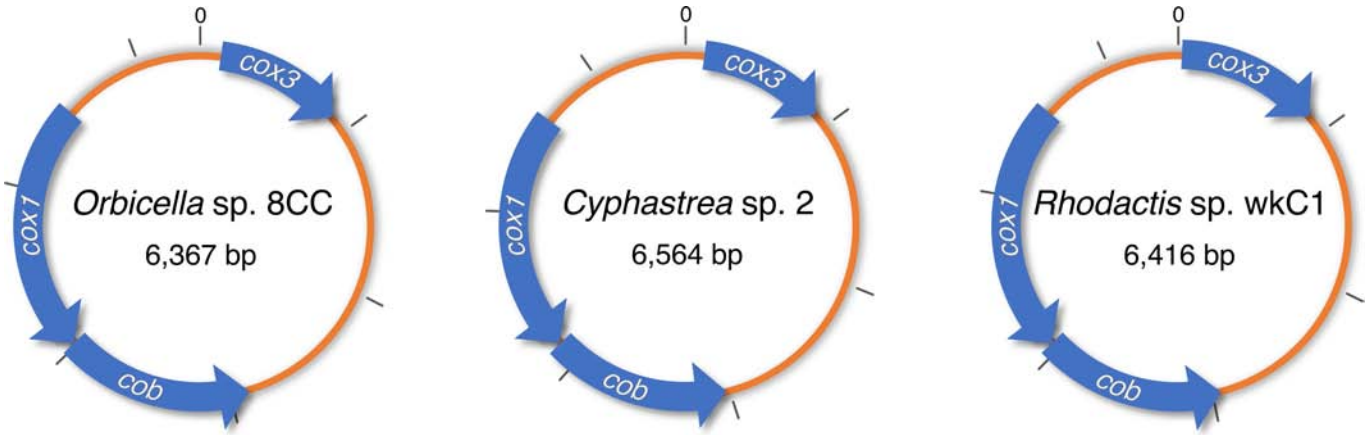
Data availability

The following are deposited in GenBank: the *Rhodactis* sp. wkC1 mitochondrial genome (accession number MH320096); corallicolid 18S, 5.8S and 28S rRNA genes from *Rhodactis* sp. wkC1 (MH304760, MH304761), *Orbicella* sp. TRC (MH304758) and *Cyphastrea* sp. 2 (MH304759); corallicolid mitochondrial genomes from *Rhodactis* sp. wkC1 (MH320093), *Orbicella* sp. 8CC (MH320094) and *Cyphastrea* sp. 2 (MH320095); and the corallicolid plastid genome from *Rhodactis* sp. wkC1 (MH304845).

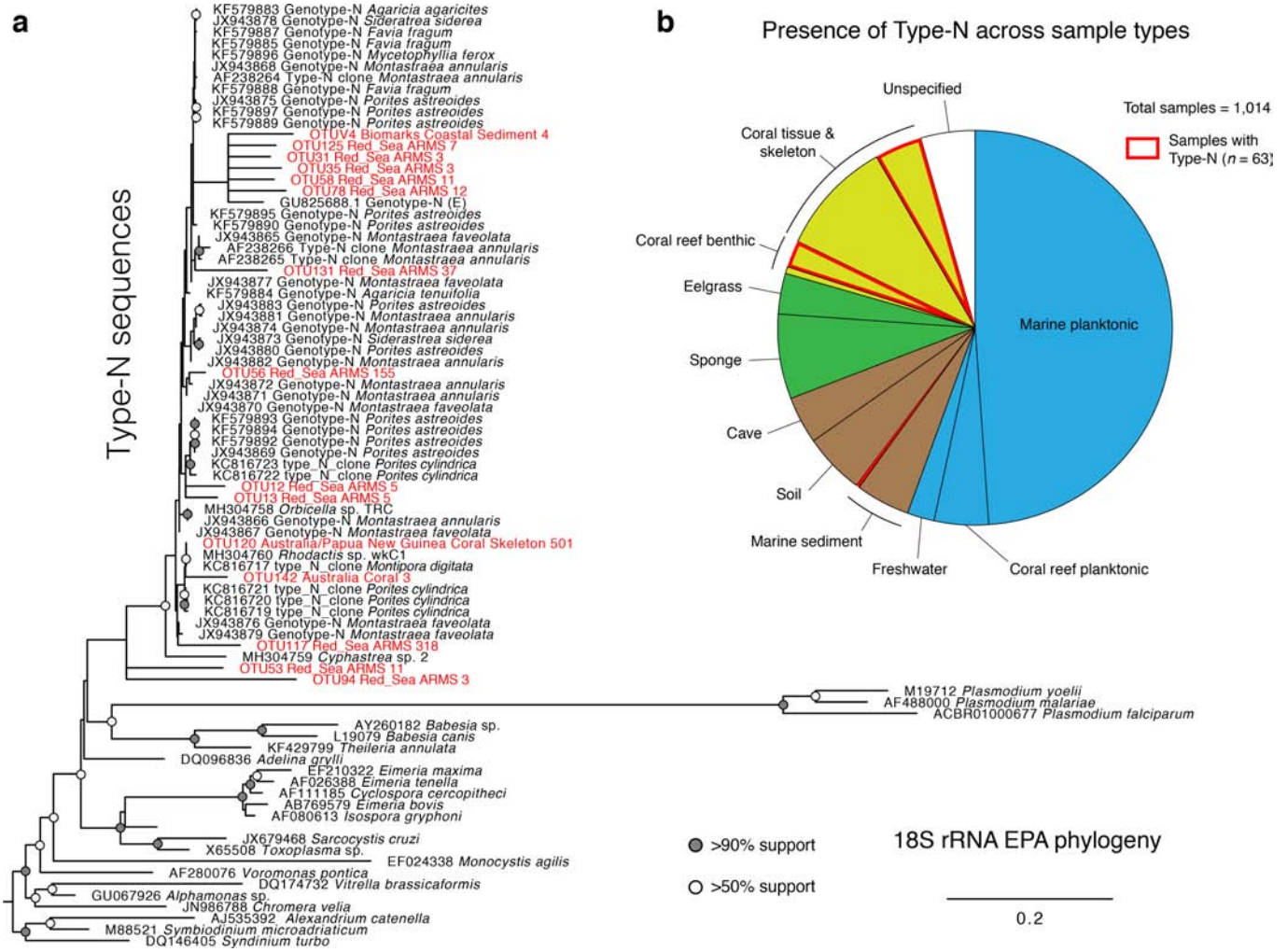
The 18S rRNA and 16S rRNA gene amplicon reads are deposited in the NCBI Sequence Read Archive (PRJNA482746).

31. Comeau, A. M., Li, W. K. W., Tremblay, J.-É., Carmack, E. C. & Lovejoy, C. Arctic Ocean microbial community structure before and after the 2007 record sea ice minimum. *PLoS ONE* **6**, e27492 (2011).
32. Bower, S. M. et al. Preferential PCR amplification of parasitic protistan small subunit rDNA from metazoan tissues. *J. Eukaryot. Microbiol.* **51**, 325–332 (2004).
33. del Campo, J. et al. A universal set of primers to study animal associated microeukaryotic communities. Preprint at https://www.biorxiv.org/content/10.1101/485532v1 (2018).
34. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, e2584 (2016).
35. Caporaso, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
36. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2013).
37. del Campo, J., Pombert, J. F., Šlapeta, J., Larkum, A. & Keeling, P. J. The 'other' coral symbiont: *Ostreobium* diversity and distribution. *ISME J.* **11**, 296–299 (2017).
38. Huse, S. M. et al. VAMPS: a website for visualization and analysis of microbial population structures. *BMC Bioinformatics* **15**, 41 (2014).
39. Massana, R. et al. Marine protist diversity in European coastal waters and sediments as revealed by high-throughput sequencing. *Environ. Microbiol.* **17**, 4035–4049 (2015).
40. de Vargas, C. et al. Eukaryotic plankton diversity in the sunlit ocean. *Science* **348**, 1261605 (2015).
41. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
42. Flegontov, P. et al. Divergent mitochondrial respiratory chains in phototrophic relatives of apicomplexan parasites. *Mol. Biol. Evol.* **32**, 1115–1131 (2015).

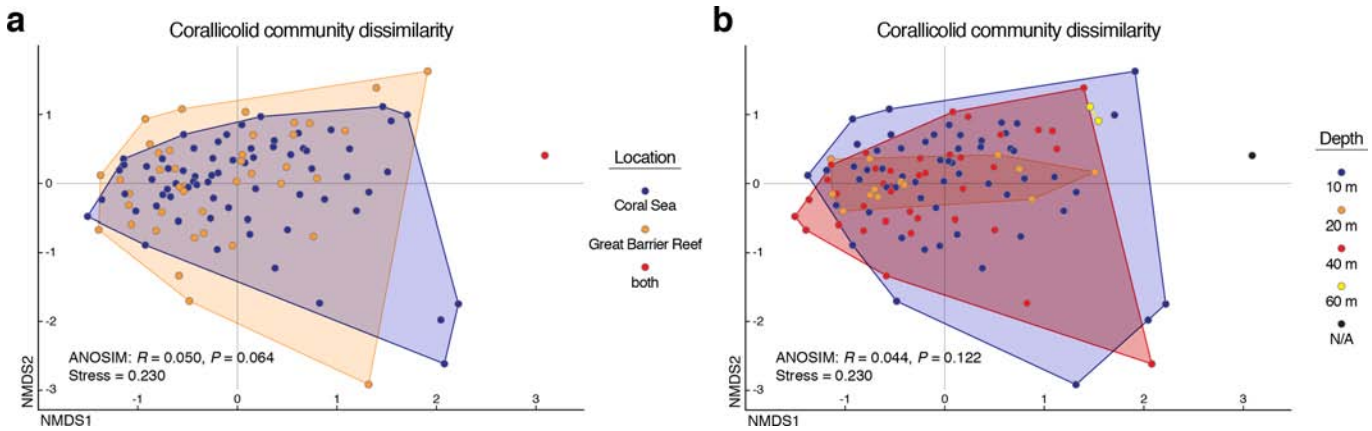
43. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).
44. Li, D. et al. MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods* **102**, 3–11 (2016).
45. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
46. Pruesse, E., Peplies, J. & Glöckner, F. O. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**, 1823–1829 (2012).
47. Ronquist, F. et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol*. **61**, 539–542 (2012).
48. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol*. **33**, 1870–1874 (2016).
49. Kayal, E., Roure, B., Philippe, H., Collins, A. G. & Lavrov, D. V. Cnidarian phylogenetic relationships as revealed by mitogenomics. *BMC Evol. Biol*. **13**, 5 (2013).
50. Zapata, F. et al. Phylogenomic analyses support traditional relationships within Cnidaria. *PLoS ONE* **10**, e0139068 (2015).
51. Kayal, E. et al. Phylogenomics provides a robust topology of the major cnidarian lineages and insights on the origins of key organismal traits. *BMC Evol. Biol*. **18**, 68 (2018).
52. Simpson, C., Kiessling, W., Mewis, H., Baron-Szabo, R. C. & Müller, J. Evolutionary diversification of reef corals: a comparison of the molecular and fossil records. *Evolution* **65**, 3274–3284 (2011).
53. Stolarski, J. et al. The ancient evolutionary origins of Scleractinia revealed by azooxanthellate corals. *BMC Evol. Biol*. **11**, 316 (2011).
54. Park, E. et al. Estimation of divergence times in cnidarian evolution based on mitochondrial protein-coding genes and the fossil record. *Mol. Phylogenet. Evol*. **62**, 329–345 (2012).
55. Baliński, A., Sun, Y. & Dzik, J. 470-million-year-old black corals from China. *Naturwissenschaften* **99**, 645–653 (2012).
56. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol*. **24**, 1586–1591 (2007).
57. Zhang, Z. et al. KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. *Genomics Proteomics Bioinformatics* **4**, 259–263 (2006).
58. Kořený, L., Sobotka, R., Janouskovec, J., Keeling, P. J. & Oborník, M. Tetrapyrrole synthesis of photosynthetic chromerids is likely homologous to the unusual pathway of apicomplexan parasites. *Plant Cell* **23**, 3454–3462 (2011).



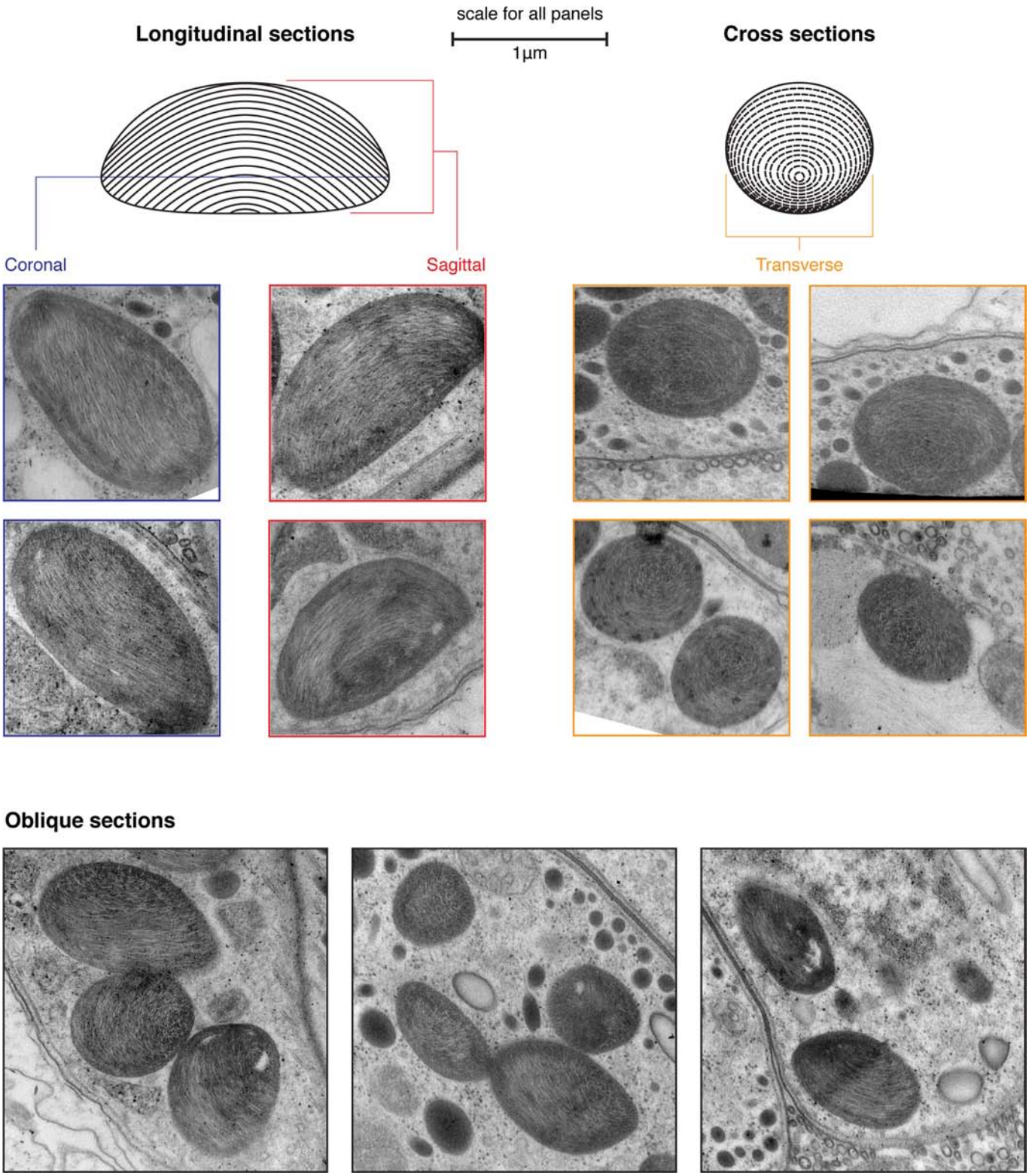
Extended Data Fig. 1 | Mitochondrial genomes of corallicolids. Names denote the host coral from which the genomes were retrieved. The three mitochondria-encoded genes are shown in blue. Tick marks (moving clockwise) denote 1,000 bp. It is unclear whether the genomes are circular (as depicted), or tandem linear.



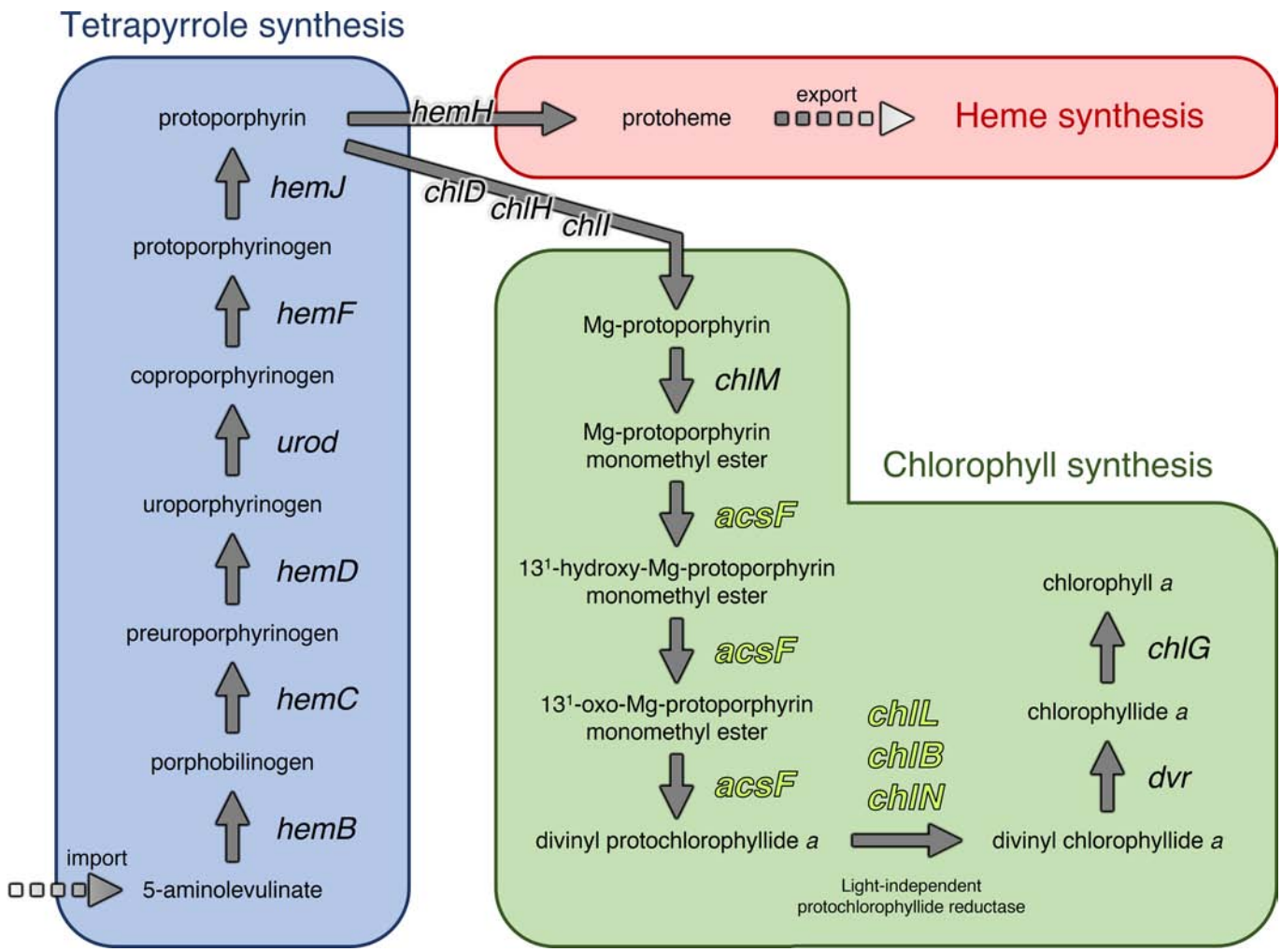
Extended Data Fig. 2 | Distribution and diversity of corallicolid type-N from eukaryotic microbiome surveys. a, Phylogenetic placement of short-amplicon OTUs (red) and near-full-length sequences (black), which show the diversity of the type-N clade. Coral host species are indicated. Values at nodes denote maximum likelihood bootstrap support ($n = 1{,}000$). Relationships between type-N lineages were generally poorly resolved. **b**, Presence of type-N reads in 18S rRNA gene surveys from environmental and host-associated samples, which shows that type-N is largely restricted to corals. Surveys included in this analysis are listed in Supplementary Table 5.



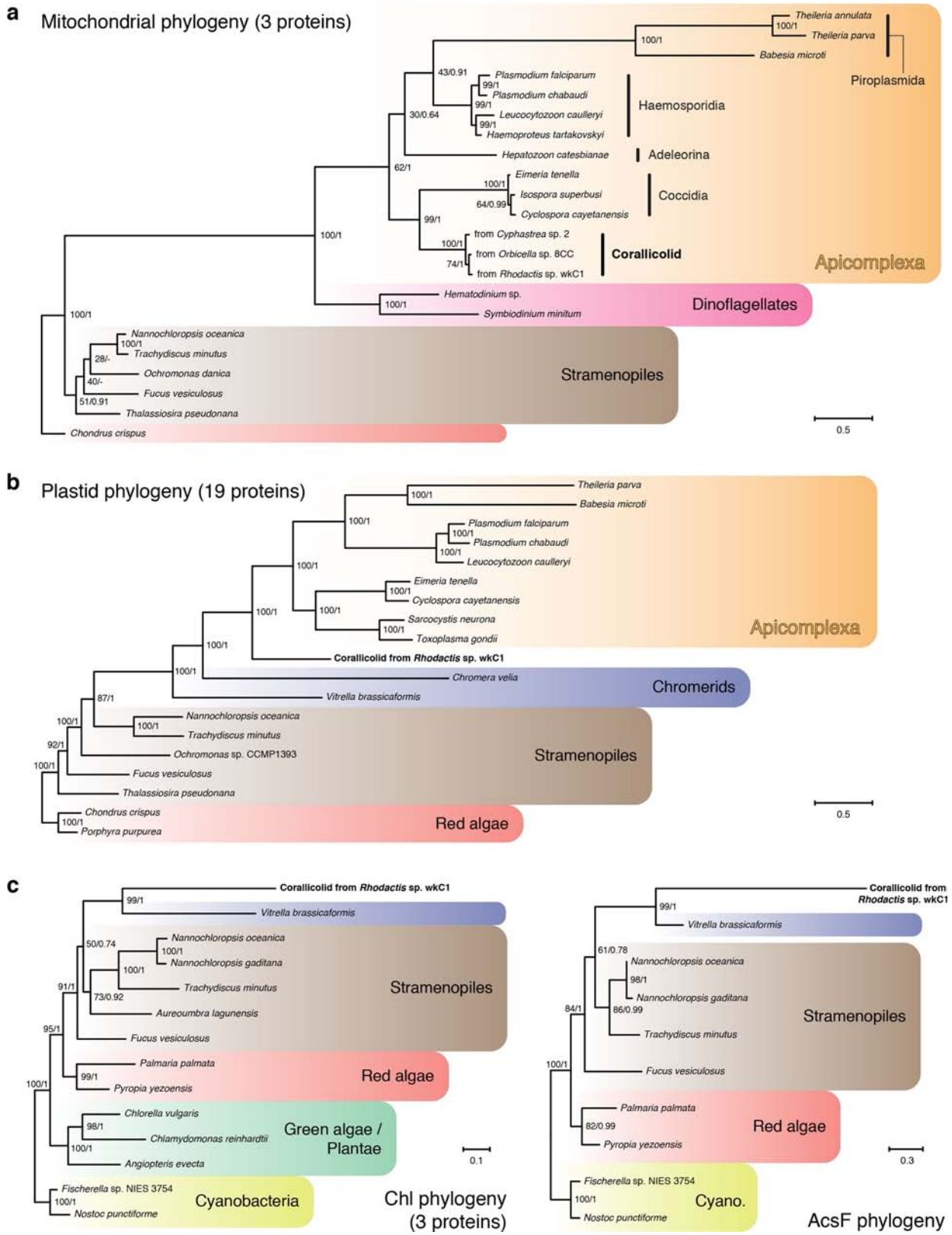
Extended Data Fig. 3 | No correlation of ARL-V community structure with abiotic factors. a, Geographical location. **b**, Water depth. Correlation calculated using ANOSIM with 999 permutations. N/A, not available.



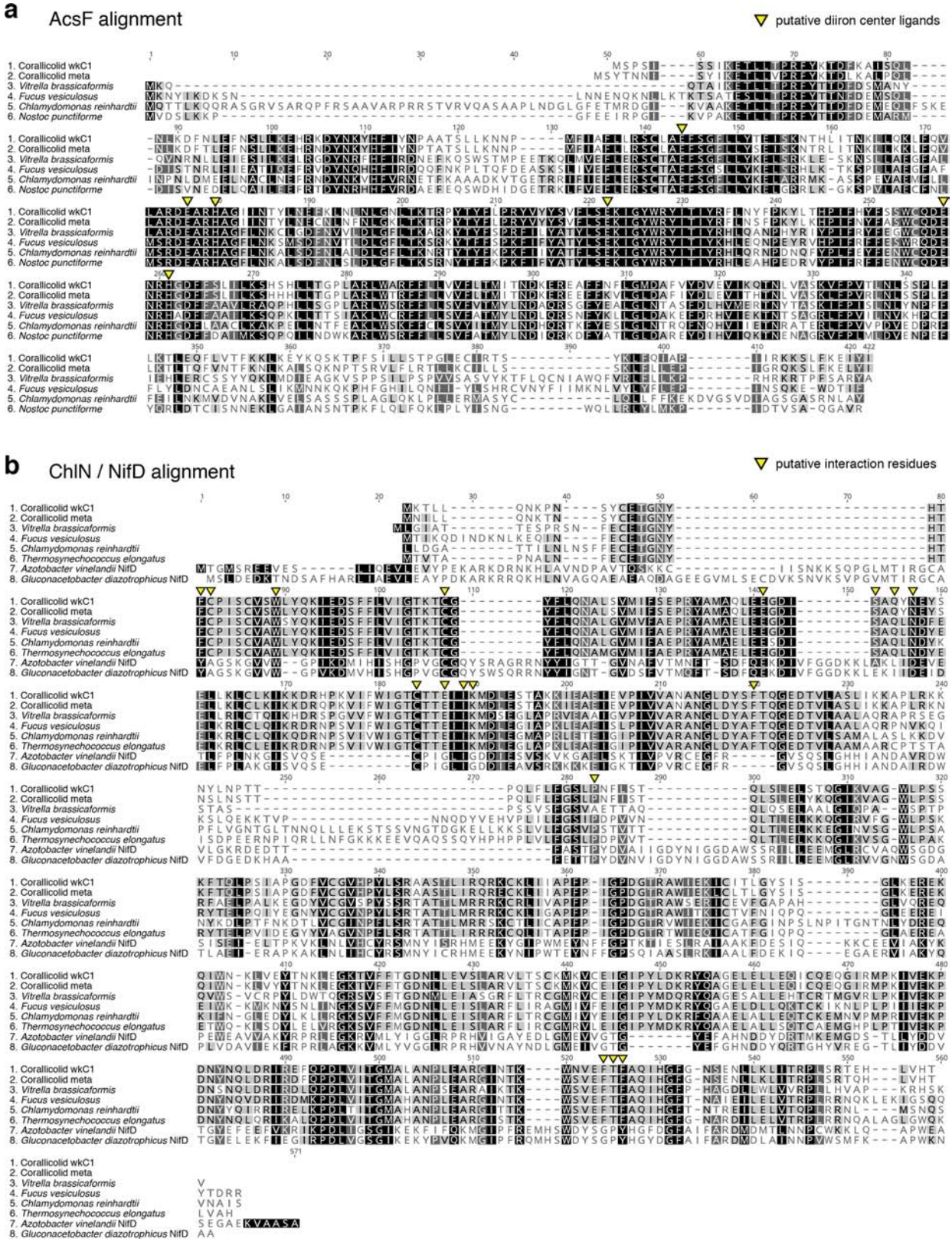
Extended Data Fig. 4 | Transmission electron micrographs of darkly stained organelles in corallicolid cells, showing distinctive internal structures. Structure and orientations (sagittal and transverse sections illustrated at the top) were inferred from viewing multiple organelles from several cells. Imaging was conducted in triplicate; representative results are shown.



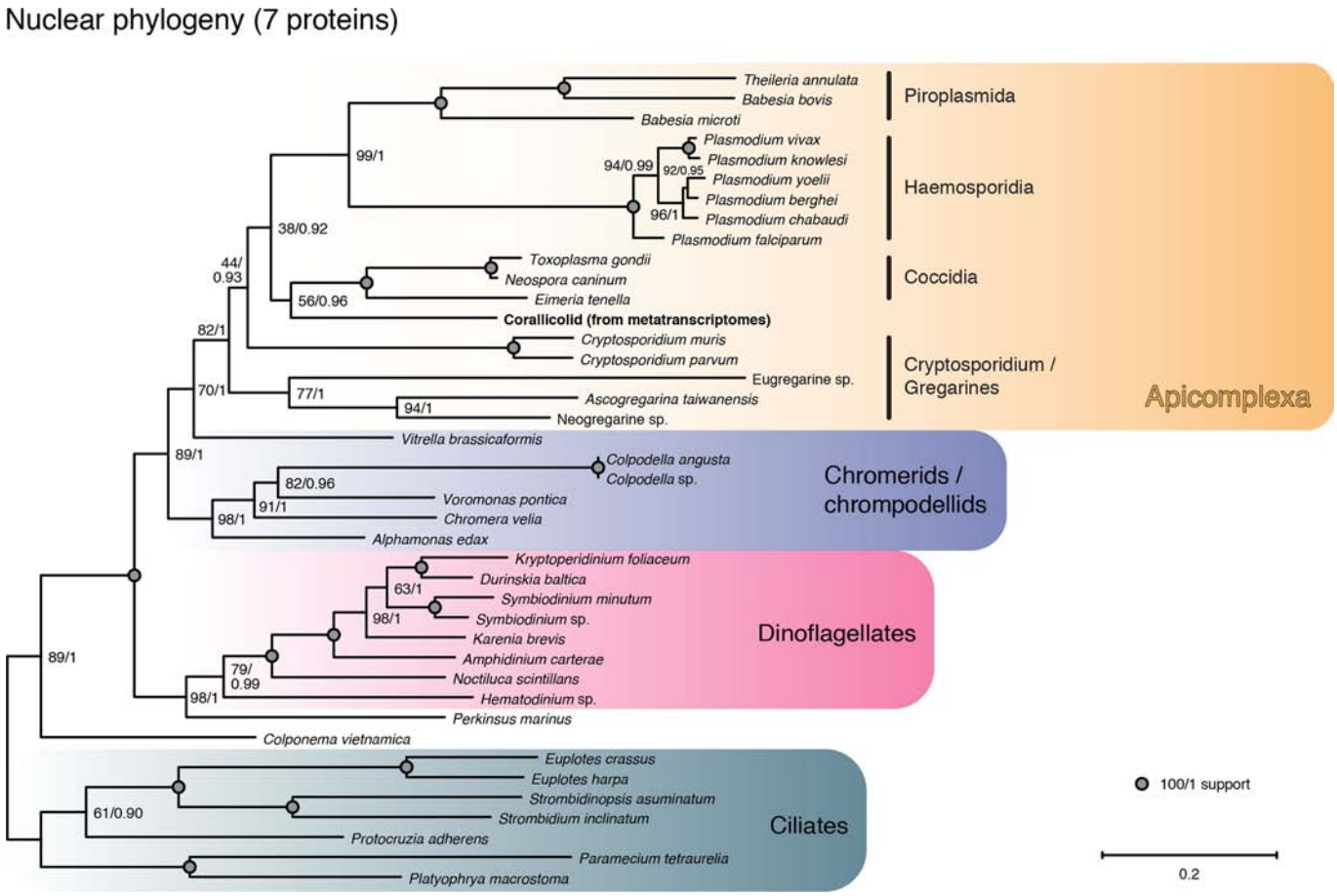
Extended Data Fig. 5 | Tetrapyrrole and chlorophyll biosynthesis pathways, showing putative function of genes. Genes retained in the corallicolid plastid genome (*acsF*, *chlL*, *chlN* and *chlB*) are highlighted. All enzymatic steps depicted here are inferred to occur within the apicomplexan plastid[58].



Extended Data Fig. 6 | Position of corallicolids in phylogenetic trees. **a**, Phylogenetic placement of corallicolids, based on mitochondria-encoded proteins. **b**, Phylogenetic placement of corallicolids, based on plastid-encoded proteins. **c**, Phylogenetic placement of corallicolid chlorophyll biosynthesis proteins (concatenation of proteins ChlL, ChlN and ChlB on the left; AcsF on the right). All phylogenetic trees shown were produced with the maximum likelihood algorithm; values at nodes denote maximum likelihood bootstrap support percentages ($n = 1{,}000$ replicates) and Bayesian posterior probabilities (see Methods).



Extended Data Fig. 7 | Conservation of key amino acid residues implies conservation of protein function in chlorophyll biosynthesis genes. **a**, **b**, Sequence alignment of AcsF (**a**) and ChlN (**b**) proteins. For ChlN alignment, a comparison to NifD is shown. 'Corallicolid meta' sequence is derived from metagenomics and metatranscriptomics assembly; 'Corallicolid wkC1' is from the complete plastid sequence.



Extended Data Fig. 8 | Phylogenetic placement of corallicolids, based on putative nucleus-encoded proteins. Based on concatenation of HSP90, RPL3, RPL27A, RPS8, RPS19, RPS21 and RPS27 proteins (Supplementary Table 6). The tree shown was produced with the maximum likelihood algorithm; values at nodes denote maximum likelihood bootstrap support percentages ($n = 1{,}000$ replicates) and Bayesian posterior probabilities (see Methods).

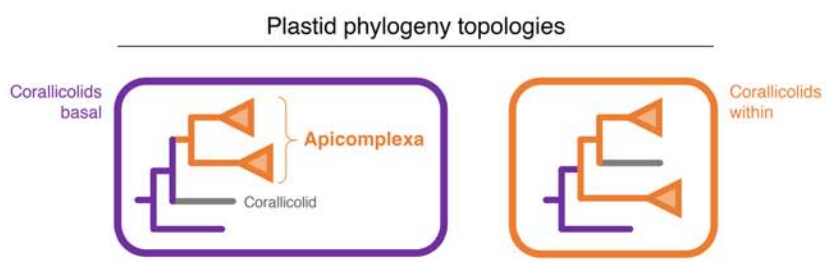
a

	Chromera velia	*Ochromonas* sp. CCMP 1393	*Vitrella brassicaformis*	*Fucus vesiculosus*	*Thalassiosira pseudonana*	*Trachydiscus minutus*	*Nannochloropsis oceanica*	*Chondrus crispus*	*Porphyra purpurea*	**Corallicolid**	*Cyclospora cayetanensis*	*Eimeria tenella*	*Toxoplasma gondii*	*Sarcocystis neurona*	*Leucocytozoon caulleryi*	*Plasmodium falciparum*	*Plasmodium chabaudi*	*Theileria parva*	*Babesia microti*
Chromera velia		13.154	13.732	13.616	13.354	13.78	12.599	13.617	13.569	14.043	12.689	11.918	12.628	12.687	11.728	11.82	11.793	9.558	9.506
Ochromonas sp. CCMP 1393	0.6335		27.928	41.023	38.841	37.554	34.77	42.288	41.995	27.964	24.795	23.845	24.289	24.427	22.171	22.368	22.562	18.639	18.372
Vitrella brassicaformis	0.6869	0.5605		29.368	27.735	28.403	26.132	30.329	30.53	30.282	25.385	23.698	25.29	25.956	21.8	21.865	21.707	18.732	17.346
Fucus vesiculosus	0.6297	0.288	0.5488		45.959	42.132	39.334	51.377	51.364	30.164	26.23	25.028	26.25	26.12	23.894	24.165	24.488	19.951	19.588
Thalassiosira pseudonana	0.6228	0.2874	0.5482	0.2523		38.996	36.916	48.717	48.844	27.927	23.881	22.803	23.931	23.967	21.414	21.154	21.486	18.313	17.571
Trachydiscus minutus	0.6274	0.3243	0.5516	0.3009	0.3055		49.986	42.322	42.269	27.771	24.812	23.646	24.088	23.947	21.66	21.733	21.818	19.011	18.226
Nannochloropsis oceanica	0.6575	0.321	0.5587	0.3059	0.3098	0.2018		40.36	40.322	26.248	23.585	22.493	23.065	23.16	20.37	20.482	20.775	17.898	17.355
Chondrus crispus	0.6228	0.2813	0.5358	0.2399	0.2335	0.3054	0.3016		74.511	30.92	26.423	25.38	25.729	26.379	23.93	23.682	23.737	19.884	19.527
Porphyra purpurea	0.6373	0.2821	0.5416	0.2426	0.2342	0.3018	0.3026	0.0994		30.22	25.673	24.41	24.755	25.508	23.216	22.731	22.852	19.259	18.375
Corallicolid	0.6196	0.5237	0.5565	0.5035	0.5173	0.5361	0.5379	0.5162	0.5197		37.251	34.997	36.561	37.189	30.67	30.476	30.766	22.765	22.403
Cyclospora cayetanensis	0.6685	0.6111	0.655	0.6032	0.6091	0.611	0.6199	0.6133	0.6259	0.4508		67.084	40.907	41.748	32.589	33.895	33.349	24.357	25.064
Eimeria tenella	0.6875	0.6126	0.6809	0.6083	0.6137	0.6244	0.6305	0.6197	0.6361	0.452	0.12		38.722	39.632	31.378	32.574	31.965	24.402	25.144
Toxoplasma gondii	0.695	0.6221	0.6577	0.6035	0.6104	0.6363	0.6371	0.624	0.6381	0.4472	0.3955	0.3989		67.888	32.408	33.333	32.805	23.911	24.551
Sarcocystis neurona	0.6871	0.6113	0.6517	0.604	0.6058	0.6287	0.6323	0.6082	0.6222	0.4433	0.3792	0.3809	0.1486		32.648	34.074	33.344	24.691	24.199
Leucocytozoon caulleryi	0.772	0.7152	0.7827	0.7	0.7151	0.7136	0.7356	0.7064	0.7324	0.5924	0.5544	0.5547	0.5548	0.5497		69.662	69.963	26.37	27.391
Plasmodium falciparum	0.7618	0.7226	0.7907	0.7029	0.7119	0.72	0.7303	0.712	0.7389	0.5837	0.5415	0.5411	0.5538	0.5357	0.1385		75.918	27.056	28.133
Plasmodium chabaudi	0.7601	0.7162	0.7884	0.6927	0.7139	0.7202	0.7271	0.7121	0.731	0.5765	0.551	0.5581	0.5521	0.5485	0.1482	0.0974		26.991	27.582
Theileria parva	0.8695	0.7729	0.8939	0.7693	0.7641	0.7897	0.807	0.7903	0.7974	0.7301	0.7007	0.7015	0.721	0.6932	0.6772	0.6732	0.6685		27.586
Babesia microti	0.8741	0.798	0.9235	0.7948	0.8065	0.8164	0.8394	0.8053	0.8314	0.7452	0.6856	0.6837	0.725	0.7237	0.6632	0.6468	0.6724	0.6531	

*d*N values
(number of non-synonymous substitutions per non-synonymous site)
>0.6 0.5 0.4 0.3 0.2 <0.2

Amino acid identity (%)
<20 20 30 40 >50

b



Gene or Protein	NJ	MP	Bayesian (Poisson)	Bayesian (cpREV)	Bayesian (GTR)	ML (cpREV)	ML (GTR)
rrs (16S)			N/A	N/A		N/A	
rrl (23S)			N/A	N/A		N/A	
ClpC							
Rpl2							
Rpl4							
Rpl6							
Rpl14							
Rpl16							
RpoB							
RpoC1							
RpoC2							
Rps2							
Rps3							
Rps4							
Rps7							
Rps8			-	-			
Rpl11							
Rps12				-			
Rps17							
Rps19							
SufB							
TufA							
Rps5							
Rpl11							
Rpl36	-	-	-	-	-	-	-

Extended Data Fig. 9 | Corallicolid phylogenetic placement using plastid data shows ambiguity. a, Pairwise amino acid identities and *d*N values support a close relationship between corallicolids and the Coccidia. **b**, Phylogenetic analysis of single plastid genes and proteins to test alternative topologies of corallicolid placement. Results vary by gene and methodology: although most plastid genes show a basal placement for corallicolids, a few support the grouping of corallicolids within the Apicomplexa. Tree construction methods are indicated at top, with the model of evolution in parentheses. NJ, neighbour joining; MP, maximum parsimony; ML, maximum likelihood. A dash indicates a lack of support for either topology. N/A, not applicable.

nature research

Corresponding author(s): Waldan K. Kwong

Last updated by author(s): Feb 23, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data from public databases were downloaded using the NCBI SRA Toolkit v2.9.1, and the JGI IMG/ER v5.0 web platform.
Data analysis	The following software was used for data analysis: QIIME v1.9.1, Megahit v1.1.2, Bowtie2, RAxML v8.2.10, Geneious R9, BLAST v2.7.1+, FastTree v2.1.5, MrBayes v3.2, PAML v4.9, KaKs_Calculator, VSEARCH, MEGA 7, SINA 1.2.11, MUSCLE, PyNAST, USEARCH, and UCLUST.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The following are deposited in GenBank: The Rhodactis sp. wkC1 mitochondrial genome (accession no. MH320096); Corallicolid 18S/5.8S/28S rRNA genes from Rhodactis sp. wkC1 (MH304760, MH304761), Orbicella sp. TRC (MH304758), Cyphastrea sp. 2 (MH304759); Corallicolid mitochondrial genomes from Rhodactis sp. wkC1 (MH320093), Orbicella sp. 8CC (MH320094), Cyphastrea sp. 2 (MH320095); Corallicolid plastid genome from Rhodactis sp. wkC1 (MH304845). The 18S rRNA and 16S rRNA gene amplicon reads are deposited in the NCBI Sequence Read Archive (PRJNA482746).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	This study involved new collection of corals to determine their microbial community composition, as well as a survey of existing coral and environmental datasets to ascertain the prevalence and distribution of corallicolids. Samples were not assigned to experimental/ treatment groups (not applicable). Microscopy data were also collected.
Research sample	Wild adult scleractinian corals (phylum Cnidaria, class Anthozoa, subclass Hexacorallia, order Scleractinia) and commercial aquarium-sourced adult anthozoans (phylum Cnidaria, class Anthozoa) were newly collected for this study. Existing, publicly available datasets were also analyzed as part of this study; these data sources are listed in Extended Data Tables 3, 4, and 5.
Sampling strategy	For collection of wild corals, sample size was chosen such that each of the most common species of scleractinian coral found in Curaçao were sampled at least once. Aquarium samples were purchased from commercial livestock vendors, with the aim of collecting at least one representative from each major Anthozoa clade. Samples were collected by mechanically severing fragments from coral colonies. Samples were immediately returned to the laboratory and frozen before subsequent DNA extraction.
Data collection	DNA sequence data from public databases were downloaded and analyzed by WKK and VM. New sequence data was obtained and analyzed by WKK and JdC, from amplicon sequencing performed at the Integrated Microbiome Resource facility at the Centre for Comparative Genomics and Evolutionary Bioinformatics at Dalhousie University, metagenomic sequencing performed at The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Canada, and Sanger sequencing performed at the University of British Columbia. Microscopy data was generated by WKK at the University of British Columbia.
Timing and spatial scale	Collection of wild corals was performed at several locations in Curaçao in April 2015 (12.122266, -68.969362; 12.108323, -68.953381; 12.036822, -68.777855).
Data exclusions	No data were excluded from the analyses.
Reproducibility	Microscopy data were collected in triplicate, using separate individuals. All new sequence data were deposited in publicly-available online repositories, and support values for phylogenetic analyses are reported in the paper.
Randomization	Not applicable, as individuals were not allocated into groups in this study.
Blinding	Blinding was not used during data acquisition and analysis. All analyzed samples and datasets are presented in this study without omission.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions	Coral samples were collected on near shore, fringing reefs in Curaçao, at a depth of between 0 and 30 m.
Location	Corals were collected from several locations in Curaçao in April 2015 (12.122266, -68.969362; 12.108323, -68.953381; 12.036822, -68.777855), at a depth of between 0 and 30 m.
Access and import/export	Corals were collected under the permits of the Dutch Antillean Government (Government reference: 2012/48584) provided to the CARMABI Foundation (CITES Institution code AN001), issued on February 13, 2013. Corals were accessed by shore diving as well as by boat, when land access was unavailable. Access and import/export was conducted in compliance with local, national, and international laws.
Disturbance	Disturbance was minimized by taking only small fragments of coral (< 2 cm^2) and by carefully avoiding physical contact with reef bottom and other organisms while sampling.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Commercially purchased Rhodactis sp. corallimorphs (phylum Cnidaria, class Anthozoa, subclass Hexacorallia, order Corallimorpharia) were maintained in the laboratory.
Wild animals	Samples of wild corals were collected by mechanically severing fragments from coral colonies. Samples were immediately returned to the laboratory and frozen before subsequent DNA extraction.
Field-collected samples	Rhodactis sp. were housed in saltwater tanks (salinity 1.025, 25°C) with a photoperiod of 7.5 hr light/16.5 hr dark. Animals that were destructively sampled were disposed in laboratory biowaste.
Ethics oversight	No ethical guidance or approval was required, as the animals used were invertebrates and fall outside of CCAC (Canadian Council on Animal Care) guidelines.

Note that full information on the approval of the study protocol must also be provided in the manuscript.