environmental microbiology

Environmental Microbiology (2018) **20**(8), 2824–2833 doi:10.1111/1462-2920.14134

Global diversity and distribution of close relatives of apicomplexan parasites

Varsha Mathur ,[1]* Javier del Campo,[1,2] Martin Kolisko[1,3] and Patrick J. Keeling[1]

[1]*Department of Botany, University of British Columbia, Vancouver, British Columbia, Canada.*

[2]*Department of Marine Biology and Oceanography, Institut de Ciències del Mar (CSIC), Barcelona, Spain.*

[3]*Institute of Parasitology, Biology Centre, Czech Academy of Sciences, Branišovská 31, 370 05 České Budějovice, Czech Republic.*

Summary

Apicomplexans are a group of obligate intracellular parasites, but their retention of a relict non-photosynthetic plastid reveals that they evolved from free-living photosynthetic ancestors. The closest relatives of apicomplexans include photosynthetic chromerid algae (e.g., *Chromera* and *Vitrella*), non-photosynthetic colpodellid predators (e.g., *Colpodella*) and several environmental clades collectively called Apicomplexan-Related Lineages (ARLs). Here we investigate the global distribution and inferred ecology of the ARLs by expansively searching for apicomplexan-related plastid small ribosomal subunit (SSU) genes in large-scale high-throughput bacterial amplicon surveys. Searching more than 220 million sequences from 224 geographical sites worldwide revealed 94 324 ARL plastid SSU sequences. Meta-analyses confirm that all ARLs are coral reef associated and not to marine environments generally, but only a subset is actually associated with coral itself. Most unexpectedly, *Chromera* was found exclusively in coral biogenous sediments, and not within coral tissue, indicating that it is not a coral symbiont, as typically thought. In contrast, ARL-V is the most diverse, geographically widespread and abundant of all ARL clades and is strictly associated with coral tissue and mucus. ARL-V was found in 19 coral species in reefs, including azooxanthellate corals at depths greater than 500 m. We suggest this is indicative of a parasitic or commensal relationship, and not of photosynthetic symbiosis, further underscoring the importance of isolating ARL-V and determining its relationship with the coral host.

Received 19 February, 2018; revised 3 April, 2018; accepted 5 April, 2018. *For correspondence. E-mail varsha.mathur@botany.ubc.ca; Tel. +1 (604) 822-2845; Fax +1 (604) 822-6089.

Introduction

Apicomplexans are a large group of parasitic protists that cause devastating diseases such as malaria, toxoplasmosis and cryptosporidiosis. Although apicomplexans are obligate intracellular parasites of animals, they have been found to harbour a non-photosynthetic plastid, known as an apicoplast (McFadden *et al.*, 1996). This discovery was particularly intriguing as it suggested that apicomplexan parasites evolved from free-living, photosynthetic ancestors. Our understanding of this transition was greatly aided by the subsequent discovery of *Chromera velia*, a photosynthetic relative of apicomplexans that was isolated from the stony coral, *Plesiastrea versipora*, in Sydney harbour (Moore *et al.*, 2008). A second photosynthetic relative of apicomplexans, *Vitrella brassicaformis*, was also discovered from the stony coral, *Leptastrea purpurea*, from the Great Barrier Reef (Oborník *et al.*, 2012). Together, *Chromera* and *Vitrella* are often referred to as 'chromerids', a non-monophyletic group that belongs to a larger clade (chrompodellids) that includes non-photosynthetic predators called colpodellids (Janouškovec *et al.*, 2015).

The ecology of the chromerids remains surprisingly speculative, given they are the first new group of algae to be described in almost 100 years. As they were both isolated from coral, they are widely assumed to be coral symbionts (Moore *et al.* 2008; Okamoto and McFadden, 2009; Oborník *et al.*, 2011), but this has not actually been shown in nature. The genomes, evolutionary history and biochemical pathways of chromerids have been extensively studied (Janouskovec *et al.*, 2010; Gile and Slamovits, 2014; Flegontov *et al.*, 2015; Woo *et al.*, 2015). In contrast, however, their complete sexual life cycle, whether they in fact live within coral, and if so their role within the coral holobiont, have all yet to be determined beyond a few preliminary observations. Cumbo and colleagues (2013) showed that *C. velia* can be transmitted vertically from adult *Montipora digitata* to the coral's egg. And Mohamed *et al.* (2018) showed that the coral's

transcriptomic response to *Chromera* might suggest a parasitic relationship to coral.

These questions became even more complex with the first molecular analyses of their distribution in nature. Traditionally, coral reefs have not been considered a significant habitat for apicomplexans: only one species of coral-infecting apicomplexan has been formally described based on morphology alone, *Gemmocystis cylindrus* (Upton and Peters, 1986), and one clade of environmental sequences has been described as coral-associated, called 'Genotype-N' (Toller *et al*., 2002). Kirk *et al*. (2013a, 2013b) showed some prevalence, seasonal differences and potential transmission strategies of 'Genotype-N'. However, identifying eukaryotic plastid contamination in environmental surveys aimed at bacteria revealed that there is actually a large diversity of plastid sequences related to apicomplexan plastid homologues, which are globally distributed and strongly associated with corals (Janouškovec *et al*., 2012). These Apicomplexan-Related Lineages (ARLs) include sequences related to cultured chromerids, but also include several abundant groups with no characterized members. In particular, ARL-V – the most abundant – is closely related to the parasitic apicomplexans and has been inferred to be a potential photosynthetic coral symbiont (Janouškovec *et al*., 2012; 2013).

The ability to detect plastid rRNA genes in bacterial surveys opens the door to detailed analyses of how eukaryotic algae are distributed in nature (del Campo *et al*., 2017). However, the initial surveys that identified ARL-V were limited in scope because of their use of a few available full-length environmental sequences (Janouškovec *et al*., 2012), as opposed to the short-read database, which has exploded in size since the identification of ARLs. While some association between ARL-V and coral was shown in their study, there were insufficient data for any solid conclusions about the distribution of either of the characterized chromerids. Here, we take advantage of the now-extensive short read survey data to curate a more comprehensive dataset based on 50 bacterial rRNA amplicon surveys comprising 220 million sequences from 224 unique geographical locations. This data emphasizes surveys not only from several environments associated with coral but also includes a wide variety of non-coral environments to test the association of ARLs both with coral environments versus marine environments more generally, and with the coral host versus the coral reef more generally.

Experimental procedures

Amplicon sequence retrieval

Illumina and 454 amplicon sequencing studies that targeted bacterial SSU rRNA genes in coral reefs (coral tissue, mucus and associated seawater and sediments), marine shallow and deep water column, hydrothermal vents, oceanic sediments and freshwater environments, were retrieved from the NCBI Sequence Read Archive (SRA: Leinonen *et al*., 2011) (Agogue *et al*., 2011; Amaral-Zettler *et al*., 2010; Andersson *et al*., 2010; Brazelton *et al*., 2010; Campbell *et al*., 2011; Campbell *et al*., 2015; Dove *et al*., 2013; Eiler *et al*., 2012; Foster *et al*., 2014; Galand *et al*., 2009; Gilbert *et al*., 2012; Glasl *et al*., 2016; Glasl *et al*., 2017; Goodwin *et al*., 2016; Hamdan *et al*., 2013; Herlemann *et al*., 2011; Huber *et al*., 2007; Jorgensen *et al*., 2012; Kellogg *et al*., 2016; Kellogg *et al*., 2017; Kerfahi *et al*., 2014; Kirchman *et al*., 2010; Lee *et al*., 2012; Lema 2014; Li *et al*., 2013; McCliment *et al*., 2012; McNally *et al*., 2017; Meistertzheim *et al*., 2016; Meyer *et al*., 2016; Morrow *et al*., 2015; Ng *et al*., 2015; Pavloudi *et al*., 2016; Pavloudi *et al*., 2017; Rogozin *et al*., 2017; Ruff *et al*., 2014; Shore-Maggio *et al*., 2015; Slapeta and Linares, 2013; Sogin *et al*., 2006; Somboonna *et al*., 2017; Spietz *et al*., 2015; Staley *et al*., 2017; Tripathi *et al*., 2015; Van Bleijswijk *et al*., 2015; Van De Water *et al*., 2015; Vezzulli *et al*., 2013; Walsh *et al*., 2016; Welch and Huse 2011; Zhu *et al*., 2013; Zinger *et al*., 2011) The full list of studies and their corresponding environment type is available in Supporting Information Table 1. The SRA toolkit and fastq-dump utility was used to extract the sequences (settings: -split-files -skip-technical) (Leinonen *et al*., 2011). Paired-end reads were merged using PEAR v0.9.8 (Paired-End Read Merger) (Zhang *et al*., 2014).

Selecting and clustering plastid 16S rRNA gene sequences

Putative apicomplexan related sequences were identified using BLASTn (Camacho *et al*., 2009) with a BLAST database constructed from the aligned environmental sequences from Janouškovec and colleagues (2012) in addition to publicly available plastid SSU rRNA sequences from apicomplexans and chromerids. Sequences were selected on the criteria that they had a minimum 85% sequence similarity to apicomplexan-related lineages (ARLs) and that the length of the alignment covered at least 90% of the length of the query sequence. Selected sequences were then clustered to a 97% threshold using CD-HIT (Li and Godzik, 2006).

Building reference tree and alignment

SSU rRNA sequences of all known ARLs (including *V. brassicaformis* and *C. velia*) were extracted from GenBank and clustered at 97% identity using USEARCH v7.0.1090 (Edgar, 2010). Sequences were aligned using MAFFT auto mode (Katoh and Standley, 2013) using a set of representative sequences as outgroups (apicomplexans, dinoflagellates and colpodellids). Alignments were checked using AliView (Larsson, 2014) and highly variable regions of the alignment were removed using trimAl 1.2 (settings: -gt 0.3 -st 0.001) (Capella-Gutiérrez *et al*., 2009). Maximum likelihood (ML) phylogenetic trees were constructed with RAxML v8 (Stamatakis, 2014; GTR-CAT-I substitution model, 1000 independent tree searches starting from distinct random topology, and 1000 standard non-parametric bootstraps). The reference tree is shown in Supporting Information Figure 2.

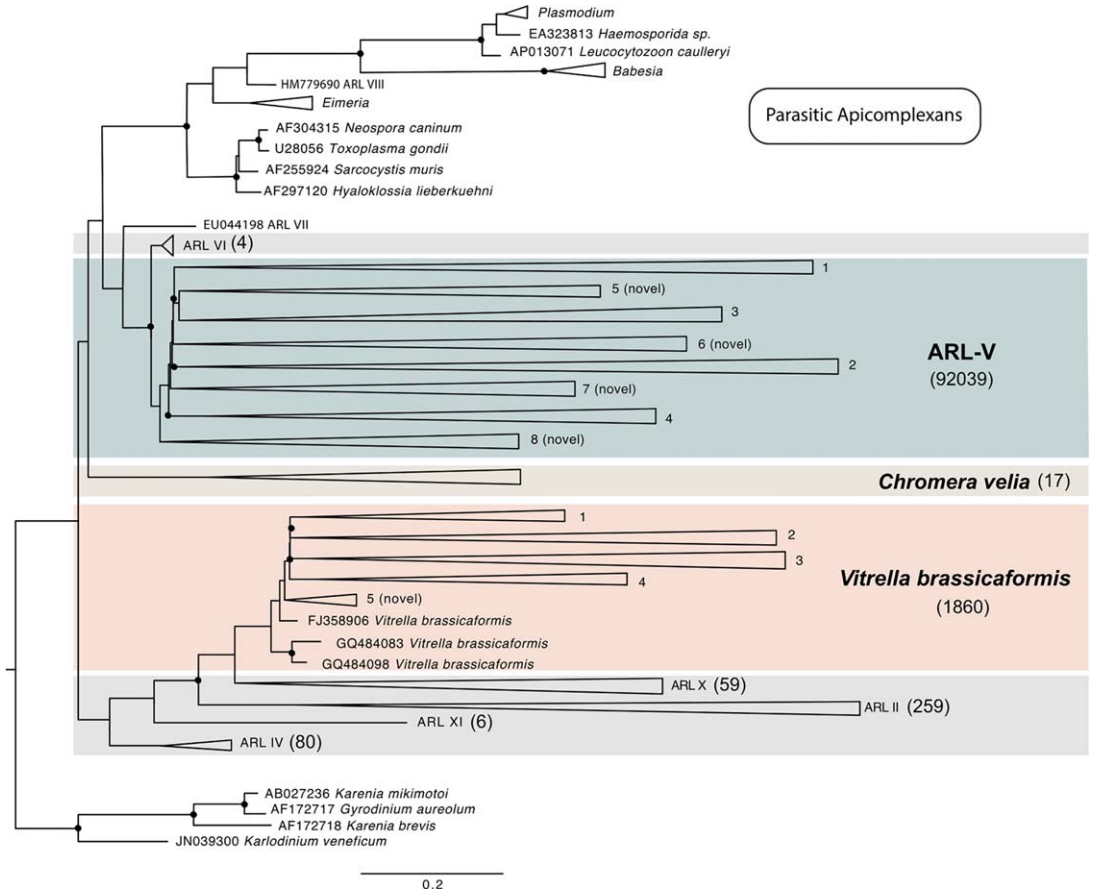

Fig. 1. Maximum likelihood phylogenetic tree inferred from the plastid 16S rRNA gene.
The coloured clades show the recovered genes of Apicomplexan-Related Lineages (ARLs) in this study in relation to parasitic apicomplexans, and a dinoflagellate outgroup. The non-colored taxa are the phylogenetic framework from a reference phylogeny (Supporting Information Figure 1). For each plastid lineage, the total number of reads recovered are shown in parentheses. The tree was constructed using RAxML Evolutionary Placement Algorithm with a reference framework of known 16S rRNA genes of apicomplexans and ARLs. Black circles correspond to nodes with bootstrap support greater than 70%.

Short reads assignment and EPA tree reconstruction

QIIME (Quantitative Insights into Microbial Ecology) v1.4.0 (Caporaso *et al*., 2010) was used to construct a phylogenetic tree to curate the selected reads and discard those that branched within apicomplexans and dinoflagellates. This was done using the following QIIME scripts: (i) align_seqs.py, the BLAST selected sequences and the reference alignment were then merged and aligned and (ii) filter_alignment.py, the merged sequences were filtered (a similar process to trimming) (settings: filter_alignment.py, -g 0.99 -s -e 0.0001). RAxML Evolutionary Placement Algorithm (EPA) (Stamatakis, 2014; model GTR-CAT-I) was used to place the sequences onto the fixed topology of the reference tree. Long branches and sequences that branched within apicomplexans and dinoflagellates were discarded. This process was carried out until only reads that branched within the apicomplexan-related lineages (ARLs) remained, and these curated sequences were then used for all subsequent analyses.

QIIME analysis: operational taxonomic unit (OTU) picking

VSEARCH v2.4.2 (Rognes *et al*., 2016) was used to quality filter and remove any chimeric reads. The sequences were then clustered to 97% identity using USEARCH v8.1.0 (Edgar, 2010). To assign taxonomy to the reads, we used QIIME's OTU (operational taxonomic unit) picking pipeline (Caporaso *et al*., 2010). OTUs were first picked based on sequence similarity within the reads using pick_otus.py. As each OTU was made up of many related sequences, a representative sequence from each OTU was then picked using pick_rep_set.py. These representative sequences were used for taxonomic annotation of the OTU and phylogenetic alignment in downstream analyses. We assigned taxonomy to each OTU representative sequence with assign_taxonomy.py, using the UCLUST consensus taxonomy classifier (Edgar, 2010). Using this annotation, a table of OTU abundances in each sample with taxonomic identifiers for each OTU was constructed. OTUs were further filtered according to their observation counts and OTUs that were observed fewer than

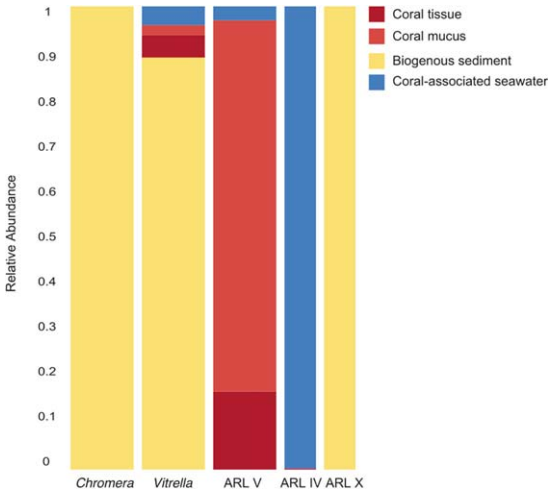

Fig. 2. The proportion of Apicomplexan-Related Lineages (ARLs) recovered from specific coral reef environments (coral tissue, coral mucus, coral-associated seawater and biogenous sediments). The proportion has been normalized to the total number of reads retrieved from that environment. Individual clades of each ARL lineage are shown as separate histogram bars. Shades of red indicate coral tissue and coral mucus. Blue indicates coral-associated seawater and yellow indicates biogenous sediments. Outliers were excluded from this histogram, refer to Supporting Information Figure 3 for outlier analysis.

two times (i.e., singletons) were discarded. The final phylogenetic tree was constructed using these curated sequences. They were assigned positions on to the fixed topology of the reference tree using the EPA algorithm implemented in RAxML (Stamatakis, 2006) assuming the GTR-CAT-I substitution model. New clades were assigned if reads did not branch with any of the reference taxa. Tree figures were edited with FigTree v1.4 (http://tree.bio.ed.ac.uk/software/figtree/) (Fig. 1).

Ecological meta-analyses

A metadata table was collated by manually downloading all the ecological information associated with the curated sequences, such as geographical location, environmental material, host, etc., from the NCBI SRA (Leinonen *et al.*, 2011). This was used to generate a mapping file (table 2 Supporting Information). Using the OTU taxonomy and the associated metadata, QIIME summarize_taxa_through_plots.py was used to identify patterns and differences in the relative abundance of the recovered ARLs in different coral hosts and environments (Figs 2 and 3).

Results and discussion

We retrieved 219 103 555 bacterial SSU sequences from 50 high-throughput amplicon sequencing surveys from the NCBI SRA (Supporting Information table 1). From this dataset, we recovered 94 324 plastid SSU rRNA sequences representing 4072 $OTUs_{97}$ of apicomplexan-related lineages (ARLs). By far the most abundant and diverse lineage retrieved was ARL-V. This lineage was comprised of 92 039 sequences and 3835 $OTUs_{97}$ that formed eight distinct clades, four of which were previously unrecognized

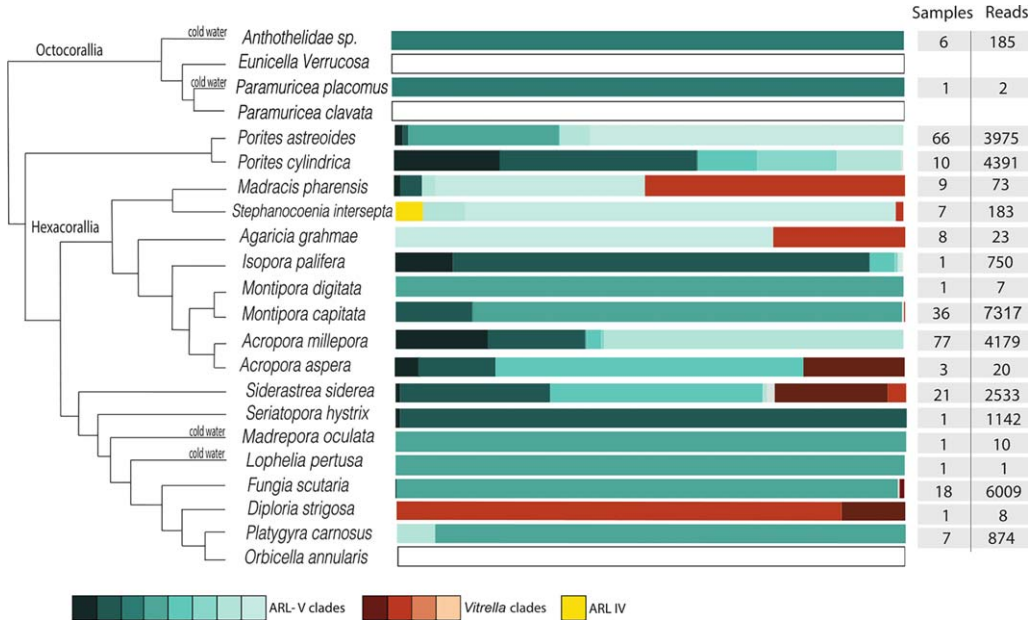

Fig. 3. The absolute abundances of the Apicomplexan-Related Lineages (ARL) sequences recovered in 24 coral species. Cladogram (left) indicates relatedness of coral species (adapted from Fukami *et al.*, 2008 and Kayal *et al.*, 2013). Table (right) shows the number of coral samples and corresponding number of reads that were retrieved from those samples. The histogram shows absolute abundances of ARL reads. Shades of blue indicate ARL-V clades, from ARLV-1 (darkest) to ARLV-8 (lightest). Shades of red indicate *Vitrella* clades, from Vit-1 (darkest) to Vit-4 (lightest). Yellow indicates ARL-IV. Clear histogram bars indicate no ARLs reads retrieved from that species.

(Fig. 1). The next most abundant group was *Vitrella* and its close relatives, where 1860 sequences and 210 $OTUs_{97}$ forming four distinct clades were recovered. Other ARLs were in much lower abundance (Fig. 1), including *Chromera*, where only 17 reads representing 7 $OTUS_{97}$ were identified and ARL-II with 259 reads and 5 $OTUs_{97}$. In addition, two new lineages were also identified, ARL-X and ARL-XI, with ARL-X branching as a close sister to *Vitrella* (Fig. 1).

Distribution of apicomplexan-related lineages confirms association with coral reefs

It has been proposed that ARLs are specifically coral reef associated (Janouškovec *et al*., 2012). However, these conclusions were based on only 121 ARL sequences. We tested this association with this larger dataset and with numerous environmental outgroups (70% of sequences in our dataset were from coral reef studies, and 30% consist of surveys of other marine and freshwater environments). Our findings confirm that the link between the ARLs and coral reefs is robust. We analysed samples from 224 unique geographical sampling sites worldwide and ARLs were found in only 29 of these locations, all which correspond to coral reefs (Supporting Information Figure 1). One exceptional result was the presence of ARL-V reads in landlocked lakes in central Asia (Baatar *et al*., 2016). However, we noted that coral bacterial communities were being analysed by the same authors at a similar time, and we conclude that these low-abundance reads most likely represent cross-contaminants (Lee *et al*., 2016).

Chromera and Vitrella are not coral symbionts

As its isolation from coral using methods used to extract *Symbiodinium*, *Chromera* has been widely implied and assumed to be a coral symbiont (Moore *et al*., 2008; Cumbo *et al*., 2013). However, this has never been shown directly and, according to a recent publication, *Chromera* is more likely an opportunistic parasite than a mutualistic symbiont (Mohamed *et al*. 2018). Here, we show that *Chromera* is relatively rare in coral reefs worldwide, and, interestingly, is also never recovered from the actual coral: *Chromera* reads were only detected in biogenous sediments surrounding corals, and never from coral tissue itself (Fig. 2). A possible limitation is the absence of survey data from the coral species where *Chromera* was originally isolated, *Plesiastrea versipora* and *Leptastrea purpurea*. However, given how strongly ARL-V sequences are associated with specifically coral mucus and tissue (see below), it seems likely that if *Chromera* were indeed a symbiont, a similarly tight association to the coral host should be detected. Therefore, we suggest that *Chromera* is not a coral symbiont per se. Instead, *Chromera* may be a symbiont of other reef-dwellers that have yet to be sampled, or a free-living autotroph in the reef biogenous sediments. Its association with the reef environment could be due to the distribution of another invertebrate host, or perhaps the *Chromera* life cycle includes a transient stage within coral that has not been observed but ties it to the reef environment. Sediments are known to harbour a diverse array of primary producers, including cyanobacteria and photosynthetic protists (e.g., dinoflagellates and diatoms) and have been reported to have a largely uncharacterized bacterial SSU diversity (Werner *et al*., 2008). However, their role in the reef ecosystem is still comparatively understudied. Greater observation of *Chromera* in nature is needed to understand its basic biology, which may help illuminate the ecological conditions driving the evolution of apicomplexans. *Chromera* may well be tied to the microbial processes in biogenous sediments and their contribution to reef primary production, nutrient cycling and maintenance of overall coral reef health.

Interestingly, *Vitrella*-related groups are not only more widely distributed, as sequences were recovered predominantly from the biogenous sediments, like *Chromera*, but also found at lower abundances in coral tissue, mucus and in coral-associated seawater (Fig. 2). *Vitrella* has been shown to have both photosynthetic and predatory stages, as well as a large sporangium stage that has not been seen in *Chromera* (Oborník *et al*., 2012). The more varied distribution we observe might reflect a more complex life cycle with different trophic stages that take advantage of different parts of the coral reef. Once again, this emphasizes the importance of elucidating the complete life cycle of *Vitrella* in nature to provide some insights into its roles on the reef, and also in the evolution of parasitism in apicomplexans.

Refining ARL-V and its association with coral

Apicomplexan Related Lineage-V is the most abundant ARL, but remains an environmental clade described solely from plastid SSU gene sequences. Based on the initial 121 environmental sequences, phylogenetic analyses placed ARL-V in an intermediate position between the photosynthetic chromerids and parasitic apicomplexans and showed a very tight association with coral reefs (Janouškovec *et al*., 2012). A follow-up survey showed ARL-V to be enriched in healthy coral tissue in shallow reefs, suggestive of a photosynthetic symbiotic relationship with coral (Janouškovec *et al*., 2013). Here, we retrieved 92 039 sequences of ARL-V, which allows us to test inferences about the distribution and ecology of this enigmatic group with greater confidence. First, the new dataset clearly confirms that ARL-V is indeed exclusively coral reef-associated and globally distributed in reefs worldwide. The relative abundance of ARL-V compared with all other ARLs is also clearly confirmed:

ARL-V sequences were more than 50-fold more abundant than any other ARL sequences. Lastly, and in striking contrast to chromerids, we confirmed that ARL-V is strongly associated with the coral host directly (in tissue and mucus); it is completely absent in the sediments, and only a few sequences were found in seawater surrounding coral (Fig. 2). In addition to confirming its abundance, the larger dataset has also confirmed and extended the high level of diversity of ARL-V. We not only identified eight distinct phylogenetic subgroups of ARL-V but also 3835 $OTUs_{97}$. Altogether, ARL-V represents several distinct, species-rich lineages, suggesting it may have a high degree of biological diversity as well.

To examine the host range of ARL-V, we searched sequencing surveys of 22 distinct coral species, out of which ARLs were present in 19 species (Fig. 3). ARL-V was the most widespread ARL retrieved and was found in every coral host that was positive for any ARL. In addition, ARL-V was retrieved in highest abundance compared with any other ARL in all individual coral species, except for *Diploria strigosa* (Fig. 3). Interestingly, the coral species positive for ARL-V do not form a monophyletic group, but rather included a diverse species of both hexacorals and octacorals (soft corals). In contrast to Janouškovec and colleagues (2012), we also found that ARL-V is not depth-dependent: ARL-V was retrieved from corals at depths of 522m as well as in shallow corals at 1m, as well as a range of depths between these extremes (Supporting Information table 2). Moreover, ARL-V was found in four species of deep, cold-water azooxanthellate corals, indicating no obligatory co-occurrence between ARL-V and *Symbiodinium*.

Based on these findings, we propose that the distribution of ARL-V is more consistent with an intracellular parasite or commensal, rather than a photosynthetic symbiont, or that at least some fraction of ARL-V diversity is non-photosynthetic. Incomplete metadata did not allow any comprehensive correlation with coral health, but ARL-V was retrieved from samples of both healthy and diseased coral tissue (Supporting Information table 2), and ARL-V has previously been found in healthy corals (Janouškovec *et al*., 2013). This suggests that ARL-V collectively does not cause any one distinctive coral disease. The small number of sequences found in seawater surrounding coral (Fig. 2) might indicate the presence of infective stages, however, it is also possible these are dead or dying cells released from coral, and direct observation of ARL-V and its biology will be needed to elucidate the exact nature of its interaction with coral. This may be problematic given the diversity of ARL-V (Fig. 1), which might reflect a high degree of host-specificity or complex population structure within each coral host species, the details of which will require careful sampling of many individual coral hosts.

Overall, ARL-V is the most abundant apicomplexan-related lineage, one that is both phylogenetically and ecologically distinct from the chromerids. While initially speculated to be a photosynthetic symbiont, here we show its distribution to be more indicative of a non-photosynthetic coral parasite or commensal. This bolsters the idea that the ARL-V clade might correspond to the coral apicomplexan Genotype-N (Toller *et al*., 2002). Genotype-N was discovered based on the nuclear 18S SSU gene, whereas the ARL-V is based on plastid 16S SSU gene. This incompatibility in molecular data prevents us from making any conclusive decision as to whether these organisms are the same (in particular as relatively few nuclear SSU rRNA surveys of coral reefs have been conducted). Another possibility is that ARL-V corresponds to *Gemmocystis cylindrus*, a coccidian has only been identified by morphology (Upton and Peters, 1986). However, once again the data are impossible to compare as neither ARL-V nor Genotype-N cells have knowingly been observed. The relationship between these three entities is a question that remains to be tested directly.

Conclusions

The association between coral and its archetypal dinoflagellate symbiont, *Symbiodinium*, is very well-studied (Yellowlees *et al*., 2008; Thornhill *et al*., 2017), yet relatively little is known about the interactions between coral and other microbial eukaryotes (del Campo *et al*. 2017; Ainsworth *et al*., 2017). The *Symbiodinium* model looms large in our thinking about coral-protist interactions, so it is perhaps not surprising that chromerids were assumed to be photosynthetic symbionts of a similar sort when first discovered. This also might have played a role in the surprisingly sparse attention paid to their ecological role in nature compared with their biochemistry and evolution. Our analyses suggest, however, that these interactions are both complex, not what we expected, and in need of direct observations in nature. *Chromera* and *Vitrella* are indeed coral reef associated but live within the biogenous sediments and not with coral itself, suggesting a role in reef sediment primary production and some unclear link to the reef ecosystem that is subtler than simply living inside coral cells. We also confirmed ARL-V to be the most abundant, phylogenetically diverse ARL lineage and to associate almost exclusively with coral mucus and tissue of a very broad range of coral hosts. But the environmental distribution of ARL-V is not consistent with a photosynthesis-based symbiosis either, suggesting instead parasitism or commensalism with coral that also requires direct observation. These organisms have attracted a great deal of attention because they are related to an important and evolutionarily interesting group of parasites, but evidence is mounting that the ecology of apicomplexan related lineages is not what we anticipated and observing them in nature with specific data on their interactions with

coral is a next, critical step to understanding both these organisms and their contribution to coral reefs.

Acknowledgements

We would like to thank Racquel Singh and Nick Irwin for their valuable help on the data analyses. This work was supported by a grant from the Canadian Institute for Health Research to PJK (MOP 42517). JdC and MK were supported by a grant to the Centre for Microbial Diversity and Evolution from the Tula Foundation and JdC was supported by a Marie Curie Outgoing Fellowship grant (FP7-PEOPLE-2012-IOF – 331450 CAARL).

Conflict of Interest

The authors declare that they have no conflict of interest.

References

Agogue, H., Lamy, D., Neal, P.R., Sogin, M.L., and Herndl, G.J. (2011) Water mass-specificity of bacterial communities in the North Atlantic revealed by massively parallel sequencing. *Mol Ecol* **20:** 258–274.

Ainsworth, T.D., Fordyce, A.J., and Camp, E.F. (2017) The other microeukaryotes of the coral reef microbiome. *Trends Microbiol* **25:** 980–991.

Amaral-Zettler, L., Artigas, L.F., Baross, J., Bharathi, P.A.L., Boetius, A., and Chandramohan, D. (2010). A global census of marine microbes. In *Life in the World's Oceans: Diversity, Distribution, and Abundance*. Oxford, UK: Wiley-Blackwell, pp. 221–245.

Andersson, A.F., Riemann, L., and Bertilsson, S. (2010) Pyrosequencing reveals contrasting seasonal dynamics of taxa within Baltic Sea bacterioplankton communities. *ISME J* **4:** 171–181.

Baatar, B., Chiang, P.W., Rogozin, D.Y., Wu, Y.T., Tseng, C.H., Yang, C.Y., *et al*. (2016) Bacterial communities of three saline meromictic lakes in Central Asia Chiang T-Y (ed). *PLoS One* **11:** e0150847.

Brazelton, W.J., Ludwig, K.A., Sogin, M.L., Andreishcheva, E.N., Kelley, D.S., Shen, C.-C., *et al*. (2010) Archaea and bacteria with surprising microdiversity show shifts in dominance over 1,000-year time scales in hydrothermal chimneys. *Proc Natl Acad Sci USA* **107:** 1612–1617.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., *et al*. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10:** 421.

Campbell, A.M., Fleisher, J., Sinigalliano, C., White, J.R., and Lopez, J.V. (2015) Dynamics of marine bacterial community diversity of the coastal waters of the reefs, inlets, and wastewater outfalls of Southeast Florida. *Microbiologyopen* **4:** 390–408.

Campbell, B.J., Yu, L., Heidelberg, J.F., and Kirchman, D.L. (2011) Activity of abundant and rare bacteria in a coastal ocean. *Proc Natl Acad Sci USA* **108:** 12776–12781.

Capella-Gutierrez, S., Silla-Martinez, J.M., and Gabaldon, T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25:** 1972–1973.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., *et al*. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7:** 335–336.

Cumbo, V.R., Baird, A.H., Moore, R.B., Negri, A.P., Neilan, B.A., Salih, A., *et al*. (2013) Chromera velia is endosymbiotic in larvae of the reef corals *Acropora digitifera* and *A. tenuis*. *Protist* **164:** 237–244.

del Campo, J., Pombert, J.-F., Šlapeta, J., Larkum, A., and Keeling, P.J. (2017) The 'other' coral symbiont: ostreobium diversity and distribution. *ISME J* **11:** 296–299.

Dove, S.G., Kline, D.I., Pantos, O., Angly, F.E., Tyson, G.W., and Hoegh-Guldberg, O. (2013) Future reef decalcification under a business-as-usual CO2 emission scenario. *Proc Natl Acad Sci USA* **110:** 15342–15347.

Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26:** 2460–2461.

Eiler, A., Heinrich, F., and Bertilsson, S. (2012) Coherent dynamics and association networks among lake bacterioplankton taxa. *ISME J* **6:** 330–342.

Flegontov, P., Michálek, J., Janouškovec, J., Lai, D.H., Jirků, M., Hajdušková, E., *et al*. (2015) Divergent mitochondrial respiratory chains in phototrophic relatives of apicomplexan parasites. *Mol Biol Evol* **32:** 1115–1131.

Foster, C., Portman, N., Chen, M., and Šlapeta, J. (2014) Increased growth and pigment content of Chromera velia in mixotrophic culture. *FEMS Microbiol Ecol* **88:** 121–128.

Fukami, H., Chen, C.A., Budd, A.F., Collins, A., Wallace, C., Chuang, Y.Y., *et al*. (2008) Mitochondrial and nuclear genes suggest that stony corals are monophyletic but most families of stony corals are not (Order Scleractinia, Class anthozoa, phylum cnidaria). *PLoS One* **3:** e3222.

Galand, P.E., Casamayor, E.O., Kirchman, D.L., and Lovejoy, C. (2009) Ecology of the rare microbial biosphere of the Arctic Ocean. *Proc Natl Acad Sci USA* **106:** 22427–22432.

Gilbert, J.A., Steele, J.A., Caporaso, J.G., Steinbrück, L., Reeder, J., Temperton, B., *et al*. (2012) Defining seasonal marine microbial community dynamics. *ISME J* **6:** 298–308.

Gile, G.H., and Slamovits, C.H. (2014) Transcriptomic analysis reveals evidence for a cryptic plastid in the colpodellid Voromonas pontica, a close relative of chromerids and apicomplexan parasites Waller RF (ed). *PLoS One* **9:** e96258.

Glasl, B., Bongaerts, P., Elisabeth, N.H., Hoegh-Guldberg, O., Herndl, G.J., and Frade, P.R. (2017) Microbiome variation in corals with distinct depth distribution ranges across a shallow–mesophotic gradient (15–85 m). *Coral Reefs* **36:** 447–452.

Glasl, B., Herndl, G.J., and Frade, P.R. (2016) The microbiome of coral surface mucus has a key role in mediating holobiont health and survival upon disturbance. *ISME J* **10:** 2280–2292.

Goodwin, S., McPherson, J.D., and McCombie, W.R. (2016) Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* **17:** 333–351.

Hamdan, L.J., Coffin, R.B., Sikaroodi, M., Greinert, J., Treude, T., and Gillevet, P.M. (2013) Ocean currents shape the microbiome of Arctic marine sediments. *ISME J* **7:** 685–696.

Herlemann, D.P.R., Labrenz, M., Jürgens, K., Bertilsson, S., Waniek, J.J., and Andersson, A.F. (2011) Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J* **5:** 1571–1579.

Huber, J.A., Mark Welch, D.B., Morrison, H.G., Huse, S.M., Neal, P.R., Butterfield, D.A., *et al*. (2007) Microbial population structures in the deep marine biosphere. *Science (80-)* **318:** 97–100.

Janouškovec, J., Horák, A., Barott, K.L., Rohwer, F.L., and Keeling, P.J. (2012) Global analysis of plastid diversity reveals apicomplexan-related lineages in coral reefs. *Curr Biol* **22:** R518–R519.

Janouškovec, J., Horák, A., Barott, K.L., Rohwer, F.L., and Keeling, P.J. (2013) Environmental distribution of coral-associated relatives of apicomplexan parasites. *ISME J* **7:** 444–447.

Janouskovec, J., Horak, A., Obornik, 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.

Janouškovec, J., Tikhonenkov, D.V., Burki, F., Howe, A.T., Kolísko, M., Mylnikov, A.P., *et al*. (2015) Factors mediating plastid dependency and the origins of parasitism in apicomplexans and their close relatives. *Proc Natl Acad Sci USA* **112:** 10200–10207.

Jorgensen, S.L., Hannisdal, B., Lanzen, A., Baumberger, T., Flesland, K., Fonseca, R., *et al*. (2012) Correlating microbial community profiles with geochemical data in highly stratified sediments from the Arctic Mid-Ocean Ridge. *Proc Natl Acad Sci USA* **109:** E2846–E2855.

Katoh, K., and Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30:** 772–780.

Kayal, E., Roure, B., Philippe, H., Collins, A.G., and Lavrov, D.V. (2013) Cnidarian phylogenetic relationships as revealed by mitogenomics. *BMC Evol Biol* **13:** 5.

Kellogg, C.A., Goldsmith, D.B., and Gray, M.A. (2017) Biogeographic comparison of Lophelia-associated bacterial communities in the Western Atlantic reveals conserved core microbiome. *Front Microbiol* **8:** 796.

Kellogg, C.A., Ross, S.W., and Brooke, S.D. (2016) Bacterial community diversity of the deep-sea octocoral *Paramuricea placomus*. *PeerJ* **4:** e2529.

Kerfahi, D., Hall-Spencer, J.M., Tripathi, B.M., Milazzo, M., Lee, J., and Adams, J.M. (2014) Shallow water marine sediment bacterial community shifts along a natural CO2 gradient in the Mediterranean Sea off Vulcano, Italy. *Microb Ecol* **67:** 819–828.

Kirchman, D.L., Cottrell, M.T., and Lovejoy, C. (2010) The structure of bacterial communities in the Western Arctic Ocean as revealed by pyrosequencing of 16S rRNA genes. *Environ Microbiol* **12:** 1132–1143.

Kirk, N.L., Ritson-Williams, R., Coffroth, M.A., Miller, M.W., Fogarty, N.D., and Santos, S.R. (2013a) Tracking transmission of apicomplexan symbionts in diverse Caribbean corals. *PLoS One* **8:** e80618.

Kirk, N.L., Thornhill, D.J., Kemp, D.W., Fitt, W.K., and Santos, S.R. (2013b) Ubiquitous associations and a peak fall prevalence between apicomplexan symbionts and reef corals in Florida and the Bahamas. *Coral Reefs* **32:** 847–858.

Larsson, A. (2014) AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30:** 3276–3278.

Lee, S.T.M., Davy, S.K., Tang, S.L., and Kench, P.S. (2016) Mucus sugar content shapes the bacterial community structure in thermally stressed *Acropora muricata*. *Front Microbiol* **7:** 371.

Lee, O.O., Yang, J., Bougouffa, S., Wang, Y., Batang, Z., Tian, R., *et al*. (2012) Spatial and species variations in bacterial communities associated with corals from the Red Sea as revealed by pyrosequencing. *Appl Environ Microbiol* **78:** 7173–7184.

Leinonen, R., Sugawara, H., and Shumway, M. (2011) The sequence read archive. *Nucleic Acids Res* **39:** D19–D21.

Lema, A.K. (2014). Diversity, stability, and uptake of diazotrophic bacterial communities associated with corals of the Great Barrier Reef. PhD Thesis. Queensland, Australia: James Cook University. [WWW document]. URL https://researchonline.jcu.edu.au/39865/

Li, W., and Godzik, A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22:** 1658–1659.

Li, J., Chen, Q., Zhang, S., Huang, H., Yang, J., Tian, X.P., *et al*. (2013) Highly heterogeneous bacterial communities associated with the South China Sea reef corals *Porites lutea*, *Galaxea fascicularis* and *Acropora millepora* Schuch R (ed). *PLoS One* **8:** e71301.

McCliment, E.A., Nelson, C.E., Carlson, C.A., Alldredge, A.L., Witting, J., and Amaral-Zettler, L.A. (2012) An all-taxon microbial inventory of the Moorea coral reef ecosystem. *ISME J* **6:** 309–319.

McFadden, G.I., Reith, M.E., Munholland, J., and Lang-Unnasch, N. (1996) Plastid in human parasites. *Nature* **381:** 482.

McNally, S.P., Parsons, R.J., Santoro, A.E., and Apprill, A. (2017) Multifaceted impacts of the stony coral Porites astreoides on picoplankton abundance and community composition. *Limnol Oceanogr* **62:** 217–234.

Meistertzheim, A.L., Lartaud, F., Arnaud-Haond, S., Kalenitchenko, D., Bessalam, M., Le Bris, N., *et al*. (2016) Patterns of bacteria-host associations suggest different ecological strategies between two reef building cold-water coral species. *Deep Res Part I Oceanogr Res Pap* **114:** 12–22.

Meyer, J.L., Rodgers, J.M., Dillard, B.A., Paul, V.J., and Teplitski, M. (2016) Epimicrobiota associated with the decay and recovery of Orbicella corals exhibiting dark spot syndrome. *Front Microbiol* **7:** 893.

Mohamed, A.R., Cumbo, V.R., Harii, S., Shinzato, C., Chan, C.X., Ragan, M.A., *et al*. (2018) Deciphering the nature of the coral–Chromera association. *ISME J* **12:** 776–790.

Moore, R.B., Oborník, M., Janouškovec, J., Chrudimský, T., Vancová, M., Green, D.H., *et al*. (2008) A photosynthetic alveolate closely related to apicomplexan parasites. *Nature* **451:** 959–963.

Morrow, K.M., Bourne, D.G., Humphrey, C., Botté, E.S., Laffy, P., Zaneveld, J., *et al*. (2015) Natural volcanic CO2seeps reveal future trajectories for host-microbial associations in corals and sponges. *ISME J* **9:** 894–908.

Ng, J.C.Y., Chan, Y., Tun, H.M., Leung, F.C.C., Shin, P.K.S., and Chiu, J.M.Y. (2015) Pyrosequencing of the bacteria associated with Platygyra carnosus corals with skeletal growth anomalies reveals differences in bacterial community composition in apparently healthy and diseased tissues. *Front Microbiol* **6:** 1142.

Oborník, M., Modrý, D., Lukeš, M., Černotíková-Stříbrná, E., Cihlář, J., Tesařová, M., *et al*. (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.

Oborník, M., Vancová, M., Lai, D.H., Janouškovec, J., Keeling, P.J., and Lukeš, J. (2011) Morphology and ultrastructure of multiple life cycle stages of the photosynthetic relative of apicomplexa, *Chromera velia*. *Protist* **162:** 115–130.

Okamoto, N., and McFadden, G.I. (2008) The mother of all parasites. *Future Microbiol* **3:** 391–395.

Pavloudi, C., Oulas, A., Vasileiadou, K., Kotoulas, G., De Troch, M., Friedrich, M.W., *et al*. (2017) Diversity and abundance of sulfate-reducing microorganisms in a Mediterranean lagoonal complex (Amvrakikos Gulf, Ionian Sea) derived from dsrB gene. *Aquat Microb Ecol* **79:** 209–219.

Pavloudi, C., Oulas, A., Vasileiadou, K., Sarropoulou, E., Kotoulas, G., and Arvanitidis, C. (2016) Salinity is the major factor influencing the sediment bacterial communities in a Mediterranean lagoonal complex (Amvrakikos Gulf, Ionian Sea). *Mar Genomics* **28:** 71–81.

Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016) VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4:** e2584.

Rogozin, D., Zadereev, E., Prokopkin, I., Tolomeev, A., Barkhatov, Y., Khromechek, E., *et al*. (2017). *Comparative study of the stability of stratification and the food web structure in the Meromictic Lakes Shira and Shunet (South Siberia, Russia)*. Cham, Switzerland: Springer, pp. 89–124.

Ruff, S.E., Probandt, D., Zinkann, A.C., Iversen, M.H., Klaas, C., Würzberg, L., *et al*. (2014) Indications for algae-degrading benthic microbial communities in deep-sea sediments along the Antarctic Polar Front. *Deep Res Part II Top Stud Oceanogr* **108:** 6–16.

Shore-Maggio, A., Runyon, C.M., Ushijima, B., Aeby, G.S., and Callahan, S.M. (2015) Differences in bacterial community structure in two color morphs of the Hawaiian reef coral Montipora capitata. *Appl Environ Microbiol* **81:** 7312–7318.

Šlapeta, J., and Linares, M.C. (2013) 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.

Sogin, M.L., Morrison, H.G., Huber, J.A., Mark Welch, D., Huse, S.M., Neal, P.R., *et al*. (2006) Microbial diversity in the deep sea and the underexplored 'rare biosphere'. *Proc Natl Acad Sci USA* **103:** 12115–12120.

Somboonna, N., Wilantho, A., Monanunsap, S., Chavanich, S., Tangphatsornruang, S., and Tongsima, S. (2017) Microbial communities in the reef water at Kham Island, lower Gulf of Thailand. *PeerJ* **5:** e3625.

Spietz, R.L., Williams, C.M., Rocap, G., and Horner-Devine, M.C. (2015) A dissolved oxygen threshold for shifts in bacterial community structure in a seasonally hypoxic estuary Vopel KC (ed). *PLoS One* **10:** e0135731.

Staley, C., Kaiser, T., Gidley, M.L., Enochs, I.C., Jones, P.R., Goodwin, K.D., *et al*. (2017) Differential impacts of land-based sources of pollution on the microbiota of Southeast Florida coral reefs. *Appl Environ Microbiol* **83:** e03378-16.

Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30:** 1312–1313.

Thornhill, D.J., Howells, E.J., Wham, D.C., Steury, T.D., and Santos, S.R. (2017) Population genetics of reef coral endosymbionts (Symbiodinium, Dinophyceae). *Mol Ecol* **26:** 2640–2659.

Toller, W.W., Rowan, R., and Knowlton, N. (2001) Repopulation of zooxanthellae in the Caribbean corals *Montastraea annularis* and *M. faveolata* following experimental and disease-associated bleaching. *Biol Bull* **201:** 360–373.

Toller, W.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.

Tripathi, B.M., Kim, M., Tateno, R., Kim, W., Wang, J., Lai-Hoe, A., *et al*. (2015) Soil pH and biome are both key determinants of soil archaeal community structure. *Soil Biol Biochem* **88:** 1–8.

Upton, S.J., and Peters, E.C. (1986) A new and unusual species of coccidium (Apicomplexa: Agamococcidiorida) from Caribbean scleractinian corals. *J Invertebr Pathol* **47:** 184–193.

Van Bleijswijk, J.D.L., Whalen, C., Duineveld, G.C.A., Lavaleye, M.S.S., Witte, H.J., and Mienis, F. (2015) Microbial assemblages on a cold-water coral mound at the SE Rockall Bank (NE Atlantic): interactions with hydrography and topography. *Biogeosciences* **12:** 4483–4496.

Van De Water, J.A.J.M., Ainsworth, T.D., Leggat, W., Bourne, D.G., Willis, B.L., and Van Oppen, M.J.H. (2015) The coral immune response facilitates protection against microbes during tissue regeneration. *Mol Ecol* **24:** 3390–3404.

Vezzulli, L., Pezzati, E., Huete-Stauffer, C., Pruzzo, C., and Cerrano, C. (2013) 16SrDNA pyrosequencing of the Mediterranean Gorgonian *Paramuricea clavata* Reveals a link among alterations in bacterial holobiont members, anthropogenic influence and disease outbreaks Medina M (ed). *PLoS One* **8:** e67745.

Walsh, E.A., Kirkpatrick, J.B., Rutherford, S.D., Smith, D.C., Sogin, M., and D'Hondt, S. (2016) Bacterial diversity and community composition from seasurface to subseafloor. *ISME J* **10:** 979–989.

Welch, D.B.M., and Huse, S.M. (2011) Microbial diversity in the deep sea and the underexplored 'rare biosphere'. *Handb Mol Microb Ecol II Metagenomics Differ Habitats* **103:** 243–252.

Werner, U., Blazejak, A., Bird, P., Eickert, G., Schoon, R., Abed, R.M.M., *et al*. (2008) Microbial photosynthesis in coral reef sediments (Heron Reef, Australia). *Estuar Coast Shelf Sci* **76:** 876–888.

Woo, Y.H., Ansari, H., Otto, T.D., Klinger, C.M., Kolisko, M., Michálek, J., *et al*. (2015) Chromerid genomes reveal the evolutionary path from photosynthetic algae to obligate intracellular parasites. *Elife* **4:** 1–41.

Yellowlees, D., Rees, T.A.V., and Leggat, W. (2008) Metabolic interactions between algal symbionts and invertebrate hosts. *Plant Cell Environ* **31:** 679–694.

Zhang, J., Kobert, K., Flouri, T., and Stamatakis, A. (2014) PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30:** 614–20.

Zhu, D., Tanabe, S.-H., Yang, C., Zhang, W., and Sun, J. (2013) Bacterial community composition of South China Sea sediments through pyrosequencing-based analysis of 16S rRNA genes. *PLoS One* **8:** e78501.

Zinger, L., Amaral-Zettler, L.A., Fuhrman, J.A., Horner-Devine, M.C., Huse, S.M., Welch, D.B.M., *et al.* (2011) Global patterns of bacterial beta-diversity in seafloor and seawater ecosystems Gilbert JA (ed). *PLoS One* **6:** e24570.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. 1. Map distribution
Geographical distribution of all samples. Blue indicates presence of apicomplexan-related lineages (ARLs), red indicates absence of ARLs, and purple indicates an overlap of presence and absence in samples from the same site.
Fig. 2. Reference Phylogeny
Reference maximum likelihood (ML) tree of plastid SSU rRNA gene. The tree includes sequences from all known Apicomplexan-Related Lineages with representative sequences of apicomplexans and dinoflagellate outgroups from GenBank. The tree was constructed using RAxML and the GTR-CAT-I substitution model and 1000 independent tree searches. Bootstrap percentages are shown on nodes.
Fig. 3. Outlier Analysis
Outliers removed from distribution analysis (figure 2) are shown in red for each environment.
Each point reflects the log transformed proportion of reads in each sample that correspond to ARL-V. Outliers were defined as points falling outside 1.5 times the interquartile range.
Table 1. List of Studies
List of Illumina and 454 amplicon sequencing surveys analysed. All studies sequenced bacterial SSU rRNA genes and were retrieved from the NCBI Sequence Read Archive. The title of the study, SRA accession number, general environment type (coral reef, freshwater, seawater, sediment) and corresponding reference are shown.
Table 2. Mapping File
Metadata table associated with samples positive for Apicomplexan-Related Lineages (ARLs). The table includes geographical location and coordinates, depth, environment, material, coral species, coral health status, coral disease and life stage.
Table 3. Operational Taxonomic Unit (OTU) Table
Table showing the number of sequences in each OTU, for each sample, and the taxonomy of that OTU.