Genomic adaptations to an endolithic lifestyle in the coral-associated alga *Ostreobium*

Cintia Iha[a*], Javier A. Varela[b], Viridiana Avila[c], Christopher J. Jackson[a], Kenny A. Bogaert[d], Yibi Chen[e], Louise M. Judd[f], Ryan Wick[f], Kathryn E. Holt[f,g], Marisa M. Pasella[a], Francesco Ricci[a], Sonja I. Repetti[a], Mónica Medina[c], Vanessa R. Marcelino[h], Cheong X. Chan[e], Heroen Verbruggen[a*]

[a]School of BioSciences, University of Melbourne, Victoria 3010, Australia

[b]School of Microbiology/Centre for Synthetic Biology and Biotechnology/Environmental Research Institute/APC Microbiome Institute, University College Cork, Cork T12 YN60, Ireland

[c]Pennsylvania State University, University Park, PA, 16802, USA

[d]Phycology Research Group, Ghent University, Krijgslaan 281 S8, 9000 Gent, Belgium

[e]School of Chemistry and Molecular Biosciences and Australian Centre for Ecogenomics, The University of Queensland, Brisbane, Queensland 4072, Australia

[f]Department of Infectious Diseases, Central Clinical School, Monash University, Melbourne, 3004, Australia

[g]London School of Hygiene & Tropical Medicine, London, WC1E 7HT, UK

[h]Centre for Innate Immunity and Infectious Diseases, Hudson Institute of Medical Research, Clayton 3168, Victoria, Australia

*corresponding author: cintiaiha@gmail.com, heroen@unimelb.edu.au

Abstract

The green alga *Ostreobium* is an important coral holobiont member, playing key roles in skeletal decalcification and providing photosynthate to bleached corals that have lost their dinoflagellate endosymbionts. *Ostreobium* lives in the coral's skeleton, a low-light environment with variable pH and O_2 availability. We present the *Ostreobium* nuclear genome and a metatranscriptomic analysis of healthy and bleached corals to improve our understanding of *Ostreobium*'s adaptations to its extreme environment and its roles as a coral holobiont member. The *Ostreobium* genome has 11,735 predicted protein-coding genes and shows adaptations for life in low and variable light conditions and other stressors in the endolithic environment. It presents an unusually rich repertoire of light harvesting complex proteins, lacks many genes for photoprotection and photoreceptors, and has a large arsenal of genes for oxidative stress response. An expansion of extracellular peptidases suggests that *Ostreobium* may supplement its energy needs and compensate for its lysine auxotrophy by feeding on the organic skeletal matrix, and a diverse set of fermentation pathways allow it to live in the anoxic skeleton at night. *Ostreobium* depends on other holobiont members for biotin and vitamin B12, and our metatranscriptomes identify potential bacterial sources. Metatranscriptomes showed *Ostreobium* becoming a dominant agent of photosynthesis in bleached corals and provided evidence for variable responses among coral samples and different *Ostreobium* genotypes. Our work provides a comprehensive understanding of the adaptations of *Ostreobium* to its extreme environment and an important genomic resource to improve our comprehension of coral holobiont resilience, bleaching and recovery.

Introduction

Coral health depends on the harmonious association between the coral animal and its microbial associates, together known as the holobiont. A wealth of studies shows that climate change threatens coral health by disrupting the association between the coral and its

photosynthetic dinoflagellate endosymbionts (Symbiodiniaceae), culminating in coral bleaching and death. The role of other microbes in coral resilience is just starting to be understood, and our current knowledge is largely based on correlations between coral health status and the presence of microbial taxa inferred from metabarcoding. Whole-genome sequences of corals and their symbionts are an invaluable resource to obtain a mechanistic understanding of the functioning and resilience of the holobiont. Recent genomic studies have shown that the coral host, dinoflagellate symbionts, and prokaryotes living in the coral tissue have complementary pathways for nutrient exchange, highlighting the interdependence and possible co-evolution between these members of the coral holobiont [1,2]. To date, most work has focused on the coral animal and microbiota associated with its living tissue, with very little work done on the highly biodiverse and functionally important microbiota inhabiting the skeleton of the coral [3,4].

The green alga *Ostreobium* sp. is an important eukaryotic symbiont living inside the coral skeleton (Fig. 1a) [4,5]. This endolithic alga is the principal agent of coral reef bioerosion, burrowing through the limestone skeleton of corals and other reef structures and dissolving up to a kilogram of reef $CaCO_3$ per m^2 per year [6]. These algae bloom in the skeleton when corals bleach (Fig. 1b) [7,8], boosted by the extra light, and – hypothetically – the extra nitrogen and CO_2 that may reach the skeleton in the absence of Symbiodiniaceae. Part of the carbohydrates produced by *Ostreobium* photosynthesis make their way into the coral tissue, extending the time it can survive without its dinoflagellate partner [9,10]. While these ecological and physiological phenomena have been described, our knowledge of the molecular mechanisms involved is scarce, severely limiting our understanding of key processes in healthy and bleached holobionts. What is the mechanism of skeletal deterioration and how tightly does *Ostreobium* metabolism integrate with that of the coral and associated microbiota? Developing this knowledge will be essential to understand and manage the roles that skeletal microbiota play during coral bleaching.

Ostreobium has an extreme lifestyle for a green alga [11]. It lives in a very dimly lit environment, with mostly the low-energy far-red wavelengths not used up by Symbiodiniaceae available [12]. The action spectrum of *Ostreobium* photosynthesis extends into far-red

wavelengths but the underlying molecular mechanisms are little known [13,14]. The daily rhythm of oxygenic photosynthesis and respiration within the skeletal matrix leads to strong fluctuations of pH and O_2, ranging from total anoxia at night to ca. 60% of air saturation during the day [11,15] and the skeleton limits diffusion of O_2 and other compounds. These stressful conditions are not normally encountered by free-living algae and the mechanisms allowing *Ostreobium* to thrive in this extreme habitat are virtually unknown.

From an evolutionary perspective, *Ostreobium* is a member of the green plant lineage (Viridiplantae), which includes the land plants that originated in the Streptophyta lineage and a broad diversity of algae in the Chlorophyta lineage[16] (Fig. 1c). *Ostreobium* is in the Bryopsidales, an order of marine algae that has evolved in the Chlorophyta. While *Ostreobium* forms small microscopic filaments, many representatives of this order are larger seaweeds[17].

So far, the genomic resources available for coral holobiont research have been limited to the coral animal (7 genomes), its dinoflagellate photosymbionts (14 genomes) and a small fraction of the prokaryotes associated with its tissue (52 metagenome-assembled genomes[2]). Here, we present the first nuclear genome of an *Ostreobium* species to extend the available genomic toolkit into the coral skeleton, an element of the holobiont crucial for our comprehension of coral resilience, bleaching, and recovery. We expect that these genomic resources will spur new insights into processes of coral bleaching encompassing the entire holobiont, as this knowledge will be essential to safeguard the future of coral reefs in a changing climate. Based on comparative analyses of the *Ostreobium* genome with those of other green algae, we show *Ostreobium*'s innovations in light harvesting antennae and derive insights and hypotheses about its functions as a coral symbiont, and as an alga living in an extreme environment.

Results and Discussion

We obtained the nuclear genome data of *Ostreobium quekettii* (SAG culture collection strain 6.99, non-axenic) by assembling the sequenced reads generated by Illumina and Nanopore

platforms (Supplementary Table 1). This strain was discovered growing on a culture of a small tropical red alga and isolated into culture. The strain is nested in a lineage of *Ostreobium* species found in scleractinian coral[18] and it readily colonizes coral skeleton when that is provided as a substrate. This clearly shows that the strain is an appropriate representative for this limestone-burrowing coral-associated genus, despite it being initially isolated from a non-coral source. *Ostreobium* is known to produce flagellated spores and we think that a spore residing on the surface of the red alga is the most likely source of the strain.

We assembled the nuclear genome of *Ostreobium quekettii* in 2,997 scaffolds (total 150.4 Mb of assembled bases; N50 length 71,393 bp), with an average sequence coverage of 116x for short-read data. The GC content of the *Ostreobium* nuclear genome is 52.3%, which is higher than 40.4% of *Caulerpa lentillifera* (the closest relative of *Ostreobium* for which a nuclear genome has been sequenced), but less than most other green algae (Supplementary Table 2). The annotation process resulted in 11,735 predicted protein-coding genes, of which ~52% were supported by RNA sequencing data. We recovered 74.3% of complete core conserved eukaryote genes (based on BUSCO [19]) in the assembled genome, comparable to the 75.9% in both *C. lentillifera* and *Ulva mutabilis* (Supplementary Fig. 1).

Photobiology in a dark place

Ostreobium has developed adaptations to the peculiar light environment in which it lives, but its mechanisms are not well understood[13,14]. Our investigation of the light harvesting complex (LHC) proteins shows that *Ostreobium* has the highest number of LHC proteins observed among green algae and *Arabidopsis thaliana* (Fig. 2a, Supplementary Figs 1 and 2). This expansion of the light harvesting protein arsenal is found in both photosystems, with duplications of the *Lhca1* and *Lhca6* gene families associated with PSI (Supplementary Fig. 2) and the presence of both *Lhcp* and *Lhcb* families associated with PSII (Supplementary Fig. 3).

Most green algae possess a single *Lhca1* gene, while *Ostreobium* has four (Fig. 2); two adjacent genes were found on two different scaffolds, suggesting that this gene family expanded via tandem duplication and translocation (Fig. 2b). The specific amino acid residue as a ligand

for Chl in LCH proteins is essential in determining the chromophore organization, which affects light spectral absorption[20]. The Lhca1 protein of green plants typically uses histidine as the chlorophyll-binding residue (A5 site in Fig. 2c), but *A. thaliana* mutants use asparagine at the A5 site, yielding red-shifted absorption spectra [20,21]. Although most green algae use histidine at the A5 site, the siphonous green algae (*Ostreobium*, *C. lentillifera* and *Bryopsis corticulans*) all have asparagine (Fig. 2c), hinting at their capability to use far-red wavelengths. Lhca6 is located in the outer LHC belt, where it forms heterodimers (with Lhca5), and their long C-terminal loops facilitate interactions between the inner and outer LHC belts[22,23]. The *Lhca6* gene occurs as a single copy in most microalgae and is absent in the seaweeds *C. lentillifera* and *U. mutabilis.* Remarkably, we recovered six copies of *Lhca6* in *Ostreobium* (Supplementary Fig. 2). This result is supported by transcriptomic evidence for all *Lhca6* genes and different chromosomal contexts (i.e. flanking genes) on the contigs.

In the PSII-associated LHC, *Ostreobium* possesses an unusual combination of both the Lhcp and major light-harvesting complex (Lhcb) protein families, and has more copies of both genes than do other green algae (Fig. 2a, Supplementary Fig.3). Lhcb is found in most species of the green lineage except prasinophytes that have the Lhcp family instead [24]. Only the streptophyte *Mesostigma viridis* is known to encode proteins of both families (Supplementary Fig. 3) suggesting that Lhcp and Lhcb were both part of the LHCII antenna system of the green plant ancestor and the families were differentially lost in different green algal and land plant lineages[25,26].

Although *Ostreobium* shows a high diversity of LHC genes, it lacks many genes for photoprotection and photoreceptors. The non-photochemical quenching (NPQ) genes for LHCSR and PsbS are both absent from *Ostreobium* and *C. lentillifera* genomes (Supplementary Fig. 3 and Supplementary Table 3). Energy-dependent quenching (qE) was not observed in other siphonous green algae [27] and our genomic data provide clear evidence that these algae are not capable of this mechanism. *Ostreobium* also lacks the light-harvesting complex-like (LIL) genes coding for OHP1 and OHP2 (Supplementary Table 3). While the function of LIL proteins has not been comprehensively determined, their involvement in response to light stress is

known[28]. The loss of genes involved in high-light sensitivity also extends to the chloroplast genome of *Ostreobium,* which lacks the chloroplast envelope membrane protein gene (*cem*A) that is needed in *Chlamydomonas* to persist in high-light conditions[29].

Ostreobium has fewer known photoreceptors than do most other green algae. We found three blue light photoreceptor genes: a phototropin (not shown), a plant cryptochrome and a photolyase/blue-light receptor 2 (PHR2; Supplementary Fig. 4). Phototropins are widespread in green plants [30]. In *C. reinhardtii*, phototropins are involved in the sexual cycle [31], phototactic response, light-dependent expression of LHC proteins, chlorophyll and carotenoid biosynthesis, and LHCSR3-based non-photochemical quenching [32]. However, the function of phototropin in *Ostreobium* remains unknown, and we did not recover LHCSR in the genome. Surprisingly, given the predominance of far-red light in coral skeletons, no genes coding for the red/far-red phytochromes were found in the *Ostreobium* genome. Although phytochromes are widely found in the green plant lineage, they have been lost in most of the green algal lineage Chlorophyta[33], although not in *Micromonas*, a prasinophyte microalga[34]. Most green algae appear to have no specific photoreceptors for red light [35]. However, *Chlamydomonas* has an animal-like cryptochrome that can be activated by either red or blue light [36]. The function of this protein is related to transcription of genes involved in photosynthesis, pigment biosynthesis, cell cycle control and circadian clock [37]. This gene is not found in *Ostreobium,* which also lacks the Cry-DASH-type cryptochromes observed in many other green algae (Supplementary Fig. 4). Functional information for Cry-DASH is still scarce [35], but this protein is known to be involved in DNA repair in *Arabidopsis* [38]. *Ostreobium* does have two genes similar to the *Arabidopsis* putative blue-light receptor protein called PAS/LOV (Uniprot O64511) [39]. The UV resistance locus8 (UVR8), an important photoreceptor to induce UV-B acclimation, is also absent in *Ostreobium* as well as *C. lentillifera*. We did not find different types of rhodopsin-like photoreceptors in *Ostreobium.*

Besides the photoprotective LHC genes and photoreceptors, *Ostreobium* and *C. lentillifera* appear to have lost the light-dependent protochlorophyllide oxidoreductase (LPOR), another important light signaling gene involved in chlorophyll biosynthesis. The reduction of

protochlorophyllide to chlorophyllide can be catalysed by either of two non-homologous enzymes: the nuclear encoded LPOR and the plastid-encoded light-independent (DPOR) protochlorophyllide oxidoreductase [40]. Most green algae have both systems, and DPOR has been lost in many eukaryotes [41]. The *Ostreobium* genome provides the first evidence for the loss of LPOR in any eukaryote, and a screening of *C. lentillifera* also came back negative, suggesting that the loss may have occurred in the common ancestor of Bryopsidales. Both genera encode DPOR in their chloroplast genomes [29,42]. Although both enzymes catalyse the same reaction, they have different features: DPOR is encoded, synthesized and active in the plastid, and is highly sensitive to oxygen [43], while the nucleus-encoded LPOR is synthesized in the cytosol and active in the plastid, and requires light to be activated [44]. DPOR might be an advantage over LPOR for endolithic photosynthetic organisms because of the low-light, low-oxygen environments they inhabit [4].

The *Ostreobium* genome clearly reflects its evolutionary trajectory into a peculiar light habitat, with an unparallelled arsenal of LHC proteins but few known mechanisms to sense the light or defend itself against excessive light. Some of these genome features are shared with other Bryopsidales, including the loss of LPOR, qE-type non-photochemical quenching and the UV resistance locus 8. This suggests that the common ancestor of Bryopsidales may have been a low-light-adapted organism, possibly an endolithic alga like *Ostreobium* is now, a hypothesis supported by *Ostreobium* being the sister lineage of all other Bryopsidales [45] and other bryopsidalean lineages also containing old, largely endolithic families [46]. Bryopsidales originated in the Neoproterozoic with *Ostreobium* diverging in the early Paleozoic [17,47]. One could speculate that the bryopsidalean ancestor inhabited a low-light environment, possibly on the dimly lit seafloor beneath Cryogenian ice sheets. The different lineages emerging from this ancestor could then have followed different evolutionary trajectories during the onset of Paleozoic grazing, with the *Ostreobium* lineage fully committing to an endolithic lifestyle while other bryopsidalean lineages engaged in an evolutionary arms race with grazers to form larger and chemically defended macroalgae.

The unusually large arsenal of LHC proteins appears to be confined to the *Ostreobium* lineage. The genus is present in a diverse range of light environments, from old oyster shells in the intertidal, where it experiences full spectrum sunlight, to coral skeleton (far-red enriched) and in mesophotic habitats (blue light), hinting at its capability to harvest energy from photons of many different wavelengths [48]. While from the genome data alone, we cannot derive that the different LHC proteins convey a capability for photosynthesis at different wavelengths, it is known that different molecular arrangement of LHC proteins and the specific pigments binding to them have an impact on their spectral properties [49], which could help *Ostreobium* acclimate to different light environments.

Life in an extreme environment

The endolithic environment, and the coral skeleton, is an extreme environment in many ways. Oxygen levels vary strongly, from high concentrations caused by photosynthesis during the day to complete anoxia due to respiration during the night, and this trend is mirrored in strong diurnal pH fluctuations [4,15]. Reactive oxygen species (ROS) can be produced in high quantities in these conditions, particularly in the morning when photosynthesis starts [50].

The *Ostreobium* genome exhibits a strong genetic capacity for oxidative stress response, with ROS scavenging and neutralizing genes present in large numbers compared to other green algae (Fig. 3). We found six genes coding for superoxide dismutases (SOD; Fig. 3 and Supplementary Fig. 5), metal-containing enzymes that form the first line of defence to scavenging ROS by catalysing the very reactive superoxide radical ($O_2^{\cdot-}$) into hydrogen peroxide (H_2O_2) [51]. All three categories of SOD are present in *Ostreobium.*

Hydrogen peroxide is commonly processed by catalase (CAT), of which *Ostreobium* has eight copies while most other green algae have one or two (Fig. 3), or none in the case of prasinophytes. Seven of the *Ostreobium* catalases formed a unique lineage in our phylogenetic analysis (Supplementary Fig. 5) and are found in tandem on three different scaffolds, indicating diversification of this gene in the *Ostreobium* lineage. Hydrogen peroxide can also be processed through the glutathione-ascorbate cycle, a metabolic pathway neutralizing hydrogen peroxide

through successive oxidations and reductions of ascorbate, glutathione, and NADPH[51] (Fig. 3). *Ostreobium* featured high copy numbers of the enzymes to quickly purify ascorbate, with eight copies of ascorbate peroxidase (APX) and seven monodehydroascorbate reductases (MDHAR) (Supplementary Fig. 5). The glutathione-ascorbate cycle is important to keep ascorbate free for H_2O_2 scavenging (Fig. 3).

Most algae live in environments with higher oxygen concentrations and can produce energy via the respiratory electron transport chain. In the coral skeleton, however, any oxygen produced through photosynthesis is completely consumed by respiration within an hour of the onset of darkness[15], and the environment is anoxic for long periods of time, hence we expect *Ostreobium* to use other electron acceptors to produce ATP and re-oxidize NAD(P)H and $FADH_2$. *Ostreobium* possesses the enzymes required to produce succinate, lactate, formate, acetate, ethanol, alanine, and glycerol (Supplementary Fig. 6), but lacks H_2 and acetate production from acetyl-CoA (PAT1/PAT2 and ACK1/ACK2). Several fermentation-related genes are present in multiple copies (Supplementary Table 4), including two lactate dehydrogenases, four tandem copies of ALDH (aldehyde dehydrogenase), and eight copies of malate dehydrogenase, implying that acetate, succinate and lactate are the preferred fermentation products in *Ostreobium*. We did not find the gene for pyruvate decarboxylase, which catalyses the production of acetaldehyde, the substrate of ALDH. This gene is also absent in *C. lentillifera*, *U. mutabilis* and prasinophytes, indicating it has been lost multiple times, and it is reasonable to assume that another yet uncharacterised enzyme may catalyse this reaction in these algae.

A comparison of the number of genes having particular InterPro annotations between *Ostreobium* and *Caulerpa* showed a large number of depleted IPR terms in *Ostreobium,* which can be attributed to an enrichment in the *Caulerpa* genome, as counts in *Ostreobium* are comparable to those of other green algae. However, genes coding peptidases were strongly enriched in *Ostreobium* (e.g. Peptidase S1, PA clan – 179 genes; Serine-proteases trypsin domain – 142 genes; Supplementary Fig. 7). About 40% of the genes are predicted to be secreted (signal peptide), and 25% are membrane bound, compared to 31% and 12% (of total 57

genes) respectively in *C. lentillifera*, suggesting an expanded potential for external protein degradation in *Ostreobium*. It is known that *Ostreobium* can penetrate the organic-inorganic composite material of the black pearl oyster nacreous layer [52], and the combination of proteolytic enzyme proliferation and its growth at excessively low light intensities lends support to the hypothesis that this alga could complement its energy needs by feeding on the organic matrix of the coral skeleton and shells.

The coral holobiont

Ostreobium plays a number of important roles in the coral holobiont, particularly during periods of coral bleaching [4], but current knowledge is far from complete and the genome can help define hypotheses of how the species may interact with other holobiont members.

Molecular mechanism of CaCO3 dissolution

Ostreobium is the main agent of microbial bioerosion of the coral skeleton. While the molecular mechanism behind this phenomenon is not yet known[53], the *Ostreobium* genome reveals the molecular toolkit available for this process and allows us to make conjectures about how bioerosion by *Ostreobium* might occur (Fig. 4a). A working model for microbial carbonate excavation was first described for cyanobacteria, involving passive uptake of Ca^{2+} at the boring front, decreasing the ion concentration in the extracellular area below calcite saturation levels and leading to dissolution of adjacent calcium carbonate [54]. Imported Ca^{2+} is transported along the cyanobacterial filament and excreted away from the growing tip, likely by P-type Ca^{2+}-ATPases that pair transport of Ca^{2+} with counter transport of protons[54]. Whereas the Ca^{2+}-ATPase mechanism appears to be somewhat widespread among endolithic cyanobacteria, it is unclear whether this mechanism extends to microboring algae like *Ostreobium*[55].

Ostreobium has an expanded repertoire (33 genes) of calcium transporters (Supplementary Table 5), including 18 voltage-dependent calcium channels and several transient receptor potential transporters, two-pore channels and calcium-transporting ATPases. Calcium uptake, possibly combined with acidification via a Ca^{2+}/H^{+} transporter would promote

297 decalcification (Fig. 4a), allowing *Ostreobium* to burrow into the coral skeleton. Calcium
298 toxicity can be avoided either by accumulation in a vacuole and/or posterior transport out of the
299 cell. In addition to calcium transport, bicarbonate uptake could also play a role in the burrowing
300 mechanism. *Ostreobium* carries two orthologs of the CIA8 transporter responsible for
301 bicarbonate transport in *C. reinhardtii* [56], and the imported bicarbonate may be further
302 transported into the chloroplast to be fixed. *Ostreobium* has a carbonic anhydrase predicted to
303 be targeted outside the cell that may further assist the decalcification process. In natural
304 communities dominated by *Ostreobium*, $CaCO_3$ dissolution was higher at night, suggesting that
305 *Ostreobium* takes advantage of the lower external pH at night [53].

306 Whilst the *Ostreobium* genome does not provide confirmation of a specific mechanism
307 for $CaCO_3$ dissolution, it lends support to a growing hypothesis of how bioerosion might work
308 in *Ostreobium*, likely bearing similarities to the process described in cyanobacteria. A detailed
309 characterisation of this dissolution process should be a priority, as *Ostreobium* is responsible for
310 ca. 30-90% of microbial dissolution of skeletal $CaCO_3$. Higher temperatures and lower pH boost
311 this activity, suggesting that this yet unknown process will lead to major reef deterioration in
312 future ocean conditions[53].

313 *Interactions with other holobiont members*

314 As far as holobiont functioning goes, interactions between the coral animal and its algal
315 and bacterial symbionts are best understood, and interactions between the Symbiodiniaceae and
316 bacteria are just starting to come into focus [57], but little is known about interactions involving
317 *Ostreobium*. It is well known that several algae lack genes related to biosynthetic pathways of,
318 and are auxotrophic for, certain nutrients. The *Ostreobium* genome shows that the metabolic
319 pathways involved in the production of vitamins B1, B2, B6 and B9 are complete, but it lacks
320 the BioA and LysA genes involved in biotin and lysine synthesis. These genes were found in all
321 other green algae used for comparison, suggesting that this gene loss is specific to *Ostreobium.*
322 It is likely that the alga obtains these nutrients from bacteria, and the dependency on external
323 lysine may also explain the enrichment of extracellular proteases encoded in the *Ostreobium*

genome. Many algae are auxotrophic for vitamin B12 (cobalamin), a cofactor involved in the synthesis of methionine, instead obtaining this compound from associated bacteria [58]. The *Ostreobium* genome encodes two gene copies of a B12-independent methionine synthase (METE) in addition to the B12-dependent version (METH, [59]), so while the alga most likely does not strictly require B12 for growth, the presence of METH suggests that it uses B12 provided by other holobiont members. Corals are also auxotrophic for several vitamins and amino acids that are produced by holobiont members [2] (Fig. 4b).

The nature of metabolic exchanges between *Ostreobium* and the coral animal is an open question in holobiont research, and a potentially critical one, as coral bleaching and subsequent blooming of the endolithic *Ostreobium* algae become increasingly common due to ocean warming. Typical algal-animal metabolic exchanges include nitrogen and CO_2 provision by the animal to the alga and carbohydrate provision to the animal by the alga [60,61]. Coral polyps are known to secrete nitrogen in the form of ammonia. An expanded repertoire of ammonia transporters was identified in *Ostreobium* (Supplementary Table 5), potentially reflecting an adaptation to increase and diversify ammonia uptake in the alga. This observation is also in line with the presence of diazotrophic bacteria facilitating the conversion of N_2 into ammonia in marine limestones (including coral skeletons) and ammonia being the most abundant form of inorganic nitrogen in skeletal pore waters [62], and adds to the evidence for the roles that endolithic organisms play in the holobiont N cycle. During carbon cycling in the holobiont, glucose has been postulated as one of the main carbohydrates exchanged between Symbiodiniaceae and corals [60]. While carbon compounds fixed by endoliths are known to be transferred to the coral animal and subsequently assimilated, neither the exact transferred molecules nor the molecular mechanisms involved in their translocation have been characterised, but the *Ostreobium* genome encodes two genes coding for H^+-glucose transporters that might be involved in this process.

Probing changes in holobiont processes

While these investigations of the *Ostreobium* genome allowed us to evaluate hypotheses about interactions within the holobiont, gaining deeper insight will require approaches that study multiple partners simultaneously. As a first step towards understanding the molecular mechanisms in the coral holobiont during coral bleaching, we screened transcriptomes of healthy and bleached coral holobionts. We exposed fragments of the coral *Orbicella faveolata* to elevated temperatures leading to bleaching, followed by re-acclimation of the bleached samples to ambient temperatures (post-bleaching condition), during which the corals remained bleached and an *Ostreobium* bloom occurred. Total metatranscriptomes (including coral tissue and skeleton) were generated for healthy control samples that were kept under ambient temperature (no bleaching), and the bleached samples.

From the KEGG annotation of the assembled transcriptome, it is clear that the number of transcripts coding for the *Ostreobium* photosynthetic apparatus exceeds that of Symbiodiniaceae (Supplementary Fig. 8), in line with the expanded set of photosynthetic genes (particularly LHC) observed in the *Ostreobium* genome. While the metatranscriptome sequencing depth was not designed to track up- and down-regulation of individual genes, it was clear that expression of the photosynthesis genes *psb*A and *rbc*L from Symbiodiniaceae was drastically lower in bleached samples while expression of these genes in *Ostreobium* increased (Supplementary Fig. 9a), supporting the notion that *Ostreobium* becomes a dominant agent of photosynthesis in the bleached holobiont. The bleached state increases the light available to *Ostreobium*, which likely leads to the need for higher repair rates of PSII protein D1 (encoded by *psb*A) and higher rates of carbon fixation (facilitated by RuBisCO, encoded by *rbc*L). The expression of these genes by Symbiodiniaceae in the bleached samples, albeit at low levels, suggests that some of these endosymbionts have remained or that some re-colonisation has occurred during the post-bleaching period.

We detected multiple haplotypes of *psb*A (Supplementary Fig. 9b), indicating that several strains of *Ostreobium* were present in the *O. faveolata* skeletons[63]. While expression levels for

these haplotypes tended to increase, some did not change significantly while others differed by an order of magnitude or more. These differences suggest that *Ostreobium* strains may differ physiologically or change in relative abundance during the experiment. Such differences in the microbiome – whether *Ostreobium* strains or other holobiont members – may result in high variability among coral samples in experimental work. Indeed, we found considerable differences in expression levels between samples within conditions, suggesting that future metatranscriptome experiments should be planned to use generous replication.

The metatranscriptomes also provide some hints as to where *Ostreobium* may source its needs for vitamin B12 and biotin. We detected 35 bacterial transcripts from the vitamin B12 pathway, including several Proteobacteria (e.g. Rhodobacteraceae), Bacteroidetes (e.g. Flavobacteriaceae) and Cyanobacteria that are known to be abundant in the coral skeleton [46]. Seven bacterial transcripts associated with the biotin pathway were found, with taxonomic affinities to Actinobacteria, Cyanobacteria and Proteobacteria (e.g. Endozoicomonadaceae, Rhodospirillaceae). This result narrows down the potential mutualistic relationships between bacteria and *Ostreobium* in the holobiont (Fig. 4b).

Conclusion

The complexity of the coral holobiont presents an interesting challenge to reconstruct a comprehensive model of metabolic exchanges and other interactions among its component organisms. Recent progress in building such models from genomes of coral and tissue-associated prokaryotes [2] has not been mirrored in the skeleton. Our results allowed refining hypotheses from previous physiological work, illuminating the biology of skeletal inhabitants and their interactions. Despite this progress, many questions remain about the exact mechanisms at play, their immediate physiological effects on the partners and longer-term ecological consequences. One challenge in this area of research is that the roles and impacts of *Ostreobium* vary in time. In terms of photosynthesis, *Ostreobium* makes a minor contribution to holobiont photosynthesis in healthy corals [10], but our results show that *Ostreobium* blooms during

bleaching cause a shift of expression levels between microbiome members. Specifically, *Ostreobium* takes over the major role of photosynthesis from the Symbiodiniaceae now lost from the holobiont. These changes likely have implications flowing through the entire microbiome interaction network, but our knowledge of this is in its infancy. From a physiological viewpoint, the question remains as to what extent the *Ostreobium* bloom is simply an opportunistic response, and whether it is actually leading to a beneficial metabolite exchange with the coral host. Similarly, the fitness costs/benefits of *Ostreobium* in the skeleton need to be evaluated across the entire coral life cycle and might differ drastically between the period of early skeleton formation following larval settlement and at later stages. The fact that there are >80 different species-level operational taxonomic units in *Ostreobium*, and virtually nothing is known their physiological features, complicates the matter further [46,63]. Genome-scale data from *Ostreobium* and other interacting partners critically link these physiological features to underlying molecular mechanisms. The results and datasets generated in this study provide a foundational reference for future research into the biology of this key holobiont member, and the intricate role that *Ostreobium* plays in coral biology. This will be of particular importance in the light of global climate change, as the increased frequency of bleaching and lower pH will boost *Ostreobium* populations and carbonate dissolution rates.

Methods

Culturing and nucleic acid extraction

Ostreobium quekettii (SAG culture collection strain 6.99, non-axenic) was cultured in F/2 media on a 14H - 8H light/dark cycle at ~19 degrees celsius. Total DNA was extracted using a modified cetyl trimethylammonium bromide (CTAB) method described in Cremen and collaborators[64]. Total RNA was extracted using Plant RNA reagent (Thermofisher, Waltham, MA, USA).

Illumina sequencing and assembly

Total DNA was sequenced using Illumina sequencing technology (HiSeq2000, 150 bp paired-end reads, ~40 GB data) (Supplementary Table 1), at Novogene, Beijing. Reads were trimmed with Cutadapt v1.12[65] using the parameters -e 0.1 -q 10 -O 1 -a AGATCGGAAGAGC. To remove plDNA, mtDNA and bacterial sequences prior to assembly all reads were mapped against the previously published *Ostreobium* chloroplast and mitochondrial sequences and a database of bacterial genomes, using BWA with the bwa-mem algorithm at default settings[66]. The bacterial database was constructed from 5460 prokaryotic representative genomes (https://www.ncbi.nlm.nih.gov/refseq/about/prokaryotes/#representative_genomes) downloaded from NCBI using the tool ncbi-genome-download (https://github.com/kblin/ncbi-genome-download). If one or both pair-end reads mapped to any database sequence, both reads were removed from the full Illumina dataset using the BBtools 'filterbyname.sh' program (https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/). The resulting dataset was assembled using Spades v3.12.0 with the --careful option[67]. Contigs produced by Spades were further filtered to identify any remaining bacterial or fungal sequences by BLAST searches of open reading frames against the nr database.

For RNA sequencing, a strand-specific 100 bp paired-end library was constructed and sequenced using Illumina HiSeq 2500 (ENA study accession number PRJEB35267). Quality filtering of reads was performed using Trimmomatic v0.39[68] with the following settings: LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20. Transcriptome data were assembled using Trinity v2.8.3[69] in both genome-guided mode and de novo mode.

Nanopore sequencing and hybrid assembly

Total DNA was extracted as above and sequenced using nanopore sequencing technology (MinION, Oxford Nanopore Technologies), producing 1,889,814 reads and ~9.5 GB data. To identify sequences originating from bacteria, all reads were used as blastn queries (megablast

algorithm) against the prokaryotic representative genome dataset. Reads with a hit alignment >1000 bp long and an e-value < 1e-100 were removed from the nanopore dataset.

A hybrid (combined long-read and short-read) genome assembly was performed with MaSuRCA v3.2.7 [70], using the filtered Illumina reads as short-read input. For long-read input, the nanopore reads were complemented with the contigs produced from the Spades assembly of Illumina data and treated as 'pseudo-nanopore' reads.

Ab initio prediction of protein-coding genes

We adapted the workflow from [71] for ab initio prediction of protein-coding genes in the *Ostreobium* genome assembly. Repetitive elements in the genome assembly were first predicted de novo using RepeatModeler v1.0.11 (http://www.repeatmasker.org/RepeatModeler/). These repeats were combined with known repeats in the RepeatMasker database (release 20171107) to generate a customized repeat library. All repetitive elements in the assembled genome scaffolds were then masked using RepeatMasker 4.0.7 (http://www.repeatmasker.org/) based on the customized repeat library, before they were subjected to prediction of genes.

The assembled transcripts were used as transcriptome evidence, and vector sequences were removed using SeqClean [72] based on UniVec (build 10) database. PASA pipeline v2.3.3 [73] and TransDecoder v5.2.0 [74] were first used to predict protein-coding genes (and the associated protein sequences) from the vector-trimmed transcriptome assemblies (hereinafter transcript-based genes). The predicted protein sequences from multi-exon transcript-based genes with complete 5′ and 3′-ends were searched (BLASTp, E ≤ 10-20) against RefSeq proteins (release 88). Genes with significant BLASTp hits (>80% query coverage and >80% subject coverage) were retained. Transposable elements were identified using HHblits [75] and Transposon-PSI (http://transposonpsi.sourceforge.net/), searching against the JAMg transposon database (https://github.com/genomecuration/JAMg). Proteins putatively identified as transposable elements were removed. Those remaining were clustered using CD-HITS (ID=75%) [76] to yield a non-redundant protein set, and the associated transcript-based genes were kept. These genes were further processed by the Prepare_golden_genes_for_predictors.pl script from the JAMg

478 package (https://github.com/genomecuration/JAMg). This step yielded a set of high-quality
479 "golden" genes, which were used as a training set for gene prediction using AUGUSTUS [77] and
480 SNAP [78]. We also employed GeneMark-ES v4.38 [79] to generate prediction from the genome
481 scaffolds, and MAKER protein2genome v2.31.10 [80] to make predictions based on homology
482 evidence to SwissProt proteins (downloaded 27 June 2018).
483 Subsequently, all genes predicted using GeneMark-ES, MAKER, PASA, SNAP and
484 AUGUSTUS were integrated into a combined set using EvidenceModeler v1.1.1 [81], following a
485 weighting scheme of GeneMark-ES 2, MAKER 8, PASA 10, SNAP 2, AUGUSTUS 6. The
486 resulting genes from the EVidenceModeler prediction were retained if they were constructed
487 using evidence from PASA, or using two or more other prediction methods.
488 Functional information of translated predicted genes was retrieved using search BLASTP
489 against UniProt databases (Swiss-Prot and TrEMBL), KEGG's annotation tool BlastKOALA [82].
490 Gene models of the genome dataset were annotated using InterProScan 5.39 [83] using InterPro
491 (version 77.0 databases) of Pfam (32.0), SUPERFAMILY (1.75) and TIGRFAMs (15.0) [84]. The
492 *Ostreobium* genomic data is deposited in https://cloudstor.aarnet.edu.au/plus/s/X9xS30hE7.

493 ## Orthogroups, phylogenetic analysis and BUSCO analysis

494 For comparative genomic analyses, we built a dataset containing genomes and
495 annotations of 20 green algae and two land plants (Supplementary Table 4). We used the
496 OrthoFinder 2.3.7 [85] pipeline (default parameters) to cluster the potential orthologous protein
497 families. Enrichment and depletion of domains in *Ostreobium* versus *Caulerpa lentillifera* were
498 identified using Fisher's exact tests with a false discovery rate correction (Benjamini-Hochberg
499 FDR method) of 0.05. All statistical tests were carried out in R [86]. Subcellular localization of
500 sequences of interest was performed using PSORT [87], using the WoLF PSORT web server
501 (https://wolfpsort.hgc.jp/) with the 'plant' option, and PredAlgo [88], using default settings.
502 We performed phylogenetic analyses for protein families relevant for photobiology and
503 oxidative stress response, based on the orthogroups obtained from the OrthoFinder analysis
504 described above. All phylogenetic trees were reconstructed using IQTREE 1.6.12 with the built-

in model selection function, and branch support estimated using ultrafast bootstrap with 1,000 bootstrap replicates [89]. To identify the light-harvesting protein families associated with PSI (Lhca), we also included Lhca protein sequences from *Bryopsis corticulans* [23] and *Chlamydomonas reinhardtii* [22].

We ran BUSCO v3.1.0 [19] on the *Ostreobium* genome using the Eukaryota dataset. This analysis identifies complete, duplicated, fragmented and missing genes that are expected to be present in a set of single-copy genes in the dataset. We ran identical BUSCO analyses on the *C. lentillifera*, *Ulva mutabilis* and *C. reinhardtii* genomes to allow direct comparison.

Metatranscriptome analysis of healthy and bleached corals

Experiment

Orbicella faveolata fragments (4cm^2) from three different colonies were collected in Petempiche Puerto Morelos, Quintana Roo, Mexico (N 20° 54'17.0", W 86° 50'11.9") at a depth between eight and nine meters in June 2013 (Permit registration MX-HR-010-MEX, Folio 036). The coral skeleton and live tissue were collected using a hammer and chisel and were transported in seawater to perform the experiment at the Instituto de Ciencias del Mar y Limnología, UNAM. Three fragments were placed in each control and "experimental" tank at ~28ºC. Following 18 days of acclimation, heaters were turned on in the treatment tank to reach ~32ºC. After 5 days of severe heat exposure, bleached corals were moved back into the control tank to recover at 28ºC, where they remained for 38 days (post-bleaching period). Post-bleached and controlled coral fragments were flash frozen and preserved in liquid nitrogen.

RNA library preparation and sequencing and data availability

Coral fragments were ground to a fine powder in liquid nitrogen. Total RNA was extracted using the mirVana miRNA Isolation Kit (Life Technologies) since this kit resulted in higher coral holobiont RNA yield and quality. RNA was purified and concentrated using the RNA Clean and Concentrator kit (Zymo Research, Irvine, USA). RNA quantification was

assessed on a Nanodrop and Qubit using 2.0 RNA Broad Range Assay Kit (Invitrogen). Quality was verified using the Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, USA). Total RNA samples were sent for metatranscriptome sequencing to the US Department of Energy's Joint Genome Institute (JGI), California. Samples were depleted in ribosomal RNA and enriched in mRNA from whole holobionts employing the RiboZero kits (Epicenter). Following the RiboZero protocol, mRNA was converted to cDNA and amplified. The libraries were sequenced on the Illumina HiSeq 2000 platform using 2 X 151 bp overlapping paired-end reads. Raw and filtered metatranscriptome sequence data, statistics and quality sequencing reports for the experiment are available at the US Department of Energy Joint Genome Institute (JGI)'s genome portal (https://genome.jgi.doe.gov/portal/) with accession codes 1086604, 1086606, 1086608, 1086610, 1086612, 1086614, Community Sequencing Project No. 1622).

Metatranscriptome analysis

The raw reads were quality-trimmed to Q10, adapter-trimmed and filtered for process artifacts using BBDuk[90]. Ribosomal RNA reads were removed by mapping against a trimmed version of the Silva database using BBmap (http://sourceforge.net/projects/bbmap). To generate a de novo reference metatranscriptome, cleaned reads from all samples per species (replicates from control and treatment fragments) were pooled and assembled using Trinity version 2.1.1[74].

We performed a BLASTn search to identify and separate the different members of the holobiont using a local database. This database was built with the *Ostreobium* genome, the sea anemones *Nematostella vectensis*[91] and *Exaiptasia diaphana* (syn. *Aiptasia pallida*)[92], the corals *Acropora digitifera*[93] and *Orbicella faveolata*[94], and the Symbiodiniaceae genomes *Breviolum minutum*[95], *Fugacium kawagutii*[1] and *Symbiodinium microadriaticum*[96], as well as several *Breviolum* spp. transcriptomes[97].

Kallisto[98] was used to perform a pseudoalignment and quantify transcript abundances, using the *Ostreobium* contigs derived from the metatranscriptome assembly as a reference. A comparison of counts per million, a correlation matrix and principal component analysis among samples was performed for a quality check of the replicates per species. Expression analysis

was performed using the DESeq2 software [99]. Differential Expressed Genes (DEGs) were defined by using a cutoff threshold of False Discovery Rate FDR <0.001 and log fold change of 2. Enzyme Commision numbers (EC) were retrieved from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database using MEGAN5 [100] for genes from each holobiont member (i.e. host, Symbiodiniaceae and *Ostreobium*).

Acknowledgements

Funding was provided by the Australian Research Council (FT110100585 to HV, DP150100705 to HV and CXC, DP200101613 to HV and MM), the University of Melbourne (CBRI to HV & KAH), the DoE Joint Genome Institute (CSP grant 1622 to MM and VAM) and the National Science Foundation (OCE 1442206 and IOS 0644438 to MM), Pennsylvania State University (to MM), the Canon Foundation (to MM) and CONACyT (216837 to VAM). Computational resources were provided through the Nectar Research Cloud. We thank Roberto Iglesias-Prieto, Claudia Tatiana Galindo and Michele Weber for assisting with field experiments and John Beardall, J. Clark Lagarias and Andrew H. Knoll for discussions. We thank Alexander Fordyce for the bleached coral photograph.

References

1 Lin, S. *et al.* The Symbiodinium kawagutii genome illuminates dinoflagellate gene expression and coral symbiosis. *Science* **350**, 691-694, doi:10.1126/science.aad0408 (2015).

2 Robbins, S. J. *et al.* A genomic view of the reef-building coral Porites lutea and its microbial symbionts. *Nat Microbiol* **4**, 2090-2100, doi:10.1038/s41564-019-0532-4 (2019).

3 Marcelino, V. R., Morrow, K. M., van Oppen, M. J. H., Bourne, D. G. & Verbruggen, H. Diversity and stability of coral endolithic microbial communities at a naturally high pCO2 reef. *Mol Ecol* **26**, 5344-5357, doi:10.1111/mec.14268 (2017).

4 Ricci, F. *et al.* Beneath the surface: community assembly and functions of the coral skeleton microbiome. *Microbiome* **7**, 159, doi:10.1186/s40168-019-0762-y (2019).

5 Verbruggen, H. & Tribollet, A. Boring algae. *Curr Biol* **21**, R876-877, doi:10.1016/j.cub.2011.09.014 (2011).

6 Tribollet, A. in *Current Developments in Bioerosion* Ch. Chapter 4, 67-94 (2008).
7 Diaz-Pulido, G. & McCook, L. J. The fate of bleached corals: patterns and dynamics of algal recruitment. *Marine Ecology Progress Series* **232**, 115-128 (2002).
8 Fine, M., Roff, G., Ainsworth, T. D. & Hoegh-Guldberg, O. Phototrophic microendoliths bloom during coral "white syndrome". *Coral Reefs* **25**, 577-581, doi:10.1007/s00338-006-0143-4 (2006).
9 Schlichter, D., Zscharnack, B. & Krisch, H. Transfer of photoassimilates from endolithic algae to coral tissue. *Naturwissenschaften* **82**, 561-564, doi:10.1007/BF01140246 (1995).
10 Fine, M. & Loya, Y. Endolithic algae: an alternative source of photoassimilates during coral bleaching. *Proc Biol Sci* **269**, 1205-1210, doi:10.1098/rspb.2002.1983 (2002).
11 Shashar, N. & Stambler, N. Endolithic algae within corals - life in an extreme environment. *J. Exp. Mar. Biol. Ecol.* **163**, 277-286 (1992).
12 Magnusson, S. H., Fine, M. & Kuhl, M. Light microclimate of endolithic phototrophs in the scleractinian corals Montipora monasteriata and Porites cylindrica. *Marine Ecology Progress Series* **332**, 119-128 (2007).
13 Koehne, B., Elli, G., Jennings, R. C., Wilhelm, C. & Trissl, H. Spectroscopic and molecular characterization of a long wavelength absorbing antenna of Ostreobium sp. *Biochim Biophys Acta* **1412**, 94-107, doi:10.1016/s0005-2728(99)00061-4 (1999).
14 Wilhelm, C. & Jakob, T. Uphill energy transfer from long-wavelength absorbing chlorophylls to PS II in Ostreobium sp. is functional in carbon assimilation. *Photosynth Res* **87**, 323-329, doi:10.1007/s11120-005-9002-3 (2006).
15 Kuhl, M., Holst, G., Larkum, A. W. & Ralph, P. J. Imaging of Oxygen Dynamics within the Endolithic Algal Community of the Massive Coral Porites Lobata(1). *J Phycol* **44**, 541-550, doi:10.1111/j.1529-8817.2008.00506.x (2008).
16 Leliaert, F. *et al.* Phylogeny and Molecular Evolution of the Green Algae. *Critical Reviews in Plant Sciences* **31**, 1-46, doi:10.1080/07352689.2011.615705 (2012).
17 Verbruggen, H. *et al.* A multi-locus time-calibrated phylogeny of the siphonous green algae. *Mol Phylogenet Evol* **50**, 642-653, doi:10.1016/j.ympev.2008.12.018 (2009).
18 Gutner-Hoch, E. & Fine, M. Genotypic diversity and distribution of Ostreobium quekettii within scleractinian corals. *Coral Reefs* **30**, 643-650, doi:10.1007/s00338-011-0750-6 (2011).
19 Waterhouse, R. M. *et al.* BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Mol Biol Evol* **35**, 543-548, doi:10.1093/molbev/msx319 (2018).
20 Morosinotto, T., Breton, J., Bassi, R. & Croce, R. The nature of a chlorophyll ligand in Lhca proteins determines the far red fluorescence emission typical of photosystem I. *J Biol Chem* **278**, 49223-49229, doi:10.1074/jbc.M309203200 (2003).
21 Morosinotto, T., Castelletti, S., Breton, J., Bassi, R. & Croce, R. Mutation analysis of Lhca1 antenna complex. Low energy absorption forms originate from pigment-pigment interactions. *J Biol Chem* **277**, 36253-36261, doi:10.1074/jbc.M205062200 (2002).
22 Suga, M. *et al.* Structure of the green algal photosystem I supercomplex with a decameric light-harvesting complex I. *Nat Plants* **5**, 626-636, doi:10.1038/s41477-019-0438-4 (2019).
23 Qin, X. *et al.* Structure of a green algal photosystem I in complex with a large number of light-harvesting complex I subunits. *Nat Plants* **5**, 263-272, doi:10.1038/s41477-019-0379-y (2019).
24 Six, C., Worden, A. Z., Rodriguez, F., Moreau, H. & Partensky, F. New insights into the nature and phylogeny of prasinophyte antenna proteins: Ostreococcus tauri, a case study. *Mol Biol Evol* **22**, 2217-2230, doi:10.1093/molbev/msi220 (2005).
25 Koziol, A. G. *et al.* Tracing the evolution of the light-harvesting antennae in chlorophyll a/b-containing organisms. *Plant Physiol* **143**, 1802-1816, doi:10.1104/pp.106.092536 (2007).

26 Neilson, J. A. & Durnford, D. G. Structural and functional diversification of the light-harvesting complexes in photosynthetic eukaryotes. *Photosynth Res* **106**, 57-71, doi:10.1007/s11120-010-9576-2 (2010).
27 Christa, G. *et al.* Photoprotection in a monophyletic branch of chlorophyte algae is independent of energy-dependent quenching (qE). *New Phytol* **214**, 1132-1144, doi:10.1111/nph.14435 (2017).
28 Neilson, J. A. & Durnford, D. G. Evolutionary distribution of light-harvesting complex-like proteins in photosynthetic eukaryotes. *Genome* **53**, 68-78, doi:10.1139/g09-081 (2010).
29 Marcelino, V. R., Cremen, M. C., Jackson, C. J., Larkum, A. A. & Verbruggen, H. Evolutionary Dynamics of Chloroplast Genomes in Low Light: A Case Study of the Endolithic Green Alga Ostreobium quekettii. *Genome Biol Evol* **8**, 2939-2951, doi:10.1093/gbe/evw206 (2016).
30 Li, F. W. *et al.* The origin and evolution of phototropins. *Front Plant Sci* **6**, 637, doi:10.3389/fpls.2015.00637 (2015).
31 Huang, K. & Beck, C. F. Phototropin is the blue-light receptor that controls multiple steps in the sexual life cycle of the green alga Chlamydomonas reinhardtii. *Proc Natl Acad Sci U S A* **100**, 6269-6274, doi:10.1073/pnas.0931459100 (2003).
32 Petroutsos, D. *et al.* A blue-light photoreceptor mediates the feedback regulation of photosynthesis. *Nature* **537**, 563-566, doi:10.1038/nature19358 (2016).
33 Rockwell, N. C. & Lagarias, J. C. Phytochrome evolution in 3D: deletion, duplication, and diversification. *New Phytol* **225**, 2283-2300, doi:10.1111/nph.16240 (2020).
34 Duanmu, D. *et al.* Marine algae and land plants share conserved phytochrome signaling systems. *Proc Natl Acad Sci U S A* **111**, 15827-15832, doi:10.1073/pnas.1416751111 (2014).
35 Kottke, T., Oldemeyer, S., Wenzel, S., Zou, Y. & Mittag, M. Cryptochrome photoreceptors in green algae: Unexpected versatility of mechanisms and functions. *J Plant Physiol* **217**, 4-14, doi:10.1016/j.jplph.2017.05.021 (2017).
36 Beel, B. *et al.* A flavin binding cryptochrome photoreceptor responds to both blue and red light in Chlamydomonas reinhardtii. *Plant Cell* **24**, 2992-3008, doi:10.1105/tpc.112.098947 (2012).
37 Duanmu, D., Rockwell, N. C. & Lagarias, J. C. Algal light sensing and photoacclimation in aquatic environments. *Plant Cell Environ* **40**, 2558-2570, doi:10.1111/pce.12943 (2017).
38 Pokorny, R. *et al.* Recognition and repair of UV lesions in loop structures of duplex DNA by DASH-type cryptochrome. *Proc Natl Acad Sci U S A* **105**, 21023-21027, doi:10.1073/pnas.0805830106 (2008).
39 Ogura, Y., Tokutomi, S., Wada, M. & Kiyosue, T. PAS/LOV proteins: A proposed new class of plant blue light receptor. *Plant Signal Behav* **3**, 966-968, doi:10.4161/psb.6150 (2008).
40 Suzuki, J. Y. & Bauer, C. E. A prokaryotic origin for light-dependent chlorophyll biosynthesis of plants. *Proc Natl Acad Sci U S A* **92**, 3749-3753, doi:10.1073/pnas.92.9.3749 (1995).
41 Hunsperger, H. M., Randhawa, T. & Cattolico, R. A. Extensive horizontal gene transfer, duplication, and loss of chlorophyll synthesis genes in the algae. *BMC Evol Biol* **15**, 16, doi:10.1186/s12862-015-0286-4 (2015).
42 Cremen, M. C. M., Leliaert, F., Marcelino, V. R. & Verbruggen, H. Large Diversity of Nonstandard Genes and Dynamic Evolution of Chloroplast Genomes in Siphonous Green Algae (Bryopsidales, Chlorophyta). *Genome Biol Evol* **10**, 1048-1061, doi:10.1093/gbe/evy063 (2018).
43 Yamazaki, S., Nomata, J. & Fujita, Y. Differential operation of dual protochlorophyllide reductases for chlorophyll biosynthesis in response to environmental oxygen levels in the cyanobacterium Leptolyngbya boryana. *Plant Physiol* **142**, 911-922, doi:10.1104/pp.106.086090 (2006).

44 Gálová, E. *et al.* A short overview of chlorophyll biosynthesis in algae. *Biologia* **63**, doi:10.2478/s11756-008-0147-3 (2008).
45 Verbruggen, H., Marcelino, V. R., Guiry, M. D., Cremen, M. C. M. & Jackson, C. J. Phylogenetic position of the coral symbiont Ostreobium (Ulvophyceae) inferred from chloroplast genome data. *J Phycol* **53**, 790-803, doi:10.1111/jpy.12540 (2017).
46 Marcelino, V. R. & Verbruggen, H. Multi-marker metabarcoding of coral skeletons reveals a rich microbiome and diverse evolutionary origins of endolithic algae. *Sci Rep* **6**, 31508, doi:10.1038/srep31508 (2016).
47 Del Cortona, A. *et al.* Neoproterozoic origin and multiple transitions to macroscopic growth in green seaweeds. *Proc Natl Acad Sci U S A* **117**, 2551-2559, doi:10.1073/pnas.1910060117 (2020).
48 Dullo, W.-C. *et al.* Factors controlling holocene reef growth: An interdisciplinary approach. *Facies* **32**, 145-188, doi:10.1007/bf02536867 (1995).
49 Grossman, A. R., Bhaya, D., Apt, K. E. & Kehoe, D. M. Light-harvesting complexes in oxygenic photosynthesis: Diversity, Control, and Evolution. *Annu Rev Genet* **29**, 231-288 (1995).
50 Foyer, C. H. Reactive oxygen species, oxidative signaling and the regulation of photosynthesis. *Environ Exp Bot* **154**, 134-142, doi:10.1016/j.envexpbot.2018.05.003 (2018).
51 Mallick, N. & Mohn, F. H. Reactive oxygen species: response of algal cells. *Journal of Plant Physiology* **157**, 183-193, doi:10.1016/s0176-1617(00)80189-3 (2000).
52 Mao Che, L. *et al.* Biodegradation of shells of the black pearl oyster, Pinctada margaritifera var. cumingii, by microborers and sponges of French Polynesia. *Marine Biology* **126**, 509-519, doi:10.1007/bf00354633 (1996).
53 Tribollet, A., Chauvin, A. & Cuet, P. Carbonate dissolution by reef microbial borers: a biogeological process producing alkalinity under different pCO2 conditions. *Facies* **65**, doi:10.1007/s10347-018-0548-x (2019).
54 Garcia-Pichel, F., Ramirez-Reinat, E. & Gao, Q. Microbial excavation of solid carbonates powered by P-type ATPase-mediated transcellular Ca2+ transport. *Proc Natl Acad Sci U S A* **107**, 21749-21754, doi:10.1073/pnas.1011884108 (2010).
55 Ramirez-Reinat, E. L. & Garcia-Pichel, F. Prevalence of Ca(2)(+)-ATPase-mediated carbonate dissolution among cyanobacterial euendoliths. *Appl Environ Microbiol* **78**, 7-13, doi:10.1128/AEM.06633-11 (2012).
56 Machingura, M. C. *et al.* Identification and characterization of a solute carrier, CIA8, involved in inorganic carbon acclimation in Chlamydomonas reinhardtii. *J Exp Bot* **68**, 3879-3890, doi:10.1093/jxb/erx189 (2017).
57 Matthews, J. L. *et al.* Symbiodiniaceae-bacteria interactions: rethinking metabolite exchange in reef-building corals as multi-partner metabolic networks. *Environ Microbiol*, doi:10.1111/1462-2920.14918 (2020).
58 Croft, M. T., Lawrence, A. D., Raux-Deery, E., Warren, M. J. & Smith, A. G. Algae acquire vitamin B12 through a symbiotic relationship with bacteria. *Nature* **438**, 90-93, doi:10.1038/nature04056 (2005).
59 Helliwell, K. E., Wheeler, G. L., Leptos, K. C., Goldstein, R. E. & Smith, A. G. Insights into the evolution of vitamin B12 auxotrophy from sequenced algal genomes. *Mol Biol Evol* **28**, 2921-2933, doi:10.1093/molbev/msr124 (2011).
60 Burriesci, M. S., Raab, T. K. & Pringle, J. R. Evidence that glucose is the major transferred metabolite in dinoflagellate-cnidarian symbiosis. *J Exp Biol* **215**, 3467-3477, doi:10.1242/jeb.070946 (2012).
61 Radecker, N., Pogoreutz, C., Voolstra, C. R., Wiedenmann, J. & Wild, C. Nitrogen cycling in corals: the key to understanding holobiont functioning? *Trends Microbiol* **23**, 490-497, doi:10.1016/j.tim.2015.03.008 (2015).
62 Ferrer, L. M. & Szmant, A. M. Nutrient regeneration by the endolithic community in coral skeletons. *Proc. 6th Coral Reef Symp.* **3**, 1-4 (1988).

63 Del Campo, J., Pombert, J. F., Slapeta, J., Larkum, A. & Keeling, P. J. The 'other' coral symbiont: Ostreobium diversity and distribution. *ISME J* **11**, 296-299, doi:10.1038/ismej.2016.101 (2017).

64 Cremen, M. C. M., Huisman, J. M., Marcelino, V. R. & Verbruggen, H. Taxonomic revision of Halimeda (Bryopsidales, Chlorophyta) in south-western Australia. *Australian Systematic Botany* **29**, doi:10.1071/sb15043 (2016).

65 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, doi:10.14806/ej.17.1.200 (2011).

66 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

67 Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**, 455-477, doi:10.1089/cmb.2012.0021 (2012).

68 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120, doi:10.1093/bioinformatics/btu170 (2014).

69 Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* **29**, 644-652, doi:10.1038/nbt.1883 (2011).

70 Zimin, A. V. *et al.* The MaSuRCA genome assembler. *Bioinformatics* **29**, 2669-2677, doi:10.1093/bioinformatics/btt476 (2013).

71 Chen, Y., Gonzalez-Pech, R. A., Stephens, T. G., Bhattacharya, D. & Chan, C. X. Evidence That Inconsistent Gene Prediction Can Mislead Analysis of Dinoflagellate Genomes. *J Phycol* **56**, 6-10, doi:10.1111/jpy.12947 (2020).

72 Chen, Y. A., Lin, C. C., Wang, C. D., Wu, H. B. & Hwang, P. I. An optimized procedure greatly improves EST vector contamination removal. *BMC Genomics* **8**, 416, doi:10.1186/1471-2164-8-416 (2007).

73 Haas, B. J. *et al.* Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res* **31**, 5654-5666, doi:10.1093/nar/gkg770 (2003).

74 Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8**, 1494-1512, doi:10.1038/nprot.2013.084 (2013).

75 Remmert, M., Biegert, A., Hauser, A. & Soding, J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods* **9**, 173-175, doi:10.1038/nmeth.1818 (2011).

76 Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658-1659, doi:10.1093/bioinformatics/btl158 (2006).

77 Stanke, M. *et al.* AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res* **34**, W435-439, doi:10.1093/nar/gkl200 (2006).

78 Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59, doi:10.1186/1471-2105-5-59 (2004).

79 Lomsadze, A., Gemayel, K., Tang, S. & Borodovsky, M. Modeling leaderless transcription and atypical genes results in more accurate gene prediction in prokaryotes. *Genome Res* **28**, 1079-1089, doi:10.1101/gr.230615.117 (2018).

80 Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491, doi:10.1186/1471-2105-12-491 (2011).

81 Haas, B. J. *et al.* Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol* **9**, R7, doi:10.1186/gb-2008-9-1-r7 (2008).

82 Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol* **428**, 726-731, doi:10.1016/j.jmb.2015.11.006 (2016).

83 Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236-1240, doi:10.1093/bioinformatics/btu031 (2014).

84 Mitchell, A. L. *et al.* InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res* **47**, D351-D360, doi:10.1093/nar/gky1100 (2019).
85 Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* **20**, 238, doi:10.1186/s13059-019-1832-y (2019).
86 R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, Vienna, Austria, 2018).
87 Horton, P. *et al.* WoLF PSORT: protein localization predictor. *Nucleic Acids Res* **35**, W585-587, doi:10.1093/nar/gkm259 (2007).
88 Tardif, M. *et al.* PredAlgo: a new subcellular localization prediction tool dedicated to green algae. *Mol Biol Evol* **29**, 3625-3639, doi:10.1093/molbev/mss178 (2012).
89 Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**, 268-274, doi:10.1093/molbev/msu300 (2015).
90 Bushnell, B. *Bbtools Software Package*, <http://sourceforge.net/projects/bbmap> (2019).
91 Putnam, N. H. *et al.* Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* **317**, 86-94, doi:10.1126/science.1139158 (2007).
92 Baumgarten, S. *et al.* The genome of Aiptasia, a sea anemone model for coral symbiosis. *Proc Natl Acad Sci U S A* **112**, 11893-11898, doi:10.1073/pnas.1513318112 (2015).
93 Shinzato, C. *et al.* Using the Acropora digitifera genome to understand coral responses to environmental change. *Nature* **476**, 320-323, doi:10.1038/nature10249 (2011).
94 Prada, C. *et al.* Empty Niches after Extinctions Increase Population Sizes of Modern Corals. *Curr Biol* **26**, 3190-3194, doi:10.1016/j.cub.2016.09.039 (2016).
95 Shoguchi, E. *et al.* Draft assembly of the Symbiodinium minutum nuclear genome reveals dinoflagellate gene structure. *Curr Biol* **23**, 1399-1408, doi:10.1016/j.cub.2013.05.062 (2013).
96 Aranda, M. *et al.* Genomes of coral dinoflagellate symbionts highlight evolutionary adaptations conducive to a symbiotic lifestyle. *Sci Rep* **6**, 39734, doi:10.1038/srep39734 (2016).
97 Parkinson, J. E. *et al.* Gene Expression Variation Resolves Species and Individual Strains among Coral-Associated Dinoflagellates within the Genus Symbiodinium. *Genome Biol Evol* **8**, 665-680, doi:10.1093/gbe/evw019 (2016).
98 Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology* **34**, 525-527, doi:10.1038/nbt.3519 (2016).
99 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
100 Huson, D. H. *et al.* MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. *PLoS Comput Biol* **12**, e1004957, doi:10.1371/journal.pcbi.1004957 (2016).

Figures

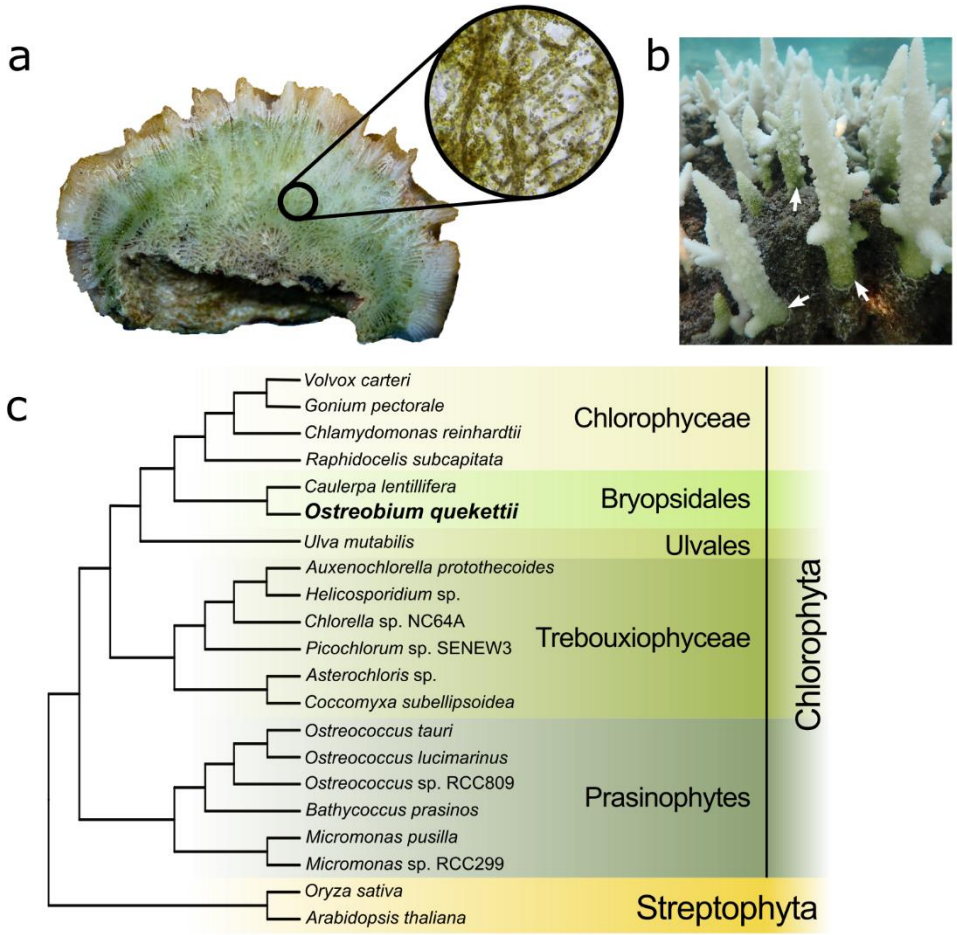

Figure 1. (a) Cross-section of *Paragoniastrea australensis* coral showing *Ostreobium* that inhabits the skeleton. The inset shows *Ostreobium* filaments after skeletal decalcification. **(b)** Bleached coral with evident *Ostreobium* bloom (indicated by white arrows). Photograph by Alexander Fordyce. **(c)** Phylogenetic tree of Viridiplantae (Streptophyta + Chlorophyta) showing the position of *Ostreobium quekettii*.

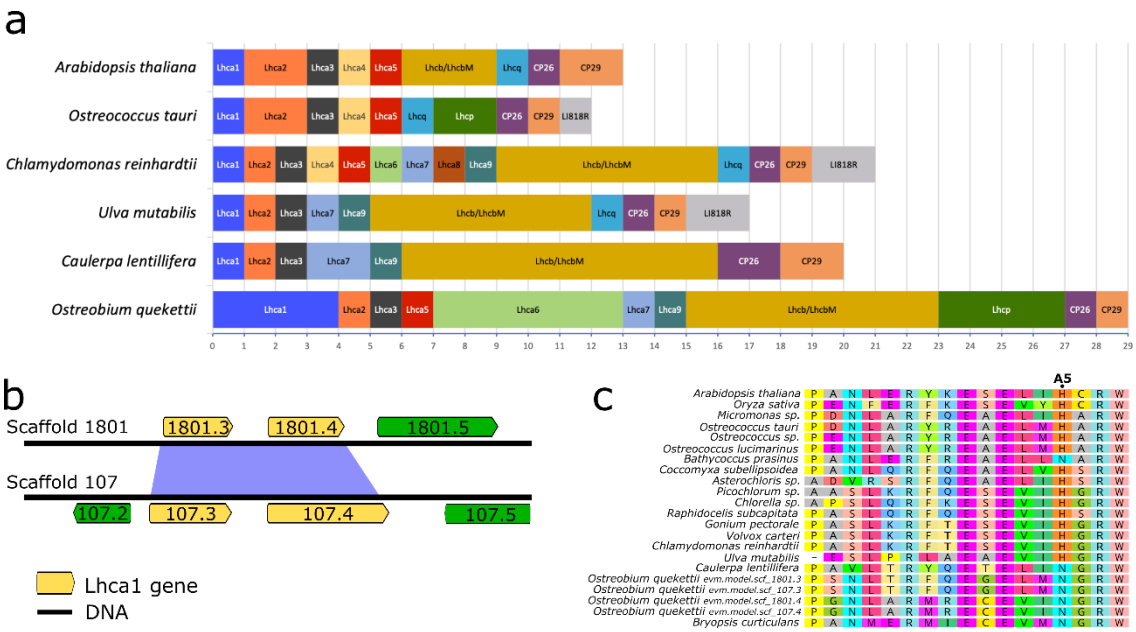

Figure 2. (a) Number of different light harvesting complex (LHC) proteins in *Ostreobium*, other green algae and *Arabidopsis thaliana*. **(b)** Synteny of *Lhca*1 copies in the *Ostreobium* genome. **(c)** Amino acid sequence comparison between Lhca1 proteins, showing asparagine (N) at the chlorophyll-binding residue A5 in *Ostreobium.*

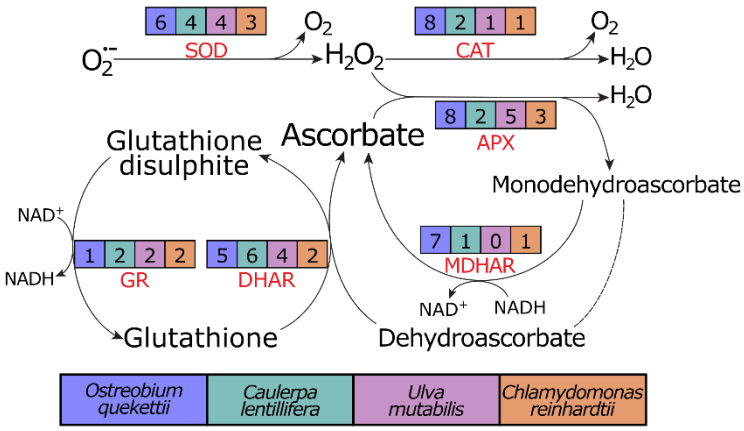

Figure 3. Simplified oxidative response pathway comparing the number of genes for enzymes found in the genomes from *Ostreobium* and some other green algae. Compared with *C. lentillifera*, *U. mutabilis* and *C. reinhardtii*, *Ostreobium* does have more copies of genes related to quick response to neutralize ROS, such as superoxide dismutase (SOD), catalase (CAT), ascorbate peroxidase (APX) and monodehydroascorbate reductase (MDHAR). DHAR, dehydroascorbate reductase. GR, glutathione reductase.

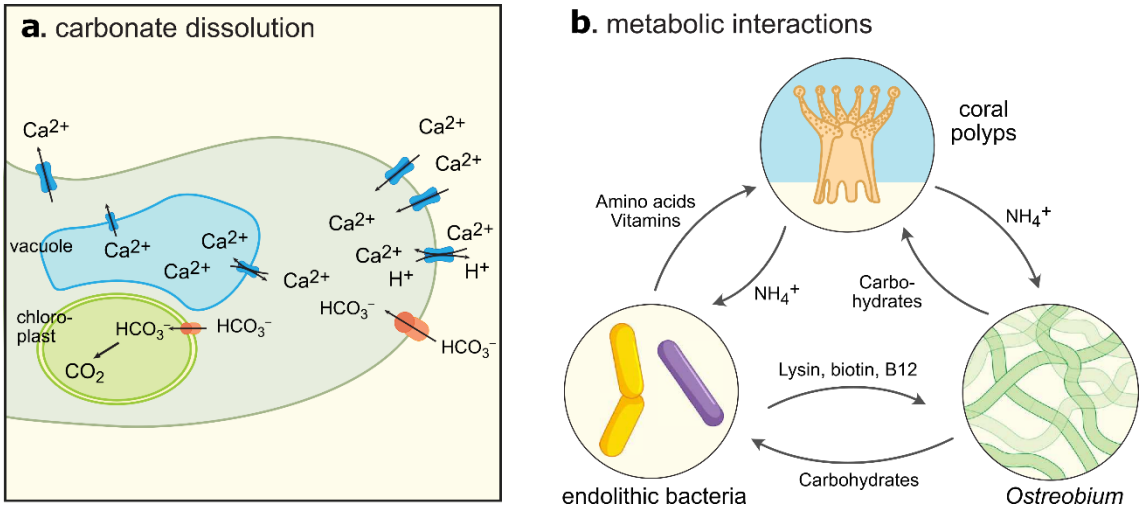

Figure 4. Roles of *Ostreobium* in the coral holobiont. **(a)** Potential mechanisms available to *Ostreobium* for excavation of the $CaCO_3$ skeleton of corals near the growing tip of *Ostreobium* filament. **(b)** Possible interactions between members of the holobiont derived from genome sequence data.