BMC Genomics

RESEARCH Open Access

The genome of a giant clam zooxanthella (*Cladocopium infistulum*) offers few clues to adaptation as an extracellular symbiont with high thermotolerance

Raúl A. González-Pech[1,2*], Jihanne Shepherd[3,4], Zachary L. Fuller[3], Todd C. LaJeunesse[2,5] and John Everett Parkinson[1*]

Abstract

Background *Cladocopium infistulum* (Symbiodiniaceae) is a dinoflagellate specialized to live in symbiosis with western Pacific giant clams (Tridacnidae). Unlike coral-associated symbionts, which reside within the host cells, *C. infistulum* inhabits the extracellular spaces of the clam's digestive diverticula. It is phylogenetically basal to a large species complex of stress-tolerant *Cladocopium*, many of which are associated with important reef-building corals in the genus *Porites*. This close phylogenetic relationship may explain why *C. infistulum* exhibits high thermotolerance relative to other tridacnid symbionts. Moreover, past analyses of microsatellite loci indicated that *Cladocopium* underwent whole-genome duplication prior to the adaptive radiations that led to its present diversity.

Results A draft genome assembly of *C. infistulum* was produced using long- and short-read sequences to explore the genomic basis for adaptations underlying thermotolerance and extracellular symbiosis among dinoflagellates and to look for evidence of genome duplication. Comparison to three other *Cladocopium* genomes revealed no obvious over-representation of gene groups or families whose functions would be important for maintaining *C. infistulum*'s unique physiological and ecological properties. Preliminary analyses support the existence of partial or whole-genome duplication among *Cladocopium*, but additional high-quality genomes are required to substantiate these findings.

Conclusion Although this investigation of *Cladocopium infistulum* revealed no patterns diagnostic of heat tolerance or extracellular symbiosis in terms of overrepresentation of gene functions or genes under selection, it provided a valuable genomic resource for comparative analyses. It also indicates that ecological divergence among *Cladocopium* species, and potentially among other dinoflagellates, is partially governed by mechanisms other than gene content. Thus, additional high-quality, multiomic data are needed to explore the molecular basis of key phenotypes among symbiotic microalgae.

Keywords Dinoflagellate, Genome duplication, Symbiodiniaceae, Symbiosis, Tridacnidae

*Correspondence:
Raúl A. González-Pech
rag5851@psu.edu
John Everett Parkinson
jparkinson@usf.edu
[1]Department of Integrative Biology, University of South Florida, Tampa, FL 33620, USA
[2]Department of Biology, Pennsylvania State University, University Park, PA 16802, USA
[3]Department of Biological Sciences, Columbia University, New York, NY 10027, USA
[4]Miller School of Medicine, University of Miami, Miami, FL 33136, USA
[5]Institute of Energy and the Environment, The Pennsylvania State University, University Park, PA 16802, USA

Background

Dinoflagellates in the family Symbiodiniaceae–commonly known as 'zooxanthellae'–establish mutualisms with corals and many other invertebrates [1]. These photosynthetic dinoflagellates provide beneficial organic nutrients and oxygen to their hosts; in turn, the hosts provide inorganic nutrients and a sheltered habitat that maximizes light exposure [2–5]. Such associations facilitate host growth in nutrient-poor waters, and calcification in the case of reef-building corals, making them foundational to the existence of coral reef ecosystems worldwide. Given the ecological abundance and importance of symbiodiniacean mutualisms, along with their sensitivity to thermal anomalies, ongoing research seeks to understand the cellular and molecular bases of these endosymbioses, especially in the context of global climate change (e.g., [6–8]). To this end, comparative genomics can provide further insight [9, 10].

For coral-Symbiodiniaceae associations, where the symbiont cell resides within a host cell vacuole, mechanisms that facilitate successful mutualisms include host recognition [11], phagosome/lysosome arrest [12, 13], and metabolic exchange [14], among others [15]. The genomes of symbiotic species exhibit possible adaptations important to intracellular symbiosis, such as an extensive repertoire of transmembrane transporters [16]; and structural changes to the genome potentially as a byproduct of an intracellular lifestyle, such as an increased frequency of duplicated regions, structural rearrangements, mobile elements, and pseudogenes [9, 17]. However, not all symbiodiniacean mutualisms are intracellular. A minority are extracellular, where symbionts reside outside of the host cell within specialized tubular structures [5, 18–20], and their genomic characteristics have yet to be explored with this ecophysiological context in mind.

The family Symbiodiniaceae encompasses multiple genera, each containing numerous species with distinct physiologies and unique ecological niches [21]. The genus *Cladocopium* (formerly known as *Symbiodinium* Clade C) is the most evolutionarily derived and genetically diverse genus of symbiotic dinoflagellates. It also has the broadest geographic and ecological distribution [22–25]. One of the few *Cladocopium* species that grows in artificial culture media is *Cladocopium infistulum* (*ITS2* type C2; [26]). This specialized symbiont occurs only in giant clams (Tridacnidae) in the western tropical Pacific Ocean. Unlike the intracellular symbionts of other hosts, this species lives extracellularly within digestive diverticula, which are fine, densely branching tubules in the clam's mantle [5, 18, 20]. The siphonal mantle of giant clams evolved these structures and enlarged lobes that function like "solar panels" to facilitate symbiont photosynthesis [5, 20, 27].

Cladocopium infistulum dominates clams that occur in warm lagoonal habitats of Palau, indicating that it is adapted to high temperature [26, 28]. Phylogenetically, it is positioned at the base of the C15 radiation [22]; this group consists of multiple species that associate with Indo-Pacific *Porites* corals, which are known to be relatively heat-tolerant [29]. There is broad speculation regarding the genetic underpinnings of symbiodiniacean thermotolerance [30, 31], but mechanisms related to general stress response appear to play a central role, especially among members of the genus *Cladocopium* [32]. Although most dinoflagellates are haploid [33], isoclonal cultures of *Cladocopium* show consistent patterns of biallelic microsatellite loci [34, 35]. Such patterns suggest either large partial duplication or whole-genome duplication (WGD) in the ancestral lineage of this genus. Such a duplicated genome may have contributed to the eventual ecological and geographic dominance of *Cladocopium*.

The unique biology and phylogenetic position of *C. infistulum* justify genome sequencing and comparative analyses to better understand aspects of Symbiodiniaceae ecology and evolution. Here, a draft genome assembly of *C. infistulum* (culture RT-203) was generated and compared to congeners with different life histories (extracellular and intracellular), hosts (clams and corals), and physiologies (thermosensitive and thermotolerant). The analyses focused on two key questions: is there overrepresentation of genes in cellular processes indicative of extracellular symbiosis and thermotolerance? And do patterns of synteny and paralog divergence provide evidence for genome duplication?

Methods

The experimental protocol is summarized here briefly; for a complete description of all steps and programs, see the Supplementary Materials and Methods. The associated code is available at https://github.com/RaulGoch/C.infistulum_genome.

Microalgal culture and DNA extraction

Cladocopium infistulum strain RT-203 (Fig. 1) was originally isolated from a giant clam (*Hippopus hippopus*) in Palau in 1982 and has been maintained in continuous culture via bimonthly transfers to fresh ASP8-A media in an incubator held at 26 °C and illuminated by white fluorescent light (4∼100 µmol photons m^{-2} s^{-1} under a 12:12 h light: dark cycle). To acquire high molecular weight DNA for genome sequencing, a cesium-chloride extraction was performed as described by [36].

Genome sequencing and assembly

Whole-genome sequencing was conducted at the Pennsylvania State University Genomics Core Facility. Short-read sequence data (2×149 bp reads, 300 bp insert

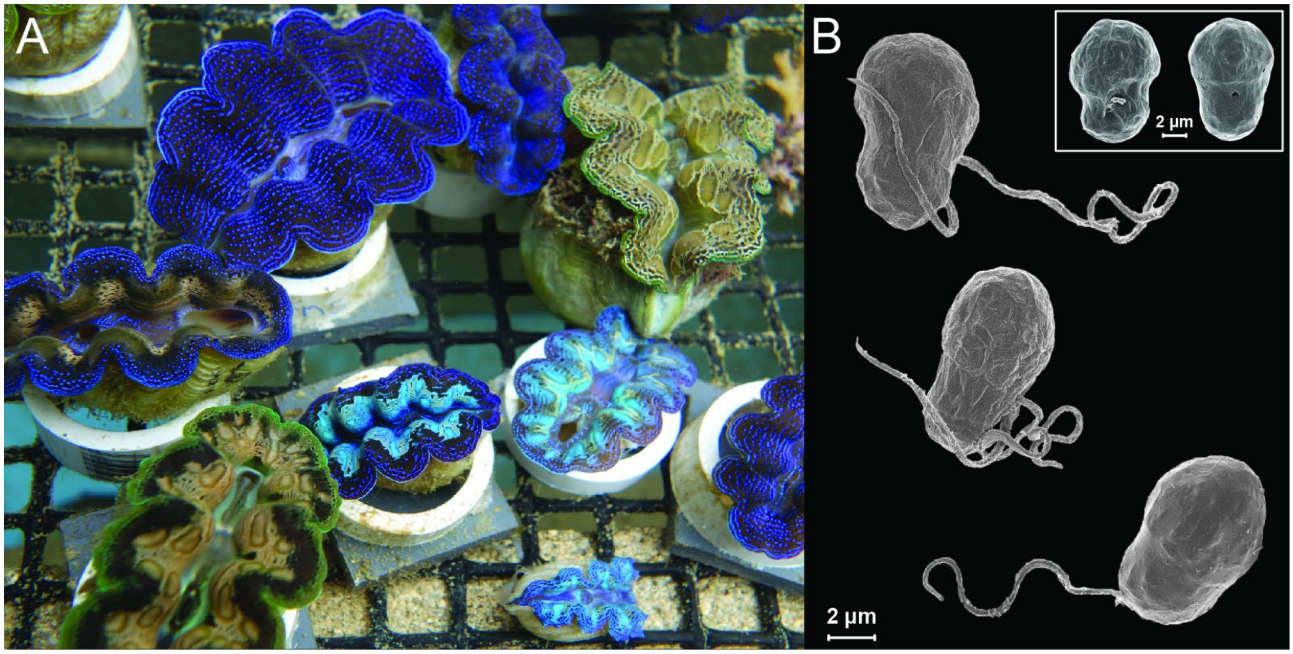

Fig. 1 *Cladocopium infistulum* is a symbiont of giant clams. **(A)** Color variants of the smallest giant clam species, *Tridacna crocea*, from the rock islands habitats of Palau. **(B)** Scanning electron microscopy (SEM) images of motile cells of *C. infistulum*. Inset shows the ventral and dorsal sides of this dinoflagellate species from [26]. Clam images were taken by Allison M. Lewis, Pennsylvania State University, and SEM micrographs were taken by Sung Yeon Lee, Seoul National University

length) were generated using paired-end libraries on the Illumina NextSeq 550 platform; long-read sequence data were generated on a PacBio Sequel system (Supplementary Table 1). After cleaning and quality control, short and long reads were combined and run through two different programs to create two hybrid de novo genome assemblies, which were then merged.

Gene prediction

For consistency in genome comparisons, gene prediction was completed using the same comprehensive gene prediction workflow that has been implemented in most Symbiodiniaceae genome assemblies to date [37], available at https://github.com/TimothyStephens/Dinoflagellate_Annotation_Workflow. In addition to ab initio prediction, the workflow incorporated transcript evidence taken from available transcriptome data from other members of *Cladocopium* (Supplementary Table 2). The workflow used a series of programs to identify and remove repeats and transposable elements prior to gene prediction. It also contained modifications to recognize unconventional dinoflagellate intron-exon boundaries [38]. As input for gene prediction, the pipeline used a custom repeat library, protein sequences from the curated Swiss-Prot database, previously predicted Symbiodiniaceae protein sequences from other studies (Supplementary Table 3), and a high-confidence set of genes derived from transcript alignment for training. The output from multiple programs was then weighted, and only those genes predicted with transcript evidence or by at least two ab initio predictors were included in the final gene models.

Removal of microbial and organellar sequences

Putative microbial contaminant sequences were removed from the genome assembly using a blob plot (Supplementary Fig. 1), which visualized taxonomic assignment, G+C content, and read coverage. After inspecting the blob plot, sequences were removed if they were annotated as belonging to bacteria or viruses or if both GC content>50% and average read coverage>100-fold, as they were unlikely to be of Symbiodiniaceae origin. To identify and remove chloroplast sequences, the genome was searched using *Cladocopium* C3 minicircle sequences as queries [39]; no similar BLASTn hits were found. Mitochondrial sequences were identified and removed by screening with putative *Cladocopium* mitochondrial sequences retrieved from the NCBI database (Supplementary Table 4). With the aim to recover sequences after removal, contaminant and organellar sequences were further scrutinized by looking at the number of predicted genes they contained (Supplementary Fig. 2). Eight scaffolds with mitochondrial BLAST hits were retained as part of the nuclear genome because they also contained other non-mitochondrial genes.

Completeness assessment, annotation of gene functions, and pseudogene screening

Completeness of the *C. infistulum* and other *Cladocopium* genome assemblies [10, 40, 41] was assessed using BUSCO v5.7.1 [42]. Gene function was annotated based on sequence similarity searches against the manually curated Swiss-Prot protein database; those with no hits were searched against the TrEMBL database. The best hit with associated Gene Ontology terms was used to annotate the query protein. Pfam domains and KEGG Orthology terms were assigned to each protein using their respective search tools. A genome-wide search for pseudogenes was performed following an approach previously implemented for Symbiodiniaceae [9, 43], in which the predicted proteins from each genome were queried against the corresponding gene-masked genome sequences as targets. Matches with ≥75% identity were treated as pseudogene regions and matches on the same strand and within 1 kbp were considered as part of the same pseudogene.

Orthogroup analyses and selection tests

To identify orthogroups from the *C. infistulum* and other Symbiodiniaceae genomes (Supplementary Table 5), inferred proteins were assessed pairwise for sequence similarity. The functions encoded in orthogroups and single-copy genes exclusive to *C. infistulum* were retrieved from the annotations described above. Orthogroup expansion/reduction in *C. infistulum* relative to other *Cladocopium* genomes was assessed pairwise using an adjusted two-tailed Fisher's exact test. Following a pipeline that has previously been applied to alveolates [44], likelihood-ratio tests between branch models were used to identify protein-coding genes under selection in *C. infistulum* relative to the other Symbiodiniaceae (Supplementary Table 5). The pipeline included sequence alignment, construction of a Maximum Likelihood tree for each orthogroup, and calculation of d_N/d_S ratios (where $d_N/d_S > 1$ is indicative of positive selection). These ratios were also calculated to test for directional selection pairwise between sets of extracellular species (*C. infistulum* and *Cladocopium* sp. Y103) and intracellular species (*Cladocopium proliferum* and *Cladocopium* C15).

Genome size estimation and analyses of genome duplication and synteny

Genome size and sequence-read coverage were estimated based on *k*-mers (i.e., all possible substrings of length *k* within the DNA sequence), specifically the *k*-mer frequency distribution from the trimmed short read-pairs, single-end reads, and corrected long reads (Supplementary Table 6). Evidence of whole-genome duplication (WGD) was assessed first by comparing fitted *k*-mer models between the two distinct ploidies (1 for haploid and 2 for diploid), and second by comparing synonymous substitution rates (K_S) of gene paralogs. As a preliminary exploration of genome duplication, the unmasked *C. infistulum* assembly was subject to whole-genome sequence alignment against itself, followed by a segmental duplication analysis using the program SEDEF [45]. Then genes that overlapped with the duplicated regions were identified and analyzed for functional enrichment of GO terms and Pfam domains. The analysis used adjusted one-tailed Fisher's exact tests, in which annotated genes overlapping with duplicated regions were compared against all annotated *C. infistulum* genes as background. Additional searches were performed to detect intragenomic collinear syntenic blocks of at least five genes in the same order as well as genes repeated *in tandem*.

Assessment of genome divergence across *Cladocopium*

To explore genome similarity and divergence within the genus, genome data from four *Cladocopium* species (Supplementary Table 5) were used: *C. infistulum* ('Cin'; this study), *C. proliferum* ('Cpr') [10, 37], *Cladocopium* C15 ('C15') [40], and *Cladocopium* sp. Y103 ('Csp') [41]. These species encompass a range of symbiont niches and hosts. *Cladocopium proliferum* is a host generalist among scleractinian corals in the Indo-Pacific [25], and the strain SCF055-01 was originally recovered from an *Acropora tenuis* coral at Magnetic Island, Australia, and has shown resilience to heat stress under experimental conditions [46]. Members of *Cladocopium* C15, which encompasses several species, broadly associate with *Porites* corals in the Indo-Pacific [22]. *Cladocopium* C15 commonly confer thermotolerance to their coral hosts [29, 47] and they have not been successfully isolated and cultured [48]. The dataset used here was produced as a metagenome assembly from *Porites lutea* collected in Orpheus Island, Australia [40]. *Cladocopium* sp. Y103 was isolated from an infaunal cardiid clam (*Fragum* sp.) in Okinawa, Japan [41], where it exists as an extracellular symbiont inhabiting the innermost part of the mantle, and accesses light through an opening in the host's shell [49]. In the gradient of heat tolerance, these taxa appear to be at the upper range, most likely with *C. infistulum* at the top, followed by *C. proliferum* and then the other species, though additional physiological experiments would need to be performed to confirm these rankings.

Whole-genome sequence alignment of the four assemblies was conducted. Then, filtered short reads of *C. infistulum* were mapped against the other species assemblies, and against itself, to corroborate sequence divergence. Genome sequence divergence within *Cladocopium* was assessed by implementing an alignment-free approach based on *k*-mers [50], including genomes of *Symbiodinium* as an outgroup [9, 41, 51]. The analysis found $k=25$ to be suitable based on the uniqueness level among

the assessed genomes (Supplementary Fig. 3). Finally, mutational bias and codon usage were compared in the full-length predicted coding sequences of all four *Cladocopium* genomes.

Targeted characterization of genes related to host-symbiont interactions, light response, and stress response

Confidence intervals (95%) from the *t*-distribution were estimated for each GO term across all four *Cladocopium* genomes to identify functions overrepresented in different niches of these species. The symbiosome is a specialized compartment where intracellular symbionts reside, which consists of both host- and symbiont-derived membranes [15]. The number of orthologs of putative symbiosome membrane proteins found in extracellular and intracellular *Cladocopium* genomes were compared by querying the predicted protein sequences against algal and microbial symbiosome sequences [52] (Supplementary Table 7). Additional searches were performed for genes with GO terms related to functions that are relevant to host-symbiont interactions (Supplementary Table 8), including nitrate/nitrite reduction and transport of sterol, ammonium, nitrate, urea, amino acids, oligopeptides, sulfate, bicarbonate, and glucose [11, 15, 53, 54]. Searches were also made for orthologs of genes implicated in light response that have been characterized in symbiodiniacean symbionts of clams ([55–61]; Supplementary Table 9), as well as GO terms associated with response to stress ([9]; Supplementary Table 10). Representation of GO terms related to host-symbiont interactions and stress response were assessed through heatmaps and hierarchical clustering by species.

Results

The *Cladocopium infistulum* genome is typical of the genus

The genome assembly of *Cladocopium infistulum* RT-203 had a total length of 0.9 Gbp, which is ~45–60% of the total estimated size of 1.5–2 Gbp (Table 1 and Supplementary Table 6). The assembly combined both short- and long-read sequence data to achieve ~22-fold coverage. Its contiguity (contig N50=66,314 bp) and completeness (BUSCO hits=75.61%; Supplementary Table 11) were comparable to the other three *Cladocopium* spp. genome assemblies (Fig. 2A; Table 1). A total of 19,834 protein-coding genes were predicted, of which 15,752 (79.4%) were annotated with at least one type of function (i.e., UniProt, GO, KEGG and/or Pfam; Supplementary Fig. 4), and 11,800 (59.5%) were supported by transcript data (Supplementary Table 12). Of the predicted genes, 438 were single-copy and unique to *C. infistulum*, while the remaining 19,396 were assigned to 11,617 orthogroups (Supplementary Fig. 5). Of these orthogroups, 37 (containing 99 genes) were unique to *C. infistulum* as they could not be found in other Symbiodiniaceae genomes. Only 130 out of the 537 *C. infistulum*-specific genes could be annotated with any function. In no instance were orthogroups significantly smaller (reduced) relative to the three other species, but three were larger (expanded; Supplementary Fig. 6). One of the expanded orthogroups contained genes coding mostly for kinase activity, and two of them encompassed constituents of the kinesin complex and microtubules (Supplementary Table 13).

Pairwise whole-genome alignment between all four *Cladocopium* genomes revealed appreciable sequence divergence (Fig. 2B), which was confirmed with the alignment qualities of short reads (Supplementary Fig. 7). A larger portion of the *C. infistulum* genome aligned to the *Cladocopium* C15 genome than to any other species. This similarity is reflected in the highly supported *k*-mer phylogeny (Fig. 2C; Supplementary Figs. 8–9), which is in agreement with phylogenetic reconstructions based on the sequence alignments of several conserved nuclear and plastid genes [25]. Most genomes shared roughly equivalent intron and intergenic region lengths (Fig. 2D). The exception was *C. infistulum*, where intergenic regions appeared longer, though perhaps only for methodological reasons (e.g., incorporation of long reads into this assembly and not the others, and reliance on congeneric transcriptome data for gene prediction, which

Table 1 Genome assembly statistics of *Cladocopium infistulum* and other Symbiodiniaceae. Genome sizes for *C. infistulum* and *Cladocopium* sp. Y03 were calculated in this study. Details from the *Symbiodinium* genomes in Fig. 2C are provided for comparison. 'Cin': *C. infistulum*, 'Cpr': *C. proliferum*, 'C15': *Cladocopium* C15, 'Csp', *Cladocopium* sp. Y03, 'Smi': *S. microadriaticum*, 'Sne': S. *necroappetens*, 'Sli': *S. linucheae*, 'Str': *S. tridacnidorum*, 'N.D.': not determined

Metric	Cin	Cpr	C15	Csp	Smi	Sne	Sli	Str
Number of scaffolds	11,092	41,235	34,589	22,508	189	104,583	37,772	16,175
Total assembly length (Gbp)	0.90	0.97	0.62	0.70	0.81	0.77	0.69	0.76
Scaffold N50 length (bp)	160,719	91,131	50,690	247,555	7,224,000	14,528	58,075	132,488
Contig N50 length (bp)	66,314	6,142	20,331	38,432	34,595	11,420	11,147	52,520
Gaps (%)	5.90	13.93	0.45	4.37	7.70	0.56	1.35	1.31
Predicted genes	19,834	39,006	22,508	33,421	29,725	35,672	32,053	25,808
Estimated genome size (Gbp)	1.49-2.00	1.19	N.D.	0.87	1.10–1.40	1.00	0.91	N.D.
Assembled fraction (%)	44.38–60.10	81.42	N.D.	80.56	57.73–73.48	76.26	75.96	N.D.

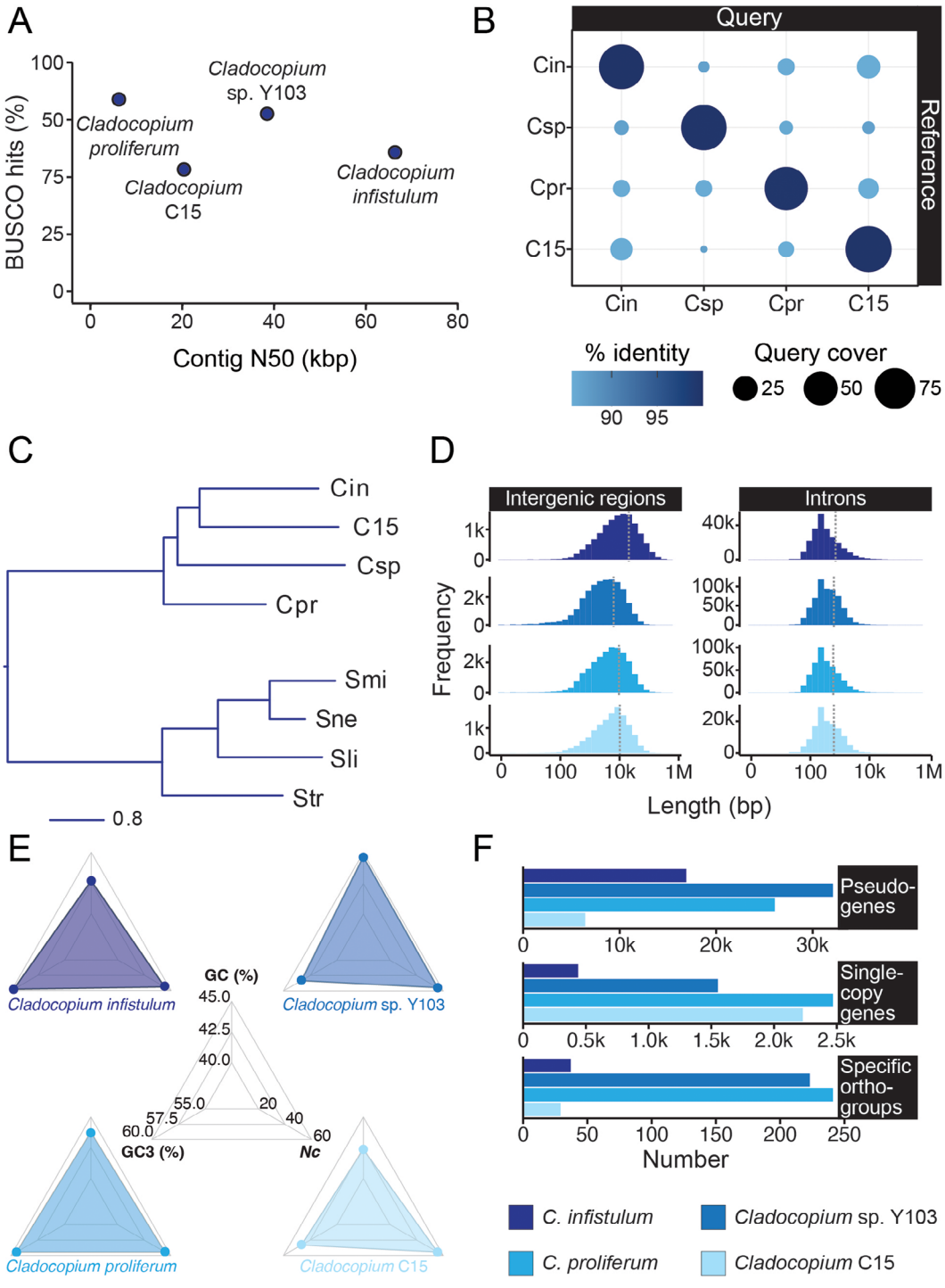

Fig. 2 The *C. infistulum* genome assembly is of similar quality compared to other *Cladocopium* assemblies. **(A)** Comparison of the *C. infistulum* genome with other available *Cladocopium* genomes in terms of contiguity (*x*-axis) and completeness (*y*-axis). **(B)** Pairwise alignments between all four *Cladocopium* genomes. The color and size of each dot represent the average percentage identity and fraction of the query aligned to the reference, respectively, according to the bottom legend. **(C)** Original NJ tree built with shared and unique 25-mers. All nodes are supported by 100% of the jackknife pseudoreplicates (Supplementary Fig. 9); scale shows D_2^S distance. *Symbiodinium* genomes were included as an outgroup and a reference of divergence in another genus. **(D)** Length distribution of introns and intergenic regions in the four genomes; the dashed gray line shows the mean in each histogram. **(E)** Three-dimensional radar plots showing genomic G+C content (GC), G+C content in third codon position of coding sequences (GC3) and effective number of codons used (*Nc*) per genome; the center plot shows the scales for the three axes. **(F)** Number of pseudogenes (top), species-specific single-copy genes (middle), and orthogroups (bottom). Colors in **(D)**, **(E)** and **(F)** represent each of the four genomes according to the bottom right legend

possibly lowers gene density). All *Cladocopium* genomes exhibited low biases in nucleotide composition and codon usage, with similar values for the effective number of codons used in a gene ($Nc \approx 60$; Fig. 2E). *Cladocopium* C15 showed lower genomic and third-codon position G+C content compared to the other species. No consistent pattern was found in the number of pseudogenes with respect to intracellular *versus* extracellular species (Fig. 2F). However, intracellular species appeared to have more unique single-copy genes than extracellular species. Both *C. infistulum* and *Cladocopium* C15, which exhibit extra- and intracellular host relationships, respectively, showed low numbers of genome-specific orthogroups (Fig. 2F).

Based on d_N/d_S ratios, there were several genes under positive selection in *C. infistulum* when compared separately against the other three *Cladocopium* (42 against *Cladocopium* C15, 110 against *Cladocopium* sp. Y103, and 157 against *C. proliferum*). However, none of these genes were shared across all pairwise comparisons (Supplementary Fig. 10). Likewise, no genes under positive selection were common to contrasts of intracellular *versus* extracellular species (Supplementary Fig. 10). When compared to available Symbiodiniaceae data sets (Supplementary Table 5), including genomes from other genera, no genes under positive selection were detected using the branch model test. However, a total of 140 *C. infistulum* genes appeared to be under negative selection; of these, ion transport was the most represented biological process.

Genes related to host-symbiont interactions, light response, and stress response

No GO terms were significantly overrepresented simultaneously in genomes of intracellular or extracellular symbionts (Supplementary Table 14). Comparing putative symbiosome membrane gene orthologs (Supplementary Fig. 11 and Supplementary Table 7) across *Cladocopium* spp. indicated that, on average, the abundance of genes coding for these membrane interaction proteins were similar in the genomes of intracellular and extracellular symbiont species. However, when considering higher-order GO annotations rather than absolute gene numbers, most gene functions of relevance to host-symbiont interactions were underrepresented in the genome of the extracellular *C. infistulum*, whereas the majority of these GO terms were overrepresented in the genome of the intracellular *C. proliferum* (Fig. 3 and Supplementary Table 8); no clear pattern was observed for these GO terms in *Cladocopium* C15 and *Cladocopium* sp. Y103. Notably, nitrite and nitrate reduction were generally less represented in extracellular symbionts and glucose transport was more represented in intracellular symbionts. The dendrogram resulting from the hierarchical clustering of the genomes based on these GO terms recapitulates the topology of the phylogenomic tree (Fig. 2C).

Comparing gene orthologs implicated in light response also revealed no consistent differences in abundance across species (Supplementary Fig. 12 and Supplementary Table 9). While there were no clusters clearly separating species by symbiont niche, the representation of GO terms related to stress response appears to be most alike between extracellular symbionts (Fig. 4 and Supplementary Table 10), with a general trend for underrepresentation of terms related to apoptosis, protein folding, glutathione metabolism, and response to oxidative and UV stress. Conversely, these GO terms were most represented in *C. proliferum* except for '*cellular response to heat*', which was most represented in *Cladocopium* C15.

Duplicated genomic regions are prevalent in *C. infistulum*

The *k*-mer frequency distribution resembled the traditional unimodal shape of haploid genomes (Fig. 5A). Typically, haploid genomes that have undergone WGD exhibit a bimodal distribution of *k*-mer frequencies that more closely resembles the two-peak pattern of a diploid genome, with one peak representing homozygous regions and another representing heterozygous regions (the *k*-mer frequency distribution from the duplicated genome of the dinoflagellate *Polarella glacialis* [62] is shown for comparison; Fig. 5B). Additionally, the haploid model better fit the *k*-mer frequency distribution than the diploid model (Fig. 5A and Supplementary Fig. 13), and under the diploid model the genome's estimated heterozygosity of duplicated regions was close to zero. A second analysis explored synonymous substitution rates (K_S) in gene paralogs, revealing only a single peak consisting of low-divergence paralogs near $K_S = 0$ in *C. infistulum* (Fig. 5C) and the other *Cladocopium* (Supplementary Fig. 14), which is a common feature in most genome assemblies across a variety of organisms [63]. Under a WGD scenario, a second peak at $K_S = 1–2$, indicating widespread and simultaneous gene duplications having occurred at about the same time, would be expected. Altogether, these results indicate that the genome of *C. infistulum* has not undergone WGD.

Signals of genome duplication were searched for by performing a reciprocal alignment of all scaffolds/contigs (Supplementary Fig. 15); the large number and lengths of repetitive hits indicate the presence of numerous duplicated regions (1–20 kbp; Supplementary Fig. 16). Further analysis with SEDEF, which detects duplicated genomic regions arising from segmental duplication, estimated that duplicated regions occupied 210.70 Mbp (23.5%) of the genome assembly. Most duplicated regions had only one additional copy (Fig. 5D), spanned from 1 to 20 kbp (Fig. 5E), and featured low Kimura divergence scores (Fig. 5F). In total, 6,657 genes (33.6% of all genes) were

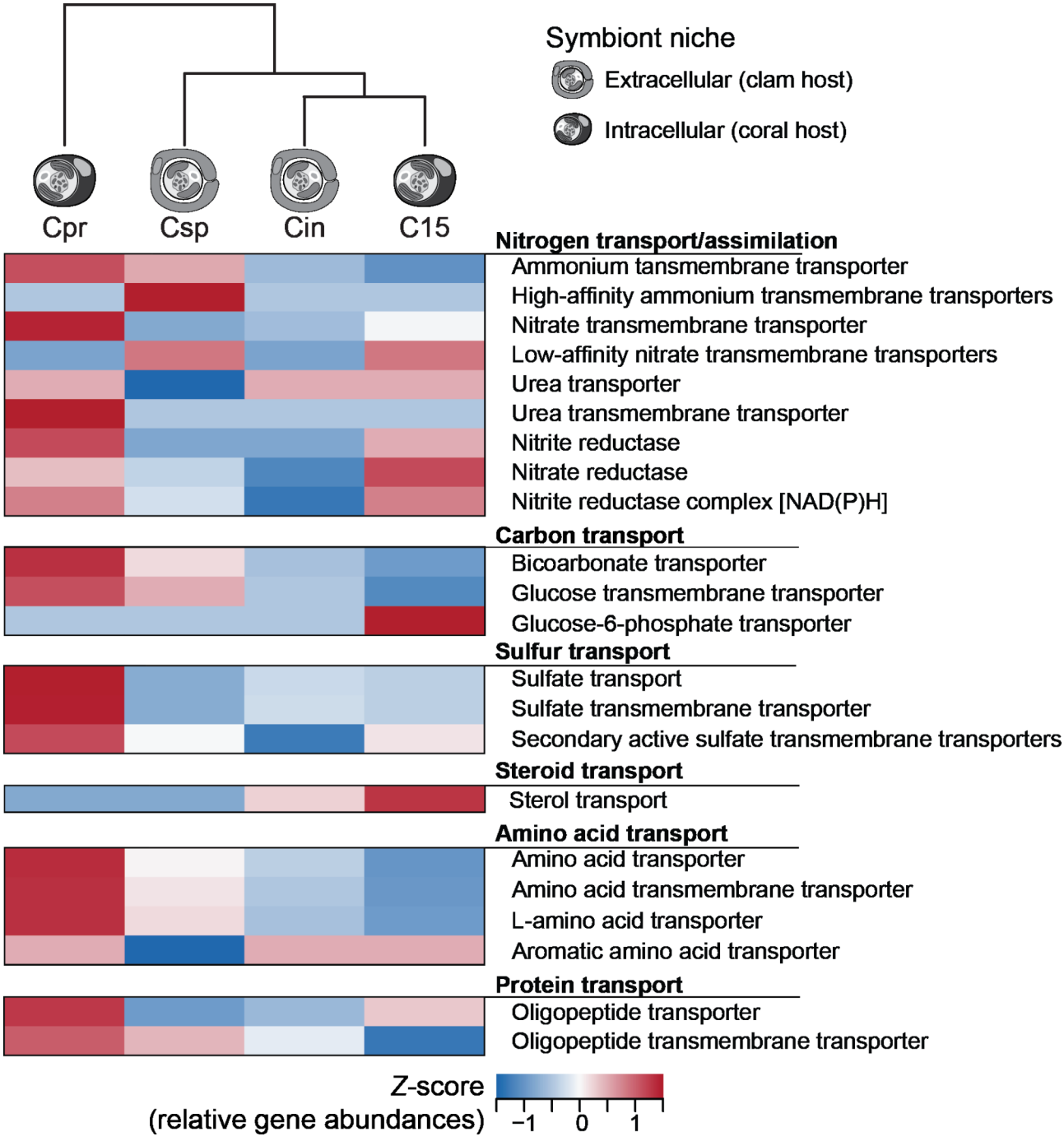

Fig. 3 Representation of gene functions in four *Cladocopium* important to nutrient exchange in the mutualism. Heatmap based on gene content of GO terms relevant to different nutrient transfer processes (Supplementary Table 8). Gene counts (rows) were *Z*-transformed, i.e., scaled to have mean zero and one standard deviation. The similarity dendrogram above the heat map was produced by complete-linkage hierarchical clustering of Euclidean distances

found within these duplicated regions. Gene functions enriched in duplicated regions were related to mobile elements, protein domain repeats, and photosynthesis (Supplementary Table 15). Most of the 23,363 duplicated regions (20,276) overlapped with interspersed repeats (often associated with mobile elements; Supplementary Fig. 17). The high abundance of a single additional copy of duplicated regions (Fig. 5D) and their similar divergence time (Fig. 5F) indicate a single duplication event. No duplicated gene blocks or arrays of genes *in tandem* were identified.

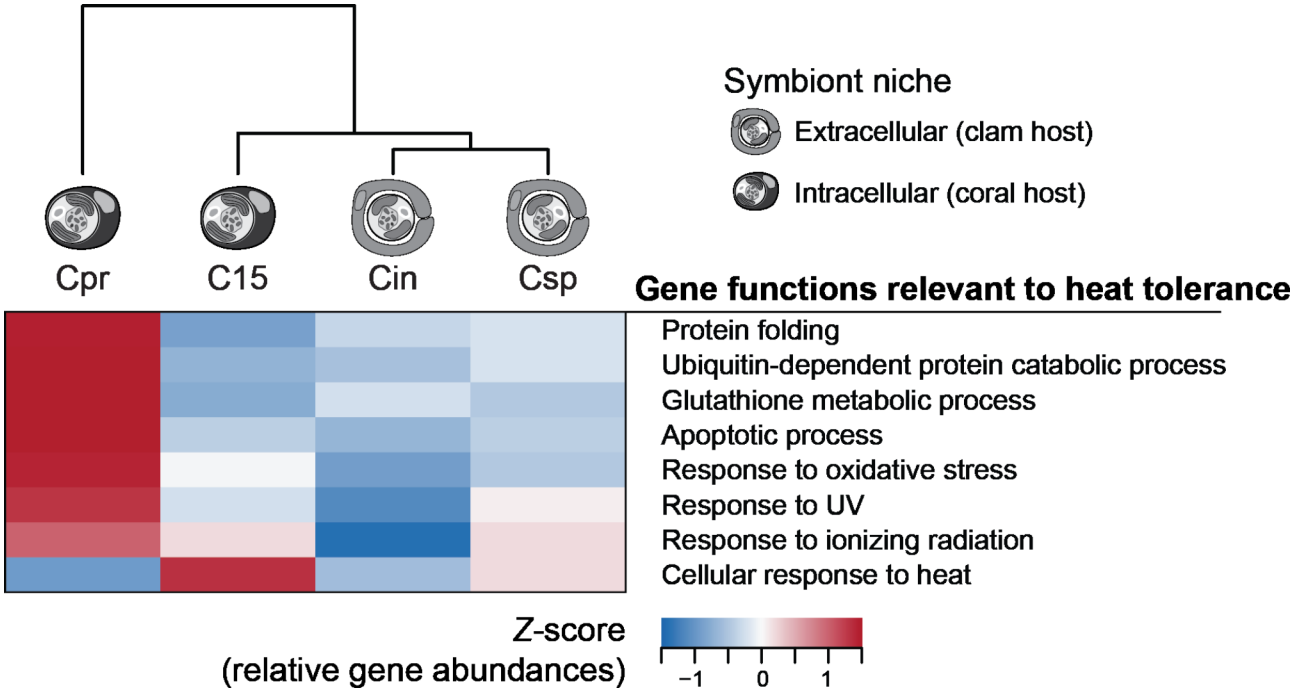

Fig. 4 Representation of genes important in thermal stress response shows no clear phylogenetic or niche patterns. Heatmap based on gene counts of relevant GO terms (Supplementary Table 10). Gene counts (rows) were *Z*-transformed, i.e., scaled to have mean zero and one standard deviation. The similarity dendrogram above the heat map was produced by complete-linkage hierarchical clustering of Euclidean distances

Discussion

The characterization of a fourth *Cladocopium* genome broadens new opportunities to identify genes, gene functions, and genomic changes implicated in ecological and evolutionary diversification within this large genus of symbiotic dinoflagellates. Despite relatively low sequence coverage, incorporating both short- and long-read sequence data into the *Cladocopium infistulum* genome assembly provided contiguity and completeness comparable to the other available *Cladocopium* assemblies (Fig. 2A; Table 1). While preliminary, these findings indicate that there are surprisingly few genomic differences between species with intra- and extra-cellular lifestyles (Fig. 3 and Supplementary Table 8), meaning that *Cladocopium* are highly versatile and/or the intra- and extra-cellular endosymbiotic environments examined here lack sufficient physiological differences to greatly influence microalgal genome evolution. Moreover, high thermal tolerance may have distinct genomic underpinnings across Symbiodiniaceae, even within the same genus (Fig. 4 and Supplementary Table 10), which are discussed below. The identification of extensive duplicated regions, most likely arising from a single event in the past, is not necessarily an indicator of whole-genome duplication (Fig. 5). These comparisons contribute to a growing understanding of the substantial genomic and adaptive differences even among closely related symbiodiniacean taxa. Note that, owing to low sequencing depth (~22-fold; Fig. 5A), partial representation of the total genome size in the assembly (~45–60%; Table 1), and gene prediction limited to transcriptome data from other *Cladocopium* species rather than a *C. infistulum* isolate, results should be interpreted conservatively. Some combination of these biases likely impacted the other Symbiodiniaceae assemblies used in the analyses, as these are common issues in comparative genomics of non-model organisms.

Representation of gene functions associated with host-symbiont interactions recapitulates the *Cladocopium* phylogeny, not the symbiont niche

Of the four *Cladocopium* species with sequenced genomes, two of them (*C. infistulum* and *Cladocopium* sp. Y103) live extracellularly when in symbiosis with their animal hosts [26, 64]. While some host-symbiont interaction genes were underrepresented in their genomes, these differences were not consistent for most other genes in this category (Fig. 3), and hierarchical clustering of GO terms did not group these species together as most similar. The concordant topology of the phylogenomic tree (Fig. 2C) and the hierarchical clustering dendrogram (Fig. 3) indicates that evolutionary relatedness between *Cladocopium* genomes, rather than symbiont niche, has a larger effect on the representation of these gene functions.

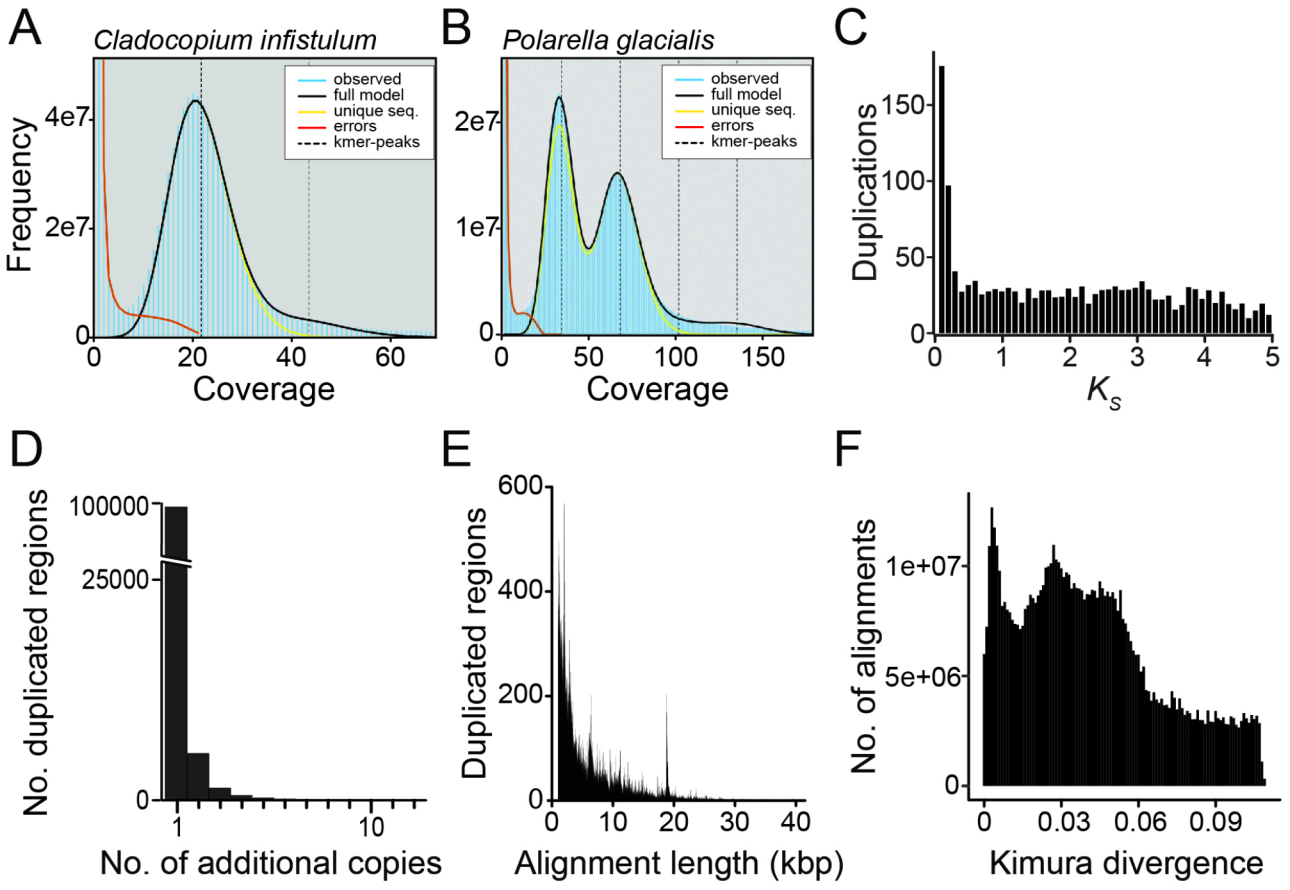

Fig. 5 Duplicated genomic regions appear to have originated at a single event in the past. **(A)** GenomeScope2 25-mer frequency distribution with ploidy set to 2. For comparison, **(B)** a GenomeScope plot of *P. glacialis* indicative of whole-genome duplication; modified from [62]. **(C)** Distribution of synonymous substitution rates (K_s) calculated from sequence alignments between pairs of gene paralogs. (**D**) Exponential graph showing the distribution of duplicated region sizes found in *C. infistulum*, ranging from 1 kbp to >40 kbp. **(E)** The alignment length distribution of duplicated segments with at least 1 kbp shows that while many span thousands of base pairs, few exceed 20 kbp. **(F)** Kimura divergence for pairwise alignments of duplicated regions

An obvious feature unique to the biology of intracellular Symbiodiniaceae is the presence of a symbiosome separating the algal cell from the host cytoplasm, consisting of both symbiont- and host-derived membranes [15]. Preliminary proteomic characterization of symbiosome membranes identified a set of proteins with homology to animal, algal, and microbial proteins [52]. Using non-animal protein sequences (Supplementary Table 7) to identify algal orthologs revealed that genes coding for these putative symbiosome proteins appear evenly represented across the *Cladocopium* genomes (Supplementary Fig. 11), irrespective of the symbiont's niche. While the symbiosome environment is unique to intracellular associations, extracellular Symbiodiniaceae living within molluscan digestive diverticula abut and interact directly with host cell membranes. The failure to detect distinct gene profiles for extra- and intracellular symbionts may reflect that, when engaged in endosymbiosis, their surrounding environments are effectively the same. Additionally, both types of symbionts need to survive outside of the host for at least a portion of their lifespans, which may select for retention of similar gene sets in support of a transient free-living ecology.

Representation of gene functions associated with response to stress shows no clear pattern with respect to *Cladocopium* phylogeny or symbiont niche

There were no obvious genes characteristic of what was expected to be the most thermally tolerant species, *C. infistulum* (Fig. 4 and Supplementary Fig. 12). While the species analyzed here exhibit a range of sensitivity to heat stress, all four species were isolated from hosts in warm environments, and therefore, might all have adaptations to cope with elevated temperatures [65]. For example, *C. proliferum* is particularly common to various corals from warmer low latitude reefs on the Great Barrier Reef [46]. Given the lagoonal and shallow habitats where corals in the genus *Porites* dominate, it is expected that associated *Cladocopium* C15 have evolved resilience to thermal stress. Clams hosting *C. infistulum* live in inshore

shallow equatorial reefs in the western Pacific [26]. These environments also experience larger diurnal temperature fluctuations than other reef habitats. Similarly, *Cladocopium* sp. Y103 lives in clams from shallow habitats with warm temperatures and high light regimes in Japan [41, 49]. While there may ultimately be differences in their maximum thermal tolerance, both clam symbionts are likely to possess genes conveying some level of resistance to heat.

In addition to the symbiont's innate physiology, the host environment also has a great capacity to influence thermal stress responses among marine mutualisms, possibly limiting Symbiodiniaceae specialization. Changes in the concentration and distribution of fluorescent pigments, mycosporine-like amino acids, heat shock proteins, and antioxidants have been proposed as mechanisms through which the host may buffer the symbiont community from heat and other environmental stressors [66]. Buffering has been inferred in associations between *Durusdinium trenchii* and *Acropora*, *Coelastrea*, *Cyphastrea*, and *Pachyseris* corals in the Indo-Pacific [67]; and in *Symbiodinium* A4–*Porites* [68] and *Symbiodinium* '*fitti*'–*Acropora* [69] associations in the Caribbean. Such host buffering could dampen stress-related ecophysiological and genomic divergence among symbiotic Symbiodiniaceae species.

Overall, *C. infistulum* and *Cladocopium* sp. Y103 had fewer genes associated with radical oxygen scavenging, protein chaperoning, or glutathione metabolism, which indicates that these extracellular clam symbionts may employ alternative mechanisms to cope with high seawater temperatures. While radical scavenging and protein chaperoning genes were underrepresented in *Cladocopium* C15, this taxon did have the most genes with functions related to '*cellular response to heat*' (i.e., coding for heat shock proteins or their transcription factors). *Cladocopium proliferum* had the highest representation of all other stress-response gene functions, which corresponds with evidence that this species is also thermotolerant [25, 70].

These findings indicate that the genetic underpinnings of thermotolerance among members of a single genus of Symbiodiniaceae may arise from diverse origins. In addition to the presence, absence, and abundance of certain genes, the basis of thermotolerance may reside in other processes, including post-transcriptional regulation via microRNA interference [71] or RNA editing [72], as well as differential methylation [73]. Thus, epigenomic, transcriptomic, proteomic, and metabolomic approaches may offer further insight into physiological capacity for thermal tolerance [32, 74–76].

Evidence for a genome duplication event remains inconclusive for *C. infistulum*

Previously, the high frequency of two alleles observed at different microsatellite loci among various *Cladocopium* spp. provided preliminary support for the occurrence of an ancestral WGD event [34, 35], possibly ~5.2 Mi years ago [21]. However, there was limited evidence in favor of WGD in the genome of *C. infistulum* based on independent analyses (Fig. 5A-C). Instead, multiple duplicated regions are spread throughout large parts of the genome (Fig. 5D-F), corresponding to ~23% of the total sequence and encompassing ~33% of all putative genes, while the rest of the genome was single copy. A recent study explored WGD in *Durusdinium trenchii*, a thermotolerant coral symbiont also within the family Symbiodiniaceae [77]. Although the authors concluded WGD *did* occur in that lineage based on the abundance of duplicated gene blocks, there was also no evidence of WGD from the synonymous substitution rate of gene paralogs. However, the recent divergence of the lineage leading to *D. trenchii* (~1 MYA) may not have allowed enough time for the accumulation of synonymous substitutions in duplicated regions. For *Cladocopium*, the ~10–20-million-year age of the genus [21] should be sufficient time for nucleotide substitutions to accumulate, presuming WGD occurred prior to the diversification of extant lineages. The possibility of low divergence masking WGD in *C. infistulum* cannot be discarded, but the detection of duplicated regions in this study makes this less likely. The apparent occurrence of genomic duplication as a single ancestral event (Fig. 5D, F) also supports the notion of WGD. Alternatively, there could have been widespread segmental duplication rather than WGD, perhaps driven in part by mobile element activity. The genomes of both *C. infistulum* (Supplementary Figs. 17 and 18, Supplementary Table 15) and *Fugacium kawagutii* [78] feature an overabundance of (retro)transposons in duplicated regions, supporting the notion that mobile elements are involved in duplication processes across the family Symbiodiniaceae.

Conclusions

The genomic characterization of *Cladocopium infistulum* revealed no patterns diagnostic of heat tolerance or extracellular symbiosis, at least with respect to overrepresentation of gene functions or genes under selection. These findings imply that adaptations among members of the genus *Cladocopium* are likely dictated by mechanisms beyond gene content. While genome duplication is prevalent in this species, the analyses could not identify a causative mechanism, and it remains unclear whether a whole-genome duplication event occurred in the ancestor to the *Cladocopium* lineage. Echoing previous investigations, this research highlights the need for functional

characterization of unknown genes and other genomic elements, as well as access to additional high-quality Symbiodiniaceae genomes. Such data are essential to better explain patterns and processes of genetic change in the eco-evolutionary diversification of dinoflagellates.

Abbreviations

GO	Gene Ontology
MYA	Million years ago
WGD	Whole–genome duplication

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-10822-0.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Acknowledgements
The authors thank Prof. Steven Maere and Prof. Claude DePamphilis for their advice regarding whole-genome duplication analyses, and Dr. Shumpei Maruyama for his guidance on gene functions of host-symbiont interactions. The authors are also thankful to Dr. Molly Przeworski for helpful comments and discussions during the initial planning of the project. The authors thank Dr. Christian Voolstra and Dr. Yi Jin Liew for making the assembled genome sequences, predicted genes and annotations available on Reef Genomics.

Author contributions
J.E.P. supervised the project. Z.L.F. and T.C.L. generated the initial sequencing data. R.A.G.P., J.S. and Z.L.F. conducted all genomic analyses. R.A.G.P., T.C.L and J.E.P. wrote the manuscript. All authors edited and approved the final manuscript.

Funding
This work was supported by the University of South Florida Research & Innovation Internal Awards Program (grant 0142687 to J.E.P); the Agouron Institute (to Z.L.F and Dr. Przeworski); and the Eberly College of Science at The Pennsylvania State University (Eberly Postdoctoral Research Fellowship to R.A.G.P.).

Data availability
The data generated for this study are available under NCBI BioProject PRJNA1103863. Raw sequencing data for *Cladocopium infistulum* strain RT-203 (NCBI BioSample SAMN41063414) are deposited in the NCBI Short Read Archive (SRA) under SRX24349346 and SRX24349347. Custom analysis scripts can be found at the following GitHub repository: https://github.com/RaulGoch/C.infistulum_genome. The genome assembly, along with predicted coding sequences, proteins and annotations are available at Reef Genomics: http://cinf.reefgenomics.org.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 15 May 2024 / Accepted: 20 September 2024
Published online: 01 October 2024

References

1. Trench R. Microalgal-invertebrate symbiosis - a review. Endocyt Cell Res. 1993;9:135–75.
2. Muscatine L. Glycerol excretion by symbiotic algae from corals and *Tridacna* and its control by the host. Science. 1967;156(3774):516–9.
3. Muscatine L, Porter JW. Reef corals: mutualistic symbioses adapted to nutrient-poor environments. Bioscience. 1977;27(7):454–60.
4. Enríquez S, Méndez ER, Iglesias-Prieto R. Multiple scattering on coral skeletons enhances light absorption by symbiotic algae. Limnol Oceanogr. 2005;50(4):1025–32.
5. Holt AL, Vahidinia S, Gagnon YL, Morse DE, Sweeney AM. Photosymbiotic giant clams are transformers of solar flux. J R Soc Interface. 2014;11(101):20140678.
6. Gómez-Campo K, Enríquez S, Iglesias-Prieto R. A road map for the development of the bleached coral phenotype. Front Mar Sci 2022, 9.
7. Avila-Magaña V, Kamel B, DeSalvo M, Gómez-Campo K, Enríquez S, Kitano H, Rohlfs RV, Iglesias-Prieto R, Medina M. Elucidating gene expression adaptation of phylogenetically divergent coral holobionts under heat stress. Nat Commun. 2021;12(1):5731.
8. Haydon TD, Matthews JL, Seymour JR, Raina J-B, Seymour JE, Chartrand K, Camp EF, Suggett DJ. Metabolomic signatures of corals thriving across extreme reef habitats reveal strategies of heat stress tolerance. Proc R Soc B Biol Sci. 2023;290(1992):20221877.
9. González-Pech RA, Stephens TG, Chen Y, Mohamed AR, Cheng Y, Shah S, Dougan KE, Fortuin MDA, Lagorce R, Burt DW, et al. Comparison of 15 dinoflagellate genomes reveals extensive sequence and structural divergence in family Symbiodiniaceae and Genus *Symbiodinium*. BMC Biol. 2021;19(1):73.
10. Liu H, Stephens TG, González-Pech RA, Beltran VH, Lapeyre B, Bongaerts P, Cooke I, Aranda M, Bourne DG, Forêt S, et al. *Symbiodinium* genomes reveal adaptive evolution of functions related to coral-dinoflagellate symbiosis. Comm Biol. 2018;1(1):95.
11. Rosset SL, Oakley CA, Ferrier-Pagès C, Suggett DJ, Weis VM, Davy SK. The molecular language of the cnidarian-dinoflagellate symbiosis. Trends Microbiol. 2021;29(4):320–33.
12. Mohamed AR, Andrade N, Moya A, Chan CX, Negri AP, Bourne DG, Ying H, Ball EE, Miller DJ. Dual RNA-sequencing analyses of a coral and its native symbiont during the establishment of symbiosis. Mol Ecol. 2020;29(20):3921–37.
13. Jacobovitz MR, Hambleton EA, Guse A. Unlocking the complex cell biology of coral–dinoflagellate symbiosis: a model systems approach. Annu Rev Genet. 2023;57(1):null.
14. Kemp DW, Hoadley KD, Lewis AM, Wham DC, Smith RT, Warner ME, LaJeunesse TC. Thermotolerant coral–algal mutualisms maintain high rates of nutrient transfer while exposed to heat stress. Proc R Soc B Biol Sci 2023;290(2007):20231403.
15. Davy SK, Allemand D, Weis VM. Cell biology of cnidarian-dinoflagellate symbiosis. Microbiol Mol Biol R. 2012;76(2):229–61.
16. Aranda M, Li Y, Liew YJ, Baumgarten S, Simakov O, Wilson MC, Piel J, Ashoor H, Bougouffa S, Bajic VB, et al. Genomes of coral dinoflagellate symbionts highlight evolutionary adaptations conducive to a symbiotic lifestyle. Sci Rep. 2016;6(1):39734.
17. González-Pech RA, Bhattacharya D, Ragan MA, Chan CX. Genome evolution of coral reef symbionts as intracellular residents. Trends Ecol Evol. 2019;34(9):799–806.
18. Trench RK, Wethey DS, Porter JW. Observations on the symbiosis with zooxanthellae among the Tridacnidae (Mollusca, Bivalvia). Biol Bull. 1981;161(1):180–98.
19. Mansour K. Communication between the dorsal edge of the mantle and the stomach of *Tridacna*. Nature. 1946;157(3999):844–844.
20. Norton JH, Shepherd MA, Long HM, Fitt WK. The zooxanthellal tubular system in the giant clam. Biol Bull. 1992;183(3):503–6.
21. LaJeunesse TC, Parkinson JE, Gabrielson PW, Jeong HJ, Reimer JD, Voolstra CR, Santos SR. Systematic revision of Symbiodiniaceae highlights the antiquity and diversity of coral endosymbionts. Curr Biol. 2018;28(16):2570–e25802576.
22. LaJeunesse TC. Species radiations of symbiotic dinoflagellates in the Atlantic and Indo-Pacific since the Miocene-Pliocene transition. Mol Biol Evol. 2005;22(3):570–81.
23. Turnham KE, Wham DC, Sampayo E, LaJeunesse TC. Mutualistic microalgae co-diversify with reef corals that acquire symbionts during egg development. ISME J. 2021;15(11):3271–85.
24. Thornhill DJ, Lewis AM, Wham DC, LaJeunesse TC. Hots-specialist lineages dominate the adaptive radiation of reef coral endosymbionts. Evolution. 2014;68(2):352–67.

25. Butler CC, Turnham KE, Lewis AM, Nitschke MR, Warner ME, Kemp DW, et al. Formal recognition of host-generalist species of dinoflagellate (*Cladocopium*, Symbiodiniaceae) mutualistic with indo-pacific reef corals. J Phycol. 2023;59(4):698–711.
26. Lee SY, Jeong HJ, Lajeunesse TC. *Cladocopium infistulum* sp. nov. (Dinophyceae), a thermally tolerant dinoflagellate symbiotic with giant clams from the western Pacific Ocean. Phycologia. 2020;59(6):515–26.
27. Mies M. Evolution, diversity, distribution and the endangered future of the giant clam–Symbiodiniaceae association. Coral Reefs. 2019;38(6):1067–84.
28. Grégoire V, Schmacka F, Coffroth MA, Karsten U. Photophysiological and thermal tolerance of various genotypes of the coral endosymbiont *Symbiodinium* sp. (Dinophyceae). J Appl Phycol. 2017;29(4):1893–905.
29. Brener-Raffalli K, Vidal-Dupiol J, Adjeroud M, Rey O, Romans P, Bonhomme F, Pratlong M, Haguenauer A, Pillot R, Feuillassier L et al. Gene expression plasticity and frontloading promote thermotolerance in *Pocillopora* corals. Peer Commun J 2022, 2.
30. Bellantuono AJ, Dougan KE, Granados-Cifuentes C, Rodriguez-Lanetty M. Free-living and symbiotic lifestyles of a thermotolerant coral endosymbiont display profoundly distinct transcriptomes under both stable and heat stress conditions. Mol Ecol. 2019;28(24):5265–81.
31. Shoguchi E, Beedessee G, Hisata K, Tada I, Narisoko H, Satoh N, Kawachi M, Shinzato C. A new dinoflagellate genome illuminates a conserved gene cluster involved in sunscreen biosynthesis. Genome Biol Evol 2021, 13(2).
32. Levin RA, Beltran VH, Hill R, Kjelleberg S, McDougald D, Steinberg PD, van Oppen MJH. Sex, scavengers, and chaperones: transcriptome secrets of divergent *Symbiodinium* thermal tolerances. Mol Biol Evol. 2016;33(9):2201–15.
33. Santos SR, Coffroth MA. Molecular genetic evidence that dinoflagellates belonging to the genus *Symbiodinium* Freudenthal are haploid. Biol Bull. 2003;204(1):10–20.
34. Wham DC, Carmichael M, LaJeunesse TC. Microsatellite loci for *Symbiodinium goreaui* and other Clade C *Symbiodinium*. Conserv Genet Resour. 2014;6(1):127–9.
35. Wham DC, LaJeunesse TC. *Symbiodinium* population genetics: testing for species boundaries and analysing samples with mixed genotypes. Mol Ecol. 2016;25(12):2699–712.
36. LaJeunesse TC. Investigating the biodiversity, ecology and phylogeny of endosymbiotic dinoflagellates in the genus *Symbiodinium* using ITS region: in search of a species level marker. J Phycol. 2001;37(5):866–80.
37. Chen Y, González-Pech RA, Stephens TG, Bhattacharya D, Chan CX. Evidence that inconsistent gene prediction can mislead analysis of dinoflagellate genomes. J Phycol. 2020;56(1):6–10.
38. Shoguchi E, Shinzato C, Kawashima T, Gyoja F, Mungpakdee S, Koyanagi R, et al. Draft assembly of the *Symbiodinium minutum* nuclear genome reveals dinoflagellate gene structure. Curr Biol. 2013;23(15):1399–408.
39. Barbrook AC, Voolstra CR, Howe CJ. The chloroplast genome of a *Symbiodinium* sp. Clade C3 isolate. Protist. 2014;165(1):1–13.
40. Robbins SJ, Singleton CM, Chan CX, Messer LF, Geers AU, Ying H, Baker A, Bell SC, Morrow KM, Ragan MA, et al. A genomic view of the reef-building coral *Porites lutea* and its microbial symbionts. Nat Microbiol. 2019;4(12):2090–100.
41. Shoguchi E, Beedessee G, Tada I, Hisata K, Kawashima T, Takeuchi T, Arakaki N, Fujie M, Koyanagi R, Roy MC, et al. Two divergent *Symbiodinium* genomes reveal conservation of a gene cluster for sunscreen biosynthesis and recently lost genes. BMC Genomics. 2018;19(1):458.
42. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol Biol Evol. 2021;38(10):4647–54.
43. Song B, Morse D, Song Y, Fu Y, Lin X, Wang W, Cheng S, Chen W, Liu X, Lin S. Comparative genomics reveals two major bouts of gene retroposition coinciding with crucial periods of *Symbiodinium* evolution. Genome Biol Evol. 2017;9(8):2037–47.
44. Goodswen SJ, Kennedy PJ, Ellis JT. A gene-based positive selection detection approach to identify vaccine candidates using *Toxoplasma Gondii* as a test case protozoan pathogen. Front Genet 2018, 9.
45. Numanagić I, Gökkaya AS, Zhang L, Berger B, Alkan C, Hach F. Fast characterization of segmental duplications in genome assemblies. Bioinformatics. 2018;34(17):i706–14.
46. Howells EJ, Beltran VH, Larsen NW, Bay LK, Willis BL, van Oppen MJH. Coral thermal tolerance shaped by local adaptation of photosymbionts. Nat Clim Change. 2012;2(2):116–20.
47. Swain TD, Chandler J, Backman V, Marcelino L. Consensus thermotolerance ranking for 110 *Symbiodinium* phylotypes: an exemplar utilization of a novel iterative partial-rank aggregation tool with broad application potential. Funct Ecol. 2017;31(1):172–83.
48. Krueger T, Gates RD. Cultivating endosymbionts — host environmental mimics support the survival of *Symbiodinium* C15 *ex hospite*. J Exp Mar Biol Ecol. 2012;413:169–76.
49. Ohno T, Katoh T, Yamasu T. The origin of algal-bivalve photo-symbiosis. Palaeontology. 1995;38(1):1–22.
50. Bernard G, Stephens TG, González-Pech RA, Chan CX. Inferring phylogenomic relationship of microbes using scalable alignment-free methods. In: Bacterial pangenomics: Methods and protocols. Edited by Mengoni A, Bacci G, Fondi M. New York, NY: Springer US; 2021:69–76.
51. Nand A, Zhan Y, Salazar OR, Aranda M, Voolstra CR, Dekker J. Genetic and spatial organization of the unusual chromosomes of the dinoflagellate *Symbiodinium microadriaticum*. Nat Genet. 2021;53(5):618–29.
52. Peng S-E, Wang Y-B, Wang L-H, Chen W-NU, Lu C-Y, Fang L-S, Chen C-S. Proteomic analysis of symbiosome membranes in Cnidaria–dinoflagellate endosymbiosis. Proteomics. 2010;10(5):1002–16.
53. Sproles AE, Kirk NL, Kitchen SA, Oakley CA, Grossman AR, Weis VM, Davy SK. Phylogenetic characterization of transporter proteins in the cnidarian-dinoflagellate symbiosis. Mol Phylogenet Evol. 2018;120:307–20.
54. Xiang T, Lehnert E, Jinkerson RE, Clowez S, Kim RG, DeNofrio JC, Pringle JR, Grossman AR. Symbiont population control by host-symbiont metabolic interaction in Symbiodiniaceae-cnidarian associations. Nat Commun. 2020;11(1):108.
55. Fam RRS, Hiong KC, Choo CYL, Wong WP, Chew SF, Ip YK. Molecular characterization of a novel algal glutamine synthetase (GS) and an algal glutamate synthase (GOGAT) from the colorful outer mantle of the giant clam, *Tridacna squamosa*, and the putative GS-GOGAT cycle in its symbiotic zooxanthellae. Gene. 2018;656:40–52.
56. Ip YK, Chew SF. Light-dependent phenomena and related molecular mechanisms in giant clam-dinoflagellate sssociations: a review. Front Mar Sci 2021, 8.
57. Pang CZ, Ip YK, Chew SF. Using transcript levels of nitrate transporter 2 as molecular indicators to estimate the potentials of nitrate transport in *Symbiodinium*, *Cladocopium*, and *Durusdinium* of the fluted giant clam, *Tridacna Squamosa*. Front Mar Sci. 2021;8.
58. Pang CZ, Boo MV, Ip YK, Chew SF. Symbiotic dinoflagellates of the giant clam, *Tridacna squamosa*, express ammonium transporter 2 at the plasma membrane and increase ots expression levels during illumination. Front Mar Sci 2022, 9.
59. Poo JST, Boo MV, Chew SF, Ip YK. Using form II ribulose-1,5-bisphosphate carboxylase/oxygenase to estimate the phototrophic potentials of *Symbiodinium*, *Cladocopium* and *Durusdinium* in various organs of the fluted giant clam, *Tridacna squamosa*, and to evaluate their responses to light upon isolation from the host. Coral Reefs. 2021;40(1):233–50.
60. Mani R, Boo MV, Ng SY, Chew SF, Ip YK. Symbiotic dinoflagellates of the giant clam, *Tridacna squamosa*, express an extracellular alpha carbonic anhydrase associated with the plasma membrane to promote HCO3 – dehydration and CO2 uptake during illumination. Coral Reefs. 2022;41(4):1097–113.
61. Teng GCY, Boo MV, Lam SH, Pang CZ, Chew SF, Ip YK. Molecular characterization and light-dependent expression of glycerol facilitator (GlpF) in coccoid Symbiodiniaceae dinoflagellates of the giant clam *Tridacna squamosa*. Gene Rep. 2022;27:101623.
62. Stephens TG, González-Pech RA, Cheng Y, Mohamed AR, Burt DW, Bhattacharya D, et al. Genomes of the dinoflagellate *Polarella glacialis* encode tandemly repeated single-exon genes with adaptive functions. BMC Biol. 2020;18(1):56.
63. Zwaenepoel A, Van de Peer Y. Wgd—simple command line tools for the analysis of ancient whole-genome duplications. Bioinformatics. 2018;35(12):2153–5.
64. Karim W, Nakaema S, Hidaka M. Temperature effects on the growth rates and photosynthetic activities of *Symbiodinium* cells. J Mar Sci Eng. 2015;3(2):368–81.
65. D'Angelo C, Hume BCC, Burt J, Smith EG, Achterberg EP, Wiedenmann J. Local adaptation constrains the distribution potential of heat-tolerant *Symbiodinium* from the Persian/Arabian Gulf. ISME J. 2015;9(12):2551–60.
66. Baird AH, Bhagooli R, Ralph PJ, Takahashi S. Coral bleaching: the role of the host. Trends Ecol Evol. 2009;24(1):16–20.
67. Hoadley KD, Lewis AM, Wham DC, Pettay DT, Grasso C, Smith R, Kemp DW, LaJeunesse TC, Warner ME. Host–symbiont combinations dictate the photophysiological response of reef-building corals to thermal stress. Sci Rep. 2019;9(1):9985.

68. Kenkel CD, Goodbody-Gringley G, Caillaud D, Davies SW, Bartels E, Matz MV. Evidence for a host role in thermotolerance divergence between populations of the mustard hill coral (*Porites astreoides*) from different reef environments. Mol Ecol. 2013;22(16):4335–48.
69. Parkinson JE, Banaszak AT, Altman NS, LaJeunesse TC, Baums IB. Intraspecific diversity among partners drives functional variation in coral symbioses. Sci Rep. 2015;5(1):15667.
70. Abrego D, Ulstrup KE, Willis BL, van Oppen MJH. Species–specific interactions between algal endosymbionts and coral hosts define their bleaching response to heat and light stress. Proc R Soc B Biol Sci. 2008;275(1648):2273–82.
71. Baumgarten S, Bayer T, Aranda M, Liew YJ, Carr A, Micklem G, Voolstra CR. Integrating microRNA and mRNA expression profiling in *Symbiodinium microadriaticum*, a dinoflagellate symbiont of reef-building corals. BMC Genomics. 2013;14(1):704.
72. Liew YJ, Li Y, Baumgarten S, Voolstra CR, Aranda M. Condition-specific RNA editing in the coral symbiont *Symbiodinium microadriaticum*. PLoS Genet. 2017;13(2):e1006619.
73. de Mendoza A, Bonnet A, Vargas-Landin DB, Ji N, Li H, Yang F, Li L, Hori K, Pflueger J, Buckberry S, et al. Recurrent acquisition of cytosine methyltransferases into eukaryotic retrotransposons. Nat Commun. 2018;9(1):1341.
74. Matthews JL, Cunning R, Witson-Williams R, Oakley CA, Lutz A, Roessner U, Grossman AR, Weis VM, Gates RD, Davy SK. The metabolic significance of symbiont community composition in the coral-algal symbiosis. In: *Applied Environmental Metabolomics*. Edited by Beale DJ, Hillyer KE, Warden AC, Jones OAH: Academic Press; 2022: 211–229.
75. Oakley CA, Newson GI, Peng L, Davy SK. The *Symbiodinium* proteome response to thermal and nutrient stresses. Plant Cell Physiol. 2022;64(4):433–47.
76. Davies SW, Gamache MH, Howe-Kerr LI, Kriefall NG, Baker AC, Banaszak AT, Bay LK, Bellantuono AJ, Bhattacharya D, Chan CX, et al. Building consensus around the assessment and interpretation of Symbiodiniaceae diversity. PeerJ. 2023;11:e15023.
77. Dougan KE, Bellantuono AJ, Kahlke T, Abbriano RM, Chen Y, Shah S, et al. Whole-genome duplication in an algal symbiont bolsters coral heat tolerance. Sci Adv. 2024;10(29):eadn2218.
78. Lin S, Cheng S, Song B, Zhong X, Lin X, Li W, Li L, Zhang Y, Zhang H, Ji Z, et al. The *Symbiodinium kawagutii* genome illuminates dinoflagellate gene expression and coral symbiosis. Science. 2015;350(6261):691–4.

Publisher's note