Article

Whole-genome duplication in an algal symbiont bolsters coral heat tolerance

Katherine E. Dougan[1,2]*, Anthony J. Bellantuono[3], Tim Kahlke[4], Raffaela M. Abbriano[4], Yibi Chen[1], Sarah Shah[1], Camila Granados-Cifuentes[2], Madeleine J. H. van Oppen[5,6], Debashish Bhattacharya[7], David J. Suggett[4,8], Mauricio Rodriguez-Lanetty[2]*, Cheong Xin Chan[1]*

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

[2]Florida International University, Department of Biological Sciences, Miami, FL 33099, U.S.A.

[3]Florida International University, Department of Biological Sciences, Biomolecular Science Institute, Miami, FL 33099, U.S.A.

[4]University of Technology Sydney, Climate Change Cluster, NSW 2007, Australia

[5]School of Biosciences, The University of Melbourne, Parkville, VIC 3010, Australia

[6]Australian Institute of Marine Science, Townsville, QLD 4810, Australia

[7]Rutgers University, Department of Biochemistry and Microbiology, New Brunswick, NJ 08901, U.S.A.

[8]King Abdullah University of Science and Technology, KAUST Reefscape Restoration Initiative (KRRI) and Red Sea Research Center (RSRC), Thuwal, 23955, Saudi Arabia

*Authors to whom correspondence and material requests should be addressed: Katherine E. Dougan (k.dougan@uq.edu.au), Mauricio Rodriguez-Lanetty (rodmauri@fiu.edu), and Cheong Xin Chan (c.chan1@uq.edu.au)

Abstract

The algal endosymbiont *Durusdinium trenchii* enhances the resilience of coral reefs under thermal stress[1,2]. As an endosymbiont, *D. trenchii* is generally expected to have a reduced genome compared to its free-living relatives, due in part to the lack of selective pressure for maintaining redundant gene functions in a stable intracellular environment within the host[3]. However, *D. trenchii* can live freely or in endosymbiosis, and the analysis of genetic markers[4] suggests that this species has undergone whole-genome duplication (WGD). Here we present genome assemblies for two *D. trenchii* isolates, confirm WGD in these taxa, and examine how selection has shaped the duplicated genome regions. We assess how the competing free-living versus endosymbiotic lifestyles of *D. trenchii* have contributed to the retention and divergence of duplicated genes, and how these processes have enhanced thermotolerance of corals hosting these symbionts. We find that lifestyle is the driver of post-WGD evolution in *D. trenchii*, with the free-living phase being most important, followed by endosymbiosis. Adaptations to both lifestyles collectively result in increased cellular fitness for *D. trenchii*, which provides enhanced thermal stress protection to the host coral. Beyond corals, this polyploid alga is a valuable model for understanding how genome-wide selective forces act to balance the often, divergent constraints imposed by competing lifestyles.

Main text

Uncovering the foundations of biotic interactions, particularly symbiosis, remains a central goal for research, given that virtually no organism lives in isolation. Coral reefs are marine biodiversity hotspots that are founded upon symbioses involving dinoflagellate algae in the Family Symbiodiniaceae[5]. These symbionts are the “solar power plants” of reefs, providing photosynthetically fixed carbon and other metabolites to the coral host[6,7]. Breakdown of the

46 coral-dinoflagellate symbiosis (i.e. coral bleaching), often due to ocean warming, puts corals at
47 risk of starvation, disease, and eventual death. Symbiodiniaceae microalgae are diverse, with at
48 least 15 clades including 11 named genera[5,8-10], encompassing a broad spectrum of symbiotic
49 associations and host-specificity. Most of these taxa are facultative symbionts (i.e. they can live
50 freely or in symbiosis), although exclusively symbiotic or free-living species also exist[5].
51 Genomes of Symbiodiniaceae are believed to reflect the diversification and specialization of
52 these taxa to inhabit distinct ecological niches[3,11]. The genomes of symbionts, due to spatial
53 confinement, are predicted to undergo structural rearrangements, streamlining, and rapid genetic
54 drift (e.g. pseudogenization)[3]. These traits are present in symbiotic Symbiodiniaceae[11].

55 Whole-genome duplication (WGD) is an evolutionary mechanism for generating functional
56 novelty and genomic innovation[12,13], and can occur due to errors in meiosis, i.e. via
57 autopolyploidy. Following WGD, the evolutionary trajectory of duplicated sequence regions
58 generally proceeds from large-scale purging, temporary retention and/or divergence, to
59 fixation[14]. WGD-derived duplicated genes (i.e. *ohnologs*[15,16]) that are retained can provide a
60 selective advantage and enhance fitness through increased gene dosage, specialization in
61 function, and/or the acquisition of novel functions[14].

62 WGD has been described in free-living unicellular eukaryotes such as yeast[17-19], ciliates[20,21], and
63 diatoms[22,23], but not in symbiotic species. Evidence of WGD is absent in the Symbiodiniaceae,
64 except for the genus *Durusdinium*, as observed in microsatellite sequence data[4]. This genus
65 includes the thermotolerant species *Durusdinium trenchii* (Fig. 1a), a facultative symbiont that
66 confers heat-tolerance to corals, thereby enhancing holobiont resilience under thermal stress[24].
67 We hypothesize that WGD played a critical role in enhancing heat-tolerance in this species.
68 Specifically, the facultative lifestyle (i.e. free-living or symbiotic) of *D. trenchii* favoured

69 fixation of WGD during the free-living phase as an adaptation to fluctuating environmental
70 conditions, with the expanded gene inventory being further modified by the coral symbiosis. To
71 test this "dual lifestyle" hypothesis, we generated *de novo* genome assemblies from two isolates
72 of *D. trenchii* and analysed their evolutionary trajectories. Predictions of our hypothesis were
73 tested against the null model of a single, free-living, lifestyle for this species. Based on gene
74 expression profiles, we elucidate how the facultative lifestyle has contributed to the fate of
75 ohnologs in these microalgae, and how natural selection acting on gene families has increased
76 thermotolerance of corals hosting *D. trenchii* symbionts. These data provide strong evidence for
77 the dual lifestyle hypothesis as a driver of post-WGD genome evolution.

78 **Whole-genome duplication in a coral endosymbiont**

79 We generated *de novo* genome assemblies from *D. trenchii* CCMP2556 (total length = 1.71 Gb;
80 N50 = 774.26 kb) and *D. trenchii* SCF082 (total length = 1.64 Gb; N50 = 398.48 kb) using 10X
81 Genomics linked reads (Tables S1 and S2). The two genomes are highly similar in terms of
82 whole-genome sequence (Table S3, ~99.7% shared identity), size (Table S4), and repeat
83 landscapes (Fig. 1b, Fig. S1), yielding ~54,000 protein-coding genes (Table S5) with a
84 comparable level of data completeness to other genome assemblies of Symbiodiniaceae (Table
85 S6; see Methods). To assess WGD in *D. trenchii*, we followed González-Pech *et al.*[11] to identify
86 collinear gene blocks within each genome (Fig. 1c, see Methods); these blocks likely arose via
87 segmental duplication and/or WGD. We identified 864 blocks implicating 27,597 (49.46% of the
88 total 55,799) genes in CCMP2556, and 776 blocks implicating 18,209 (34.02% of the total
89 53,519) genes in SCF082 (Tables S7 and S8). The proportion of genes present in collinear blocks
90 in *D. trenchii* is ~49-fold greater (Fig. 1d) than that in other Symbiodiniaceae and the outgroup
91 dinoflagellate *Polarella*, which have not experienced WGD. We also observed a high extent of

92 conserved synteny (22,041 CCMP2556 genes syntenic with 21,094 SCF082 genes), with
93 ohnologs predominant in these syntenic blocks (CCMP2556: 15,395 [69.85%]; SCF082: 12,617
94 [59.31%]) (Fig. 1e, Figure S2, and Table S9). Using homologous protein sets derived from
95 available whole-genome data, our inference of lineage-specific duplicated genes (see Methods)
96 revealed 7,945 gene duplication events specific to *D. trenchii*, which is an order of magnitude
97 greater than in other Symbiodiniaceae (Fig. 1f).

98 Examination of the overall distribution of DNA synonymous substitutions (K_S) showed a distinct
99 peak (Fig. S3), as expected following WGD; the small peak values are explained by the recency
100 of this event in *D. trenchii*[25]. The WGD likely occurred after the split of *D. trenchii* from its
101 sister *Durusdinium glynnii* 0.11–1.93 million years ago (MYA), based on LSU rDNA genetic
102 divergence estimates[5]. Our analysis of whole-genome data following Ladner *et al.*[26] aligns with
103 these estimates of a Pleistocene origin in the Indo-Pacific (Supplementary Information), a period
104 of frequent sea-level changes in this region[27]. These results, based on independently assembled
105 genomes from two isolates, combined with the extent and size of the gene blocks (Table S7 and
106 Fig. S2), provide unambiguous evidence for WGD in *D. trenchii*.

107 **Asymmetric divergence of ohnolog-pair expression**

108 To assess putative ohnolog functions in *D. trenchii*, we analysed transcriptome data of
109 CCMP2556[28] that were generated from free-living cells in culture and from cells in
110 endosymbiosis with the anemone *Exaiptasia pallida*, both under ambient (28°C) and thermal
111 stress (34°C) conditions. We focused on 6,147 expressed ohnolog-pairs that were supported by
112 10 or more mapped transcripts in ≥50% of the samples, and inferred gene co-expression
113 networks (Fig. S4 and Table S10) using weighted gene co-expression network analysis
114 (WGCNA). Most (4,412 [71.7%] of 6,147) ohnolog-pairs were recovered in different networks,

115 indicating the prevalence of expression divergence between duplicates post-WGD. We then
116 classified ohnolog pairs into five groups based on their differential expression (DE) patterns
117 (Figs 2A, S5-S9; see Methods). Each group exhibited different characteristics (Table S11)
118 relative to expression (Fig. 2a-b), sequence similarity (Fig. S10), gene structure (i.e. exon
119 gain/loss; Fig. 2c), and/or alternative splicing (Fig. S11-S12); see Supplementary Information.
120 Ohnolog-pairs that were differentially expressed between lifestyles observed in only one copy
121 (Group 2; 2,244 [36.5%]%]; Fig. 2d), and those with opposing differential expression observed
122 in any one comparison (Group 5; 100 [1.6%]; Fig. 2e; Table S12) from strongly contrasting
123 expression profiles (most Pearson correlation coefficients < 0; Fig. 2b), showed significantly
124 elevated levels of positive selection, exon gain/loss, sequence divergence, and differential exon
125 usage (DEU; Table S13) relative to the other three groups (all pairwise Wilcox $p < 0.05$;
126 Supplementary Information); these differences were not attributed to the differing number of
127 splice junctions per gene, and ohnologs show greater extent of alternative splicing than
128 singletons (Fig. S13 and Table S14).

129 Divergence of expression between ohnologs within a pair can have different outcomes, including
130 the change in expression specificity, an important mechanism for adaptation after WGD. We
131 assessed expression specificity using the τ index[29] that ranges between 0 (i.e. broad expression,
132 low specificity) to 1 (i.e. narrow expression, high specificity) for all genes that passed the
133 WGCNA quality filtering. We identified 3,508 genes of high expression specificity ($\tau > 0.7$), of
134 which 1,893 (53.96%) were ohnologs (Table S15). Compared to singletons and other duplicate
135 types (except for proximal duplicates), the ohnologs exhibited significantly elevated τ (Fig. S14,
136 Kruskal-Wallis test; $p < 10^{-5}$), indicating narrow expression profiles that are more specialized to
137 distinct conditions. This divergence in expression was observed in Group 2 pairs (Fig. S6), for

138 which the differentially expressed copy in each pair showed higher τ and variance in expression
139 relative to its counterpart (Kruskal-Wallis test; $p < 10^{-15}$). Whereas most instances of specialized
140 expression are associated with the free-living lifestyle (Table S16), this specialization reflects
141 response to temperatures among the dispersed duplicates (i.e. duplicates separated by >20 genes;
142 Chi-square test post-hoc: p=0, residuals=6.15 at 34°C) and the ohnologs (p <0.05,
143 residuals=3.10 at 28°C). Interestingly, only ohnologs displayed a tendency towards expression
144 specialization in the symbiotic lifestyle at both 28°C (Chi-square test post-hoc: $p < 0.01$,
145 residuals=3.61) and 34°C ($p < 0.01$, residuals=3.89; Table S16). Although the post-WGD
146 specialization described here relates to lifestyle, parallels are known in multicellular organisms
147 whereby the partitioning of expression across spatiotemporal scales is often observed in different
148 tissues, organs, or developmental stages[14,30]. In post-WGD yeasts, this trend may represent the
149 uncoupling of noise and plasticity in gene expression that enables dynamic gene-expression
150 responses in one of the two duplicates[31]. In *D. trenchii*, this trait may provide greater flexibility
151 in gene expression when cells are free-living or experiencing temperature stress.

152 The up-regulation or specialization of gene expression by some ohnolog-pairs to different
153 lifestyles in *D. trenchii* appears to either be mediated by, or coincide with, alterations in exon
154 organization. Evolution of WGD-derived genes can lead to loss and/or diversification of
155 alternative spliced forms, and the partitioning of ancestral splice forms between gene duplicates.
156 We investigated this issue by examining the interplay of sequence conservation in exonic
157 sequences with patterns of splice junction conservation and DEU within ohnolog pairs. Based on
158 mean percentage of shared exons per pair, the ohnolog pairs in Group 5 (13.35%) had lower
159 exonic conservation compared to other groups (>16%). DEU across all groups were biased
160 toward functions associated with the free-living lifestyle. Nearly all ohnolog pairs in Group 5 (95

161 of 100; Fig. 2e and Fig. S9) displayed contrasting differential expression between free-living and
162 symbiotic phases at one or both temperatures, underscoring lifestyle as a strong driver of
163 expression divergence (Supplementary Information). Group 5 ohnologs that are specialized for
164 the symbiotic lifestyle exhibited lower overall DEU, and possessed fewer exons than their
165 counterparts that were up-regulated under the free-living lifestyle (Wilcoxon rank sum test, p =
166 0.015, V = 2435.5). These ohnologs also contained exons that were more dominantly expressed
167 during the symbiotic lifestyle (Fig. 3a, Wilcoxon rank sum test, p = 0.02798, V = 2540). Such a
168 bias in DEU composition towards symbiosis-specialized exons was not observed in the other
169 groups, e.g. Group 2 (Fig. 3b-c). Consequently, the symbiosis-specific DEU, together with the
170 overall decrease in per-gene exons and DEU among symbiosis-associated ohnologs in Group 5,
171 suggest a symbiosis-specific streamlining of gene function. Together with our observation of
172 RNA editing (Tables S17 and S18, Figure S15; Supplementary Information), these results
173 collectively indicate that alterations to gene structure and alternative splicing drive expression
174 divergence of ohnologs in *D. trenchii* that are explained by algal lifestyle.

175 Although genomic streamlining is usually associated with obligate endosymbionts rather than
176 facultative symbionts, gene duplication may facilitate streamlining in one of the two duplicates
177 in favour of symbiotic lifestyle. Ohnolog pairs of Group 5 were significantly enriched for key
178 functions (Table S19), such as the processing of glutamine and production of the key antioxidant
179 of glutathione, which have been linked to nitrogen cycling associated with symbiosis[32,33]; the
180 implicated genes include glutamine synthetase and S-formylglutathione hydrolase (Table S12).
181 These results suggest that following WGD, specialization of gene expression to distinct
182 conditions may also be enabled by the streamlining of functions and specialization to symbiosis.

183 In contrast, for their duplicated counterparts, the evolution of greater functional flexibility may
184 reflect selection during the free-living phase.

185 **Partitioned functionality in central metabolic pathways**

186 WGD enables the retention of complete expression networks. Of the 19 inferred co-expression
187 networks (Table S10), different gene duplication types displayed preferential distributions to
188 WGCNA modules ($p < 2.2 \times 10^{-16}$, $\chi^2 = 525.63$). Singletons and ohnologs were biased towards
189 contrasting co-expression networks, with singletons predominantly associated with networks
190 linked to the symbiotic lifestyle (M1, M8, and M17 in Table S20), and ohnologs with networks
191 linked to a free-living lifestyle (M2, M5, and M6). This result suggests that genes preferentially
192 retained as ohnologs were expressed at contrasting times, compared to those that were lost such
193 that the remaining copies become singletons. Differential expression of ohnologs was observed
194 at the greatest magnitude between lifestyles during heat stress at 34°C (Chi-square test post-hoc:
195 $p < 10^{-3}$, Residuals=4.03; Table S21); this may explain in part how *D. trenchii* can establish itself
196 or increase in abundance in new hosts both during and after heat waves[34-38]. These contrasting
197 patterns of singleton and ohnolog membership across co-expression networks indicate a strong
198 association of ohnolog retention with expression networks that are tightly linked to the free-
199 living lifestyle.

200 We investigated retention of complete metabolic pathways in both *D. trenchii* isolates. Of the 98
201 pathways retained in duplicate (Table S22), specialization driven by lifestyle was detected in
202 central metabolic pathways (Figs S16-S23), such as glycolysis/gluconeogenesis (Fig. 3d and Fig.
203 S16). Ohnolog specialization in glycolysis/gluconeogenesis reflects the contrasting functions of
204 this pathway during the symbiotic *versus* free-living phases. That is, a high rate of
205 gluconeogenesis, inferred using ohnolog expression data, supplies glucose for translocation to

206 the coral host during symbiosis, whereas a high rate of glycolysis fuels the energetic needs of
207 free-living cells that tolerate more variable environments[3]. Although most enzymes were
208 encoded by Group 2 ohnologs (for which one gene copy was differentially expressed between
209 lifestyles; Fig. 3e), a key rate-limiting enzyme of gluconeogenesis and the Calvin cycle, fructose
210 1,6-bisphosphatase, was differentially expressed in response to heat stress in symbiosis.
211 Development of minor or partitioned functionality following WGD has been described in
212 duplicate glycolysis pathways[39]. In yeast, these pathways diverged and became semi-
213 independent, with each specialized for low and high glucose levels[39]. In *D. trenchii*, this might
214 allow fine-tuning of carbon metabolism to the contrasting energetic needs of a dual lifestyle.

215 **Concluding remarks**

216 Our results provide strong evidence that the dual lifestyle has been a key driver of post-WGD
217 genome evolution in the dinoflagellate *D. trenchii*. Our working hypothesis is illustrated in Fig.
218 4. Under the null hypothesis of a solely free-living lifestyle, we expect post-WGD adaptations to
219 primarily be driven by fluctuating environmental conditions (e.g., nutrient availability). Under
220 the hypothesis of a dual lifestyle that includes symbiosis, adaptations will also strengthen the
221 maintenance of a stable host-symbiont relationship, and efficient nutrient/metabolite exchange
222 within the coral holobiont. Although our results provide stronger support for the free-living
223 phase as the primary driving force behind post-WGD evolution, both lifestyles impact the
224 maintenance and expression divergence of ohnologs. These combined selective forces increase
225 overall fitness in *D. trenchii*, with the greater expression divergence of ohnologs under elevated
226 temperatures a contributor to the high thermotolerance of this species when it is in symbiosis
227 with corals[40]. Benefits conferred by WGD to a free-living lifestyle in more-variable
228 environments, as well as tailoring of post-WGD duplicates to different lifestyles, primed *D.*

229 *trenchii* to persist longer in the coral holobiont when faced with thermal stress. Whether
230 symbiosis may also have negative effects on fitness post-WGD is unknown[41]. It should be noted
231 that the dual lifestyle is widespread in Symbiodiniaceae[5], but WGD is not. Although other
232 facultative symbionts within Symbiodiniaceae (e.g., *Cladocopium thermophilum*)[42,43] are also
233 known for their thermotolerance, WGD was not implicated in these lineages[11,44]. Therefore, the
234 key feature of *D. trenchii* that we are addressing is not dual lifestyle alone, but rather how the
235 capacity for dynamically switching between the symbiotic *versus* free-living phase impacts post-
236 WGD genome evolution and adaptation. Because Symbiodiniaceae propagate to very high
237 densities in coral tissues (10^5–10^6 cells/cm^2)[45,46], the symbiotic phase of *D. trenchii* allows a
238 rapid increase in the population size, particularly of fast-growing genotypes, while resident in
239 host tissues. Consequently, genotypes that have faster growth rates or greater resilience to heat
240 due to WGD-derived adaptations can re-seed free-living populations upon dissociation from the
241 coral due to colony death, bleaching, or other mechanisms of symbiont population control.
242 Repeated cycles of symbiosis followed by the free-living phase may therefore increase the
243 overall fitness of *D. trenchii* populations under the dual lifestyle[47]. Retention of multiple gene
244 copies combined with fixed, adaptive changes likely makes *D. trenchii* more capable of
245 metabolic maintenance under dynamic, often stressful environments, and hence a more-resilient
246 symbiont. Such factors may in turn explain the large geographic and expanded host range of *D.*
247 *trenchii*[24] and its well-known capacity for increasing coral survival under heat waves. Therefore,
248 in an intriguing and unexpected twist, WGD, primarily driven by selection under a free-living
249 life phase has converted *D. trenchii* into a coral symbiont able to protect the host coral from
250 thermal stress during symbiosis. *D. trenchii* is also a valuable model for studying the genome-
251 wide impacts of facultative lifestyles.

Methods

De novo genome assembly and prediction of protein-coding genes

Durusdinium trenchii strains CCMP2556 and SCF082 (previously designated UTSD amur-D-MI) originally isolated from an *Orbicella faveolate* and *Acropora muricata* coral colonies, respectively, were each separately cultured and genomic DNA extracted for genomic sequencing (see Supplementary Information). Chromium libraries were generated for 10X linked-read sequencing and yielded a total of 236.45 Gbp for CCMP2556 and 212.03 Gbp for SCF082. We assessed the ploidy of *D. trenchii* using *k*-mers and GenomeScope2[48], which revealed a distinctive single peak in both isolates indicating a haploid genome as seen in other Symbiodiniaceae (Figure S24), with a predicted heterozygosity of 0.31% and 0.20% in CCMP2556 and SCF082, respectively.

For each isolate, a preliminary draft genome was assembled *de novo* using 10X Genomics Supernova v2.1.1. For CCMP2556, the estimated genome coverage (~100×) exceeded the optimal range (38–56×) of the Supernova assembler; we subsampled the 1.6B reads to 600M reads (~60× coverage). For SCF082, coverage estimates were observed to be impacted due to the presence of contaminant DNA from microbial sources in the sequencing reads; the *de novo* assembly was generated using all 1.4B reads with the flag *–accept_extreme_coverage*.

Presence of putative contaminant scaffolds in the supernova assemblies was investigated using a comprehensive approach adapted from Iha *et al.*[49] informed by read coverage, G+C content, taxonomic designation, and *de novo* transcriptome mapping. Taxon-annotated G+C-coverage plots (Figure S25) were generated using the BlobTools suite v1.1[50] to identify scaffolds in each assembly that deviated by read coverage, taxonomic sequence similarity, and/or G+C content.

274 Read coverage was assessed using BWA v0.7.17, based on mapping of quality-trimmed reads
275 (Longranger v2.2.2[51] ran at default setting) to the genome assembly. The taxonomic identity of
276 scaffolds was assigned based on BLASTN search ($E \leq 10^{-20}$) against genome sequences from
277 bacteria, archaea, viruses, and alveolates in the NCBI nt database (release 2021-05-10). *De novo*
278 transcriptome assemblies were mapped to the genome assemblies using minimap2 v2.18[52] within
279 which we have modified the codes to account for non-canonical splice sites of dinoflagellates.
280 Scaffolds that were designated as non-dinoflagellate were removed from the assemblies if they
281 lacked mapped transcripts from the corresponding *de novo* transcriptome assembly, or when
282 <10% of mapped transcripts indicate evidence of introns in the genomes. We considered a
283 scaffold as a putative contaminant if (a) its sequence coverage or G+C content is not within the
284 1.5 × interquartile range, and (b) it lacks any transcript support defined above. Upon removal of
285 these putative contaminant sequences from the CCMP2556 assembly, the filtered assembly was
286 incorporated in the database as the *D. trenchii* reference for assessing the assembled scaffolds of
287 SCF082 using the same approach.

288 Publicly available RNA-Seq data from previous studies of CCMP2556[28] and SCF082[53] were
289 used to further scaffold the assembled genome sequences (see Supplementary Information).
290 RNA-Seq reads for both isolates were first quality-trimmed using fastp[54] (mean Phred quality ≥
291 30 across a 4bp window; minimum read length of 50bp). For each isolate, the filtered reads were
292 assembled *de novo* using Trinity v2.11.0[55] independently for each treatment. The transcriptome
293 assemblies for CCMP2556 (791,219 total transcripts) and those for SCF082 (355,411 total
294 transcripts) were mapped to the filtered genome assemblies using minimap2 v2.18[52] that was
295 modified to recognize the non-canonical splice sites of dinoflagellates. The mapped transcripts

296 were then used to scaffold the filtered genome assemblies with L_RNA_Scaffolder[56] at default
297 parameters.

298 A second round of scaffolding was then performed with ARBitR[57], which incorporates the
299 distance information from linked-read sequencing data when merging and scaffolding
300 assemblies. Longranger BASIC quality-trimmed linked genome reads (outputs from the standard
301 10X Genomics data workflow) were mapped to the scaffolded genome assemblies for ARBitR
302 scaffolding, yielding the final genome assemblies: CCMP2556 (assembly size = 1.70 Gb; N50 =
303 750 Kb; 29,137 scaffolds) and SCF082 (assembly size = 1.64 Gb; N50 = 398.5 Kb; 44,682
304 scaffolds) **(**Table S2). The CCMP2556 assembly is the most contiguous reported in
305 Symbiodiniaceae aside from the recent chromosome-level assemblies for *Symbiodinium*
306 *microadriaticum*[58] and *Breviolum minutum*[59].

307 Genome and gene features of dinoflagellates are highly idiosyncratic and atypical of eukaryotes,
308 due in part to non-canonical splice sites[60]. Therefore, the prediction of protein-coding genes from
309 dinoflagellate genomes requires a comprehensive workflow
310 (https://github.com/TimothyStephens/Dinoflagellate_Annotation_Workflow/) tailored for these
311 features, guided by high-confidence evidence[61]. Here, we adopted a customised workflow
312 integrating the results from multiple methods, guided by available transcript and protein
313 sequences, independently for CCMP2556 and SCF082; see Supplementary Information for
314 detail.

315 **Analysis of whole-genome duplication**

316 We first searched for evidence of collinear gene blocks using MCScanX[62] in intra-species mode
317 (*-b 1*) to identify putative duplicate gene blocks within each genome (i.e. segmental and/or

318 whole-genome duplication), and in inter-species mode (*-b 2*) to identify syntenic gene blocks
319 between the two genomes. A collinear block is defined as at least five genes conserved in the
320 same orientation and order as a result of segmental duplication and/or WGD events. For each
321 comparison, all-vs-all BLASTP search results were restricted to the top five hits (query or
322 subject coverage > 50%; $E \leq 10^{-5}$). Predicted genes from each genome were classified using
323 *duplicate_gene_classifier* (within MCScanX) into singleton, dispersed duplicates (i.e. duplicates
324 separated by >20 genes), proximal duplicates (i.e. duplicates separated by <20 genes), tandem
325 duplicates, and WGD/segmental duplicates (i.e. ohnologs).

326 Second, we assessed the reconciliation between each gene tree and the species tree; the
327 topological incongruence between the two trees indicates history of gene duplication or loss[63]
328 OrthoFinder v2.3.10[64] was first used to infer homologous gene sets among Suessiales species
329 using BLASTP (E-value $\leq 10^{-5}$). Multiple sequence alignments were performed with MAFFT
330 v7.487[65] (*-linsi*), from which phylogenetic trees were inferred using FastTree v2.1.11[66] at default
331 parameters. Reconciliation of gene-tree and species-tree within OrthoFinder was then used to
332 identify lineage-specific duplication events; those specific to *D. trenchii* indicative of WGD-
333 derived duplicated genes (i.e. ohnologs).

334 Third, we assessed the impact of WGD on the rate of synonymous substitution (K_s) among all
335 homologous gene sets, using CCMP2556 as the reference, following the wgd pipeline[67]. Briefly,
336 homologous protein clusters were inferred using a Markov Clustering algorithm[68] from the
337 previous all-versus-all BLASTP search (used for MCScanX), and aligned using MAFFT[65].
338 Phylogenetic tree for each homologous protein cluster was inferred using FastTree2[66] and used
339 to estimate K_s values for each cluster using codeml implemented in PAML[69]. A Gaussian-
340 mixture model was applied to the K_s distribution, using a four-component model that provided

341 the best fit for the data according to Akaike information criterion (AIC), yielding a final node-
342 averaged histogram of K_s distribution. To estimate the timing of WGD, we first calculated the
343 estimated substitution rate (*r*) per year in Symbiodiniaceae adapting the approach of Ladner *et*
344 *al.*[26] to incorporate genome data and the updated divergence time estimates from LaJeunesse *et*
345 *al.*[5].

346 We followed Aury *et al.*[20] to infer metabolic pathways that were preferentially retained in
347 duplicate following WGD using PRIAM v2 (January 2018 release). Briefly, we identified
348 metabolic enzymes that had been uniquely retained as ohnologs or singletons. We then compared
349 the proportion of enzymes uniquely retained as ohnologs to singletons, to the background
350 proportion of the number of ohnologs and singletons annotated in the genome. This tests whether
351 the number of uniquely retained metabolic enzymes for a particular pathway exceeds the
352 background levels that would be expected to occur by random. We additionally required (a) five
353 or more distinct enzymatic proteins to be identified as uniquely retained in either duplicate or
354 singleton, and (b) pathways to be significantly overrepresented in both isolates. The proportion
355 of enzymes coded by genes that were uniquely retained as ohnologs or singletons, compared to
356 their overall proportions in the genome, was used to determine which KEGG pathways were
357 preferentially retained in duplicate following WGD.

358 **Evolution of ohnolog expression**

359 Trimmed RNA-Seq reads (above) were mapped to the corresponding genome using HISAT2
360 v2.2.1 (*--concordant-only*) with a Hierarchical Graph FM index informed by annotated exon and
361 splice sites. Counts of uniquely mapped paired-end (PE) reads overlapping with CDS regions
362 were then enumerated using *featureCounts* (*-p --countReadPairs --B -C*) implemented in
363 Subread v0.2.3[70]. The raw counts were filtered to remove lowly expressed genes using the

364 *filtrByExpr* function in edgeR. Differential gene expression analysis was performed with edgeR
365 using a generalized linear model. We considered genes to be differentially expressed when false
366 discovery rate (FDR) < 0.01 and the absolute value of $\log_2$(fold-change) > 1. We compared the
367 difference between lifestyles at two temperatures, i.e. symbiosis *versus* free-living at 34°C (L-
368 34), and symbiosis *versus* free-living at 28°C (L-28), and the response to temperature stress in
369 the two lifestyles, i.e. 34°C *versus* 28°C at free-living (T-Fr), and 34°C *versus* 28°C in symbiosis
370 (T-Sy).

371 A weighted gene co-expression network analysis (WGCNA) was performed on all genes in R
372 using the WGCNA package. Variance of normalised counts were calculated using the standard
373 DESeq2 workflow followed by its *varianceStabilizingTransformation*. Because symbiosis is a
374 strong driver of expression in Symbiodiniaceae, using the inferred soft-thresholding power for
375 reducing noise and setting a required threshold for gene correlations would have yielded a mean
376 connectivity of over 4,000 at the inferred power of 6. Therefore, a weighted, unidirectional co-
377 expression network was inferred using a power of 18, the recommended value for signed
378 networks with less than 20 samples. Co-expression modules were inferred using the function
379 *blockwiseModules* that collectively infers signed networks (networkType="signed",
380 TOMtype="signed", maxBlockSize=10000, corType="bicor", maxPoutliers=0.05,
381 pearsonFallback="individual", deepSplit=2, dcuth=0.999, minModuleSize=30,
382 reassignThreshold=0.1, cutHeight=0.2).

383

384 We calculated the adjacencies using a signed network with *bicor* robust correlation coefficient
385 (power=18, type="signed", corFnc=bicor, maxPoutliers=0.1, pearsonFallback="individual"). A

386 topological association matrix was then inferred with a signed network and *dissTOM* computed
387 from the product. A hierarchical dendrogram of genes was inferred using *hclust*
388 (method="average"). The dendrogram was cut using *cutreeDynamic* (deepSplit=2,
389 minClusterSize=15, cutHeight=0.999) and the cut dendrogram merged with *mergeCloseModules*
390 (cutHeight=0.15, corFnc=bicor, maxPoutliers=0.1, pearsonFallback="individual"). Preferential
391 distribution of the different gene duplication categories to WGCNA modules was assessed with a
392 chi-square test and a post hoc analysis performed with the R package *chisq.posthoc.test*
393 (https://github.com/ebbertd/chisq.posthoc.test.git).

394 Expression specificity of ohnologs was assessed using the tau (τ) index[29], where $\tau = 1$ indicates
395 highly specific expression, and $\tau = 0$ indicates broad expression. The log-normalised fragments
396 per kilobase million (FPKM) counts were used to calculate τ index scores for those genes with a
397 $\log_2(\text{FPKM} + 1) > 1$ in at least one condition following Yanai *et al.*[29]. The τ indices for the
398 different MCScanX duplication categories were compared using a Kruskal-Wallis rank sum test;
399 pairwise comparisons using Wilcoxon rank sum test with continuity correction and holm *p*-value
400 adjustment were performed to determine differences between the duplication categories. A chi-
401 square test of all significant τ indices ($\tau \geq 0.7$) was conducted to assess potential biases in
402 expression specificity for treatments among the duplication categories.

403 **Analysis of post-transcriptional regulation**

404 All-versus-all BLASTN search (query or subject coverage > 50%; $E \leq 10^{-20}$) was used to
405 identify shared exonic sequences that have been retained since WGD. For inferring differential
406 exon usage (DEU) within genes among the treatment conditions, gene models were first broken
407 up into exon "counting bins" using the Python script *dexseq_prepare_annotation.py* from
408 DEXSeq. The relative usage of each exon bin, i.e. the number of transcripts mapping to the bin

409 or to the gene, was then calculated from the HISAT2 BAM file using *dexseq_count.py*. The
410 DEXSeq R package was then used to infer differential exon usage within genes using a
411 generalised linear model, correcting for significance at the gene level using the Benjamini-
412 Hochberg method[71].

413 To examine the conservation of splice junctions in ohnolog pairs, all *de novo* assembled
414 transcripts were first aligned to the genome using a minimap2 v2.20[52] with code modified to
415 recognize alternative splice sites in dinoflagellates, from which splices sites were identified and
416 annotated using PASA[72]. Splice sites categorized as alternative acceptor, alternative donor,
417 alternative exon, retained exon, and skipped exon were retained for subsequent analysis. Each
418 identified splice event was assigned two unique identifiers to represent the upstream and
419 downstream positions of the splice event, along with its gene identifier and genomic location.
420 The upstream and downstream 300bp-region for each splice event were then extracted using the
421 bedtools v2.30 *flank* and *getfasta* functions. An all-*versus*-all BLASTN search of the extracted
422 splice junction sequences was used to identify sequence similarity ($E \leq 10^{-5}$) between the
423 sequences. Custom Python scripts were used to filter the BLASTN results to identify conserved
424 splice junctions, in which both upstream and downstream regions for a splice event in two
425 ohnologs were significantly similar ($E \leq 10^{-5}$). The splice junction profile for each ohnolog pair
426 was then converted to a binary representation, where the presence of a splice junction in an
427 ohnolog was represented as 1 and the absence of a splice junction represented as 0 (i.e.,
428 conserved splice junctions represented as 1 in both ohnologs compared to 0 for those that were
429 not conserved). A Kendall's rank correlation was then conducted in R to identify ohnolog pairs
430 that exhibited high level of conservation in splice junctions. An exact binomial test was also

431 performed to identify ohnolog pairs that had diverged in terms of total splice junctions ($p <$

432 0.05).

433 **Data availability**

434 The genome data generated from this study for the two *D. trenchii* isolates are available at NCBI

435 through BioProject accession PRJEB66001. The assembled genomes, predicted gene models,

436 and proteins for *D. trenchii* CCMP2556 and SCF082 are available

437 at https://doi.org/10.48610/27da3e7.

438 ## References

439 1 Rowan, R. Thermal adaptation in reef coral symbionts. *Nature* **430**, 742 (2004).

440 2 Berkelmans, R. & van Oppen, M. J. H. The role of zooxanthellae in the thermal tolerance
441 of corals: a 'nugget of hope' for coral reefs in an era of climate change. *Proc. R. Soc. B*
442 **273**, 2305-2312 (2006).

443 3 González-Pech, R. A., Bhattacharya, D., Ragan, M. A. & Chan, C. X. Genome evolution of
444 coral reef symbionts as intracellular residents. *Trends in Ecology & Evolution* **34**, 799-806
445 (2019).

446 4 Pettay, D. T. & Lajeunesse, T. C. Microsatellite loci for assessing genetic diversity,
447 dispersal and clonality of coral symbionts in 'stress-tolerant'clade D *Symbiodinium*. *Mol.*
448 *Ecol. Resour.* **9**, 1022-1025 (2009).

449 5 LaJeunesse, T. C. *et al.* Systematic revision of Symbiodiniaceae highlights the antiquity
450 and diversity of coral endosymbionts. *Curr. Biol.* **28**, 2570-2580 (2018).

451 6 Rädecker, N. *et al.* Heat stress destabilizes symbiotic nutrient cycling in corals. *Proc. Natl.*
452 *Acad. Sci. U. S. A.* **118**, e2022653118 (2021).

453 7 Xiang, T. *et al.* Symbiont population control by host-symbiont metabolic interaction in
454 Symbiodiniaceae-cnidarian associations. *Nat. Commun.* **11**, 108 (2020).

455 8 LaJeunesse, T. C. *et al.* Revival of *Philozoon* Geddes for host-specialized dinoflagellates,
456 'Zooxanthellae', in animals from coastal temperate zones of Northern and Southern
457 Hemispheres. *Eur. J. Phycol.*, 1-15 (2021).

458 9 Pochon, X. & LaJeunesse, T. C. *Miliolidium* n. gen, a new symbiodiniacean genus whose
459 members associate with soritid foraminifera or are free-living. *J. Eukaryot. Microbiol.* **68**,
460 e12856 (2021).

461 10 Nitschke, M. R. *et al.* Description of *Freudenthalidium* gen. nov. and *Halluxium* gen. nov.
462 to formally recognize clades Fr3 and H as genera in the family Symbiodiniaceae
463 (Dinophyceae). *J. Phycol.* **56**, 923-940 (2020).

464 11 González-Pech, R. A. *et al.* Comparison of 15 dinoflagellate genomes reveals extensive
465 sequence and structural divergence in family Symbiodiniaceae and genus *Symbiodinium*.
466 *BMC Biol.* **19**, 73 (2021).

467 12 Van de Peer, Y., Maere, S. & Meyer, A. The evolutionary significance of ancient genome
468 duplications. *Nat. Rev. Genet.* **10**, 725-732 (2009).

469 13 Schranz, M. E., Mohammadin, S. & Edger, P. P. Ancient whole genome duplications,
470 novelty and diversification: the WGD Radiation Lag-Time Model. *Curr. Opin. Plant Biol.*
471 **15**, 147-153 (2012).

472 14 Conant, G. C. & Wolfe, K. H. Turning a hobby into a job: how duplicated genes find new
473 functions. *Nat. Rev. Genet.* **9**, 938-950 (2008).

474 15 Ohno, S., Wolf, U. & Atkin, N. B. Evolution from fish to mammals by gene duplication.
475 *Hereditas* **59**, 169-187 (1968).

476 16 Singh, P. P., Arora, J. & Isambert, H. Identification of ohnolog genes originating from
477 whole genome duplication in early vertebrates, based on synteny comparison across
478 multiple genomes. *PLoS Comput. Biol.* **11**, e1004394 (2015).

479 17 Kellis, M., Birren, B. W. & Lander, E. S. Proof and evolutionary analysis of ancient
480 genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**, 617-624 (2004).

481 18 Gallone, B. *et al.* Domestication and divergence of *Saccharomyces cerevisiae* beer yeasts.
482 *Cell* **166**, 1397-1410 (2016).

483 19 Peter, J. *et al.* Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature*
484 **556**, 339-344 (2018).

485 20 Aury, J.-M. *et al.* Global trends of whole-genome duplications revealed by the ciliate
486 *Paramecium tetraurelia*. *Nature* **444**, 171-178 (2006).

487 21 McGrath, C. L., Gout, J.-F., Doak, T. G., Yanagi, A. & Lynch, M. Insights into three
488 whole-genome duplications gleaned from the *Paramecium caudatum* genome sequence.
489 *Genetics* **197**, 1417-1428 (2014).

490 22 Maeda, Y. *et al.* Chromosome-scale genome assembly of the marine oleaginous diatom
491 *Fistulifera solaris*. *Mar. Biotechnol.* **24**, 788-800 (2022).

492 23 Tanaka, T. *et al.* Oil accumulation by the oleaginous diatom *Fistulifera solaris* as revealed
493 by the genome and transcriptome. *Plant Cell* **27**, 162-176 (2015).

494 24 Pettay, D. T., Wham, D. C., Smith, R. T., Iglesias-Prieto, R. & LaJeunesse, T. C. Microbial
495 invasion of the Caribbean by an Indo-Pacific coral zooxanthella. *Proc. Natl. Acad. Sci. U.*
496 *S. A.* **112**, 7513-7518 (2015).

497 25 Tiley, G. P., Barker, M. S. & Burleigh, J. G. Assessing the performance of Ks plots for
498 detecting ancient whole genome duplications. *Genome Biol. Evol.* **10**, 2882-2898 (2018).

499 26 Ladner, J. T., Barshis, D. J. & Palumbi, S. R. Protein evolution in two co-occurring types
500 of *Symbiodinium*: an exploration into the genetic basis of thermal tolerance in
501 *Symbiodinium* clade D. *BMC Evol. Biol.* **12**, 217 (2012).

502 27 Voris, H. K. Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems
503 and time durations. *J. Biogeogr.* **27**, 1153-1167 (2000).

504 28 Bellantuono, A. J., Dougan, K. E., Granados-Cifuentes, C. & Rodriguez-Lanetty, M. Free-
505 living and symbiotic lifestyles of a thermotolerant coral endosymbiont display profoundly
506 distinct transcriptomes under both stable and heat stress conditions. *Mol. Ecol.* **28**, 5265-
507 5281 (2019).

508 29 Yanai, I. *et al.* Genome-wide midrange transcription profiles reveal expression level
509 relationships in human tissue specification. *Bioinformatics* **21**, 650-659 (2005).

510 30 Prince, V. E. & Pickett, F. B. Splitting pairs: the diverging fates of duplicated genes. *Nat.*
511 *Rev. Genet.* **3**, 827-837 (2002).

512 31 Lehner, B. Conflict between noise and plasticity in yeast. *PLoS Genet.* **6**, e1001185 (2010).

513 32 Matthews, J. L. *et al.* Partner switching and metabolic flux in a model cnidarian–
514 dinoflagellate symbiosis. *Proc. R. Soc. B* **285** (2018).

515 33 Yuyama, I., Ishikawa, M., Nozawa, M., Yoshida, M.-a. & Ikeo, K. Transcriptomic changes
516 with increasing algal symbiont reveal the detailed process underlying establishment of
517 coral-algal symbiosis. *Sci. Rep.* **8**, 16802 (2018).

518 34 Silverstein, R. N., Cunning, R. & Baker, A. C. Tenacious D: *Symbiodinium* in clade D
519 remain in reef corals at both high and low temperature extremes despite impairment. *J.*
520 *Exp. Biol.* **220**, 1192-1196 (2017).

521 35 Claar, D. C. *et al.* Dynamic symbioses reveal pathways to coral survival through prolonged
522 heatwaves. *Nat. Commun.* **11**, 6097 (2020).

523 36 Abrego, D., Willis, B. L. & van Oppen, M. J. Impact of light and temperature on the uptake
524 of algal symbionts by coral juveniles. *PLoS ONE* **7**, e50311 (2012).

525 37 Herrera, M. *et al.* Temperature transcends partner specificity in the symbiosis
526 establishment of a cnidarian. *ISME J.* **15**, 141-153 (2021).

527 38 Matsuda, S. B. *et al.* Temperature-mediated acquisition of rare heterologous symbionts
528 promotes survival of coral larvae under ocean warming. *Glob. Chang. Biol.* **28**, 2006-2025
529 (2022).

530 39 Conant, G. C. & Wolfe, K. H. Functional partitioning of yeast co-expression networks after
531 genome duplication. *PLoS Biol.* **4**, e109 (2006).

532 40 Chakravarti, L. J. & van Oppen, M. J. Experimental evolution in coral photosymbionts as a
533 tool to increase thermal tolerance. *Front. Mar. Sci.* **5**, 227 (2018).

534 41 Carretero-Paulet, L. & Van de Peer, Y. The evolutionary conundrum of whole-genome
535 duplication. *Am. J. Bot.* **107**, 1101 (2020).

536 42 Howells, E. J. *et al.* Corals in the hottest reefs in the world exhibit symbiont fidelity not
537 flexibility. *Mol. Ecol.* **29**, 899-911 (2020).

538 43 Hume, B. C. *et al. Symbiodinium thermophilum* sp. nov., a thermotolerant symbiotic alga
539 prevalent in corals of the world's hottest sea, the Persian/Arabian Gulf. *Sci. Rep.* **5**, 8562
540 (2015).

541 44 Dougan, K. E. *et al.* Genome-powered classification of microbial eukaryotes: focus on
542 coral algal symbionts. *Trends Microbiol.* **30**, 831-840 (2022).

543 45 Drew, E. A. The biology and physiology of alga-invertebrates symbioses. II. The density of
544 symbiotic algal cells in a number of hermatypic hard corals and alcyonarians from various
545 depths. *J. Exp. Mar. Biol. Ecol.* **9**, 71-75 (1972).

546 46 Porter, J. W., Muscatine, L., Dubinsky, Z. & Falkowski, P. G. Primary production and
547 photoadaptation in light- and shade-adapted colonies of the symbiotic coral, *Stylophora*
548 *pistillata*. *Proc. R. Soc. B* **222**, 161-180 (1984).

549 47 Bhattacharya, D., Stephens, T. G., Chille, E. E., Benites, L. F. & Chan, C. X. Facultative
550 lifestyle drives diversity of coral algal symbionts. *Trends Ecol. Evol.*, accepted; in press
551 (2024).

552 48 Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. GenomeScope 2.0 and Smudgeplot
553 for reference-free profiling of polyploid genomes. *Nat. Commun.* **11**, 1432 (2020).

554 49 Iha, C. *et al.* Genomic adaptations to an endolithic lifestyle in the coral-associated alga
555 *Ostreobium*. *Curr. Biol.* **31**, 1393-1402 (2021).

556 50 Laetsch, D. R. & Blaxter, M. L. BlobTools: Interrogation of genome assemblies.
557 *F1000Res.* **6**, 1287-1287 (2017).

558 51 Marks, P. *et al.* Resolving the full spectrum of human genome variation using linked-reads.
559 *Genome Res.* **29**, 635-645 (2019).

560 52 Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-
561 3100 (2018).

562 53 Camp, E. F. *et al.* Proteome metabolome and transcriptome data for three Symbiodiniaceae
563 under ambient and heat stress conditions. *Sci. Data* **9**, 153 (2022).

564 54 Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor.
565 *Bioinformatics* **34**, i884-i890 (2018).

566 55 Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the
567 Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494-1512 (2013).

568 56 Xue, W. *et al.* L_RNA_scaffolder: scaffolding genomes with transcripts. *BMC Genomics*
569 **14**, 604 (2013).

570 57 Hiltunen, M., Ryberg, M. & Johannesson, H. ARBitR: an overlap-aware genome assembly
571 scaffolder for linked reads. *Bioinformatics* **37**, 2203-2205 (2021).

572 58 Nand, A. *et al.* Genetic and spatial organization of the unusual chromosomes of the
573 dinoflagellate *Symbiodinium microadriaticum*. *Nat. Genet.* **53**, 618-629 (2021).

574 59 Marinov, G. K. *et al.* Transcription-dependent domain-scale three-dimensional genome
575 organization in the dinoflagellate *Breviolum minutum*. *Nat. Genet.* **53**, 613-617 (2021).

576 60 Wisecaver, J. H. & Hackett, J. D. Dinoflagellate genome evolution. *Annu. Rev. Microbiol.*
577 **65**, 369-387 (2011).

578 61 Chen, Y., González-Pech, R. A., Stephens, T. G., Bhattacharya, D. & Chan, C. X.
579 Evidence that inconsistent gene prediction can mislead analysis of dinoflagellate genomes.
580 *J. Phycol.* **56**, 6-10 (2020).

581 62 Wang, Y. *et al.* MCScanX: a toolkit for detection and evolutionary analysis of gene
582 synteny and collinearity. *Nucleic Acids Res.* **40**, e49 (2012).

583 63 Goodman, M., Czelusniak, J., Moore, G. W., Romero-Herrera, A. E. & Matsuda, G. Fitting
584 the gene lineage into its species lineage, a parsimony strategy illustrated by cladograms
585 constructed from globin sequences. *Syst. Biol.* **28**, 132-163 (1979).

586 64 Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative
587 genomics. *Genome Biol.* **20**, 238 (2019).

588 65 Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7:
589 improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772-780 (2013).

590 66 Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2–approximately maximum-likelihood
591 trees for large alignments. *PLoS ONE* **5**, e9490 (2010).

592 67 Zwaenepoel, A. & Van de Peer, Y. wgd—simple command line tools for the analysis of
593 ancient whole-genome duplications. *Bioinformatics* **35**, 2153-2155 (2019).

594 68 Enright, A. J., Van Dongen, S. & Ouzounis, C. A. Graph clustering by flow simulation.
595 *Nucleic Acids Res.* **30**, 1575-1584 (2000).

596 69 Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**,
597 1586-1591 (2007).

598 70 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for
599 assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014).

600 71 Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and
601 powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **57**, 289-300 (1995).

602 72 Haas, B. J. *et al.* Automated eukaryotic gene structure annotation using EVidenceModeler
603 and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, R7 (2008).

604

605 **Acknowledgements**

606 This Project was supported by funding from the Australian Research Council grants
607 DP190102474 awarded to C.X.C. and D.B. and FL180100036 to M.J.H.v.O., the Australian
608 Academy of Science Thomas Davies Grant for Marine, Soil, and Plant Biology awarded to
609 C.X.C., the National Science Foundation grant CAREER1453519 awarded to M.R.L., the City
610 University of New York grant PSC-CUNY 69757-0047 awarded to C.G.C., the National Institute
611 of Food and Agriculture-US Department of Agriculture Hatch grant NJ01180 awarded to D.B.,
612 and funding from the University of Technology Sydney to T.K. and D.J.S. This project is also
613 supported by computational resources of the National Computational Infrastructure (NCI)
614 National Facility systems through the NCI Merit Allocation Scheme (Project d85) awarded to
615 C.X.C. We thank Brian Kemish and David Green for their technical assistance during genome
616 assembly as well as Joel Burke and Michelle Havlik for technical assistance with the algal
617 cultures.

618 **Author contributions**

619 K.E.D., A.J.B., D.J.S., C.X.C., and M.R.L. conceptualized the study; K.E.D., D.B., and C.X.C.
620 drafted the initial manuscript with A.J.B., M.J.H.v.O., D.J.S., and M.R.L. helped finalize the
621 text; A.J.B. and R.M.A. maintained the cultures and extracted genomic DNA samples; K.E.D.
622 performed all bioinformatic analyses except for the analyses on RNA editing and organellar
623 genomes done by Y.C. and S.S., respectively. Y.C. performed RNA editing analysis and S.S.
624 organellar genome analysis. All authors read and approved the manuscript.

625 **Competing interests**

626 The authors declare no competing interests.

627

628 **Figure legends**

629 **Figure 1. WGD in a facultative coral endosymbiont.** (**a**) Microscopic images of a free-living
630 *D. trenchii* cell and a *Exaiptasia pallida* anemone hosting *D. trenchii* under fluorescence, with
631 red indicating the presence of *D. trenchii*. (**b**) Repeat landscapes shown separately for the
632 CCMP2556 and SCF082 genomes. (**c**) Circle plot depicting the location of syntenic blocks
633 containing collinear gene blocks (i.e. ohnologs) between the CCMP2556 and SCF082 genomes.
634 Ribbons indicate syntenic gene blocks identified with MCScanX that overlap with putative
635 WGD-duplicated regions in both isolates (blue; n=2,427), one isolate only (red; n=612), or
636 neither isolate (black; n=35). (**d**) The percentage of genes in duplicated collinear gene blocks
637 relative to the number of duplicated collinear gene blocks identified within the genomes of
638 Suessiales species. (**e**) Number of genes and syntenic genes recovered for each gene duplication
639 category for the two isolates. (**f**) Phylogenetic tree of Order Suessiales showing the number of
640 lineage-specific gene-family duplications at each node.

641 **Figure 2. Ohnolog expression post-WGD.** (**a**) Three-dimensional scatterplot of the five groups
642 of ohnologs pairs based on their pattern of differential expression (DE), i.e. pairs for which:
643 neither gene showed DE (Group 1; blue); only one showed DE (Group 2; orange); both ohnologs
644 showed DE at the same time in the same manner (Group 3; green); both ohnologs showed DE
645 but at different times (Group 4; purple); and both ohnologs showed DE at the same time but in
646 opposing directions (Group 5; red). (**b**) Pearson correlation coefficients showing correlation of
647 expression patterns between each ohnolog pair within each of the five groups. (**c**) Exon gain/loss
648 between each ohnologs pair within each of the five groups. Heatmaps depicting the normalized
649 gene expression (*z*-score) for (**d**) Group 2 and (**e**) Group 5.

650 **Figure 3. Exon reorganization underlies functional divergence.** (**a**) Three-dimensional
651 scatterplot depicting directionality of DEU among ohnologs pairs of Group 2 and Group 5, which
652 reflects the pattern of gene-level DEU. The *z*-axis shows the absolute change in DEU (i.e. overall
653 DEU) within each ohnolog pair from Groups 2 and 5, the *x*-axis shows the relative change in DEU
654 (i.e. proportion of DEU), and the *y*-axis indicates Pearson correlation coefficient of the gene
655 expression. An ohnolog-pair from Group 2 with (**b**) DEU in the ohnolog with gene-level DEU, and
656 (**c**) no DEU in its counterpart. (**d**) Glycolysis and gluconeogenesis pathways for which genes
657 indicated in red were implicated by differentially expressed ohnologs, and an asterisk indicating rate-
658 limiting or key enzymes. (**e**) The $\log_2$(fold-change) in expression of differentially expressed ohnologs
659 across distinct comparison of growth conditions with their corresponding scenario indicated on the
660 right.

661 **Figure 4. Model of post-WGD divergence in a facultative endosymbiont.** Putative selective
662 constraints faced by free-living and symbiotic Symbiodiniaceae under the dual lifestyle are
663 shown, with a focus on post-WGD ohnolog sequence divergence and differential gene
664 expression.

665 **List of supplementary materials**

666 Supplementary Information

667 Figs. S1-S25

668 Tables S1-S22

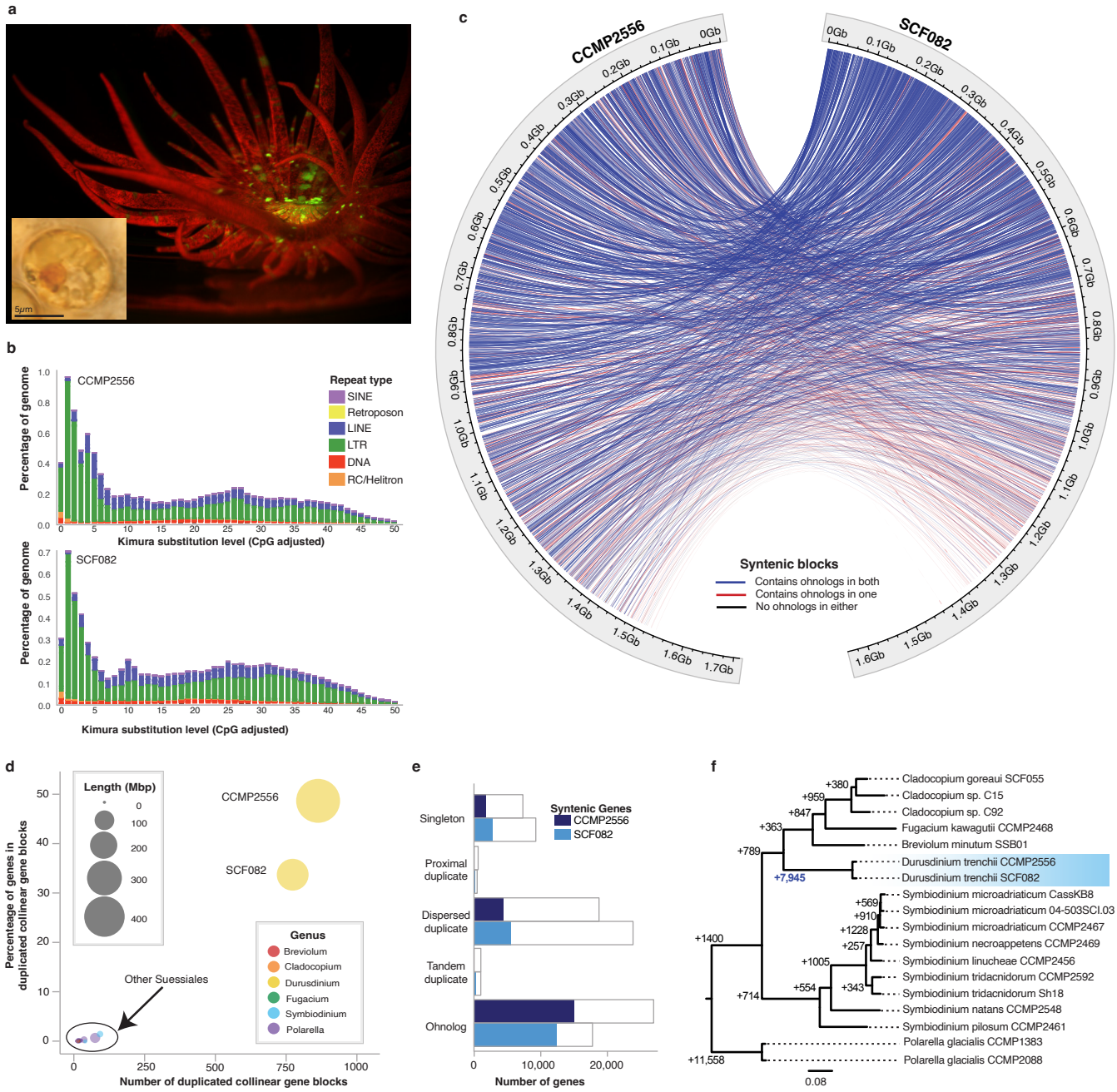

Figure 1. WGD in a facultative coral endosymbiont. (**a**) Microscopic images of a free-living *D. trenchii* cell and a *Exaiptasia pallida* anemone hosting *D. trenchii* under fluorescence, with red indicating the presence of *D. trenchii*. (**b**) Repeat landscapes shown separately for the CCMP2556 and SCF082 genomes. (**c**) Circle plot depicting the location of syntenic blocks containing collinear gene blocks (i.e. ohnologs) between the CCMP2556 and SCF082 genomes. Ribbons indicate syntenic gene blocks identified with MCScanX that overlap with putative WGD-duplicated regions in both isolates (blue; n=2,427), one isolate only (red; n=612), or neither isolate (black; n=35). (**d**) The percentage of genes in duplicated collinear gene blocks relative to the number of duplicated collinear gene blocks identified within the genomes of Suessiales species. (**e**) Number of genes and syntenic genes recovered for each gene duplication category for the two isolates. (**f**) Phylogenetic tree of Order Suessiales showing the number of lineage-specific gene-family duplications at each node.

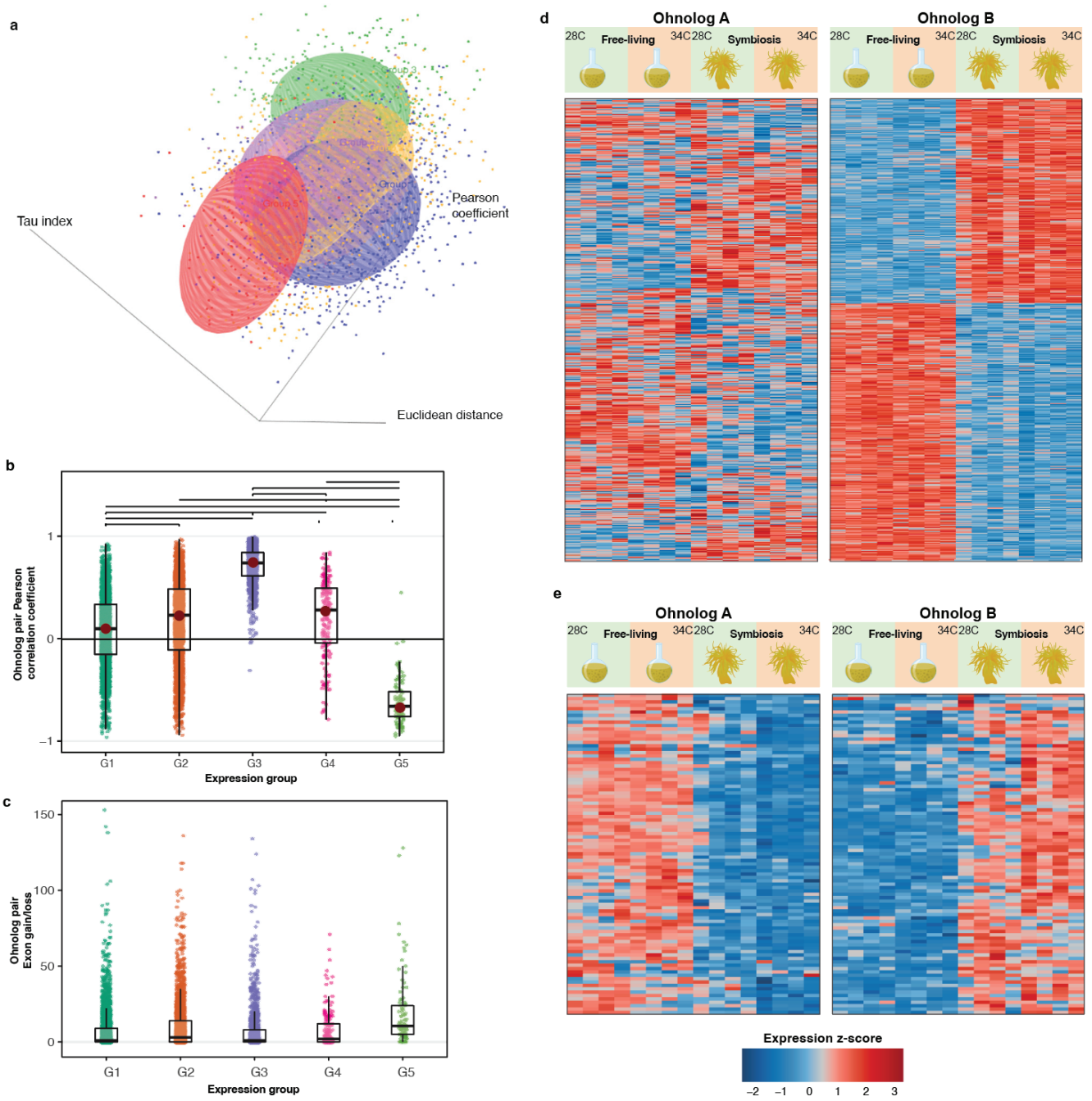

Figure 2. Ohnolog expression post-WGD. (**a**) Three-dimensional scatterplot of the five groups of ohnologs pairs based on their pattern of differential expression (DE), i.e. pairs for which: neither gene showed DE (Group 1; blue); only one showed DE (Group 2; orange); both ohnologs showed DE at the same time in the same manner (Group 3; green); both ohnologs showed DE but at different times (Group 4; purple); and both ohnologs showed DE at the same time but in opposing directions (Group 5; red). (**b**) Pearson correlation coefficients showing correlation of expression patterns between each ohnolog pair within each of the five groups. (**c**) Exon gain/loss between each ohnologs pair within each of the five groups. Heatmaps depicting the normalized gene expression (*z*-score) for (**d**) Group 2 and (**e**) Group 5.

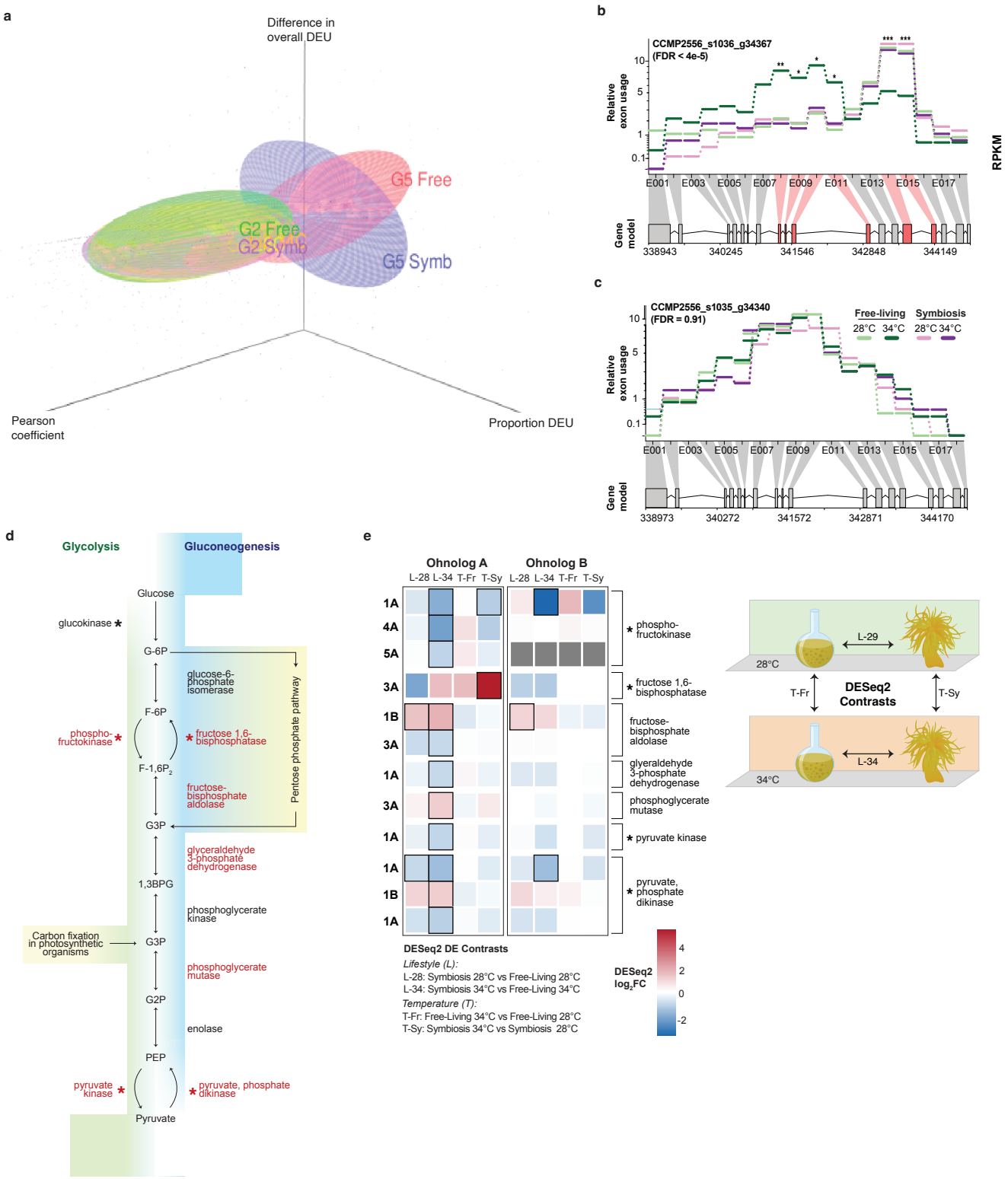

Figure 3. Exon reorganization underlies functional divergence. (**a**) Three-dimensional scatterplot depicting directionality of DEU among ohnologs pairs of Group 2 and Group 5, which reflects the pattern of gene-level DEU. The z-axis shows the absolute change in DEU (i.e. overall DEU) within each ohnolog pair from Groups 2 and 5, the x-axis shows the relative change in DEU (i.e. proportion of DEU), and the y-axis indicates Pearson correlation coefficient of the gene expression. An ohnolog-pair from Group 2 with (**b**) DEU in the ohnolog with gene-level DEU, and (**c**) no DEU in its counterpart. (**d**) Glycolysis and gluconeogenesis pathways for which genes indicated in red were implicated by differentially expressed ohnologs, and an asterisk indicating rate-limiting or key enzymes. (**e**) The $\log_2$(fold-change) in expression of differentially expressed ohnologs across distinct comparison of growth conditions with their corresponding scenario indicated on the right.

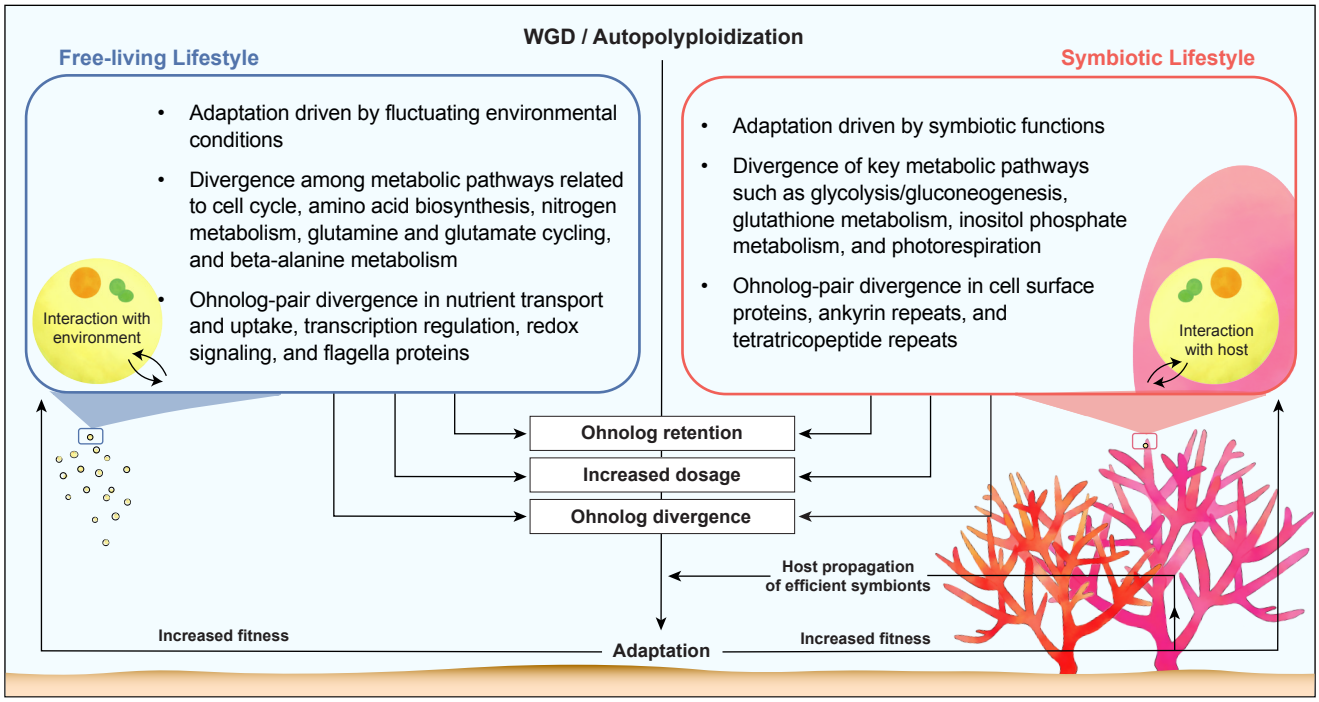

Figure 4. Model of post-WGD divergence in a facultative endosymbiont. Putative selective constraints faced by free-living and symbiotic Symbiodiniaceae under the dual lifestyle are shown, with a focus on post-WGD ohnolog sequence divergence and differential gene expression.