

1 Article

2 ***Symbiodinium* genomes reveal adaptive evolution of**

3 **functions related to symbiosis**

4 Huanle Liu¹, Timothy G. Stephens¹, Raúl A. González-Pech¹, Victor H. Beltran², Bruno
5 Lapeyre^{3,4}, Pim Bongaerts⁵, Ira Cooke³, David G. Bourne^{2,6}, Sylvain Forêt^{7,*}, David J.
6 Miller³, Madeleine J. H. van Oppen^{2,8}, Christian R. Voolstra⁹, Mark A. Ragan¹ and Cheong
7 Xin Chan^{1,10,†}

8 ¹Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD 4072,
9 Australia

10 ²Australian Institute of Marine Science, Townsville, QLD 4810, Australia

11 ³ARC Centre of Excellence for Coral Reef Studies and Department of Molecular and Cell
12 Biology, James Cook University, Townsville, QLD 4811, Australia

13 ⁴Laboratoire d'excellence CORAIL, Centre de Recherches Insulaires et Observatoire de
14 l'Environnement, Moorea 98729, French Polynesia

15 ⁵Global Change Institute, The University of Queensland, Brisbane, QLD 4072, Australia

16 ⁶College of Science and Engineering, James Cook University, Townsville, QLD 4811,
17 Australia

18 ⁷Research School of Biology, Australian National University, Canberra, ACT 2601, Australia

19 ⁸School of BioSciences, The University of Melbourne, VIC 3010, Australia

20 ⁹Red Sea Research Center, Division of Biological and Environmental Science and
21 Engineering, King Abdullah University of Science and Technology (KAUST), Thuwal
22 23955-6900, Kingdom of Saudi Arabia

23 ¹⁰School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane,
24 QLD 4072, Australia

25 *Deceased 17 December 2016

26 †Correspondence and requests for materials should be addressed to Cheong Xin Chan
27 (c.chan1@uq.edu.au)

28 **Abstract**

29 Symbiosis between dinoflagellates of the genus *Symbiodinium* and reef-building corals forms
30 the trophic foundation of the world's coral reef ecosystems. Here we present the first draft
31 genome of *Symbiodinium goreauii* (Clade C, type C1: 1.03 Gbp), one of the most ubiquitous
32 endosymbionts associated with corals, and an improved draft genome of *Symbiodinium*
33 *kawagutii* (Clade F, strain CS-156: 1.05 Gbp), previously sequenced as strain CCMP2468, to
34 further elucidate genomic signatures of this symbiosis. Comparative analysis of four available
35 *Symbiodinium* genomes against other dinoflagellate genomes led to the identification of 2460
36 nuclear gene families that show evidence of positive selection, including genes involved in
37 photosynthesis, transmembrane ion transport, synthesis and modification of amino acids and
38 glycoproteins, and stress response. Further, we identified extensive sets of genes for meiosis
39 and response to light stress. These draft genomes provide a foundational resource for
40 advancing our understanding *Symbiodinium* biology and the coral-algal symbiosis.

41

42 Introduction

43 Coral reefs provide habitats for one-quarter to one-third of all marine species¹. Although
44 typically surrounded by nutrient-poor waters, coral reefs show high rates of primary
45 productivity, with the fixed carbon supporting not only the biomass of reef organisms but
46 also commercial and recreational fisheries and aquaculture. Reef-building corals rely on the
47 symbiosis between the coral animal *per se* and photosynthetic dinoflagellates of the genus
48 *Symbiodinium*. This symbiosis is based on mutual nutrient exploitation, with corals providing
49 shelter and inorganic nutrients to their algal partners, while *Symbiodinium* supply their coral
50 hosts with photosynthates that can meet up to 95% of the corals' energy requirements^{2,3}.

51 The relationship between *Symbiodinium* and their host determines not only the rate of
52 coral-reef growth (calcium carbonate deposition), but also how the system responds to
53 environmental stress³. Many studies have shown that coral-*Symbiodinium* mutualism is
54 susceptible to environmental factors including temperature, light and salinity. Exposure to
55 ultraviolet radiation, thermal stress or a combination of both can initiate photoinhibition,
56 decoupling of carbon flow between symbiont and host, oxidative damage and breakdown of
57 the symbiosis, a phenomenon known as coral bleaching⁴. Unless the symbiosis is soon re-
58 established the coral host is at risk of starvation, disease and eventual death⁵. In recent
59 decades, coral bleaching has led to large-scale mortality on coral reefs around the world, with
60 the most recent global coral bleaching event (2014-2016) now confirmed as the longest and
61 most severe on record^{6,7}.

62 Despite the critical importance of this coral-dinoflagellate symbiosis, little is known
63 about the underlying molecular mechanisms (apart from photosynthesis and carbon
64 exchange), largely due to the lack of comprehensive understanding of what molecules,
65 pathways and functions *Symbiodinium* can contribute. Genomes of *Symbiodinium* (and of

66 dinoflagellates more broadly) are known for their idiosyncratic features including non-
67 canonical splice sites, extensive methylation⁸ and large sizes, up to 250 Gbp⁹. Their plastid
68 genomes occur as plasmid-like minicircles¹⁰⁻¹²; their mitochondrial genomes harbor only
69 three protein-coding genes and lack stop codons, and both mitochondrial¹³⁻¹⁵ and nuclear¹⁶
70 transcripts are extensively edited.

71 *Symbiodinium* are classified into nine clades¹⁷⁻¹⁹, with members of Clades A, B, C and
72 D responsible for the vast majority of associations with scleractinian corals²⁰. Draft genomes
73 have been published for representatives of Clades A, B and F¹⁷⁻¹⁹, with sequence comparisons
74 demonstrating them to be highly divergent¹⁸. Genome sequences are still lacking for Clade C,
75 the most ubiquitous and diverse clade associated with tropical reef corals²¹, at least some sub-
76 clades (“types”) of which are ecologically partitioned²².

77 Here we report draft genomes of two *Symbiodinium* from the Pacific Ocean: *S. goreau*
78 (type C1; isolated from the acroporid coral *Acropora tenuis*) from the Great Barrier Reef, and
79 *S. kawagutii* CS-156 (=CCMP2468, Clade F) from Hawaii. *Symbiodinium* type C1 is one of
80 two “living ancestors” (along with type C3) of Clade C²¹, and one of the most dominant types
81 associated with reef corals in both Indo-Pacific and Caribbean waters²⁰. *S. goreau* has been
82 reported from >150 coral species on Australia’s Great Barrier Reef, representing >80% of the
83 studied coral genera in this region²³ across environments from reef flats to lower mesophotic
84 depths²³⁻²⁵. In contrast, *S. kawagutii* CS-156 (=CCMP2468) was isolated during attempts to
85 culture the symbiont from *Montipora verrucosa* (Todd LaJeunesse, *personal*
86 *communication*). This isolate has yet to be verified to occur in mutualistic symbiosis with any
87 coral, and appears incapable of establishing experimental symbiosis with cnidarian hosts²⁶.
88 Instead *S. kawagutii* may be exclusively a symbiont of foraminifera, or occur free-living at
89 low environmental densities but proliferate opportunistically in culture. As some genome data

90 have been published for *S. kawagutii* CCMP2468¹⁸, we used these in combination with new
91 data from the present study to generate a refined genome assembly. The genomes of *S.*
92 *goreaui* and *S. kawagutii* offer a platform for comparative genomic analyses between two of
93 the most-recently diverged *Symbiodinium* lineages Clades C and F, and published genome
94 sequences in the more-basal Clades A and B.

95 Adopting a comparative approach using both genome and transcriptome data, we
96 systematically investigated genes and functions that are specific to *Symbiodinium* vis-à-vis
97 other dinoflagellates, and their association with the establishment and maintenance of
98 symbiosis. We also computationally identify genes and functions for which there is evidence
99 of adaptive selection in *Symbiodinium*. This is the most-comprehensive comparative analysis
100 so far of *Symbiodinium* genomes, and the first to include a prominent endosymbiont of corals
101 of Indo-Pacific and Caribbean reefs.

102 **Results**

103 ***Genomes of S. goreaui and S. kawagutii***

104 We sequenced and generated two draft *Symbiodinium* genome assemblies *de novo*, for *S.*
105 *goreaui* (Clade C, 1.03 Gbp) and for *S. kawagutii* (Clade F, 1.05 Gbp). Details of data
106 generation and assembly statistics are shown in Supplementary Tables S1 and S2
107 respectively. Our *S. goreaui* assembly consists of 41,289 scaffolds (N50 length 98,034 bp).
108 For *S. kawagutii*, we first verified that our data (from isolate CS-156) and the published data
109 (from the synonym isolate CCMP2468) are indeed from the same culture of origin (see
110 Supplementary Methods and Supplementary Figure S1). Compared to the published assembly
111 (Lin et al.¹⁸), independent mapping of their ten fosmid sequences¹⁸ onto our preliminary CS-
112 156 assembly yielded up to 43-fold and 37-fold fewer gaps and mismatches respectively
113 (Supplementary Figure S2). We later combined both datasets in a single *de novo* assembly,

114 yielding 16,959 scaffolds (N50 length 268,823 bp). Genome-size estimates based on *k*-mer
115 coverage are 1.19 Gbp for *S. goreau* and 1.07 Gbp for *S. kawagutii* (Supplementary Table
116 S3), comparable to those for other sequenced *Symbiodinium* genomes. We also recovered
117 sequences putatively derived from their plastid genomes (Supplementary Tables S4, S5 and
118 S6) including their distinct core conserved regions (Supplementary Table S7), and from their
119 mitochondrial genomes; see Supplementary Note for details.

120 The repeat content of the assembled genomes ranged from 16.0% (*S. kawagutii*) to
121 27.9% (*S. microadriaticum*); a large peak in transposable element (TE) abundance observed
122 at high divergence (Kimura distance²⁷ 15-25) in all genomes (Supplementary Figure S3)
123 suggests that most extant transposable elements are remnants of an ancient burst of TE
124 activity. TE activity has been broadly linked to genome size in plants²⁸⁻³⁰, so reduced TE
125 activity may be connected with the relative compactness of *Symbiodinium* genomes in
126 comparison to those of other dinoflagellates. However, as these genomes are still in draft, the
127 impact of assembly completeness on the patterns of repeat divergence cannot be dismissed.

128 Using a stringent threshold to remove genome scaffolds of potential bacterial or viral
129 origin (Methods), we predict 35,913 and 26,609 high-quality gene models respectively for *S.*
130 *goreau* and *S. kawagutii* (Supplementary Table S8). Usage profiles of codons and amino
131 acids are shown in Supplementary Figures S4 and S5 respectively, and non-canonical splice
132 sites in Supplementary Table S9 and Supplementary Figure S6. Although we report fewer
133 genes, the majority (67.0% and 64.4% respectively for *S. goreau* and *S. kawagutii*) have
134 transcriptome support, and we generally recovered more of the 458 conserved core eukaryote
135 genes (Supplementary Figure S7), 371 of which are common to all four *Symbiodinium* based
136 on the predicted gene models (Figure 1a; Supplementary Table S10); similar results are
137 observed for the corresponding genome sequences (Supplementary Figure S7). About 94% of

138 the predicted genes have introns, similar to *S. microadriaticum* (98.2%) and *S. minutum*
139 (95.3%); the earlier *S. kawagutii* genome assembly¹⁸ may have underestimated the proportion
140 of intron-containing genes (Supplementary Table S8) due to a less-stringent approach to gene
141 prediction. All coding sequences have higher G+C content (56.7% in *S. goreau* and 55.0% in
142 *S. kawagutii*) than does the genome overall, comparable to coding sequences from other
143 *Symbiodinium* (57.7% in *S. microadriaticum* and 52.7% in *S. minutum*).

144 ***Sequence divergence and synteny***

145 Despite the seemingly high number of protein-coding genes, an earlier analysis of syntenic
146 blocks¹⁷ found only several hundred blocks conserved in a pairwise manner among three
147 published *Symbiodinium* genomes. Here we included our two new genome sequences in this
148 analysis, and focused further on syntenic collinear blocks, requiring each block to contain the
149 same genes in the same order and orientation with no gene losses (Methods). The *S. goreau*
150 and *S. kawagutii* genomes share the most collinear blocks, 889 blocks implicating 8621
151 genes; 62 of these blocks are of size >15, with the largest containing 76 genes
152 (Supplementary Table S11). Thus substantial proportions of genes in these two genomes
153 occur in clusters: for cluster size ≥ 5 genes, 32.4% and 24.0% of *S. kawagutii* and *S. goreau*
154 genes respectively. These are likely to be underestimates, as the assemblies remain
155 fragmentary. At the other end of the spectrum, the genomes of *S. microadriaticum* and *S.*
156 *goreau* share only 86 collinear blocks of size ≥ 5 , with maximum size 12 and implicating
157 588 genes in total (Supplementary Table S11). These results suggest that (a) although Clades
158 C and F are divergent, they nonetheless are the most-closely related among the four analysed
159 *Symbiodinium* genomes (in line with their phylogenetic relationship); and (b) C and F are
160 more divergent from Clade A than from Clade B (in line with their phylogenetic
161 relationship). Thus whole-genome sequences support and extend earlier conclusions, based

162 on common marker sequences, about phylogenetic relationships among *Symbiodinium*
163 clades³¹⁻³³ and the remarkable divergence among *Symbiodinium* lineages^{17,18}.

164 ***Gene and protein functions***

165 All annotated genes from *S. goreau* and *S. kawagutii* genomes are listed in Supplementary
166 Tables S12 and S13 respectively. Of the 35,913 proteins predicted in *S. goreau*, 31,646
167 (88.1%) show similarity (BLASTP, $E \leq 10^{-5}$) to sequences in UniProt; among these, 29,198
168 (81.3% of 35,913) and 19,718 (54.9%) are annotated with Gene Ontology (GO) terms or
169 Pfam domains (Supplementary Tables S12 and S14). For *S. kawagutii*, 21,947 of 26,609
170 proteins (82.5%) find a match in UniProt (Supplementary Tables S13 and S14). *Protein*
171 *kinase* (Pfam PF00069), *reverse transcriptase* (PF07727), *ion transport protein* (PF00520)
172 and *ankyrin repeats* (PF12796) are among the most-abundant domains in *Symbiodinium*,
173 appearing among the ten most-abundant for each of the four genomes (Supplementary Table
174 S15). Ankyrin repeat motifs are important in the recognition of surface proteins, and more
175 generally in protein-protein interactions³⁴. Thus proteins potentially involved in host-
176 symbiont interaction (with phosphorylation, ion transport and protein recognition/interaction
177 domains) are well represented in the predicted *Symbiodinium* proteomes.

178 We compared functions of proteins predicted from these four *Symbiodinium* genomes
179 to a set of 27 more-narrowly scoped eukaryotes: 17 alveolates (ten other dinoflagellates, four
180 ciliates, two apicomplexans and *Perkinsus marinus*), stramenopiles (two diatoms) and
181 Archaeplastida (four rhodophytes, three chlorophytes and *Arabidopsis*). This 31-taxon set
182 (1,136,347 proteins; Supplementary Tables S16 and S17) represents lineages in which one or
183 more endosymbioses are implicated in plastid origin³⁵⁻³⁷; these proteins were clustered (based
184 on sequence similarity) into 56,530 groups of size two or greater (Supplementary Table S17;
185 see Methods). Using this 31-taxon dataset as background, we assessed the over- or under-

186 representation of protein domains within our various groups of *Symbiodinium* proteins. We
187 found 270 domains (Supplementary Table S18) to be significantly overrepresented (adjusted
188 $p \leq 0.05$) in *Symbiodinium*. Interestingly, many domains e.g. *C-5 cytosine-specific DNA*
189 *methylase* (PF00145), *planctomycete cytochrome c* (PF07635) and RNA polymerase *sigma-*
190 *70 region 2* (PF04542) are also overrepresented in the four *Symbiodinium* genomes in a
191 similar comparison against 880,909 proteins in a 15-taxon set that includes ten other
192 dinoflagellates and the immediate outgroup *Perkinsus marinus* (Supplementary Table S19).
193 Thus compared to related eukaryotes and to other dinoflagellates, *Symbiodinium* is enriched
194 in functions involved in methylation of cytosine, (photosynthetic) energy production and
195 RNA polymerisation. Hydroxymethylation of uracil is common (12-70%) in dinoflagellate
196 genomes³⁸; while cytosine methylation has been described in *Symbiodinium*^{39,40}, our findings
197 suggest that cytosine methylation is more prominent in *Symbiodinium* than in these other
198 dinoflagellates.

199 Activation of some retrotransposons is part of the stress-response mechanism in
200 diatoms⁴¹, plants^{42,43} and other eukaryotes^{44,45}. The *reverse transcriptase* domain (PF07727)
201 is enriched in *Symbiodinium* in both the 31-taxon and 15-taxon sets, suggesting that
202 retrotransposition could be a prominent mechanism of stress response in *Symbiodinium* and
203 dinoflagellates. Although we set a stringent threshold for removing putative bacterial or viral
204 sequences (see Methods), 40 (~0.1%) of the final 41,289 genome scaffolds of *S. goreau* have
205 significant hits (BLASTN $E \leq 10^{-20}$) to the virus genomes⁴⁶ isolated from the same *S. goreau*
206 (type C1) strain, with 16 identical regions (76-609 bp) distributed in nine scaffolds of lengths
207 ranging from 1092 to 7,338,656 bp. Whether this indicates introgression of viral sequences
208 remains to be determined.

209 ***Positive selection of Symbiodinium genes***

210 Using a branch-site model based on the ratio of substitution rates in non-synonymous (dN) to
211 synonymous (dS) sites⁴⁷ (Methods and Supplementary Figure S8), we identified
212 *Symbiodinium* genes under positive selection in comparison to ten other dinoflagellates, with
213 *Perkinsus marinus* as the outgroup (15 taxa: Supplementary Tables S16 and S17). The
214 reference species tree (Figure 2a) was computed following Price and Bhattacharya⁴⁸. We then
215 based our analysis of adaptive evolution on all orthologous sets for which the protein tree is
216 topologically congruent with our reference tree (Methods).

217 Of our 44,282 homologous sets, 2460 containing 7987 *Symbiodinium* proteins show
218 evidence of positive selection in one or more *Symbiodinium* lineages; 1069 of these sets are
219 annotated with GO terms (Supplementary Table S20). Figure 2b shows the terms (level 3) in
220 the three GO hierarchies that are shared by $\geq 5\%$ of these 1069 sets. In the Biological Process
221 hierarchy, metabolic processes are highly represented (*primary metabolic process* [292] and
222 *macromolecule compound metabolic process* [243] are among the four most-frequent terms),
223 followed by *oxidation reduction* [96] and transport (*establishment of localization* [90],
224 *transport* [90], and *transmembrane transport* [80]). Highly represented terms in the
225 Molecular Function hierarchy point to binding of diverse molecules and ions e.g. *protein*
226 *binding* [173] and metabolism (*hydrolase* [390], *transferase* [344]). In Cellular Component,
227 *cell part* [113], *membrane* [86] and *membrane part* [59] are the most frequent. Thus in
228 *Symbiodinium* as represented by these four assemblies, broad aspects of metabolism, and
229 transport including across membranes, show evidence of positive selection, in line with their
230 recognised importance in cnidarian-dinoflagellate symbioses¹⁷.

231 We further assessed enrichment of GO terms against all annotated terms specifically in
232 the four *Symbiodinium* genomes (Supplementary Table S21) compared with the other

233 dinoflagellates in this study. Here we consider themes among Biological Process terms. The
234 first theme is that functions associated with photosynthetic light reactions are enriched among
235 the positively selected *Symbiodinium* genes; *photosynthesis*, *light reaction* and *Photosystem*
236 *II assembly* are significantly over-represented (adjusted $p \leq 0.05$), as are Cellular Component
237 terms related to plastid functions e.g. *thylakoid*, *photosynthetic membrane*, *intracellular*
238 *membrane-bounded organelle* (Supplementary Table S21). Coral bleaching has been
239 associated with the loss of light-harvesting proteins and the subsequent inactivation of
240 photosystem II (PSII) in *Symbiodinium* under combined light and temperature stress⁴⁹⁻⁵¹.
241 These authors reported that coral bleaching associated with algal photobleaching can be
242 ameliorated, at least in part, by thermal acclimation of *Symbiodinium* to improve the thermal
243 tolerance of PSII.

244 The second emerging theme involves the transport of ions and metabolites across
245 membranes. *Intracellular transport*, *cytosolic transport*, *transition metal ion transport* and
246 *copper ion transport* as well as terms related to transmembrane transport of amino acids,
247 organic acids and carboxylic acids are significantly enriched (Supplementary Table S21);
248 these functions underpin multiple physiological processes, including but not limited to pH
249 regulation, calcification and photosynthetic carbon fixation⁵². Investigated *Symbiodinium* are
250 enriched in membrane transporters compared with other dinoflagellates¹⁷. These biological
251 processes are especially relevant to the maintenance and regulation of coral-dinoflagellate
252 symbiosis^{52,53}, possibly including its sensitivity and/or response to environmental stress.

253 The third theme is the enrichment of functions related to the biosynthesis and
254 modification of amino acids and glycoproteins (Supplementary Table S21) e.g. *protein*
255 *phosphorylation*, *peptide biosynthesis process*, *protein ADP-ribosylation*, *protein*
256 *glycosylation*, *D-amino acid metabolic process* and *glycoprotein biosynthetic process*. Corals

257 lack the capacity to synthesise a number of amino acids (e.g. cysteine in *Acropora*
258 *digitifera*⁵⁴), thus selection acting on the synthesis of amino acids may indicate the critical
259 role of *Symbiodinium* in supplying amino acids both for self-preservation and for the coral
260 host. Glycoprotein molecules are often surface-localised and in microbes form the basis of
261 microbe-associated molecular patterns (MAMPs) which, in conjunction with a host-
262 associated pattern recognition receptor, mediate recognition by a host⁵⁵. Both in culture and
263 *in hospite*, *Symbiodinium* exude glycoconjugates⁵⁶⁻⁵⁹. Where investigated, cell-surface glycan
264 profiles are stable over time within a *Symbiodinium* culture but can differ between clades
265 within a host⁶⁰. *N*-acetyl and mannosyl residues are prominent constituents of *Symbiodinium*
266 cell-surface glycans, making them candidates for MAMPs that could participate in the
267 establishment of symbiosis. Lin et al.¹⁸ reported a *S. kawagutii* glycan biosynthesis pathway
268 distinct from that of *S. minutum*, again pointing to a possible role of glycans in specificity of
269 host recognition^{60,61}. Neubauer et al.⁶² demonstrated that the thrombospondin type 1 repeat
270 (TSR) from the sea anemone *Aiptasia pallida* contains binding sites for glycosaminoglycan,
271 and that blocking TSR led to decreased colonisation by *S. minutum*. Our results offer the first
272 evidence of positive selection of functions underlying the biosynthesis and modification of
273 amino acids and glycoproteins, suggesting that these functions are critical in the
274 establishment of cnidarian-dinoflagellate symbioses.

275 Our fourth emerging theme is stress response. Enriched terms annotated for the
276 positively selected genes include *cell redox homeostasis*, *translation initiation* and 22 terms
277 describing the negative regulation of gene expression, transcription, RNA biosynthesis and
278 cellular biosynthetic and metabolic processes (Supplementary Table S21). Environmental
279 stressors can disrupt the cellular redox homeostasis and break down the coral-dinoflagellate
280 symbiosis. Negative regulation of transcription may represent a stress response that
281 safeguards the genome from accumulating DNA damage⁶³; a similar stress response has been

282 observed in coral⁶⁴. Other enriched functions that may be related to stress response include
283 *mitotic nuclear division, translation*, and various processes of nucleotide biosynthesis and
284 modification e.g. *RNA methylation, rRNA methylation, DNA replication, RNA processing*,
285 and *deoxyribonucleotide biosynthetic process*. Our results thus provide evidence that stress
286 response is under positive selection in *Symbiodinium*, presumably (given their lifestyle) in
287 relation to the establishment and/or maintenance of symbiosis.

288 ***Do Symbiodinium have sex?***

289 *Symbiodinium* have been hypothesised to reproduce sexually and to have a diploid life stage⁶⁵
290 but definitive evidence for sex, e.g. karyogamy and meiosis, has yet to be observed^{31,66-68}.
291 The ability to reproduce sexually offers increased efficiency of selection and adaptation⁶⁹. So
292 far, the strongest evidence supporting meiotic potential in *Symbiodinium* comes from patterns
293 of population-genetic variation revealed in allozymes, randomly amplified polymorphic DNA
294 and other molecular markers^{22,31,70-73}. Indeed, for some markers a higher degree of genetic
295 variation has been observed in certain *Symbiodinium* clades than in dinoflagellates known to
296 reproduce sexually⁷⁰. More recently, differential gene expression analysis⁷⁴ using a
297 heterologous culture from which our sequenced *S. goreau* was derived revealed an
298 enrichment of gene functions related to meiosis under thermal stress, suggesting a switch
299 from mitosis to meiosis under stress conditions.

300 Schurko and Logsdon⁷⁵ described a “meiosis detection toolkit”, a set of 51 genes
301 specific or related to meiosis^{76,77} that collectively point to a capacity for meiosis even in
302 divergent or specialised eukaryotic genomes⁷⁸. Incomplete genome coverage or assembly,
303 sequence divergence, paralogy, patterns of overlapping function and evolutionary
304 specialisation mean that not all 51 need to be present or detectable for a lineage to be
305 assessed as probably sexual, or only recently asexual^{75,77}. Thirty-one of these genes were

306 earlier identified in *Symbiodinium* Clades A and B⁷⁶. Here, BLASTP search ($E \leq 10^{-5}$) of
307 predicted proteins in these four *Symbiodinium* genomes recovered matches corresponding to
308 48 of the of 51 toolkit genes in *S. microadriaticum*, 47 in *S. minutum* and *S. goreau*, and 46
309 in *S. kawagutii* (Figure 3a and Supplementary Table S22). Eight of the 11 meiosis-specific
310 proteins were detected in all four *Symbiodinium*. REC114, SAD1 and XRS2 found weaker
311 matches ($E \geq 10^{-14}$) in one to two genomes, although confirmatory UniProt domains were
312 usually present (Supplementary Table S22). RAD17 is the *Schizosaccharomyces pombe*
313 homolog of *S. cerevisiae* RAD24⁷⁹, for which we find highly significant matches ($E \leq 10^{-127}$)
314 in all four *Symbiodinium*. Moreover, 15 of the 51 genes show evidence of positive selection
315 in *Symbiodinium* against other dinoflagellates (Supplementary Table S22). Our data imply
316 that these four *Symbiodinium* are, or until recently have been, capable of meiosis.

317 ***Response to light stress***

318 Mycosporine-like amino acids (MAAs) are UV-protective compounds that, in corals and
319 other marine organisms, also act as antioxidants scavenging reactive oxygen species (ROS).
320 Up to five MAAs have been reported in *Symbiodinium* (Clades A, B and C) isolated from
321 cnidarian hosts^{80,81}. The MAA biosynthetic pathway involves dehydroquinase synthase
322 (DHQS), *O*-methyltransferase (O-MT), an ATP-grasp and non-ribosomal peptide synthetase
323 (NRPS)^{82,83}. In cyanobacteria, a short-chain dehydrogenase may play a role in converting
324 shinorine to palythine-serine⁸⁴. Genes encoding these four MAA-biosynthetic enzymes were
325 reported absent from the *S. kawagutii* genome¹⁸. Here, using known proteins in bacteria,
326 fungi and cnidarians as queries, we recovered all five enzymes including the short-chain
327 dehydrogenase from the *S. microadriaticum*, *S. goreau* and *S. kawagutii* genomes
328 (Supplementary Table S23); ATP-grasp was not found in *S. minutum*. These enzymes were
329 earlier reported absent from *S. kawagutii*, and it was proposed that their absence can be
330 compensated via coral-*Symbiodinium* co-evolution¹⁸; this hypothesis remains to be

331 investigated, but we note that this *S. kawagutii* isolate has not been observed in association
332 with an animal host²⁶.

333 Scytonemin is a UV-blocker first reported in terrestrial cyanobacteria^{83,85}, and in
334 contrast to MAAs was thought to be synthesised exclusively by cyanobacteria⁸⁶. The genome
335 of *Nostoc punctiforme* contains an 18-gene operon that specifies proteins of scytonemin
336 biosynthesis and regulation, including the synthesis of aromatic amino-acid precursors^{87,88}. Its
337 expression is up-regulated by UV radiation⁸⁹. Homologs of six of these 18 genes have been
338 described in the coral *Acropora ditigifera*, and were considered putative instances of lateral
339 genetic transfer⁹⁰. We find 12 of these 18 genes in the genomes of *S. goreau* and in *S.*
340 *kawagutii*, 11 in *S. microadriaticum* and ten in *S. minutum* (Figure 3b, Supplementary Table
341 S24).

342 Genes responsible for biosynthesis of tryptophan (*trpA*, *trpB*, *trpC*, *trpD* and *trpE*) and
343 the two key enzymes of chorismate (and aromatic amino acid) biosynthesis, *aroG* and *aroB*
344 (dehydroquinate synthase, also important for MAA biosynthesis), are found in all
345 *Symbiodinium* genomes, albeit so far in different scaffolds; these genes are also present in
346 *Arabidopsis thaliana* although not in corals or *Hydra* which, like most other animals, are
347 unable to synthesise tryptophan. The recovery of more of these 18 genes in *Symbiodinium*
348 than in corals or other animals (Figure 3b) could reflect the impact of endosymbiotic
349 association of ancestral cyanobacteria during the course of plastid evolution in all
350 photosynthetic eukaryotes including dinoflagellates³⁵⁻³⁷. The presence of multiple gene copies
351 (Supplementary Table S24) also implicates genetic duplication. Our findings suggest that
352 *Symbiodinium* has the capacity to produce scytonemin, either by itself or jointly within the
353 holobiont.

354 **Discussion**

355 *Symbiodinium* can form associations with a wide range of cnidarian hosts (as well as some
356 other marine invertebrates and protists) across broad geographic and time scales⁹¹. The
357 symbiosis between corals and *Symbiodinium* relies on compatible host-symbiont recognition
358 and sustainable nutrient exchange, both of which are vulnerable to external environmental
359 factors including temperature and light. A sustainable coral-*Symbiodinium* association
360 requires an adaptive capacity in the face of environmental extremes.

361 In this study we generated the first draft genome of *S. goreau* (Clade C), a much-
362 improved draft genome of *S. kawagutii* (Clade F) and high-quality gene models for both.
363 Comparative analysis revealed remarkable divergence among the genomes of *Symbiodinium*
364 from four clades, consistent with previous single-gene phylogenetic relationships. We
365 identified 2460 *Symbiodinium* gene families that are under positive selection, many of which
366 encode functions directly relevant to the establishment and/or maintenance of symbiosis with
367 the coral host. We also identified complete, or near-complete, sets of genes indicative of the
368 presence of meiosis and several mechanisms of stress tolerance, functions that have until now
369 been considered absent from *S. kawagutii*. Our results demonstrate the remarkable genomic
370 capacity of *Symbiodinium* to synthesize key metabolites that are essential to the
371 establishment of symbiosis with coral hosts, and to respond to environmental stress.

372 *S. goreau* (type C1) belongs to one of the most globally dominant clades (Clade C) on
373 coral reefs, and analysis of its draft genome has provided important new insights into coral-
374 algal symbiosis. This genomic resource is already demonstrating utility in the identification
375 of symbiont fractions in *de novo* sequencing of coral tissues^{92,93}, and will provide a
376 foundation for targeted studies into the molecular biology, physiology and of this crucial
377 symbiosis and its adaptation to a changing environment.

378 **Methods**

379 ***Biological materials and sequencing data***

380 *Symbiodinium goreau* (Clade C, type C1; AIMS-aten-C1-MI-cfu-B2, now AIMS culture
381 collection SCF055-01) is a single-cell monoclonal culture first isolated from the coral
382 *Acropora tenuis* at Magnetic Island (Queensland, Australia) at 3 m depth⁹⁴. *Symbiodinium*
383 *kawagutii* CCMP2468 (Clade F) was maintained as a monoclonal culture. Genomic DNA
384 was extracted from these isolates using the Qiagen DNeasy Plant Mini Kit. We generated a
385 total of 116.0 Gb of sequence data (2 paired-end libraries of 230- and 500-bp inserts, plus 3
386 mate-pair libraries of 3-, 6- and 9-Kbp inserts) for *S. goreau*, and a total of 92.2 Gb of
387 sequence data (1 paired-end library of 230-bp inserts, plus 3 mate-pair libraries of 4-, 6- and
388 9-Kbp inserts) for *S. kawagutii*, in each case using the Illumina HiSeq 2500 Rapid Chemistry
389 platform. See Supplementary Methods and Supplementary Table S1 for details.

390 ***Genome assembly and annotation***

391 We adopted a combined approach to *de novo* genome assembly (Supplementary Methods) in
392 which multiple assembly programs (CLC Genomics Workbench (Qiagen), SPAdes⁹⁵ and
393 ALLPATHS-LG⁹⁶) were first used independently. The quality of each assembly was assessed
394 based on full-length recovery of phylogenetic markers and known coding sequences. Once
395 the best assembly (the master assembly) was identified, other assemblies were used to refine
396 it via merging scaffolds and filling gaps. We adopted a comprehensive *ab initio* approach for
397 gene prediction using all available dinoflagellate proteins, as well as all *Symbiodinium* genes
398 and transcriptomes, as guiding evidence. Our approach combines evidence-based methods i.e.
399 PASA (with TransDecoder)⁹⁷, AUGUSTUS⁹⁸ and MAKER⁹⁹, and unsupervised machine-
400 learning (GeneMark-ES¹⁰⁰ and SNAP¹⁰¹; alternative splice sites were specified in these
401 methods by modifying the relevant scripts. We then used EvidenceModeler¹⁰² to combine

402 multiple sets of predicted genes, allocating a heavier weight for evidence-based predictions
403 than for those produced by unsupervised approaches; for details see Supplementary Methods.
404 Final genome assemblies, predicted gene models and proteins are available at
405 <https://cloudstor.aarnet.edu.au/plus/index.php/s/6yziMf2ygWjGu0L>.

406 We adopted multiple approaches to assess genome completeness. Established methods
407 including CEGMA¹⁰³ and BUSCO¹⁰⁴ are based on conserved genes in a limited number of
408 eukaryote model organisms that are distantly related to dinoflagellates. The use of these
409 methods resulted in relatively low recovery of conserved eukaryote genes in *Symbiodinium*
410 (e.g. 33-42% of BUSCO genes; Supplementary Figure S7) when run at default setting. We
411 further assessed completeness using BLAST based on predicted proteins from the gene
412 models and the assembled genome scaffolds. For each genome, we followed Baumgarten et
413 al.¹⁰⁵ and searched (BLASTP, $E \leq 10^{-5}$) against the predicted proteins using the 458 CEGMA
414 proteins¹⁰³. We also searched against the CEGMA proteins using the genome scaffolds
415 (BLASTX $E \leq 10^{-5}$), against genome scaffolds using the 458 CEGMA proteins (TBLASTN,
416 $E \leq 10^{-5}$), and against genome scaffolds using the 458 CEGMA transcripts (TBLASTX, $E \leq$
417 10^{-5}) (Supplementary Table S10, Supplementary Figure S7).

418 ***Identification and removal of bacterial and viral sequences***

419 Bacterial and viral sequences were identified and removed following Aranda et al.¹⁷. Briefly,
420 we used our genome scaffolds (BLASTN, $E \leq 10^{-20}$) to query the complete and draft bacterial
421 genomes in NCBI, and the viral genomes in NCBI and PhAnToMe (<http://phantome.org>). We
422 applied more-stringent criteria than did Aranda et al.¹⁷ to identify putative contaminating
423 sequences, removing from the final assembly any scaffold sequence that contains regions that
424 match a bacterial or viral genome (BLASTN bit score > 1000 , $E \leq 10^{-20}$) where such regions

425 constitute >10% of the overall scaffold length (Supplementary Methods, Supplementary
426 Figure S9).

427 ***Analysis of genome synteny and collinearity***

428 Using all predicted genes and their associated genomic positions, we assessed the number of
429 syntenic collinear blocks (i.e. regions with the same genes coded in the same order, free from
430 rearrangement or loss) shared pairwise among genomes of *S. microadriaticum* (Clade A)¹⁷, *S.*
431 *minutum* (B)¹⁹, *S. goreauii* (C) and *S. kawagutii* (F). We used BLASTP ($E \leq 10^{-5}$) to search
432 for similar proteins within each genome, and among all of them. Next we used MCScanX¹⁰⁶
433 with parameter -s 5 to sort the BLASTP matches (alignments) based on genomic positions; to
434 minimise the number of collinear gene pairs arising from tandem repeats, we report only
435 collinear blocks that consist of five or more genes.

436 ***Functional annotation of gene models***

437 We adopted a similar approach to Aranda et al.¹⁷ to annotate gene models. Protein sequences
438 predicted using the standard genetic code were used as query (BLASTP, $E \leq 10^{-5}$) first
439 against Swiss-Prot, and those with no Swiss-Prot hits subsequently against TrEMBL (both
440 databases from UniProt release 2016_10). Gene Ontology (<http://geneontology.org/>) terms
441 associated with Swiss-Prot and TrEMBL hits were obtained using the UniProt-GOA mapping
442 (release 2016_10).

443 ***Identification of homologous protein groups and gene families***

444 Our workflow for delineation of sets of putatively homologous proteins, multiple sequence
445 alignment, generation of protein-family and reference trees, and analysis of selection is
446 shown in Supplementary Figure S8. Using predicted proteins from 31 phylogenetically diverse
447 organisms including *Symbiodinium* for which genome and/or transcriptome data are available
448 (Supplementary Table S25), we identified sets of putatively homologous proteins using

449 OrthoFinder¹⁰⁷ and consider the corresponding gene sets (families) to be homologs.
450 Following Harlow et al.¹⁰⁸ and Beiko et al.¹⁰⁹, we consider sequences within single-copy sets
451 (i.e. those in which each genome is represented no more than once) to be orthologs. We refer
452 to sets that contain proteins only from *Symbiodinium*, plus the *Symbiodinium* singletons, as
453 *Symbiodinium*-specific. For enrichment analysis of annotated features (GO terms or Pfam
454 domains), we compared the features within the *Symbiodinium*-specific set against those in
455 each background set (i.e. the 31-taxon set and, separately, the 15-taxon set) using a
456 hypergeometric test; features with an adjusted¹¹⁰ p -value < 0.05 were considered significant.

457 ***Analysis of positive selection in Symbiodinium genes***

458 For the 15-taxon set we sorted the 311,651 protein sets into 1,654 single-copy (ortholog) and
459 16,836 multi-copy sets. Multiple sequence alignments were carried out using MAFFT v7.245
460 at -linsi mode¹¹¹; questionably aligned columns and rows were removed from these
461 alignments using trimAl¹¹² with the -automated1 option.

462 Branch-site models (BSMs; see below) require a reference topology. We follow Price
463 and Bhattacharya⁴⁸ to generate the reference species tree. The trimmed single-copy protein
464 alignments were concatenated prior to maximum-likelihood (ML) inference of the species
465 phylogeny using IQTREE¹¹³; each alignment represents a partition for which the best
466 evolutionary model was determined independently. Support for each node was assessed using
467 2000 rapid bootstraps. The species tree so generated (Figure 2a) is similar to that of Price and
468 Bhattacharya⁴⁸, with very strong support ($\geq 96\%$ bootstrap replicates) for all internal nodes;
469 the *Symbiodinium* and Suessiales (*Symbiodinium* + *Polarella glacialis*) clades received 100%
470 bootstrap support.

471 Of all trimmed protein alignments, those with ≥ 60 aligned positions and ≥ 4 sequences
472 were used in subsequent analysis. For multi-copy protein sets, we imposed further filtering

473 criteria. We first inferred individual ML trees for the multi-copy sets using IQ-TREE, and
474 each resulting protein tree was compared with the reference species tree. Those congruent
475 with the reference species tree at genus level, and in which all *Symbiodinium* are resolved as
476 an exclusive monophyletic clade, were judged paralog-free and used in subsequent BSM
477 analysis (Supplementary Figure S8). Among the 16,836 multi-copy sets of the 15-taxon
478 analysis, 1788 (10.6%) resolve all *Symbiodinium* sequences into an exclusive monophyletic
479 clade and are topologically congruent at genus level with the reference species tree (i.e.
480 contain co-orthologs but not paralogs) and were retained, while the remaining 15,048 failed
481 one or both of these filtering criteria (i.e. contain presumed paralogs) and were not analysed
482 further (Supplementary Figure S8). The percentages of missing data and parsimoniously
483 informative sites in all filtered protein alignments for the 15-taxon set are detailed in
484 Supplementary Table S26. For each filtered alignment, we used the corresponding coding-
485 sequence alignment (codon alignment) generated using PAL2NAL¹¹⁴ in the BSM analysis.
486 Some predicted protein sequences in MMETSP¹¹⁵ do not match their corresponding CDS,
487 sometimes due to problematic translation and other times due to a frameshift. For these, we
488 used MACSE¹¹⁶ to derive the codon alignments.

489 We applied the branch-site model (BSM) implemented in the *codeml* program in
490 PAML 4.9¹¹⁷ to detect positive selection signal unique to the *Symbiodinium* lineage. BSMs
491 allow the dN/dS ratio ω to vary among both sites and branches, making it possible to infer
492 selection at both. We computed two models: a null model with fixed $\omega = 1$, and an alternative
493 model that estimates ω in our defined foreground branches (here, the node that leads to all
494 *Symbiodinium* lineages). We then compared the likelihoods of these two models to determine
495 the better fit. To reduce false positives we applied *q*-value estimation for false discovery rate
496 control, as implemented in R package *qvalue* to adjust *p*-values. Instances with an adjusted *p*

497 ≤ 0.05 were considered significant. See Supplementary Note and Supplementary Figure S10
498 for analysis of gene gain and gene loss in *Symbiodinium*.

499 **References**

- 500 1. Plaisance, L., Caley, M.J., Brainard, R.E. & Knowlton, N. The diversity of coral reefs:
501 what are we missing? *PLoS ONE* **6**, e25026 (2011).
- 502 2. Falkowski, P.G., Dubinsky, Z., Muscatine, L. & Porter, J.W. Light and the
503 bioenergetics of a symbiotic coral. *Bioscience* **34**, 705-709 (1984).
- 504 3. Muscatine, L. & Porter, J.W. Reef corals: mutualistic symbioses adapted to nutrient-
505 poor environments. *Bioscience* **27**, 454-460 (1977).
- 506 4. Glynn, P.W. Coral reef bleaching: facts, hypotheses and implications. *Global Change*
507 *Biol.* **2**, 495-509 (1996).
- 508 5. Hoegh-Guldberg, O. Climate change, coral bleaching and the future of the world's coral
509 reefs. *Mar. Freshw. Res.* **50**, 839-866 (1999).
- 510 6. Albright, R. et al. Ocean acidification: linking science to management solutions using
511 the Great Barrier Reef as a case study. *J. Environ. Manage.* **182**, 641-650 (2016).
- 512 7. Hughes, T.P. et al. Global warming and recurrent mass bleaching of corals. *Nature* **543**,
513 373-377 (2017).
- 514 8. Lin, S. Genomic understanding of dinoflagellates. *Res. Microbiol.* **162**, 551-569 (2011).
- 515 9. LaJeunesse, T.C., Lambert, G., Andersen, R.A., Coffroth, M.A. & Galbraith, D.W.
516 *Symbiodinium* (Pyrrophyta) genome sizes (DNA content) are smallest among
517 dinoflagellates. *J. Phycol.* **41**, 880-886 (2005).
- 518 10. Barbrook, A.C., Voolstra, C.R. & Howe, C.J. The chloroplast genome of a
519 *Symbiodinium* sp. Clade C3 isolate. *Protist* **165**, 1-13 (2014).
- 520 11. Howe, C.J., Nisbet, R.E.R. & Barbrook, A.C. The remarkable chloroplast genome of
521 dinoflagellates. *J. Exp. Bot.* **59**, 1035-1045 (2008).
- 522 12. Zhang, Z., Green, B.R. & Cavalier-Smith, T. Single gene circles in dinoflagellate
523 chloroplast genomes. *Nature* **400**, 155-159 (1999).
- 524 13. Nash, E.A., Nisbet, R.E.R., Barbrook, A.C. & Howe, C.J. Dinoflagellates: a
525 mitochondrial genome all at sea. *Trends Genet.* **24**, 328-335 (2008).
- 526 14. Waller, R.F. & Jackson, C.J. Dinoflagellate mitochondrial genomes: stretching the rules
527 of molecular biology. *BioEssays* **31**, 237-245 (2009).
- 528 15. Jackson, C.J. et al. Broad genomic and transcriptional analysis reveals a highly derived
529 genome in dinoflagellate mitochondria. *BMC Biol.* **5**, 41 (2007).

- 530 16. Liew, Y.J., Li, Y., Baumgarten, S., Voolstra, C.R. & Aranda, M. Condition-specific
531 RNA editing in the coral symbiont *Symbiodinium microadriaticum*. *PLoS Genet.* **13**,
532 e1006619 (2017).
- 533 17. Aranda, M. et al. Genomes of coral dinoflagellate symbionts highlight evolutionary
534 adaptations conducive to a symbiotic lifestyle. *Sci. Rep.* **6**, 39734 (2016).
- 535 18. Lin, S. et al. The *Symbiodinium kawagutii* genome illuminates dinoflagellate gene
536 expression and coral symbiosis. *Science* **350**, 691-694 (2015).
- 537 19. Shoguchi, E. et al. Draft assembly of the *Symbiodinium minutum* nuclear genome
538 reveals dinoflagellate gene structure. *Curr. Biol.* **23**, 1399-1408 (2013).
- 539 20. LaJeunesse, T.C. et al. Low symbiont diversity in southern Great Barrier Reef corals,
540 relative to those of the Caribbean. *Limnol. Oceanogr.* **48**, 2046-2054 (2003).
- 541 21. LaJeunesse, T.C. "Species" radiations of symbiotic dinoflagellates in the Atlantic and
542 Indo-Pacific since the Miocene-Pliocene transition. *Mol. Biol. Evol.* **22**, 570-581
543 (2005).
- 544 22. Thornhill, D.J., Lewis, A.M., Wham, D.C. & LaJeunesse, T.C. Host-specialist lineages
545 dominate the adaptive radiation of reef coral endosymbionts. *Evolution* **68**, 352-367
546 (2014).
- 547 23. Tonk, L., Bongaerts, P., Sampayo, E.M. & Hoegh-Guldberg, O. *SymbioGBR*: a web-
548 based database of *Symbiodinium* associated with cnidarian hosts on the Great Barrier
549 Reef. *BMC Ecol.* **13**, 7 (2013).
- 550 24. Bongaerts, P. et al. Prevalent endosymbiont zonation shapes the depth distributions of
551 scleractinian coral species. *R. Soc. Open Sci.* **2**, 140297 (2015).
- 552 25. Chan, Y.L. et al. Generalist dinoflagellate endosymbionts and host genotype diversity
553 detected from mesophotic (67-100 m depths) coral *Leptoseris*. *BMC Ecol.* **9**, 21 (2009).
- 554 26. Yuyama, I., Higuchi, T. & Mezaki, T. *Symbiodinium kawagutii* (clade F) coats the
555 surface of *Aropora solitaryensis*, resulting in the formation of a sheet-like crust. *Proc.*
556 *13th Intl Coral Reef Symp.*, 49-56 (2016).
- 557 27. Kimura, M. A simple method for estimating evolutionary rates of base substitutions
558 through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**, 111-120 (1980).
- 559 28. Vitte, C., Panaud, O. & Quesneville, H. LTR retrotransposons in rice (*Oryza sativa*, L.):
560 recent burst amplifications followed by rapid DNA loss. *BMC Genomics* **8**, 218 (2007).
- 561 29. Tenailon, M.I., Hufford, M.B., Gaut, B.S. & Ross-Ibarra, J. Genome size and
562 transposable element content as determined by high-throughput sequencing in maize
563 and *Zea luxurians*. *Genome Biol. Evol.* **3**, 219-229 (2011).
- 564 30. Lee, S.I. & Kim, N.S. Transposable elements and genome size variations in plants.
565 *Genomics Inform* **12**, 87-97 (2014).

- 566 31. LaJeunesse, T.C. Investigating the biodiversity, ecology, and phylogeny of
567 endosymbiotic dinoflagellates in the genus *Symbiodinium* using the ITS region: in
568 search of a “species” level marker. *J. Phycol.* **37**, 866-880 (2001).
- 569 32. Pochon, X. & Gates, R.D. A new *Symbiodinium* clade (Dinophyceae) from soritid
570 foraminifera in Hawai'i. *Mol. Phylogenet. Evol.* **56**, 492-497 (2010).
- 571 33. van Oppen, M.J.H., Mieog, J.C., Sánchez, C.A. & Fabricius, K.E. Diversity of algal
572 endosymbionts (zooxanthellae) in octocorals: the roles of geography and host
573 relationships. *Mol. Ecol.* **14**, 2403-2417 (2005).
- 574 34. Mosavi, L.K., Cammett, T.J., Desrosiers, D.C. & Peng, Z.Y. The ankyrin repeat as
575 molecular architecture for protein recognition. *Protein Sci.* **13**, 1435-1448 (2004).
- 576 35. Chan, C.X., Gross, J., Yoon, H.S. & Bhattacharya, D. Plastid origin and evolution: new
577 models provide insights into old problems. *Plant Physiol.* **155**, 1552-1560 (2011).
- 578 36. Howe, C.J., Barbrook, A.C., Nisbet, R.E.R., Lockhart, P.J. & Larkum, A.W.D. The
579 origin of plastids. *Phil. Trans. R. Soc. B* **363**, 2675-2685 (2008).
- 580 37. Yoon, H.S. et al. Tertiary endosymbiosis driven genome evolution in dinoflagellate
581 algae. *Mol. Biol. Evol.* **22**, 1299-1308 (2005).
- 582 38. Rae, P.M. Hydroxymethyluracil in eukaryote DNA: a natural feature of the Pyrrophyta
583 (dinoflagellates). *Science* **194**, 1062-1064 (1976).
- 584 39. Blank, R.J., Huss, V.R. & Kersten, W. Base composition of DNA from symbiotic
585 dinoflagellates: a tool for phylogenetic classification. *Arch. Microbiol.* **149**, 515-520
586 (1988).
- 587 40. ten Lohuis, M.R. & Miller, D.J. Hypermethylation at CpG-motifs in the dinoflagellates
588 *Amphidinium carterae* (Dinophyceae) and *Symbiodinium microadriaticum*
589 (Dinophyceae): evidence from restriction analyses, 5-azacytidine and ethionine
590 treatment. *J. Phycol.* **34**, 152-159 (1998).
- 591 41. Maumus, F. et al. Potential impact of stress activated retrotransposons on genome
592 evolution in a marine diatom. *BMC Genomics* **10**, 624 (2009).
- 593 42. Ito, H. et al. A stress-activated transposon in *Arabidopsis* induces transgenerational
594 abscisic acid insensitivity. *Sci. Rep.* **6**, 23181 (2016).
- 595 43. Ramallo, E., Kalendar, R., Schulman, A.H. & Martínez-Izquierdo, J.A. *Remel*, a *Copia*
596 retrotransposon in melon, is transcriptionally induced by UV light. *Plant Mol. Biol.* **66**,
597 137-150 (2008).
- 598 44. Faure, E., Best-Belpomme, M. & Champion, S. X-irradiation activates the *Drosophila*
599 1731 retrotransposon LTR and stimulates secretion of an extracellular factor that
600 induces the 1731-LTR transcription in nonirradiated cells. *J Biochem* **120**, 313-319
601 (1996).

- 602 45. Hunter, R.G., Gagnidze, K., McEwen, B.S. & Pfaff, D.W. Stress and the dynamic
603 genome: steroids, epigenetics, and the transposome. *Proc. Natl. Acad. Sci. U. S. A.* **112**,
604 6828-6833 (2015).
- 605 46. Weynberg, K.D. et al. Prevalent and persistent viral infection in cultures of the coral
606 algal endosymbiont *Symbiodinium*. *Coral Reefs* **36**, 773-784 (2017).
- 607 47. Kimura, M. Preponderance of synonymous changes as evidence for the neutral theory
608 of molecular evolution. *Nature* **267**, 275-276 (1977).
- 609 48. Price, D.C. & Bhattacharya, D. Robust Dinoflagellata phylogeny inferred from public
610 transcriptome databases. *J. Phycol.* **53**, 725-729 (2017).
- 611 49. Takahashi, S., Yoshioka-Nishimura, M., Nanba, D. & Badger, M.R. Thermal
612 acclimation of the symbiotic alga *Symbiodinium* spp. alleviates photobleaching under
613 heat stress. *Plant Physiol.* **161**, 477-485 (2013).
- 614 50. Lesser, M.P. Elevated temperatures and ultraviolet radiation cause oxidative stress and
615 inhibit photosynthesis in symbiotic dinoflagellates. *Limnol. Oceanogr.* **41**, 271-283
616 (1996).
- 617 51. Lesser, M.P. Oxidative stress causes coral bleaching during exposure to elevated
618 temperatures. *Coral Reefs* **16**, 187-192 (1997).
- 619 52. Barott, K.L., Venn, A.A., Perez, S.O., Tambutte, S. & Tresguerres, M. Coral host cells
620 acidify symbiotic algal microenvironment to promote photosynthesis. *Proc. Natl. Acad.*
621 *Sci. U. S. A.* **112**, 607-612 (2015).
- 622 53. Zoccola, D. et al. Bicarbonate transporters in corals point towards a key step in the
623 evolution of cnidarian calcification. *Sci. Rep.* **5**, 9983 (2015).
- 624 54. Shinzato, C. et al. Using the *Acropora digitifera* genome to understand coral responses
625 to environmental change. *Nature* **476**, 320-323 (2011).
- 626 55. Davy, S.K., Allemand, D. & Weis, V.M. Cell biology of cnidarian-dinoflagellate
627 symbiosis. *Microbiol. Mol. Biol. Rev.* **76**, 229-261 (2012).
- 628 56. Lin, K.L., Wang, J.T. & Fang, L.S. Participation of glycoproteins of zooxanthellal cell
629 walls in the establishment of a symbiotic relationship with the sea anemone, *Aiptasia*
630 *pulchella*. *Zool. Stud.* **39**, 172-178 (2000).
- 631 57. Markell, D.A., Trench, R.K. & Iglesias-Prieto, R. Macromolecules associated with the
632 cell walls of symbiotic dinoflagellates. *Symbiosis* **12**, 19-31 (1992).
- 633 58. Markell, D.A. & Trench, R.K. Macromolecules exuded by symbiotic dinoflagellates in
634 culture: amino acid and sugar composition. *J. Phycol.* **29**, 64-68 (1993).
- 635 59. Markell, D.A. & Wood-Charlson, E.M. Immunocytochemical evidence that symbiotic
636 algae secrete potential recognition signal molecules *in hospite*. *Mar. Biol.* **157**, 1105-
637 1111 (2010).

- 638 60. Logan, D.D.K., LaFlamme, A.C., Weis, V.M. & Davy, S.K. Flow-cytometric
639 characterization of the cell-surface glycans of symbiotic dinoflagellates (*Symbiodinium*
640 spp.). *J. Phycol.* **46**, 525-533 (2010).
- 641 61. Wood-Charlson, E.M., Hollingsworth, L.L., Krupp, D.A. & Weis, V.M. Lectin/glycan
642 interactions play a role in recognition in a coral/dinoflagellate symbiosis. *Cell.*
643 *Microbiol.* **8**, 1985-1993 (2006).
- 644 62. Neubauer, E.F. et al. A diverse host thrombospondin-type-1 repeat protein repertoire
645 promotes symbiont colonization during establishment of cnidarian-dinoflagellate
646 symbiosis. *eLife* **6**, e24494 (2017).
- 647 63. Ljungman, M. The transcription stress response. *Cell Cycle* **6**, 2252-2257 (2007).
- 648 64. Mohamed, A.R. et al. The transcriptomic response of the coral *Acropora digitifera* to a
649 competent *Symbiodinium* strain: the symbiosome as an arrested early phagosome. *Mol.*
650 *Ecol.* **25**, 3127-3141 (2016).
- 651 65. Fitt, W.K. & Trench, R.K. Endocytosis of the symbiotic dinoflagellate *Symbiodinium*
652 *microadriaticum* Freudenthal by endodermal cells of the scyphistomae of *Cassiopeia*
653 *xamachana* and resistance of the algae to host digestion. *J. Cell Sci.* **64**, 195-212
654 (1983).
- 655 66. Freudenthal, H.D. *Symbiodinium* gen. nov. and *Symbiodinium microadriaticum* sp.
656 nov., a zooxanthella: taxonomy, life cycle, and morphology. *J. Protozool.* **9**, 45-52
657 (1962).
- 658 67. Taylor, F.J.R. Implications and extensions of the serial endosymbiosis theory of the
659 origin of eukaryotes. *Taxon* **23**, 229-258 (1974).
- 660 68. Trench, R.K. Diversity of symbiotic dinoflagellates and the evolution of microalgal-
661 invertebrate symbioses. *Proc. 8th Intl Coral Reef Symp.* **2**, 1275-1286 (1997).
- 662 69. de Visser, J.A. & Elena, S.F. The evolution of sex: empirical insights into the roles of
663 epistasis and drift. *Nat. Rev. Genet.* **8**, 139-149 (2007).
- 664 70. Baillie, B.K., Belda-Baillie, C.A. & Maruyama, T. Conspecificity and Indo-Pacific
665 distribution of *Symbiodinium* genotypes (Dinophyceae) from giant clams. *J. Phycol.* **36**,
666 1153-1161 (2000).
- 667 71. Pettay, D.T., Wham, D.C., Pinzón, J.H. & LaJeunesse, T.C. Genotypic diversity and
668 spatial-temporal distribution of *Symbiodinium* clones in an abundant reef coral. *Mol.*
669 *Ecol.* **20**, 5197-5212 (2011).
- 670 72. Reichman, J.R., Wilcox, T.P. & Vize, P.D. PCP gene family in *Symbiodinium* from
671 *Hippopus hippopus*: low levels of concerted evolution, isoform diversity, and spectral
672 tuning of chromophores. *Mol. Biol. Evol.* **20**, 2143-2154 (2003).
- 673 73. Santos, S.R. & Coffroth, M.A. Molecular genetic evidence that dinoflagellates
674 belonging to the genus *Symbiodinium* Freudenthal are haploid. *Biol. Bull.* **204**, 10-20
675 (2003).

- 676 74. Levin, R.A. et al. Sex, scavengers, and chaperones: transcriptome secrets of divergent
677 *Symbiodinium* thermal tolerances. *Mol. Biol. Evol.* **33**, 2201-2215 (2016).
- 678 75. Schurko, A.M. & Logsdon, J.M., Jr. Using a meiosis detection toolkit to investigate
679 ancient asexual "scandals" and the evolution of sex. *BioEssays* **30**, 579-589 (2008).
- 680 76. Chi, J., Parrow, M.W. & Dunthorn, M. Cryptic sex in *Symbiodinium* (Alveolata,
681 Dinoflagellata) is supported by an inventory of meiotic genes. *J. Eukaryot. Microbiol.*
682 **61**, 322-327 (2014).
- 683 77. Malik, S.B., Pightling, A.W., Stefaniak, L.M., Schurko, A.M. & Logsdon, J.M., Jr. An
684 expanded inventory of conserved meiotic genes provides evidence for sex in
685 *Trichomonas vaginalis*. *PLoS ONE* **3**, e2879 (2008).
- 686 78. Speijer, D., Lukeš, J. & Eliáš, M. Sex is a ubiquitous, ancient, and inherent attribute of
687 eukaryotic life. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 8827-8834 (2015).
- 688 79. Griffiths, D.J., Barbet, N.C., McCready, S., Lehmann, A.R. & Carr, A.M. Fission yeast
689 *rad17*: a homologue of budding yeast *RAD24* that shares regions of sequence similarity
690 with DNA polymerase accessory proteins. *EMBO J.* **14**, 5812-5823 (1995).
- 691 80. Banaszak, A.T., LaJeunesse, T.C. & Trench, R.K. The synthesis of mycosporine-like
692 amino acids (MAAs) by cultured, symbiotic dinoflagellates. *J. Exp. Mar. Biol. Ecol.*
693 **249**, 219-233 (2000).
- 694 81. Rosic, N.N. & Dove, S. Mycosporine-like amino acids from coral dinoflagellates. *Appl.*
695 *Environ. Microbiol.* **77**, 8478-8486 (2011).
- 696 82. Balskus, E.P. & Walsh, C.T. The genetic and molecular basis for sunscreen
697 biosynthesis in cyanobacteria. *Science* **329**, 1653-1656 (2010).
- 698 83. Gao, Q. & Garcia-Pichel, F. Microbial ultraviolet sunscreens. *Nat. Rev. Microbiol.* **9**,
699 791-802 (2011).
- 700 84. D'Agostino, P.M. et al. Comparative profiling and discovery of novel glycosylated
701 mycosporine-like amino acids in two strains of the cyanobacterium *Scytonema* cf.
702 *crispum*. *Appl. Environ. Microbiol.* **82**, 5951-5959 (2016).
- 703 85. Rastogi, R.P., Sonani, R.R. & Madamwar, D. Cyanobacterial sunscreen scytonemin:
704 role in photoprotection and biomedical research. *Appl. Biochem. Biotechnol.* **176**, 1551-
705 1563 (2015).
- 706 86. Garcia-Pichel, F. & Castenholz, R.W. Characterization and biological implications of
707 scytonemin, a cyanobacterial sheath pigment. *J. Phycol.* **27**, 395-409 (1991).
- 708 87. Balskus, E.P. & Walsh, C.T. Investigating the initial steps in the biosynthesis of
709 cyanobacterial sunscreen scytonemin. *J. Am. Chem. Soc.* **130**, 15260-15261 (2008).
- 710 88. Soule, T., Stout, V., Swingley, W.D., Meeks, J.C. & Garcia-Pichel, F. Molecular
711 genetics and genomic analysis of scytonemin biosynthesis in *Nostoc punctiforme*
712 ATCC 29133. *J. Bacteriol.* **189**, 4465-4472 (2007).

- 713 89. Soule, T., Garcia-Pichel, F. & Stout, V. Gene expression patterns associated with the
714 biosynthesis of the sunscreen scytonemin in *Nostoc punctiforme* ATCC 29133 in
715 response to UVA radiation. *J. Bacteriol.* **191**, 4639-4646 (2009).
- 716 90. Shinzato, C., Mungpakdee, S., Satoh, N. & Shoguchi, E. A genomic approach to coral-
717 dinoflagellate symbiosis: studies of *Acropora digitifera* and *Symbiodinium minutum*.
718 *Front. Microbiol.* **5**, 336 (2014).
- 719 91. Frankowiak, K. et al. Photosymbiosis and the expansion of shallow-water corals. *Sci.*
720 *Adv.* **2**, e1601122 (2016).
- 721 92. Bongaerts, P. et al. Deep reefs are not universal refuges: reseeded potential varies
722 among coral species. *Sci. Adv.* **3**, e1602373 (2017).
- 723 93. Thomas, L., Kennington, W.J., Evans, R.D., Kendrick, G.A. & Stat, M. Restricted gene
724 flow and local adaptation highlight the vulnerability of high-latitude reefs to rapid
725 environmental change. *Global Change Biol.* **23**, 2197-2205 (2017).
- 726 94. Howells, E.J. et al. Coral thermal tolerance shaped by local adaptation of
727 photosymbionts. *Nat. Clim. Change* **2**, 116-120 (2012).
- 728 95. Bankevich, A. et al. SPAdes: a new genome assembly algorithm and its applications to
729 single-cell sequencing. *J. Comput. Biol.* **19**, 455-477 (2012).
- 730 96. Gnerre, S. et al. High-quality draft assemblies of mammalian genomes from massively
731 parallel sequence data. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 1513-1518 (2011).
- 732 97. Haas, B.J. et al. Improving the *Arabidopsis* genome annotation using maximal
733 transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654-5666 (2003).
- 734 98. Stanke, M. et al. AUGUSTUS: *ab initio* prediction of alternative transcripts. *Nucleic*
735 *Acids Res.* **34**, W435-439 (2006).
- 736 99. Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database
737 management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491
738 (2011).
- 739 100. Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y.O. & Borodovsky, M. Gene
740 identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids*
741 *Res.* **33**, 6494-6506 (2005).
- 742 101. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
- 743 102. Haas, B.J. et al. Automated eukaryotic gene structure annotation using
744 EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**,
745 R7 (2008).
- 746 103. Parra, G., Bradnam, K. & Korf, I. CEGMA: a pipeline to accurately annotate core
747 genes in eukaryotic genomes. *Bioinformatics* **23**, 1061-1067 (2007).

- 748 104. Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. & Zdobnov, E.M.
749 BUSCO: assessing genome assembly and annotation completeness with single-copy
750 orthologs. *Bioinformatics* **31**, 3210-3212 (2015).
- 751 105. Baumgarten, S. et al. The genome of *Aiptasia*, a sea anemone model for coral
752 symbiosis. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 11893-11898 (2015).
- 753 106. Wang, Y. et al. MCScanX: a toolkit for detection and evolutionary analysis of gene
754 synteny and collinearity. *Nucleic Acids Res.* **40**, e49 (2012).
- 755 107. Emms, D.M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome
756 comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**,
757 157 (2015).
- 758 108. Harlow, T.J., Gogarten, J.P. & Ragan, M.A. A hybrid clustering approach to
759 recognition of protein families in 114 microbial genomes. *BMC Bioinformatics* **5**, 45
760 (2004).
- 761 109. Beiko, R.G., Harlow, T.J. & Ragan, M.A. Highways of gene sharing in prokaryotes.
762 *Proc. Natl. Acad. Sci. U. S. A.* **102**, 14332-14337 (2005).
- 763 110. Benjamini, Y. & Yekutieli, D. The control of the false discovery rate in multiple testing
764 under dependency. *The Annals of Statistics* **29**, 1165-1188 (2001).
- 765 111. Katoh, K., Kuma, K.-I., Toh, H. & Miyata, T. MAFFT version 5: improvement in
766 accuracy of multiple sequence alignment. *Nucleic Acids Res.* **33**, 511-518 (2005).
- 767 112. Capella-Gutiérrez, S., Silla-Martínez, J.M. & Gabaldón, T. trimAl: a tool for automated
768 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972-1973
769 (2009).
- 770 113. Nguyen, L.T., Schmidt, H.A., von Haeseler, A. & Minh, B.Q. IQ-TREE: a fast and
771 effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol.*
772 *Biol. Evol.* **32**, 268-274 (2015).
- 773 114. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: robust conversion of protein sequence
774 alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, W609-612
775 (2006).
- 776 115. Keeling, P.J. et al. The Marine Microbial Eukaryote Transcriptome Sequencing Project
777 (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans
778 through transcriptome sequencing. *PLoS Biol.* **12**, e1001889 (2014).
- 779 116. Ranwez, V., Harispe, S., Delsuc, F. & Douzery, E.J. MACSE: Multiple Alignment of
780 Coding SEquences accounting for frameshifts and stop codons. *PLoS ONE* **6**, e22594
781 (2011).
- 782 117. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**,
783 1586-1591 (2007).

784

785 **Acknowledgements**

786 We thank Manuel Aranda for advance access to the *S. microadriaticum* genome data, and
787 Todd LaJeunesse for information on the original isolation of *S. kawagutii*. This project was
788 supported by the Reef Future Genomics (ReFuGe) 2020 International Consortium and the
789 Great Barrier Reef Foundation. The data reported in this work were supported by funding
790 from Bioplatforms Australia through the Australian Government National Collaborative
791 Research Infrastructure Strategy (NCRIS). In memory of SF, our friend and colleague who is
792 sorely missed.

793 **Author contributions**

794 HL, MAR and CXC conceived the study and designed the experiments. HL, TGS, RGP and
795 CXC conducted all computational analyses. VHB and BL established the algal cultures and
796 extracted the DNA. HL, TGS, RGP, IC, MAR and CXC analysed and interpreted the results.
797 HL and CXC prepared all figures, tables, and the first draft of this manuscript. SF and CRV
798 provided analytical tools and scripts. HL, TGS, MAR and CXC wrote the manuscript. PB, IC,
799 DGB, DJM, MJHvO and CRV assisted in experimental design and writing of the manuscript.
800 All authors reviewed, commented on and approved the final manuscript.

801 **Data availability**

802 All sequence data generated from this study will be available at the NCBI Short Read
803 Archive (accessions XXXX and XXXX). Assembled genomes, predicted gene models and
804 proteins are available at <https://cloudstor.aarnet.edu.au/plus/index.php/s/6yziMf2ygWjGu0L>.

805 **Code availability**

806 The customised scripts for AUGUSTUS and PASA used in this study were previously
807 published in Aranda et al.¹⁷; they are available at <http://smic.reefgenomics.org/download/>.

808 **Competing financial interests**

809 The authors declare no competing financial interests.

810 **Figure legends**

811 **Figure 1. Comparison of *Symbiodinium* genomes.** (a) Number of recovered core eukaryote
812 genes in each genome based on CEGMA. (b) Number of identified syntenic collinear blocks
813 (and the corresponding number of implicated genes) between each genome pair.

814 **Figure 2. Testing for positive selection acting on *Symbiodinium* genomes.** (a) The
815 reference 15-species supertree of *Symbiodinium*, dinoflagellates and *Perkinsus marinus* (as
816 outgroup) based on single-copy orthologous genes. Support based on 2000 rapid bootstraps is
817 shown on each internal node, and the branch length is the number of substitutions per site. (b)
818 Percentage of positively selected genes with annotated GO (level 3) terms in *Symbiodinium*,
819 shown for principal hierarchies Biological Function, Molecular Function and Cellular
820 Component.

821 **Figure 3. Recovery of genes in *Symbiodinium*.** (a) Meiosis-related genes recovered in the
822 genomes of *S. microadriaticum* (Clade A), *S. minutum* (Clade B), *S. goreaui* (Clade C) and *S.*
823 *kawagutii* (Clade F). (b) Scytonemin biosynthesis genes in *Symbiodinium* genomes relative to
824 the coral *Acropora digitifera*, sea anemone *Nematostella vectensis*, hydra (*Hydra*
825 *magnipapillata*) and the green plant *Arabidopsis thaliana*. The order of the 18-gene cluster in
826 the cyanobacteria *Nostoc punctiforme* is used as a reference, with presence (+) and absence (-
827) of a gene in each species are shown. Figure modified from Shinzato et al.⁹⁰.

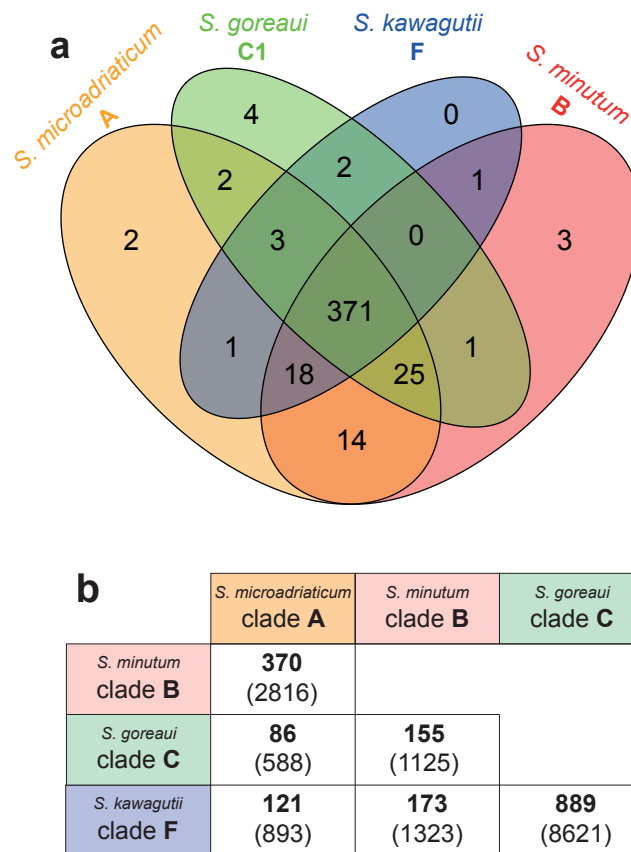


Figure 1. Comparison of *Symbiodinium* genomes. (a) Number of recovered core eukaryote genes in each genome based on CEGMA. **(b)** Number of identified syntenic collinear blocks (and the corresponding number of implicated genes) between each genome pair.

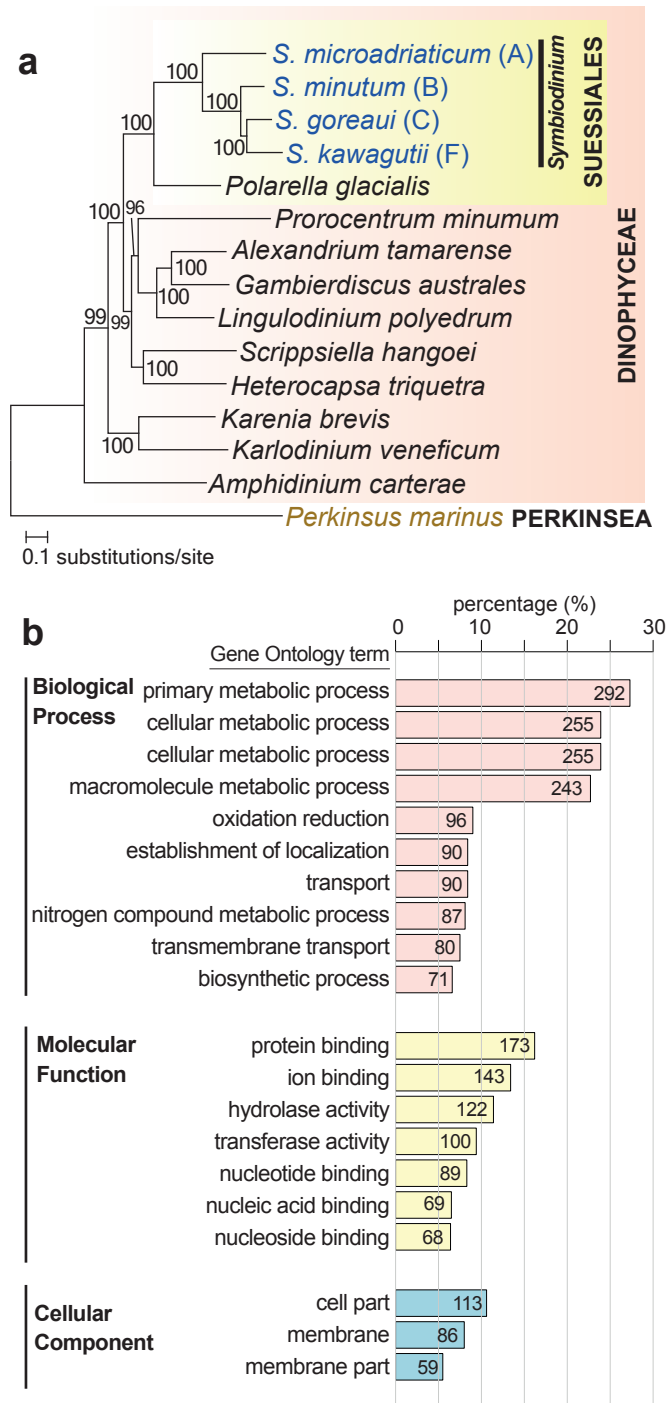


Figure 2. Testing for positive selection acting on *Symbiodinium* genomes. (a) The reference 15-species supertree of *Symbiodinium*, dinoflagellates and *Perkinsus marinus* (as outgroup) based on single-copy orthologous genes. Support based on 2000 rapid bootstraps is shown on each internal node, and the branch length is the number of substitutions per site. (b) Percentage of positively selected genes with annotated GO (level 3) terms in *Symbiodinium*, shown for principal hierarchies Biological Process, Molecular Function and Cellular Component.

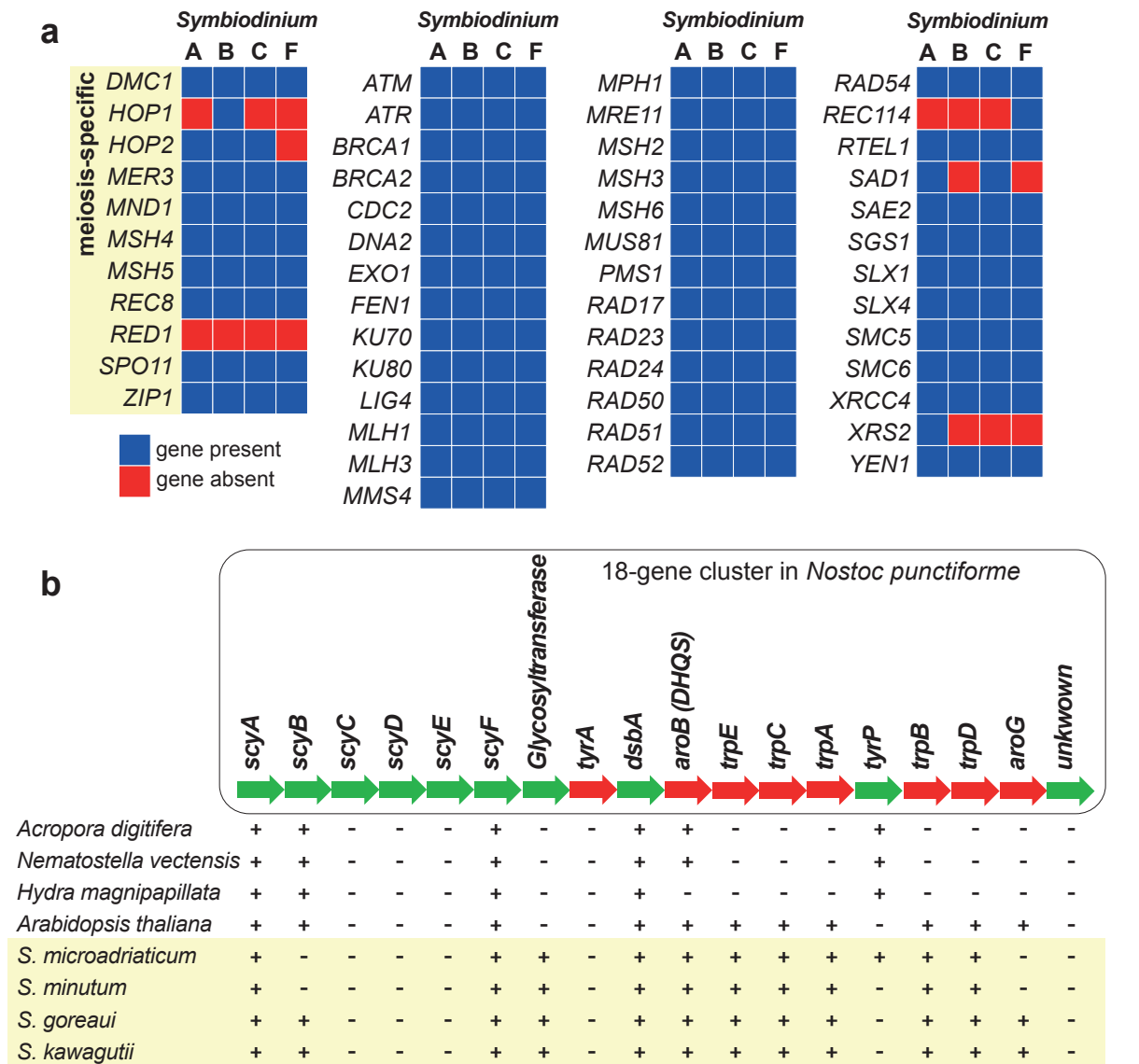


Figure 3. Recovery of genes in *Symbiodinium*. (a) Meiosis-related genes recovered in the genomes of *S. microadriaticum* (Clade A), *S. minutum* (Clade B), *S. goreau* (Clade C) and *S. kawagutii* (Clade F). (b) Scytonemin biosynthesis genes in *Symbiodinium* genomes relative to the coral *Acropora digitifera*, sea anemone *Nematostella vectensis*, hydra (*Hydra magnipapillata*) and the green plant *Arabidopsis thaliana*. The order of the 18-gene cluster in the cyanobacteria *Nostoc punctiforme* is used as a reference, with presence (+) and absence (-) of a gene in each species are shown. Figure modified from Shinzato et al.⁹⁰.