Article - Discoveries

Ploidy variation and its implications for reproduction and population dynamics in two sympatric Hawaiian coral species

Timothy G. Stephens[1,*], Emma L. Strand[2], Amin R. Mohamed[3], Amanda Williams[1,4], Eric N. Chiles[5,6], Xiaoyang Su[5,6], Debashish Bhattacharya[1,*], and Hollie M. Putnam[2,*]

[1]Department of Biochemistry and Microbiology, Rutgers University; New Brunswick, NJ 08901, USA.

[2]Department of Biological Sciences, University of Rhode Island; Kingston, RI 02881, USA.

[3]Institute for Molecular Bioscience, The University of Queensland; Brisbane, QLD 4072, Australia.

[4]Microbial Biology Graduate Program, Rutgers University; New Brunswick, NJ 08901, USA.

[5]Metabolomics Shared Resource, Rutgers Cancer Institute of New Jersey; Rutgers University, New Brunswick, NJ 08901, USA.

[6]Department of Medicine, Division of Endocrinology, Robert Wood Johnson Medical School, Rutgers University; New Brunswick, NJ 08901, USA.

*Corresponding authors: Debashish Bhattacharya, Hollie M. Putnam, and Timothy G. Stephens

Email: dbhattac@rutgers.edu, hputnam@uri.edu, ts942@sebs.rutgers.edu

Author Contributions: HMP and DB conceptualized the project; HMP, ES, ARM, AW, XS, ECN, and TGS designed the research; ES, ARM, and TGS performed the research; and TGS, ELS, ARM, AW, ENC, XS, DB, and HMP wrote/edited the paper.

Competing Interest Statement: Authors declare that they have no competing interests.

Classification: Biological Sciences, Environmental Sciences

Keywords: Coral, ploidy, genome evolution

Abstract

The response of coral reef ecosystems to anthropogenic climate change is driven by a complex interaction between location, stress history, species composition, and genetic background of the reef system. The latter two factors are particularly relevant when considering the different reproductive strategies used by coral species. We studied the stress resistant coral *Montipora capitata* and the more stress sensitive *Pocillopora acuta* from Kāneʻohe Bay, Oʻahu, Hawaiʻi. High-quality genome assemblies were generated for both species with the *M. capitata* assembly at chromosome-level resolution and the *P. acuta* assembly derived from a triploid colony, representing the first non-diploid genome generated from a coral. We report significant differences in the reproductive strategies of these coral species that not only affect the genetic structure of their populations in Kāneʻohe Bay, but also impact algal symbiont composition. Single-nucleotide polymorphism analysis shows that *P. acuta* comprises at least nine distinct genotypes in the bay with ancestral diploid and derived triploid lineages. In contrast, *M. capitata* colonies are diploids with nearly all being genotypically distinct. Genotype has a strong effect on gene expression profiles in these species, largely outweighing the effects of environmental stress. Our insights advance understanding of how reproductive strategy and ploidy can vary between different coral species and among local populations, how these factors constrain coral holobiont genetic diversity, and how genotype constrains genome-wide gene expression.

Introduction

As sea surface temperature continues to rise against a background of declining pH, it is critical to understand how heat waves may extend thermal anomalies given daily and seasonal fluctuations (Cant et al. 2021; Fischer et al. 2021). For coral reef ecosystems, which are built on the nutritional symbiosis between scleractinian coral hosts and their single celled dinoflagellate (algal) endosymbionts in the family Symbiodiniaceae (LaJeunesse et al. 2018), thermal stress may lead to dysbiosis and mortality. This phenomenon is known as coral "bleaching", whereby symbiotic cells and pigments are expelled or lost from the host tissue, leaving the bright white color of the underlying coral animal body and skeleton (van Oppen and Lough 2018). Bleaching is the primary cause of mass coral mortality (Hughes et al. 2017). Coral reefs are also threatened by the increased amount of CO_2 in the atmosphere that dissolves in the surface ocean, changing the carbonate chemistry and lowering the pH, which has been termed ocean acidification (Hoegh-Guldberg et al. 2007) (OA). Understanding coral response to long-term environmental stress is challenging, given the genetically diverse collection of organisms (cnidarian animal host, algal symbionts, prokaryotic microbiome, fungi and other eukaryotes, and viruses) that make up the holobiont and contribute to its health and resilience (Veron 2000). Furthermore, corals are impacted by continual abiotic changes (e.g., diurnal, and seasonal light and temperature variation) and a plethora of interacting taxa (e.g., algae, fish, viruses) that are of non-holobiont provenance.

In this study, high-quality genome assemblies were generated from the two analyzed coral species; for *Montipora capitata* a chromosome-level assembly was generated, with the 14 largest chromosomes in this assembly likely representing the 14 chromosomes predicted in other *Montipora* species (Kenyon 1997). The *Pocillopora acuta* genome assembly is from a triploid individual and represents the first non-diploid genome assembly generated from corals. We generated RNA-seq data from fragments of *M. capitata* (stress resistant) and *P. acuta* (stress sensitive) coral colonies (figs. 1A and B, respectively). The sequenced fragments were derived

from six reefs in Kāneʻohe Bay (fig. 1C), a 45 km² sheltered water body in Oʻahu, Hawaiʻi. Single-nucleotide polymorphisms (SNPs) were identified in each sample and showed that genetic diversity, ploidy, and reproductive strategy differ between the two species in the bay. The RNA-seq data was also used to assess the affect that population structure has on the response of the coral transcriptome to environmental stressors.

Results

Two new genome assemblies of Hawaiian corals

Genome assemblies of *M. capitata* and *P. acuta* were generated to assist with analysis of the RNA-Seq data and to support further research on these key Hawaiian coral species. For *M. capitata*, an initial assembly was generated using published PacBio long-read and Illumina short-read sequencing data (Shumaker et al. 2019); 56.5 million read-pairs of Dovetail Genomics Omni-C sequencing data was generated and used with the HiRise assembly workflow to scaffold the initial assembly. The resulting scaffolded assembly (hereinafter the “new genome assembly”) has fewer assembled bases (781 Mbp vs. 886 Mbp, respectively) and scaffolds (1,701 vs. 3,043), and has a vastly improved N50 (47.7 Mbp vs. 0.54 Mbp; supplementary table S1) compared to the Hawaiian *M. capitata* assembly that was previously published by our group (Shumaker et al. 2019). The 14 largest scaffolds in this assembly, ranging in size from ~22 to ~69 Mbp, likely represent the 14 chromosomes predicted in other *Montipora* species (fig. 2A and B) (Kenyon 1997). These putative chromosomes total 679.7 Mbp of assembled sequence, which is only slightly larger than the estimated genome size of 643.3 Mbp (fig. 2C; estimated by GenomeScope2 (Ranallo-Benavidez et al. 2020) using *k*-mers of size 21 bp). Compared to the published *M. capitata* genome, the new assembly has slightly improved BUSCO completeness for both the Metazoa (from 95.2% to 95.7%, respectively) and Eukaryota (from 97.7% to 99.2%, respectively) datasets (supplementary table S1), and a reduced number of duplicated BUSCO genes for both the Metazoa (from 21.2% to 1.6%, respectively) and Eukaryota (from 22.0% to 1.2%, respectively) datasets. The high duplication rate in the published version could result from haplotigs (i.e., sequences derived from different

homologous chromosomes) that evaded removal during the assembly process; this problem appears to have been resolved in the new assembly. Because the same PacBio and Illumina libraires were used to assemble the new and previously published genomes, the significant improvement observed in the new assembly is attributed to the use of a different hybrid assembly approach, combined with the Dovetail Omni-C library prep and scaffolding with the HiRise software. The *M. capitata* isolate that was sequenced appears to be a diploid, with a good fit between the theoretical diploid model implemented in GenomeScope2 and the *k*-mer frequency histogram (black line in fig. 2C and supplementary fig. S1A), and a clear "smudge" (bright yellow region in supplementary fig. S1B and C) of *k*-mer pairs with a coverage of 2n and a normalized coverage of 1/2; all of which suggests that the sample is diploid.

For *P. acuta*, the genome assembly generated in this study (hereinafter the "new genome assembly") was constructed using 17.5 Gbp (1,083,200 reads with an N50 of 17.18 Kbp) of PacBio HiFi data. This new assembly is larger (403.8 Mbp and 550 scaffolds; supplementary table S1; supplementary fig. S2) than the available *P. acuta* (Vidal-Dupiol et al. 2020) (352 Mbp; isolated from Indonesia) and *P. damicornis* (Cunning et al. 2018) (234.3 Mbp; isolated from Panama) assemblies, and is larger than its estimated genome size of 353.3 Mbp (estimated by GenomeScope2 using *k*-mers of size 21 bp, generated from 23 Gbp of Illumina short read data [generated in this study]; fig. 2D). The *P. acuta* isolate that was sequenced is a putative triploid; the major peaks in the *k*-mer frequency histogram (at ~17x and ~51x) appear to fit a triploid model (black line fig. 2D and supplementary fig. S3C) better than a diploid model (black line supplementary fig. S3A and B) and there is a clear "smudge" (bright yellow region in supplementary fig. S3D and E) of *k*-mer pairs with a coverage of ~3n and a normalized coverage of 1/3; all of which suggests that the sample is likely a triploid. The BUSCO completeness of the new assembly is improved for both the Metazoa (95.6%), and Eukaryota (98.8%) datasets (supplementary table S1) compared to the published *P. acuta* and *P. damicornis* assemblies (both have ~90% and 91.4% completeness of the Metazoa and Eukaryota datasets, respectively). However, the new assembly

126 has a high proportion of duplicated BUSCO genes (>15% in both datasets) compared with the
127 published datasets (both have <1% in both datasets), which suggests that the new *P. acuta*
128 assembly contains haplotigs that have not been removed. The genome assembly workflow
129 implemented by DNA Link Sequencing Lab does include a haplotig removal stage, which in this
130 case, appears to have been ineffective at removing all duplicate scaffolds. This could have been
131 caused by the *P. acuta* sample that was sequenced being a putative triploid, which possibly
132 confounded the haplotig identification process. The program used for haplotig removal, Purge
133 Haplotigs (Roach et al. 2018), was designed for use with diploid species and the added
134 complexities of a triploid genome have likely prevented it from fully resolving the assembly and
135 removing the remaining haplotigs.
136
137 **Predicted genes in the new Hawaiian coral genome assemblies**
138 Genes were predicted in the new *M. capitata* assembly using Liftoff (Shumate and Salzberg 2020),
139 which transferred over the gene models from the published genome assembly to the new assembly
140 (preserving the original gene names), and in the new *P. acuta* assembly using BRAKER (Bruna et
141 al. 2021), which was used to *ab initio* predict a new set of genes using evidence provided by the
142 alignment of the RNA-Seq data generated in this study. For *M. capitata*, 55,086 genes were
143 reconstructed in the new assembly compared with the 63,227 predicted in the published version
144 (supplementary table S2). The BUSCO completeness of the new predicted genes is reduced in the
145 new assembly (86.9% of the Metazoa and 85.9% for the Eukaryota BUSCO datasets) compared
146 with the published assembly (94.0% and 93.3%, respectively) however, the number of duplicated
147 BUSCO genes is reduced in the new assembly (0.8% for both BUSCO datasets) compared to the
148 published (18.2% and 18.8%, respectively). Of the 55,086 new gene models, 42,043 (76.32%) had
149 "valid" open reading frames (ORFs: i.e., they contained both start and stop codons with no in-frame
150 stop codons) and 13,043 had "invalid" ORFs (i.e., they are missing start or stop codons, or have
151 in-frame stop codons).
152

153 The reduction in the number of predicted genes in the new *M. capitata* assembly, compared with
154 the published version (supplementary table S2), is likely driven by its reduced assembly size, with
155 many of the missing genes arising from retained haplotigs that were removed in the new assembly.
156 There are also a significant number of genes with invalid ORFs in the new *M. capitata* genome that
157 are missing start or stop codons or have in-frame stop codons. Because the new *M. capitata*
158 genome was constructed using the same genome sequencing libraries as the published version
159 these invalid gene models may represent variation between the different haplotypes in *M. capitata*
160 (i.e., these genes could be from regions where the published and new assemblies have resolved
161 different haplotypes). They could also result from regions of the genome that have chimeric phasing
162 (i.e., a region is a mix of different haplotypes), or it could be that these genes are actually non-
163 coding RNA that contain valid ORFs in one haplotype and not the other. The differences between
164 the new and published predicted genes remains to be further investigated however, it does not
165 affect our downstream analysis because these genes are only used as a guide for selecting SNPs
166 that are from putative functional regions of the genome. The reduction in the BUSCO completeness
167 of the new *M. capitata* predicted genes is likely a reflection of these challenges, with the differences
168 between the assembled haplotypes, or chimeric haplotypes, the putative primary cause.
169
170 There are 36,794 genes in the new *P. acuta* genome assembly (supplementary table S2), which is
171 10,717 more than predicted in the *P. damicornis* genome (Cunning et al. 2018) but roughly
172 equivalent to the 36,140 genes (63,181 alternatively spliced transcripts) in the published *P. acuta*
173 "experimental" set (Vidal-Dupiol et al. 2020). Because the *P. acuta* experimental gene set was
174 predicted directly from aligned RNA-Seq data it does not have predicted ORFs, and thus likely
175 included a number of non-coding RNA. However, it represents a better approximation of the
176 structure of genes in *P. acuta* then the other "*ab initio*" set (with 64,558 genes) which was reported
177 by the authors of the study to be highly fragmented (Vidal-Dupiol et al. 2020) and has a very low
178 BUSCO completeness (39.7% for the Metazoa and 36.8 % for the Eukaryota datasets). The
179 BUSCO completeness of the predicted genes in the new *P. acuta* assembly is higher (95.9% for

180 the Metazoa and 97.2% for the Eukaryota BUSCO datasets) than in *P. damicornis* (87.9% and
181 86.3%, respectively) and in the *P. acuta* experimental set (84.7% and 83.5%, respectively)
182 however, it also has a higher proportion of duplicated BUSCO genes (26.4% and 22.7%,
183 respectively) than *P. damicornis* (0.2% and 0.4%, respectively) but a lower proportion than in the
184 *P. acuta* experimental set (37.5% and 37.6%, respectively), which has a high number of predicted
185 alternatively spliced transcripts that is likely driving this increase. Other statistics about the
186 predicted genes in the new *P. acuta* assembly are largely congruent between the *P. damicornis*
187 and *P. acuta* experimental datasets. For example, the mean number of introns per gene is similar,
188 albeit slightly higher, in new assemblies predicted genes (6.84) compared with the *P. damicornis*
189 (6.17) and *P. acuta* experimental (5.24) gene sets.
190
191 **Genotyping of *P. acuta* RNA-Seq samples**
192 For *P. acuta*, RNA-Seq libraries were generated from 119 coral colonies collected from six reefs
193 across Kāneʻohe Bay. These libraries were genotyped and filtered (samples assumed as diploid;
194 see Materials and Methods), resulting in the identification of 12,969 variant sites. These were
195 converted into an alignment with 13,948 parsimony-informative, 1,835 singleton, and 10,155
196 constant sites (25,938 total). An unrooted phylogenetic tree (fig. 3A) constructed using the
197 parsimony-informative and singleton sites (13,948 total) shows that many of the *P. acuta* samples
198 are grouped into well supported (100% bootstrap support [BS]) clades with short internal branches,
199 indicating they are derived from the same genotype. We estimate that there are at least six distinct
200 *P. acuta* genotypes (clades PA1-6), that were sequenced in multiple samples, that have spread
201 throughout the Bay *via* asexual reproduction. The three other clades highlighted in the tree (PA7-9
202 in fig. 3A) are separated from the other major clades by well-supported nodes (100% BS) and
203 longer branches. Genotype PA9 was sampled once in our analysis. Clades PA7 and PA8 include
204 a small number of samples and are more diverged than the six major clades (e.g., branch lengths
205 between the samples in clade PA5 are small [<0.01] when compared to, for example, the two
206 samples in PA8 [~0.28]); the more diverged samples in these clades may representing district

genotypes. A tree constructed using only SNPs that overlap with the exons of genes predicted in the new *P. acuta* genome (3,730 parsimony-informative and 483 singleton sites; 4,213 total; supplementary fig. S4A) and a tree constructed directly from the 12,969 variant sites (supplementary fig. S4B) using the identity-by-state proportions calculated for each sample are consistent with results of the larger analysis. The topology of an additional tree constructed using SNPs identified in all *P. acuta* samples plus six *P. damicornis* individuals collected in Indonesia (fig. 3B; 1,856 parsimony-informative, 74 singleton, and 71 constant sites [2,640 total]) again, supports the robust grouping of samples shown in figure 3A. The *P. damicornis* samples are positioned near PA7, suggesting that PA6-9 include the initial colonizers of Kāneʻohe Bay (fig. 3B), with subsequent expansion of *P. acuta* populations, resulting in the more derived (e.g., PA5, PA3, PA1) clades.

Analysis of genotype distribution provides some (weak) evidence for clustering based on sample location (fig. 3B). An example is clade PA4 that is predominantly found in the south end of the bay, which has longer water residence times (Lowe et al. 2009), with all samples in this clade collected from Lilipuna, except for one collected from reefs 42&43 (most northern reefs sampled). The PA4 sample from the Lilipuna site may have spread locally *via* asexual reproduction (budding, fragmentation, or self-fertilization brooding: e.g., as described for local *P. damicornis* (Jokiel 1985)), with transport to more northern reefs in the Bay occurring through infrequent stochastic processes. Another example is clades PA6 and PA8 that are predominantly found in the north end of the bay, with the majority collected from reefs 42&43 and 35&36. One sample was collected from reef 18 and one from reefs 11&13 (both reefs are in the middle of the bay), and no samples were from the southern Lilipuna and HIMB reefs.

Genotyping of *M. capitata* RNA-Seq samples

For *M. capitata*, RNA-Seq libraries were generated from 132 samples. These libraries were genotyped and filtered (samples assumed as diploid; see Materials and Methods), resulting in the identification of 14,191 variant sites. These sites were converted into an alignment with 25,541

234 parsimony-informative, 1,165 singleton and 1,676 constant sites (28,382 total). An unrooted
235 phylogenetic tree built using the parsimony-informative and singleton sites (26,706 total) (fig. 4A)
236 demonstrates the stark contrast between *P. acuta* and *M. capitata*. The *M. capitata* samples do not
237 group into diverged clades like those from *P. acuta* (fig. 3A), and the clades that were putatively
238 identified (fig. 4A) are not robust when the data are reduced to SNPs that overlap with the exons
239 from the predicted genes in the new *M. capitata* assembly (supplementary fig. S5). The between-
240 sample branch lengths in the *M. capitata* tree are also quite large (e.g., 0.295 between two closely
241 related samples; fig. 4A) and comparable to the distances between different clades in *P. acuta*
242 (e.g., 0.21 between PA3 and PA4, or 0.327 between PA5 and PA6; fig. 3A). This suggests that
243 each of the *M. capitata* samples represents a distinct genotype, with the only exceptions being
244 Mcapitata_HTAC_TP11_2380 and Mcapitata_ATAC_TP11_2302 (bottom red backet in fig. 4A),
245 and Mcapitata_ATAC_TP12_2403 and Mcapitata_ATAC_TP6_2402 (top red backet in fig. 4A)
246 which are likely to be isolates of single genotypes. The association between these pairs of samples
247 is robust even if the phylogenetic tree is constructed using SNPs that overlap with exons
248 (supplementary fig. S5). Furthermore, these genotypes were isolated from a single reef (Reef 18;
249 fig. 4B). The clades identified in *M. capitata* do not show a correlation with a particular sample site
250 or region of the Bay.
251
252 **The evolution of triploidy and asexual reproduction in *P. acuta***
253 The proportion of reads which support each allele at a given site (as a fraction of all reads that align
254 to that site) was plotted separately for each sample. These SNP allele frequency distributions can
255 be used to assess the ploidy of each sample. SNP variant sites covered by ≥ 20 aligned reads were
256 extracted from the *P. acuta* RNA-Seq samples, the *P. acuta* reference genome, and the *P.*
257 *damicornis* RNA-Seq samples. All samples were run using the ploidy estimated after the first round
258 of genotype analysis to avoid biasing the distributions of triploid samples (see Materials and
259 Methods). The SNP allele frequency distributions (Dataset S1) for *P. acuta* show that some
260 samples have a distribution with a single peak at roughly 0.5, which is expected for diploids (i.e., a

261 site can have two alleles, each present in ~50% of the reads), whereas other samples have
262 distributions with two peaks at roughly 0.33 and 0.66 (fig. 3C and Dataset S1), which is expected
263 for triploids (Weiss et al. 2018) (i.e., a site can have three alleles which would each be present in
264 ~33% of the reads at a site). The two peaks in the *P. acuta* triploid SNP allele frequency distributions
265 all have roughly similar abundances, suggesting that most variant sites are biallelic. There were 44
266 *P. acuta* RNA-Seq samples with a single (diploid) peak and 75 samples with two (triploid) peaks.
267 The *P. acuta* reference genome, which was shown via other approaches to be a triploid
268 (supplementary fig. S3), has two clear peaks at 0.33 and 0.66. The *P. damicornis* samples used as
269 a phylogenetic outgroup in figure 3B all have a single peak centered around 0.5, demonstrating
270 that they are likely diploids. The *P. acuta* diploids likely comprise the ancestral populations that
271 colonized Kāneʻohe Bay because they (i.e., PA5-9) are the deepest splits in the tree rooted with
272 diploid *P. damicornis* samples (fig. 3B). Triploid PA1-4, which are likely to be infertile, appear to
273 have evolved more recently and comprise the majority of collected samples (76/119 [63.9%]; fig.
274 3B). The distributions of *P. damicornis* samples have a less smooth pattern compared to *P. acuta*,
275 this is likely attributed to the smaller number of variant sites that passed filtering in these samples
276 (the numbers shown at the bottom of each plot in Dataset S1), which is likely a result of them having
277 small library sizes (max 12.9 million trimmed read pairs in *P. damicornis* [supplementary table S3]
278 compared to an average of 18.3 million trimmed read pairs in *P. acuta* sample [supplementary table
279 S4]). The diploid *P. acuta* RNA-Seq samples are all in clades PA5-9 with the only exception,
280 Pacuta_HTHC_TP5_1415, being part of PA2. The triploid samples are all in clades PA1-4 and are
281 well resolved from each other (fig. 3A and B). In contrast, all *M. capitata* samples have a single
282 (diploid) peak (Dataset S2).
283
284 We also produced a set of variants from the *P. acuta* RNA-Seq samples and the *P. acuta* reference
285 genome that considered the estimated ploidy of each sample during haplotype calling (see
286 Materials and Methods). This produced a total of 12,888 variant sites, of which 12,584 were biallelic
287 (i.e., only two alleles were observed across all samples). These biallelic sites were converted into

288 an alignment with 13,859 parsimony-informative, 1,363 singleton, and 9,946 constant sites (25,168
289 total). An unrooted phylogenetic tree (supplementary fig. S4C) constructed using the parsimony-
290 informative and singleton sites (15,222 total) shows that the *P. acuta* reference genome groups
291 within clade PA2 and is most closely related to the triploid Pacuta_HTHC_TP11_2185 sample. This
292 result is consistent with the observation that all triploid samples are contained within clades PA1-
293 PA4. A tree was also constructed by clustering samples by their patterns of shared alleles across
294 all variant sites (supplementary fig. S4B). It shows that the genotypes and the separation of
295 samples by ploidy is robust, with the only major differences being the placement of the triploid
296 Pacuta_HTHC_TP11_2185 outside PA2; the diploid Pacuta_HTHC_TP5_1415 sample still groups
297 with the other samples of PA2. Constructing a tree in this way removes the need to represent the
298 variants as positions in an alignment (which is challenging when ploidy varies across samples or
299 across sites, and when alleles are not fully phased) and to select an appropriate evolutionary model
300 for the alignment.
301
302 **Conformation of sample ploidy using nQuire**
303 The nQuire program was used as an alternative approach to study the ploidy of each *P. acuta*
304 and *M. capitata* RNA-Seq sample, as well as the *P. acuta* reference genome, and the *P.*
305 *damicornis* samples used as an outgroup in figure 3B. For all *P. acuta* samples, the ploidy model
306 with the lowest nQuire delta Log-Likelihood value (supplementary fig. S6A; supplementary table
307 S5), which indicates that a particular model is the best fit for the sample (Weiss et al. 2018),
308 matches the ploidy estimated from the SNP allele frequency distributions (Dataset S1). nQuire
309 also confirmed that a triploid model is the best fit for the *P. acuta* reference genome, and a diploid
310 model is the best fit for the *P. damicornis* samples. The delta Log-Likelihood values of the
311 samples after denoising (supplementary fig. S6B; see Materials and Methods) also supports
312 these results, with the only anomalies being the triploid Pacuta_ATAC_TP11_2306 and
313 Pacuta_HTAC_TP1_2414 which have extremely high delta Log-Likelihood values compared with
314 the other samples. Regardless, a triploid model was still the best fit for these samples (in

agreement with our other analysis) and because these sample are not outliers in supplementary figure S6A it is possible that this is an artifact introduced by the denoising process. For *M. capitata*, all samples were identified as diploids except for Mcapitata_HTAC_TP12_1632 and Mcapitata_ATAC_TP11_1644 which were identified as a tetraploid and a triploid, respectively (supplementary table S5; supplementary fig. S6C samples highlighted in red). Whereas it is possible that these samples are a tetraploid and a triploid, the SNP allele frequency distributions (Dataset S2; supplementary fig. S6E) do not support this hypothesis. Tetraploid samples would have three peaks in the distribution at approximately 0.25, 0.5, and 0.75 along the x-axis and triploid samples would have two peaks in the distribution at approximately 0.33, and 0.66 along the x-axis (Weiss et al. 2018), which is not observed for these samples (supplementary fig. S6E). Instead, the distributions have a single peak at ~0.5 and an increase in the frequency of alleles at the tails of the distribution (increasing as the distribution approaches 0.1 and 0.9 along the x-axis); this is in contrast with the other *M. capitata* samples where there is a single peak at 0.5 and the tails of the distribution do not show an increase in allele frequency. The increase in frequency of alleles that occupy the tails of the distribution could result from the samples being a mix of (at least two) different genotypes. If one of the genotypes in the sample was dominant (present at a much higher frequency than the others) then we would see an increase in the frequency of SNP alleles with support around 0.1 and 0.9. These SNPs, rather than originating from heterozygous regions of the different haplotypes within a single genotype, represent heterozygous regions between different genotypes, and thus the proportion of reads that support the alleles reflects the relative abundances of the different genotypes and not the ploidy of the sample. Only Mcapitata_HTAC_TP12_1632 was still identified as a tetraploid in the denoised data (supplementary fig. S6D), Mcapitata_ATAC_TP11_1644 was identified as a diploid, but did have a higher delta Log-Likelihood value than the other samples, showing that the diploid model did not fit this sample as well as it did for the other samples with distributions that match the expected

340 pattern (i.e., the samples without an increase in the frequency of alleles with support around 0.1
341 and 0.9).
342
343 **Physiological response of corals to stress conditions**
344 To investigate the relationship between genotype and gene expression, we exposed *M. capitata*
345 and *P. acuta* coral fragments to abiotic stress that mirrored real-world conditions. We found that
346 both corals paled in color significantly in response to treatment through time (3-way interaction of
347 Temperature, pCO_2, and Time for *M. capitata* [$F_{1,14}$ = 60.45, p = 0.00000001] and *P. acuta* [$F_{1,10}$ =
348 71.95, p < 0.00000000002]; supplementary table S6). There was a much more substantial
349 reduction in color score under high temperature in *P. acuta* than *M. capitata* (fig. 1D). The high
350 temperature and low pH conditions, however, led to relatively little mortality with no differences in
351 survivorship by treatment in *M. capitata* for high temperature (p = 0.20), high pCO_2 (p = 0.77), or
352 their interaction (p = 0.97). In contrast, there was >80% mortality in *P. acuta* exposed to high
353 temperature by the end of the experiment (fig. 1E), with significant mortality in response to high
354 temperature (p = 0.029) and high pCO_2 (p = 0.030), and their interaction (p = 0.013, supplementary
355 table S7).
356
357 **Effects of genotype on RNA-Seq expression profiles**
358 The reads from the *P. acuta* RNA-Seq samples were aligned against the predicted genes of the
359 published *P. acuta* genome (Vidal-Dupiol et al. 2020) and the correlation between each sample
360 was calculated based on the resulting gene expression profiles. For the heatmaps shown in figure
361 5A and supplementary figure S7, each panel provides different information about the samples and
362 makes clear that sample expression profiles do not cluster based on treatment condition, timepoint,
363 or the collection location. Rather, they cluster (fig. 5A) based on the putative clades/genotypes that
364 were identified (fig. 3A). There is a preponderance of samples from PA1, PA3, and PA5 (the largest
365 clades) in the three major blocks (encapsulated by red squares in fig. 5A and supplementary fig.
366 S7) of correlated genes in the heatmap. The samples from clade PA2 are clustered with those from

367 PA1, which is expected because these two clades are more closely related to each other then they
368 are to others in the tree (fig. 3A). Samples from PA6, PA7, and PA8 also cluster based on their
369 positions in the tree.
370
371 The same approach, using the published *M. capitata* genome (Shumaker et al. 2019), was
372 undertaken to assess the correlation between the expression profiles of the *M. capitata* RNA-Seq
373 samples. From this analysis, it is clear that, similar to *P. acuta*, the expression profiles of *M. capitata*
374 do not cluster based on treatment, timepoint, or collection site (supplementary fig. S8). However,
375 they also do not cluster strongly based on the clades (fig. 5B) that were identified (fig. 4A), which
376 is unsurprising given that almost all *M. capitata* samples are distinct genotypes; in contrast many
377 of the *P. acuta* clades were composed of one, or only a few, distinct genotypes. The only clear case
378 of *M. capitata* sample clustering are two pairs of samples (Mcapitata_HTAC_TP11_2380 and
379 Mcapitata_ATAC_TP11_2302, and Mcapitata_ATAC_TP12_2403 and
380 Mcapitata_ATAC_TP6_2402; encapsulated by red squares in fig. 5B and supplementary fig. S8)
381 that are putatively derived from the same genotypes.
382
383 **Coral symbiont composition**
384 Algal composition differed greatly between these species with *P. acuta* dominated by clade C1
385 (*Cladocopium* spp.), whereas *M. capitata* housed clades C17/C31 or D1 (*Durusdinium* spp.), or a
386 mix of these algae (supplementary fig. S9). These results indicate that under the conditions used,
387 *P. acuta* is a relatively stress sensitive coral that displays substantial bleaching and high mortality
388 under extended heat stress alone, and in combination with low pH, which is supported by in situ
389 bleaching studies (Bahr et al. 2015; Jokiel and Brown 2004). *P. acuta* is, however, widely distributed
390 in Kāneʻohe Bay and can recruit under projected levels of ocean warming (Bahr et al. 2020).
391

392 **Symbiont holotranscriptome**

393 There is a diverse collection of Symbiodiniaceae symbionts in the coral colonies we studied
394 (supplementary fig. S9) for which we have no reference genomes available to discriminate between
395 genes and paralogs in different lineages. Therefore, we analyzed the photosymbionts as a holo-
396 transcriptome (see Materials and Methods) to identify samples that share the same expression
397 pattern of non-coral-derived transcripts. The *P. acuta* holotranscriptome consists of 1,066,584
398 transcripts (contig N50 of 971 bp; supplementary table S8) that were assembled from ca. 533
399 million read pairs (supplementary table S9) that did not align to the published *P. acuta* genome
400 assembly (Vidal-Dupiol et al. 2020). BLASTn analysis using custom databases of Symbiodiniaceae
401 and coral-derived sequences found that 261,082 (24.28%) of the *P. acuta* holotranscriptome
402 sequences have top hits to the coral database, 248,841 (23.33%) have top hits to the symbiont
403 database, and 556,661 (52.19%) do not have top hits to either database (i.e., are derived from an
404 unknown source). The *M. capitata* holotranscriptome consists of 1,293,200 transcripts (contig N50
405 of 1,143 bp; supplementary table S8) that were assembled from ca. 628 million read pairs
406 (supplementary table S9) that did not align to the published genome (Shumaker et al. 2019).
407 BLASTn analysis of the holotranscriptome identified 217,816 (16.84%) coral-derived and 238,663
408 (18.46) symbiont-derived transcripts, the remaining 836,721 (64.70%) transcripts are derived from
409 other unknown sources.

410

411 Clustering of the expression profiles of just the algal symbiont-derived transcripts from the
412 holotranscriptome of each species showed that for *P. acuta*, the samples grouped not based on
413 treatment, timepoint, or collection site but primarily by host genotype (fig. 5C). Whereas the effect
414 of genotype is not as pronounced in the symbiont data as it is for the host (fig. 5A), it is still present
415 with samples from PA5, PA3, and PA1 forming clusters. This loose grouping is also observed in
416 the expression profiles of the non-coral-derived transcripts (supplementary fig. S10A; which
417 includes unknown and symbiont-derived transcripts) but is weaker than in the analysis using only
418 symbiont-derived transcripts. The algal Internal Transcribed Spacer 2 (ITS2) profiles of the *P. acuta*

samples are almost always dominated by a single ITS2 profile, with *Cladocopium* "C1d/C42.2−C1−C3cg" ITS2 being the most common type observed across the samples. Only three samples have the *Durusdinium* ("D1/D4−D6−D1ab") ITS2 profiles (fig. 5C), and all *Cladocopium* "C1d−noName−C42.2−C1−noName" ITS2 profiles were detected in samples from clade PA5 (red boxes in fig. 5C and supplementary fig. S10A; supplementary table S10). The *M. capitata* samples have more defined groups within the heatmap that do not correlate with treatment, timepoint, collection site, or host clade, but with the ITS2 profiles (supplementary fig. S10B; supplementary table S10). There are three major groups of samples in the heatmap of the non-coral-derived transcripts (supplementary fig. S10B, highlighted with four red boxes) that represents (*i*) samples with a mix of algal symbionts that always includes a significant fraction of *Durusdinium* sequences (top and bottom red boxes), (*ii*) samples dominated by a single *Cladocopium* ITS2 type sequence (second red box from the top; the only exception in this group is Mcapitata_HTAC_TP10_1315 which contains a small fraction of *Durusdinium*), and (*iii*) samples with all, or a majority, or their ITS2 sequences from *Durusdinium* (second red box from the bottom). The strong separation of samples in the *M. capitata* heatmap is likely caused by orthologs in different symbiont lineages being resolved as separate transcripts. This means that even if the same suite of algal genes were being expressed across all samples, their reads would assemble and map to different transcripts, causing the expression profiles of samples with differing symbiont compositions to have very low similarity and low correlation in the heatmap, and would cause the expression profiles of samples with the same symbiont composition to have relatively high similarity and high correlation in the heatmap.

Discussion

In this study, we generated high quality genome assemblies and RNA-Seq data from colonies of the stress sensitive *P. acuta* and stress tolerant *M. capitata* from six reefs in Kāneʻohe Bay. The *M. capitata* assembly is a significant improvement over the existing version (Shumaker et al. 2019), and in the case of *P. acuta*, represents the first Hawaiian isolate sequenced for this species and

446 the first genome assembly from a non-diploid coral colony. We used SNPs to interrogate the
447 population structure of the two species across the bay, identifying robust genotypes and variation
448 in ploidy in *P. acuta*, and using various stress treatments, demonstrated the strong linkage between
449 coral genotype and transcriptome response to abiotic stress.
450
451 **Two high-quality genome assemblies from key Hawaiian corals**
452 The chromosome-level assembly of *M. capitata* is an important resource for comparative and
453 resequencing efforts of corals from Hawaii and other reefs across the world. The 1,687 "orphan"
454 scaffolds (totaling ca. 101 Mbp of assembled sequence), that are not part of the 14 putative
455 chromosomes, are of unknown provenance and remain to be further explored. The *P. acuta*
456 genome assembly was generated from a triploid isolate and is larger than the estimated genome
457 size (403.8 Mbp vs. 353.3 Mbp, respectively), suggesting that some haplotigs still remain, which is
458 unsurprising given the complexities associated with processing a triploid vs. a diploid genome
459 (which most tools are benchmarked against). Regardless, the generation of an assembly from a
460 triploid coral in particular where closely related diploid and triploid individuals live in close proximity,
461 enables research into how polyploidization affects the adaptability of corals and raises larger
462 questions about the plasticity of coral genomes. The estimated sizes of the published *P. acuta*
463 (isolated from Indonesia), *P. damicornis* (from Panama) and new *P. acuta* genomes are similar
464 (Cunning et al. 2018; Vidal-Dupiol et al. 2020) suggesting that globally, *Pocillopora* species have a
465 genome size of ca. 350 Mbp.
466
467 ***P. acuta* has limited genetic diversity and frequently undergoes asexual reproduction**
468 Our results suggest that after introduction, *P. acuta* genotypes have spread to most sites in
469 Kāneʻohe Bay with active transport between reefs. This "genotypes everywhere" result has
470 previously been found for Kāneʻohe Bay *P. damicornis* populations (Gorospe and Karl 2013). Using
471 microsatellite data, Gorospe and Karl (Gorospe and Karl 2013) studied a single reef and found that
472 >70% of the colonies comprised seven genotypes with high clonal propagation. Neighboring reefs

473 however conformed to a genetic panmixia model with no inter-reef genetic structure. Our results
474 support this model, showing the existence of at least nine distinct *P. acuta* genotypes with
475 widespread distribution of the major lineages in Kāneʻohe Bay. The absence of isolation by
476 distance, even when individual reefs show genetic structure is puzzling, and may be explained by
477 microhabitat variability that selects for particular genotypes that occupy these niches in different
478 reefs (Gorospe and Karl 2011): i.e., presuming no major barriers exist to larval dispersal in the bay.
479 This result might also be explained by a genetic bottleneck. A recent natural event, such as severe
480 bleaching that caused mass coral mortality (e.g., the 2014-2015 Kāneʻohe Bay bleaching event
481 (Bahr et al. 2015)), may have removed much of the *P. acuta* from this region. The subsequent
482 repopulation of Kāneʻohe Bay by surviving corals, or recolonization from different regions, coupled
483 with asexual reproduction, would result in the observed, low genetic diversity.
484
485 ***M. capitata* is a highly diverse outbreeder**
486 The genetically diverse *M. capitata* colonies (fig. 4A) stand in stark contrast to the distinct
487 genotypes observed in *P. acuta* (fig. 3A). Despite the *M. capitata* tree having been constructed
488 from a larger input alignment then the *P. acuta* tree, and many of the nodes having strong bootstrap
489 support, the *M. capitata* clades are not robust when the data are reduced to variants that overlap
490 with the exons from the predicted genes in the new assembly (supplementary fig. S5). This
491 instability is likely a result of the high genetic diversity and sexual reproduction among Kāneʻohe
492 Bay *M. capitata*. The "star radiation" apparent in the *M. capitata* analyses suggest that we are not
493 dealing with potential cryptic or closely related species of *Montipora* in our dataset because of the
494 lack of structure in these trees (fig. 4A; supplementary fig. S5) with each lineage equally diverged
495 from the others. These results are supported by previous studies (Caruso et al. 2021; Concepcion
496 et al. 2014), with a recent survey of almost 600 *M. capitata* colonies in Kāneʻohe Bay that found
497 very few clonal individuals and no evidence of isolation by distance (Caruso et al. 2021). The few

clonal colonies that were identified were almost exclusively found at the same collection site, which is what we also observed in our analysis.

Genotype impacts RNA-seq expression profile

Genotype has a strong impact on the overall gene expression profile of coral samples, even though they were subjected to highly contrasting treatments. This is highlighted by *P. acuta*, where the grouping of samples by gene expression profile is explained predominantly by host clade (which for *P. acuta* may represent individual genotypes), and obscures differences based on stress treatment (fig. 5A). This pattern is apparent when considering samples from the control (ATAC) and high temperature/high pCO_2 (HTHC) treatments (supplementary fig. S7). These coral fragments cluster in the heatmap based on genotype even though the extended HTHC treatment led to *P. acuta* mortality. The *M. capitata* expression profiles also reflect genotype and not the experimental design. The *M. capitata* samples, most of which are distinct genotypes, do not form clear clusters that associate with any of the analyzed metainformation (fig. 5B and supplementary fig. S8). The only clear cases of clustering are the two pairs of samples that are putatively derived from the same genotypes.

Ploidy varies between species and may affect *P. acuta* fitness

Our results suggest that the diploid *P. acuta* are both sexual outbreeders and generate asexual brooded larvae, as previously described (Nakajima et al. 2018; Richmond and Jokiel 1984; Schmidt-Roach et al. 2014). The evolution of triploid genotypes in this species may be an adaptation to local conditions in Kāneʻohe Bay, possibly allowing them to outcompete the ancestral diploid genotypes. Although triploids are rare in wild populations, they are frequent in plants and in animals such as oysters, often confer beneficial commercial traits such as improved growth, pathogen resistance, and through infertility, protection of superior, adapted genotypes (Kang et al. 2013). Triploids may also enhance the rate of autotetraploid formation (Husband 2004). Triploids occur in the coral *Acropora palmata* and may be a path to generating different ploidy levels in

525 different members of this genus (Iliana et al. 2005; Kenyon 1997). A previous study of Hawaiian
526 *Pocillopora* species showed the presence of three haplotypes per colony for several genes,
527 suggesting the existence of triploidy in this genus (Flot et al. 2008). Interestingly, the presence of
528 a single sample (Pacuta_HTHC_TP5_1415) in clade PA2 that is diploid despite all other members
529 of this clade being triploid, suggests that this *P. acuta* lineage may have reverted to the ancestral
530 state of sexual reproduction.
531
532 Whereas it is possible that the difference in ploidy between *P. acuta* samples could be distorting
533 the structure of the tree and inflating the separation between clades PA1-4 and clades PA5-9, this
534 scenario seems unlikely given that we see structure within the two ploidy groups (i.e., clear
535 subclades) and one diploid sample groups within PA2 with strong node support (100% BS). These
536 data suggest that the separation of different ploidies in the tree is the outcome of evolutionary
537 history and not analysis artifacts. It is also interesting that the two-peaks in the triploid *P. acuta*
538 SNP allele frequency distributions are of similar abundance (Dataset 1). If all three alleles in these
539 samples were evolving independently then we would expect to see a larger peak at 0.33 (whereby
540 each of the chromosome copies have different alleles) than at 0.66 (where two of the chromosome
541 copies encode the same allele). This result suggests that triploid *P. acuta* may have arisen from
542 the self-fertilization of a *P. acuta* egg, followed by fertilization by a foreign sperm, or that one of the
543 two gametes was diploid and provided two closely related sets of alleles. Alternatively, failure of
544 the ovum to extrude the second polar body after fertilization could lead to triploidy. These are the
545 most common mechanisms for generating triploid plants and animals (Carson et al. 2018;
546 Rosenbusch 2008). It is also conceivable that the two-peaks in the SNP allele frequency
547 distributions are explained by chimeric *P. acuta* colonies. There is evidence of chimerism in corals
548 through fusion of two or more genetically distinct individuals (Oury et al. 2020; Rinkevich et al.
549 2016; Willis et al. 2006) as well as mosaicism *via* somatic cell mutations (Schweinsberg et al. 2015;
550 Willis et al. 2006). The two-peaked distributions of *P. acuta* are however difficult to explain under
551 chimerism because the fused colonies would have to comprise one haploid and one diploid

individual to generate this result. If both colonies were diploid, as expected, then there would be a 4-allele (tetraploid) profile with peaks at 0.25, 0.5, and 0.75 (Weiss et al. 2018). Given these considerations, we suggest that the most likely scenario to explain our results is triploid *P. acuta* individuals rather than fused/mixed diploids.

Host genotype constrains symbiont diversity

The tendency for *P. acuta* to contain the same dominant symbiont type highlights how the reproductive strategy of the host (asexual brooding or fragmentation) may constrain holobiont genetic diversity. The close association between *P. acuta* and *Cladocopium* "C1d/C42.2−C1−C3cg" might provide a selective advantage as the two partners co-evolve, producing a more adaptive partnership (e.g., in terms of photosynthate exchange and metabolic synchronization). This idea is supported by the presence of *Cladocopium* "C1d/C42.2−C1−C3cg" in both diploid and triploid samples, suggesting that this association predates the introduction of *P. acuta* to Kāneʻohe Bay and that the few samples with different algal ITS2 profiles are (potentially infrequent) examples of symbiont switching. Because we are unable to resolve the relationship between samples within PA1-5, due to their high within-clade relatedness, it is not possible to identify how often symbiont switching has occurred in Kāneʻohe Bay *P. acuta*. For example, PA5 is the only clade to contain "C1d−noName−C42.2−C1−noName" and it is possible that this alga became the dominant symbiont in a single ancestral PA5 colony that gave rise to all other colonies in the Bay with this ITS2 profile.

Conclusions

Our results demonstrate that two sympatric coral species collected from the same reefs in Kāneʻohe Bay have very different population structures, genetic histories, and reproductive strategies. These factors affect their transcriptomic response to stress and the diversity of their algal symbiont populations. *P. acuta* comprises a limited number of distinct genotypes, which is consistent with a previous analysis of population structure for *P. damicornis* (Gorospe and Karl 2013) from this bay.

579 The most frequent *P. acuta* genotypes in Kāneʻohe Bay have spread through asexual reproduction
580 and share high genetic similarity, symbiont composition, and transcriptomic profiles despite being
581 subjected to different stress conditions. The variation in ploidy within these samples, with triploid
582 samples likely propagating asexually, reveals the complex genetic history of this species. In
583 contrast, *M. capitata* has uniform (relatively) high genetic diversity between samples, is a sexual
584 outbreeder, and has limited identifiable genotypes (that is, each sample is a distinct genotype). Key
585 questions that remain to be explored include: how will the different reproductive strategies of the
586 two species affect their ability to recolonize an area after a major bleaching event and what selective
587 advantage (if any) has allowed triploid *P. acuta* to spread throughout the bay? Perhaps most
588 importantly, our results demonstrate a clear need to thoroughly characterize the genetic structure
589 and reproductive behavior of corals for informed interpretation of 'omics' data. The underpinnings
590 of coral resistance and resilience involve not only algal symbiont and prokaryotic microbiome
591 contributions, but also fundamentally reflect host genotype(s) and their provenance. This is
592 particularly pertinent with the acceleration of the search for transferrable mechanisms for use as
593 human interventions to assist coral resistance and resilience and maintain these invaluable coral
594 reef ecosystems under increasing global change.
595
596 **Materials and Methods**
597 **Coral Collections**
598 This study focused on two dominant coral species that vary in their environmental tolerance in
599 Kāneʻohe Bay: *Montipora capitata* and *Pocillopora acuta* (Bahr et al. 2016; Gibbin et al. 2015).
600 About 900 coral samples (one ~5x5cm fragment per colony from 75 colonies per species per site)
601 were collected from six reef areas ranging across the north to south span and fringing to patch
602 reefs of Kāneʻohe Bay (Lilipuna Fringe: 21°25'45.9"N 157°47'28.0"W; HIMB: 21°26'09.8"N
603 157°47'12.7"W; Reef 18: 21°27'02.9"N 157°48'40.1"W; Reef 11, 13: 21°27'02.9"N
604 157°47'41.8"W; Reef 35, 36: 21°28'26.0"N 157°50'01.2"W; Reef 42, 43: 21°28'37.9"N
605 157°49'36.8"W; supplementary fig. S11A and B) under Hawaiʻi Department of Aquatic Resources

606 Special Activity Permit 2019-60, between 4-10 September 2018. Corals were affixed in an upright
607 position to individually numbered plastic coral fragment mounting plugs by hot gluing the base of
608 the skeleton to the plug. Coral fragments were allowed to acclimate for ~14 days and then were
609 randomly allocated to mesocosm tanks. The 450 samples collected from each species were
610 randomly allocated to different projects, the results of which are not presented here; after
611 treatment 132 *M. capitata* and 119 *P. acuta* samples had RNA-Seq libraries generated (the
612 increased mortality of the *P. acuta* samples, and problems with the extraction of some samples,
613 resulted in the difference in the number of samples between the two species; a more detailed
614 description is presented in a following section).
615
616 **Experimental Setup**
617 Treatment conditions (n=3 tanks treatment^{-1}) were randomly assigned to twelve outdoor
618 mesocosm tanks (122 cm x 122 cm x 30 cm; 510 L). Flow rates, measured daily with a graduated
619 cylinder and timer, averaged 84.36 ± 1.20 mL second^{-1} (mean ± sem; n=826), resulting in the
620 mesocosm tank turnover every ~1.6 hours. Temperature treatment conditions were programmed
621 to mimic the natural daily fluctuations (0.75 °C ± 0.06) of the environment at the collection sites in
622 Kāneʻohe Bay, Hawaiʻi (NOAA Moku o Loe Buoy data from September 2018). A Neptune Apex
623 control system was set-up to generate a pH-stat feedback system. Seawater temperature and pH
624 were constantly monitored using an Apex Aquacontroller (Neptune System). A microprocessor-
625 controlled power strip (Apex Neptune Energy Bar 832) and wifi base unit (Apex Neptune
626 Controller Base Unit) controlled 12 individual control units (Apex Neptune PM1 pH/ORP Probe
627 Module). Each tank was measured using Neptune Systems Apex Extended Life Temperature
628 Probe (accuracy = ± 0.05 °C) and a Neptune Systems Apex pH Probe (accuracy = ± 0.01 pH)
629 that measured temperature and pH every second. Temperature was monitored with Apex, but
630 controlled separately by underwater heaters (ProHeat D-1500 Heater Controllers, precision ± 1
631 °C), whereas pH was both monitored and controlled with the Apex system.
632

633 Daily fluctuating pH levels were generated with an independent pH-stat feedback system in each
634 tank in order to simulate predicted future higher CO_2 concentration conditions (Pachauri et al.
635 2014). Two 99.99% food-grade CO_2 cylinders were connected to an automatic gas cylinder
636 changeover system (Assurance Valve Systems, Automatic Gas Changeover Eliminator Valves
637 #6091) to prevent an abrupt shortage of CO_2 supply. CO_2 was brought into the system on-
638 demand, based on the pH reading in the tank (Neptune Apex control system) using gas flow
639 solenoids (Milwaukee MA955) with airlines plumbed into a venturi injector (Forfuture-go G1/2
640 Garden Irrigation Device Venturi Fertilizer Injector), which was connected to a water circulating
641 pump (Pondmaster Pond-mag Magnetic Drive Water Pump Model 5). Gas injected into the
642 system was either CO_2 or ambient air and bubbling was constant due to the pressure driven
643 pump moving water and gas through the venturi injector.
644
645 OnSet HOBO loggers (HOBO Water Temp Pro v2, accuracy = ± 0.21 °C, resolution = 0.02 °C)
646 were placed in each tank at the same height as the coral fragments for the duration of the
647 experiment in order to log temperature in 10 minute intervals. Mesocosm tanks were 60% shaded
648 from full irradiance, and photosynthetically active radiation was measured continuously with the
649 Apex cosine corrected PAR Sensor (accuracy = ± 5%) that was cross calibrated to the Li-Cor
650 cosine corrected PAR sensor (LI-192). Measurements were also made in 6 positions in each tank
651 using the Li-Cor 193 spherical underwater quantum sensor (LI-193), which revealed there was no
652 significant difference in light between the tank positions (n=4, p=0.948). Light was then assessed
653 in the center of the tank ~daily throughout the experiment using the Li-Cor 193 spherical
654 underwater quantum sensor. To further reduce any potential position effects with respect to
655 incoming water, heater position, or bubble stream, the positions of the coral fragments in the tank
656 were changed weekly during buoyant weighing.
657
658
659

660 **Experimental Timeline**

661 Coral fragments were exposed to four experimental conditions: Ambient Temperature Ambient
662 pCO_2 (ATAC: 27.47 °C ± 0.13; 8.03 pH ± 0.01, 396 µatm ± 12), Ambient Temperature High pCO_2
663 (ATHC: 27.46 °C ± 0.12; 7.68 pH ± 0.014, 1045 µatm ± 35), High Temperature Ambient pCO_2
664 (HTAC: 29.37 °C ± 0.06; 8.01 ± 0.01, 418 µatm ± 13), and High Temperature High pCO_2 (HTHC:
665 29.51 °C ± 0.06; 7.64 ± 0.02, 1174 µatm ± 55) for two months (22 September - 17 November
666 2018), and then ramped down to ambient conditions over two months (18 November 2018 - 12
667 January 2019) (supplementary fig. S11C and D). During the two-month stress exposure period,
668 one fragment per species per tank (n=3 per treatment) was sampled at random for each
669 molecular time point (0h, 6h, 12h, 24h, and 1, 2, 4, 6, 8, 12, and 16 weeks; supplementary table
670 S11). Sampled fragments were placed in sterile whirlpak bags and snap frozen in liquid nitrogen
671 at 13:00 each day (with the exception of 6 h, 12 and 24 hr time points) and stored at -80 °C.

672

673 **Total Alkalinity and Carbonate Chemistry**

674 To assess the efficacy of the treatments, tank parameters (temperature in °C, pH on the total
675 scale, and salinity in psu) were tracked up to several times daily using a handheld digital
676 thermometer (Fisherbrand Traceable Platinum Ultra-Accurate Digital Thermometer, accuracy = ±
677 0.05 °C, resolution = 0.001 °C) and a portable multiparameter meter (Thermo Scientific Orion
678 Star A series A325). A pH probe (Mettler Toledo InLab Expert Pro pH probe #51343101;
679 accuracy = ± 0.2 mV, resolution = 0.1 mV) and conductivity probe (Orion DuraProbe 4-Electrode
680 Conductivity Cell Model 013010MD; accuracy = 0.5% of psu reading, resolution = 0.01 psu) were
681 used with the Orion A Star meter. To ensure pH (total scale) measurement quality, a tris standard
682 (Dickson Laboratory Tris Batch T27 Bottle 269, 236 and Batch T26 Bottle 198) was measured
683 daily to assess mV as a function of temperature.

684

685 To determine carbonate chemistry of each tank, pH, temperature, and salinity measurements
686 were taken daily (as described above; supplementary table S12). Twice a week, water samples

687 were taken simultaneously to measure total alkalinity (TA). An automated titrator (Mettler Toledo
688 T50) was used to titrate water samples with 0.1M hydrochloric acid (Dickson Laboratory Titrant
689 A3, A14) and accuracy was determined using certified reference material (Dickson Laboratory
690 CO_2 CRM Batches 132, 137, 176). A non-linear, least-squares procedure of the Gran approach
691 (SOP 3b (Dickson et al. 2007)) was used to calculate TA (μmol kg^{-1} seawater). Carbonate values
692 were calculated with SEACARB(v3.0.11) (Gattuso et al. 2015) in R Studio.
693
694 **Mortality and Color Score**
695 Survivorship was recorded daily and analyzed using a Cox proportional hazards regression
696 model including Temperature and pCO_2, with individual fragment ID assayed through time with
697 the calls coxph and Surv in the package survival (Therneau 2020) (supplementary table S13). To
698 assess tissue color change over time, each fragment was photographed once a week for all 16
699 weeks with a red, blue, green color standard ruler. ImageJ (Schneider et al. 2012) was used to
700 extract the mean blue, green, and red color score for each coral fragment and these values were
701 normalized to the mean of the red, blue, and green color standards. The color score was
702 quantified as PC 1 from principal component analysis (Edmunds et al. 2003) (PCA) and analyzed
703 statistically using a repeated measures ANOVA with Type III Wald Chi Square test (lmer function
704 in the lme4 package (Bates et al. 2015) in R), for the fixed factors of Temperature, pCO_2, and
705 Time, with individual fragment ID as the repeated measure (supplementary table S6).
706
707 **Coral samples used for RNA-Seq**
708 During the stress experiment 450 samples per species, were subjected to one of four treatment
709 conditions. Half of the samples collected (225 per species) were used for a separate physiology
710 experiment, the other half underwent DNA and RNA extraction (described below). During the
711 course of the treatment the *P. acuta* corals severely bleached, resulting in the mortality of nubbins
712 from the final timepoint (TP12; 16 weeks). No samples were therefore available for processing
713 and this timepoint is absent from all *P. acuta* treatments. In addition, one *P. acuta* sample, from

714 the HTHC treatment at timepoint 11 (12 weeks) failed during RNA extraction and processing;
715 leaving two samples from this timepoint. For *M. capitata*, due to sample degradation, one sample
716 for HTAC timepoint 1 (0hr) could not be sequenced and a sample from timepoint 4 (12hr) was
717 sequenced in its place; there are now n=2 for *M.* capitata HTAC timepoint 1 and n=4 for timepoint
718 4. In total, 119 RNA-Seq samples from *P. acuta* and 132 from *M. capitata* were sequenced and
719 analyzed in this study (supplementary tables S4 and S11).
720
721 **DNA and RNA Extractions**
722 DNA and RNA were extracted simultaneously from the coral fragments that had been snap frozen
723 and stored at -80 °C. A small piece was clipped off using clippers sterilized in 10% bleach,
724 deionized water, isopropanol, and RNAse free water, and then placed in 2 mL microcentrifuge
725 tube containing 0.5mm glass beads (Fisher Scientific Catalog. No 15-340-152) with 1000 µL of
726 DNA/RNA shield. A two-step extraction protocol was used to extract RNA and DNA, with the first
727 step as a "soft" homogenization to reduce shearing of RNA or DNA. Tubes were vortexed at high
728 speed for 1 and 2 minutes for *Pocillopora acuta* and *Montipora capitata* fragments, respectively.
729 The supernatant was removed and designated as the "soft extraction". Second, 500 µL of
730 DNA/RNA shield was added to the bead tubes and placed in a Qiagen TissueLyser for 1 minute
731 at 20 Hz. The supernatant was removed and designated as the "hard extraction". Subsequently,
732 300 µL of sample from both soft and hard homogenate was extracted with the Zymo Quick-
733 DNA/RNA Miniprep Plus Kit Protocol with the following modifications. A first elution of 10 µL of
734 warmed 10 mM Tris HCl was added to DNA columns for a 15-minute room temperature
735 incubation and centrifuged at 16,000 rcf(g) for 30 seconds. A second elution of 100 µL of warmed
736 10 mM Tris HCl was added for a 5-minute incubation and centrifuged at 16,000 rcf(g) for 30
737 seconds. Only the second DNA elution, a higher quality portion of the sample, was kept for
738 downstream analysis. Broad-range DNA and high-sensitivity RNA quantity (ng_µL) were
739 measured with a ThermoFisher Qubit Fluorometer, DNA quality was assessed using gel
740 electrophoresis, and RNA quality was measured with an Agilent TapeStation System. RNA-Seq

741 samples were sequenced by GENEWIZ (https://www.genewiz.com) using the Illumina NovaSeq
742 6000 platform. DNA from the "soft" extract of a single *P. acuta* sample was send to DNA Link
743 Sequencing Lab (https://www.dnalinkseqlab.com) for sequencing on their PacBio Sequal 2 and
744 Illumina NovaSeq 6000 platforms. A whole *M. capitata* nubbin was frozen and shipped to Dovetail
745 Genomics (https://dovetailgenomics.com) for processing using their Omni-C assay and workflow.
746

747 **Internal Transcribed Spacer 2 Sequencing**

748 "Hard" DNA was diluted to 3.33 ng/uL and 3 uL (10 ng total) was added to a master mix solution
749 containing 50 uL 2X Phusion HiFi Mastermix (Thermo Scientific F531s), 46 µL of DNase/RNase
750 free water, and 2 µL each of 10 uM Forward (GAATTGCAGAACTCCGTG) and 10 uM Reverse
751 (GGGATCCATATGCTTAAGTTCAGCGGGT) ITS2 Primers (Lajeunesse and Trench 2000).
752 Samples were run in three 33 uL replicates (3.33 ng in each replicate) in a polymerase chain
753 reaction (PCR) under the following conditions: 3 minutes at 95 °C, then 35 cycles of 30 seconds
754 at 95 °C, 30 seconds at 52 °C, and 30 seconds at 72 °C, followed by 2 minutes at 72 °C and held
755 indefinitely at 4 °C. Triplicate PCR products were pooled, run on a 2% agarose gel to confirm
756 ~300 bp fragment length, and stored at -20 °C. 45 µL of PCR product was sequenced at the
757 University of Rhode Island Sequencing center (URI GSC). URI GSC performed a second
758 amplification to attach Nextra indices and adapters (5-10 PCR cycles) and amplicons were
759 cleaned with Ampure XP beads. Products were visualized using gel electrophoresis and
760 quantified with qPCR before sequencing as paired-end with 300 cycles double check with HP for
761 each read pair on a Illumina MiSeq. Sequences were identified to "Type Profile" using SymPortal
762 (Hume et al. 2019).
763

764 **Genome size estimation**

765 The genome sizes and ploidies of the *M. capitata* and *P. acuta* samples used to generate the
766 reference genomes were estimated using the GenomeScope2 and Smudgeplot tools (Ranallo-
767 Benavidez et al. 2020). For *M. capitata*, Illumina genome sequencing data was retrieved from

768 NCBI SRA (Run SRR8497577 from BioProject PRJNA509219), trimmed using Trimmomatic
769 (Bolger et al. 2014) (v0.39; ILLUMINACLIP:adapters.fa:2:30:10 SLIDINGWINDOW:4:20
770 LEADING:20 TRAILING:20 MINLEN:36), and decomposed into *k*-mers using Jellyfish (Marcais
771 and Kingsford 2011) (v2.3.0; k=21). The *k*-mer frequency histogram produced by Jellyfish (using
772 the 'jellyfish histo' command) was imported into GenomeScope2 with a theoretical diploid model
773 fitted over the data (fig. 2C and supplementary fig. S1A). Smudgeplot was run (following the
774 workflow from https://github.com/KamilSJaron/smudgeplot/wiki/manual-of-smudgeplot-with-
775 jellyfish) using the *k*-mers extracted by Jellyfish with a lower *k*-mer coverage cutoff of 4 (just after
776 the minimum between the initial error peak and the first major peak) and an upper cutoff of 102
777 (8.5 times the coverage of the first major peak at around 12x coverage) chosen based of the
778 GenomeScope2 profile shown in figure 2C; smudgeplots were generated assuming a haploid
779 coverage of 12x ("smudgeplot.py plot -n 12"). For *P. acuta*, the raw Illumina genome sequencing
780 reads generated as part of this study were decomposed into *k*-mers using Jellyfish (k=21). The *k*-
781 mer frequency histogram was imported into GenomeScope2 twice, once using a theoretical
782 diploid model (supplementary fig. S3A and B), and a second time using a theoretical triploid
783 model (fig. 2D and supplementary fig. S3C). Smudgeplot was run using the *k*-mers extracted by
784 Jellyfish with a lower k-mer coverage cutoff of 7 (just after the minimum between the initial error
785 peak and the first major peak) and an upper cutoff of 153 (8.5 times the coverage of the first
786 major peak around at 18x coverage) chosen based of the GenomeScope2 profile shown in figure
787 2D; smudgeplots were generated assuming a haploid coverage of 18 ("smudgeplot.py plot -n
788 18").
789
790 **Genome assembly and predict of protein-coding genes for SNP analysis**
791 The *P. acuta* genome assembly used in this study was sequences using PacBio HiFi and
792 assembled using FALCON-Unzip and Purge Haplotigs (sequencing and assembly was done by
793 DNA Link Sequencing Lab). Genes were predicted in the genome using the BRAKER pipeline
794 (Altschul et al. 1990; Barnett et al. 2011; Bruna et al. 2021; Camacho et al. 2009; Hoff et al. 2016;

795 Hoff et al. 2019; Li et al. 2009; Lomsadze et al. 2014; Stanke et al. 2008; Stanke et al. 2006)
796 (v2.1.6; --softmasking --gff3) with aligned reads from all 119 *P. acuta* RNA-Seq samples (mapped
797 using STAR (Dobin et al. 2013) [v2.7.8a; --outSAMtype BAM SortedByCoordinate]) provided as
798 evidence for prediction.
799
800 An initial long-read *M. capitata* genome assembly was constructed using canu (Koren et al. 2017)
801 (2.2-development; default parameters) from the PacBio reads available from NCBI BioProject
802 PRJNA509219 (Shumaker et al. 2019). Short-read data (from NCBI BioProject PRJNA509219)
803 was aligned against the genome with bowtie2 (v2.4.2; default options) and was used for polishing
804 by Pilon (Walker et al. 2014) (v1.23; default options). The polished canu long-read assembly had
805 repeats softmasked using RepeatModeler (Flynn et al. 2020) (v2.0.1; default options) and
806 RepeatMasker (Smit et al. 2013-2015) (v4.1.1; default options) before putative haplotigs were
807 removed using HaploMerger2 (Huang et al. 2017) (v3.6; default options using the softmasked
808 genome). The haplotig purged long-read assembly was sent to Dovetail Genomics (plus an
809 associated coral nubbin for high molecular weight DNA extraction) for analysis using their Omni-C
810 and HiRise assembly workflows. This produced a final genome assembly that was of putative
811 chromosome level resolution. The genes from the published *M. capitata* genome (Shumaker et al.
812 2019) (same isolate) were transferred using Liftoff (Shumate and Salzberg 2020) (v1.6.1;
813 Minimap2 (Li 2018) v2.17-r941; -a 0.95 -s 0.95 -polish -cds) to the new chromosome resolution
814 assembly.
815
816 Completeness of the new reference genomes and the predicted genes was assessed using
817 BUSCO v5.0.0 (--mode protein, genome, or transcriptome depending on the dataset being
818 assessed) with the eukaryota_odb10 and metazoa_odb10 datasets (Manni et al. 2021).
819
820
821

822 **RNA-Seq read processing**

823 Adapters and low-quality regions were trimmed from the RNA-Seq data generated in this study
824 using Cutadapt v2.9 (Martin 2011) (--nextseq-trim 10 --minimum-length 25 -a
825 AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC -A
826 AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTA; supplementary table S4). A second round
827 of trimming with Cutadapt, using the output from the first round, was used to remove poly-G
828 regions from the 5′-ends of the second read in each pair (-G G(20) -e 0.0 -n 10 --minimum-length
829 25). Read quality was assessed at each stage using FastQC v0.11.7 (default parameters;
830 http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and MultiQC (Ewels et al. 2016)
831 (v1.9).

832

833 **Identification of Open Reading Frames in published *P. acuta* reference genes**

834 Open Reading Frames (ORFs) were predicted in the available *P. acuta* reference transcripts (the
835 RNA-Seq based "experimental" genes reported in (Vidal-Dupiol et al. 2020)) using TransDecoder
836 v5.5.0 (https://github.com/TransDecoder/TransDecoder); HMMER v3.1b2 was used to search the
837 candidate ORFs against the Pfam database (release 33.1; i-Evalue < 0.001) and BLASTp
838 (v2.10.1; -max_target_seqs 1 -evalue 1e-5) was used to search candidate ORFs against the
839 SwissProt database (release 2020_05), with the resulting homology information used by
840 TransDecoder to guide ORF prediction.

841

842 **Quantification of gene expression**

843 Expression of the published *M. capitata* genes (Shumaker et al. 2019) and the ORFs predicted in
844 the *P. acuta* "experimental" genes were quantified over the sequenced timepoints using Salmon
845 (Patro et al. 2017) v1.10 (--allowDovetail --validateMappings --seqBias --gcBias). The sample
846 correlation heatmaps shown in figure 5, and supplementary figures S7, S8 and S10 were
847 generated using the scripts available as part of the Trinity package (Haas et al. 2013) (see

848 section “Compare Replicates Across Samples” at

849 https://github.com/trinityrnaseq/trinityrnaseq/wiki/QC-Samples-and-Biological-Replicates).

850

851 **Identification of variants in RNA-Seq data**

852 A workflow diagram summarizing the variant identification and analysis pipeline used in this study

853 is presented in supplementary figure S12. Variants (single-nucleotide polymorphisms [SNPs] and

854 insertion and deletion variants [INDELs]) were identified using the GATK (McKenna et al. 2010)

855 (v4.2.0.0) framework, adhering to their best practices workflow

856 (https://gatk.broadinstitute.org/hc/en-us/articles/360035531192?id=4067) whenever possible. For

857 each sample, the trimmed RNA-Seq reads were aligned against the new genomes (generated by

858 this study) using STAR (v2.7.8a; --sjdbOverhang 149 --outSAMtype BAM SortedByCoordinate --

859 twopassMode Basic). Read-group information was extracted from the read names by rgsam

860 (v0.1; https://github.com/djhshih/rgsam; --qnformat illumina-1.8) and added to the aligned reads

861 using gatk MergeBamAlignment (--INCLUDE_SECONDARY_ALIGNMENTS false --

862 VALIDATION_STRINGENCY SILENT). Duplicate reads (that had originating from the same DNA

863 fragment) were removed using gatk MarkDuplicates (--CREATE_INDEX true --

864 VALIDATION_STRINGENCY SILENT) before reads that spanned intron-exon boundaries were

865 split using gatk SplitNCigarReads (default parameters). The base quality recalibration steps

866 suggested in the GATK best practices workflow could not be applied to our samples because a

867 set of expected high confidence variants (which is required for the recalibration process) is not

868 available for either of the species being studied. Finally, haplotypes were called using the aligned

869 reads (with read-group information added, duplicate reads marked, and reads spanning exon

870 boundaries split) using gatk HaplotypeCaller (-dont-use-soft-clipped-bases -ERC GVCF). For

871 each species, the VCF files produced by HaplotypeCaller were combined using gatk

872 CombineGVCFs before being jointly genotyped using gatk GenotypeGVCFs (-stand-call-conf 30 -

873 -annotation AS_MappingQualityRankSumTest --annotation AS_ReadPosRankSumTest).

874

875 The variants produced by GenotypeGVCFs were separated into SNPs and INDELs before two
876 rounds of filtering were applied. In the first round, variant sites were removed if, in the 'INFO'
877 annotation, the variant confidence normalized by depth (QD) was < 20, mapping quality (MQ)
878 was < 50 (<45 for INDELs), allele-specific strand bias estimated by the Fisher's exact test (FS)
879 was >5, and allele-specific strand bias estimated by the symmetric odds ratio test (SOR) was
880 <2.5 (gatk VariantFiltration --filter-name "VarScores_filter_QD" --filter-expression "QD < 20.000" -
881 -filter-name "VarScores_filter_MQ" --filter-expression "MQ < 50.000" --filter-name
882 "VarScores_filter_FS" --filter-expression "FS > 5.000" --filter-name "VarScores_filter_SOR" --
883 filter-expression "SOR > 2.500"). The density distribution of these metrics across the samples
884 from *M. capitata* and *P. acuta* was visualized and used to guild selection of the thresholds used
885 for filtering (supplementary fig. S13). In the second round, for each sample a variant was
886 converted to a no call (gatk SelectVariants --set-filtered-gt-to-nocall) if, in the 'FORMAT'
887 annotation, the read depth (DP) was <20 (gatk VariantFiltration --genotype-filter-name "DP_filter"
888 --genotype-filter-expression "DP < 20.000"). For each sample, the read depth of the variants that
889 passed the first round of filtering were visualized and used to guild selection of the threshold used
890 for the second round of filtering (Datasets S3 and S4). Variants were converted to VCF table
891 format using gatk VariantsToTable (-F CHROM -F POS -GF GT); variant sites where at least one
892 sample did not have a genotype called (i.e., there was insufficient read depth covering that site in
893 a sample for it to be considered for analysis) were removed, leaving just variant sites where a
894 genotype could be confidently called (having passed all quality filtering thresholds) in all samples
895 generated from that species.
896
897 For *P. acuta* a second round of haplotype calling was undertaken using information about the
898 putative ploidy of each sample; ploidy was estimated from the distribution of SNP alleles with a
899 given proportion of supporting reads (Dataset S1). For triploid *P. acuta* samples the gatk
900 HaplotypeCaller stage of the workflow was rerun using the same input reads as the first round but
901 with the "--sample-ploidy 3" parameter set; for diploid *P. acuta* samples the haplotypes called in

902 the first found were reused. The VCF files produced by HaplotypeCaller using the correct ploidy
903 were combined using gatk CombineGVCFs before being jointly genotyped using gatk
904 GenotypeGVCFs (-stand-call-conf 30 --annotation AS_MappingQualityRankSumTest --annotation
905 AS_ReadPosRankSumTest). Variants were filtered using the same thresholds.
906
907 **SNP phylogeny**
908 The filtered SNP VCF table (composed of just sites with variants called in all samples) was
909 converted to a fasta representation (where the genotype of each variant in a sample is
910 represented by a one-character code) using the "vcfTab_to_callsTab.py" and
911 "callsToFastaPhy_speed.py" scripts available from https://github.com/evodify/genotype-files-
912 manipulations. The one-character codes (which represent the genotype called at each site in a
913 sample) were converted into two unambiguous nucleotide characters (A becomes AA, T becomes
914 TT, G:GG, C:CC, R:AG, Y:CT, M:AC, K:GT, S:CG, W:AT, N:NN). This new fasta formatted
915 representation of the sample's genotypes was used as the input alignment for phylogenetic
916 analysis; IQTREE (Nguyen et al. 2015) (v1.6.12) was used to remove the invariant sites from the
917 alignment before being used to construct a maximum-likelihood phylogenetic tree (-m GTR+ASC
918 -bb 2000 -quiet; using the alignment without invariant sites) with node support tested via 2,000
919 ultrafast phylogenetic bootstraps (Hoang et al. 2018) (BS). A second maximum-likelihood
920 phylogenetic tree was constructed using the same workflow except, it only used filtered SNPs
921 that overlapped with the exons of the genes predicted in the new reference genomes. The filtered
922 SNPs were used with the SNPRelate R package (Zheng et al. 2012) (v1.24.0) to construct a
923 phylogenetic tree without the need to first build an alignment representation of the variants. First,
924 SNPs in VCF format were converted to GDS format (using the "snpgdsVCF2GDS" function), then
925 the identity-by-state (IBS) proportion was calculated for each pair of samples ("snpgdsIBS";
926 autosome.only=FALSE) and used to for clustering ("snpgdsHCluster").
927
928

929 **Identification of variants in *P. damicornis* RNA-Seq data**

930 Six RNA-Seq samples from *P. damicornis* (two from SRA Study SRP011059.1 (Vidal-Dupiol et al.
931 2013) and four from SRA Study SRP029998 (Vidal-Dupiol et al. 2014)) were retrieved from SRA
932 and trimmed using Cutadapt v2.9 (-q 10 --minimum-length 25 -a
933 GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG -A
934 AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTA). The *P. damicornis* samples had
935 haplotypes called using the same workflow as the *P. acuta* RNA-Seq samples, with the only
936 exception being that read-group information (shown in supplementary table S3) was added
937 manually to the reads (using gatk AddOrReplaceReadGroups) instead of using rgsam to extract
938 the information automatically from the read names.

939

940 The *P. damicornis* haplotypes were combined with the *P. acuta* haplotypes (run assuming all
941 samples were diploid) using gatk CombineGVCFs. All downstream genotyping and filtering
942 processes were conducted as described in the previous section. A SNP based phylogeny of the
943 *P. damicornis* and *P. acuta* RNA-Seq samples was constructed using the same approach as
944 described in the previous section expect that all sites (including invariant sites) were used for the
945 analysis and the GTR model, not the GTR+ASC model, was used for construction of the
946 phylogenetic tree. This difference in the workflow is because the inclusion of the *P. damicornis*
947 samples resulted in fewer sites (with variants called in all samples) from which an alignment could
948 be constructed, the removal of invariant sites resulted in identical sequences in the alignment that
949 were excluded from analysis by IQTREE.

950

951 **Identification of variants in *P. acuta* PacBio HiFi genome sequencing data**

952 The PacBio HiFi genome sequencing reads were aligned against the new genome using
953 Minimap2 (Li 2018) (v2.20; -ayYL --MD --eqx -x map-hifi; the -R option was used to set the read-
954 group information to be the same as in the unaligned BAM file). Samtools (Danecek et al. 2021)
955 (v1.11) was used to remove unmapped reads and secondary alignments from the Minimap2

956 output, it was also used to sort the output by alignment coordinates. Aligned reads with insertions
957 >10 Kbp were found to cause errors when analyzed by gatk HaplotypeCaller, so they were
958 removed using gatk FilterSamReads (--JAVASCRIPT_FILE script.js --FILTER includeJavascript -
959 -CREATE_INDEX --COMPRESSION_LEVEL 0; the script.js script was coded so FilterSamReads
960 would remove reads with inserts >10 Kbp). The filtered reads had haplotypes called using gatk
961 HaplotypeCaller (-dont-use-soft-clipped-bases -ERC GVCF --sample-ploidy 3); the reference
962 genome haplotype VCF file was combined with the haplotype VCF files from the *P. acuta* RNA-
963 Seq samples (created using the estimated ploidy assigned to each sample) using gatk
964 CombineGVCFs. The combined haplotype VCF file was jointly genotyped using gatk
965 GenotypeGVCFs (-stand-call-conf 30 --annotation AS_MappingQualityRankSumTest --annotation
966 AS_ReadPosRankSumTest). All downstream genotyping and filtering processes were conducted
967 as described in the previous section. A phylogeny of the samples was constructed using SNPs
968 from sites that passed the filtering criteria and that had variants called in all samples. As the
969 samples being analyzed had different ploidies, only variant sites with two alleles observed across
970 all samples (biallelic sites) were used, and the three haplotypes called in the triploid samples
971 were converted to two haplotypes so they could be converted into an alignment with the diploid
972 samples that was then used to construct a phylogeny. At each biallelic site, triploid homozygous
973 alleles (e.g., A/A/A) were converted to diploid homozygous allele (e.g., A/A), and triploid
974 heterozygous alleles (e.g., A/A/C) were converted to diploid heterozygous alleles (e.g., A/C) by
975 collapsing the redundant allele. The condensed biallelic variants were directly converted into a
976 fasta representation (A/A:AA; C/C:CC; G/G:GG; T:T/TT; A/C or C/A:AC; A/G or G/A:AG; A/T or
977 T/A:AT; C/G or G/C:CG; C/T or T/C:CT; G/T or T/G:GT) and used as the input alignment for
978 phylogenetic analysis; IQTREE (Nguyen et al. 2015) (v1.6.12) was used to remove the invariant
979 sites from the alignment before being used to construct a maximum-likelihood phylogenetic tree (-
980 m GTR+ASC -bb 2000 -quiet; using the alignment without invariant sites) with node support
981 tested via 2,000 ultrafast phylogenetic bootstraps (Hoang et al. 2018) (BS).
982

983 **Analysis of sample ploidy**

984 The ploidy of each RNA-Seq sample from *M. capitata* and *P. acuta*, as well as the *P. acuta*
985 reference genome and the *P. damicornis* RNA-Seq libraries (used as the outgroup in fig. 3B) was
986 initially estimated by visually inspecting the distribution of alleles with a given proportion of read
987 support (i.e., the allelic depth of the variants in each sample). The alleles used for ploidy
988 estimation were extracted from the variants produced by GenotypeGVCFs (supplementary fig.
989 S12); variants with a read depth (DP) < 20 were removed from each sample (gatk
990 VariantFiltration --genotype-filter-name "DP_filter" --genotype-filter-expression "DP < 20"). The
991 alleles we extracted at this stage of the workflow, and not after final filtered step, because we
992 wanted to capture as many sites as possible in each sample, and the filtering that was applied in
993 the main workflow, which also filtered out variants based on the overall quality of a site across all
994 of the sample from a species, was too stringent given that we are not interested in comparing
995 sites across samples.

996

997 The results of the initial ploidy estimates were confirmed by nQuire (Weiss et al. 2018) (retrieved
998 on 7/7/2021 from https://github.com/clwgg/nQuire), which was run using the aligned and
999 processed RNA-Seq reads that were used to call haplotypes (i.e., the processed reads given to
1000 gatk HaplotypeCaller), and the *P. acuta* PacBio HiFi reads aligned against the new *P. acuta*
1001 genome assembly using MiniMap2. Briefly, the aligned BAM files were converted into "BIN" files,
1002 filtering for reads with a minimum mapping quality of 20 and sites with a minimum coverage of 20
1003 ("nQuire create -q 20 -c 20 -x"). Denoised BIN files were created using the "nQuire denoise"
1004 command run on the initial BIN files. The delta Log-Likelihood values for each ploidy model was
1005 calculated by the "nQuire lrdmodel" command for each of the initial and denoised BIN files.

1006

1007 **Assembly of coral holotranscriptomes**

1008 A holotranscriptome was assembled for each coral species and was used to assess the
1009 expression patterns of the symbiont community. Reads from *M. capitata* and *P. acuta* RNA-Seq

1010 samples were aligned against the published *M. capitata* (Shumaker et al. 2019) and *P. acuta*
1011 (Vidal-Dupiol et al. 2020) genome assemblies using HISAT2 (Kim et al. 2019) (v2.2.1; "-q --
1012 phred33 --dta --rf --very-sensitive"). Samtools (Danecek et al. 2021) (v1.11; "view -f 12") was
1013 used to extract reads that were unmapped and that had mates that were unmapped; the resulting
1014 pairs of unaligned reads were separated based on orientation (i.e., as either 'up' or 'down' stream
1015 reads). The unaligned reads from all samples from the same coral species were assembled using
1016 Trinity (Grabherr et al. 2011) (v2.11.0; "--SS_lib_type RF") into a "holotranscriptome". Expression
1017 of the assembled transcripts in each holotranscriptome was quantified by Salmon (Patro et al.
1018 2017) (v1.10; --validateMappings --seqBias --gcBias --libType ISR) using the unaligned RNA-Seq
1019 reads from each sample.
1020
1021 The taxonomic provenance of the transcripts in the assembled holotranscriptomes was assessed
1022 by comparison against two databases. The first is comprised of transcripts from a diverse
1023 collection of Symbiodiniaceae datasets (supplementary table S13) (Aranda et al. 2016; Chen et
1024 al. 2020; González-Pech et al. 2021; Johnson et al. 2019; Keeling et al. 2014; Liu et al. 2018;
1025 Robbins et al. 2019; Shoguchi et al. 2018; Shoguchi et al. 2013), which includes sequences from
1026 assembled transcriptomes as well as predicted genes from the published genome assemblies.
1027 The second database was comprised of species-specific coral sequences. The *P. acuta* species-
1028 specific database was comprised of CDS from the published *P. acuta* genome "experimental"
1029 predicted gene set (Vidal-Dupiol et al. 2020), the predicted protein-coding genes from the *P.*
1030 *damicornis* genome (Cunning et al. 2018), and transcripts assembled by StringTie2 (Kovaka et al.
1031 2019) (v2.0.6; --rf) using the reads from the RNA-Seq samples generated in this study aligned by
1032 HISAT2 (v2.2.1; -q --phred33 --no-unal --dta --rf --very-sensitive) against the new *P. acuta*
1033 genome assembly. The *M. capitata* species-specific database was comprised of CDS from the
1034 predicted protein-coding genes in the published *M. capitata* genome assembly (Shumaker et al.
1035 2019) and transcripts assembled by StringTie2 and HISAT2 using the same approach as for *P.*
1036 *acuta*. Transcripts were compared using BLASTn (v2.10.1; default parameters) separately

against the Symbiodiniaceae and the appropriate coral databases. The top hit for each transcript from each database was extracted, these top hits were then combined and filtered so that only best top hit (with the highest bitscore) was retained for each transcript. Finally, top hits were removed if they had an *e*-value $<1x10^{-5}$. Transcript that have a best top hit to a sequence from the Symbiodiniaceae database are considered to be putatively symbiont derived, transcripts that have best top hits to a sequence from the coral species-specific database are considered to be putatively coral derived, and transcripts without hits to either database are considered to be of unknown (non-coral) taxonomic provenance.

Data and materials availability

All data, code, and materials used in the analysis are available at https://osf.io/gvhkx/, https://github.com/hputnam/Coral_Stress_Phenome, and from NCBI's SRA repository (RNA-Seq data are available from PRJNA731596; genome sequencing data from PRJNA761443; algal ITS2 sequencing data from PRJNA761780). Genome assemblies and predicted genes are available from http://cyanophora.rutgers.edu/montipora/ and http://cyanophora.rutgers.edu/Pocillopora_acuta/.

Acknowledgments

We acknowledge support of the research and administrative staff at the Hawaii Institute of Marine Biology where this work was done. This work is/was supported by the USDA National Institute of Food and Agriculture, Hatch Formula project accession number 1017848, the National Science Foundation grant NSF-OCE 1756616, the Catalyst Science Fund grant 2020-008, the National Institute of Food and Agriculture and United States Department of Agriculture Hatch grant NJ01180, and the National Aeronautics and Space Administration grant 80NSSC19K0462, which were awarded to DB. This work was also supported by the National Science Foundation grant NSF-OCE 1756623 awarded to HMP.

1064 **References**

1065 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment
1066 search tool. J Mol Biol 215(3):403-10.

1067 Aranda M, Li Y, Liew YJ, Baumgarten S, Simakov O, Wilson MC, Piel J, Ashoor H,
1068 Bougouffa S, Bajic VB et al. . 2016. Genomes of coral dinoflagellate symbionts
1069 highlight evolutionary adaptations conducive to a symbiotic lifestyle. Sci Rep
1070 6:39734.

1071 Bahr KD, Jokiel PL, Rodgers KS. 2015. The 2014 coral bleaching and freshwater flood
1072 events in Kane'ohe Bay, Hawai'i. PeerJ 3:e1136.

1073 Bahr KD, Jokiel PL, Rodgers KuS. 2016. Relative sensitivity of five Hawaiian coral
1074 species to high temperature under high-pCO2 conditions. Coral Reefs 35(2):729-
1075 738.

1076 Bahr KD, Tran T, Jury CP, Toonen RJ. 2020. Abundance, size, and survival of recruits of
1077 the reef coral *Pocillopora acuta* under ocean warming and acidification. PLoS
1078 One 15(2):e0228168.

1079 Barnett DW, Garrison EK, Quinlan AR, Stromberg MP, Marth GT. 2011. BamTools: A
1080 C++ API and toolkit for analyzing and managing BAM files. Bioinformatics
1081 27(12):1691-2.

1082 Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting Linear Mixed-Effects Models
1083 Using lme4. 2015 67(1):48.

1084 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina
1085 sequence data. Bioinformatics 30(15):2114-20.

1086 Bruna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2: Automatic
1087 eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported
1088 by a protein database. NAR Genom Bioinform 3(1):lqaa108.

1089 Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL.
1090 2009. BLAST+: Architecture and applications. BMC Bioinformatics 10:421.

1091 Cant J, Salguero-Gomez R, Kim SW, Sims CA, Sommer B, Brooks M, Malcolm HA,
1092 Pandolfi JM, Beger M. 2021. The projected degradation of subtropical coral
1093 assemblages by recurrent thermal stress. J Anim Ecol 90(1):233-247.

1094 Carson JC, Hoffner L, Conlin L, Parks WT, Fisher RA, Spinner N, Yatsenko SA,
1095 Bonadio J, Surti U. 2018. Diploid/triploid mixoploidy: A consequence of
1096 asymmetric zygotic segregation of parental genomes. Am J Med Genet A
1097 176(12):2720-2732.

1098 Caruso C, de Souza MR, Ruiz-Jones L, Conetta D, Hancock J, Hobbs C, Hobbs C,
1099 Kahkejian V, Kitchen R, Marin C et al. . 2021. Genetic patterns in *Montipora*
1100 *capitata* across an environmental mosaic in Kāne'ohe Bay.
1101 bioRxiv:2021.10.07.463582.

1102 Chen YB, Gonzalez-Pech RA, Stephens TG, Bhattacharya D, Chan CX. 2020. Evidence
1103 that inconsistent gene prediction can mislead analysis of dinoflagellate genomes.
1104 Journal of Phycology 56(1):6-10.

1105 Concepcion GT, Baums IB, Toonen RJ. 2014. Regional population structure of
1106 *Montipora capitata* across the Hawaiian Archipelago. Bulletin of Marine Science
1107 90(1):257-275.

1108 Cunning R, Bay RA, Gillette P, Baker AC, Traylor-Knowles N. 2018. Comparative
1109 analysis of the *Pocillopora damicornis* genome highlights role of immune system
1110 in coral evolution. Sci Rep 8(1):16134.

1111 Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane
1112 T, McCarthy SA, Davies RM et al. . 2021. Twelve years of SAMtools and
1113 BCFtools. Gigascience 10(2).

1114 Dickson A, Chris S, Christian JR. 2007. Guide to Best Practices for Ocean CO2
1115 Measurements. North Pacific Marine Science Organization.

1116 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M,
1117 Gingeras TR. 2013. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics
1118 29(1):15-21.

1119 Edmunds P, Gates R, Gleason D. 2003. The tissue composition of *Montastraea franksi*
1120 during a natural bleaching event in the Florida Keys. Coral Reefs 22:54-62.

1121 Ewels P, Magnusson M, Lundin S, Kaller M. 2016. MultiQC: Summarize analysis results
1122 for multiple tools and samples in a single report. Bioinformatics 32(19):3047-8.

1123 Fischer K, Kreyling J, Beaulieu M, Beil I, Bog M, Bonte D, Holm S, Knoblauch S, Koch
1124 D, Muffler L et al. . 2021. Species-specific effects of thermal stress on the
1125 expression of genetic variation across a diverse group of plant and animal taxa
1126 under experimental conditions. Heredity (Edinb) 126(1):23-37.

1127 Flot JF, Magalon H, Cruaud C, Couloux A, Tillier S. 2008. Patterns of genetic structure
1128 among Hawaiian corals of the genus *Pocillopora* yield clusters of individuals that
1129 are compatible with morphology. C R Biol 331(3):239-47.

1130 Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020.
1131 RepeatModeler2 for automated genomic discovery of transposable element
1132 families. Proc Natl Acad Sci U S A 117(17):9451-9457.

1133 Gattuso J, Epitalon J, Lavigne H, Orr J, Gentili B, Hagens M, Hofmann A, Mueller J-D,
1134 Proye A, Rae J et al. . 2015. SEACARB: Seawater Carbonate Chemistry.

1135 Gibbin EM, Putnam HM, Gates RD, Nitschke MR, Davy SK. 2015. Species-specific
1136 differences in thermal tolerance may define susceptibility to intracellular acidosis
1137 in reef corals. Marine Biology 162(3):717-723.

1138 González-Pech RA, Stephens TG, Chen Y, Mohamed AR, Cheng Y, Shah S, Dougan
1139 KE, Fortuin MDA, Lagorce R, Burt DW et al. . 2021. Comparison of 15

1140 dinoflagellate genomes reveals extensive sequence and structural divergence in
1141 family Symbiodiniaceae and genus *Symbiodinium*. BMC Biology 19(1):73.

1142 Gorospe KD, Karl SA. 2011. Small-scale spatial analysis of *in situ* sea temperature
1143 throughout a single coral patch reef. Journal of Marine Biology 2011:719580.

1144 Gorospe KD, Karl SA. 2013. Genetic relatedness does not retain spatial pattern across
1145 multiple spatial scales: dispersal and colonization in the coral, *Pocillopora*
1146 *damicornis*. Mol Ecol 22(14):3721-36.

1147 Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan
1148 L, Raychowdhury R, Zeng Q et al. . 2011. Full-length transcriptome assembly
1149 from RNA-Seq data without a reference genome. Nat Biotechnol 29(7):644-52.

1150 Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB,
1151 Eccles D, Li B, Lieber M et al. . 2013. *De novo* transcript sequence reconstruction
1152 from RNA-seq using the Trinity platform for reference generation and analysis.
1153 Nat Protoc 8(8):1494-512.

1154 Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2:
1155 Improving the ultrafast bootstrap approximation. Mol Biol Evol 35(2):518-522.

1156 Hoegh-Guldberg O, Mumby PJ, Hooten AJ, Steneck RS, Greenfield P, Gomez E, Harvell
1157 CD, Sale PF, Edwards AJ, Caldeira K et al. . 2007. Coral reefs under rapid
1158 climate change and ocean acidification. Science 318(5857):1737-42.

1159 Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. 2016. BRAKER1:
1160 Unsupervised RNA-Seq-based genome annotation with GeneMark-ET and
1161 AUGUSTUS. Bioinformatics 32(5):767-9.

1162 Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. 2019. Whole-Genome Annotation with
1163 BRAKER. Methods Mol Biol 1962:65-95.

1164 Huang S, Kang M, Xu A. 2017. HaploMerger2: Rebuilding both haploid sub-assemblies
1165 from high-heterozygosity diploid genome assembly. Bioinformatics 33(16):2577-
1166 2579.

1167 Hughes TP, Kerry JT, Alvarez-Noriega M, Alvarez-Romero JG, Anderson KD, Baird
1168 AH, Babcock RC, Beger M, Bellwood DR, Berkelmans R et al. . 2017. Global
1169 warming and recurrent mass bleaching of corals. Nature 543(7645):373-377.

1170 Hume BCC, Smith EG, Ziegler M, Warrington HJM, Burt JA, LaJeunesse TC,
1171 Wiedenmann J, Voolstra CR. 2019. SymPortal: A novel analytical framework and
1172 platform for coral algal symbiont next-generation sequencing *ITS2* profiling. Mol
1173 Ecol Resour 19(4):1063-1080.

1174 Husband BC. 2004. The role of triploid hybrids in the evolutionary dynamics of mixed-
1175 ploidy populations. Biological Journal of the Linnean Society 82(4):537-546.

1176 Iliana BB, Colin RH, Michael EH. 2005. Mendelian microsatellite loci for the Caribbean
1177 coral *Acropora palmata*. Marine Ecology Progress Series 288:115-127.

1178 Johnson LK, Alexander H, Brown CT. 2019. Re-assembly, quality evaluation, and
1179 annotation of 678 microbial eukaryotic reference transcriptomes. Gigascience
1180 8(4).

1181 Jokiel PL. 1985. Lunar periodicity of planula release in the reef coral *Pocillopora*
1182 *damicornis* in relation to various environmental factors. Proceedings of the Fifth
1183 International Coral Reef Congress, Tahiti 4:307-316.

1184 Jokiel PL, Brown EK. 2004. Global warming, regional trends and inshore environmental
1185 conditions influence coral bleaching in Hawaii. Global Change Biology
1186 10(10):1627-1641.

1187 Kang JH, Lim HJ, Kang HS, Lee JM, Baby S, Kim JJ. 2013. Development of genetic
1188 markers for triploid verification of the Pacific oyster, *Crassostrea gigas*. Asian-
1189 Australas J Anim Sci 26(7):916-20.

1190 Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, Amaral-Zettler LA, Armbrust EV,
1191 Archibald JM, Bharti AK, Bell CJ et al. . 2014. The Marine Microbial Eukaryote
1192 Transcriptome Sequencing Project (MMETSP): Illuminating the functional
1193 diversity of eukaryotic life in the oceans through transcriptome sequencing. Plos
1194 Biology 12(6).

1195 Kenyon JC. 1997. Models of reticulate evolution in the coral genus *Acropora* based on
1196 chromosome numbers: Parallels with plants. Evolution 51(3):756-767.

1197 Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome
1198 alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol
1199 37(8):907-915.

1200 Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu:
1201 Scalable and accurate long-read assembly via adaptive *k*-mer weighting and
1202 repeat separation. Genome Res 27(5):722-736.

1203 Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, Pertea M. 2019.
1204 Transcriptome assembly from long-read RNA-seq alignments with StringTie2.
1205 Genome Biol 20(1):278.

1206 LaJeunesse TC, Parkinson JE, Gabrielson PW, Jeong HJ, Reimer JD, Voolstra CR,
1207 Santos SR. 2018. Systematic revision of Symbiodiniaceae highlights the antiquity
1208 and diversity of coral endosymbionts. Curr Biol 28(16):2570-2580 e6.

1209 Lajeunesse TC, Trench RK. 2000. Biogeography of two species of *Symbiodinium*
1210 (Freudenthal) inhabiting the intertidal sea anemone *Anthopleura elegantissima*
1211 (Brandt). Biol Bull 199(2):126-34.

1212 Li H. 2018. Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics
1213 34(18):3094-3100.

1214 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G,
1215 Durbin R, Genome Project Data Processing S. 2009. The Sequence
1216 Alignment/Map format and SAMtools. Bioinformatics 25(16):2078-9.

1217 Liu H, Stephens TG, Gonzalez-Pech RA, Beltran VH, Lapeyre B, Bongaerts P, Cooke I,
1218 Aranda M, Bourne DG, Foret S et al. . 2018. *Symbiodinium* genomes reveal
1219 adaptive evolution of functions related to coral-dinoflagellate symbiosis.
1220 Commun Biol 1:95.

1221 Lomsadze A, Burns PD, Borodovsky M. 2014. Integration of mapped RNA-Seq reads
1222 into automatic training of eukaryotic gene finding algorithm. Nucleic Acids Res
1223 42(15):e119.

1224 Lowe RJ, Falter JL, Monismith SG, Atkinson MJ. 2009. A numerical study of circulation
1225 in a coastal reef-lagoon system. Journal of Geophysical Research: Oceans
1226 114(C6).

1227 Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO Update:
1228 Novel and streamlined workflows along with broader and deeper phylogenetic
1229 coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Molecular
1230 Biology and Evolution.

1231 Marcais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting
1232 of occurrences of *k*-mers. Bioinformatics 27(6):764-70.

1233 Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing
1234 reads. EMBnet.journal 17(1):3.

1235 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K,
1236 Altshuler D, Gabriel S, Daly M et al. . 2010. The Genome Analysis Toolkit: a
1237 MapReduce framework for analyzing next-generation DNA sequencing data.
1238 Genome Res 20(9):1297-303.

1239 Nakajima Y, Chuang PS, Ueda N, Mitarai S. 2018. First evidence of asexual recruitment
1240 of *Pocillopora acuta* in Okinawa Island using genotypic identification. PeerJ
1241 6:e5915.

1242 Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A fast and
1243 effective stochastic algorithm for estimating maximum-likelihood phylogenies.
1244 Mol Biol Evol 32(1):268-74.

1245 Oury N, Gelin P, Magalon H. 2020. Together stronger: Intracolonial genetic variability
1246 occurrence in *Pocillopora* corals suggests potential benefits. Ecol Evol
1247 10(12):5208-5218.

1248 Pachauri RK, Allen MR, Barros VR, Broome J, Cramer W, Christ R, Church JA, Clarke
1249 L, Dahe Q, Dasgupta P et al. . 2014. Climate Change 2014: Synthesis Report.
1250 Contribution of Working Groups I, II and III to the Fifth Assessment Report of
1251 the Intergovernmental Panel on Climate Change. Geneva, Switzerland: IPCC.

1252 Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and
1253 bias-aware quantification of transcript expression. Nat Methods 14(4):417-419.

1254 Ranallo-Benavidez TR, Jaron KS, Schatz MC. 2020. GenomeScope 2.0 and Smudgeplot
1255 for reference-free profiling of polyploid genomes. Nat Commun 11(1):1432.

1256 Richmond R, Jokiel P. 1984. Lunar periodicity in larva release in the reef coral
1257 *Pocillopora damicornis* at Enewetak and Hawaii. Bulletin of Marine Science
1258 34:280-287.

1259 Rinkevich B, Shaish L, Douek J, Ben-Shlomo R. 2016. Venturing in coral larval
1260 chimerism: a compact functional domain with fostered genotypic diversity. Sci
1261 Rep 6:19493.

1262 Roach MJ, Schmidt SA, Borneman AR. 2018. Purge Haplotigs: allelic contig
1263 reassignment for third-gen diploid genome assemblies. BMC Bioinformatics
1264 19(1):460.

1265 Robbins SJ, Singleton CM, Chan CX, Messer LF, Geers AU, Ying H, Baker A, Bell SC,
1266 Morrow KM, Ragan MA et al. . 2019. A genomic view of the reef-building coral
1267 *Porites lutea* and its microbial symbionts. Nature Microbiology 4(12):2090-2100.

1268 Rosenbusch BE. 2008. Mechanisms giving rise to triploid zygotes during assisted
1269 reproduction. Fertil Steril 90(1):49-55.

1270 Schmidt-Roach S, Miller KJ, Lundgren P, Andreakis N. 2014. With eyes wide open: A
1271 revision of species within and closely related to the *Pocillopora damicornis*
1272 species complex (Scleractinia; Pocilloporidae) using morphology and genetics.
1273 Zoological Journal of the Linnean Society 170(1):1-33.

1274 Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of
1275 image analysis. Nat Methods 9(7):671-5.

1276 Schweinsberg M, Weiss LC, Striewski S, Tollrian R, Lampert KP. 2015. More than one
1277 genotype: How common is intracolonial genetic variability in scleractinian
1278 corals? Mol Ecol 24(11):2673-85.

1279 Shoguchi E, Beedessee G, Tada I, Hisata K, Kawashima T, Takeuchi T, Arakaki N, Fujie
1280 M, Koyanagi R, Roy MC et al. . 2018. Two divergent *Symbiodinium* genomes
1281 reveal conservation of a gene cluster for sunscreen biosynthesis and recently lost
1282 genes. BMC Genomics 19(1):458.

1283 Shoguchi E, Shinzato C, Kawashima T, Gyoja F, Mungpakdee S, Koyanagi R, Takeuchi
1284 T, Hisata K, Tanaka M, Fujiwara M et al. . 2013. Draft assembly of the
1285 *Symbiodinium minutum* nuclear genome reveals dinoflagellate gene structure.
1286 Curr Biol 23(15):1399-408.

1287 Shumaker A, Putnam HM, Qiu H, Price DC, Zelzion E, Harel A, Wagner NE, Gates RD,
1288 Yoon HS, Bhattacharya D. 2019. Genome analysis of the rice coral *Montipora*
1289 *capitata*. Sci Rep 9(1):2571.

1290 Shumate A, Salzberg SL. 2020. Liftoff: Accurate mapping of gene annotations.
1291 Bioinformatics.

1292 Smit A, Hubley R, Green P. 2013-2015. RepeatMasker Open-4.0.

1293 Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntenically
1294 mapped cDNA alignments to improve *de novo* gene finding. Bioinformatics
1295 24(5):637-44.

1296 Stanke M, Schoffmann O, Morgenstern B, Waack S. 2006. Gene prediction in eukaryotes
1297 with a generalized hidden Markov model that uses hints from external sources.
1298 BMC Bioinformatics 7:62.

1299 Therneau MT. 2020. A Package for Survival Analysis in R.

1300 van Oppen M, Lough J. 2018. Synthesis: Coral Bleaching: Patterns, Processes, Causes
1301 and Consequences. p. 343-348.

1302 Veron JEN. 2000. Corals of the world. Vol 1-3. Australian Institute of Marine Science,
1303 Townsville, Qld, Australia.

1304 Vidal-Dupiol J, Chaparro C, Pratlong M, Pontarotti P, Grunau C, Mitta G. 2020.
1305 Sequencing, *de novo* assembly and annotation of the genome of the scleractinian
1306 coral, *Pocillopora acuta*. bioRxiv:698688.

1307 Vidal-Dupiol J, Dheilly NM, Rondon R, Grunau C, Cosseau C, Smith KM, Freitag M,
1308 Adjeroud M, Mitta G. 2014. Thermal stress triggers broad *Pocillopora*
1309 *damicornis* transcriptomic remodeling, while *Vibrio coralliilyticus* infection
1310 induces a more targeted immuno-suppression response. PLoS One 9(9):e107672.

1311 Vidal-Dupiol J, Zoccola D, Tambutte E, Grunau C, Cosseau C, Smith KM, Freitag M,
1312 Dheilly NM, Allemand D, Tambutte S. 2013. Genes related to ion-transport and
1313 energy production are upregulated in response to CO2-driven pH decrease in
1314 corals: new insights from transcriptome analysis. PLoS One 8(3):e58652.

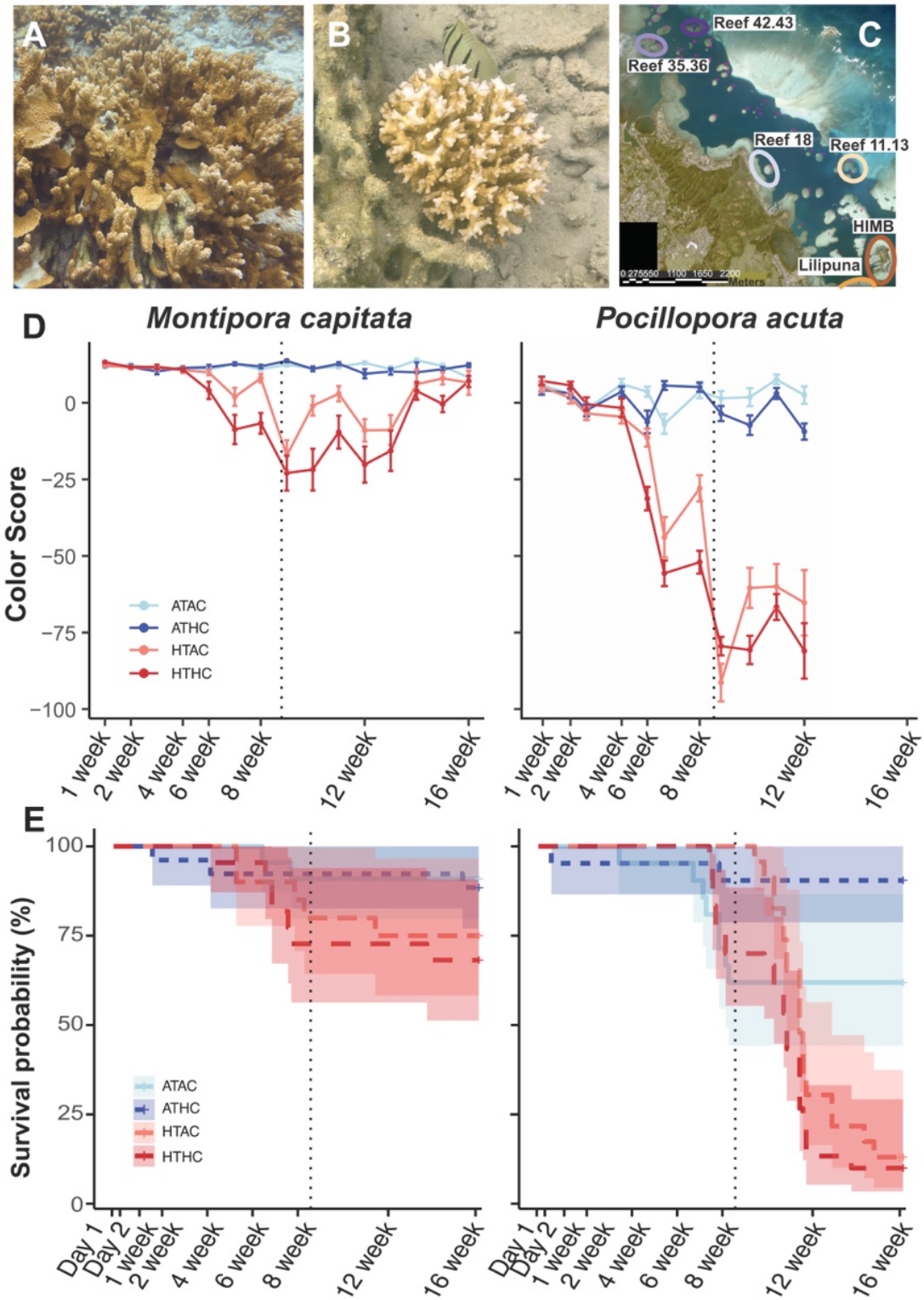
1315 Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng
1316 Q, Wortman J, Young SK et al. . 2014. Pilon: An integrated tool for
1317 comprehensive microbial variant detection and genome assembly improvement.
1318 PLoS One 9(11):e112963.

1319 Weiss CL, Pais M, Cano LM, Kamoun S, Burbano HA. 2018. nQuire: A statistical
1320 framework for ploidy estimation using next generation sequencing. BMC
1321 Bioinformatics 19(1):122.

1322 Willis BL, van Oppen MJH, Miller DJ, Vollmer SV, Ayre DJ. 2006. The role of
1323 hybridization in the evolution of reef corals. Annual Review of Ecology,
1324 Evolution, and Systematics 37(1):489-517.

1325 Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. 2012. A high-performance
1326 computing toolset for relatedness and principal component analysis of SNP data.
1327 Bioinformatics 28(24):3326-8.
1328

1329 **Figures and Tables**



1330

1331 **Figure 1. Study corals in Kāneʻohe Bay, Oʻahu, Hawaiʻi and physiological results.**
1332 Colonies of *Montipora capitata* (*A*), the rice coral, and *Pocillopora acuta* (*B*), the lace coral.
1333 Images taken by D. Bhattacharya. (*C*) Aerial image of Kāneʻohe Bay with the six reefs where the
1334 coral samples were collected highlighted using colored circles and labels. The color of the circles
1335 surrounding the collection sites corresponds to the colors used in figures 3 and 4 and in the
1336 supplementary figures. The legend for distance in meters is shown at the bottom of the image.
1337 (*D*) Color change of the corals indicating visual bleaching that was quantified using color
1338 standardized photographs. (*E*) Probability of survivorship under each treatment during the 16-
1339 week study period. Treatment conditions are represented in (*D*) and (*E*) as light blue (Ambient
1340 Temperature Ambient pCO_2; ATAC), dark blue (Ambient Temperature High pCO_2; ATHC), light
1341 red (High Temperature Ambient pCO_2; HTAC), and dark red (High Temperature High pCO_2;
1342 HTHC) lines. The dashed vertical lines in (*D*) and (*E*) indicate the transition from treatment to
1343 recovery periods.

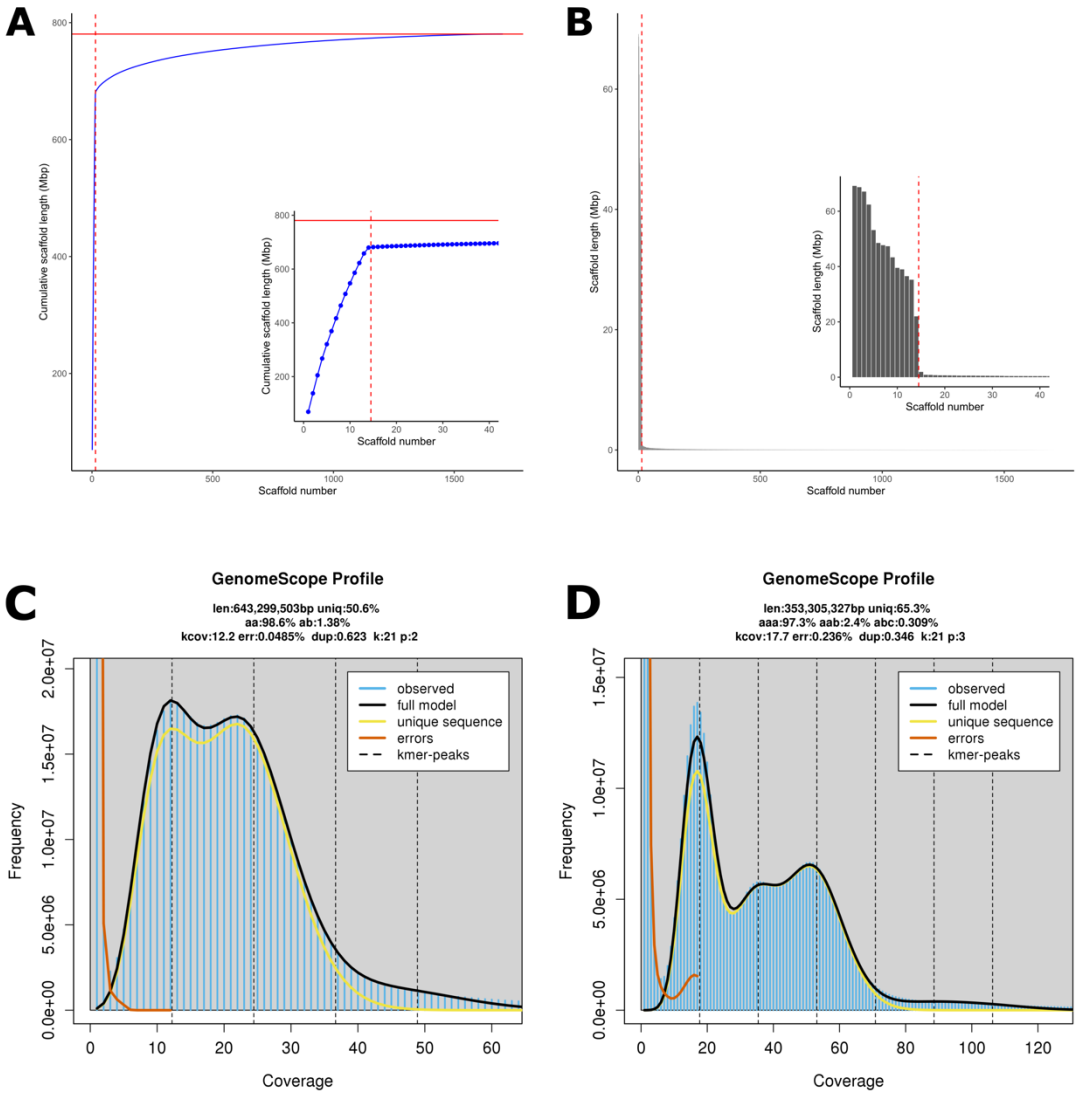
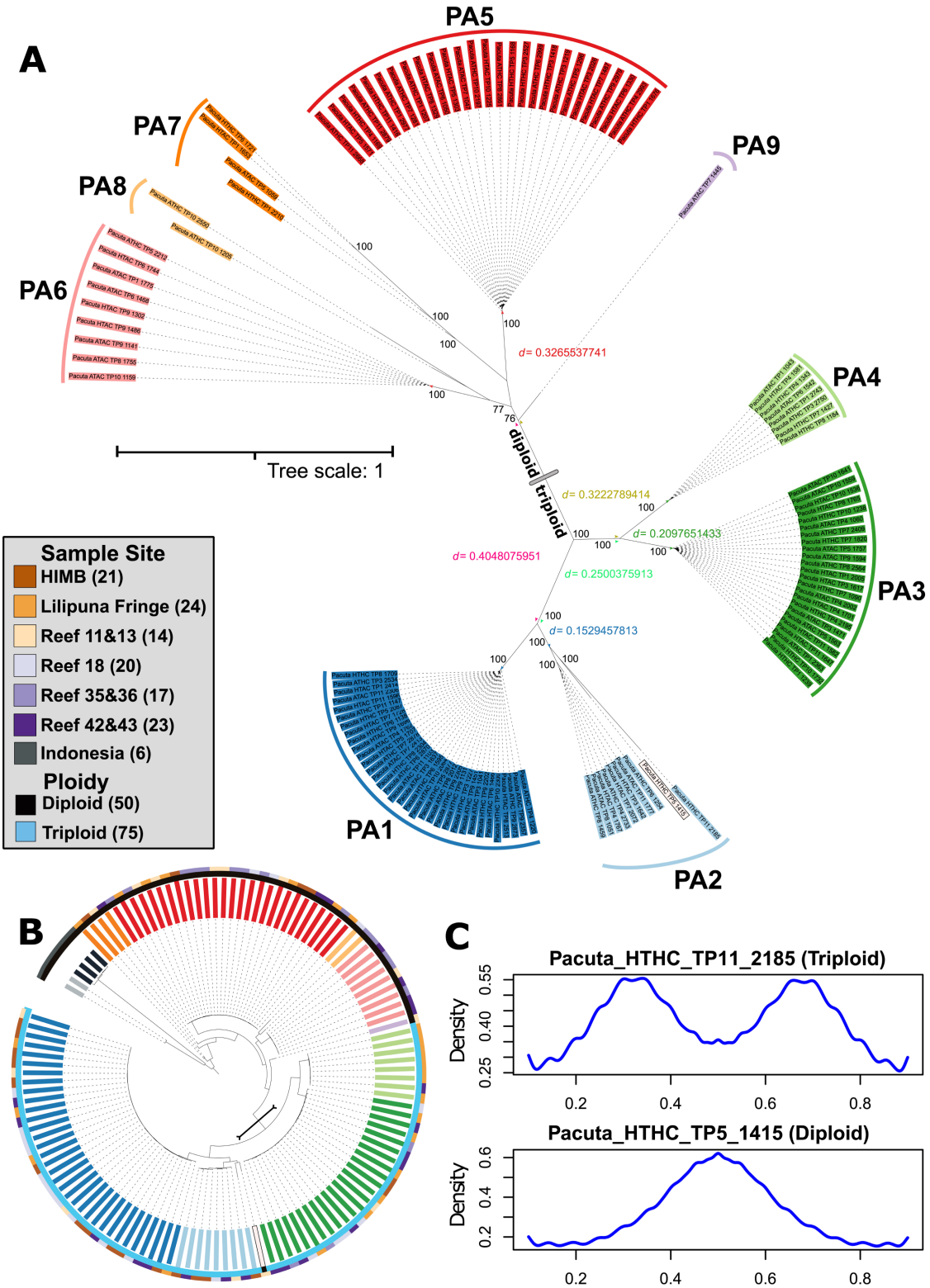


Figure 2. High quality genome assemblies

Cumulative (*A*) and individual (*B*) length of scaffolds in the new *M. capitata* genome assembly. Scaffolds were sorted by length in descending order; each point along the x-axis of (*A*) and (*B*) represents a scaffold, with the longest scaffold being the first and the shortest being the last on the axis of each plot. In (*A*) and (*B*) a zoomed-in section of the larger plot is shown on the right highlights the 40 largest scaffolds; a horizontal red line in (*A*) shows the total assembled bases in the new genome and a vertical dashed line in (*A*) and (*B*) is positioned after the 14th largest scaffold. GenomeScape2 linear *k*-mer distributions of *M. capitata* (*C*) and *P. acuta* (*D*) with theoretical diploid (*C*) and triploid (*D*) models shown by the black lines.

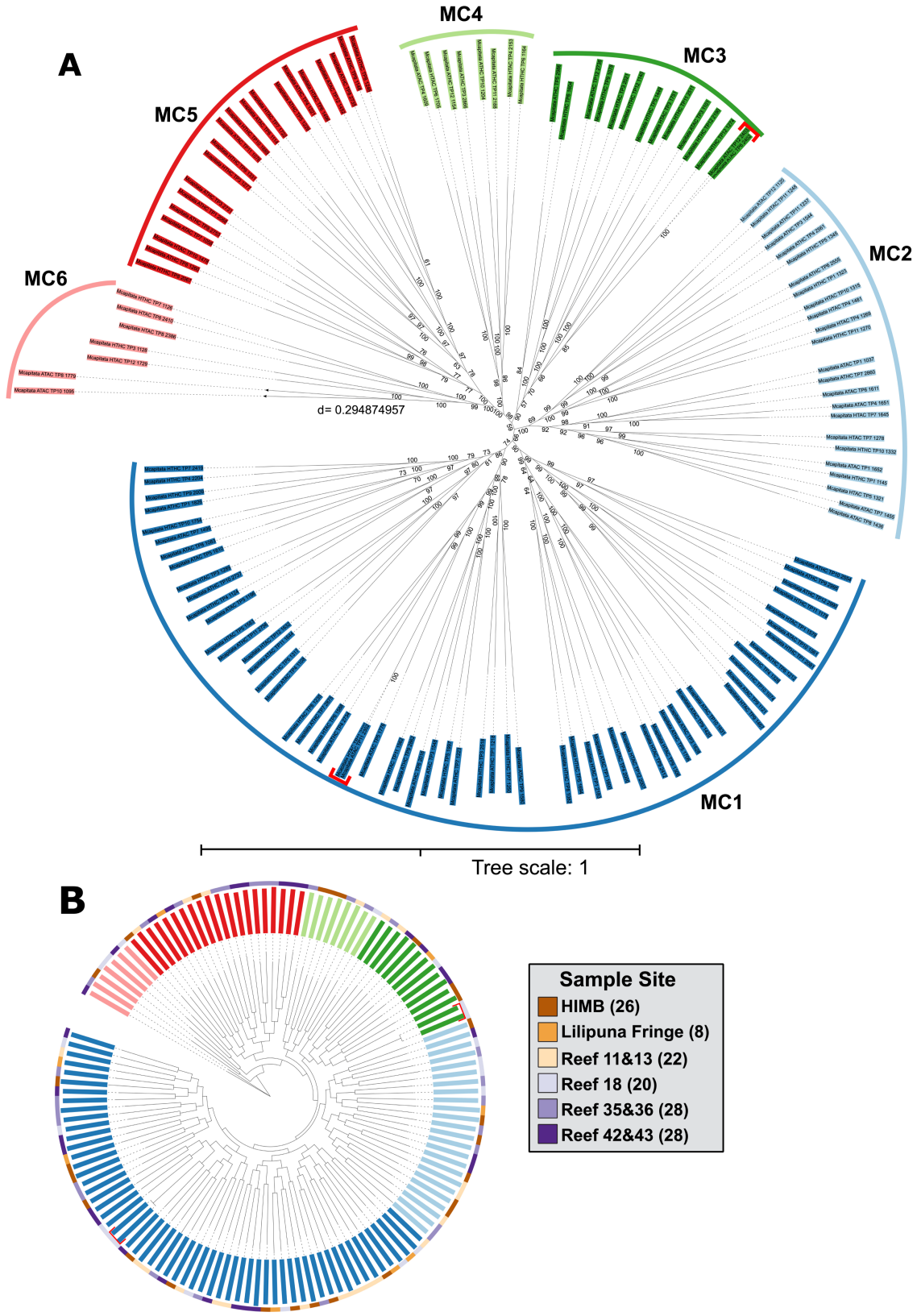

A
PA5
PA7
PA8
PA6
PA9
PA4
PA3
PA1
PA2
diploid
triploid
d= 0.3265537741
d= 0.3222789414
d= 0.2097651433
d= 0.4048075951
d= 0.2500375913
d= 0.1529457813
Tree scale: 1
Sample Site
HIMB (21)
Lilipuna Fringe (24)
Reef 11&13 (14)
Reef 18 (20)
Reef 35&36 (17)
Reef 42&43 (23)
Indonesia (6)
Ploidy
Diploid (50)
Triploid (75)
B
C
Pacuta_HTHC_TP11_2185 (Triploid)
Density
Pacuta_HTHC_TP5_1415 (Diploid)
Density

1354

1355

1356 **Figure 3. Phylogenetic analysis of *P. acuta* samples.**

1357 (*A*) An unrooted phylogenetic tree constructed from an alignment (13,948 parsimony-informative
1358 and 1,835 singleton sites; 13,948 total) built using the SNPs identified in the *P. acuta* RNA-Seq
1359 data. The samples have been colored and labeled according to the putative clades identified in
1360 the tree. Support values for the internal nodes are based on 2000 ultrafast bootstrap
1361 approximations. The distance (*d*) between major clades in the tree are shown along with colored
1362 arrows indicating which nodes were compared. The putative split between ancestral diploid
1363 lineages and derived, triploid lineages is shown in the tree. (*B*) *P. acuta* phylogenetic tree
1364 constructed from an alignment (1,856 parsimony-informative, 74 singleton sites, and 71 constant
1365 sites; 2,640 total) built using SNPs identified in all *P. acuta* and six *P. damicornis* samples (Vidal-
1366 Dupiol et al. 2014; Vidal-Dupiol et al. 2013). This tree is rooted using the six *P. damicornis*
1367 samples. The putative ploidy and sample collection site information for each sample is annotated
1368 using the colored inner and outer rings, respectively. A legend explaining each color used in the
1369 figure is presented in the middle left of the image with the number of samples in each category
1370 shown in brackets. The putative split between ancestral diploid lineages and derived, triploid
1371 lineages is shown in the tree. (*C*) Density plots for samples Pacuta_HTHC_TP11_2185 (putative
1372 triploid) and Pacuta_HTHC_TP5_1415 (putative diploid) showing the proportion of reads at a
1373 given site that support each SNP allele.


A
MC1
MC2
MC3
MC4
MC5
MC6
d= 0.294874957
Tree scale: 1
B
Sample Site
HIMB (26)
Lilipuna Fringe (8)
Reef 11&13 (22)
Reef 18 (20)
Reef 35&36 (28)
Reef 42&43 (28)

1374

1375 **Figure 4. Phylogenetic analysis of *M. capitata* samples.**

1376 (*A*) An unrooted phylogenetic tree constructed from an alignment (25,541 parsimony-informative
1377 and 1,165 singleton sites; 26,706 total) built using the SNPs identified in the *M. capitata* RNA-Seq
1378 data. The samples have been colored and labeled according to the putative clades identified in
1379 the tree. Support values for the internal nodes are based on 2000 ultrafast bootstrap
1380 approximations. The distance (*d*) between some samples in the tree are highlighted to show the
1381 level of divergence between the *M. capitata* samples. Samples that are putatively from the same
1382 genotype are indicated by the red brackets. (*B*) Phylogenetic tree shown in (*A*) with sample
1383 collection site information indicated by the colored ring. A legend explaining each color used in
1384 the figure is presented in the bottom right of the image with the number of samples in each
1385 category shown in brackets.

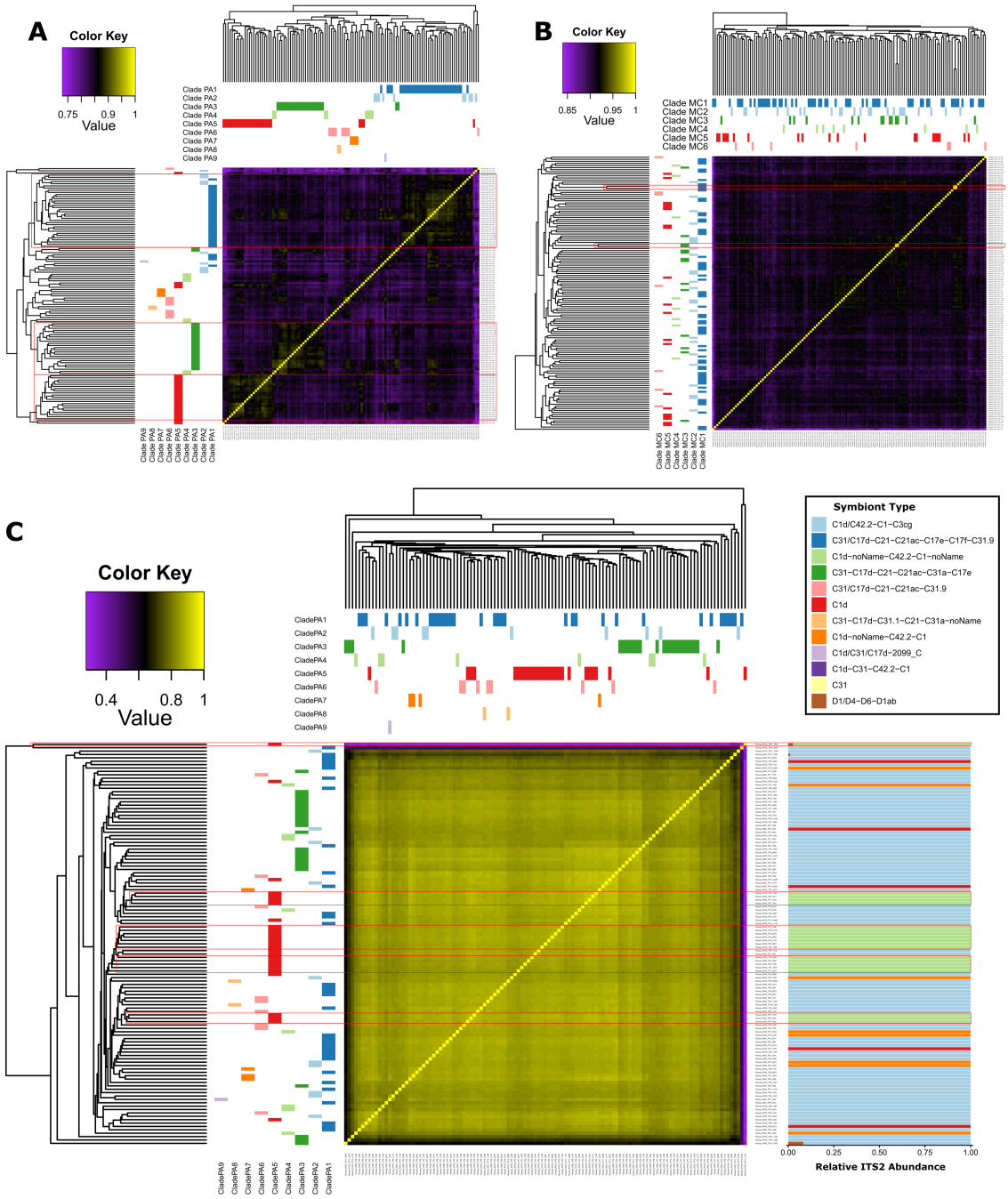


Figure 5. Analysis of gene expression profiles from host and symbiont-derived transcripts. Heatmaps showing the (*A*) *P. acuta* gene expression correlation results annotated with the putative clades identified in figure 3A, and (*B*) the *M. capitata* gene expression correlation results annotated with the putative clades identified in figure 4A. Groups of samples in (*A*) that cluster by genotype and the two pairs of samples in (*B*) that are putatively derived from the same genotypes

are encapsulated by red squares. (*C*) Symbiont-derived holotranscriptome expression profiles of *P. acuta* with host clades (fig. 3A) and algal ITS2 (supplementary fig. S9) data shown along the sides of the heatmap. A legend explaining each color used for the ITS2 data in the bar plot is presented in the top right of the panel. Samples with ITS2 profiles that contain *Cladocopium* "C1d−noName−C42.2−C1−noName" are encapsulated by red squares. The order of the samples and the dendrograms plotted along each axis were determined by hierarchical clustering of the distances calculated between the samples in each heatmap.