

1 **Reduced representation sequencing for symbiotic anthozoans: are reference**  
2 **genomes necessary to eliminate endosymbiont contamination and make**  
3 **robust phylogeographic inference?**

4

5 Benjamin M. Titus<sup>1,2,\*</sup> & Marymegan Daly<sup>1</sup>

6

7 <sup>1</sup>Department of Evolution, Ecology, and Organismal Biology, The Ohio State University,

8 Columbus, Ohio, 43212, USA

9

10 <sup>2</sup>Division of Invertebrate Zoology, American Museum of Natural History, New York, NY

11 10024, USA

12

13 \*Corresponding author: [bentitus3@gmail.com](mailto:bentitus3@gmail.com); [btitus@amnh.org](mailto:btitus@amnh.org)

14

15 Keywords: Cnidarians, phylogeography, allele frequency spectrum, coral reefs, *Symbiodinium*,

16 Symbiodiniaceae

17

18

19

20

21

22

23 **Abstract**

24 Anthozoan cnidarians form the backbone of coral reefs. Their success relies on endosymbiosis  
25 with photosynthetic dinoflagellates in the family Symbiodiniaceae. Photosymbionts represent a  
26 hurdle for researchers using population genomic techniques to study these highly imperiled and  
27 ecologically critical species because sequencing datasets harbor unknown mixtures of anthozoan  
28 and photosymbiont loci. Here we use range-wide sampling and a double-digest restriction-site  
29 associated DNA sequencing (ddRADseq) of the sea anemone *Bartholomea annulata* to explore  
30 how symbiont loci impact the interpretation of phylogeographic patterns and population genetic  
31 parameters. We use the genome of the closely related *Exaiptasia diaphana* (previously *Aiptasia*  
32 *pallida*) to create an anthozoan-only dataset from a genomic dataset containing both *B. annulata*  
33 and its symbiodiniacean symbionts and then compare this to the raw, holobiont dataset. For each,  
34 we investigate spatial patterns of genetic diversity and use coalescent model-based approaches to  
35 estimate demographic history and population parameters. The Florida Straits are the only  
36 phylogeographic break we recover for *B. annulata*, with divergence estimated during the last  
37 glacial maximum. Because *B. annulata* hosts multiple members of Symbiodiniaceae, we  
38 hypothesize that, under moderate missing data thresholds, *de novo* clustering algorithms that  
39 identify orthologs across datasets will have difficulty identifying shared non-coding loci from the  
40 photosymbionts. We infer that, for anthozoans hosting diverse members of Symbiodiniaceae,  
41 clustering algorithms act as *de facto* filters of symbiont loci. Thus, while at least some  
42 photosymbiont loci remain, these are swamped by orders of magnitude greater numbers of  
43 anthozoan loci and thus represent genetic “noise,” rather than contributing genetic signal.

44

45

## 46 **1. Introduction**

47           The study of the distribution of genetic diversity across broad geographic space (i.e.  
48 phylogeography) can shed light on the historical and contemporary processes responsible for the  
49 generation and maintenance of biodiversity within species and ecosystems (Arbogast, 2001;  
50 Avise, 2009; Avise, Bowen, & Ayala, 2016; Knowles, 2009). Phylogeographic surveys  
51 demarcate barriers to dispersal, routinely recover cryptic species, and with increasing dataset  
52 sizes and statistical approaches, can estimate important demographic parameters such as  
53 effective population size, divergence time, migration rates, and historical changes in population  
54 size (e.g. Avise et al., 2016; Carstens, Pelletier, Reid, & Satler, 2013; Knowles, 2009; Pante et  
55 al., 2015; Pelletier & Carstens, 2014; Smith et al., 2017). To that end, high-throughput  
56 sequencing, which can generate thousands of unlinked single nucleotide polymorphisms (SNPs)  
57 across the genome, has been particularly powerful, allowing for greater statistical and  
58 explanatory power into complex evolutionary and demographic histories (e.g. Carstens,  
59 Lemmon, & Lemmon, 2012; Excoffier, Doupenloup, Huerta-Sánchez, Sousa, & Foll, 2013;  
60 McCormack, Hird, Zellmer, Carstens, & Brumfield, 2013).

61           Although the field of phylogeography has a long history in marine systems (e.g. Bowen  
62 et al., 1992, 1994; Reeb & Avise, 1990), cnidarians in the class Anthozoa (i.e. corals, sea  
63 anemones, zoanthids, corallimorpharians), which form the backbone of coral reefs and a major  
64 component of its biodiversity, have been historically challenging to work with at the population  
65 level. In addition to large range sizes and the logistical difficulties of sampling underwater,  
66 mitochondrial DNA barcodes (mtDNA), the molecular marker of choice for metazoan  
67 phylogeographic studies from the field's outset, evolve too slowly in most anthozoans to be  
68 useful for intraspecific studies (e.g. Allio, Donega, Galtier, & Nabholz, 2017; Daly, Gusmão,

69 Reft, & Rodríguez, 2010; Shearer, Van Oppen, Romano, & Wörheide, 2002;). Further, the  
70 overwhelming majority of tropical anthozoans found on coral reefs form endosymbioses with  
71 photosynthetic dinoflagellates in the family Symbiodinaceae, which allows these animals to  
72 thrive in oligotrophic habitats (e.g. Baker, 2003; Gates & Edmunds, 1999; Muscatine,  
73 McCloskey, & Marian, 1981; Rowan & Powers, 1991; Santos, 2016). In field-collected samples,  
74 contamination from symbiodiniaceans is unavoidable, and resulting DNA extractions harbor a  
75 mix of anthozoan and dinoflagellate DNA (termed “holobiont DNA”). The combination of  
76 slowly evolving mtDNA and dinoflagellate contamination complicates the development of  
77 molecular markers suitable for population level questions (e.g. Shearer, Gutiérrez-Rodríguez, &  
78 Coffroth, 2005). No broadly useful phylogeographic markers have ever been developed for  
79 anthozoans, and thus, most population genetic studies of tropical anthozoans rely on species-  
80 specific microsatellite loci to make population-level inferences (e.g. Andras, Rypien, & Harvell,  
81 2013; Baums, Miller, & Hellberg, 2005; Foster et al., 2012; Rippe et al., 2017; Titus et al.,  
82 2017a).

83         The generation of datasets targeting thousands of single nucleotide polymorphisms  
84 (SNPs) from anonymous loci via high-throughput sequencing is affordable and provides  
85 genome-scale data for non-model organisms. However, marine scientists interested in studying  
86 symbiotic anthozoans must still contend with symbiodiniacean contamination in genomic  
87 sequence data because there are no simple or reliable ways to completely separate symbiont and  
88 host DNA before sequencing. For studies using transcriptomic approaches, anthozoan and  
89 dinoflagellate DNA can be parsed bioinformatically, as assembled contigs are long, and  
90 conserved, enough to map to published genomic resources (e.g. Davies, Marchetti, Ries, &  
91 Castillo, 2016; Kenkel & Matz, 2016; Kenkel, Moya, Strahl, Humphrey, & Bay, 2018).

92 However, the reduced representation sequencing approaches most commonly used in population-  
93 level phylogeographic studies (e.g. RADseq, GBS) produce anonymous loci, have short read  
94 lengths (e.g. 50-100 bp), and are expected to be recovered largely from non-coding regions.  
95 Thus, currently available anthozoan reference genomes will be of limited use to separate  
96 dinoflagellate from anthozoan SNPs bioinformatically unless the reference species is closely  
97 related to the focal taxa. Likewise, the currently available genomic resources for Symbiodinaceae  
98 are also of limited use for parsing reduced representation SNP datasets because of the genetic  
99 diversity within its members: long considered to belong to a single genus (*Symbiodinium*), the  
100 photosymbiotic dinoflagellates are now recognized to represent 7-15 genus-level lineages  
101 (LaJeunesse et al., 2018), with genetic distances between many of these on par with order-level  
102 divergences (LaJeunesse et al., 2018; Rowan & Powers, 1992; Santos, 2016). Thus, any  
103 symbiodiniacean reference genome used in an attempt to disambiguate endosymbiont and host  
104 DNA needs to be very closely related to the specific endosymbiotic dinoflagellate found within  
105 the focal anthozoan species to effectively identify dinoflagellate sequences within reduced  
106 representation datasets.

107       Because of the complexity of disentangling host and symbiont sequences from reduced  
108 representation sequencing of holobiont DNA, these sequencing approaches have been applied in  
109 only limited ways to a small number of photosymbiotic anthozoan species. The majority of these  
110 studies come from the scleractinian coral genus *Acropora* (e.g. Devlin-Durante & Baums, 2017;  
111 Drury et al., 2017; Shinzato, Mungpakdee, Arakaki, & Satoh, 2015; Rosser et al., 2017;), which  
112 have circumvented symbiont contamination by mapping RADseq or GBS loci to the congeneric  
113 *Acropora digitifera* reference genome (Shinzato et al., 2011). Others have mapped anonymous  
114 loci to conspecific or congeneric transcriptomes and used only the resulting protein-coding SNP

115 datasets for interspecific phylogenetic reconstruction and hybridization studies (e.g. Combosch  
116 & Vollmer 2015; Forsman et al., 2017; Johnston et al., 2017). One study employed a subtraction  
117 library approach, spinning down homogenized tissue in an effort to remove dinoflagellate cells  
118 prior to DNA extraction and creating a separate reduced representation dinoflagellate reference  
119 library (Bongaerts et al., 2017). Leydet, Grupstra, Coma, Ribes, & Hellberg (2018) targeted  
120 anthozoan RADseq loci by including a congeneric, aposymbiotic, species in their library prep  
121 and sequencing- acting as a *de facto* reference library. Each of these studies recognized the  
122 importance of removing symbiodiniacean sequences from their reduced representation datasets,  
123 acknowledging that successful interpretation of patterns or population parameters requires  
124 knowing the extent to which each organism is contributing to the observed patterns.

125         While true in theory, in practice, how important is it to account for and remove 100% of  
126 endosymbiotic dinoflagellate loci from reduced representation datasets? Are reference genomes,  
127 or other approaches, always required in order to obtain anthozoan datasets that lead to robust  
128 phylogeographic inference and that do not lead spurious results? We hypothesize that, in many  
129 commonly encountered circumstances, the unique combination of anthozoan biology, diversity  
130 of the endosymbionts, and the manner in which *de novo* SNP-calling programs (i.e. pyRAD,  
131 *Stacks*) identify orthologous loci in reduced representation datasets will alleviate the need for  
132 anthozoan reference genomes to separate anthozoan from dinoflagellate DNA. This hypothesis  
133 rests on several observations. First, many tropical anthozoans have flexible associations that  
134 involve diverse lineages of Symbiodiniaceae (e.g. Santos, 2016; Silverstein, Correa, & Baker,  
135 2012). Members of the same host species can harbor different lineages of Symbiodiniaceae  
136 (previously called Clades and Types of *Symbiodinium*) in different habitats and across broad  
137 geographic space, and even within the same individual or colony (Baker, 2003; Silverstein et al.,

138 2012; Santos, 2016). Second, as outlined above, the genetic divergences between members of  
139 Symbiodiniaceae are comparable to order-level differences seen in other dinoflagellates,  
140 representing divergences as old as the mid-Jurassic (LaJeunesse et al., 2018, Santos, 2016).  
141 Thus, given an adequate sampling distribution, many reduced representation datasets produced  
142 for tropical anthozoans will harbor multiple lineages of Symbiodiniaceae. Finally, reduced  
143 representation datasets should be comprised of largely short, non-coding DNA fragments. When  
144 *de novo* SNP-calling programs cluster DNA sequence fragments and call orthologous loci, the  
145 user specifies a missing data threshold before a locus is incorporated into a final dataset. A  
146 moderately conservative missing data threshold may be enough to filter out the majority of  
147 symbiodiniacean sequences because the program cannot find enough mutationally-conserved,  
148 orthologous loci present across the genetically divergent Symbiodiniaceae to meet the missing  
149 data thresholds. For example, imagine a RADseq dataset consisting of 10 individuals of an  
150 anthozoan from Florida that harbor *Symbiodinium* (previously “Clade A”), and 10 individuals  
151 from Bermuda that harbor *Breviolum* (previously *Symbiodinium* “Clade B”), The genetic  
152 divergence between *Symbiodinium* and *Breviolum* is large enough (pairwise distance for LSU  
153 DNA = 0.37, estimated divergence ~170 mya: LaJeunesse et al., 2018) that few (if any) non-  
154 coding orthologous DNA sequences from *Symbiodinium* and *Breviolum* would be retained under  
155 a pyRAD missing data threshold requiring a locus to be present in 75% of all individuals. Thus,  
156 the loci that would be retained in the final dataset would primarily be from the host anthozoan,  
157 which represents intraspecific diversity at shallower evolutionary timescales.

158 To test this, we used double digest restriction-site associated DNA sequencing  
159 (ddRADseq) to reconstruct the range-wide phylogeographic history of the corkscrew sea  
160 anemone *Bartholomea annulata*- a species known to harbor multiple members of

161 Symbiodiniaceae throughout the Tropical Western Atlantic (TWA) (see Grajales, Rodríguez,  
162 Thornhill, 2016). We then leverage the genome of the sea anemone *Exaiptasia diaphana*  
163 (previously *Aiptasia pallida*; see Grajales & Rodríguez, 2014; ICZN, 2017) (Baumgarten et al.,  
164 2015), a closely related species from the same family (Aiptasiidae; Grajales & Rodriguez, 2016),  
165 to create an aposymbiotic SNP dataset for *B. annulata*. We compare the spatial genetic structure  
166 of the aposymbiotic and full holobiont SNP datasets throughout the region, and use coalescent  
167 simulation and model selection to understand whether the two datasets are similarly structured  
168 and whether they are interpreted as having similar patterns of demographic history with the same  
169 parameter estimates (i.e. effective population size, migration) as the aposymbiotic data. We  
170 discuss the competing phylogeographic reconstructions and their implications for future studies  
171 on symbiotic anthozoans, and further our understanding of the phylogeographic history of  
172 Caribbean coral reef taxa.

173

## 174 **2. Methods**

### 175 *2.1. Focal taxon*

176 The corkscrew sea anemone, *B. annulata*, is the most abundant large-bodied species of  
177 anemone on coral reef habitats throughout the TWA (Briones-Fourzán, Pérez-Ortiz, Negrete-  
178 Soto, Barradas-Ortiz, & Lozano-Álvarez, 2012). Like many tropical anthozoans, it is symbiotic  
179 with multiple members of Symbiodiniaceae throughout its range- *Symbiodinium* (formerly Clade  
180 A) in Bermuda and Florida and *Cladocopium* (formerly Clade C) in Florida, Mexico, and  
181 Panama (Grajales et al., 2016). Symbiodiniaceans are obtained horizontally after planktonic  
182 larvae metamorphose and settle to the benthos, or vertically during pedal laceration. Sexual  
183 reproduction occurs twice per year, but asexual reproduction occurs year round (Jennison, 1981).



184 Although the contribution of sexual and asexual reproduction can vary by habitat type, *B.*  
185 *annulata* appears to rely primarily on sexual reproduction (Titus et al., 2017a). This species also  
186 appears to have highly dynamic populations with rapid turnover and a maximum estimated  
187 lifespan of ~2 years (O'Reilly & Chadwick, 2017; O'Reilly, Titus, Nelsen, Ratchford, &  
188 Chadwick, In Press).

189 Ecologically, *B. annulata* serves as an important host to the most diverse community of  
190 crustacean ectosymbionts of any TWA sea anemone (Briones-Fourzán et al., 2012; Titus & Daly,  
191 2017), including cleaner shrimps that remove parasites from more than 20 families of reef fishes  
192 (Huebner & Chadwick, 2012a, b; Titus, Daly, & Exton, 2015a, b; Titus, Vondriska, & Daly,  
193 2017b, Titus, Palombit, & Daly, 2017c). Thus, this species forms the hub of a complex multi-  
194 level symbiosis that has potentially radiating effects across multiple trophic levels on TWA reef  
195 systems. Finally, this species and its crustacean symbionts are collected commercially by the  
196 ornamental aquarium trade along the Florida Reef Tract and are listed by the Florida Fish and  
197 Wildlife Conservation Commission as “biologically vulnerable” and “species of conservation  
198 concern.”

199

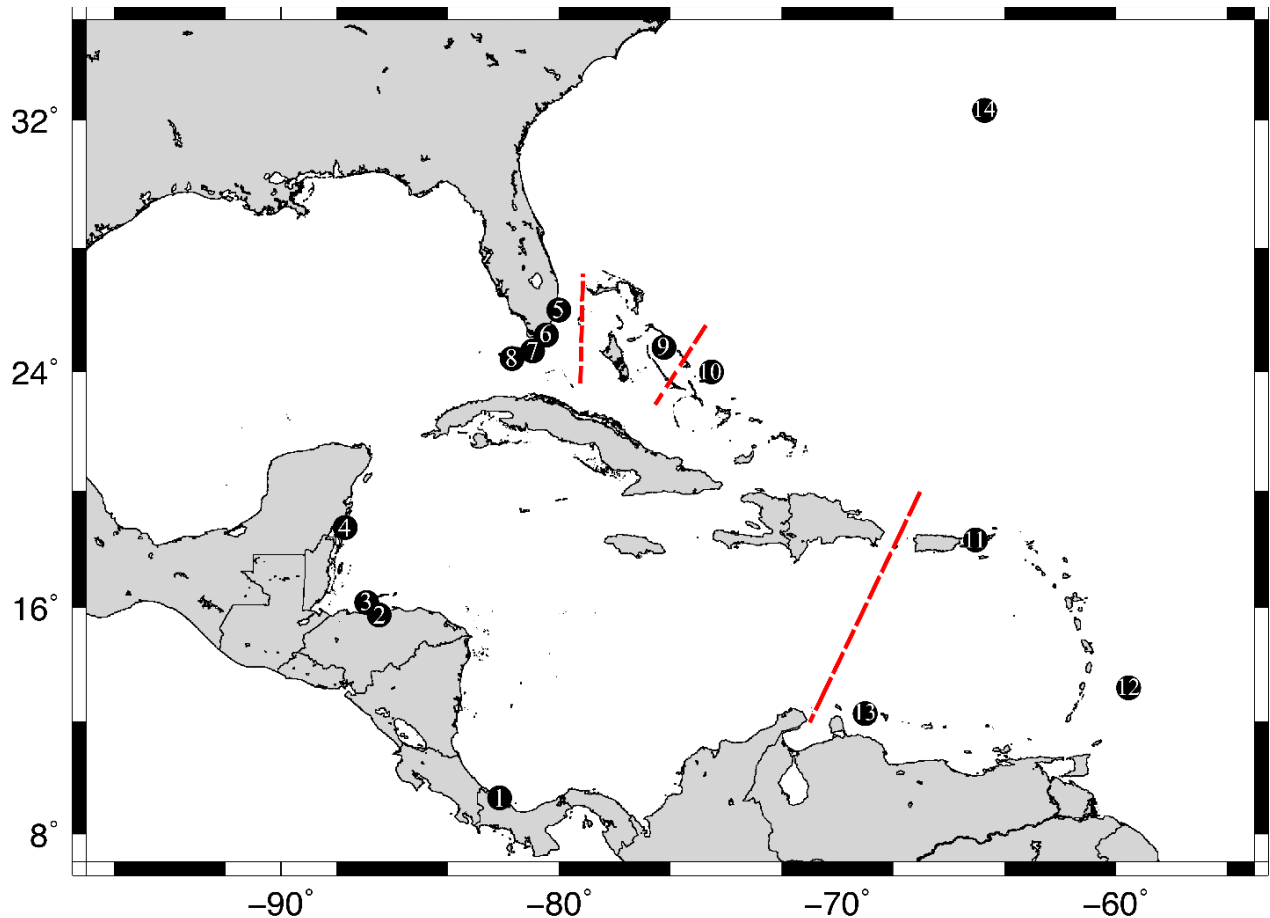
## 200 *2.2. Sample collection, DNA isolation, library preparation, and data processing*

201 Tissue samples (i.e. tentacle clippings and whole animals) were collected using SCUBA  
202 from 14 localities encompassing the entire geographic range of *B. annulata*, and from localities  
203 separated by known phylogeographic barriers (Table 1; Fig. 1; Table S1; reviewed by DeBiasse,  
204 Richards, Shivji, & Hellberg, 2016). Samples were collected by hand from coral reef habitats  
205 between 5- and 15-m depth and preserved on shore using RNAlater. 20-30 samples were  
206 collected per locality and transferred back to The Ohio State University for DNA extraction,

207 library preparation, and sequencing. Genomic DNA was isolated using DNeasy Blood and  
208 Tissue Kits (Qiagen Inc.) and stored at -20°C. DNA degradation was assessed for each sample  
209 using gel electrophoresis, and only samples with high molecular weight DNA were carried  
210 forward for ddRADseq library preparation. DNA concentrations were quantified (ng/uL) using a  
211 Qubit 2.0 (ThermoFisher) fluorometer and dsDNA broad-range assay kits. 20uL aliquots, each  
212 with 200ng of DNA, were prepared for each sample and used for ddRADseq library preparation.

213       Between 12-15 individual *B. annulata* samples per locality were carried forward for  
214 ddRADseq library preparation. Genomic DNA was digested using two restriction enzymes  
215 (*EcoRI*-HF and *psti*-HF), Illumina compatible barcodes were annealed to restriction cut sites,  
216 samples were size selected manually using a 400-800 bp size range, and then cleaned using  
217 Nucleospin Gel and PCR clean up kits (Macherey-Nagel). Following size selection, each  
218 individual sample was amplified using polymerase chain reaction (PCR), cleaned using AMPure  
219 XP beads (Agilent), and then quantified via quantitative PCRs (qPCR) to inform the pooling of  
220 individual samples into final libraries. A total of 141 individuals (Table 1) met all quality control  
221 steps and were pooled across five separate libraries. Samples were sequenced on an Illumina  
222 HiSeq 2500 using single-end 100 base pair reads at The Ohio State University Genomics Shared  
223 Resource.

224



225  
226 Figure 1. Map of sampling localities throughout the Tropical Western Atlantic for the populations of corkscrew sea  
227 anemone *Bartholomea annulata* studied here. 1. Bocas del Toro, Panama, 2) Cayos Cochinos, Honduras, 3) Utila,  
228 Honduras, 4) Mahahual, Mexico, 5) Ft. Lauderdale, Florida, 6) Upper Keys, Florida, 7) Middle Keys, Florida, 8)  
229 Lower Keys, Florida, 9) Eleuthera, Bahamas, 10) San Salvador, Bahamas, 11) St. Thomas, US Virgin Islands, 12)  
230 Barbados, 13) Curacao, 14) Bermuda. Red dashed lines denote previously recovered major phylogeographic breaks  
231 in the region.  
232

### 233 2.3. Data processing and aposymbiotic dataset assembly

234 Raw sequence reads were demultiplexed, aligned, and assembled *de novo* using the  
235 program pyRAD v3.0.66 (Eaton, 2014). We required a minimum base call Phred score of 20 and  
236 set the maximum number of bases in a locus with Phred scores < 20 (NQual) to five. Low quality  
237 base calls were replaced with Ns. We set the clustering threshold (Wclust) to 0.90 to assemble  
238 reads into loci, and required a minimum coverage depth of seven to call a locus (Mindepth).  
239 Finally, we required a locus to be present in 75% of all individuals to be retained in the final

240 dataset. RADseq protocols are known to be susceptible to missing data due to mutations in  
241 restriction cut sites and allelic dropout (e.g. Arnold, Corbett-Detig, Hartl, & Bomblies, 2013), but  
242 biases can also arise when datasets are overly conservative (i.e. no missing data allowed; Huang  
243 & Knowles, 2014). Thus we allowed some missing data in our final dataset.

244 Previously, we delimited two cryptic lineages of *B. annulata* co-distributed throughout  
245 the range in our ddRADseq dataset (Titus, Blischak, & Daly, 2018). After running pyRAD to  
246 completion, members of the infrequently sampled lineage (Clade 1, N = 18 individuals; Titus et  
247 al., 2018) and individuals with low sequencing coverage (< 500,000 reads; N = 21 individuals),  
248 were removed from the dataset, leaving only the more well sampled lineage (Clade 2; n = 101  
249 individuals; Table 1; Table S1). After accounting for cryptic diversity this initial SNP dataset  
250 represented our “holobiont dataset,” sequences that were, putatively, a combination of anemone  
251 and algal DNA.

252 To create an anemone-only “aposymbiotic dataset,” we mapped polymorphic loci from  
253 the holobiont ddRADseq dataset to the genome of the closely related *Exaiptasia diaphana*  
254 (Baumgarten et al., 2015) to identify anemone-only sequences. *Exaiptasia diaphana* and *B.*  
255 *annulata* are members of the same family and are closely related (Grajales & Rodriguez, 2016),  
256 and polymorphic microsatellites have previously been designed from *E. diaphana* that amplify in  
257 *B. annulata* (Titus et al., 2017a). To map polymorphic *B. annulata* loci to *E. diaphana*, we  
258 downloaded the *E. diaphana* genome and created a local BLAST database. After initially  
259 running pyRAD to completion, a python script (parse\_loci.py, available on Dryad doi:XXX) was  
260 written to select the first DNA sequence from each locus in the .loci output file, and create a  
261 .fasta file that could then be BLAST-ed against the *E. diaphana* genome (BLAST+; Camacho et  
262 al., 2009). We used an 85% identity threshold to call a locus as putatively anemone in origin.

263 Next, a separate python script (blast2loci.py, available on Dryad doi:XXX) was used to read  
264 through the BLAST output file, pull all sequences in all loci that met the 85% identity threshold,  
265 and create a new .loci file with the same file name as the original. The original .loci file was then  
266 replaced with the new anemone-only file, at which point the final step of pyRAD (step 7) was re-  
267 run to create our final anemone-only output files (i.e. unlinked SNPs and alleles files) for  
268 downstream analyses. As a final check, we created additional local BLAST databases by  
269 downloading publicly available endosymbiotic dinoflagellate genomes: *Symbiodinium*  
270 *micradriaticum* (Aranda et al., 2016 as *Symbiodinium micradriaticum* “Clade A”), *Breviolum*  
271 *minutum* (Shoguchi et al., 2013 as *Symbiodinium minutum* “Clade B”), *Cladocopium goreau*  
272 (Liu et al., 2018, as *Symbiodinium goreau* “Clade C”), and *Fugacium kawagutii* (Liu et al.,  
273 2018, as *Symbiodinium kawagutii* “Clade F”). We mapped both our holobiont and aposymbiotic  
274 datasets to the symbiodiniacean genomes to see if we could 1) identify any symbiodiniacean  
275 sequences in the holobiont data, and 2) confirm that no loci in our aposymbiotic dataset mapped  
276 to both symbiodiniacean and *Exaiptasia* genomes. Lastly, we mapped our holobiont dataset to  
277 the genome of the distantly related starlet sea anemone, *Nematostella vectensis* (Putnam et al.,  
278 2007), to gauge the extent to which intra-order (Actiniaria) genomic resources could be used to  
279 effectively identify anemone-only 100bp ddRADseq loci. All scripts for mapping and parsing  
280 anemone from symbiodiniacean DNA, along with full details and instructions for using them,  
281 can be found on Dryad (doi:XXX).

282

#### 283 2.4. Population genetic structure

284 We used the clustering program Structure v2.3.4 (Pritchard, 2000) to infer population  
285 genetic structure across the Tropical Western Atlantic. For both holobiont and aposymbiotic

286 datasets, we collapsed bi-allelic data into haplotypes at each locus, thus using information  
287 contained in linked SNPs when more than one SNP was present in a locus. Structure analyses  
288 were conducted using the admixture model and correlated allele frequencies. Each MCMC chain  
289 for each value of  $K$  was run with a burnin of  $1 \times 10^5$  generations and sampling period of  $2 \times 10^5$   
290 generations.

291 We initially conducted two separate Structure analyses for both the holobiont and  
292 aposymbiotic datasets. First, we conducted three iterations of a broad range of  $K$  values (1- 6) to  
293 gain an initial snapshot of the data across the region. In both initial analyses we used the peak  $\ln$   
294  $\text{Pr}(D|K)$  and the  $\Delta K$  (Evanno et al. 2005) to inform the selection of the best  $K$  value. We then re-  
295 ran Structure using a narrower range of  $K$  values (1-4) but with more iterations ( $n = 10$ ). Each  
296 MCMC chain for each value of  $K$  was run with a burnin of  $1 \times 10^5$  generations and sampling  
297 period of  $2 \times 10^5$  generations. Again, we used  $\ln \text{Pr}(D|K)$  and  $\Delta K$  to select the best value of  $K$ .

298 We conducted an analysis of molecular variance (AMOVA) in Arlequin v.3.5 (Excoffier  
299 & Lischer, 2010) to test for hierarchical partitioning of genetic diversity across the region.  
300 Following our Structure results (see Results), we partitioned samples into Eastern and Western  
301 regions. We tested for hierarchical structure among sample localities ( $\varphi_{ST}$ ), among sample  
302 localities within a region ( $\varphi_{SC}$ ), and between regions ( $\varphi_{CT}$ ). Calculations in Arlequin v3.5 were  
303 made using haplotype data and distance matrices calculated using the number of different alleles  
304 per locus. Statistical significance was assessed with 10,000 permutations. Pairwise  $\varphi_{ST}$  values  
305 were calculated to test for differentiation among sample localities. Genetic diversity summary  
306 statistics and pairwise  $\varphi_{ST}$  values were also calculated in Arlequin for all sample localities. All  
307 calculations were conducted for both aposymbiotic and holobiont datasets.

308

309 2.5. *Demographic modeling selection and parameter estimation*

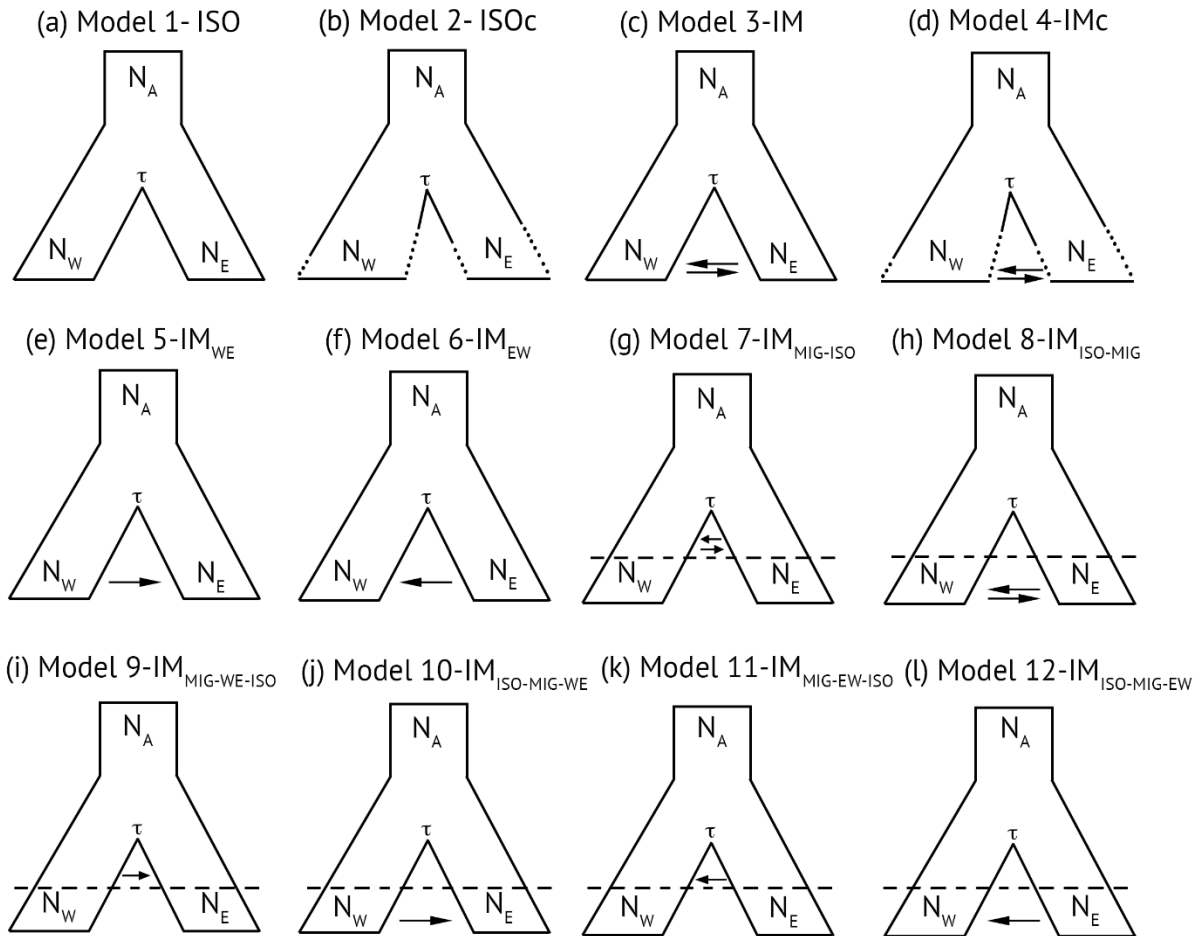
310 While broad-scale patterns of spatial genetic structure may be robust to some levels of  
311 dinoflagellate contamination in reduced representation sequencing datasets, we expect that  
312 demographic model selection approaches that make inferences regarding patterns of  
313 demographic history, and that generate important population parameter estimates (i.e. effective  
314 population sizes, migration rates), should be highly sensitive to the incorporation of data from  
315 taxa with different evolutionary histories. Thus, we conducted model selection using the allele  
316 frequency spectrum (AFS) and coalescent simulations in the program *fastsimcoal2* (FSC2;  
317 Excoffier et al., 2013). FSC2 uses coalescent simulations to calculate the composite likelihood of  
318 arbitrarily complex demographic models under a given AFS. The best fit model can then be  
319 selected using Akaike information criterion (AIC). We developed 12 user-specified demographic  
320 models (Figure 2), all variants of a two-population isolation-migration model as Structure  
321 delimited  $K = 2$  as the best clustering scheme (see Results). Models differed in the directionality  
322 of gene flow, population size changes following divergence, and whether they exhibit patterns of  
323 secondary contact following divergence. Genetic clusters in Structure were largely partitioned  
324 East and West in the TWA, and 25 individuals from each putative population (50 individuals  
325 total; Table S2) were selected to generate two-population, joint-folded, AFS. We conducted  
326 model selection on both aposymbiotic and holobiont datasets.

327 Two-population, joint-folded AFS were generated from pyRAD output files and  
328 previously published python scripts (see Satler & Carstens, 2017). One of the assumptions of  
329 FSC2 is that SNPs are in linkage equilibrium (Excoffier et al., 2013), and thus, only one SNP per  
330 locus was selected to produce the AFS. Further, AFS calculations in FSC2 require fixed numbers  
331 of alleles from all populations (i.e. no missing data). As meeting this latter requirement would

332 greatly decrease our dataset size, and thus likely bias our analyses, we followed the protocol of  
333 Satler and Carstens (2017) and Smith et al., (2017) by requiring a locus in our AFS to be present  
334 in 85% of all individuals. To account for missing data without violating the requirements of the  
335 AFS we built our AFS as follows: 1) if a locus had fewer alleles than our threshold it was  
336 discarded, 2) if a locus had the exact number of alleles as the threshold, the minor allele  
337 frequency was recorded, and 3) if a locus exceeded the threshold, alleles were down-sampled  
338 with replacement until the number of alleles met the threshold, at which point the minor allele  
339 frequency was counted. This approach allowed us to maximize the number of SNPs used to build  
340 the AFS, but also has the potential to lead to monomorphic alleles based on the down-sampling  
341 procedure (see Satler & Carstens 2017). Thus, we repeated the AFS building procedure 10 times,  
342 allowing us to account for variation in the down-sampling process during model selection, and  
343 also allowing us to calculate confidence intervals on our parameter estimates (Satler & Carstens,  
344 2017; Smith et al., 2017).

345 Each simulation analysis in FSC2 (i.e. each AFS replicate per model; 12 models x 10  
346 replicates) was repeated 50 times, and we selected the run with the highest composite likelihood  
347 for each AFS replicate and model. The best-fit model was then calculated using the AIC and  
348 model probabilities calculated following Burnham and Anderson (2002). Because FSC2 requires  
349 a per generation mutation rate to scale parameter estimates into real values, we used the  
350 substitution per site per generation mutation rate of  $4.38 \times 10^{-8}$  proposed for tropical anthozoans  
351 (Prada et al., 2017) and a generation time of 1 year for *B. annulata* (Jennison, 1981). All analyses  
352 were conducted on the Oakley cluster at the Ohio Supercomputer Center (<http://osc.edu>).





353

354 Figure 2. Models used in FSC2 to understand the demographic processes leading to the two-population pattern of  
 355 diversification in the corkscrew anemone *Bartholomea annulata* across the Tropical Western Atlantic. Each model  
 356 is a two-population isolation-migration (IM) model that varies in the degree and directionality of gene flow and  
 357 effective population size. Models are as follows: a) isolation only, b) isolation with population size changes  
 358 following divergence, c) IM model with symmetric migration, d) IM model with symmetric migration and  
 359 population size changes, e) IM model with migration from the Western to Eastern population, f) IM model with  
 360 migration from population Eastern to Western, g) IM model with symmetric migration between populations  
 361 immediately following divergence followed by more contemporary isolation, h) IM model with isolation  
 362 immediately following divergence followed by secondary contact and symmetric migration, i) IM model with  
 363 migration from population Western to Eastern immediately following divergence followed by more contemporary  
 364 isolation, j) IM model with isolation immediately following divergence followed by secondary contact and  
 365 migration from population Western to Eastern, k) IM model with migration from population Eastern to Western  
 366 immediately following divergence followed by more contemporary isolation, and l) IM model with isolation  
 367 immediately following divergence followed by secondary contact and migration from population Eastern to  
 368 Western.

369

370

### 371 3. RESULTS

#### 372 3.1. Dataset assembly

373 Double digest RADseq library preparation and sequencing resulted in a total of 186.7  
374 million sequence reads across 141 individuals, 175.6 million of which passed quality control  
375 filtering and were retained to create the final dataset. Accounting for individuals with low  
376 sequence reads (< 500,000 reads) and cryptic species level diversity (Titus et al., 2018) resulted  
377 in a final intraspecific dataset of 101 individuals (Table 1; Table S1). Requiring a locus to be  
378 present in a minimum of 75% of all individuals resulted in a final holobiont data set of 11,331  
379 parsimony-informative sites distributed across 3854 unlinked loci. After BLASTing these loci to  
380 the *Exaiptasia diaphana* genome, we retained 1402 loci that had matched with high confidence  
381 ( $\geq 85\%$  identity) and these were used as the final aposymbiotic SNP dataset. A total of 59 of the  
382 3854 holobiont ddRADseq loci ( $\sim 1.5\%$ ) mapped to Symbiodiniaceae genomes (Table S3),  
383 confirming the presence of at least some symbiont DNA in our holobiont dataset. Of these, 58  
384 mapped to the *S. microadriaticum* genome and one mapped to the *C. goreau* genome (Table  
385 S3). Only five *B. annulata* loci mapped to the starlet sea anemone *N. vectensis* genome (Table  
386 S4). SNP files and raw data for both holobiont and aposymbiotic datasets are available on Dryad  
387 (Dryad doi:XXX).

388

### 389 *3.2. Population genetic structure*

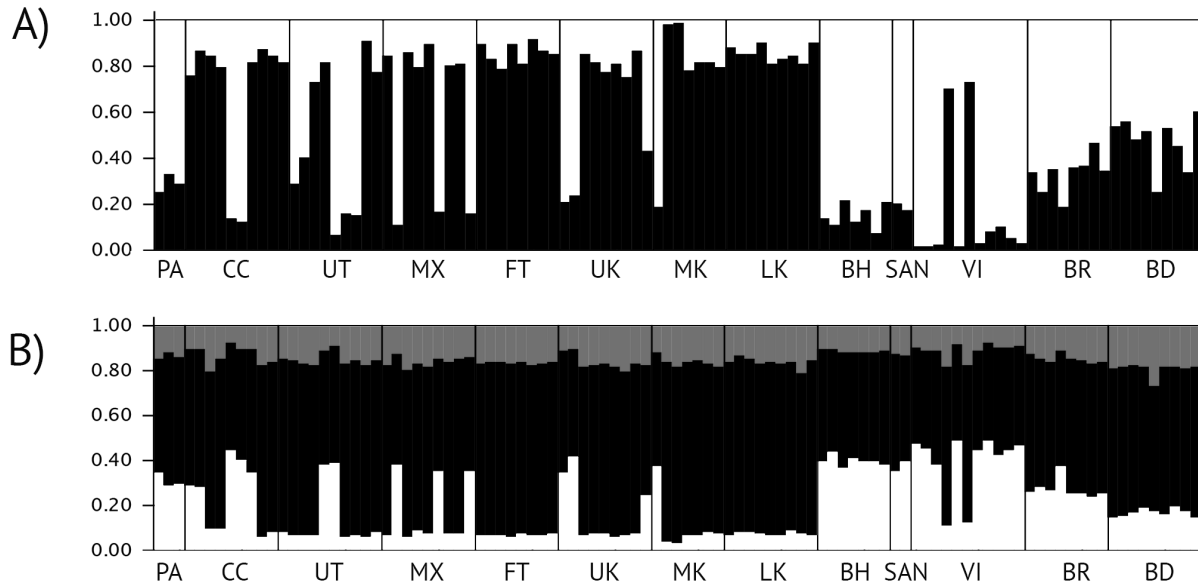
390 Genetic clustering analysis in Structure resolved similar patterns across the TWA for both  
391 aposymbiotic and holobiont datasets. For the aposymbiotic dataset,  $K = 2$  was selected by  
392 Structure using both  $\ln P(K)$  and  $\Delta K$  as the best clustering scheme (Fig. 3; Table S5). Diversity  
393 was largely binned into Western and Eastern partitions, but with admixture (Figure 3). The most  
394 notable genetic break was that between the Lower Keys (LK) and Eleuthera, Bahamas (BH),  
395 sample localities in close proximity and bisected by the Florida Straits (Figs. 1 & 3). The

396 holobiont dataset recovered similar geographic partitioning, but Structure selected  $K = 3$  as the  
397 best partitioning scheme using  $\ln P(K)$  and  $\Delta K$  (Fig. 3; Table S6). The additional genetic cluster  
398 did not illuminate any unrecovered geographic partitioning across the region beyond what was  
399 recovered by a  $K = 2$  partitioning scheme (Fig. 3). The West-East genetic break across the TWA,  
400 with admixture, is still largely resolved in the holobiont dataset with the most notable break  
401 again between the LK and BH sample localities (Fig. 3b).

402       Population genetic analyses in Arlequin reflect nearly identical results for both datasets.  
403 AMOVA results indicate low, but significant, population genetic structure at all hierarchical  
404 levels for both datasets, and both datasets have similar patterns of genetic variation at each  
405 hierarchical level (Table 2). Similarly, pairwise  $\phi_{ST}$  values calculated by Arlequin were low, but  
406 significant, among many sample localities for both datasets (Table 3), and there were no major  
407 differences in the genetic diversity summary statistics for both datasets across the range (Table  
408 4).

409

410



411

412 Figure 3. A) Genetic clustering results ( $K=2$ ) for the aposymbiotic *Bartholomea annulata* RADseq dataset. B)  
413 Genetic clustering results ( $K=3$ ) for the holobiont *Bartholomea annulata* RADseq dataset. Samples are partitioned  
414 by sample locality in a largely West to East (left to right) geographic layout. 1. Bocas del Toro, Panama, 2) Cayos  
415 Cochinos, Honduras, 3) Utila, Honduras, 4) Mahahual, Mexico, 5) Ft. Lauderdale, Florida, 6) Upper Keys, Florida,  
416 7) Middle Keys, Florida, 8) Lower Keys, Florida, 9) Eleuthera, Bahamas, 10) San Salvador, Bahamas, 11) St.  
417 Thomas, US Virgin Islands, 12) Barbados, 13) Bermuda.

418

### 419 3.3. Demographic model selection

420 Coalescent modeling in FSC2 returned identical model selection results between  
421 aposymbiotic and holobiont datasets (Table 5). For both, Akaike Information Critereon (AIC)  
422 selected model 6, an IM model with unidirectional gene flow from East to West, as the best-fit  
423 model (Fig. 2). According to Akaike model weights, model 6 received over 0.70 of the support  
424 (Table 5) in both aposymbiotic and holobiont datasets. A secondary contact model (Model 10;  
425 Fig. 2), with isolation immediately after divergence followed by secondary contact and  
426 unidirectional West-East gene flow, received the next highest amount of support according to  
427 AIC, although the Akaike weight differed in how much support was given to each, with the  
428 aposymbiotic dataset having a clearer preference for this model over the next best one, compared  
429 to the holobiont data set (Table 5).

430 Parameter values and confidence intervals for effective population size, divergence time,  
431 and migration rate estimated from FSC2 simulations were entirely overlapping between  
432 aposymbiotic and holobiont datasets (Table 6). For both datasets, FSC2 estimated that Eastern *B.*  
433 *annulata* populations had greater effective population sizes than Western populations, and that  
434 per-generation migration rate was low (Table 6). Divergence time estimates varied more than  
435 other parameter values but still had overlapping confidence intervals. The aposymbiotic dataset  
436 had an estimated a mean divergence time between Eastern and Western populations at ~39,000  
437 ybp, whereas the holobiont dataset had an estimated a mean divergence time between  
438 populations at ~58,000 ybp.

439

#### 440 **4. DISCUSSION**

##### 441 *4.1. Reduced representation sequencing for symbiotic anthozoans*

442 Symbiodiniacean DNA contamination has represented a substantial hurdle for researchers  
443 working on tropical anthozoans (e.g. Shearer et al., 2005; Bongaerts et al., 2017; Leydet et al.,  
444 2018), and high-throughput sequencing has done little to alleviate these issues in population-  
445 level studies. Based on the analyses we conduct here, however, we fail to reject our hypothesis  
446 that, anthozoan reference genomes may not always be necessary for making reliable spatial and  
447 demographic phylogeographic inference including, importantly, population parameter estimates.  
448 Instead, we find broadly similar interpretations from a holobiont dataset and from one “cleaned”  
449 of symbiont sequences. Understanding the evolutionary and historical processes that have shaped  
450 the diversity of tropical anthozoans has been, and will continue to be, an important research  
451 priority for marine phylogeographers (Bowen, Rocha, Toonen, Karl, & ToBo Lab, 2016; Bowen  
452 et al., 2016). Our results present a promising framework and way forward for researchers

453 wishing to employ these reduced representation sequencing approaches on symbiotic anthozoan  
454 species that are not closely related to species with currently available reference genomes.

455         The framework and experimental design of our study, effectively a single-species  
456 phylogeographic study that spans the entire range of our focal taxon, is representative of many  
457 studies that examine the spatial and demographic history of a given species. Although the degree  
458 to which symbiotic anthozoans are specific to a particular lineage of Symbiodiniaceae is  
459 unresolved, evidence is mounting that these associations are spatially and temporally variable,  
460 particularly in stony corals, where much of this research has focused (e.g. Silverstein et al.,  
461 2012). Thus, we believe that given a broad sampling scheme with respect to geography and  
462 habitat, *de novo* assembly and SNP-calling programs will act as *de facto* filtering programs for  
463 symbiodiniaceans in many reduced representation datasets produced from symbiotic anthozoans.  
464 Resulting datasets will be overwhelmingly comprised of anthozoan DNA loci.

465         In our analyses, we would expect that more than doubling of our dataset (~1400 to 3800  
466 SNPs) by the inclusion of putative symbiont loci in our holobiont dataset would lead to major  
467 differences in interpretation. The ~2400 uncharacterized loci that did not map to the *E. diaphana*  
468 genome represent some combination of anemone and symbiont sequences. Because each  
469 anthozoan tentacle cell can contain dozens of *Symbiodinium* cells, and thus *Symbiodinium* nuclei  
470 often outnumber anemone nuclei (reviewed by Davy, Allemand, & Weis, 2012), the holobiont  
471 dataset could reflect a greater contribution from dinoflagellates than from *B. annulata*. If even  
472 only half of the 2400 uncharacterized loci our holobiont dataset were from members of  
473 Symbiodiniaceae, we would expect them to greatly influence our holobiont analyses, especially  
474 our parameter estimates, which should be the most sensitive to the incorporation of sequence  
475 data from multiple species with different evolutionary histories. That we recover

476 indistinguishable phylogeographic histories with completely overlapping diversity indices and  
477 parameter estimates leads us to hypothesize that we have very few symbiodiniacean loci in our  
478 holobiont dataset. The use of a sea anemone reference genome from the same family rather than  
479 a congeneric or conspecific reference genome is the most likely explanation for why ~2300 loci  
480 remain uncharacterized in our holobiont dataset: these loci are not shared between *E. diaphana*  
481 and *B. annulata*, and so are not included in the aposymbiotic dataset (because that uses the *E.*  
482 *diaphana* genome as a probe for putative anemone loci). Although mapping our reads to  
483 genomic resources from members of Symbiodiniaceae confirms we do have some dinoflagellate  
484 sequence data in our holobiont dataset (at least ~1.5% of all loci), these are such a small fraction  
485 of the SNPs that they may simply be genetic “noise,” swamped out by orders of magnitude more  
486 anthozoan SNPs.

487         Conspecific, or congeneric, reference genomes clearly represent the best approach to  
488 removing symbiont loci from reduced representation datasets. However, to date, there are only a  
489 handful of published anthozoan genomes (Baumgarten et al. 2015; Prada et al., 2017; Putnam et  
490 al. 2007; Shinzato et al. 2011; Wang, Liew, Li, Zoccola, Tambutte, & Aranda, 2017; Voolstra et  
491 al., 2017). Our study demonstrates that reference genomes within the same family may serve as  
492 adequate genomic resources, but reference genomes that are simply within the same order are  
493 likely too distant to serve in the same capacity, at least for actinarians: only five loci from *B.*  
494 *annulata* (suborder Anthemonae, superfamily Metridioidea, family Aiptasiidae) mapped to the  
495 genome of *Nematostella vectensis* (suborder Anenthemonae, superfamily Edwardsioidiea,  
496 Family Edwardsiidae). This point parallels the observation that using a reference genome of the  
497 endosymbiotic dinoflagellate without concern for the particular lineage of symbiodiniacean  
498 harbored by a particular anthozoan is unlikely to remove all dinoflagellate loci.

499           From a practical standpoint, we recommend that studies employing reduced  
500 representation approaches for symbiotic anthozoans without genomic resources from closely  
501 related species 1) employ extensive geographic sampling, or sample broadly across ecologically  
502 disjunct habitats (i.e. depth, temperature, nutrient concentration) to maximize the likelihood of  
503 sampling hosts that harbor diverse symbiodiniaceans and 2) demonstrate empirically that  
504 multiple lineages of Symbiodiniaceae are represented in the collected samples via PCR or  
505 sequencing (e.g. ITS, cp23s). In host species with highly specific endosymbiont associations, the  
506 approach to sampling and sequencing we describe here would likely be ineffective, as  
507 orthologous symbiodiniacean loci would be present in all samples and sample localities, and *de*  
508 *novo* clustering programs would not filter these out. In these cases, employing approaches like  
509 those of Bongaerts et al. (2017) or Leydet et al. (2018) may be required. Finally, for a host  
510 species where population genetic differentiation is driven by a handful of SNPs under selection,  
511 incorporating even a small number of symbiont loci could mask important signal. This is  
512 unknowable *a priori*, and studies wanting to analyze holobiont DNA at the population level  
513 should acknowledge these limitations, and follow up studies should be conducted once reference  
514 genomes are available.

515

#### 516 *4.2. Phylogeographic history of Bartholomea annulata*

517           The phylogeographic history of coral reef communities in the TWA most often revolves  
518 around a major barrier to dispersal at the Mona Passage, separating Hispanola from Puerto Rico  
519 (e.g. Baums et al., 2005; Hellberg, 2007; DeBiasse et al. 2016). This barrier has been well  
520 resolved for a number of stony corals, fishes, and other invertebrates (reviewed by DeBiasse et  
521 al. 2016). Our range-wide phylogeographic analysis demonstrates that the corkscrew sea



522 anemone *Bartholomea annulata* shows subtle, but significant, genetic structure across the TWA,  
523 with the Florida Straits, rather than the Mona Passage, being the most well resolved  
524 phylogeographic break in the region. Further, while we demonstrate a number of low, but  
525 significant,  $\varphi_{ST}$  values across many sample localities, genetic clustering loosely groups *B.*  
526 *annulata* into Eastern and Western populations (Fig. 3). The Bahamas and the Florida Keys,  
527 sample localities immediately to the East and West of the Florida Straits are separated by ~100  
528 km, but is the region with the clearest genetic partitioning (Figure 3). Sample localities further  
529 East (e.g. Barbados, Bermuda) and West (e.g. Mexico, Honduras) exhibit more genetic  
530 admixture and may have experienced more historical and contemporary gene flow. No other  
531 major genetic structuring was recovered across the TWA, although phylogeographic breaks and  
532 regions with unique genetic diversity are known for other groups of organisms, including the  
533 Southern Caribbean phylogeographic break between Panama and Curacao, a proposed Central  
534 Bahamas phylogeographic break, and regions such as the Meso-American Barrier Reef, Panama,  
535 and Bermuda (reviewed by DeBiasse et al. 2016).

536       Across the Florida Straits, demographic model selection suggests that the best fit for  
537 these data among the models we tested is a two-population pattern with continuous  
538 unidirectional gene flow from East to West following divergence (Table 5). An important note is  
539 that as the coalescent is a backwards-in-time framework, a model with gene flow from East to  
540 West reflects forward-in-time gene flow from West to East. This largely fits with the prevailing  
541 currents in the TWA, as currents that deflect North in the Western Caribbean basin ultimately  
542 form the Loop Current in the Gulf of Mexico and then are forced East through the narrow stretch  
543 of sea between Florida and Cuba before turning North again and forming the Gulf Stream.

544 Contemporary gene flow from East to West, would most likely occur in the Southern Caribbean  
545 where equatorial currents flow westward near the Southern Windward Islands.

546 Divergence time estimates from FSC2 between Eastern and Western populations given  
547 the current estimate of mutation rate suggests a recent divergence between populations of *B.*  
548 *annulata* sometime within the last 30,000-50,000 years (Table 6), firmly within the most recent  
549 glacial maxima (15,000-100,000 years before present). During this time, sea level would have  
550 been as much as 120 m below present day levels, and both the Florida peninsula and the  
551 Bahamas platform would have been sub-aerially exposed, significantly increasing the amount of  
552 dry land subdividing the region and also decreasing available reef habitat (reviewed by Ludt &  
553 Rocha 2015). This would have been especially true for Eleuthera, which would have been  
554 isolated from the Florida Straits by two large portions of the then-dry-land Bahamas Banks, and  
555 two enclosed deep water trenches (Tongue of the Ocean and the Exuma Sound). Water  
556 exchange, and thus potential for dispersal and gene flow, would have been greatly reduced  
557 during this period, allowing for allopatric divergence and local retention of larvae. This scenario  
558 would fit well with a phylogeographic model of divergence followed by a period of isolation,  
559 then secondary contact and migration during more recent interglacial periods which coincided  
560 with sea level rise. A secondary contact model was the next best fit to our data according to AIC  
561 (Table 5). However, we either 1) do not have enough signature in the data for it to be selected as  
562 the best fit, or, 2) even though the Florida and Bahamas populations would have been largely  
563 isolated, other sample localities in the Eastern (i.e. Virgin Islands, Barbados) and Western (i.e.  
564 Mexico, Honduras) would not have been isolated to the same extent, and gene flow between  
565 these localities could be responsible for the unidirectional gene flow we see in our best-fit  
566 models.

567            Interestingly, FSC2 simulations and population summary statistics estimate larger  
568 effective population sizes in the Eastern Caribbean than in the West (Table 6). At face value, this  
569 seems to be at odds with the current geography of the Western Caribbean basin as there is more  
570 submerged continental-shelf shallow-water habitat in the Western Caribbean and Florida than  
571 there is in the Eastern Caribbean (Ludt & Rocha, 2015), where coral reef habitat is largely  
572 limited to small fringing reefs around islands of volcanic origin. Unidirectional gene flow from  
573 the West to the East, as recovered by our best-fit model, could be responsible for this increase in  
574 effective population size, with the Eastern Caribbean effectively a sink of a genetic diversity. In  
575 addition, the Bahamas are a large, shallow, archipelago and likely capable of supporting  
576 immense census population sizes of *B. annulata*. As a habitat generalist, *B. annulata* can  
577 colonize hard bottom, seagrass, mangrove, and coral-dominated habitats (e.g. Briones-Fourzán et  
578 al. 2012; O'Reilly & Chadwick, 2017; Titus et al., 2017a), and is thus, not limited strictly to fore  
579 reefs. Large habitat space with genetic input from Western population could be driving this  
580 pattern.

581

#### 582 4.3. Conclusions

583            Our study demonstrates that the corkscrew sea anemone, *Bartholomea annulata*, exhibits  
584 weak genetic structure across the Tropical Western Atlantic, and that demographic modeling of  
585 this species suggests that unidirectional gene flow from the western to eastern Caribbean can  
586 largely explain the observed patterns of genetic diversity. Interestingly, we recover the same  
587 spatial and demographic patterns, including entirely overlapping parameter estimates, regardless  
588 of whether we use an aposymbiotic ddRADseq dataset or whether we use our putative holobiont  
589 dataset. Although we can confirm that at least ~1.5% of the loci in the holobiont dataset are from

590 members of Symbiodiniaceae, we hypothesize that the remaining ~2400 uncharacterized loci are  
591 primarily from *B. annulata*, representing SNPs not shared with *Exaiptasia diaphana*. Because of  
592 the diversity of dinoflagellate lineages hosted by *B. annulata* across its range and the genetic  
593 divergence within and among lineages within Symbiodiniaceae, we believe the manner in which  
594 *de novo* reduced representation clustering algorithms assemble RADseq datasets effectively  
595 removes most of the SNPs from the photosymbionts.

596 To further test our hypothesis, this study should be repeated with an anthozoan species  
597 with flexible *Symbiodinium* associations and that has a conspecific reference genome available  
598 (e.g., in *Exaiptasia diaphana*, *Acropora digitifera*, *Stylophora pistillata*). This would allow the  
599 exact number of anthozoan SNPs identified in the final dataset to be quantified rather than  
600 leaving a large fraction of SNPs uncharacterized. Nonetheless, our findings represent an  
601 important avenue along which future research on symbiotic anthozoans can continue until greater  
602 numbers of reference genomes can be sequenced, annotated, and made publicly available.  
603 Tropical anthozoans form the foundation of ecosystems that rival rainforests in diversity,  
604 perform important ecological roles, have commercial value, and are especially vulnerable to  
605 climate change (e.g. Hughes et al. 2017; Palumbi et al., 2014). As selection can act rapidly on  
606 standing genetic diversity (Przeworski et al., 2005; Barrett & Schluter, 2008; Reid et al., 2016),  
607 understanding the historical processes that have shaped contemporary distributions of diversity  
608 can help set conservation priorities in a rapidly changing climate.

609

## 610 **Acknowledgements**

611 We are grateful to Erich Bartels, Annelise del Rio, Jose Diaz, Dan Exton, Lisle Gibbs, Natalie  
612 Hamilton, Alex Hunter, Anna Klompen Jason Macrander, Spencer Palombit, Stephen Ratchford,

613 Nancy Sheridan Nuno Simoes, Jill Titus, Cory Walter, Eric Witt, Clay Vondriska, and the  
614 Operation Wallacea dive staff for assistance in the field and laboratory. We also thank Paul  
615 Blischak, Jordan Satler, Megan Smith, and Bryan Carstens for bioinformatic assistance and  
616 advice regarding *fastsimcoal2* analyses and model selection. Bellairs Research Station, the  
617 Bermuda Institute of Ocean Science, Cape Eleuthera Institute, CARMABI, Coral View Dive  
618 Center, Gerace Research Centre, the Honduran Coral Reef Foundation, Mote Marine Laboratory,  
619 Smithsonian Tropical Research Institute, and the University of the Virgin Islands provided  
620 valuable logistical support in the field. Specimens were collected from throughout the Tropical  
621 Western Atlantic under permits: SE/A-88- 15, PPF/DGOPA-127/14, CZ01/9/9, FKNMS-2012-  
622 155, SAL-12-1432A-SR, STT037-14, 140408, MAR/FIS/17, and 19985. This research was  
623 supported by funding from a National Science Foundation-Doctoral Dissertation Improvement  
624 Grant DEB-1601645 and Florida Fish and Wildlife Conservation Commission awards to B.M.T.  
625 & M.D. Operation Wallacea, American Philosophical Society, International Society for Reef  
626 Studies Graduate Fellowship, PADI Foundation Grant, and American Museum of Natural  
627 History Lerner Gray Funds supported field research for B.M.T. Additional funding was provided  
628 through the Trautman Fund of The OSU Museum of Biological Diversity, The Ohio State  
629 University, and National Science Foundation DEB-1257796 to MD.

630

631 **Data Accessibility Statement:** Raw sequence data, Python scripts, and all files for all analyses  
632 will be archived in Dryad upon final acceptance of this manuscript.

633

634 **Author Contributions:** B.M.T. and M.D. conceived the study and collected samples, B.M.T.  
635 conducted laboratory work and analyzed the data. B.M.T. and M.D. wrote and edited the  
636 manuscript.

637

## 638 **References**

- 639 Allio, R., Donega, S., Galtier, N., & Nabholz, B. (2017). Large variation in the ratio of  
640 mitochondrial to nuclear mutation rate across animals: implications for genetic diversity  
641 and the use of mitochondrial DNA as a molecular marker. *Molecular Biology and*  
642 *Evolution*, *34*, 2762-2772.
- 643 Andras, J. P., Rypien, K. L., & Harvell, C. D. (2013). Range-wide population genetic structure of  
644 the Caribbean sea fan coral, *Gorgonia ventalina*. *Molecular Ecology*, *22*, 56-73.
- 645 Aranda, M., Li, Y., Liew, Y. J., Baumgarten, S., Simakov, O., Wilson, M. C., Piel, J., Ashoor,  
646 H., Bougouffa, S., Bajic, V. B., Ryu, T., Bayer, T., Micklem G., Kim, H., Bhak, J.,  
647 LaJeunesse, T., Voolstra, C., R. Genomes of coral dinoflagellate symbionts highlight  
648 evolutionary adaptations conducive to a symbiotic lifestyle. *Scientific Reports*, *6*, 39734.
- 649 Arbogast, B. S. (2001). Phylogeography: the history and formation of species. *American*  
650 *Zoologist*, *41*, 134-135.
- 651 Arnold, B., Corbett-Detig, R. B., Hartl, D., & Bomblies, K. (2013). RADseq underestimates  
652 diversity and introduces genealogical biases due to nonrandom haplotype  
653 sampling. *Molecular Ecology*, *22*, 3179-3190.
- 654 Avise, J. C. (2009). Phylogeography: retrospect and prospect. *Journal of Biogeography*, *36*, 3-  
655 15.
- 656 Avise, J. C., Bowen, B. W., & Ayala, F. J. (2016). In the light of evolution X: comparative  
657 phylogeography. *Proceedings of the National Academy of Sciences*, *113*, 7957-7961.
- 658 Baker, A. C. (2003). Flexibility and specificity in coral-algal symbiosis: diversity, ecology, and  
659 biogeography of *Symbiodinium*. *Annual Review of Ecology, Evolution, and*  
660 *Systematics*, *34*, 661-689.
- 661 Barrett, R. D. & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in*  
662 *Ecology and Evolution*, *23*, 38-44.
- 663 Baumgarten, S., Simakov, O., Esherick, L. Y., Liew, Y. J., Lehnert, E. M., Michell, C. T., Li, Y.,  
664 Hambleton, E. A., Guse, A., Oates, M. E., & Gough, J. (2015). The genome of *Aiptasia*, a  
665 sea anemone model for coral symbiosis. *Proceedings of the National Academy of*  
666 *Sciences*, *112*, 11893-11898.
- 667 Baums, I. B., Miller, M. W., & Hellberg, M. E. (2005). Regionally isolated populations of an  
668 imperiled Caribbean coral, *Acropora palmata*. *Molecular Ecology*, *14*, 1377-1390.
- 669 Bongeaerts, P., Riginos, C., Brunner, R., Englebort, N. Smith, S. R. Hoegh-Guldberg, O. (2017).  
670 Deep reefs are not universal refuges: reseeded potential varies among coral species.  
671 *Science Advances*, *3*, e1602373.
- 672 Bowen, B. W., Meylan, A. B., Ross, J. P., Limpus, C. J., Balazs, G. H., & Avise, J. C. (1992).  
673 Global population structure and natural history of the green turtle (*Chelonia mydas*) in  
674 terms of matriarchal phylogeny. *Evolution*, *46*, 865-881.

- 675 Bowen, B. W., Kamezaki, N., Limpus, C. J., Hughes, G. R., Meylan, A. B., & Avise, J. C.  
676 (1994). Global phylogeography of the loggerhead turtle (*Caretta caretta*) as indicated by  
677 mitochondrial DNA haplotypes. *Evolution*, *48*, 1820-1828.
- 678 Bowen, B. W., Shanker, K., Yasuda, N., Celia, M., Malay, M. C. M. D., von der Heyden, S.,  
679 Paulay, G., Rocha, L. A., Selkoe, K. A., Barber, P. H., & Williams, S. T. (2014).  
680 Phylogeography unplugged: comparative surveys in the genomic era. *Bulletin of Marine*  
681 *Science*, *90*, 13-46.
- 682 Bowen, B. W., Rocha, L. A., Toonen, R. J., Karl, S. A., & ToBo Lab (2013). The origins of  
683 tropical marine biodiversity. *Trends in Ecology and Evolution*, *28*, 359-366.
- 684 Bowen, B. W., Gaither, M. R., DiBattista, J. D., Iacchei, M., Andrews, K. R., Grant, W. S.,  
685 Toonen, R. J., & Briggs, J. C. (2016). Comparative phylogeography of the ocean  
686 planet. *Proceedings of the National Academy of Sciences*, *113*, 7962-7969.
- 687 Briones-Fourzán, P., Pérez-Ortiz, M., Negrete-Soto, F., Barradas-Ortiz, C., & Lozano-Álvarez,  
688 E. (2012). Ecological traits of Caribbean sea anemones and symbiotic  
689 crustaceans. *Marine Ecology Progress Series*, *470*, 55-68.
- 690 Burnham, K. P., & Anderson, D. R. (2002). Model selection and multimodal inference: A  
691 practical information theoretic approach (2nd ed.). New York, NY: Springer-Verlag.
- 692 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T.  
693 L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, *10*, 421.
- 694 Carstens, B., Lemmon, A. R., & Lemmon, E. M. (2012). The promises and pitfalls of next-  
695 generation sequencing data in phylogeography. *Systematic Biology*, *61*, 713-715.
- 696 Carstens, B. C., Pelletier, T. A., Reid, N. M., & Satler, J. D. (2013). How to fail at species  
697 delimitation. *Molecular Ecology*, *22*, 4369-4383.
- 698 Combosh, D. J., Vollmer, S. V. (2015). Trans-Pacific RAD-seq population genomics confirms  
699 introgressive hybridization in Eastern Pacific *Pocillopora* corals. *Molecular*  
700 *Phylogenetics and Evolution*, *88*, 154-162.
- 701 Daly, M., Gusmão, L. C., Reft, A. J., & Rodríguez, E. (2010). Phylogenetic signal in  
702 mitochondrial and nuclear markers in sea anemones (Cnidaria, Actiniaria). *Integrative*  
703 *and Comparative Biology*, *50*, 371-388.
- 704 Davies, S. W., Marchetti, A., Ries, J. B., & Castillo, K. D. (2016). Thermal and pCO<sub>2</sub> stress  
705 elicit divergent transcriptomic responses in a resilient coral. *Frontiers in Marine*  
706 *Science*, *3*, 112.
- 707 Davy, S. K., Allemand, D., & Weis, V. M. (2012). Cell biology of the cnidarian-dinoflagellate  
708 symbiosis. *Microbiology and Molecular Biology Reviews*, *7*, 229-261.
- 709 DeBiasse, M. B., Richards, V. P., Shivji, M.S., & Hellberg, M. E. (2016). Shared  
710 phylogeographical breaks in a Caribbean coral reef sponge and its invertebrate  
711 commensals. *Journal of Biogeography*, *43*, 2136-2146.
- 712 Devlin-Durante, M. K., & Baums, I. B. (2017). Genome-wide survey of single-nucleotide  
713 polymorphisms reveals fine-scale population structure and signs of selection in the  
714 threatened Caribbean elkhorn coral, *Acropora palmata*. *PeerJ*, *5*, e4077.
- 715 Drury, C., Schopmeyer, S., Goergen, E., Bartels, E., Nedimyer, K., Johnson, M., Maxwell, K.,  
716 Galvan, V., Manfrino, C., Lirman, D. (2017). Genomic patterns in *Acropora cervicornis*  
717 show extensive population structure and variable genetic diversity. *Ecology and*  
718 *Evolution*, *7*, 6188-6200.
- 719 Eaton, D. A. (2014). PyRAD: assembly of de novo RADseq loci for phylogenetic  
720 analyses. *Bioinformatics*, *30*, 1844-1849.

- 721 Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: a new series of programs to  
722 perform population genetics analyses under Linux and Windows. *Molecular Ecology*  
723 *Resources*, 10(3), 564-567.
- 724 Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust  
725 demographic inference from genomic and SNP data. *PLOS Genetics*, 9, e1003905.
- 726 Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals  
727 using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14, 2611-  
728 2620.
- 729 Forsman, Z. H. Coral hybridization or phenotypic variation? Genomic data reveal gene flow  
730 between *Porites lobata* and *P. compressa*. *Molecular Phylogenetics and Evolution*. 111,  
731 132-148.
- 732 Foster, N. L., Paris, C. B., Kool, J. T., Baums, I. B., Stevens, J. R., Sanchez, J. A., Bastidas, C.,  
733 Agudelo, C., Bush, P., Day, O., & Ferrari, R. (2012). Connectivity of Caribbean coral  
734 populations: complementary insights from empirical and modelled gene flow. *Molecular*  
735 *Ecology*, 21, 1143-1157.
- 736 Gates, R. D., & Edmunds, P. J. (1999). The physiological mechanisms of acclimatization in  
737 tropical reef corals. *American Zoologist*, 39, 30-43.
- 738 Grajales, A. & Rodríguez, E. (2014). Morphological revision of the genus *Aiptasia* and the  
739 family Aiptasiidae (Cnidaria, Actiniaria, Metridioidea). *Zootaxa*, 3826, 55-100.
- 740 Grajales, A., Rodríguez, E. (2016). Elucidating the evolutionary relationships of the Aiptasiidae,  
741 a widespread cnidarian–dinoflagellate model system (Cnidaria: Anthozoa: Actiniaria:  
742 Metridioidea). *Molecular Phylogenetics and Evolution*, 94, 252-263.
- 743 Grajales, A, Rodríguez, E., Thornhill, D. J. (2015). Patterns of *Symbiodinium* spp. associations  
744 within the family Aiptasiidae, a monophyletic lineage of symbiotic sea anemones  
745 (Cnidaria, Actiniaria). *Coral Reefs*, 35, 345-355.
- 746 Hammerman, N. M., Rivera-Vicens, R. E., Galaska, M. P., Weil, E., Appeldoorn, R. S., Alfaro,  
747 M., & Schizas, N. V. (2018). Population connectivity of the plating coral *Agaricia*  
748 *lamarcki* from southwest Puerto Rico. *Coral Reefs*, 37(1), 183-191.
- 749 Hellberg, M. E. (2007). Footprints on water: the genetic wake of dispersal among reefs. *Coral*  
750 *Reefs*, 26(3), 463-473.
- 751 Huang, H, Knowles, L. L. (2014). Unforeseen consequences of excluding missing data from  
752 next-generation sequences: simulation study of RAD sequences. *Systematic Biology*, 65,  
753 357-365
- 754 Huebner, L. K. & Chadwick, N. E. (2012a) Patterns of cleaning behaviour on coral reef fish by  
755 the anemoneshrimp *Ancylomenes pedersoni*. *Journal of the Marine Biological*  
756 *Association of the United Kingdom* 92, 1557-1562.
- 757 Huebner, L. K. & Chadwick, N. E. (2012b) Reef fishes use sea anemones as visual cues for  
758 cleaning interactions with shrimp. *Journal of Experimental Marine Biology &*  
759 *Ecology*. 416, 237-242.
- 760 Hughes, T.P., Kerry, J.T., Álvarez-Noriega, M., Álvarez-Romero, J.G., Anderson, K.D., Baird,  
761 A.H., Babcock, R.C., Beger, M., Bellwood, D.R., Berkelmans, R. & Bridge, T.C. (2017).  
762 Global warming and recurrent mass bleaching of corals. *Nature*, 543(7645), p.373.
- 763 ICZN, I., 2017. Opinion 2404 (Case 3633)–*Dysactis pallida* Agassiz in Verrill, 1864 (currently  
764 *Aiptasia pallida*; Cnidaria: Anthozoa: Hexacorallia: Actiniaria): precedence over *Aiptasia*  
765 *diaphana* (Rapp, 1829), *Aiptasia tagetes* (Duchassaing de Fombressin & Michelotti,  
766 1864), *Aiptasia mimosa* (Duchassaing de Fombressin & Michelotti, 1864) and *Aiptasia*



- 767 *inula* (Duchassaing de Fombressin & Michelotti, 1864) not approved. *Bulletin of*  
768 *Zoological Nomenclature*, 72, 130-131.
- 769 Jennison, B. L. (1981). Reproduction in three species of sea anemones from Key West,  
770 Florida. *Canadian Journal of Zoology*, 59, 1708-1719.
- 771 Johnston, E. C., Forsman, Z. H., Flot, J. F., Schmidt-Roach, S., Pinzon, J. H., Knapp, I. S. S.,  
772 Toonen, R. J. (2017). A genomic glance through the fog of plasticity and diversification in  
773 *Pocillopora*. *Scientific Reports*, 7, 5991.
- 774 Kenkel, C. D., & Matz, M. V. (2017). Gene expression plasticity as a mechanism of coral  
775 adaptation to a variable environment. *Nature Ecology & Evolution*, 1, 0014.
- 776 Kenkel, C. D., Moya, A., Strahl, J., Humphrey, C., & Bay, L. K. (2018). Functional genomic  
777 analysis of corals from natural CO<sub>2</sub>-seeps reveals core molecular responses involved in  
778 acclimatization to ocean acidification. *Global Change Biology*, 24, 158-171.
- 779 Knowles, L. L. (2009). Statistical phylogeography. *Annual Review of Ecology, Evolution, and*  
780 *Systematics*, 40, 593-612.
- 781 LaJeunesse, T.C., Parkinson, J.E., Gabrielson, P.W., Jeong, H.J., Reimer, J.D., Voolstra, C.R. &  
782 Santos, S.R. (2018). Systematic Revision of Symbiodiniaceae Highlights the Antiquity  
783 and Diversity of Coral Endosymbionts. *Current Biology*.  
784 <https://doi.org/10.1016/j.cub.2018.07.008>
- 785 Leydet, K. P., Grupstra, C. G., Coma, R., Ribes, M., & Hellberg, M. E. (2018). Host-targeted  
786 RAD-Seq reveals genetic changes in the coral *Oculina patagonica* associated with range  
787 expansion along the Spanish Mediterranean coast. *Molecular Ecology*, 27, 2529-2543.
- 788 Liu, H., Stephens, T. G., Gonzalez-Pech, R.A., Beltran, V. H., Lapeyre, B., Bongaerts, P.,  
789 Cooke, I., Aranda, M., Bourne, D. G., Foret, S., & Miller, D. J. (2018). *Symbiodinium*  
790 genomes reveal adaptive evolution of functions related to coral-dinoflagellate symbiosis.  
791 *Nature Communications Biology*, 1, 95
- 792 Ludt, W. B., & Rocha, L. A. (2015). Shifting seas: the impacts of Pleistocene sea-level  
793 fluctuations on the evolution of tropical marine taxa. *Journal of Biogeography*, 42(1), 25-  
794 38.
- 795 McCormack, J. E., Hird, S. M., Zellmer, A. J., Carstens, B. C., & Brumfield, R. T. (2013).  
796 Applications of next-generation sequencing to phylogeography and  
797 phylogenetics. *Molecular Phylogenetics and Evolution*, 66, 526-538.
- 798 Muscatine, L., R. McCloskey, L., & E. Marian, R. (1981). Estimating the daily contribution of  
799 carbon from zooxanthellae to coral animal respiration. *Limnology and Oceanography*, 26,  
800 601-611.
- 801 Palumbi, S. R., Barshis, D. J., Traylor-Knowles, N. & Bay, R.A., 2014. Mechanisms of reef coral  
802 resistance to future climate change. *Science*, 344, 895-898.
- 803 Pante, E., Puillandre, N., Viricel, A., Arnaud-Haond, S., Aurelle, D., Castelin, M., Chenuil, A.,  
804 Destombe, C., Forcioli, D., Valero, M. & Viard, F. (2015). Species are hypotheses: avoid  
805 connectivity assessments based on pillars of sand. *Molecular Ecology*, 24, 525-544.
- 806 Pelletier, T. A., & Carstens, B. C. (2014). Model choice for phylogeographic inference using a  
807 large set of models. *Molecular Ecology*, 23, 3028-3043.
- 808 Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using  
809 multilocus genotype data. *Genetics*, 155, 945-959.
- 810 Prada, C., Hanna, B., Budd, A. F., Woodley, C. M., Schmutz, J., Grimwood, J., Iglesias-Prieto,  
811 R., Pandolfi, J. M., Levitan, D., Johnson, K. G., & Knowlton, N. (2016). Empty niches

- 812 after extinctions increase population sizes of modern corals. *Current Biology*, 26, 3190-  
813 3194.
- 814 Putnam, N. H., Srivastava, M., Hellsten, U., Dirks, B., Chapman, J., Salamov, A., Terry, A.,  
815 Shapiro, H., Lindquist, E., Kapitonov, V. V. & Jurka, J. (2007). Sea anemone genome  
816 reveals ancestral eumetazoan gene repertoire and genomic organization. *Science*, 317, 86-  
817 94.
- 818 Prezeworski, M., Coop, G., & Wall, J. D. (2005). The signature of positive selection on standing  
819 genetic variation. *Evolution*, 59, 2312-2323.
- 820 O'Reilly, E. E., Chadwick, N. E. (2017). Population dynamics of corkscrew sea anemones  
821 *Bartholomea annulata* in the Florida Keys. *Marine Ecology Progress Series*, 567, 109-  
822 123.
- 823 O'Reilly, E. E., Titus, B. M., Nelsen, M. W., Ratchford, S., Chadwick, N. E. (In Press). Giant  
824 ephemeral anemones? Rapid growth and high mortality of corkscrew sea anemones  
825 *Bartholomea annulata* (Le Sueur, 1817) under variable conditions. *Journal of*  
826 *Experimental Marine Biology and Ecology*.
- 827 Reeb, C. A., & Avise, J. C. (1990). A genetic discontinuity in a continuously distributed species:  
828 mitochondrial DNA in the American oyster, *Crassostrea virginica*. *Genetics*, 124, 397-  
829 406.
- 830 Reid, N. M., Proestou, D. A., Clark, B. W., Warren, W. C., Colbourne, J. K., Shaw, J. R.,  
831 Karchner, S. I., Hahn, M. E., Nacci, D., Oleksiak, M.F., & Crawford, D.L. (2016). The  
832 genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild  
833 fish. *Science*, 354, 1305-1308.
- 834 Rippe, J. P., Matz, M. V., Green, E. A., Medina, M., Khawaja, N. Z., Pongwarin, T., Pinzón C, J.  
835 H., Castillo, K. D., & Davies, S. W., 2017. Population structure and connectivity of the  
836 mountainous star coral, *Orbicella faveolata*, throughout the wider Caribbean  
837 region. *Ecology and Evolution*, 7, 9234-9246.
- 838 Rosser, N. L., Thomas, L., Stankowski, S., Richards, Z. T., Kennington, W.J., Johnson, M. S.  
839 (2017). Phylogenomics provides new insight into evolutionary relationships and  
840 genealogical discordance in the reef-building coral genus *Acropora*. *Proceedings of the*  
841 *Royal Society B-Biological Sciences*. 284, 20162182.
- 842 Rowan, R. O. B., & Powers, D. A. (1991). A molecular genetic classification of zooxanthellae  
843 and the evolution of animal-algal symbioses. *Science*, 251, 1348-1351.
- 844 Rowan, R., & Powers, D. A. (1992). Ribosomal RNA sequences and the diversity of symbiotic  
845 dinoflagellates (zooxanthellae). *Proceedings of the National Academy of Sciences*, 89,  
846 3639-3643.
- 847 Santos, S. R. (2016). From one to many: the population genetics of cnidarian-*Symbiodinium*  
848 symbioses. In *The Cnidaria, Past, Present and Future* (pp. 359-373). Springer, Cham.
- 849 Satler, J. D., & Carstens, B.C. (2017). Do ecological communities disperse across biogeographic  
850 barriers as a unit? *Molecular Ecology*, 26, 3533-3545.
- 851 Shearer, T. L., Van Oppen, M. J. H., Romano, S. L., & Wörheide, G. (2002). Slow mitochondrial  
852 DNA sequence evolution in the Anthozoa (Cnidaria). *Molecular Ecology*, 11, 2475-2487.
- 853 Shearer, T. L., Gutiérrez-Rodríguez, C., & Coffroth, M. A. (2005). Generating molecular  
854 markers from zooxanthellate cnidarians. *Coral Reefs*, 24, 57-66.
- 855 Shinzato, C., Mungpakdee, S., Arakaki, N., & Satoh, N. (2015). Genome-wide SNP analysis  
856 explains coral diversity and recovery in the Ryukyu Archipelago. *Scientific Reports*, 5,  
857 18211.

- 858 Shoguchi, E. et al. (2013). Draft assembly of the *Symbiodinium minutum* nuclear genome reveals  
859 dinoflagellate gene structure. *Current Biology*, 23, 1399–1408. ADD other authors  
860 Silverstein, R. N., Correa, A. M., & Baker, A. C. (2012). Specificity is rarely absolute in coral–  
861 algal symbiosis: implications for coral response to climate change. *Proceedings of the*  
862 *Royal Society of London B: Biological Sciences*, 279, 2609–2618.
- 863 Smith, M. L., Ruffley, M., Espíndola, A., Tank, D. C., Sullivan, J., & Carstens, B. C. (2017).  
864 Demographic model selection using random forests and the site frequency  
865 spectrum. *Molecular Ecology*, 26, 4562–4573.
- 866 Thomas, L., Kennington, W. J., Evans, R. D., Kendrick, G. A., & Stat, M. (2017). Restricted  
867 gene flow and local adaptation highlight the vulnerability of high-latitude reefs to rapid  
868 environmental change. *Global Change Biology*, 23, 2197–2205.
- 869 Titus, B. M. & Daly, M. (2017). Specialist and generalist symbionts show counterintuitive levels  
870 of genetic diversity and discordant demographic histories along the Florida Reef Tract.  
871 *Coral Reefs*, 36, 339–354.
- 872 Titus, B. M., Daly, M., & Exton, D. A. (2015a). Do reef fish habituate to diver presence?  
873 Evidence from two reef sites with contrasting historical levels of SCUBA intensity in the  
874 Bay Islands, Honduras. *PLOS ONE*. 10(3): e0119645.
- 875 Titus, B. M., Daly, M., & Exton, D. A. (2015b). Temporal patterns of Pederson shrimp  
876 (*Ancylomenes pedersoni* Chace 1958) cleaning interactions of Caribbean coral reefs.  
877 *Marine Biology*, 162, 1651–1664.
- 878 Titus, B. M., Daly, M., Macrander, J., Del Rio, A., Santos, S. R., & Chadwick, N. E. (2017a).  
879 Contrasting abundance and contribution of clonal proliferation to the population structure  
880 of the corkscrew sea anemone *Bartholomea annulata* in the tropical Western Atlantic.  
881 *Invertebrate Biology*, 136: 62–74.
- 882 Titus, B. M. Vondriska, C., & Daly, M. (2017b). Comparative behavioral observations  
883 demonstrate the spotted “cleaner” shrimp *Periclimenes yucatanicus* engages in true  
884 symbiotic cleaning interactions. *Royal Society Open Science*, 4:170078.
- 885 Titus, B. M., Palombit, S., & Daly, M. (2017c). Endemic diversification in an isolated  
886 archipelago with few endemics: an example from a cleaner shrimp species complex in the  
887 Tropical Western Atlantic. *Biological Journal of the Linnaean Society*, 122: 98–112.
- 888 Titus, B. M., Blischak, P. D., & Daly, M. (2018). Genomic signatures of sympatric speciation  
889 with historical and contemporary gene flow in a tropical anthozoan. *bioRxiv* 399360 doi:  
890 <https://doi.org/10.1101/399360>.
- 891 Voolstra, C. R., Li, Y., Liew, Y. J., Baumgarten, S., Zoccola, D., Flot, J. F., Tambutté, S.,  
892 Allemand, D. & Aranda, M., 2017. Comparative analysis of the genomes of *Stylophora*  
893 *pistillata* and *Acropora digitifera* provides evidence for extensive differences between  
894 species of corals. *Scientific Reports*, 7, 17583.
- 895 Wang, X., Liew, Y. J., Li, Y., Zoccola, D., Tambutte, S., & Aranda, M. (2017). Draft genomes of  
896 the corallimorpharians *Amplexidiscus fenestrafer* and *Discosoma* sp. *Molecular Ecology*  
897 *Resources*, 17, e187–e195.
- 898

899

900

901 **Tables**

902 Table 1. Sample localities, sample sizes, and geographic coordinates of corkscrew sea anemone  
903 *Bartholomea annulata* used in this study. Sample sizes reflect the number of samples sequenced  
904 and the number of samples retained in the final double digest Restriction-site Associated DNA  
905 sequencing (ddRADseq) dataset after accounting for low sequence reads and cryptic species  
906 diversity (in parentheses). Differences between the number of samples sequenced and retained  
907 reflects variation in the number of sequence reads and sequencing coverage in our ddRADseq  
908 dataset across all individuals.

Locality	Code	Sample sizes	Latitude	Longitude
Eleuthera, Bahamas	BH	10 (7)	24°49'44.51"N	76°16'46.11"W
San Salvador, Bahamas	SAN	10 (2)	24° 2'37.12"N	74°31'59.83"W
Barbados	BR	10 (8)	13°11'30.52"N	59°38'29.04"W
Bermuda	BD	12 (9)	32°26'53.62"N	64°45'45.42"W
Curacao	CU	10 (0)	12° 7'19.45"N	68°58'10.80"W
Ft. Lauderdale, Florida, USA	FT	9 (8)	26° 4'19.80"N	80° 5'46.68"W
Upper Keys, Florida, USA	UK	10 (9)	25° 1'57.92"N	80°22'4.45"W
Middle Keys, Florida, USA	MK	10 (7)	24°41'58.09"N	80°56'21.48"W
Lower Keys, Florida, USA	LK	10 (9)	24°33'42.39"N	81°23'31.59"W
Utila, Honduras	UT	10 (10)	16° 5'18.03"N	86°54'38.54"W
Cayos Cochinos, Honduras	CC	10 (9)	15°57'1.12"N	86°29'51.82"W
Mahahual, Mexico	MX	10 (9)	18°42'18.45"N	87°42'34.46"W
Bocas del Toro, Panama	PA	10 (3)	9°25'7.28"N	82°20'32.55"W
St. Thomas, US Virgin Islands	VI	11 (11)	18°19'0.69"N	64°59'22.59"W
<b>Total</b>		141(101)		

909

910

911

912

913

914 Table 2. Analysis of Molecular Variance (AMOVA) results for aposymbiotic and holobiont  
915 *Bartholomea annulata* RADseq datasets. Data were partitioned into Eastern and Western  
916 Regions and reflect nearly identical levels of genetic diversity partitioned at all hierarchical  
917 levels. \*\*p < 0.0005; \*p < 0.005.

<b>Dataset</b>	<b>Among localities</b>	$\varphi_{ST}$	<b>Among localities within regions</b>	$\varphi_{SC}$	<b>Among regions</b>	$\varphi_{CT}$
Aposymbiotic	96.55	0.03**	2.24	0.02**	1.21	0.01*
Holobiont	95.90	0.04**	2.62	0.03**	1.48	0.01*

918

919

920

921

922

923

924

925

926

927

928

929 Table 3. Pairwise  $\varphi_{ST}$  calculated among sample localities for aposymbiotic (below diagonal) and  
 930 holobiont (above diagonal) *Bartholomea annulata* datasets. Locality codes correspond to those in  
 931 Figure 1.  $\varphi_{ST}$  values highlighted and bolded are significant at  $p < 0.05$ . ^denotes  $\varphi_{ST}$  values  
 932 significant in the holobiont dataset only

	PA	CC	UT	MX	FT	UK	MK	LK	BH	SAN	VI	BR	BD
PA	-	<b>0.05</b>	<b>0.03</b>	<b>0.08</b>	<b>0.05</b>	<b>0.03</b>	<b>0.04</b>	<b>0.04</b>	<b>0.05</b>	<b>0.03</b>	<b>0.03</b>	<b>0.03<sup>^</sup></b>	<b>0.03</b>
CC	<b>0.05</b>	-	<b>0.05</b>	0.04	<b>0.04</b>	<b>0.04</b>	<b>0.03</b>	<b>0.07</b>	<b>0.07</b>	<b>0.03</b>	<b>0.03</b>	<b>0.03</b>	<b>0.05</b>
UT	<b>0.03</b>	<b>0.04</b>	-	<b>0.08</b>	<b>0.04</b>	<b>0.03</b>	<b>0.05</b>	<b>0.05</b>	<b>0.05</b>	<b>0.04</b>	<b>0.05</b>	<b>0.03</b>	<b>0.05</b>
MX	<b>0.06</b>	0.02	<b>0.07</b>	-	<b>0.07</b>	<b>0.05<sup>^</sup></b>	<b>0.09</b>	<b>0.09</b>	<b>0.08</b>	<b>0.07</b>	<b>0.08</b>	<b>0.05<sup>^</sup></b>	<b>0.07</b>
FT	<b>0.05</b>	<b>0.03</b>	<b>0.04</b>	<b>0.06</b>	-	<b>0.03</b>	<b>0.06</b>	<b>0.06</b>	<b>0.06</b>	<b>0.04</b>	<b>0.03<sup>^</sup></b>	<b>0.04</b>	<b>0.05</b>
UK	<b>0.03</b>	<b>0.04</b>	<b>0.02</b>	0.03	<b>0.02</b>	-	<b>0.03</b>	<b>0.02</b>	<b>0.03</b>	0.00	0.00	<b>0.01<sup>^</sup></b>	<b>0.02</b>
MK	<b>0.03</b>	<b>0.02</b>	<b>0.03</b>	<b>0.07</b>	<b>0.06</b>	<b>0.02</b>	-	<b>0.02<sup>^</sup></b>	<b>0.03</b>	0.00	0.02	0.01	<b>0.02</b>
LK	<b>0.03</b>	<b>0.06</b>	<b>0.04</b>	<b>0.07</b>	<b>0.06</b>	<b>0.02</b>	0.01	-	<b>0.03</b>	0.00	0.02	<b>0.02</b>	<b>0.02</b>
BH	<b>0.04</b>	<b>0.06</b>	<b>0.05</b>	<b>0.05</b>	<b>0.06</b>	0.02	<b>0.02</b>	<b>0.03</b>	-	0.01	0.02	<b>0.02</b>	<b>0.03</b>
SAN	<b>0.02</b>	<b>0.02</b>	<b>0.03</b>	<b>0.07</b>	<b>0.04</b>	0.00	0.00	0.00	0.00	-	<b>0.06</b>	0.00	0.00
VI	0.02	<b>0.04</b>	<b>0.05</b>	<b>0.04</b>	0.04	0.00	0.00	0.01	0.02	<b>0.07</b>	-	0.00	0.00
BR	<b>0.02</b>	<b>0.03</b>	<b>0.02</b>	0.04	<b>0.03</b>	0.01	0.01	<b>0.02</b>	<b>0.02</b>	0.00	0.00	-	<b>0.02<sup>^</sup></b>
BD	<b>0.03</b>	<b>0.04</b>	<b>0.03</b>	<b>0.05</b>	<b>0.04</b>	<b>0.01</b>	<b>0.01</b>	<b>0.02</b>	<b>0.03</b>	0.00	0.02	0.01	-

933

934

935

936

937

938

939

940

941 Table 4. Diversity indices calculated from aposymbiotic and holobiont (in parentheses) RADseq  
 942 data for *Bartholomea annulata* across the Tropical Western Atlantic. Diversity indices calculated  
 943 for each sample locality and for each genetically defined population grouping determined by  
 944 Structure. Values reflect nearly identical genetic diversity indices between aposymbiotic and  
 945 holobiont datasets all sample localities.  $N_G$  = Number of gene copies, SS = segregating sites,  $\theta$  =  
 946 theta calculated from segregating sites,  $\pi$  = nucleotide diversity, Ho = observed heterozygosity,  
 947 He = expected heterozygosity.

Sample locality	Code	$N_G$	SS	$\theta$	$\pi$	Ho	He
Eleuthera	BH	14	169 (169)	53.1 (53.1)	0.030 (0.028)	0.20 (0.18)	0.21 (0.19)
San Salvador	SAN	4	93 (84)	50.7 (45.8)	0.033 (0.029)	0.45 (0.48)	0.52 (0.51)
Barbados	BR	16	166 (140)	50.02 (42.1)	0.031 (0.026)	0.17 (0.17)	0.19 (0.18)
Bermuda	BD	18	197 (177)	57.7 (51.4)	0.035 (0.032)	0.15 (0.15)	0.17 (0.16)
Curacao	CU	-	-	-	-	-	-
Ft. Lauderdale, Florida, USA	FT	16	206 (178)	62.08 (53.6)	0.031 (0.026)	0.17 (0.15)	0.18 (0.16)
Upper Keys, Florida, USA	UK	18	221 (198)	64.2 (57.5)	0.03 (0.028)	0.15 (0.14)	0.16 (0.15)
Middle Keys, Florida, USA	MK	14	177 (151)	55.6 (47.4)	0.029 (0.027)	0.18 (0.19)	0.20 (0.20)
Lower Keys, Florida, USA	LK	18	198 (169)	57.56 (49.3)	0.033 (0.026)	0.17 (0.16)	0.18 (0.16)
Cayos Cochinos, Honduras	CC	18	194 (169)	56.4 (49.1)	0.029 (0.025)	0.15 (0.14)	0.16 (0.15)
Utila, Honduras	UT	20	179 (163)	50.4 (45.9)	0.027 (0.042)	0.13 (0.12)	0.14 (0.13)
Mexico	MX	18	152 (148)	44.1 (43.0)	0.033 (0.032)	0.14 (0.14)	0.16 (0.15)
Panama	PA	6	68 (59)	29.8 (25.8)	0.025 (0.022)	0.33 (0.35)	0.39 (0.38)
US Virgin Islands	VI	22	152 (171)	41.6 (46.9)	0.026 (0.027)	0.14 (0.13)	0.15 (0.13)

948

949

950 Table 5. Akaike Information Criterion results for model selection from FSC2 for the  
 951 aposymbiotic and holobiont (in parentheses) *Bartholomea annulata* datasets. Model rank was  
 952 identical between aposymbiotic and holobiont datasets, with broadly similar model likelihoods  
 953 and model weights. Model refers to those depicted and described in Figure 2.  $k$  = number of  
 954 parameters in the model, AIC = Akaike Information Criterion,  $\Delta_i$  = change in AIC scores, and  $w_i$   
 955 = Akaike weights. Models are listed according to their AIC rank and the highest ranked model is  
 956 highlighted.

Model	$k$	ln(Likelihood)	AIC	$\Delta_i$	Model Likelihoods	$w_i$
6 - IM <sub>EW</sub>	5	-8846.5 (-19863.5)	17703.1 (39737.0)	0 (0)	1 (1)	0.75 (0.73)
10 - IM <sub>ISO-MIG-WE</sub>	6	-8847.3 (-19863.5)	17706.7 (39739.0)	3.5 (2.0)	0.17 (0.36)	0.12 (0.26)
5 - IM <sub>WE</sub>	5	-8848.4 (-19870.0)	17706.9 (39750.1)	3.7 (13.1)	0.15 (0.001)	0.11 (0.001)
3 - IM	6	-8853.0 (-19877.1)	17718.0 (39766.3)	14.8 (29.3)	$6.0e^{-3}$ ( $4.2e^{-7}$ )	$4.0e^{-3}$ ( $3.1e^{-7}$ )
8 - IM <sub>ISO-MIG</sub>	7	-8857.7 (-19844.6)	17729.5 (39783.3)	26.3 (46.3)	$1.8e^{-6}$ ( $8.6e^{-11}$ )	$1.4e^{-6}$ ( $6.3e^{-11}$ )
1 - ISO	4	-8868.3 (-19922.8)	17744.6 (39853.7)	41.4 (116.7)	$1.0e^{-9}$ ( $4.5e^{-26}$ )	$7.7e^{-10}$ ( $3.3e^{-26}$ )
12 - IM <sub>ISO-MIG-EW</sub>	6	-8885.5 (-19963.2)	17783.0 (39938.4)	79.8 (201.3)	$4.6e^{-18}$ ( $1.8e^{-44}$ )	$3.4e^{-18}$ ( $1.3e^{-44}$ )
2 - ISOc	11	-10965.6 (-22778.0)	21953.3 (45578.1)	4250.0 (5841.1)	0 (0)	0 (0)
4 - IMc	12	-11569.8 (-24493.4)	23163.7 (49010.9)	5460.5 (9272.9)	0 (0)	0 (0)
9 - IM <sub>MIG-WE-ISO</sub>	6	-13772.4 (-30574.7)	27556.8 (61161.5)	9853.6 (21424.4)	0 (0)	0 (0)
7 - IM <sub>MIG-ISO</sub>	7	-13772.0 (-30576.1)	27558.1 (61166.2)	9854.9 (21429.1)	0 (0)	0 (0)
11 - IM <sub>MIG-EW-ISO</sub>	6	-13947.7 (-30985.4)	27907.5 (61982.8)	10204.3 (22245.8)	0 (0)	0 (0)

957



958 Table 6. Parameter estimates and 95% confidence intervals (CI) generated from FSC2 coalescent  
 959 simulations for aposymbiotic (Aposym) and holobiont (Holo) *Bartholomea annulata* datasets.  $N_e$   
 960 = effective population size,  $\tau$  = divergence time,  $Mig_{EW}$  = migration rate from Eastern to Western  
 961 populations. Values reported for  $N_e$  are in number of individuals and the values for  $\tau$  are reported  
 962 in years before present. Parameter values reflect overlapping confidence intervals between  
 963 aposymbiotic and holobiont for every parameter calculated.

<b>Dataset</b>	<b><math>N_e</math> Ancestral</b>		<b><math>N_e</math> West</b>		<b><math>N_e</math> East</b>		<b><math>\tau</math></b>		<b><math>Mig_{EW}</math></b>	
	Mean	95% CI	Mean	95% CI	Mean	95% CI	Mean	95% CI	Mean	95% CI
<b>Aposym</b>	38,179	(± 6058)	61,787	(± 5465)	100,290	(± 5339)	39,274	(± 11,079)	1.50e <sup>-4</sup>	(± 1.24 e <sup>-5</sup> )
<b>Holo</b>	33,273	(± 9477)	68,137	(± 5302)	118,113	(± 10,619)	58,140	(± 14,884)	1.25e <sup>-4</sup>	(± 9.46 e <sup>-6</sup> )

964

965