bioRxiv preprint doi: https://doi.org/10.1101/161117; this version posted July 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 **Revisiting the phylogeny of Zoanthidea (Cnidaria: Anthozoa): staggered alignment of**
2 **hypervariable sequences improves species tree inference**
3
4 Timothy D. Swain
5 Integrative Research Center, Field Museum of Natural History, Chicago, IL 60605, USA
6 Department of Civil and Environmental Engineering, Northwestern University, Evanston, IL
7 60208, USA
8 Email: tswain@fieldmuseum.org
9 Phone: 312-665-7756
10 Fax: 312-665-7516
11
12 **Abstract**
13 The recent rapid proliferation of novel taxon identification in the Zoanthidea has been
14 accompanied by a parallel propagation of gene trees as a tool of species discovery, but not a
15 corresponding increase in our understanding of phylogeny. This disparity is caused by the trade-
16 off between the capabilities of automated DNA sequence alignment and data content of genes
17 applied to phylogenetic inference in this group. Conserved genes or segments are easily aligned
18 across the order, but produce poorly resolved trees; hypervariable genes or segments contain the
19 evolutionary signal necessary for resolution and robust support, but sequence alignment is
20 daunting. Staggered alignments are a form of phylogeny-informed sequence alignment
21 composed of a mosaic of local and universal regions that allow phylogenetic inference to be
22 applied to all nucleotides from both hypervariable and conserved gene segments. Comparisons
23 between species tree phylogenies inferred from all data (staggered alignment) and hypervariable-

bioRxiv preprint doi: https://doi.org/10.1101/161117; this version posted July 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

24 excluded data (standard alignment) demonstrate improved confidence and greater topological
25 agreement with other sources of data for the complete-data tree. This novel phylogeny is the
26 most comprehensive to date (in terms of taxa and data) and can serve as an expandable tool for
27 evolutionary hypothesis testing in the Zoanthidea.

28

29 **Resumen**

30 Spanish language translation by Lisbeth O. Swain, DePaul University, Chicago, Illinois, 60604,
31 USA.

32 Aunque la proliferación reciente y acelerada en la identificación de taxones en
33 Zoanthidea ha sido acompañada por una propagación paralela de los árboles de genes como una
34 herramienta en el descubrimiento de especies, no hay una correspondencia en cuanto a la
35 ampliación de nuestro conocimiento en filogenia. Esta disparidad, es causada por la competencia
36 entre la capacidad de los alineamientos de secuencia del ácido desoxirribonucleico (ADN)
37 automatizados y la información contenida en los datos de genes que se aplican a los métodos de
38 inferencia filogenética en este grupo de Zoanthidea. Las regiones o segmentos de genes
39 conservados son fácilmente alineados dentro del orden; sin embargo, producen árboles de genes
40 con resultados paupérrimos; además, aunque estas regiones hipervariables de genes o segmentos
41 contienen las señas evolutivas necesarias para apoyar la construcción robusta y completa de
42 árboles filogenéticos, estos genes producen alineamientos de secuencia abrumadores. Los
43 alineamientos escalonados de secuencias son una forma de alineamientos informados por la
44 filogenia y compuestos de un mosaico de regiones locales y universales que permiten que
45 inferencias filogenéticas sean aplicadas a todos los nucleótidos de regiones hipervariables y de
46 genes o segmentos conservados. Las comparaciones entre especies de árboles filogenéticos que

bioRxiv preprint doi: https://doi.org/10.1101/161117; this version posted July 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

47 se infirieron de los datos de alineamientos escalonados y los datos hipervariables excluidos
48 (alineamiento estandarizado), demuestran un mejoramiento en la confiabilidad y un mayor
49 acuerdo tipológico con respecto a otras fuentes que contienen árboles filogenéticos hechos de
50 datos más completos. Esta nueva forma escalonada de filogenia es una de los más compresibles
51 hasta la fecha (en términos de taxones y datos) y que pueden servir como una herramienta de
52 amplificación para probar la hipótesis evolutiva de Zoanthidea.
53 **Keywords:** gene tree, hypervariable sequences, phylogeny-informed alignment, species tree,
54 staggered alignment
55

56 **1. Introduction**

57 Nucleotide sequence-based molecular phylogenetics have been intensively applied to the
58 Anthozoa order Zoanthidea, however our understanding of the evolutionary relationships among
59 zoanthidean species has not progressed at the same pace. Since 2004, at least 107 phylogenetic
60 trees that focus mostly or exclusively on Zoanthidea have been published in 46 reports (a rate of
61 nearly 9 trees and 4 reports per year; Table 1). Within these trees, the mean number of species
62 per tree is only 16.2, while the mean number of terminals is nearly three times as large, with
63 some published trees containing <5% unique species among their terminals (Table 1).
64 Furthermore, of the 107 trees identified, 100 (or 94%) are gene trees relying upon a single gene
65 or gene fragment rather than attempted species trees (see section 3.1), and 30% of these gene
66 trees are built solely on data from the mitochondrial cytochrome oxidase subunit I gene (COI;
67 Table 1). It has been known since at least 2002 that the rate of evolution of COI within Anthozoa
68 is insufficient for distinguishing between closely related species and is largely uninformative for
69 addressing phylogenetic questions below the family or genus-level (Shearer et al., 2002). The

bioRxiv preprint doi: https://doi.org/10.1101/161117; this version posted July 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

70 details of these published trees suggest that the primary goal of phylogenetic research on this
71 order could not be the relationships among species, but is more likely the exploration of gene
72 evolution among specimens (see section 3.1). However, these gene trees have been generally
73 used in species delimitation and are subsequently over-interpreted as species trees and molecular
74 parataxonomic evidence of evolutionary relationships among species and higher taxa in the
75 nearly complete absence of other data (see Swain et al., 2015; Swain et al., 2016; Swain and
76 Swain, 2014 for further discussion of molecular parataxonomy).

77 Of the seven published zoanthidean trees that rely upon more than a single gene to
78 support its inferences, only three (or 2.8% of the total) include attempted species trees based on
79 analysis of concatenations (See section 3.1) of more than two genes originating from more than
80 one genomic compartment (Table 1); even though this higher-level of genomic sampling (i.e.,
81 multiple genes from multiple genomic compartments) is the standard minimum practice in
82 molecular phylogenetics and is rapidly being overshadowed by the use of hundreds of genes or
83 whole genomes as the basis of phylogentic reconstructions (phylogenomics). These three
84 analyses include the 6 gene, 93 species (representing 11 genera across 5 families) tree of Swain
85 (2010), the 3 gene, 29 species (representing 12 genera across 3 families) tree of Sinniger et al.
86 (2013), and the 5 gene, 48 species (representing 14 genera across 2 families) tree of Montenegro
87 et al (2016) (Table 1). None of these analyses are as comprehensive as is currently possible
88 (Table S1), both in terms of data available for taxa (at least 144 species, 25 genera, 9 families)
89 and genes (at least 6 genes). Additionally, analyses that include genes with divergent
90 hypervariable regions (e.g. mitochondrial 16S or nuclear ITS; see sections 3.2–3.3) either ignore
91 these data (e.g., Montenegro et al., 2016) or limit the scope of research to closely related species
92 (e.g., Sinniger et al., 2013) to eliminate difficult homology assessments (but, see Swain, 2009,

bioRxiv preprint doi: https://doi.org/10.1101/161117; this version posted July 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

93 2010). With the spectacular proliferation of gene trees and discarded datasets (see sections 3.2–
94 3.3), along with poor representation of species and higher taxa (less than 65% of species and
95 56% of genera and families for which there is available sequence for at least two genes are
96 represented in any one of the previously published phylogenies), a comprehensive revision of the
97 phylogeny of Zoanthidea seems warranted. Here, I present an updated comprehensive phylogeny
98 for Zoanthidea and demonstrate that the use of a staggered sequence alignment to retain the
99 evolutionary signal contained in hypervariable genes results in a more robust and defensible
100 topology.
101

102 **2. Materials and Methods**

103 The phylogeny of Zoanthidea was revised to be current and comprehensive through
104 maximum-likelihood inference applied to a concatenated multi-gene, multi-genomic alignment
105 (see section 3.1). Nucleotide sequences of all unique Zoanthidea taxa, for which at least two
106 different genes were available from GenBank, were included in the analysis (Table S1). Several
107 of these genes contain hypervariable regions that are commonly used to differentiate species, but
108 are also usually discarded for phylogenetic inference because of challenging homology
109 assessments (see section 3.2). The effect of including hypervariable regions in tree inference (see
110 section 3.3) was assessed through paired analyses: beginning with a reconstruction based on a
111 staggered alignment that included every available nucleotide (see section 3.2, Fig. 1), followed
112 by a reconstruction based on the identical alignment with an exclusion set to remove
113 hypervariable positions and retain only conserved, universally alignable regions.

114 *2.1 Sampling strategy*

115 All Zoanthidea taxa with at least two different nucleotide sequences available in

bioRxiv preprint doi: https://doi.org/10.1101/161117; this version posted July 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

116 GenBank for the six most commonly applied genes were included in the analyses. These genes
117 include nuclear 18S, ITS, and 28S ribosomal RNA (rRNA) of the internal transcribed spacer
118 (ITS) region and mitochondrial 12S and 16s rRNA and protein-coding COI (Table S1).
119 Sequences were compiled by first targeting additional accessions available for the taxa used in
120 the alignment of Swain (2010) (TreeBASE: S10492), and then expanding to include additional
121 available taxa (that met the above criteria). Outgroup taxa were selected following Rodríguez et
122 al. (2014) to include closely related *Relicanthus daphneae* and Antipatharia species and more
123 distantly related Actiniaria species.

124 Identification of archived nucleotide sequences originating from unique species of
125 Zoanthidea is challenging because of the rapid evolution of current taxonomic designations, the
126 common practice of including unnamed and undescribed specimens without unique identifiers in
127 previous phylogenetic analyses, use of molecular parataxonomy uncoupled from accepted
128 species concepts for naming new taxa (see Mayden 1997 for a review of species concepts), and
129 misidentification of taxa or substitution of contaminate sequences (in place of targeted taxa)
130 within nucleotide databases (see Swain, 2010; Swain et al., 2015; Swain et al., 2016; Swain and
131 Swain, 2014; and Table S1 for details of name usage tracking and examples of mislabeled
132 accession identification). As a result, a simple search of GenBank for Zoanthidea will return a
133 hodgepodge collection of intractable or incorrect identification labels attached to > 2,800
134 accessions (last accessed on May 30, 2016; http://www.ncbi.nlm.nih.gov/nuccore).

135 The only way to correctly identify which nucleotide sequences originate from the same
136 specimens, and identify all available unique species (that meet the minimum genomic sampling
137 detailed above), is to follow the usage of individual sequences through the 46 publications (and
138 their associated supplemental files) and 107 molecular phylogenies that have been constructed

bioRxiv preprint doi: https://doi.org/10.1101/161117; this version posted July 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

139 for Zoanthidea (Table 1, Table S1). This required tracking GenBank accessions, publication-
140 specific taxon and specimen names, and in some cases creating species specific nucleotide
141 alignments to compare intraspecific specimens and identify mislabeled sequences (e.g. Swain
142 and Swain, 2014).

143 *2.2 Sequence alignment*

144 A concatenated alignment was assembled by integrating additional sequences into the
145 existing staggered framework of a previously published multiple alignment for Zoanthidea. This
146 included every nucleotide of all the sequences analyzed (complete-data) and is the staggered
147 alignment (See Table S1 for the genes and gene segments used for each species). Gene sequences
148 originating from GenBank were edited to remove amplification primers and single nucleotide
149 insertions from protein coding genes. Sequences were added to the alignment of Swain (2010)
150 (TreeBase: S10492) and aligned manually in Bioedit 7.2.5 (Hall, 1999) following the existing
151 staggered framework for hypervariable regions ITS1 and ITS2, and hypervariable regions within
152 16S and 12S (see section 3.2 for a description of theory, usage, and structure of staggered
153 alignments and Table 2 for the alignment positions that were staggered). The sequences of the
154 remaining three genes (18S, 28S, & COI), one gene segment (5.8S), and conserved regions of
155 16S and 12S conform to the format of a standard universal alignment within the complete-data
156 staggered alignment. The standard alignment is derived directly from the staggered alignment by
157 excluding the staggered hypervariable regions (see section 2.3 and Table 2).

158 Hypervariable regions are divergent in both sequence identity and length and are only
159 alignable with homologous sequences from closely-related species. The staggered regions of the
160 matrix are a form of phylogeny-informed alignment, where sequences are added to the matrix
161 proximal to closely-related taxa that were previously aligned (Morrison et al., 2015). This allows

bioRxiv preprint doi: https://doi.org/10.1101/161117; this version posted July 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

162 visual detection and alignment of short sequences within hypervariable regions that are
163 homologous in closely-related species, which are then isolated from divergent taxa within the
164 matrix by inserting coding for unknown character states (question marks) in parallel matrix
165 positions (see section 3.2). This creates a block of unambiguously aligned sequences and moves
166 all divergent sequences further along the alignment, where the process is then repeated, resulting
167 in the staggering of hypervariable regions. The initial assessment of sequence similarity is
168 performed at the 5′ terminus of ITS1, and is therefore dependent on the evolution of sequences in
169 this region; however similarity is continuously reassessed while proceeding in the 3′ direction
170 and individual sequences may be aligned with different groups of species along the length of the
171 alignment (all of which is completely documented in the TreeBase alignment accession:
172 S20129). Using the previous framework ('jump-starting': Morrison, 2006) greatly simplifies the
173 process as homologous sequences are already identified and adding new taxa requires relatively
174 few adjustments.

175 *2.3 Model parameters and phylogenetic analyses*

176 The complete-data staggered alignment was partitioned for model-fitting and
177 phylogenetic analysis, and species trees were independently inferred from the staggered
178 (complete-data) and standard (hypervariable-excluded) alignments. Partitioning traced the
179 boundaries of ribosomal subunits, staggered hypervariable regions, and codon positions
180 (following Li et al., 2008) for model-fitting and phylogenetic analyses (Table 2), however branch
181 length optimization was linked during tree inference due to incomplete per-partition taxon
182 sampling. Parameters of nucleotide evolution of the General Time Reversible (GTR) model with
183 gamma (+C) were determined simultaneously to phylogenetic inference using maximum-
184 likelihood analysis of RAxML v8.2.8 (Stamatakis, 2014) in the CIPRES Science Gateway v3.3

bioRxiv preprint doi: https://doi.org/10.1101/161117; this version posted July 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

185 (Miller et al., 2010) for the complete-data staggered alignment, and for the hypervariable
186 exclusion set applied to the same alignment (i.e., the standard alignment). This exclusion set
187 followed partitions applied to hypervariable regions (Table 2) and was used to assess the effect of
188 the staggered, rapidly evolving data (see section 3.3). Nonparametric bootstrap support was
189 estimated in RAxML using GTR and a categorical per-site rate heterogeneity approximation
190 (CAT) from 1000 pseudoreplicates (Stamatakis, 2014). Edwardsiid Actiniaria species were
191 designated as the outgroup.

192

193 **3. Theory**

194 *3.1 Gene trees and species trees*

195 We have known, for more than 30 years, the critical distinction between gene trees and
196 species trees (Pamilo and Nei, 1988). Gene trees, or phylogenetic inferences based upon a single
197 gene, reconstruct the relationships among homologous variants of a gene sampled from different
198 species and describes gene evolution. Species trees, or phylogenetic inferences based upon
199 multiple genes or data sources, reconstruct the evolutionary histories of species and describes
200 species evolution. Although both may use DNA sequence evolution as the basis of inference,
201 gene evolution is not equivalent to species evolution (Pamilo and Nei, 1988), and analyses of
202 species-level questions (e.g. species detection, identification, and systematics) based upon gene
203 trees falsely assume equivalence in the evolution of genes and species.

204 This disparity between gene and species trees has multiple biological causes, but is
205 ultimately due to the incomplete history of species that is provided by sampling a small
206 proportion of their genomes. Evolutionary events such as gene deletion or duplication and
207 horizontal gene transfer can cause dramatic differences in the evolutionary histories of genes and

bioRxiv preprint doi: https://doi.org/10.1101/161117; this version posted July 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

208 species, but usually occur under specific circumstances, regions of the genome, or evolutionary
209 lineages (Edwards, 2009). Conversely, differing evolutionary rates among genes, resulting in
210 incomplete lineage sorting (ILS), or deep coalescence, and the related issues of branch length
211 heterogeneity and heterotachy, appear to be universal issues that affect all genes, genomes
212 (organelle or nucleus), and lineages (Edwards, 2009).

213 There are various approaches to reconstructing species trees, but all attempt to do so by
214 incorporating multiple regions of the genome, and if possible multiple genomes, into the same
215 tree inference. The two main approaches involve concatenation of multiple genes into the same
216 alignment which is then used to infer a species tree, and alternatively inferring a tree from each
217 gene which are then used to infer a species tree under a coalescence model. Which is the best
218 approach for obtaining a highly supported and statically consistent inference is currently a source
219 of active debate, as each approach makes specific assumptions and carry specific weaknesses
220 (Edwards, 2009; Roch and Warnow, 2015; Springer and Gatesy 2016). There is a potential for
221 concatenated analysis to return statistically inconsistent trees because of discordance in gene
222 evolution (Degnan and Rosenberg, 2006; Rosenberg, 2013), which species tree-estimation based
223 on gene trees can overcome (even in the anomaly zone; Liu et al., 2010); however discordance
224 between Zoanthidea phylogenies based on nuclear and mitochondrial genes has been previously
225 demonstrated to be insignificant (Swain, 2010). This assessment was performed at the genome-
226 level rather than the gene-level because of incomplete taxon sampling of each gene, making
227 comparable gene trees possible for only a few taxa. Additionally, datasets with incomplete taxon
228 sampling (such as the case here) are predicted to be inferred with greater confidence using
229 concatenation methods (Edwards, 2009). Regardless of the approach, incorporation of a greater
230 volume and diversity of data into the inference should improve resolution and confidence in

bioRxiv preprint doi: https://doi.org/10.1101/161117; this version posted July 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

231 comparison to gene trees.

232 *3.2 Staggered Alignments*

233 Multiple sequence alignment and tree inference are both critical to molecular
234 phylogenetic analysis, however the quality and data content of tree inference is entirely
235 dependent upon the quality and data content of alignment, as a phylogeny is a representation and
236 interpretation of its underlying homology statements (Morrison, 2009a). Given the critical role of
237 alignment as the primary homology assessment in phylogenetic inference, standard automated
238 nucleotide sequence alignment is adept at assessing substitution events, but generally inadequate
239 for generating homology hypotheses for most other common evolutionary events (e.g., deletion,
240 duplication, insertion, inversion, and translocation; Morrison, 2009a). This limitation is generally
241 not problematic for protein-coding regions because much of the evolutionary change observed in
242 these sequences involves single residue substitutions and standard automated alignment based on
243 sequence similarity can adequately assign homology within these events. However, only 1–20%
244 of the genome of multicellular eukaryotes are composed of these conserved coding regions
245 (Szymanski et al., 2007), leaving most genetic material and its evolutionary events not assessable
246 by standard automated alignment (Morrison, 2009b). Additionally, many genes commonly
247 applied to phylogenetic questions are not protein-coding and therefore most standard automated
248 alignments of these genes contain misaligned sequences.

249 The main concern is over-alignment, or aligning sequences that are unlikely to be
250 homologous. This issue is usually addressed in a standard alignment by excluding sequence
251 regions with alignment challenges that are easily observable; meaning that even automated
252 alignment requires manual correction to exclude low quality homology assessments (Morrison,
253 2009b). There are multiple alternative approaches to refine alignments depending upon the

bioRxiv preprint doi: https://doi.org/10.1101/161117; this version posted July 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

254 molecular function of the nucleotides involved (such as rRNA secondary structure prediction
255 coupled with alignment correction or molecular morphometric analysis: e.g., Aguilar and
256 Reimer, 2010; Swain and Taylor, 2003; Torres-Suarez, 2014), but all are focused on the same
257 goal: retaining within the analysis as much information as possible (Morrison, 2009b).

258 One approach, which was commonly used prior to the wide-spread application of
259 automated sequence alignment (Morrison, 2006), is the staggered alignment (Barta, 1997). A
260 staggered alignment is a mosaic of local and universal nucleotide sub-alignments that are
261 partially informed by phylogeny which allow the retention of all nucleotides (complete-data)
262 without aligning non-homologues sequences (Morrison et al., 2015). This approach creates a
263 single matrix that aligns homologous nucleotides within each sub-alignment: universal sub-
264 alignments (equivalent to a standard alignment) are composed of conserved regions of the
265 genome and include all taxa in the analysis, while local sub-alignments are composed of
266 hypervariable regions of the genome and include only closely related taxa in each sub-alignment
267 which are then staggered relative to other local sub-alignments (Fig. 1). A standard alignment can
268 be derived directly from a staggered alignment by excluding the staggered local sub-alignments.

269 Like most standard automated alignment approaches, the staggered alignment relies upon
270 sequence similarity to assess alignment quality, however it also benefits from the ability to
271 examine multiple adjacent positions to specifically accommodate length variation among
272 sequences. This approach can be initiated through automated alignment as a first approximation,
273 but is completed through manual adjustment of nucleotide positioning following a single rule: do
274 not align sequences that are not obviously homologous (Morrison, 2006). The resulting staggered
275 alignment explicitly addresses our inability to simultaneously assess homology of complex
276 evolutionary events across all targeted taxa, without discarding informative data. A standard

bioRxiv preprint doi: https://doi.org/10.1101/161117; this version posted July 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

277 alignment would follow the identical procedure of automated alignment as a first approximation
278 which is then completed through manual adjustment; however with a standard alignment, the
279 manual adjustment discards hypervariable sequence data rather than making it available for
280 phylogenetic inference. Phylogenetic trees inferred from the same DNA sequences using
281 staggered alignments will differ from trees inferred from standard alignments in resolution,
282 branch order, and node support within the terminal clades, but the internal order of cladogenesis
283 will be similar because they rely upon similar data matrices (Barta 1997).

284 *3.3 Use of hypervariable regions in phylogenetic reconstruction*

285 The need for molecular markers that are informative at species and population levels has
286 driven interest in capturing information contained in hypervariable regions of genomes. There is
287 broad consensus that these regions contain informative variation, and they are widely used as
288 species-level markers in many systems (Coleman, 2007; Forsman et al., 2009). Although
289 mitochondrial outpaces nuclear sequence evolution in most organisms (Creer, 2007),
290 mitochondrial sequence evolution is extraordinarily slow among anthozoans and is often
291 invariant among its most closely-related species (Shearer et al., 2002). The nuclear ribosomal
292 ITS region contains a mosaic of secondary structural elements that can be relatively conserved
293 (stems) or variable (loops and buldges) (Hillis and Dixon, 1991), and evolve under complex and
294 varied evolutionary constraints. Hypervarible ITS sequences can differ in both length and
295 nucleotide identity, which can cause challenges in assessing homology for multiple sequence
296 alignment. Additionally, the ITS array can be repeated hundreds or thousands of times within a
297 single genome and the exact sequence of nucleotides in each copy are homogenized to variying
298 degrees of completion and precision by concerted evolution (Elder and Turner, 1995; Hillis and
299 Dixon, 1991); allowing for potential intragenomic variation.

bioRxiv preprint doi: https://doi.org/10.1101/161117; this version posted July 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

300 These two issues, alignment challenges and intragenomic variation, are widely
301 acknowledged as being the primary obstacles to sucessfully using hypervariable ITS sequences
302 in phylogentic inference. Alignment challenges can be mitigated by exclusively comparing
303 regions with similar sequence length and identity through a staggered alignement (see section
304 3.2), thereby increasing the probablity of correct homology assesment at each matrix position.
305 Intragenomic variation can be a more insidous problem, however intragenomic heterogeneity
306 that is sufficient to mask evolutionary singnal is often confined to specific taxonomic groups
307 (Coleman, 2003, 2007). For example, within the Anthozoa order Scleractinia, ITS sequences
308 have been broadly applied to phylogenetics (reviewed in Forsman et al., 2009; Kitahara et al.,
309 2016), and extreme intragenomic heterogeneity is largely confined to the genus *Acropora* (Wei et
310 al., 2006). Although intragenomic variation is not a common target for analysis within
311 Zoanthidea, direct sequencing of ITS regions results in largely unambigous sequence reads
312 (suggesting that most copies are homogenized; e.g., Swain, 2009, 2010) and, if intragenomic
313 heterogeneity is present, it appears to be generally insufficient to mask the evolutionary signal of
314 species boundaries (but see Reimer et al., 2007c for example of hybridization) and intraspecific
315 relationships (Aguilar and Reimer, 2010; Fujii and Reimer, 2011; Hibino et al., 2014; Irei et al.,
316 2015; Kise and Reimer, 2016; Montenegro et al., 2016; Montenegro et al., 2015; Reimer et al.,
317 2012a; Reimer and Fujii, 2010; Reimer et al., 2013a; Reimer et al., 2010c; Reimer et al., 2012b;
318 Reimer et al., 2014; Reimer et al., 2008a; Reimer and Sinniger, 2010b; Reimer et al., 2008c;
319 Reimer et al., 2007b; Risi and Macdonald, 2015; Risi and Macdonald, 2016; Sinniger et al.,
320 2010a; Swain, 2009, 2010).
321
322 **4. Results and Discussion**

bioRxiv preprint doi: https://doi.org/10.1101/161117; this version posted July 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

323 *4.1 Nucleotide alignment matrix*

324 The search of GenBank for novel taxon and gene sequences added 234 accessions (from
325 97 species) to the 273 accessions (from 82 species) retained from the alignment of Swain (2010),
326 for a total of 767 genes or gene segments from 144 Zoanthidea and 11 outgroup species,
327 including representatives of nearly all known Zoanthidea genera (sequences of Epizoanthidae
328 genera *Paleozoanthus* and *Thoracactis* are unavailable; Table S1). This includes 114 accessions
329 of COI; a very short sequence (<600 nt) that codes for a subunit of an enzyme of the electron
330 transport chain. As in other Anthozoa taxa, COI has long been known to be nearly useless for
331 phylogenetic inference among closely-related species as its evolutionary rate is >100 times
332 slower than most marine invertebrates (Hellberg, 2006; Shearer et al., 2002; Stampar et al.,
333 2012). These COI sequences are included here only because these data are available, not to
334 encourage continued investment in their collection, which should be seen as a misuse of limited
335 resources expended upon an uninformative marker. The remaining 7 gene segments targeted in
336 this analysis are incompletely sampled, such that the matrix contains 653 Zoanthidea genes or
337 gene segments out of a possible 1008 (144 taxa x 7 gene segments). While incomplete data
338 matrices are not desirable, maximum likelihood phylogenetic analyses are generally robust to
339 missing data issues and including incompletely sampled genes generally increases their accuracy
340 relative to excluding them (Jiang et al., 2014; Streicher et al., 2016). Collection of the missing
341 sequences and completion of this matrix, along with bolstering data for species that have
342 available sequences but did not meet the minimal requirements for inclusion here (e.g.
343 *Isozoanthus sulcatus*), should be a focus of future research on improving the resolution and
344 confidence in the Zoanthidea phylogeny. Completing this matrix will also allow a more
345 comprehensive analysis of potential gene tree discordance then is currently possible, as well as

bioRxiv preprint doi: https://doi.org/10.1101/161117; this version posted July 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

346 inferring statistically consistent species trees under a coalescence model if discordance is
347 detected.

348 Once compiled, and properly staggered, the alignment contains >2.8 million matrix
349 positions (155 rows by 18,075 columns; TreeBase: S20129). Alignment staggering represents a
350 significant time commitment over automated multiple alignment alone, as hypervariable regions
351 of small subsamples of taxa are individually assessed and manually adjusted. However, the
352 resulting staggered alignment simultaneously maximizes information content and analysis power
353 as it retains all nucleotides recovered from the original sequencing reads and allows the
354 application of nucleotide evolutionary model-informed phylogenetic inference. There are
355 alternative methods available, such as secondary structure prediction followed by alignment
356 correction or molecular morphometric analysis (e.g., Aguilar and Reimer, 2010; Swain and
357 Taylor, 2003; Torres-Suarez, 2014), but all retain less of the original sequence data or restrict the
358 diversity of taxa that can be simultaneously analyzed and employ less powerful and robust
359 analyses. Partitioning the data matrix resulted in twelve distinct models of nucleotide evolution,
360 detailed in Table 2, that were applied in ML inferences of tree topology and bootstrap support.

361 *4.2 Complete-data phylogenetic inference: analysis of the staggered alignment*

362 A search for the optimal ML tree from the partitioned staggered alignment resulted in a
363 best tree with a likelihood score of -65296 (Fig. 2). This analysis recovered highly supported
364 (bootstrap values of 100) monophylies of taxa representing order Zoanthidea and suborder
365 Brachycnemina, with suborder Macrocnemina ancestral to Brachycnemina, and an overall tree
366 topology that generally conforms to previously reported phylogenies that are of comparable
367 taxon sampling (i.e. Sinniger et al., 2005; Swain, 2010).

368 Most currently recognized families were also recovered at a high-level of certainty

bioRxiv preprint doi: https://doi.org/10.1101/161117; this version posted July 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

369 including Microzoanthidae, Epizoanthidae, Abyssoanthidae, Hydrozoanthidae, and Sphenopidae.
370 Nanozoanthidae is monospecific and therefore impossible to assess. Parazoanthidae is both
371 paraphyletic (with respect to Abyssoanthidae) and polyphyletic (with respect to *Isozoanthus*) and
372 the relationships among genera are largely unresolved (i.e., extremely weak bootstrap support)
373 except for monophyletic sister genera *Parazoanthus* and *Umimayanthus*. Other monophyletic
374 Parazoanthidae genera include *Antipathozoanthus*, *Mesozoanthus*, *Zibrowius*, *Hurlizoanthus*,
375 *Savalia*, and *Bergia*. *Corallizoanthus* is polyphyletic, and *Bullagummizoanthus, Kauluzoanthus,*
376 and *Kulamanamana* are monospecific (and therefore impossible to assess); however the topology
377 in this region of the tree generally lacks support due to the paucity of sequence data available for
378 these taxa (which is almost entirely mitochondrial; Table S1). Clarification of these relationships
379 should be a priority for future research and could be easily accomplished by adding highly
380 informative ITS and 28S genes. Zoanthidae is paraphyletic with respect to Neozoanthidae, where
381 the genus *Neozoanthus* is ancestral to *Isaurus*. Also within Zoanthidae, *Zoanthus* is paraphyletic
382 with respect to *Acrozoanthus*. Although Sphenopidae is monophyletic, its daughter genus
383 *Palythoa* is paraphyletic with respect to *Sphenopus*. Although there is much more confidence in
384 this region of the tree, ~37% of the brachycnemic taxa are represented exclusively by
385 mitochondrial genes in the data matrix and further resolution and confidence could be obtained
386 by completing the matrix with information derived from the nuclear compartment.

387 *4.3 Hypervariable-excluded phylogenetic inference: analysis of the standard alignment*

388 A search for the optimal ML tree from the partitioned standard alignment, (equivalent to
389 the staggered alignment excluding hypervariable regions; Table 2), resulted in a best tree with a
390 likelihood score of -31899 (Fig. 3). Although the exact branching order of the two trees varies,
391 the overall toplology of the hypervariable-excluded tree is largely congruent with the complete-

bioRxiv preprint doi: https://doi.org/10.1101/161117; this version posted July 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

392 data tree, and apart from a few exceptions matches the monophylies of higher taxa detailed
393 above. The hypervariable-excluded tree differs with a polyphyletic Sphenopidae, *Abyssoanthus*
394 within *Parazoanthus*, *Acrozoanthus* basal to *Zoanthus* and *Zoanthus* as monophyletic, and
395 *Sphenopus* basal to Brachycnemina. Additionally, the bootstrap values are ~15% lower in the
396 hypervariable-excluded tree, with comparable nodes falling from a mean of 80.1 in the complete-
397 data tree to 68.5 in the hypervariable-excluded tree.

398 *4.4 Preferred topology: the complete-data inference based on the staggered alignment*

399 Given these competing topologies, the complete-data tree based on the staggered
400 alignment best reflects our current understanding of molecular evolution, evolution of form, and
401 systematics of Zoanthidea. The position of *Abyssoanthus* is generally problematic (and weakly
402 supported in both inferences) because its nucleotide sequences are highly divergent (and
403 therefore rarely included in phylogenetic analyses) and its anatomy is all but unknown (Reimer
404 and Sinniger, 2010a; Reimer et al., 2007a); however, *Abyssoanthus* as sister to *Mesozoanthus* (as
405 in the complete-data tree based on the staggered alignment) agrees with previous phylogenetic
406 hypotheses (Reimer and Sinniger, 2010a; Reimer et al., 2007a) and the ecology of both taxa,
407 whereas *Abyssoanthus* within *Parazoanthus* (as in the hypervariable-excluded tree based upon
408 the standard alignment) would be a significant departure from our understanding of both genera.

409 In the hypervariable-excluded tree (based upon the standard alignment), Sphenopidae is
410 polyphyletic because *Palythoa mizigama* is inferred to be part of the Zoanthidae monophyly.
411 Zoanthidae and Sphenopidae have contrasting anatomical features (reviewed in Swain et al.,
412 2016) and there is no indication that the morphology of *P. mizigama* differs from our
413 understanding of *Palythoa* and Sphenopidae (Irei et al., 2015). *Acrozoanthus* and *Sphenopus* are
414 both basal to their respective clades in the hypervariable-excluded tree (based upon the standard

bioRxiv preprint doi: https://doi.org/10.1101/161117; this version posted July 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

415 alignment), supporting the validity of these genera; however, both are odd genera represented by
416 few species that might best fit within other genera, as inferred in the complete-data tree (based
417 upon the staggered alignment). *Acrozoanthus* is a mono-specific genus that was created because
418 of its apparent ability to build an erect skeleton (Saville-Kent, 1893), however this skeleton was
419 revealed to be the parchment-like tube of a worm in the genus *Eunice* (Ryland, 1997). Therefore
420 skeleton-building is not the odd character of *Acrozoanthus*, rather is its ability to form symbiotic
421 associations with polycheate worms (which is only common among zoanthideans in the genus
422 *Epizoanthus* (Swain, 2010, Reimer et al., 2010a) and cannot help us to understand the
423 relationship between *Acrozoanthus* and *Zoanthus*). Including *Acrozoanthus* within *Zoanthus* has
424 been previously suggested using morphology (reviewed in Ryland, 1997) and molecular
425 phylogenetics (e.g. Reimer et al., 2010c; Sinniger et al., 2005; Swain, 2010), and the complete-
426 data tree (based upon the staggered alignment) agrees with this conclusion. *Sphenopus* is odd
427 because it is solitary and azooxanthellate, whereas most of Sphenopidae are colonial and
428 zooxanthellate. However, two recently described *Palythoa* species, *P. umbrosa* and *P. mizigama*,
429 form colonies with small numbers of polyps and are uncharacteristically azooxanthellate (Irei et
430 al., 2015). *Sphenopus* is part of a monophyly with *P. umbrosa* and *P. mizigama* within the
431 *Palythoa* monophyly in the complete-data tree (based upon the staggered alignment), but these
432 taxa are dispersed across the hypervariable-excluded tree (based upon the standard alignment).
433 All of these major differences between these two trees suggest that the topology of the complete-
434 data tree (based upon the staggered alignment) is a more accurate representation.

435 *4.5 Novel hypotheses*

436 Although the complete-data tree (based upon the staggered alignment) is largely
437 congruent with previous comparable phylogenies, it also supports novel hypotheses and

bioRxiv preprint doi: https://doi.org/10.1101/161117; this version posted July 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

438 discredits others. Previous work on Microzoanthidae and Nanozoanthidae had placed these
439 families within a monophyly with the Parazoanthidae genus *Isozoanthus* and as sister to the
440 remaining Zoanthidea (Fujii and Reimer, 2011, 2013). The complete-data tree (based upon the
441 staggered alignment) and the hypervariable-excluded tree (based upon the standard alignment)
442 strongly supports the hypothesis that Microzoanthidae and Nanozoanthidae form an exclusive
443 monophyly at the base of Zoanthidea and that *Isozoanthus* is sister to *Epizoanthus* (Fig. 2, 3).

444 The region of the complete-data tree (based upon the staggered alignment) containing the
445 octocoral-symbiotic genera and their allies (*Bullagummizoanthus*, *Corallizoanthus*,
446 *Hurlizoanthus*, *Kauluzoanthus*, *Kulamanamana*, *Savalia*, *Zibrowius*) lacks strong support, mostly
447 because there are data for only 3 of 6 genes targeted in these analyses (Table S1), but suggests
448 that many of these taxa could be congeneric. Unfortunately, along with the paucity of genetic
449 data for these taxa, there is also very little anatomical data upon which a hypothesis could be
450 based (Swain et al., 2016). There is considerable information about the genetics and anatomy of
451 the three *Corallizoanthus* species included here (Swain, 2010; Swain et al., 2015; Swain et al.,
452 2016; Swain and Swain, 2014), which could serve as guide for reexamination of the remaining
453 taxa in this group. Resolving this portion of the tree should be a priority for future research.

454 At the base of the *Palythoa* monophyly is a strongly supported clade of taxa that is almost
455 entirely comprised of species that were once assigned to the genus *Protopalythoa*. This includes
456 *P. grandis*, *P. heliodiscus*, *P. variabilis*, but not others such as *P. grandiflora* and *P. mutuki*. The
457 status of this genus has been in dispute for some time (Burnett et al., 1997, Low et al.,2016;
458 Reimer et al., 2006c; Ryland and Lancaster, 2003), and the findings presented here do not settle
459 this issue, but perhaps this novel hypothesis is an opening to reconsider the validity of
460 *Protopalythoa*. Again, additional DNA sequence that could be used to complete the data matrix

bioRxiv preprint doi: https://doi.org/10.1101/161117; this version posted July 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

461 could dramatically improve our understanding of the taxa in this region.

462 *4.6 Sister of Zoanthidea*

463 Where the order Zoanthidea inserts into the Cnidaria Tree of Life has been one of the
464 targets of sustained research and a point of contention (see Rodríguez et al., 2014 and references
465 therein). Sister to Zoanthidea has been variously hypothesized to be either Actiniaria or
466 Antipatharia, and its relationship with the remaining orders is poorly understood. The recent
467 Actiniaria-focused inference by Rodríguez et al. (2014) put forth the novel hypothesis that the
468 enigmatic species *Relicanthus daphneae* may be sister to Zoanthidea and both form a monophyly
469 with Antipatharia, while Actinaria is much more distantly related. The Zoanthidea-focused
470 analysis presented here included representatives of Antipatharia, Actiniaria, and *R. daphneae*,
471 however the root of the phylogenetic inference was set only as the two undescribed Edwardsiidae
472 species, allowing the remaining taxa to be placed according to the data. In this analysis *R.*
473 *daphneae* is sister to Antipatharia and both form a monophyly with Zoanthidea (Fig. 1, 2). With
474 the highly divergent sequences of *R. daphneae* and weak bootstrap support at its insertion, along
475 with partially-overlapping taxon and datasets between the analysis presented here and Rodríguez
476 et al. (2014), it would seem to remain an open question.

477

478 **5. Conclusions**

479 Proliferation of Zoanthidea species and higher taxa descriptions since the year 2000, have
480 occurred at a rate not seen for nearly 100 years (Swain et al., 2016). This explosion of discovery
481 and identification of evolutionary patterns was triggered by the application of molecular
482 phylogenetics to an order that was largely ignored because rampant confusion at the species and
483 genus levels. Molecular tools have detected cryptic species and genera, both reinforced and

bioRxiv preprint doi: https://doi.org/10.1101/161117; this version posted July 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

484 dismantled previous taxonomic and systematic hypotheses, and greatly expanded our
485 understanding of Zoanthidea diversity. However, the tendency to, almost exclusively, construct
486 gene trees based on many specimens of the same species (Table 1) as a tool of species discovery,
487 has caused a parallel proliferation in published trees without a corresponding increase in the
488 understanding of species phylogeny. The comprehensive phylogeny presented here is intended to
489 fill that gap. Its utility is not in simply describing the proximate relationships of individual
490 species as supported by the nucleotide data, as the phylogenies presented here are generally in
491 agreement with previously published Zoanthidea species trees and many gene trees (but see
492 section 4.5 for novel hypotheses). Rather the major advancement here is the inclusion of nearly
493 all available species and nucleotide sequence data (including data that are usually discarded) into
494 a single robust phylogeny (increasing resolution and node certainty) that can then serve as an
495 expandable tool for performing higher-level assessments of evolutionary hypotheses and
496 phylogeny-corrected statistical analyses (see exemplar analyses in Swain et al., 2015; Swain et
497 al., 2016).

498 The absence of a comprehensive phylogeny can be traced to the disparity between the
499 simplicity of alignment and the data content of the genes currently being applied to this group.
500 Conserved genes can be easily aligned across taxa representing all genera and families of
501 Zoanthidea, but produces poorly resolved and poorly supported trees. Hypervariable genes can
502 provide the evolutionary signal necessary for topology resolution and confidence building, but
503 creating the homology statements (sequence alignments) necessary for phylogenetic inference is
504 a challenge. By staggering the alignment, it is possible to overcome the challenge of
505 hypervariable sequence evolution while retaining its information. The complete-data tree based
506 upon the staggered alignment provides a better supported topology that best matches our

bioRxiv preprint doi: https://doi.org/10.1101/161117; this version posted July 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

507 understanding of evolution among Zoanthidea taxa. All major incongruencies between the
508 hypervariable-excluded (based upon the standard alignment) and complete-data (based upon the
509 staggered alignment) trees favor the hypothesis offered by the complete-data tree.
510

511 **Acknowledgments**

512 This work was supported by the US National Science Foundation [CBET-0937987,
513 CBET-1240416, and OCE-0550599]; Northwestern University, Evanston, IL; the Field Museum
514 of Natural History, Chicago, IL; and Florida State University, Tallahassee, FL; and L. Marcelino,
515 and V. Backman of NU, R. Bieler of FMNH, and J. Wulff and S. Steppan of FSU. My original
516 use of staggered alignments in phylogenetic analyses grew out of conversations on analyzing
517 hypervariable sequences with S. Steppan.
518

519 **References**
520

521 Aguilar, C., Reimer, J.D., 2010. Molecular phylogenetic hypotheses of *Zoanthus* species
522 (Anthozoa:Hexacorallia) using RNA secondary structure of the internal transcribed spacer 2
523 (ITS2). Marine Biodiversity 40, 195–204.
524
525 Barta, J.R., 1997. Investigating phylogenetic relationships within the *Apicomplexa* using
526 sequence data: The search for homology. Methods-a Companion to Methods in Enzymology 13,
527 81–88.
528
529 Burnett, W.J., Benzie, J.A.H., Beardmore, J.A., Ryland, J.S., 1997. Zoanthids (Anthozoa,
530 Hexacorallia) from the Great Barrier Reef and Torres Strait, Australia: Systematics, evolution
531 and a key to species. Coral Reefs 16, 55–68.
532
533 Coleman, A.W., 2003. ITS2 is a double-edged tool for eukaryote evolutionary comparisons.
534 Trends in Genetics 19, 370–375.
535
536 Coleman, A.W., 2007. Pan-eukaryote ITS2 homologies revealed by RNA secondary structure.
537 Nucleic Acids Research 35, 3322–3329.
538
539 Creer, S., 2007. Choosing and using introns in molecular phylogenetics. Evolutionary
540 Bioinformatics 3, 99–108.

bioRxiv preprint doi: https://doi.org/10.1101/161117; this version posted July 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

541
542 Degnan, J.H., Rosenberg, N.A., 2006. Discordance of species trees with their most likely gene
543 trees. PLoS Genetics, pp. 762–768.
544
545 Edwards, S.V., 2009. Is a new and general theory of molecular systematics emerging? Evolution
546 63, 1-19.
547
548 Elder, J.F., Turner, B.J., 1995. Concerted evolution of repetitive DNA sequences in eukaryotes.
549 Quarterly Review of Biology 70, 297–320.
550
551 Forsman, Z.H., Barshis, D.J., Hunter, C.L., Toonen, R.J., 2009. Shape-shifting corals: molecular
552 markers show morphology is evolutionarily plastic in *Porites*. BMC Evolutionary Biology 9, 45.
553
554 Fujii, T., Reimer, J.D., 2011. Phylogeny of the highly divergent zoanthid family Microzoanthidae
555 (Anthozoa, Hexacorallia) from the Pacific. Zoologica Scripta 40, 418–431.
556
557 Fujii, T., Reimer, J.D., 2013. A new family of diminutive zooxanthellate zoanthids (Hexacorallia:
558 Zoantharia). Zoological Journal of the Linnean Society 169, 509–522.
559
560 Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis
561 program for Windows 95 /98 /NT. Nucleic Acid Symposium Series, pp. 95–98.
562
563 Hellberg, M.E., 2006. No variation and low synonymous substitution rates in coral mtDNA
564 despite high nuclear variation. BMC Evolutionary Biology 6, 24
565
566 Hibino, Y., Todd, P.A., Yang, S.-y., Benayahu, Y., Reimer, J.D., 2014. Molecular and
567 morphological evidence for conspecificity of two common Indo-Pacific species of *Palythoa*
568 (Cnidaria: Anthozoa). Hydrobiologia 733, 31–43.
569
570 Hillis, D.M., Dixon, M.T., 1991. Ribosomal DNA: molecular evolution and phylogenetic
571 inference. Quarterly Review of Biology 66, 411–453.
572
573 Irei, Y., Sinniger, F., Reimer, J.D., 2015. Descriptions of two azooxanthellate *Palythoa* species
574 (Subclass Hexacorallia, Order Zoantharia) from the Ryukyu Archipelago, southern Japan.
575 Zookeys 478, 1–26.
576
577 Jiang, W., Chen, S.Y., Wang, H., Li, D.Z., Wiens, J.J., 2014. Should genes with missing data be
578 excluded from phylogenetic analyses? Molecular Phylogenetics and Evolution 80, 308–318.
579
580 Kise, H., Reimer, J.D., 2016a. Unexpected diversity and a new species of *Epizoanthus*
581 (Anthozoa, Hexacorallia) attached to eunicid worm tubes from the Pacific Ocean. Zookeys 562,
582 49–71.
583
584 Kitahara, M.V., Fukami, H., Benzoni, F., Huang, D., 2016. The new systematics of Scleractinia:
585 integrating molecular and morphological evidence. In: Goffredo, S., Dubinsky, Z. (Eds.), The
586 Cnidaria, Past, Present and Future: The world of Medusa and her sisters. Springer International

bioRxiv preprint doi: https://doi.org/10.1101/161117; this version posted July 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

587 Publishing, pp. 41–59.
588
589 Koupaei, A.N., Mostafavi, P.G., Mehrabadi, J.F., Fatemi, S.M.R., 2014. Molecular diversity of
590 coral reef-associated zoanthids off Qeshm Island, northern Persian Gulf. International Aquatic
591 Research 6, 64.
592
593 Koupaei, A.N., Mostafavi, P.G., Mehrabadi, J.F., Fatemi, S.M.R., Dehghani, H., 2016. Diversity
594 of shallow water zoantharians in Hengam and Larak Islands, in the Persian Gulf. Journal of the
595 Marine Biological Association of the United Kingdom 96, 1145–1155.
596
597 Li, C., Lu, G., Orti, G., 2008. Optimal data partitioning and a test case for ray-finned fishes
598 (Actinopterygii) based on ten nuclear loci. Systematic Biology 57, 519–539.
599
600 Liu, L.A., Yu, L.L., Edwards, S.V., 2010. A maximum pseudo-likelihood approach for estimating
601 species trees under the coalescent model. BMC Evolutionary Biology 10, 302.
602
603 Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for
604 inference of large phylogenetic trees. Gateway Computing Environments Workshop (GCE), New
605 Orleans, LA, pp. 1–8.
606
607 Montenegro, J., Low, M.E.Y., Reimer, J.D., 2016. The resurrection of the genus *Bergia*
608 (Anthozoa, Zoantharia, Parazoanthidae). Systematics and Biodiversity 14, 63–73.
609
610 Montenegro, J., Sinniger, F., Reimer, J.D., 2015. Unexpected diversity and new species in the
611 sponge-Parazoanthidae association in southern Japan. Molecular Phylogenetics and Evolution
612 89, 73–90.
613
614 Morrison, D.A., 2006. Multiple sequence alignment for phylogenetic purposes. Australian
615 Systematic Botany 19, 479–539.
616
617 Morrison, D.A., 2009a. A framework for phylogenetic sequence alignment. Plant Systematics
618 and Evolution 282, 127–149.
619
620 Morrison, D.A., 2009b. Why would phylogeneticists ignore computerized sequence alignment?
621 Systematic Biology 58, 150–158.
622
623 Morrison, D.A., Morgan, M.J., Kelchner, S.A., 2015. Molecular homology and multiple-
624 sequence alignment: an analysis of concepts and practice. Australian Systematic Botany 28, 46–
625 62.
626
627 Pamilo, P., Nei, M., 1988. Relationships between gene trees and species trees. Molecular Biology
628 and Evolution 5, 568–583.
629
630 Reimer, J.D., Foord, C., Yuka Irei, Y., 2012a. Species diversity of shallow water zoanthids
631 (Cnidaria: Anthozoa: Hexacorallia) in Florida. Journal of Marine Biology 2012, 856079.
632

bioRxiv preprint doi: https://doi.org/10.1101/161117; this version posted July 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

633 Reimer, J.D., Fujii, T., 2010. Four new species and one new genus of zoanthids (Cnidaria,
634 Hexacorallia) from the Galapagos Islands. Zookeys, 1–36.
635
636 Reimer, J.D., Hirose, M., Irei, Y., Obuchi, M., Sinniger, F., 2011a. The sands of time: rediscovery
637 of the genus *Neozoanthus* (Cnidaria: Hexacorallia) and evolutionary aspects of sand incrustation
638 in brachycnemic zoanthids. Marine Biology 158, 983–993.
639
640 Reimer, J.D., Hirose, M., Nishisaka, T., Sinniger, F., Itani, G., 2010a. *Epizoanthus* spp.
641 associations revealed using DNA markers: A case study from Kochi, Japan. Zoological Science
642 27, 729–734.
643
644 Reimer, J.D., Hirose, M., Wirtz, P., 2010b. Zoanthids of the Cape Verde Islands and their
645 symbionts: previously unexamined diversity in the Northeastern Atlantic. Contributions to
646 Zoology 79, 147–163.
647
648 Reimer, J.D., Hirose, M., Yanagi, K., Sinniger, F., 2011b. Marine invertebrate diversity in the
649 oceanic Ogasawara Islands: a molecular examination of zoanthids (Anthozoa: Hexacorallia) and
650 their *Symbiodinium* (Dinophyceae). Systematics and Biodiversity 9, 133–143.
651
652 Reimer, J.D., Irei, Y., Fujii, T., Yang, S.Y., 2013a. Molecular analyses of shallow-water
653 zooxanthellate zoanthids (Cnidaria: Hexacorallia) from Taiwan and their *Symbiodinium* spp.
654 Zoological Studies 52, 38.
655
656 Reimer, J.D., Irei, Y., Naruse, T., 2013b. A record of *Neozoanthus cf. uchina* Reimer, Irei & Fujii,
657 2012 from the Yaeyama Islands, southern Ryukyu Islands, Japan. Fauna Ryukyuana 1–7.
658
659 Reimer, J.D., Ishikawa, S.A., Hirose, M., 2010c. New records and molecular characterization of
660 *Acrozoanthus* (Cnidaria: Anthozoa: Hexacorallia) and its endosymbionts (*Symbiodinium* spp.)
661 from Taiwan. Marine Biodiversity 41, 313–323.
662
663 Reimer, J.D., Lin, M., Fujii, T., Lane, D.J.W., Hoeksema, B.W., 2012b. The phylogenetic
664 position of the solitary zoanthid genus *Sphenopus* (Cnidaria: Hexacorallia). Contributions to
665 Zoology 81, 43–54.
666
667 Reimer, J.D., Lorion, J., Irei, Y., Hoeksema, B., Wirtz, P., 2014. Ascension Island shallow-water
668 Zoantharia (Hexacorallia: Cnidaria) and their zooxanthellae (*Symbiodinium*). Journal of the
669 Marine Biological Association of the United Kingdom doi:10.1017/S0025315414000654.
670
671 Reimer, J.D., Nonaka, M., Sinniger, F., Iwase, F., 2008a. Morphological and molecular
672 characterization of a new genus and new species of parazoanthid (Anthozoa: Hexacorallia:
673 Zoantharia) associated with Japanese Red Coral. Coral Reefs 27, 935–949.
674
675 Reimer, J.D., Ono, S., Fujiwara, Y., Takishita, K., Tsukahara, J., 2004. Reconsidering *Zoanthus*
676 spp. diversity: Molecular evidence of conspecifity within four previously presumed species.
677 Zoological Science 21, 517–525.
678

bioRxiv preprint doi: https://doi.org/10.1101/161117; this version posted July 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

679 Reimer, J.D., Ono, S., Iwama, A., Takishita, K., Tsukahara, J., Maruyama, T., 2006a.
680 Morphological and molecular revision of *Zoanthus* (Anthozoa: Hexacorallia) from southwestern
681 Japan, with descriptions of two new species. Zoological Science 23, 261–275.
682
683 Reimer, J.D., Ono, S., Iwama, A., Tsukahara, J., Maruyama, T., 2006b. High levels of
684 morphological variation despite close genetic relatedness between *Zoanthus aff. vietnamensis*
685 and *Zoanthus kuroshio* (Anthozoa : Hexacorallia). Zoological Science 23, 755–761.
686
687 Reimer, J.D., Ono, S., Takishita, K., Tsukahara, J., Maruyama, T., 2006c. Molecular evidence
688 suggesting species in the zoanthid genera *Palythoa* and *Protopalythoa* (Anthozoa : Hexacorallia)
689 are congeneric. Zoological Science 23, 87–94.
690
691 Reimer, J.D., Ono, S., Tsukahara, J., Iwase, F., 2008b. Molecular characterization of the zoanthid
692 genus *Isaurus* (Anthozoa : Hexacorallia) and associated zooxanthellae (*Symbiodinium* spp.) from
693 Japan. Marine Biology 153, 351–363.
694
695 Reimer, J.D., Sinniger, F., 2010a. Discovery and description of a new species of *Abyssoanthus*
696 (Zoantharia: Hexacorallia) at the Japan Trench: the world's deepest known zoanthid. Cahiers De
697 Biologie Marine 51, 451–457.
698
699 Reimer, J.D., Sinniger, F., 2010b. Unexpected diversity in Canadian Pacific zoanthids (Cnidaria:
700 Anthozoa: Hexacorallia): a molecular examination and description of a new species from the
701 waters of British Columbia. Marine Biodiversity 40, 249–260.
702
703 Reimer, J.D., Sinniger, F., Fujiwara, Y., Hirano, S., Maruyama, T., 2007a. Morphological and
704 molecular characterisation of *Abyssoanthus nankaiensis*, a new family, new genus and new
705 species of deep-sea zoanthid (Anthozoa : Hexacorallia : Zoantharia) from a north-west Pacific
706 methane cold seep. Invertebrate Systematics 21, 255–262.
707
708 Reimer, J.D., Sinniger, F., Hickman, C.P., 2008c. Zoanthid diversity (Anthozoa : Hexacorallia) in
709 the Galapagos Islands: a molecular examination. Coral Reefs 27, 641–654.
710
711 Reimer, J.D., Takishita, K., Ono, S., Maruyama, T., 2007b. Diversity and evolution in the
712 zoanthid genus *Palythoa* (Cnidaria : Hexacorallia) based on nuclear ITS-rDNA. Coral Reefs 26,
713 399–410.
714
715 Reimer, J.D., Takishita, K., Ono, S., Tsukahara, J., Maruyama, T., 2007c. Molecular evidence
716 suggesting interspecific hybridization in *Zoanthus* spp. (Anthozoa: Hexacorallia). Zoological
717 Science 24, 346–359.
718
719 Reimer, J.D., Todd, P.A., 2009. Preliminary molecular examination of zooxanthellate zoanthids
720 (Hexacorallia: Zoantharia) and associated zooxanthellae (*Symbiodinium* spp.) diversity in
721 Singapore. Raffles Bulletin of Zoology 22, 103–120.
722
723 Risi, M.M., Macdonald, A.H.H., 2015. Possible synonymies of *Zoanthus* (Anthozoa:
724 Hexacorallia) species on the east coast of South Africa with Pacific congeners. Systematics and

bioRxiv preprint doi: https://doi.org/10.1101/161117; this version posted July 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

725 Biodiversity 13, 93–103.
726
727 Risi, M.M., Macdonald, A.H.H., 2016. Molecular examination of rocky shore brachycnemic
728 zoantharians (Anthozoa: Hexacorallia) and their *Symbiodinium* symbionts (Dinophyceae) in the
729 southwest Indian Ocean. Marine Biodiversity 46, 113–127.
730
731 Roch, S., Warnow, T., 2015. On the robustness to gene tree estimation error (or lack thereof) of
732 coalescent-based species tree methods. Systematic Biology 64, 663–676.
733
734 Rodríguez, E., Barbeitos, M.S., Brugler, M.R., Crowley, L.M., Grajales, A., Gusmão, L.,
735 Häussermann, V., Reft, A., Daly, M., 2014. Hidden among sea anemones: The first
736 comprehensive phylogenetic reconstruction of the order actiniaria (Cnidaria, Anthozoa,
737 Hexacorallia) reveals a novel group of hexacorals. PLoS One 9, e96998.
738
739 Rosenberg, N.A., 2013. Discordance of species trees with their most likely gene trees: a unifying
740 principle. Molecular Biology and Evolution 30, 2709–2713.
741
742 Ryland, J.S., 1997. Budding in *Acrozoanthus* Saville-Kent, 1893 (Anthozoa: Zoanthidea). In: den
743 Hartog, J. (Ed.), 6th International Conference on Coelenterate Biology. Nationaal
744 Natuurhistorisch Museum, pp. 423–428.
745
746 Ryland, J.S., Lancaster, J.E., 2003. Revision of methods for separating species of *Protopalythoa*
747 (Hexacorallia: Zoanthidea) in the tropical West Pacific. Invertebrate Systematics 17, 407–428.
748
749 Santos, M.E.A., Kitahara, M.V., Lindner, A., Reimer, J.D., 2015. Overview of the order
750 Zoantharia (Cnidaria: Anthozoa) in Brazil. Marine Biodiversity 46, 547–559.
751
752 Saville-Kent, W., 1893. The Great Barrier Reef of Australia: its products and potentialities.
753 Allen, W.H., London. pp. 1–387.
754
755 Shearer, T.L., Van Oppen, M.J.H., Romano, S.L., Worheide, G., 2002. Slow mitochondrial DNA
756 sequence evolution in the Anthozoa (Cnidaria). Molecular Ecology 11, 2475–2487.
757
758 Shiroma, E., Reimer, J.D., 2010. Investigations into the reproductive patterns, ecology, and
759 morphology in the zoanthid genus *Palythoa* (Cnidaria: Anthozoa: Hexacorallia) in Okinawa,
760 Japan. Zoological Studies 49, 182–194.
761
762 Sinniger, F., Haussermann, V., 2009. Zoanthids (Cnidaria: Hexacorallia: Zoantharia) from
763 shallow waters of the southern Chilean fjord region, with descriptions of a new genus and two
764 new species. Organisms Diversity & Evolution 9, 23–36.
765
766 Sinniger, F., Montoya-Burgos, J.I., Chevaldonne, P., Pawlowski, J., 2005. Phylogeny of the order
767 Zoantharia (Anthozoa, Hexacorallia) based on the mitochondrial ribosomal genes. Marine
768 Biology 147, 1121–1128.
769
770 Sinniger, F., Ocana, O.V., Baco, A.R., 2013. Diversity of zoanthids (Anthozoa: Hexacorallia) on

bioRxiv preprint doi: https://doi.org/10.1101/161117; this version posted July 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

771 Hawaiian seamounts: Description of the Hawaiian gold coral and additional zoanthids. PLos One
772 8, e52607.
773
774 Sinniger, F., Reimer, J.D., Pawlowski, J., 2008. Potential of DNA sequences to identify zoanthids
775 (Cnidaria: Zoantharia). Zoological Science 25, 1253–1260.
776
777 Sinniger, F., Reimer, J.D., Pawlowski, J., 2010a. The Parazoanthidae (Hexacorallia: Zoantharia)
778 DNA taxonomy: description of two new genera. Marine Biodiversity, 57–70.
779
780 Sinniger, F., Zelnio, K.A., Taviani, M., Reimer, J.D., 2010b. Presence of *Abyssoanthus* sp.
781 (Anthozoa: Zoantharia) in the Mediterranean Sea: an indication of non-dependence of
782 *Abyssoanthus* to chemosynthetic-based ecosystems. Cahiers De Biologie Marine 51, 475–478.
783
784 Springer, M.S., Gatesy, J., 2016. The gene tree delusion. Molecular Phylogenetics and Evolution
785 94, 1–33.
786
787 Stamatakis, A., 2014. RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of
788 Large Phylogenies. Bioinformatics 30, 1312–1313.
789
790 Stampar, S.N., Maronna, M.M., Vermeij, M.J.A., Silveira, F.L.D., Morandini, A.C., 2012.
791 Evolutionary diversification of banded tube-dwelling anemones (Cnidaria; Ceriantharia;
792 *Isarachnanthus*) in the Atlantic Ocean. PLoS One 9, e109481.
793
794 Streicher, J.W., Schulte, J.A., Wiens, J.J., 2016. How should genes and taxa be sampled for
795 phylogenomic analyses with missing data? An empirical study in iguanian lizards. Systematic
796 Biology 65, 128–145.
797
798 Swain, T.D., 2009. Phylogeny-based species delimitations and the evolution of host associations
799 in symbiotic zoanthids (Anthozoa, Zoanthidea) of the wider Caribbean region. Zoological
800 Journal of the Linnean Society 156, 223–238.
801
802 Swain, T.D., 2010. Evolutionary transitions in symbioses: dramatic reductions in bathymetric
803 and geographic ranges of Zoanthidea coincide with loss of symbioses with invertebrates.
804 Molecular Ecology 19, 2587–2598.
805
806 Swain, T.D., Schellinger, J.L., Strimaitis, A.M., Reuter, K.E., 2015. Evolution of anthozoan
807 polyp retraction mechanisms: convergent functional morphology and evolutionary allometry of
808 the marginal musculature in order Zoanthidea (Cnidaria: Anthozoa: Hexacorallia). BMC
809 Evolutionary Biology 15, 123.
810
811 Swain, T.D., Strimatis, A.M., Reuter, K.E., Boudreau, W., 2016. Towards integrative systematics
812 of Anthozoa (Cnidaria): Evolution of form in the order Zoanthidea. Zoologica Scripta DOI:
813 10.1111/zsc.12195.
814
815 Swain, T.D., Swain, L.M., 2014. Molecular parataxonomy as taxon description: examples from
816 recently named Zoanthidea (Cnidaria: Anthozoa) with revision based on serial histology of

bioRxiv preprint doi: https://doi.org/10.1101/161117; this version posted July 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

817 microanatomy. Zootaxa 3796, 81–107.
818
819 Swain, T.D., Taylor, D.J., 2003. Structural rRNA characters support monophyly of raptorial
820 limbs and paraphyly of limb specialization in water fleas. Proceedings of the Royal Society B-
821 Biological Sciences 270, 887–896.
822
823 Szymanski, M., Erdmann, V.A., Barciszewski, J., 2007. Noncoding RNAs database (ncRNAdb).
824 Nucleic Acids Research 35, D162–D164.
825
826 Torres-Suarez, O.L., 2014. *Gorgonia mariae* and *Antillogorgia bipinnata* populations inferred
827 from compensatory base change analysis of the internal transcribed spacer 2. Molecular
828 Phylogenetics and Evolution 79, 240–248.
829
830 Wei, N.W.V., Wallace, C.C., Dai, C.F., Pillay, K.R.M., Chen, C.A., 2006. Analyses of the
831 ribosomal internal transcribed spacers (ITS) and the 5.8S gene indicate that extremely high
832 rDNA heterogeneity is a unique feature in the scleractinian coral genus *Acropora* (Scleractinia;
833 Acroporidae). Zoological Studies 45, 404–418.
834

835 **Figure and table captions**

836 **Table 1.** Review of published Zoanthidea molecular phylogenies inferred from nucleotide
837 sequences, including the source reference (Reference), figure number of the phylogeny within
838 that source (Fig.), number of apparent species in each phylogeny (Sp.), number of terminals in
839 each phylogeny (Term.), outgroup employed in each phylogeny (Outgroup), genes used in the
840 inference of each phylogeny (Genes), and the genomic compartment sampled by the genes used
841 to infer each phylogeny (Genome). *Reimer et al., 2007b; Fig. 4.

842

843 **Table 2.** Partition definitions and per-partition parameter estimates used to model sequence
844 evolution for phylogenetic inference. Hypervariable regions that were staggered in the staggered
845 alignment, or excluded in the standard alignment, are alignment partitions ITS1, ITS2, 16S-HV,
846 and 12S-HV.

847

848 **Figure 1.** Example matrices aligned by standard (a) and staggered (b) protocols. Note that the

bioRxiv preprint doi: https://doi.org/10.1101/161117; this version posted July 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

849 standard alignment is "over-aligned", assuming homology where it is unlikely.

850

851 **Figure 2.** Complete-data tree. Maximum likelihood phylogeny of Zoanthidea based on a
852 staggered alignment of concatenated nuclear (18S, ITS1, 5.8S, ITS2, & 28S) and mitochondrial
853 (12S & 16S) ribosomal RNA and mitochondrial protein-coding (COI) nucleotide sequences.
854 Support indicated by 1000 pseudoreplicate maximum likelihood bootstrap values. Taxonomic
855 notations: order, Ac = Actiniaria, An = Antipatharia; family, M = Microzoanthidae, black bar =
856 Nanozoanthidae, gray diagonal lines = Parazoanthidae, gray bar = Abyssoanthidae, Hydrozoanth
857 = Hydrozoanthidae, gray horizontal line = Zoanthidae, N = Neozoanthidae.

858

859 **Figure 3.** Hypervariable-excluded tree. Maximum likelihood phylogeny of Zoanthidea based on
860 standard alignment of concatenated nuclear (18S, 5.8S, & 28S) and mitochondrial (12S & 16S)
861 ribosomal RNA and mitochondrial protein-coding (COI) nucleotide sequences, with
862 hypervariable regions of 12S & 16S (12S-HV, 16S-HV; Table 2) excluded. Support indicated by
863 1000 pseudoreplicate maximum likelihood bootstrap values. Taxonomic notations: order, Ac =
864 Actiniaria, An = Antipatharia; family, M = Microzoanthidae, black bar = Nanozoanthidae, gray
865 diagonal lines = Parazoanthidae, gray bar = Abyssoanthidae, Hydrozoanth = Hydrozoanthidae,
866 gray horizontal line = Zoanthidae, N = Neozoanthidae.

867

868 **Table S1.** Zoanthidea specimens and GenBank accession numbers for each of the sequences
869 used for phylogenetic inference. Accessions of sequences not included in the alignment of Swain
870 (2010) (TreeBASE: S10492) are in bold.

bioRxiv preprint doi: https://doi.org/10.1101/161117; this version posted July 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1.

Reference	Fig.	Sp.	Term.	Outgroup	Genes	Genome
Reimer et al., 2004	4	4	29	*Palythoa*	COI	Mit
Sinniger et al., 2005	3	21	25	Actininaria	16S, 12S	Mit
Reimer et al., 2006b	2a	5	13	*Palythoa*	16S	Mit
Reimer et al., 2006b	2b	5	16	*Palythoa*	5.8S	Nuc
Reimer et al., 2006c	3	8	29	*Parazoanthus*	COI	Mit
Reimer et al., 2006c	4	8	12	*Parazoanthus*	16S	Mit
Reimer et al., 2006a	10	6	47	*Parazoanthus*	COI	Mit
Reimer et al., 2006a	11	6	15	*Parazoanthus*	16S	Mit
Reimer et al., 2007b	3	8	52	*Zoanthus*	16S	Mit
Reimer et al., 2007b	4	5	67	*Palythoa*	ITS	Nuc
Reimer et al., 2007c	3	4	50	*Parazoanthus*	COI	Mit
Reimer et al., 2007c	4	4	92	*Parazoanthus*	5.8S	Nuc
Reimer et al., 2007a	3	25	30	Scler. & Actin.	COI	Mit
Reimer et al., 2007a	4	20	23	Scler. & Actin.	16S	Mit
Reimer et al., 2008a	5	15	23	*Abyssoanthus*	16S	Mit
Reimer et al., 2008a	6	21	31	*Abyssoanthus*	COI	Mit
Reimer et al., 2008a	7	5	9	*Parazoanthus*	ITS	Nuc
Reimer et al., 2008c	2	22	40	Epizoanthidae	16S	Mit
Reimer et al., 2008c	3	26	46	Epizoanthidae	COI	Mit
Reimer et al., 2008c	4	4	15	*Parazoanthus*	ITS	Nuc
Sinniger et al., 2008	1	36	54	Epizoanthidae	COI	Mit
Reimer et al., 2008c	2	36	54	Epizoanthidae	16S	Mit
Reimer et al., 2008b	3	13	24	*Parazoanthus*	16S	Mit
Reimer et al., 2008b	4	11	27	*Parazoanthus*	COI	Mit
Sinniger and Haussermann, 2009	2	41	59	Epizoanthidae	COI, 16S	Mit
Reimer and Todd, 2009	3	11	22	*Parazoanthus*	16S	Mit
Reimer and Todd, 2009	4	10	44	*Parazoanthus*	COI	Mit
Swain, 2009	1	10	64	Actininaria	ITS	Nuc
Swain, 2009	2	17	17	Actininaria	16S	Mit
Reimer and Sinniger, 2010a	1	21	27	Epizoanthidae	16S	Mit
Sinniger et al., 2010b	1	40	56	Epizoanthidae	16S	Mit
Sinniger et al., 2010a	3	35	48	*Isozoanthus*	COI	Mit
Sinniger et al., 2010a	4	36	51	Epizoanthidae	16S	Mit
Sinniger et al., 2010a	5	38	79	Epizoanthidae	ITS	Nuc
Reimer et al., 2010a	2a	9	17	*H. gracilis*	16S	Mit
Reimer et al., 2010a	2b	10	18	*H. gracilis*	COI	Mit
Reimer et al., 2010b	3a	31	42	Parazoanthidae	16S	Mit
Reimer et al., 2010b	3b	33	47	–	COI	Mit
Shiroma and Reimer, 2010*	3	5	67	*P. heliodiscus*	ITS	Mit
Reimer et al., 2010c	3a	6	9	*P. tuberculosa*	COI long	Mit
Reimer et al., 2010c	3b	7	19	*P. tuberculosa*	COI medium	Mit
Reimer et al., 2010c	3c	8	21	*P. tuberculosa*	COI short	Mit
Reimer et al., 2010c	3d	9	16	*Palythoa*	16S	Mit
Reimer et al., 2010c	3e	6	23	P. tuberculosa	ITS2	Nuc
Reimer and Sinniger, 2010b	4	41	68	Epizoanthidae	16S	Mit
Reimer and Sinniger, 2010b	S1	13	59	Epizoanthidae	COI	Mit
Reimer and Sinniger, 2010b	S2	10	28	Hydrozoanthidae	ITS	Nuc
Reimer and Fujii, 2010	5a	17	39	Zoanthidae	16S	Mit
Reimer and Fujii, 2010	5b	17	45	Brachycnemina	COI	Mit
Reimer and Fujii, 2010	6	3	14	Hydrozoanthidae	ITS	Nuc
Swain, 2010	1	93	93	Actiniaria	18S–28S, 16S, 12S, COI	M&N
Aguilar and Reimer, 2010	2	8	16	*Palythoa*	ITS2	Nuc
Aguilar and Reimer, 2010	S2	7	9	*Palythoa*	ITS2	Nuc
Fujii and Reimer, 2011	7	10	35	Actiniaria	COI	Mit

bioRxiv preprint doi: https://doi.org/10.1101/161117; this version posted July 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Reference	Fig.	Sp.	Term.	Outgroup	Genes	Genome
Fujii and Reimer, 2011	8	3	17	*Parazoanthus*	ITS	Nuc
Fujii and Reimer, 2011	S1	11	30	Actiniaria	16S	Mit
Reimer et al., 2011a	2	16	30	*Parazoanthus*	16S	Mit
Reimer et al., 2011a	S2	11	26	Hydrozoanthidae	COI	Mit
Reimer et al., 2011b	9	12	34	H. gracilis	16S	Mit
Reimer et al., 2011b	S1	10	35	H. gracilis	COI	Mit
Reimer et al., 2012a	2a	32	46	Parazoanthus	16S	Mit
Reimer et al., 2012a	2b	17	28	Hydrozoanthidae	COI	Mit
Reimer et al., 2012a	2c	23	35	*Parazoanthus*	COI	Mit
Reimer et al., 2012a	2d	25	32	Hydrozoanthidae	ITS	Nuc
Reimer et al., 2012b	2	11	39	*Zoanthus*	16S	Mit
Reimer et al., 2012b	3	11	35	*Zoanthus*	COI	Mit
Reimer et al., 2012b	4	8	65	*Palythoa*	ITS	Nuc
Reimer et al., 2013a	1	15	135	E. illoricatus	COI	Mit
Reimer et al., 2013a	2	10	120	H. gracilis	16S	Mit
Reimer et al., 2013a	3	4	55	P. heliodiscus	ITS	Nuc
Fujii and Reimer, 2013	1	17	23	Actiniaria	16S	Mit
Fujii and Reimer, 2013	2	14	19	Actiniaria	COI	Mit
Sinniger et al., 2013	4	29	35	Epizoanthidae	18S, COI, 16S	M&N
Sinniger et al., 2013	S1	32	36	Microzoanthidae	16S	Mit
Hibino et al., 2014	2a	4	93	*Palythoa*	COI	Mit
Hibino et al., 2014	2b	5	118	*Palythoa*	16S	Mit
Hibino et al., 2014	3	3	64	*Palythoa*	ITS	Mit
Reimer et al., 2013b	1	5	9	*Zoanthus*	16S	Mit
Reimer et al., 2014	2	30	44	unrooted	16S & COI	Mit
Reimer et al., 2014	3a	12	21	*Palythoa*	ITS	Nuc
Reimer et al., 2014	3b	5	17	U. parasiticus	ITS	Nuc
Koupaei et al., 2014	3	9	37	*Parazoanthus*	16S	Mit
Montenegro et al., 2015	7	16	88	*Antipathozoanthus*	ITS	Nuc
Montenegro et al., 2015	8	14	92	*Antipathozoanthus*	COI	Mit
Montenegro et al., 2015	9	13	82	*Antipathozoanthus*	16S	Mit
Montenegro et al., 2015	10	5	61	*Parazoanthus*	ALG11	Nuc
Montenegro et al., 2015	11	16	22	*Antipathozoanthus*	ITS	Nuc
Montenegro et al., 2016	2	48	48	Epizoanthidae	18S–28S, 16S, COI	M&N
Montenegro et al., 2016	SD1	25	25	Epizoanthidae	18S	Nuc
Montenegro et al., 2016	SD1	31	53	Epizoanthidae	ITS	Nuc
Montenegro et al., 2016	SD1	18	18	Epizoanthidae	28S	Nuc
Montenegro et al., 2016	SD1	45	67	Epizoanthidae	16S	Mit
Montenegro et al., 2016	SD1	23	34	Epizoanthidae	COI	Mit
Risi and Macdonald, 2015	5	8	10	*Hydrozoanthus*	COI	Mit
Risi and Macdonald, 2015	6	8	10	*Hydrozoanthus*	16S	Mit
Risi and Macdonald, 2015	7	8	15	*Hydrozoanthus*	ITS	Nuc
Irei et al., 2015	7	13	34	*Palythoa*	ITS	Nuc
Irei et al., 2015	8	17	61	Zoanthidae	16S	Mit
Irei et al., 2015	9	17	59	Zoanthidae	COI	Mit
Santos et al., 2015	2	17	36	Epizoanthidae	16S, COI	Mit
Koupaei et al., 2016	3	13	80	*Hydrozoanthus*	16S	Mit
Koupaei et al., 2016	4	16	44	*Hydrozoanthus*	COI	Mit
Kise and Reimer, 2016	5	9	48	*Hydrozoanthus*	ITS	Nuc
Kise and Reimer, 2016	6	16	60	*Hydrozoanthus*	16S	Mit
Kise and Reimer, 2016	7	11	62	*Hydrozoanthus*	COI	Mit
Risi and Macdonald, 2016	4	14	142	*Hydrozoanthus*	16S	Mit
Risi and Macdonald, 2016	5	10	52	*Hydrozoanthus*	ITS	Nuc

Table 2

partition	concatenated alignment positions	base frequencies A	C	G	T	substitution rates (G–T = 1) A–C	A–G	A–T	C–G	C–T	gamma shape
18S	1–1798	0.2665	0.2027	0.2592	0.2714	0.8489	1.3323	0.4201	1.0333	48.6716	0.0919
ITS1	1799–6594	0.2441	0.2490	0.2516	0.2552	1.1178	1.6339	0.9692	1.2628	1.8670	1.0728
5.8S	6595–6751	0.2346	0.2116	0.2794	0.2742	1.1664	2.6667	0.5464	1.4403	3.2678	0.1733
ITS2	6752–9515	0.2313	0.2616	0.2670	0.2399	1.1315	2.2442	1.4078	1.3658	2.4699	0.7796
28S	9516–13231	0.2473	0.2388	0.2804	0.2333	0.5129	1.2414	0.4359	0.5037	9.7887	0.1520
16S	13848–13941, 14127–14459, 15294–15421, 15863–16060	0.3036	0.1996	0.2635	0.2331	0.8519	3.1614	1.8538	0.6183	5.9487	0.1885
16S-HV	13232–13847, 13942–14126, 14460–15293, 15422–15862	0.2300	0.2758	0.2732	0.2208	0.9120	2.2374	1.6214	0.5731	3.0695	0.4515
12S	16061–16243, 16307–16387, 16567–16792, 16857–16922, 16957–17188	0.3080	0.1931	0.2631	0.2356	1.2070	3.3024	1.7436	0.2116	4.3271	0.1470
12S-HV	16244–16306, 16388–16566, 16793–16856, 16923–16956, 17189–17311	0.1973	0.3163	0.3065	0.1797	1.0257	1.9514	4.7599	0.3503	3.8633	0.4427
COI-1	17312-18075\3	0.1933	0.2327	0.3088	0.2650	2.3507	7.6712	4.3862	0.2389	9.2735	0.9192
COI-2	17313-18075\3	0.2644	0.1864	0.3102	0.2388	1.1284	2.0395	0.6799	0.2599	29.7862	0.1132
COI-3	17314-18075\3	0.1367	0.2533	0.1637	0.4462	0.0001	0.0001	0.4517	783.7444	9.6590	0.0689

Fig. 1

a.

```
GTTTCAGAA-----TATTAGT
GTTTCAGAA-----TATTAGT
GTTTCAGGAG----TATTAGT
GTTTCAGGAC----TATTAGT
GTATCAGAGAT-CCTATTAGT
GTATCAGAGATGCCTATTAGT
GTATCAGAGATGCGTATTAGT
```

b.

```
GTTTCAGAA??????????TATTAGT
GTTTCAGAA??????????TATTAGT
GTTTCAG??GAG???????TATTAGT
GTTTCAG??GAC???????TATTAGT
GTATCAG?????AGAT-CCTATTAGT
GTATCAG?????AGATGCCTATTAGT
GTATCAG?????AGATGCGTATTAGT
```

Fig. 2

0.1

Microzoanthus kagerou |k1|
Microzoanthus occultus
Microzoanthus kagerou |k2|
Nanozoanthus harenaceus
Isozoanthus giganteus
Isozoanthus cf. giganteus
Epizoanthus sp. |Deep Med|
Epizoanthus aff. illoricatus
Epizoanthus beriber
Epizoanthus illoricatus
Epizoanthus inazuma
Epizoanthus paguricola
Epizoanthus sp. |Sub-Antarctic|
Epizoanthus scotinus
Epizoanthus incrustatus
Epizoanthus lindhali
Epizoanthus aff. arenaceus |S2|
Epizoanthus aff. arenaceus |S7|
Epizoanthus couchii
Epizoanthus arenaceus
Epizoanthus cf. arenaceus
Epizoanthus fiordicus
Epizoanthus ramosus
Epizoanthus cf. balanorum
Epizoanthus aff. arenaceus |C|
Epizoanthus aff. arenaceus |HI|
Antipathozoanthus sp. |M2|
Antipathozoanthus hickmani
Antipathozoanthus macaronesicus
Parazoanthid sp. |NC2|
Parazoanthid sp. |CORSARO|
Parazoanthid sp. |EBISCO|
Parazoanthid sp. |NC3|
Abyssoanthus nankaiensis
Abyssoanthus convalis
Mesozoanthus lilkweminensis
Mesozoanthus fossii
Zibrowius sp. |3242|
Zibrowius ammophilus
Parazoanthid sp. |2296|
Hurlizoanthus parrishi
Hurlizoanthus sp. |4098|
Corallizoanthus aff. tsukaharai |CA|
Bullagummizoanthus emilyacadiaarum
Corallizoanthus aff. tsukaharai |NZ|
Kulamanamana haumeaae
Corallizoanthus tsukaharai
Kauluzoanthus kerbyi
Savalia savaglia
Savalia lucifica
Umimayanthus sp. |Tas|
Umimayanthus aff. parasiticus |SUL|
Umimayanthus chanpuru
Umimayanthus aff. parasiticus |MAD|
Umimayanthus nakama
Umimayanthus miyabi
Umimayanthus sp. |47J|
Umimayanthus parasiticus
Parazoanthus cf. elongatus |NZ|
Parazoanthus cf. elongatus |Chi|
Parazoanthus juan-fernandezii
Parazoanthus aff. juan-fernandezii |FRA|
Parazoanthus aff. juan-fernandezii |CA|
Parazoanthus cf. hertwigi
Parazoanthus axinellae
Parazoanthus anguicomus
Parazoanthus capensis
Parazoanthus sp. |269|
Parazoanthus darwini
Parazoanthus aff. swiftii |Sal|
Parazoanthus aff. swiftii |NUR|
Parazoanthus swiftii
Parazoanthus sp. |1401|
Bergia catenularis
Bergia puertoricense
Bergia sp. |SUL5|
Bergia aff. catenularis |SEN|
Bergia cutressi
Terrazoanthus sp. |302|
Terrazoanthus sp. |YP|
Terrazoanthus sp. |276|
Terrazoanthus sp. |FI|
Terrazoanthus patagonichus
Terrazoanthus onoi
Terrazoanthus sp. |02-27|
Terrazoanthus minutus
Terrazoanthus californicus
Terrazoanthus sinnigeri
Hydrozoanthus gracilis
Hydrozoanthus cf. gracilis |NC2|
Hydrozoanthus cf. gracilis |NC1|
Hydrozoanthus antumbrosus
Hydrozoanthus aff. tunicans |SUL|
Hydrozoanthus tunicans
Hydrozoanthidae sp. |Mada1|
Isaurus sp. |B|
Isaurus tuberculatus
Isaurus sp. |WA|
Isaurus tuberculatus |FS05|
Neozoanthus uchina
Neozoanthus caleyi
Neozoanthus sp. |Iran|
Zoanthus sociatus
Zoanthus pulchellus
Zoanthus sansibaricus |B|
Zoanthus cf. sansibaricus
Zoanthus gigantus
Zoanthus solanderi
Zoanthus sp. |FS227|
Zoanthus praelongus
Zoanthus natalensis
Zoanthus durbanensis
Zoanthus aff. pulchellus
Zoanthus kuroshio
Zoanthus aff. vietnamensis
Zoanthus vietnamensis
Acrozoanthus australiae
Zoanthus sansibaricus
Palythoa caribaeorum
Palythoa tuberculosa
Palythoa cf. tuberculosa |289|
Palythoa aff. caesia
Palythoa mutuki |1618|
Palythoa grandiflora
Palythoa heliae
Palythoa aff. mutuki |Carib|
Palythoa cf. caribaeorum
Palythoa aff. mutuki |265|
Palythoa aff. mutuki |Iran|
Palythoa aff. mutuki |1142|
Palythoa aff. sakurajimensis
Palythoa aff. clavata
Palythoa aff. mutuki |1167|
Palythoa aff. psammophilia
Sphenopus marsupialis
Sphenopus marsupialis |JRS3|
Palythoa mizigama
Palythoa umbrosa
Palythoa heliodiscus
Palythoa singaporensis
Palythoa variabilis
Palythoa aff. variabilis
Palythoa cf. grandis
Palythoa grandis
Trissopathes pseudotristicha
Dendrobathypathes boutillieri
Tanacetipathes barbadensis
Acanthopathes thyoides
Antipathes atlantica
Leiopathes glaberrima
Relicanthus daphneae
Edwardsiid sp. |BAR|
Edwardsiid sp. |CUR|
Nematostella vectensis
Edwardsia elegans

M
Epizoanthidae
Parazoanthidae
Hydrozoanth
N
Zoanthidae
Sphenopidae
Macrocnemina
Brachycnemina
Zoanthidea
An
Ac

Fig. 2