Integrating phylogeographic and ecological niche approaches to delimitating cryptic lineages in the blue-green damselfish (*Chromis viridis*)

Shang Yin Liu [Corresp., 1, 2] , **Mao-Ning Tuanmu** [3] , **Rita Rachmawati** [4, 5] , **Gusti Ngurah Mahardika** [6] , **Paul H Barber** [5]

[1] Department of Marine Biotechnology and Resources, National Sun Yat-Sen University, Kaohsiung, Taiwan

[2] Doctoral Degree Program in Marine Biotechnology, National Sun Yat-Sen University, Kaohsiung, Taiwan

[3] Biodiversity Research Center, Academia Sinica, Taipei, Taiwan

[4] Center for Fisheries Research, Ministry of Marine Affairs and Fisheries, Jakarta, Indonesia

[5] Department of Ecology and Evolutionary Biology, University of California, Los Angeles, LA, United States

[6] The Indonesian Biodiversity Research Centre, University of Udayana, Bali, Indonesia

Corresponding Author: Shang Yin Liu
Email address: syvliu@mail.nsysu.edu.tw

Species delimitation is challenging in sibling species/cryptic lineages because of the absence of clear diagnostic traits. However, integration of different approaches such as phylogeography and ecological niche comparison offers one potential approach to tease apart recently diverged lineages. In this study, we estimate the ecological niche divergence among lineages in *Chromis viridis* in a broad-scale phylogeographic framework to test whether the combination of these two approaches can effectively distinguish recently diverged lineages. Results from Cytb and Rag2 analyses identified two cryptic lineages (*C. viridis A* and *C. viridis B*) that diverged ~3Myr ago. Estimates of ecological niche divergence with 11 environmental parameters across the broad geographic range of these lineages showed overlapping ecological niches and niche conservatism. However, regardless of the incongruence between genetic and ecological niche divergence, the substantial genetic divergence between the two clades of *C. viridis* in both mtDNA and nuclear loci strong suggest that they are cryptic taxa.

1   Integrating phylogeographic and ecological niche approaches to delimitating cryptic lineages in

2   the blue-green damselfish (*Chromis viridis*).

3   Shang-Yin Vanson Liu[1*], Mao-Ning Tuanmu[2], Rita Rachmawati[3,4], Gusti Ngurah Mahardika[5],

4   Paul H. Barber[3]

5   [1] Department of Marine Biotechnology and Resources, National Sun Yat-Sen University,

6   Kaohsiung, Taiwan

7   [2] Biodiversity Research Center, Academia Sinica, Taipei, Taiwan

8   [3] Department of Ecology and Evolutionary Biology, University of California Los Angeles, Los

9   Angeles, USA

10 [4] Center for Fisheries Research, the Ministry of Marine Affairs and Fisheries, Jakarta, Indonesia

11 [5] The Indonesian Biodiversity Research Centre, The Animal Biomedical and Molecular Biology

12 Laboratory of Udayana University, Jl Sesetan-Markisa 6, Denpasar, Bali, Indonesia

14 *Correspondence: Shang Yin Vanson Liu, Department of Marine Biotechnology and Resources,
15 National Sun Yat-Sen University, Kaohsiung, Taiwan. E-mail: syvliu@mail.nsysu.edu.tw

17 **Abstract**

18 Species delimitation is challenging in sibling species/cryptic lineages because of the absence of

19 clear diagnostic traits. However, integration of different approaches such as phylogeography and

20 ecological niche comparison offers one potential approach to tease apart recently diverged

21 lineages. In this study, we estimate the ecological niche divergence among lineages in *Chromis*

22 *viridis* in a broad-scale phylogeographic framework to test whether the combination of these two

23 approaches can effectively distinguish recently diverged lineages. Results from Cytb and Rag2

24 analyses identified two cryptic lineages (*C. viridis A* and *C. viridis B*) that diverged ~3Myr ago.

25 Estimates of ecological niche divergence with 11 environmental parameters across the broad

26 geographic range of these lineages showed overlapping ecological niches and niche

27 conservatism. However, regardless of the incongruence between genetic and ecological niche

28 divergence, the substantial genetic divergence between the two clades of *C. viridis* in both

29 mtDNA and nuclear loci strong suggest that they are cryptic taxa.

31 **Introduction**

32 Comparative phylogeography examines patterns of congruence in phylogenetic breaks across

33 species distribution. It is built on the assumption that the processes driving lineage divergence,

34 speciation, and the evolution of biodiversity mainly involves geographical, historical, and

35 environmental factors that favor isolation and limit gene flow between populations (*Avise, 2000;*

36 *Rocha et al., 2007*).

37 In the marine realm, phylogeographic studies frequently identify cryptic species,

38 morphologically indistinguishable groups that have differences in neutral genetic markers that

39 are equal or greater than those observed between species with diagnostic morphological traits

40 (Knowlton, 1993). In some cases (e.g. *Drew & Barber, 2009; Drew et al., 2010; Hubert et al.,*

41 *2012; Liu et al., 2012*) such cryptic lineages are subsequently described as new species (*Allen &*

42 *Drew, 2012; Liu et al., 2012; Allen, 2015*), indicating that such cryptic lineages represent

43 overlooked biodiversity.

45 Systematists use different characteristics to differentiate species, and the variation in these

46 characteristics can arise at different times and rates during the process of lineage diversification

47 (*de Queiroz, 2007*). However, species delimitation is particularly difficult in adaptive or recent

48 radiations, when nascent species boundaries and their defining characteristics can be unclear

49 (*Puebla et al., 2008; Wagner et al., 2012; Victor, 2015*). In such cases, integration of different

50 approaches can increase the ability to detect recently separated lineages (*Leache et al., 2009*)

51   and can provide stronger evidence for lineage separation when the results are concordant (*Liu et*

52   *al., 2012; Allen & Erdmann, 2012; Larkin et al., 2016*).

53

54   While integrating morphological diversification and genetic divergence is relatively common in

55   addressing recent lineage diversification (e.g. *Allen & Drew, 2012; Allen & Erdmann, 2012*),

56   ecological diversification (i.e. niche divergence) is seldom used to differentiate cryptic

57   lineages/species in the marine realm. Niche diversification is a foundation of speciation (*Pyron et*

58   *al., 2009*). Therefore, if niche divergence leads to lineage differentiation in the process of

59   populations adapting to new environments (*Wiens, 2004; Wiens & Graham, 2005*), then a niche

60   partitioning should be expected between lineages, particularly among recently diverged

61   overlapping cryptic lineages identified in many phylogeographic studies.

62

63   To date, only a handful of studies have found niche divergence between cryptic marine taxa. For

64   example, studies on deep-sea octocorals showed that niche partitioning between lineages is

65   associated with the depth (*Quattrini et al., 2013; Quattrini et al., 2017*) and the marine

66   cyanobacteria (genus *Prochlorococcus*) exhibits niche partitioning in associated with geography

67   and environmental conditions. The absence of niche differentiation in the study of cryptic marine

68   taxa is partially due to the lack of a centralized high-resolution spatial data representing both

69   benthic and pelagic marine environments (*Sbrocco & Barber, 2013*), resources that are now

70   available.

71   *Chromis viridis* is widely distributed Indo-Pacific coral reef fish. Both juveniles and adults

72   associate with *Acropora* corals where they school and feed on zooplankton in the water column

73   above coral heads (*Frédérich et al., 2009*). Using phylogenetic analyses based on mtDNA

74   variation, Froukh & Kochzius (2008) found three cryptic lineages in *C. viridis* across four Indo-

75   Pacific localities, and Messmer et al. (2012) documented the presence of additional cryptic

76   lineages in the Great Barrier Reef. However, it is unclear how geographical and ecological

77   processes contribute to this nascent diversification.

78

79   To better understand the processes driving lineage diversification in *C. viridis*, we conducted a

80   broad scale phylogeographic study of *C. viridis* across its distribution range. We then estimated

81   ecological niche divergence between lineages in *C. viridis* in this phylogeographic framework to

82   test whether the combination of these two approaches can effectively distinguish recently

83   diverged lineages.

84

85   **Materials & Methods**

86   **Sample collection and DNA extraction**

87   We collected 252 *Chromis viridis* between 2007 and 2017 from 15 locations across its Indo-

88   Pacific distribution (Table 1; Fig. 1). We collected specimens by hand net and clove oil, either by

89   scuba diving or snorkeling and preserved tissue samples (fin clips, a piece of muscle, or both) in

90   95% alcohol and stored at 4 °C. Large fish were released following fin clipping, but individuals

91   too small for fin clipping were euthanized with clove oil and preserved whole in 95% ethanol.

92

93   We extracted genomic DNA from tissue fragments using Geneaid Tissue Genomic DNA mini

94   Kit (Geneaid Biotech, Taiwan) following manufacturers' protocol and eluted extracted DNA in

95   TE buffer and stored at -20 °C until amplification by polymerase chain reaction (PCR).

96

97   **Amplification of genetic markers**

98

99 We amplified a portion of the mitochondrial cytochrome b (Cytb) gene using universal primers

100 GluDG-L and H16460 (*Palumbi, 1996*). PCRs reactions were 30 μL in volume, containing 10-40

101 ng template DNA, 3 μL 10X buffer, 0.2 mM dNTPs, 1.5 mM MgCl$_2$, 10 mM each primer, and

102 0.2 units of Taq polymerase (MDbio, Taipei). The thermocycling profile consisted of initial

103 denaturation at 94 °C for 2 min, followed by 41 cycles of denaturation at 94 °C for 30 s,

104 annealing at 57 °C for 30 s, and extension at 72 °C for 40 s, concluding with a final extension at

105 72 °C for 2 min.

106

107 Because mitochondrial and nuclear genes can have very different histories, we also amplified the

108 nuclear recombination-activating (Rag2) gene for a subset of samples that represented distinct

109 clades recovered in the mtDNA, using primers RAG2F and RAG2R (*Westneat & Alfaro, 2005*).

110 PCR reactions were as above, except that we used a 5mM MgCl concentration and the following

111 thermocycling parameters: 39 cycles of denaturation at 94 °C for 30 s, annealing at 56 °C for 30

112 s, and extension at 72 °C for 40 s, and a final extension at 72 °C for 2 min.

113

114 The nucleotide sequences of the PCR products of both loci were determined using ABI 3730XL

115 automated sequencer (Carlsbad CA, U.S.A.) by Genomics (https://www.genomics.com.tw).

116 Nucleotide sequences were assembled and edited using the SEQUENCHER version 4.2 software

117 (Gene Code, Ann Arbor MI, U.S.A.). Sequences were uploaded to GenBank under the accession

118 number MH743228-MH743691.

119

120 **Phylogenetic analyses**

121 Prior to any analyses, we aligned DNA sequences from each gene region in Clustal W

122 (Thompson et al., 1994) and exported these sequences to MEGA 6 (*Tamura et al., 2013*) to

123 visually inspect all alignments for accuracy. To quantify genetic diversity measures, we

124 calculated standard genetic diversity indexes including haplotype diversity (*h*) and nucleotide

125 diversity (π) in Arlequin 3.5 (*Excoffier & Lischer, 2010*). We then inferred phylogenetic trees

126 from each individual locus using maximum likelihood (ML) and Bayesian inference performed

127 on the CIPRES Science Gateway (Miller et al., 2010), and in BEAST 2.4.5 (*Bouckaert et al.,*

128 *2014*). We conducted the ML analyses in RAxML version 8.1.24 (*Stamatakis, 2014*) using the

129 GTR+G substitution model. For Bayesian inference, we used MrBayes version 3.2.2 (*Ronquist et*

130 *al., 2012*), implementing two parallel runs of four simultaneous Markov chains for 10 million

131 generations, sampling every 1000 generations and using the default parameters. Run parameters

132 employed unlinked substitution models, rated heterogeneity models, and based frequencies

133 across partitions. In the Bayesian analyses the first one million generations (10%) were discarded

134 as burn-in, based on log- likelihood tree scores. Meanwhile, the convergence diagnostic was

135 applied and the stop probability was set to 0.01. Nodal support was evaluated individually for all

136 trees using non- parametric bootstrapping with 1000 maximum likelihood replicates as employed

137 in RAxML (*Stamatakis, 2014*), and by calculation of posterior probabilities as employed in

138 MrBayes. Lastly, we generated a median-joining haplotype networks based on Cytb and Rag2

139 sequence datasets by using Popart 1.7 (http://popart.otago.ac.nz).

140

141 To date the ages of *Chromis viridis* clades, we arbitrarily subsampled 2 Cytb sequences from

142 each clade; we also included two outgroups, including *Chromis atrilobata* (AY208524.1) and

143 *Chromis multilineata* (AY208533.1) to calibrate date ranges as these two taxa diverged 3.1Mya,

144 (*Quenouille et al., 2004*). We created the XML BEAST input file using the software BEAUti

145 v.1.8.2. (as implemented in BEAST), with the setting of 50 million generations under the

146 uncorrelated relaxed clock model, and sampling trees once every 1000 generations. To test for

147 inter-run variation, we conducted two independent runs in BEAST 2.4.5 (*Bouckaert et al., 2014*),

148 and then checked these runs for convergence with the software Tracer v1.5 (available at

149 http://tree.bio.ed.ac.uk/software/tracer/). After discarding the 20% burn-in, we pooled the

150 remaining tree samples from the two runs into a combined file to calculate the maximum clade

151 credibility tree. From this tree, we calculated the posterior mean divergence ages and 95%

152 credibility intervals (CI) for all nodes using Tree Annotator v1.8.2 (implemented in the BEAST

153 package). To compare the similarity of two gene tree topologies, Phylo.io software (*Robinson et*

154 *al., 2016*) was used.

155

156 **Ecological niche characterization and comparison**

157

158 To characterize the ecological niches of *C. viridis* (*C. viridis* Clade A) and potential cryptic

159 species (*C. viridis* Clade B), we obtained geophysical, biotic, and environmental data layers for

160 global sea surface from Bio-ORACLE (downloaded on Feb. 20th, 2018; *Tyberghein et al., 2012;*

161 *Assis et al., 2018*). We then extracted the values of those factors for the locations where *C.*

162 *viridis* samples were collected. We excluded factors such as ice thickness and sea ice

163 concentration from our analyses because all sample sites were tropical and ice free.

164

165 To explore niche differentiation between *C. viridis* Clade A and *C. viridis* Clade B, we first

166 compared the values of individual environmental factors between *C. viridis* Clade A and Clade B

167   sampling sites using Mann-Whitney tests. Next, we plotted niche differences in a two-

168   dimensional non-metric multidimensional scaling (NMDS) space based on the Euclidean

169   dissimilarity of those factors. Due to the different units of the environmental factors, we

170   standardized factor values before running the NMDS analysis. Lastly, we tested equivalency of

171   ecological niches between the species based on environmental factors of *C. viridis* Clade A and

172   Clade*s* B sampling locations in a Principle Components Analysis framework. Due to potential

173   correlations among environmental factors, we ran the principal component analysis (PCA) on all

174   environmental factors and then used the resulting principal components to define niches of the

175   species. Because *C. viridis* mainly occur in coral reefs, we restricted our PCA to regions (grid

176   cells) where the maximum depth was equal or less than 50m based on the bathymetry data

177   obtained from MARSPEC (http://www.marspec.org/; *Sbrocco & Barber, 2013*). To make the

178   spatial resolution of the bathymetry data layer, which is originally at 30-by-30-arc second

179   resolution, consistent with the Bio-ORACLE data, we upscaled it to 5-by-5-arc minute resolution

180   with each pixel value in the upscaled data layer being the maximum value over 10x10 original

181   pixels within that new pixel.

182

183   To measure niche similarity between *C. viridis* Clade A Clade B, we calculated the *D* and *I*

184   statistics developed by Warren et al. (2008). These two indices range from 0 and 1 with higher

185   values indicating more similar niches. To test for niche equivalency between these two cryptic

186   lineages of *C. viridis, we* compared the observed similarity values to null distributions of

187   similarity obtained through a randomization procedure under a hypothesis of an identical niche

188   (*Warren et al., 2008*), using the "dismo" package in R (version 3.4.2).

189
190

191 **Results**

192 We sequenced 911 base pairs of mitochondrial Cytb, and 743 base pairs of nuclear gene (Rag2,

193 743 bp) from 248 individual *Chromis viridis*, representing 15 locations across the Indo-Pacific

194 (Fig. 1). In total, there were 135 unique haplotypes of Cytb; nucleotide diversity (π) ranged from

195 0.0033 to 0.0544, and haplotype diversity (h) ranged from 0.8824 to 1 among locations. For

196 Rag2, 30 samples would not amplify, resulting in a total of 218 Rag2 sequences, representing

197 fourteen unique haplotypes after omitting those sites with ambiguity codes (e.g. Y and K),

198 additionally, there were 40 parsimony informative sites after alignment and 26 (heterozygotes)

199 out of 40 were contained ambiguous signals. The nucleotide diversity (π) and haplotype diversity

200 (h) of Rag2 ranged from 0 to 0.0154 and 0.9933 to 1, respectively (Table 1). These two genetic

201 diversity indexes could be over-estimated since those ambiguous sites were been considered as

202 variable sites during computation in Arlequin 3.5.

203

204 The maximum likelihood tree based on Cytb gene revealed three deeply divergent lineages

205 within *C. viridis*; Clade A, Clade A-1, and Clade B (Fig. 2). In contrast, the corresponding Rag2

206 likelihood tree only differentiated Clades A and B, with no additional divergence within Clade 1.

207 However, Clades A and B were not concordant across Cytb and Rag2; four mtDNA Clade B

208 samples, one from Komodo and three from Lizard Island, clustered with Clade A based on Rag2.

209 Additionally, one sample from Komodo contained a Clade B Rag2 sequence, but a Clade A Cytb

210 sequence (Table S1).

211

212 Whether based on Cytb or Rag2, all Clade B haplotypes came from only three locations; Lizard

213 Island on the Great Barrier Reef, Komodo Island in Indonesia, and the island of Fiji. In contrast,

214 Clade A occurred broadly throughout the Indo-Pacific, including all three Clade B localities.

215   Based solely on mtDNA, Clade A was further divided into two clades; Clade A was broadly

216   distributed, and Clade A-1 was restricted to the Red Sea, although the Red Sea clade was not

217   recovered in the Rag2 phylogeny. The general tree topology of these two genes was highly

218   similar (Figure S1), except the Clade A-1 and those samples clustered with it, which can be

219   revealed only by Cytb.

220

221

222   The patterns from the median joining network tree were identical to the phylogenetic trees based

223   on either Cytb and Rag2. However, a few individuals assigned to Clade B based on Cytb

224   clustered with Clade A based on their Rag2 sequences (Fig. 2).

225

226   Results from BEAST indicated that *C. viridis* Clade A and Clade B diverged approximately

227   3Mya, with the divergence between Clade A and A-1 dating to 1.2Mya (Fig. 3).

228

229   **Niche divergence tests**

230

231   Results spanning 11 environmental factors (Table 2) showed no significant difference in

232   ecological niche parameters of Clade A or Clade B of *C. viridis*. Moreover, there was also no

233   significant environmental difference between the locations where two clades co-occur and those

234   where only Clade B occurs (Table 2), suggesting that the environmental factors cannot explain

235   why Clade B co-occurs with Clade A at some locations, but not at others.

236

237   The NMDS analysis also showed a lack of niche differentiation. The first two NMDS axes

238   explained 86.7% of the variation in the 11 environmental factors across the 15 sampling sites.

239   The three sampling sites where *C. viridis* B occurs fall at the center of the convex hull of all

240   sampling sites in the two-dimensional NMDS space (Fig. 4), indicating complete niche overlap

241   between Clades A and B. The first four principal components of the environmental factors,

242   which were standardized before the analysis, had eigenvalues higher than 1. Together they

243   explain about 83.1% of the total variation across all the grid cells with the maximum water depth

244   less than or equal to 50m. Based on the component scores corresponding to the locations where

245   Clades A and B were found, niche similarity indices were high, with $D = 0.841$ and $I = 0.973$.

246   Moreover, niche equivalency tests showed that none of the two values were significantly

247   different from the values obtained through a randomization process under the hypothesis of an

248   identical niche (*P*-values for D and I were 0.311 and 0.294, respectively).

249

250   **Discussion**

251   Range-wide phylogeographic analyses of the blue green damselfish, *C. viridis* revealed two

252   divergent lineages in both mitochondrial Cytb and nuclear Rag2 DNA sequence data. These

253   cryptic lineages were first reported by Froukh and Kochzius (2008) in the Coral Triangle and

254   Red Sea, and subsequently by Messmer et al. (2012) who examined Australia and French

255   Polynesia. However, these studies each only covered a fraction of the geographic range of *C.*

256   *viridis*. By examining patterns across its entire range, this study shows that these cryptic lineages

257   of *C. viridis* are sympatric over a portion of their Pacific range.

258

259   Moreover, previous studies (*Froukh & Kochzius, 2008*; *Messmer et al., 2012*) only used mtDNA

260   markers, providing incomplete insights into genetic structure due to its maternal inheritance

261   (*Prugnolle & De Meeûs, 2002*; *Daly-Engel et al., 2012*). By sequencing both mtDNA and

262   nuclear markers, this study confirms the presence of two cryptic clades in *C. viridis,* Clades A

263   and B. These clades diverged approximately 3Mya similar to the divergence of *C. atrilobata* and

264   *C. multilineata* that were separated by Isthmus of Panama (*Domingues, et al. 2005*). Given the

265   depth of this divergence and the largely concordant differentiation of Clade A and B in Cytb and

266   Rag2, it is likely that these two lineages represent distinct species, raising interesting questions

267   about the sympatry of Clades A and B in part of their range.

268

269   **Origin of lineage diversification**

270   Although the divergence of Clades A and B of *C. viridis* is clear, the origin of this divergence is

271   not. Within the Indo-Pacific region, the "Indo-Pacific Barrier" has been considered as a soft

272   barrier that could be the main driving force of marine biological provinces in this region (*Gaither*

273   *et al., 2011*). Barber & Bellwood (2005) and Cowen & Bellwood (2013) examined the

274   importance of this barrier by evaluating the extent of temporal concordance in vicariance in three

275   prominent families of reef fish, including Labridae, Pomacentridae and Chaetodontidae. Both

276   studies showed that the isolation effect of Indo-Pacific barrier on the widely distributed fishes

277   mainly occurred mostly between end Miocene and Early Pliocene (2-6 Myr), and the majority of

278   vicariance events occurred in a narrow time interval at approximately 2.5 Myr.

279

280   The date of divergence of Clade A and B broadly conforms to the onset of Plio-Pleisetocene sea

281   level fluctuations (*Voris, 2000*). However, Clade B populations of *C. viridis* occur on islands

282   surround by deep water where lowered sea levels would not result in land barriers that promote

283   vicariance, unlike other marine taxa distributed on opposite sides of the Sunda Shelf (e.g. *Barber*

284 *et al., 2006*; *Crandall et al., 2008*; *DeBoer et al., 2014a, 2014b*; *Crandall et al., 2014*; *Waldrop*

285 *et al., 2016*; *Simmonds et al., 2018*). Given the Indo-Pacific wide range of *C. viridis* and a

286 divergence time between two cryptic clades dating to 3Mya, isolation across the Indo-Pacific

287 Barrier is the most likely driver of divergence between Clades A and B. Under this scenario,

288 Clade A would be an Indian Ocean clade that has expanded into the Pacific, where it now

289 overlaps with the Pacific Clade B, a process previously noted in Neritid snails (*Crandall et al.,*

290 *2008*).

291

292 An alternate, but not mutually exclusive explanation for divergence of Clade B comes from

293 recent studies of coral associated marine taxa. Similar to phytophagus insects that undergo

294 ecological divergence associated with host switching (Berlocher & Feder, 2002; Hébert et al.,

295 2016), recent studies from the marine realm demonstrate that changes in coral host taxa can

296 promote lineage divergence, potentially leading to speciation (*Simmonds et al., 2018*). Samples

297 were not collected in a way that allows us to test this hypothesis, but future studies separating

298 samples by coral host to determine whether individuals from Clades A and B exist in mixed

299 schools in sympatric populations, and if so, whether those schools are associated with different

300 coral hosts.

301

302 The geographic distribution of Clade B is curious in that it is observed in Komodo, but not in

303 other populations in the Lesser Sunda Islands (e.g. Amed, Nusa Lembongan). Similarly, Clade B

304 is observed on the Great Barrier Reef and Fiji, but not in New Caledonia, a population located

305 between these too population. One explanation for this disjunct distribution is that Clade B

306 individuals are relatively rare, and that greater sampling intensity would reveal Clade B

307   haplotypes in adjacent ranges, as would be expected. Alternatively, it is possible that Clade A is

308   gradually displacing Clade B populations, and the areas of sympatry represent locations where

309   this process is incomplete. Similar arguments were made to explain sympatry of highly divergent

310   clades of marine snails in the Pacific Ocean (*Crandall et al., 2008*).

311

312   In contrast to the divergence of Clade A and B, the divergence of Clade A-1, unique to the Red

313   Sea, is more easily interpreted as the result of vicariance. The Red Sea is a semi-enclosed basin

314   that is frequently invoked in driving population differentiation of reef organisms between the

315   Red Sea and Indian Ocean (*Dibattista et al., 2013*). However, while mitochondrial DNA shows a

316   well-supported Red Sea sub clade within *C. viridis* Clade A, this clade is not confirmed with

317   nuclear Rag2 sequence.

318   Lack of concordance between the mtDNA and nuclear phylogenies is not surprising. First, there

319   is substantially more genetic variation in the mtDNA sequences because DNA repair

320   mechanisms are less efficient in mtDNA, resulting in much higher substitution rates (*Alexeyev et*

321   *al., 2013*). Second, because mitochondrial genome is maternally inherited and haploid, its

322   effective population size is one-fourth that of a nuclear gene, meaning that lineage sorting occurs

323   more rapidly in mtDNA (*Hare, 2001*). Thus, mitochondrial gene trees have a substantially higher

324   probability of accurately recovering recently divergence lineages with short internode distances

325   than do nuclear genes (Moore, 1995). Given that the Red Sea mtDNA clade only dates to 1.2

326   Mya, it is unsurprising that divergence in this region is not recovered in the Rag2 sequences.

327   Thus, while the Red Sea population is clearly a sub-population of *C. viridis* Clade A, it is

328   unlikely a cryptic species, like Clades A and B.

329

Cryptic Species

330

331 Many phylogeographic studies in the Indo-Pacific find highly divergent clades, often with

332 regions of sympatry (i.e. *Barber et al., 2002, 2006*; *Crandall et al., 2008*; *Gaither et al., 2011*;

333 *Liu et al., 2012, DeBoer et al., 2014*; *Bowen et al., 2016*). In many cases, these cryptic clades

334 may represent cryptic species, similar to Liu et al. (2012) where a divergent clade of

335 *Pomacentrus coelestis* in Micronesia was subsequently described as a new species (Liu et al.,

336 2013). The concordant phylogeographic patterns in mtDNA and nuclear DNA, and the depth of

337 this divergence suggests that Clades A and B of *C. viridis* may represent two cryptic species with

338 different geographic ranges; Clade A is a widely distributed Indo-Pacific, while Clade B is only

339 found in Lizard Island, Komodo, and Fiji.

340

341 Comparison of Cytb and Rag2 trees revealed a limited amount of discordance. Given the

342 relatively recent divergence, one potential explanation for this pattern is incomplete lineage

343 sorting (*Tang et al., 2012*). Incomplete lineage sorting is the simplest explanation for individuals

344 with Clade B *C. viridis* haplotypes in the mtDNA tree having Clade A *C. viridis* Rag2 sequences

345 (Table S1). However, incomplete lineage sorting could not explain the sequence of komo4

346 clustered with *C. viridis* clade B in the Rag2 tree but clustered with *C. viridis* Clade A in the

347 mtDNA tree. The later finding could likely explain by the consequence of hybridization.

348

349 Hybridization is in *C. viridis* is unsurprising as closely related Pomacentrids often have

350 overlapping geographic distributions, co-occur in the same microhabitats (e.g. colonies of

351 branching corals) (*Randall et al., 1977*), and spawn sympatrically (*Jan, 1996*). Hybridization has

352 been observed in several sibling species of damselfish, including *Abudefduf*

353   *abdominalis×Abudefduf vaigiensis* (*Coleman et al., 2014*), *Amphiprion*

354   *chrysopterus×Amphiprion sandaracinos* (*Gainsford et al., 2015*), *Amphiprion*

355   *mccullochi×Amphiprion akindynos* (*van der Meer et al., 2012*), *Dascyllus carneus×D.*

356   *marginatus* (*DiBattista et al., 2015*), and *D. reticulatus×D. aruanus* (*He et al., 2017*). As such,

357   both hybridization and incomplete lineage sorting are likely responsible for the discordant

358   patterns in the two markers.

359

360   **Incongruence between genetic and ecological niche divergence**

361

362   Hutchinson (1978) proposed that two species cannot occupy the same ecological niche, yet niche

363   conservatism predicts that closely related taxa retain ancestral ecological affiliations and persists

364   in similar environments (*Lord et al., 1995*; *Webb et al., 2002*; *Wiens & Graham, 2005*).

365   Allopatric sister taxa are often characterized by niche conservatism, but because geographic

366   isolation drives speciation (*Peterson et al., 1999*; *Peterson et al., 2001*; *Kozak & Wiens, 2006*)

367   sibling species can't compete for the same niche space, because, by definition, they have non-

368   overlapping geographic ranges.

369

370   The alternative to niche conservatism is niche divergence, in which sister taxa occupy different

371   niches (*Dayan & Simberloff, 2005*). Niche divergence is typically associated with sympatric

372   speciation as diversification results from reduction in gene flow associated with divergence in

373   traits with ecological function (e.g. habitat segregation, pollinator divergence, behavioral

374   changes, phenological shifts, and mating system shifts). Under niche divergence, young sister

375   species with high degrees of range overlap should ecologically diverge (*Dayan & Simberloff,*

376   *2005*; *Davies et al., 2007*). While ecological modeling and phylogenetics are commonly

377   integrated to understand the relationship between evolutionary and ecological divergence of

378   sibling species (*Kalkvik et al., 2012*; *Schorr et al., 2012*), few studies examine the intra-specific

379   (cryptic lineage) level (*Gutiérrez-Tapia & Palma, 2016*).

380

381   Our results show that Clade A and B of *C. viridis* are sympatric, with Clade B haplotypes

382   occurring only at three localities. There are two possible explanations for this pattern. First,

383   lineage diversification could have occurred in sympatry, as seen in Indo-Pacific coralivorous

384   snails (*Simmonds et al., 2018*). Alternatively, diversification could have occurred in allopatry

385   with secondary overlap in geographic ranges as proposed for neritid snails (*Crandall et al.,*

386   *2008*). Niche comparisons based on 11 environmental factors showed that Clade A has a broader

387   ecological niche than Clade B does (Fig 4). Despite this difference, there is no significant niche

388   differentiation between these two lineages. As such, diversification is more likely to have arisen

389   with niche conservatism rather than niche differentiation.

390

391   Given the absence of niche differentiation, our results suggest that the diversification of Clades

392   A and B most likely occurred in allopatry, and that secondary contact is the most likely

393   explanation for their current distribution. If true, niche conservatism could help explain the

394   disjunct distribution of Clade B. Under this scenario, Clade A and Clade B would compete for

395   the same niche space when they occur in sympatry. If Clade A is a superior competitor, it would

396   gradually displace Clade B populations, potentially explaining the relative rarity of Clade B

397   haptotypes and their disjunct distributions.

398

399   Alternatively, Pyron et al. (2015) suggested that niche conservatism could result from soft

400   allopatry where there is low environmental heterogeneity. As noted above, the shallow Sunda

401   and Sahul continental shelves exposed during low sea level stands (*Voris 2000*), forming long

402   land bridges that restricted larval exchange between the tropical Indian Ocean and the western

403   Pacific (reviewed in *Randall, 1998*). However, because deep water passages in what is modern

404   day Indonesia remained open, and because low sea level stands were intermittent, the Indo-

405   Pacific Barrier has been considered a soft dispersal barrier for marine taxa (*Cowman &*

406   *Bellwood, 2013*). The 11 environmental factors we used to define the niches of the two lineages

407   of *C. viridis* only reflect broad-scale environmental conditions. However, there could be subtle

408   variation in environmental factors, or variation could occur at finer scales (e.g. microhabitat),

409   resulting in soft-allopatric divergence.

410

411   It is important to note that the broad environmental variation we examine does not capture the

412   potential diet shifts observed in sibling species (*Goodheart et al., 2017*), novel traits (*Liu et al.,*

413   *2018*), and/or micro habitat preference (*Whitney et al., 2018*) that could act to drive or reinforce

414   lineage diversification. The latter is particularly intriguing given recent studies demonstrating

415   ecological speciation resulting from shifts in coral hosts (*Simmonds et al., 2018*). Further studies

416   such as stable isotope analysis and detailed morphological examinations in a phylogenetic

417   framework are needed to better understand the ecological and morphological divergence between

418   these two lineages.

419

420

421

422   **Conclusions**

423 Although cryptic diversification in widespread marine species is common (*Hubert et al., 2012*),

424 phylogeographic studies typically ignore the potential role of ecological niche partitioning on

425 lineage diversification. The present study shows evolutionary divergence between two lineages

426 of *C. viridis* that have overlapping ecological niches, supporting niche conservatism. It is unclear

427 whether this pattern results from allopatric divergence with secondary contact, or from subtle

428 differentiation in ecological niches not captured by the broad scale environmental data used to

429 compare ecological niches. Regardless of the origins, the substantial genetic divergence between

430 the two clades of *C. viridis* in both mtDNA and nuclear loci strong suggest that they are cryptic

431 taxa. Because *C. viridis* is highly exploited in the aquarium trade (*Wabnitz, 2003*), we suggest

432 that a higher conservation priority needs to be given on the Clade B lineages restricted to

433 Australia, Indonesia, and Fiji to protect this unique genetic lineage.

434

435

436

437

438

439

440

441 **Acknowledgements**

442 We thank Dr. S. Cheng, Dr. M-J Ho, Dr. Y-R Cheng, F-T Chang, N. Narendra and A. Sembiring

443 for their assistant in the field, and Dr. V. Messmer, Dr. B. Frédérich, Dr. F. Borsa, Professor S.

444 Planes and S. Johnson for sharing tissue samples. Special thanks to A. C. Bentley, for his

445 curatorial help on obtaining tissue samples from Ichthyology session of Kansas University. We

446 deeply thank the Indonesian Government and the Ministry of Research and Technology for

447 permitting our research activities in Indonesia under research permit

448 No.272/SIP/FRP/SM/VII/2013. This study was funded by Ministry of Science and Technology,

449 Taiwan (Grant Number: MOST 106-2611-M-110-009). All applicable international, national,

450   and institutional guidelines for the use of animals were followed. We declare no conflict of

451   interest.

452

453

454

455

456   **References**

457   **Alexeyev M, Shokolenko I, Wilson G, LeDoux S. 2013.** The maintenance of mitochondrial

458   DNA integrity-critical analysis and update. *Cold Spring Harbor Perspectives In Biology* **5:**

459   a012641

460

461   **Allen GR, Erdmann MV. 2012.** *Reef fishes of the East Indies*. Perth, Australia: Tropical Reef

462   Research

463

464   **Allen GR, Drew J. 2012.** *Pomacentrus maafu* a new species of damselfish from the Southwest

465   Pacific. *Aqua: International Journal of Ichthyology* **18:** 171-180

466

467   **Allen GR, Erdmann MV, Kurniasih E. 2015.** *Chrysiptera caesifrons*, a new species of

468   damselfish (Pomacentridae) from the south-western Pacific Ocean. *Journal of the Ocean Science*

469   *Foundation* **15:** 16-32

470

471   **Assis J, Tyberghein L, Bosch S, Verbruggen H, Serrão EA, De Clerck O. 2018.** Bio-

472   ORACLE v2.0: Extending marine data layers for bioclimatic modelling. *Global Ecology*

473   *Biogeography* **27:** 277–284

474

475   **Avise JC. 2000.** *Phylogeography: the history and formation of species*. Cambridge, MA:

476   Harvard university press.

477

478   **Avise JC. 2009.** Phylogeography: Retrospect and prospect. *Journal of Biogeography* **36:** 3–15

479

480   **Barber PH, Palumbi SR, Erdmann MV, Moosa MK. 2002.** Sharp genetic breaks among

481   populations of *Haptosquilla pulchella* (Stomatopoda) indicate limits to larval transport: patterns,

482   causes, and consequences. *Molecular Ecology* **11:** 659-674

483

484   **Barber PH, Palumbi SR, Erdmann MV. 2006.** Comparative phylogeography of three

485   codistributed stomatopods: origins and timing of regional lineage diversification in the Coral

486   Triangle. *Evolution* **60:** 1825-1839

487

488   **Berlocher SH, Feder JL. 2002.** Sympatric speciation in phytophagous insects: moving beyond

489   controversy? *Annual Review of Entomology* **47:** 773-815

490

491   **Bowen BW, Karl SA, Pfeiler E. 2007.** Resolving evolutionary lineages and

492   taxonomy of bonefishes (*Albula spp*.). In: Ault, J.S. (Ed.), Biology and

493   Management of the World Tarpon.

494

495   **Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut**
496   **A, Drummond AJ. 2014.** BEAST 2: A Software Platform for Bayesian Evolutionary Analysis.
497   *PLoS Computational Biology* **10:** e1003537
498

499 **Webb CO, Ackerly DD, McPeek MA, Donoghue MJ. 2002.** Phylogenies and community

500 ecology. *Annual review of ecology and systematics* **33**: 475-505

501

502 **Coleman RR, Gaither MR, Kimokeo B, Stanton FG, Bowen BW, Toonen RJ. 2014.** Large-

503 scale introduction of the Indo-Pacific damselfish *Abudefduf vaigiensis* into Hawai'i promotes

504 genetic swamping of the endemic congener *A. abdominalis*. *Molecular Ecology* **23:** 5552–5565

505

506 **Cowman PF, Bellwood DR. 2013.** The historical biogeography of coral reef fishes: global

507 patterns of origination and dispersal. *Journal of Biogeography* **40:** 209 - 224

508

509 **Crandall ED, Frey M, Grossberg RK. Barber, PH. 2008.** Contrasting demographic history

510 and phylogeographical patterns in two Indo-Pacific gastropods. *Molecular Ecology* **17:** 611-626

511

512 **Crandall ED, Treml EA, Liggins L, Gleeson L, Yasuda N, Barber PH, Gert W, Riginos, C.**

513 **2014.** Return of the ghosts of dispersal past: historical spread and contemporary gene flow in the

514 blue sea star *Linckia laevigata. Bulletin of Marine Science* **90:** 399-425

515

516 **Daly-Engel TS, Seraphin KD, Holland KN, Coffey JP, Nance HA, Toonen RJ, Bowen BW.**

517 **2012.** Global phylogeography with mixed-marker analysis reveals male-mediated dispersal in the

518 endangered scalloped hammerhead shark (*Sphyrna lewini*). *PLoS One* **7:** e29986

519

520 **Dayan T, Simberloff D. 2005.** Ecological and community-wide character displacement: the next

521 generation. *Ecology Letters* **8:** 875-894

522

523   **Davies TJ, Meiri S, Barraclough TG, Gittleman JL. 2007.** Species co-existence and character

524   divergence across carnivores. *Ecology Letters* **10:** 146–152

525

526   **DeBoer TS, Naguit MR, Erdmann MV, Ablan-Lagman MCA, Ambariyanto, Carpenter**

527   **KE, Toha AHA, Barber PH. 2014a.** Concordant phylogenetic patterns inferred from

528   mitochondrial and microsatellite DNA in the giant clam *Tridacna crocea*. *Bulletin of Marine*

529   *Science* **90:** 301-329

530

531   **DeBoer TS, Naguit MR, Erdmann MV, Ablan-Lagman MCA, Ambariyanto, Carpenter K**

532   **E, Toha AHA, Barber PH. 2014b.** Concordance between phylogeographic and biogeographic

533   boundaries in the Coral Triangle: conservation implications based on comparative analyses of

534   multiple giant clam species *Bulletin of Marine Science* **90:** 277-300

535

536   **de Queiroz K. 2007.** Species concepts and species delimitation. *Systematic Biology* **56:** 879 –

537   886

538

539   **DiBattista JD, Berumen ML, Gaither MR, Rocha LA, Eble JA, Choat JH, Craig MT,**

540   **Sinclair-Taylor TH, Bowen BW. 2013.** After continents divide: comparative phylogeography

541   of reef fishes from the Red Sea and Indian Ocean. *Journal of Biogeography* **40:** 1170-1181.

542

543   **DiBattista JD, Rocha L A, Hobbs J-PA, He S, Priest MA, Sinclair-Taylor**

544 **TH, Bowen BW, Berumen ML. 2015.** When biogeographical provinces collide: hybridization

545 of reef fishes at the crossroads of marine biogeographical provinces in the Arabian Sea. *Journal*

546 *of Biogeography* **42:** 1601– 1614

547

548 **Drew JA, Barber PH. 2009.** Sequential cladogenesis of *Pomacentrus moluccensis* (Bleeker,

549 1853) supports the peripheral origin of marine biodiversity in the Indo-Australian Archipelago.

550 *Molecular Phylogenetics and Evolution* **53:** 335–339

551

552 **Drew JA, Allen GR, Erdmann MV. 2010.** Congruence between genes and color morphs in a

553 coral reef fish: population variability in the Indo-Pacific damselfish *Chrysiptera rex* (Snyder,

554 1909). *Coral Reefs* **29:** 439-444

555

556 **Frédérich B, Fabri G, Lepoint G, Vandewalle P, Parmentier E. 2009.** Trophic niches of

557 thirteen damselfishes (Pomacentridae) at the Grand Recif of Toliara, Madagascar. *Ichthyological*

558 *Reserch* **56:** 10–17

559

560 **Froukh T, Kochzius M. 2008.** Species boundaries and evolutionary lineages in the blue green

561 damselfishes *Chromis viridis* and *Chromis atripectoralis* (Pomacentridae). *Journal of Fish*

562 *Biology* **72:** 451-457

563

564 **Gainsford A, Herwerden L, Jones GP. 2015.** Hierarchical behaviour, habitat use and species

565 size differences shape evolutionary outcomes of hybridization in a coral reef fish. *Journal of*

566 *Evolutionary Biology* **28:** 205–222

567

568 **Gaither MR, Bowen BW, Bordenave TR, Rocha LA, Newman SJ, Gomez JA, van**

569 **Herwerden L, Craig MT. 2011.** Phylogeography of the reef fish *Cephalopholis argus*

570 (Epinephelidae) indicates Pleistocene isolation across the Indo-Pacific Barrier with

571 contemporary overlap in the Coral Triangle. *BMC evolutionary biology* **11:** 189

572

573 **Goodheart JA, Bazinet AL, Valdés Á, Collins AG, Cummings MP 2017.** Prey preference

574 follows phylogeny: evolutionary dietary patterns within the marine gastropod group

575 Cladobranchia (Gastropoda: Heterobranchia: Nudibranchia). *BMC evolutionary biology* **17:** 221

576

577 **Gutiérrez-Tapia P, Palma RE. 2016.** Integrating phylogeography and species distribution

578 models: cryptic distributional responses to past climate change in an endemic rodent from the

579 central Chile hotspot. *Diversity and Distributions* **22:** 638-650

580

581 **Hare MP. 2001.** Prospects for nuclear gene phylogeography. *Trends in Ecology & Evolution* **16:**

582 700-706

583

584 **Hébert JB, Scheffer SJ, Hawthorne DJ. 2016.** Evidence for ecological speciation via a host

585 shift in the holly leaf miner, *Phytomyza glabricola* (Diptera: Agromyzidae). *Ecology and*

586 *Evolution* **6:** 6565-6577

587

588   **He S, Johansen JL, Hoey AS, Pappas MK, Berumen ML. 2017.** Molecular confirmation of

589   hybridization between *Dascyllus reticulatus×Dascyllus aruanus* from the Great Barrier Reef.

590   *Marine Biodiversity*, 1-10

591

592   **Hubert N, Meyer CP, Bruggemann HJ, Guérin F, Komeno RJL, Espiau B, Causse R,**

593   **Williams JT. Planes S. 2012.** Cryptic diversity in Indo-Pacific coral-reef fishes revealed by

594   DNA-barcoding provides new support to the centre-of-overlap hypothesis. *PLoS one* **7:** e28987

595

596   **Hutchinson GE. 1978.** An Introduction to Population Ecology (Yale Univ Press, New Haven,

597   CT).

598   **Davies JT, Meiri S, Barraclough TG, Gittleman JL. 2007.** Species coexistence and character

599   divergence across carnivores. *Ecology letters* **10:** 146-152

600

601   **Domingues VS, Bucciarelli G, Almada VC, Bernardi G. 2005.** Historical colonization and

602   demography of the Mediterranean damselfish, *Chromis chromis*. *Molecular Ecology* **14:** 4051-

603   4063

604

605   **Kalkvik HM, Stout IJ, Doonan TJ, Parkinson CL. 2012.** Investigating niche and lineage

606   diversification in widely distributed taxa: phylogeography and ecological niche modeling of the

607   *Peromyscus maniculatus* species group. *Ecography* **35:** 54-64

608

609   **Knowlton N. 1993.** Sibling species in the sea. *Annual Review of Ecology and Systematics* **24:**

610   189-216

611

612 **Kocher TD, Conroy JA, McKaye KR, Stauffer JR. 1993.** Similar morphologies of cichlid fish

613 in Lakes Tanganyika and Malawi are due to convergence. *Molecular Phylogenetic and Evolution*

614 **2:** 158–165

615

616 **Kozak KH, Wiens JJ. 2006.** Does niche conservatism promote speciation? A case study in

617 North American salamanders. *Evolution* **60:** 2604-2621

618

619 **Leaché AD, Koo MS, Spencer CL, Papenfuss TJ, Fisher RN, McGuire JA. 2009.**

620 Quantifying ecological, morphological, and genetic variation to delimit species in the coast

621 horned lizard species complex (*Phrynosoma*). *Proceedings of the National Academy of Sciences*

622 **106:** 12418-12423

623

624 **Larkin AA, Blinebry SK, Howes C, Lin Y, Loftus SE, Schmaus CA, Zinser ER, Johnson Z**

625 **I. 2016.** Niche partitioning and biogeography of high light adapted *Prochlorococcus* across

626 taxonomic ranks in the North Pacific. *The ISME journal* **10:** 1555

627

628 **Lessios HA. 2008.** The Great American Schism: divergence of marine organisms after the rise of

629 the Central American isthmus. *Annual Review of Ecology, Evolution, and Systematics* **39:** 63–91

630

631 **Liu SYV, Dai C-F, Allen GR, Erdmann MV. 2012.** Phylogeography of the neon damselfish

632 *Pomacentrus coelestis* indicates a cryptic species and different species origins in the West

633 Pacific Ocean. *Marine Ecology Progress Series* **458:** 155–167

634

635 **Liu SYV, Ho H-C, Dai C-F. 2014.** A new species of *Pomacentrus* (Actinopterygii:

636 Pomacentridae) from Micronesia, with comments on its phylogenetic relationships. *Zoological*

637 *Studies* **52:** 1–8

638

639 **Liu SYV, Frederich B, Lavoué S, Chang J, Erdmann MV, Mahardika GN, Barber PH.**

640 **2018.** Buccal venom gland associates with increased of diversification rate in the fang blenny

641 fish *Meiacanthus* (Blenniidae; Teleostei). *Molecular phylogenetics and evolution* **125:** 138-146

642

643 **Lord J, Westoby M, Leishman M. 1995.** Seed size and phylogeny in six temperate floras:

644 constraints, niche conservatism, and adaptation. *The American Naturalist* **146:** 349-364

645

646 **Marske K, Leschen R, Buckley T. 2012.** Concerted versus independent evolution and the

647 search for multiple refugia: comparative phylogeography of four forest beetles. *Evolution* **66:**

648 1862–1877

649

650 **Messmer V, Jones GP, Munday PL, Planes S. 2012.** Concordance between genetic and species

651 diversity in coral reef fishes across the Pacific Ocean biodiversity gradient. *Evolution* **66:** 3902-

652 3917

653

654 **Miller M, Pfeiffer W, Schwartz T. 2010.** Creating the CIPRES science gateway for inference

655 of large phylogenetic trees. *Gateway Computing Environments Workshop (GCE)*, 1-8

656

657   **Moore WS. 1995.** Inferring phylogenies from mtDNA variation: mitochondrial-gene trees

658   versus nuclear-gene trees. *Evolution* **49:** 718-726

659

660   **Palumbi SR. 1996.** Nucleic acid II: the polymerase chain reaction. In: Hillis, D.M., Moritz, G.,

661   Mable, B.K. (eds.). *Molecular Systematics. Sunderland, Mass.: Sinauer Associates*, 205–247

662

663   **Peterson AT, Soberón J, Sánchez-Cordero V. 1999.** Conservatism of ecological niches in

664   evolutionary time. *Science* **285:** 1265–1267

665

666   **Peterson AT, Sánchez-Cordero V, Soberon J, Bartley J, Buddemeier RW, Navarro-**

667   **Sigüenza AG. 2001.** Effects of global climate change on geographic distributions of Mexican

668   Cracidae. *Ecological modelling* **144:** 21-30

669

670   **Prugnolle F, De Meeûs T. 2002.** Inferring sex-biased dispersal from population genetic tools: a

671   review. *Heredity 88*, 161

672

673   **Puebla O, Bermingham E, Guichard F. Whiteman E. 2007.** Colour pattern as a single trait

674   driving speciation in *Hypoplectrus* coral-reef fishes? *Proceedings of the Royal Society B* **274:**

675   1265–1271

676

677   **Pyron RA, Burbrink FT. 2009.** Lineage diversification in a widespread species: roles for niche

678   divergence and conservatism in the common king snake, *Lampropeltis getula*. *Molecular*

679   *ecology* **18:** 3443-3457

680

681 **Pyron RA, Costa GC, Patten MA, Burbrink FT. 2015.** Phylogenetic niche conservatism and

682 the evolutionary basis of ecological speciation. *Biological Reviews* **90:** 1248-1262

683

684 **Quattrini AM, Georgian SE, Byrnes L, Stevens A, Falco R, Cordes EE. 2013.** Niche

685 divergence by deep-sea octocorals in the genus *Callogorgia* across the continental slope of the

686 Gulf of Mexico. *Molecular Ecology* **22:** 4123-4140

687

688 **Quattrini AM, Gómez CE, Cordes EE. 2017.** Environmental filtering and neutral processes

689 shape octocoral community assembly in the deep sea. *Oecologia* **183:** 221-236

690

691 **Quenouille B, Bermingham E, Planes S. 2004.** Molecular systematics of the damselfishes

692 (Teleostei: Pomacentridae): Bayesian phylogenetic analyses of mitochondrial and nuclear DNA

693 sequences. *Molecular phylogenetics and evolution* **31:** 66-88

694

695 **Rambaut A, Drummond AJ. 2014.** BEAST 2: A Software Platform for Bayesian Evolutionary

696 Analysis. *PLoS Computational Biology* **10:** e1003537

697

698 **Randall HA, Allen GR. 1977.** A revision of the damselfish genus *Dascyllus* (Pomacentridae)

699 with the description of a new species. *Records of the Australian Museum* **31:** 349-385

700

701 **Randall JE. 1998.** Zoogeography of shore fishes of the Indo-Pacific region. *Zoological Studies*

702 **37:** 227-268

703

704 **Robinson O, Dylus D, Dessimoz C. 2016.** Phylo.io: Interactive viewing and compar- ison of

705 large phylogenetic trees on the web. *Molecular Biology and Evolution* **33:** 2163–2166

706

707 **Rocha LA, Craig MT, Bowen BW. 2007.** Phylogeography and the conservation of coral reef

708 fishes. *Coral Reefs* **26:** 501-512

709

710 **Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L,**

711 **Huelsenbeck, JP. 2012.** MrBayes 3.2: efficient Bayesian phylogenetic inference and model

712 choice across a large model space. *Systematic Biology* **61:** 539-542

713

714 **Sbrocco EJ, Barber PH. 2013.** MARSPEC: ocean climate layers for marine spatial ecology.

715 *Ecology* **94:** 979-979

716

717 **Schorr G, Holstein N, Pearman PB, Guisan A, Kadereit JW. 2012.** Integrating species

718 distribution models (SDMs) and phylogeography for two species of Alpine Primula. *Ecology and*

719 *Evolution* **2:** 1260-1277

720

721 **Simmonds SE, Chou V, Cheng SH, Rachmawati R, Calumpong HP, Mahardika GN,**

722 **Barber PH. 2018.** Evidence of host-associated divergence from coral-eating snails (genus

723 *Coralliophila*) in the Coral Triangle. *Coral Reefs* **37:** 355-371

724 **Stamatakis A. 2014.** RAxML version 8: a tool for phylogenetic analysis and post-analysis of

725 large phylogenies. *Bioinformatics* **30:** 1312-1313

749

750 **Voris HK. 2000.** Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and

751 time durations. *Journal of Biogeography* **27:** 1153−1167

752 **Wabnitz C. 2003.** From ocean to aquarium: the global trade in marine ornamental species (No.

753 17). UNEP/Earthprint

754

755 **Waldrop E, Hobbs JPA, Randall, JE, DiBattista JD, Rocha LA, Kosaki RK, Berumen ML,**

756 **Bowen BW. 2016.** Phylogeography, population structure and evolution of coral-eating

757 butterflyfishes (Family Chaetodontidae, genus *Chaetodon*, subgenus Corallochaetodon). *Journal*

758 *of Biogeography* **43:** 1116-1129

759

760 **Warren DL, Glor RE, Turelli M. 2008.** Environmental niche equivalency versus conservatism:

761 quantitative approaches to niche evolution. *Evolution* **62:** 2868-2883

762

763 **Westneat MW, Alfaro ME. 2005.** Phylogenetic relationships and evolutionary history of the

764 reef fish family Labridae. *Molecular phylogenetics and evolution* **36:** 370-390

765

766 **Whitney JL, Donahue MJ, Karl SA. 2018.** Niche divergence along a fine-scale ecological

767 gradient in sympatric color morphs of a coral reef fish. *Ecosphere* **9:** e02015

768

769 **Wiens JJ. 2004.** Speciation and ecology revisited: phylogenetic niche conservatism and the

770 origin of species. *Evolution* **58:** 193−197

771

772   **Wiens JJ, Graham CH. 2005.** Niche conservatism: integrating evolution, ecology, and

773   conservation biology. *Annual Review of Ecology, Evolution, and Systematics* **36:** 519-539

774

775   **Wisz MS, Hijmans RJ, Li J, Peterson AT, Graham CH, Guisan A. 2008.** Effects of sample

776   size on the performance of species distribution models. *Diversity and Distributions* **14:** 763-773

777

Figure 1

Map of sampling locations. Abbreviations of the locations are given in Table 1, and the number in the parenthesis is the sample size.

Abbreviations of the locations are given in Table 1, and the number in the parenthesis is the sample size.

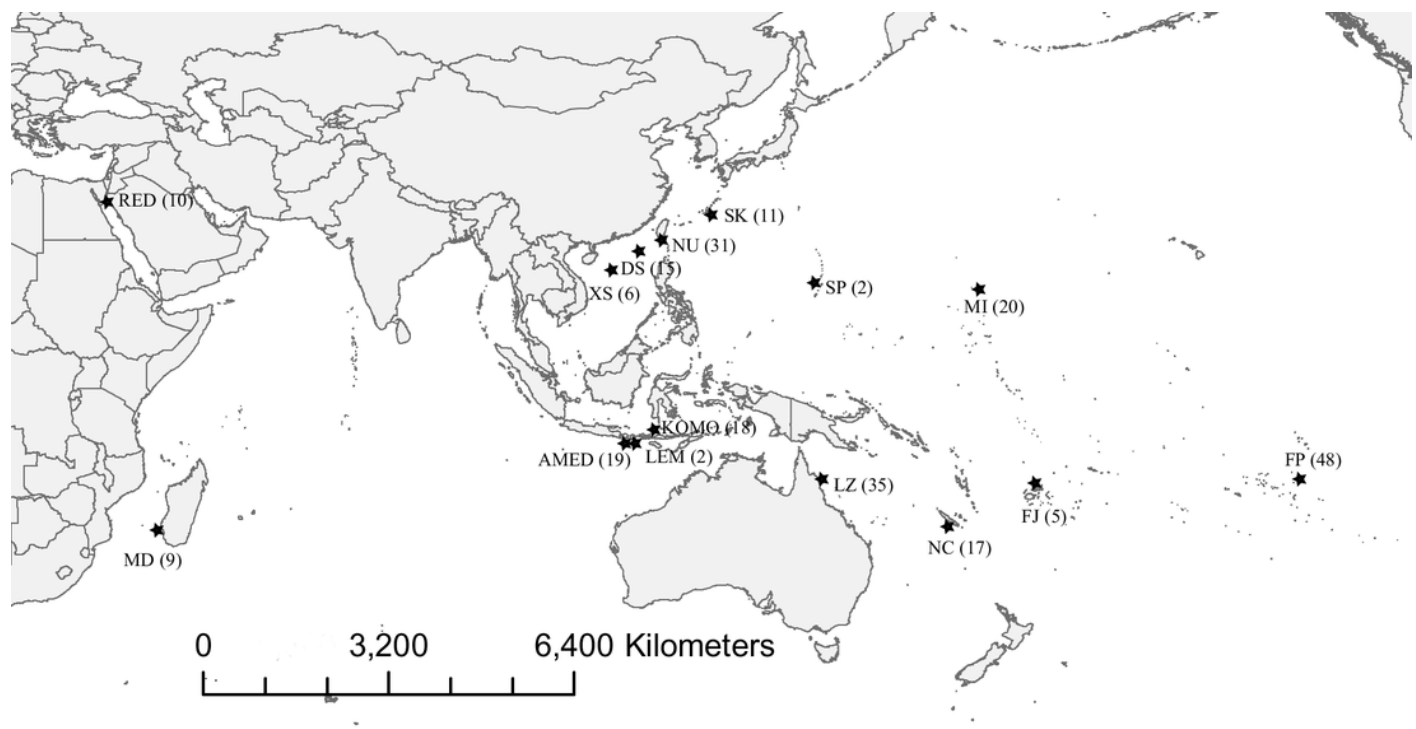



Figure 2

Bayesian phylogenetic tree and corresponding haplotype network based on two genetic markers including *Cytb* (A) and Rag2 (B).

Nodes are presented only for those with bootstrap scores >85% majority rule for maximum

likelihood and >90% majority probabilities for Bayesian probability values (BI/ML). For the

Haplotype network, different colors indicate different clades (e.g., white = *Chromis viridis*

Clade A, gray = *Chromis viridis* Clade B, and black = Red Sea)

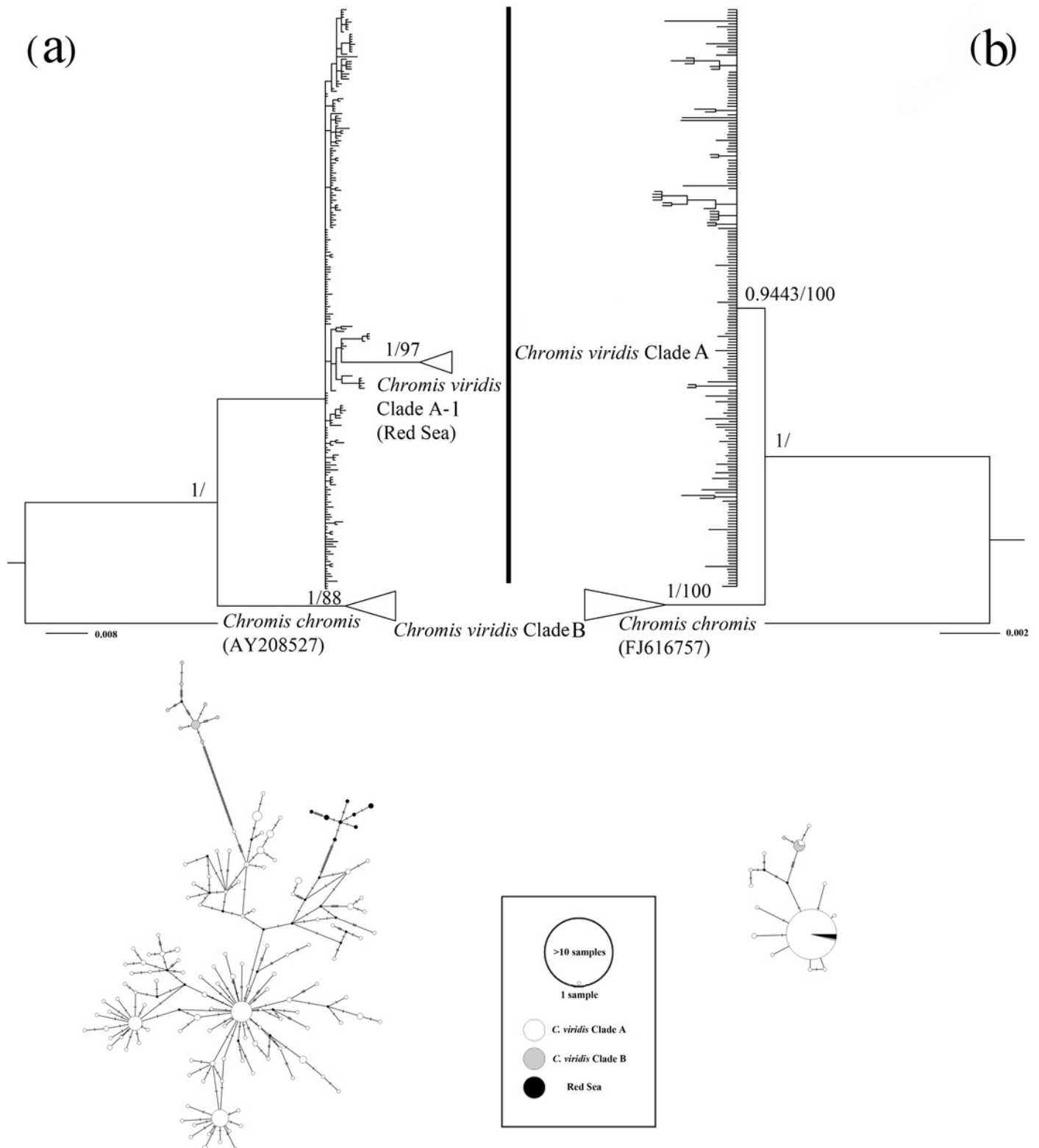



Figure 3

Time-tree of cryptic lineages among *Chromis viridis* obtained from BEAST.

With a 3.1 MYA time constraint on the node between *Chromis atrilobata* and *Chromis multilineata* (Quenouille et al., 2004). Horizontal grey bars at nodes indicate 95% posterior probability densities (HPD) intervals of age.

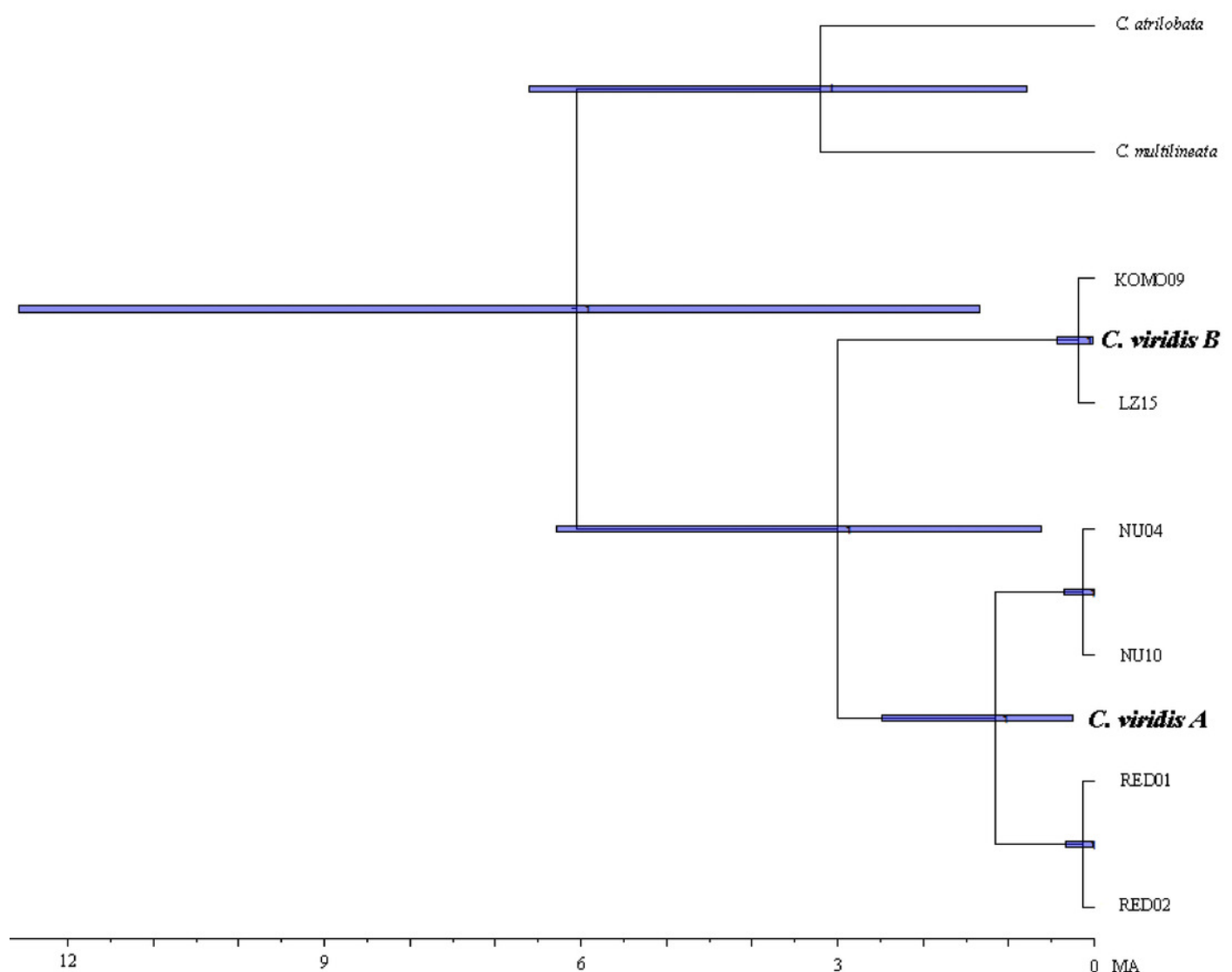
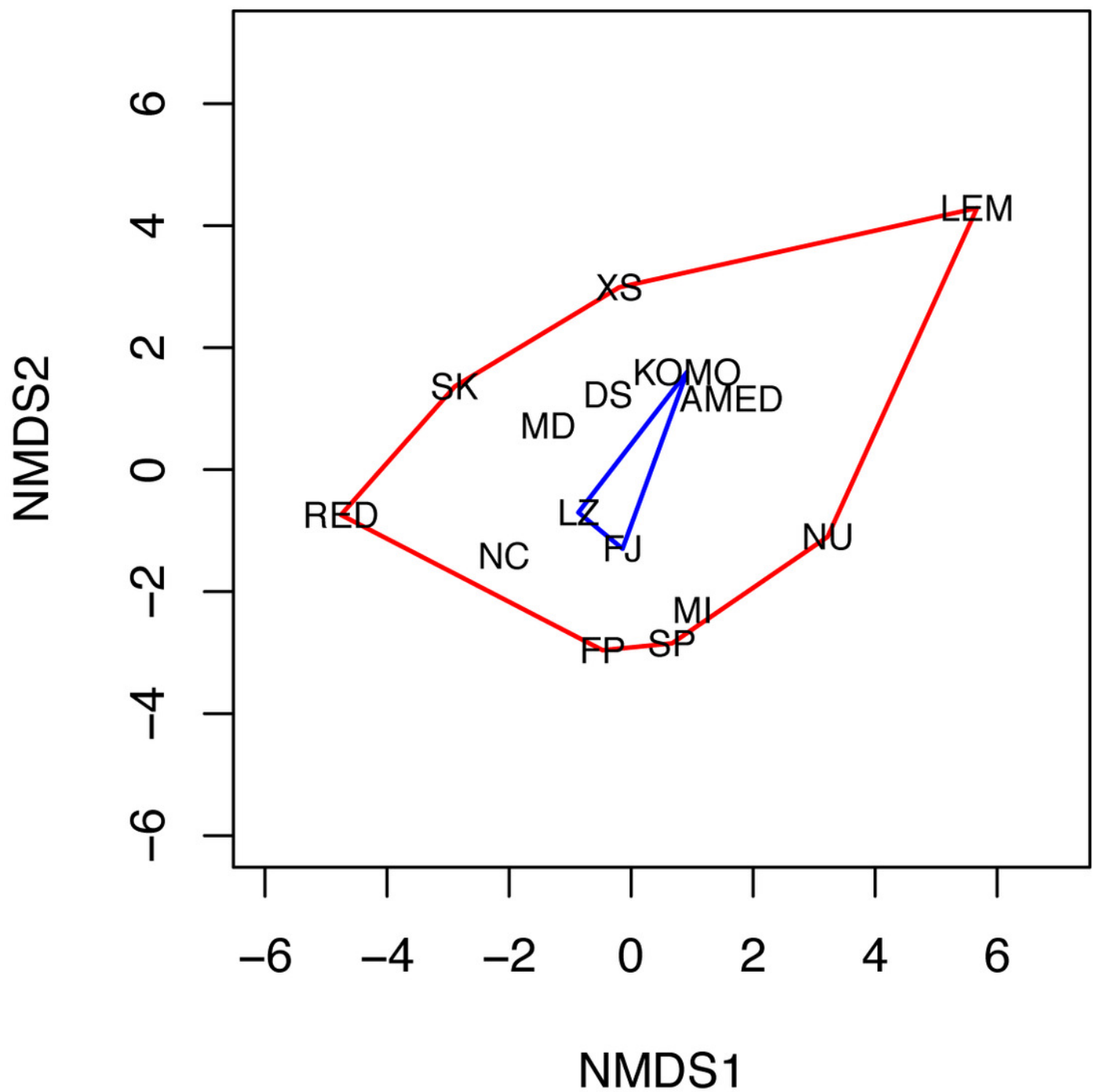



Figure 4

Non-metric multidimensional scaling (NMDS) plot of the 15 sampling sites.

The two NMDS axes (NMDS1 and NMDS2) represent environmental gradients defined by the 11 environmental factors examined across the sites. The convex hulls for the sampling sites where *C. viridis* A and *C. viridis* B were found are shown in red and blue, respectively. Please see the legend of Fig.1 for the abbreviation of the site names.

Sampling locations and diversity indices based on *Cytb* sequences (911bp) in 15 populations of *Chromis viridis* from the Indo-Pacific.

Abbreviations are as follows: N = sample size, nh = number of haplotype, h = haplotype diversity, and π = nucleotide diversity.

1   Table 1. Sampling locations and diversity indices based on *Cytb* sequences (911bp) in 15 populations of *Chromis viridis* from the

2   Indo-Pacific. Abbreviations are as follows: N = sample size, nh = number of haplotype, h = haplotype diversity, and π = nucleotide

3   diversity.

Abb.	Location	*Cytb*				Rag2			
		N	nh	h ± SD	π ± SD	n	nh	h ± SD	π ± SD
RED	Eliat, Israel	10	8	0.956 ± 0.059	0.004 ± 0.002	6	6	1.000 ± 0.096	0.003 ± 0.002
MD	Toliara, Madagascar	9	6	0.889 ± 0.091	0.007 ± 0.004	9	9	1.000 ± 0.052	0.005 ± 0.003
AMED	Amed, Bali, Indonesia	19	14	0.953 ± 0.036	0.007 ± 0.004	18	17	0.994 ± 0.021	0.004 ± 0.002
LEM	Nusa Lembongan, Bali, Indonesia	2	2	1.000 ± 0.500	0.003 ± 0.004	2	2	1.000 ± 0.500	0.006 ± 0.006
KOMO	Komodo, Indonesia	18	11	0.882 ± 0.064	0.014 ± 0.007	18	18	1.000 ± 0.019	0.005 ± 0.003
XS	Xisha, China	6	6	1.000 ± 0.096	0.006 ± 0.004	8	8	1.000 ± 0.063	0.004 ± 0.003
DS	Dongsha, Taiwan	15	11	0.933 ± 0.054	0.005 ± 0.003	15	15	1.000 ± 0.024	0.005 ± 0.003
NU	NPP Ⅲ Inlet, Taiwan	31	24	0.972 ± 0.020	0.005 ± 0.003	23	23	1.000 ± 0.013	0.004 ± 0.002
SK	SesokoIsland, Japan	11	9	0.946 ± 0.066	0.005 ± 0.003	11	11	1.000 ± 0.034	0.005 ± 0.003
SP	Saipan, USA	2	1	0.000 ± 0.000	0.000 ± 0.000	2	2	1.000 ± 0.500	0.000 ± 0.000
MI	Marshall Island, R.O. Marshall Islands	20	15	0.942 ± 0.043	0.005 ± 0.003	13	13	1.000 ± 0.030	0.004 ± 0.003
NC	New Caledonia, French	17	16	0.993 ± 0.023	0.006 ± 0.003	15	15	1.000 ± 0.024	0.004 ± 0.002
LZ	Lizard Island, Australia	35	26	0.968 ± 0.020	0.039 ± 0.019	35	35	0.998 ± 0.007	0.009 ± 0.005
FJ	Fiji	5	5	1.000 ± 0.127	0.054 ± 0.033	4	4	1.000 ± 0.177	0.015 ± 0.011
FP	Moorea Island, French Polynesia	48	21	0.7846 ± 0.0612	0.0028 ± 0.0016	39	35	0.9933 ± 0.0080	0.0042 ± 0.0025

Mann-Whitney U tests for the environmental differences between presence locations of clade A and those of clade B, and between presence and absence locations of clade A.

1 Table 2. Mann-Whitney U tests for the environmental differences between presence locations of
2 clade A and those of clade B, and between presence and absence locations of clade A.

Environmental Factor	A presence (Mean ± SD)	B presence (Mean ± SD)	A absence (Mean ± SD)	A presence vs. B presence		A presence vs. A absence	
				U	P-value	U	P-value
Temperature (°C)	27.44±0.86	27.15±1.35	27.08±1.47	20.5	0.86	20	0.84
Salinity (PSS)	34.31±0.86	34.55±1.43	34.61±1.57	22.5	1	18	1
Current velocity (m^{-1})	0.09±0.06	0.13±0.19	0.14±0.21	23.5	0.95	17	0.95
Nitrate (mol·m^{-3})	0.007±0.009	0.11±0.31	0.13±0.34	25.5	0.77	15	0.73
Phosphate (mol·m^{-3})	0.25±0.05	0.23±0.06	0.22±0.06	18.5	0.68	22	0.63
Silicate (mol·m^{-3})	3.57±1.99	3.74±1.85	3.79±1.90	23.5	0.95	17	0.95
Dissolved molecular oxygen (mol·m^{-3})	203.15±1.50	203.65±3.34	203.78±3.70	22.5	1	18	1
Iron (umol·m^{-3})	0.001±0.0003	0.0007±0.0005	0.0006±0.0005	10.5	0.17	30	0.1
Chlorophyll (mg·m^{-3})	0.14±0.04	0.15±0.12	0.15±0.13	18.5	0.68	22	0.63
Phytoplankton (umol·m^{-3})	1.19±0.16	1.14±0.48	1.13±0.54	15.5	0.44	25	0.36
Primary productivity (g·m^{-3}·day^{-1})	0.006±0.003	0.006±0.008	0.007±0.009	17.5	0.59	23	0.54

6