

Title	Ecosystem size predicts the probability of speciation in migratory freshwater fish
Author(s)	Yamasaki, Yo Y.; Takeshima, Hirohiko; Kano, Yuichi; Oseko, Naoharu; Suzuki, Toshiyuki; Nishida, Mutsumi; Watanabe, Katsutoshi
Citation	Molecular Ecology (2020), 29(16): 3071-3084
Issue Date	2020-08
URL	http://hdl.handle.net/2433/250818
Right	This is the peer reviewed version of the following article: Yamasaki YY, Takeshima H, Kano Y, et al. Ecosystem size predicts the probability of speciation in migratory freshwater fish. Mol Ecol. 2020, which has been published in final form at https://doi.org/10.1111/mec.15415 . This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.; The full-text file will be made open to the public on 19 April 2021 in accordance with publisher's 'Terms and Conditions for Self-Archiving'.; This is not the published version. Please cite only the published version.; この論文は出版社版ではありません。引用の際には出版社版をご確認ご利用ください。
Type	Journal Article
Textversion	author

1 **Title:** Ecosystem size predicts the probability of speciation in migratory freshwater fish

2

3 **Short title:** Ecosystem size predicts speciation

4

5 Authors: Yo Y. YAMASAKI^{1,2*}, Hirohiko TAKESHIMA^{3,4}, Yuichi KANO⁵, Naoharu
6 OSEKO⁶, Toshiyuki SUZUKI⁷, Mutsumi NISHIDA⁸, Katsutoshi WATANABE^{1*}

7

8 Author affiliations:

9 ¹ Graduate School of Science, Kyoto University, Kitashirakawa-oiwake-cho, Sakyo, Kyoto
10 606-8502, Japan

11 ² Ecological Genetics Laboratory, Department of Genomics and Evolutionary Biology,
12 National Institute of Genetics, Yata 1111, Mishima, Shizuoka, 411-8540, Japan

13 ³ Research Institute for Humanity and Nature, 457-4 Motoyama, Kamigamo, Kita, Kyoto
14 603-8047, Japan

15 ⁴ Department of Marine Biology, Tokai University, Shimizu 3-20-2, Shizuoka 424-0902,
16 Japan

17 ⁵ Graduate education and research training program in Decision Science for a Sustainable
18 Society, Kyushu University, 744, Motooka, Nishi, Fukuoka 819-0395, Japan

19 ⁶ 3-18-5-203, Nagata Ginowan, Okinawa 901-2212, Japan

20 ⁷ Kawanishi-midoridai Senior High School, 1-8 Kouyoudai, Kawanishi, Hyogo 666-0115,
21 Japan

22 ⁸ University of the Ryukyus, 1 Senbaru, Nishihara-cho, Nakagami-gun, Okinawa 903-0213,
23 Japan

24

25 * Corresponding authors:

26 Yo Y. Yamasaki: yo.yamasaki@terra.zool.kyoto-u.ac.jp

27 Katsutoshi Watanabe: watanak@terra.zool.kyoto-u.ac.jp

28 **ABSTRACT**

29 Predicting speciation is a fundamental goal of research in evolutionary ecology. The
30 probability of speciation is often positively correlated with ecosystem size. Although the
31 mechanisms driving this correlation are generally difficult to identify, a shared geographic and
32 ecological context provides a suitable condition to study the mechanisms that promote
33 speciation in large ecosystems by reducing the number of factors to be considered. Here, we
34 determined the correlation between speciation and ecosystem size, and discussed the
35 underlying mechanisms of this relationship, using a probable parallel ecotype formation for
36 freshwater fish. Our population genetic analysis revealed that speciation of the landlocked
37 goby, *Rhinogobius* sp. YB, of the Ryukyu Archipelago, Japan, from its migratory ancestor, *R.*
38 *brunneus*, occurred in parallel across five islands. Logistic regression analysis showed that
39 speciation probability could be predicted using island size. The results suggest that ecosystem
40 size predicts the occurrence of adaptation and reproductive isolation, likely through its
41 association with three possible factors: divergent selection strength, population persistence,
42 and occurrence probability of habitat separation.

43

44 **Key words**

45 ecological speciation, parallel evolution, island biogeography, approximate Bayesian
46 computation, speciation-area relationship, *Rhinogobius*

47

48 1 | INTRODUCTION

49 Understanding the patterns and mechanisms of speciation is the fundamental goal of
50 evolutionary ecology (Coyne & Orr, 2004; Reznick & Ricklefs, 2009). As macroevolutionary
51 patterns of speciation, such as differences in diversification rate among lineages or regions,
52 are caused by the accumulation of individual speciation events, the patterns are expected to be
53 strongly influenced by speciation mechanisms (i.e., evolution of reproductive isolation).
54 However, the relationships between macropatterns and the underlying mechanisms are not
55 well understood, because they are usually investigated separately (Rosenblum et al., 2012;
56 Cutter & Gray, 2016; Rabosky, 2016; but see Rabosky & Matute, 2013). Explicit
57 consideration of the speciation mechanism would be helpful for obtaining a causal
58 explanation of how macroevolutionary patterns of speciation have been generated (Harvey et
59 al., 2017; Kisel et al., 2012).

60 Ecosystem size, often represented by habitat area, has long been considered as one of the best
61 predictors for variation in species richness (MacArthur & Wilson, 1967). Recent studies on
62 species–area relationships also recognize “in situ speciation” as an important factor, in
63 addition to immigration and extinction rates (Losos & Parent, 2009; Losos & Schluter, 2000;
64 Wagner, Harmon, & Seehausen, 2014; Weigelt, Steinbauer, Cabral, & Kreft, 2016). The
65 positive correlation between ecosystem size and speciation rate is supported by several
66 empirical (Kisel & Barraclough, 2010; Parent & Crespi, 2006) and theoretical (Gavrilets &
67 Vose, 2005; Rosindell & Phillimore, 2011) studies.

68 Several hypotheses have been proposed that explain how speciation rate, or speciation
69 probability, is affected by ecosystem size through speciation processes. For example, the
70 opportunity for geographical isolation, which impedes gene flow, is likely to increase in
71 larger areas (Kisel & Barraclough, 2010; Losos & Schluter, 2000). As larger areas contain
72 heterogeneous environments, this phenomenon might also promote diversification to new
73 environments following speciation (Parent & Crespi, 2006). Moreover, larger areas allow
74 larger population sizes and longer persistence of newly emerged species (Kisel, McInnes,
75 Toomey, & Orme, 2011). However, these hypotheses have been rarely tested empirically,
76 because studies have usually assessed speciation by considering various mechanisms
77 simultaneously. Thus, it is difficult to evaluate the influence of different factors on the
78 speciation process. This problem could be addressed by targeting a group with parallel

79 diversification under the same mechanism (Lucek, Kristjánsson, Skúlason, & Seehausen,
80 2016; Vamosi, 2003). However, when using such an approach, it is important to first obtain
81 evidence of repeated evolution under the same mechanism as the focal system.

82 Due to clear boundaries and replications, freshwater invasion by migratory fish in island
83 habitats is a suitable case for testing the relationship between ecosystem size and occurrence
84 of speciation. The East Asian goby, genus *Rhinogobius*, is one such fish group, in which the
85 amphidromous (migratory) species *R. brunneus* might have repeatedly produced strictly
86 freshwater populations (*Rhinogobius* sp. YB sensu Akihito, Sakamoto, Ikeda, & Aizawa,
87 2013) in the Ryukyu Archipelago of Japan (Kano, Nishida, & Nakajima, 2012; Nishida, 2001;
88 Ohara, Takagi, Hashimoto, Miyazaki, & Hirashima, 2008). However, the number of times
89 and geographical scales (i.e., islands, rivers, etc.) of the origin of the freshwater form of this
90 goby have not yet been elucidated. Furthermore, the extent and factors of reproductive
91 isolation between the amphidromous and freshwater forms have not been investigated.

92 We propose that the freshwater colonization of the goby in the Ryukyu Archipelago is a good
93 model system for testing the correlation between ecosystem size and speciation probability for
94 the following reasons. First, the ancestral amphidromous populations might have shared
95 common evolutionary potential, because of possible gene flow through the sea during the
96 larval period. Second, ecosystem age is similar among islands. According to geographical
97 data, most islands in the Ryukyu Archipelago formed nearly simultaneously ~1.5 million
98 years ago by the opening of straits between the islands (Osozawa et al., 2012). Third,
99 ecological conditions are probably similar among the islands, because almost no freshwater
100 fish species, except those of *Rhinogobius*, occur in the upper reaches of the rivers in the
101 Ryukyu Archipelago (Kano et al., 2012; Yoshigou, 2014), resulting in few competitors and
102 predators, which are important components of ecological opportunity (Wellborn &
103 Langerhans, 2014; Yoder et al., 2010).

104 Using the simple system of *Rhinogobius* gobies, we primarily aimed to clarify whether the
105 speciation probability of the freshwater form can be explained by ecosystem size. We first
106 confirmed the existence of genetic isolation between the amphidromous and freshwater forms
107 inhabiting the same islands or rivers. Next, we estimated where and how many times the
108 freshwater forms originated, based on statistically explicit tests among demographic models
109 that hypothesized parallel or single origin scenarios. We then examined whether speciation

110 probability could be predicted by ecosystem size. The correlation between independent
111 origins of the freshwater form inferred from demographic modelling and ecosystem size (e.g.,
112 island area) was tested statistically (see Figure 1, concept diagram representing our approach).
113 On the basis of these results, we discuss the mechanisms producing the relationships between
114 ecosystem size and the speciation process in this system.

115 **2 | MATERIALS AND METHODS**

116 **2.1 | Study system**

117 The freshwater resident form derived from the amphidromous species *Rhinogobius brunneus*
118 is recognized as an independent species but has not yet been described because of taxonomic
119 confusion. Previous studies (e.g., Akihito et al., 2013) tentatively used the name
120 “*Rhinogobius* sp. YB” for this freshwater form. However, until its evolutionary status is
121 clarified, we will use the code “FF” (freshwater form) for *Rhinogobius* sp. YB and “AF”
122 (amphidromous form) for *R. brunneus* based on their life history modes.

123 The FF completes its life in freshwater; conversely, AF larvae hatch in rivers and immediately
124 drift down to the sea and return to rivers after 2–3 months (Mizuno, 2001). The eggs and
125 larvae of FF are larger than those of AF (Supporting Information methods and results S1).
126 These traits are assumed to be an adaptation to the river environment, where small food items
127 (i.e., plankton) are scarce and higher level of swimming ability in river currents is needed
128 (Hirashima & Tachihara, 2000; Nishida, 2001; Shinomiya, Sasabe, Sakurai, & Kishino, 2005;
129 Yamasaki, Nishida, Suzuki, Mukai, & Watanabe, 2015).

130 Although the distribution of the two species sometimes overlapped in a single river, their
131 distributions were generally different. AF mainly lived in the middle to lower reaches of
132 rivers. In fact, AF populations are found only in rivers shorter than 18 km in the central
133 Honshu region of Japan (Tamada, 2005). The larvae of AF must rapidly reach the sea to avoid
134 starvation because of the scarcity of small food in rivers (Iguchi & Mizuno, 1999; Moriyama,
135 Yanagisawa, Mizuno, & Omori, 1998). Thus, they might not be able to reproduce effectively
136 in areas far from the sea (Iguchi & Mizuno, 1999). By contrast, FF mainly live in the upper
137 reaches of rivers, and are often found in the upper areas of waterfalls where no other fish
138 species are distributed (Kano et al., 2012).

139 2.2 | Sampling

140 Samples were collected on seven islands where populations of FF have been recorded
141 previously (locality code 2. Amami-Ohshima Is., 4. Tokunoshima Is., 5. Okinoerabujima Is.,
142 6. Okinawajima Is., and 7. Kumejima Is. in Middle Ryukyu; 8. Ishigakijima Is. and 9.
143 Iriomotejima Is. in South Ryukyu; Table S1; Figure 2; Ohara et al., 2008; Yoshigo, 2014). In
144 addition, AF specimens were collected from three other islands (Figure 2). The specimens
145 were identified by body colour patterns when fresh (Kano et al., 2012; Suzuki, Shibukawa,
146 Yano, & Senou, 2004). Identification based on body colour was in agreement with the results
147 of the genetic analysis, but some samples were difficult to identify or showed inconsistent
148 genetic characteristics. When genetic data suggested that the focal specimens were not judged
149 as hybrids based on the criteria described below and that incorrect identification based on
150 morphology in the field was likely (see the next section and Supporting methods and results
151 S1), these samples were removed from all subsequent analyses. If a sample was judged to be a
152 hybrid, we tentatively named it based on morphology. As a result, 842 FF and 578 AF
153 specimens from 52 points in 42 river systems were included in the analysis. Both forms were
154 collected from nine river systems; they were sympatrically collected at seven points.

155 2.3 | Genetic data and population structure

156 Genetic data of *Rhinogobius* specimens were obtained by genotyping 20 nuclear
157 microsatellite loci (Table S2; Ohara, Takahashi, & Takagi, 2004). Methods for molecular
158 experiments and genotyping are described in the Supporting methods and results S1.
159 Genotype data were deposited in Dryad (<https://doi.org/10.5061/dryad.mpg4f4qvh>). To assess
160 the genetic features of the respective populations, we calculated the expected and observed
161 heterozygosity (H_E and H_O , respectively) and allelic richness (R_s) for the respective
162 populations. Pairwise F_{ST} and Jost's D were also calculated. We conducted analysis of
163 molecular variance (AMOVA; Excoffier, Smouse, & Quattro, 1992) separately for AF and FF,
164 to assess hierarchical genetic differentiation. To visualize relatedness among populations, we
165 constructed a neighbour-joining tree among populations based on Nei's D_a distance (Nei,
166 Tajima, & Tatenno, 1983) using poptree2 (Takezaki, Nei, & Tamura, 2010). The credibility of
167 each clade was evaluated using 1,000 bootstrap resamplings.

168 To infer the existence and extent of gene flow and genetic isolation, we analysed the genetic
169 structure of the population using structure version 2.3.4 (Lawson, van Dorp, & Falush, 2018;
170 Pritchard, Stephen, & Donnelly, 2000). We prepared two data sets: one for both FF and AF
171 populations from each island, and another for only AF populations from all islands. Evanno's
172 ΔK (Earl & von Holdt, 2012; Evanno, Regnaut, & Goudet, 2005) was used as a reference for
173 K value selection. Hybrid individuals were evaluated using q values; we arbitrarily judged
174 specimens as hybrids (or individuals with hybrid origin) in cases where minor genetic
175 components accounted for 10% or more of the aggregate (Vähä & Primmer, 2006). Detailed
176 methods are provided in the Supporting methods and results S1.

177 **2.4 | Testing the parallel formation of FF by ABC**

178 Before testing the relationship between ecosystem size and speciation probability, we needed
179 to infer the islands where speciation of FF occurred. To examine the single or parallel origins
180 of FF statistically, we conducted a demographic model selection under the Approximate
181 Bayesian Computation (ABC) framework (Beaumont, 2010). We briefly describe the
182 procedures of model construction and model selection in this section (for details, see
183 Supporting methods and results S1).

184 Our general model construction strategies for ABC analysis were as follows. First, we
185 generated separate models using three geographical scales (i.e. within-islands, regional
186 [Middle and South Ryukyu], and archipelago scales) and examined FF origins in each
187 geographical scale to reduce the number of models and populations examined in a single
188 analysis. Second, we examined both models with and without gene flow between AF and FF
189 populations. This is because a simulation study suggests that gene flow could alter apparent
190 phylogenetic relationships among populations when simple genetic distance was used (Bierne,
191 Gagnaire, & David, 2013). Demographic model comparisons that consider gene flow can
192 overcome this problem (Butlin et al., 2014).

193 In the within-islands analyses, we mainly examined two questions: whether FF had single or
194 parallel origin, and whether gene flow between AF and FF existed on each island. We
195 hypothesized several plausible divergence orders between populations based on the result of
196 structure analysis, phylogeny and geographical locations for respective islands. Thereafter, we
197 compared the models with single/parallel origins and with/without gene flow. Analyses were

198 conducted for the five islands where FF is distributed in multiple rivers (i.e., Amami-Oshima
199 Is., Tokunoshima Is., Okinawajima Is., Kumejima Is. and Iriomotejima Is.). We selected three
200 (or two) river populations for each island to reduce the model complexities. Criteria for
201 population selection are described in the Supporting methods and results S1. The constructed
202 models (a total of 34 models) are described in Figure S1a–e. Summary statistics are given in
203 Table S3.

204 For the among-islands within-region analyses, we constructed several plausible models
205 separately for the regions of Tokunoshima Is. and Okinoerabujima Is., Middle Ryukyu, and
206 South Ryukyu. Five to 10 models were constructed on the basis of the phylogenetic tree and
207 geographical distance between islands. Finally, in the between-region analyses, we compared
208 10 models generated from the results of the previous analyses. All constructed models are
209 described in the Supporting methods and results S1 and Figure S1f–I. Scripts that describe the
210 models have been deposited in Dryad (<https://doi.org/10.5061/dryad.mpg4f4qvh>).

211 For model selection, we chose the model that had the highest posterior probability in the
212 respective analyses. To evaluate the accuracy of model selection, we calculated a
213 misclassification rate (i.e., the rate at which the focal model was selected despite the other
214 model being true [type 2 error rate]).

215 **2.5 | Correlation between ecosystem size and speciation probability**

216 We first identified the islands on which the speciation of FF occurred based on the results of
217 structure (genetic isolation) and ABC model selection (independent origin). The relationship
218 between ecosystem size and speciation was analysed for the Middle and South Ryukyu
219 islands. This is because these islands are assumed to be similar in age, were never connected
220 to the continent after separation, and were not catastrophically affected by volcanic activity
221 (Osozawa et al., 2012; Supporting methods and results S1). We acquired distribution
222 information of the two forms from published literature (Yoshigou, 2014) and our own field
223 survey, and we only chose the islands where either or both forms were distributed. This is
224 because we assumed that speciation was not likely to have occurred on the island where either
225 of these species is not currently distributed. We coded the occurrence of speciation in each
226 island as follows: 1 for the island where speciation occurred, and 0 for the island where
227 speciation did not occur. As a proxy of ecosystem size, we used island area (km²; data from

228 the Geospatial Information Authority of Japan), maximum catchment area (km²), maximum
229 river length (m) or the maximum number of waterfalls on each island (Supporting methods
230 and results S1).

231 We performed one-parameter logistic regression analysis in the generalized linear model
232 framework with binomial error structure for the response variable (speciation: 1, presence; 0,
233 absence) using the glm package of r version 3.1.1 (R Core Team, 2014). Due to a strong
234 positive correlation between the four variables (Pearson's correlation coefficient, 0.592–0.993,
235 Supporting methods and results S1), we primarily used island area as the explanatory variable
236 for the main result. Explanatory variables, except the number of waterfalls, were
237 log₁₀-transformed. To test the significance of explanatory variables, we conducted a
238 likelihood ratio test using 10,000 rounds of parametric bootstrap sampling with a constant
239 model as a null model that assumes no relationship between ecosystem variables and
240 speciation probability. We calculated McFadden's pseudo R^2 .

241 Furthermore, to examine the effect of the ecosystem size parameters other than island size
242 (i.e., maximum catchment area, river length and number of waterfalls), we constructed
243 another generalized linear model for speciation probability. We used the residuals from a
244 regression between island size and each of the other ecosystem size parameters, together with
245 the island size and their interaction, as explanatory variables. Stepwise model selections by
246 Akaike information criterion (AIC) scores were conducted using the *step* function on r.

247 **3 | RESULTS**

248 **3.1 | Basic summary statistics**

249 Heterozygosity and allelic richness were generally higher in the AF populations than in the
250 FF populations (Table S1, Figure S2a, b). The AF populations showed relatively little
251 pairwise genetic differentiation within islands and among islands within regions (Tables S4–
252 S6, Figure S2c, d). In comparison, the genetic differentiation of AF populations between
253 regions was large (Tables S4–S6, Figure S2c, d). The FF populations often showed large
254 genetic differentiation within islands and among islands within regions, as well as among
255 regions (Tables S4–S6, Figure S2c, d).

256 **3.2 | Genetic isolation and hybrids**

257 In the AF data set, Evanno's ΔK showed the maximum value at $K = 2$; the populations in
258 North plus Middle Ryukyu and South Ryukyu were clearly separated. In the data set for each
259 island, including FF and AF populations, ΔK had a maximum value at $K = 2$ for all islands,
260 except one (Amami-Oshima Is.). The two genetic clusters corresponded to the two forms,
261 except on Okinawajima Is. (see below; Figure 3; Figure S3).

262 Suspected hybrid individuals were rare (2.0% of the total individuals). No suspected hybrid
263 individuals were detected from five of the seven sympatric collection points. Several
264 suspected hybrid individuals were detected in 11 FF and AF populations. Among these, all FF
265 individuals that were identified based on morphology in the FFOK-4-1 and FFOK-4-2
266 populations (Okinawajima Is.; Figures 2 and 3; Figure S3) exhibited the genetic features of
267 AF as the dominant component.

268 **3.3 | Genetic relationships among populations**

269 The neighbour-joining tree showed two divergent groups separated by the Kerama Gap in all
270 of the FF and AF populations (bootstrap value, 97%; Figure 4). Each group included both
271 forms, which we present as Group 1 (Honshu, North Ryukyu and Middle Ryukyu) and Group
272 2 (South Ryukyu), hereafter. In both groups, FF populations

273 on each island formed a monophyletic group (bootstrap value, 91%–100%), except those on
274 two islands in Group 1 (Amami-Oshima Is., 67%; Okinawajima Is., 33%). In addition, FF
275 populations on different islands (Tokunoshima Is. and Okinoerabujima Is.) formed a
276 monophyletic group (89%). The AF populations within each group were closely related to
277 each other, and the populations of each island in Group 1 tended to form a monophyletic
278 group.

279 **3.4 | Test for multiple origins of FF**

280 In the within-island-scale analysis, the single origin model of FF was selected for all islands
281 (Table 1 and Table S7; Figure 4 and Figure S4). Models including gene flow among all or
282 partial populations of the two forms showed higher posterior probability than those not
283 considering gene flow (Table 1 and Table S7; Figure 4 and Figure S4).

284 At the among-island scale, Group 1 in Middle Ryukyu and Group 2 in South Ryukyu were
285 analysed separately. For Group 1, we first examined the single origin of FF in the adjoining
286 islands, Tokunoshima Is. and Okinoerabujima Is., which was suggested by the phylogenetic
287 analysis. ABC analysis supported their single origin (Table 1 and Table S7), and thus only FF
288 on Tokunoshima Is. was used for subsequent analyses. The model selection for Group 1
289 strongly supported the four independent origins of FF (Table 1; Table S7). The model with
290 gene flow between the two forms provided a higher Bayes factor (BF) than the model not
291 considering gene flow (Table S7). In Group 2, the single origin of FF of Iriomotejima Is. and
292 Ishigakijima Is. was supported (Table 1; Table S7). The model including gene flow between
293 the AF and FF populations was better supported than that not considering gene flow (Table
294 S7).

295 For the between-region scale, the neighbour-joining tree for populations suggested
296 independent origins of FF in Groups 1 and 2. However, no conclusion could be drawn at this
297 scale from ABC, because no single model was successfully selected, probably because of
298 insufficient data (Table 1; Table S7).

299 **3.5 | Correlation between speciation probability and ecosystem size**

300 We confirmed the occurrence of AF or FF populations on 18 islands from the published
301 literature and our field survey (Table S8). As we could not specify the islands where unique
302 speciation occurred, we tested 20 alternative scenarios that hypothesized that speciation
303 occurred in different combinations of the candidate islands (see Supporting methods and
304 results S1, and Table S9 for details).

305 For all four data sets in the scenario of independent origins between the regions, speciation
306 probability was significantly explained by island area (Table 2 and Table S9; Figure 5 and S5).
307 The relationship was also significant for 12 of the 16 data sets in the nonindependent scenario
308 (Table S9). Almost identical results were obtained for maximum catchment area and river
309 length (Table 2 and Table S9; Figure S5). Furthermore, two of the four data sets for the
310 independent origin scenario and five of the 16 data sets for the nonindependent scenario
311 showed significant relationships between the number of waterfalls and speciation (Table 2
312 and Table S9). Model selection by AIC scores showed that the best model included no

313 variables, except for island area, as the predictors for speciation probability in most data sets
314 when the variables were controlled by island area (Table S9).

315 **4 | DISCUSSION**

316 Our results revealed that FF repeatedly evolved from AF in the Ryukyu Archipelago, and that
317 the occurrence of FF was highly correlated with island size. These results suggest that the
318 speciation of FF was promoted in larger ecosystems. In the following sections, we discuss the
319 mechanisms of speciation and the causes of correlation between ecosystem size and
320 speciation of FF.

321 **4.1 | Reproductive isolation between the two forms and its evolution**

322 Genetic isolation between the two forms was suggested, even at sympatric sites (Figure 3 and
323 Figure S3); the results support that FF and AF are reproductively isolated. This genetic
324 isolation between AF and FF cannot be explained by isolation by distance (IBD), because
325 geographically isolated FF populations in different river systems showed closer relationships
326 than FF and AF in the same river. This relationship was consistent in the respective islands.
327 Some behavioural mechanisms might be responsible for this reproductive isolation, as
328 reported for other species pairs of *Rhinogobius* (Mizuno, 1987), which share common
329 reproductive features with intersexual behavioural interactions (Takahashi & Kohda, 2004).
330 Differences in the breeding season might also contribute to their current reproductive isolation
331 (AF, January to May; FF, March to June; Hirashima, 2016; Katoh, 1996; Tachihara, 2009).
332 Although the presence of pre- and post-zygotic isolations has not been tested for our studied
333 pair of species and their position along a speciation continuum (Nosil, 2012) is not clear,
334 genetic isolation, as well as distinct life history divergence, support that FF populations
335 should be treated as different biological species from AF populations.

336 Although AF and FF currently exhibit genetic isolation, our ABC analysis suggested that
337 reproductive isolation between these two ecologically divergent forms evolved under the
338 existence of gene flow. This result suggests that the evolution of reproductive isolation was
339 not caused simply by allopatry and mutation-order mechanisms but by divergent ecological
340 adaptation between the AF and FF populations (Seehausen & Wagner, 2014).

341 What is the factor that led to the evolution of their reproductive isolation? Previous studies
342 have focused on the differences in food size for larvae between seas and rivers (Nishida,
343 2001). AF larvae generally feed on small food items available in seas. If larvae cannot reach
344 the sea rapidly, large larvae might be favoured by directional selection in rivers (Closs, Hicks,
345 & Jellyman, 2013; Kishi, 1979; Kondo, Maeda, Hirashima, & Tachihara, 2013; Mizuno,
346 1963). A mathematical model has shown that speciation could occur due to differences in
347 favourable egg and larval size between the upper and lower reaches (Omori et al., 2012). Such
348 empirical and theoretical evidence suggests that the primary factor in the evolution of
349 reproductive isolation is divergent selection between the optimal combinations of the
350 within-rivers distribution and larval size.

351 The role of barriers in a river should also be emphasized, because they might promote
352 adaptation to freshwater habitats through reduced gene flow and the invasion of predators and
353 competitors. During repeated sea-level fluctuations during the Pleistocene, steep marine
354 terraces of limestone or coral reef developed in the coastal areas of the Ryukyu Archipelago
355 (Machida, Ota, Kawana, Moriwaki, & Nagaoka, 2001). Such terraces would form many
356 waterfalls near river mouths. Nishida (2001) and Kano et al. (2012) hypothesized an
357 important role of waterfall formation in the landlocking process of FF. The relationship
358 between such barriers and divergent selection should be further investigated to understand the
359 early processes of landlocking.

360 Despite the existence of current reproductive isolation, suspected hybrid individuals were
361 occasionally detected. Some hybridization might be explained by artificial disturbances in the
362 distribution and migration schemes of AF. Six of the 11 populations containing suspected
363 hybrids were collected from the waters above artificial dams; they included FFOK-4-1 and
364 FFOK-4-2 populations of FF, which primarily exhibited genetic features of AF. Artificial
365 dams would promote hybridization in the following process.

366 Artificial dams have promoted the establishment of a landlocked AF population (Tachihara,
367 2009), because there are small food items in reservoirs that the larvae of AF and hybrids
368 would be able to use. This would induce a sympatric distribution of the two forms and
369 facilitate the survival of hybrids in the upper reaches of the dam. These cases might be
370 examples of human-mediated reverse speciation (Grabenstein & Taylor, 2017; Taylor et al.,
371 2006).

372 4.2 | Multiple origins of FF

373 The AF and FF populations showed large divergence between the North/Middle Ryukyu
374 (Group 1) and South Ryukyu (Group 2). The AF populations were closely related among
375 islands within each region (Figures 3 and 4), suggesting that gene flow by larval drifting is
376 effective at the regional scale, but is prevented between regions by the Kerama Gap.

377 A series of ABC analyses supported the parallel derivation of FF populations on respective
378 islands. The estimated number of times speciation occurred was four or five. The uncertainty
379 in the number is attributed to the unresolved relationship among the FF populations across the
380 Middle and South Ryukyus, which could not be resolved by ABC analysis, probably owing to
381 the limitation of microsatellite markers, such as allele size homoplasy. However, the
382 following reasons support that the FF populations originated independently in each region;
383 that is, FF originated five times. First, the independent origins of FF populations among
384 islands within regions were supported (Table 1; Figure 4). Second, the simple tree analysis
385 supported the dichotomous relationship of the populations between the North/Middle and
386 South Ryukyus. Third, the two regions are separated by a deep strait, the Kerama Gap, which
387 is considered to have not emerged above water since the Ryukyu Archipelago separated from
388 the continent (Osozawa et al., 2012). This geographical barrier has also been noted for many
389 other terrestrial animals, such as amphibians and reptiles (Matsui, Ito, et al., 2005; Matsui,
390 Shimada, et al., 2005; Ota, 1998; Tominaga, Matsui, Eto, & Ota, 2015).

391 Our results suggest that the multiple origins of FF were an example of parallel speciation,
392 which is the phenomenon whereby reproductive isolation repeatedly evolves under the same
393 mechanism (Schluter & Nagel, 1995). Nosil (2012) proposed four criteria to distinguish
394 parallel speciation: nonmonophyletic pattern of an ecotype; nonmonophyly of ecologically
395 similar forms representing multiple origins, rather than the effect of gene flow or incomplete
396 lineage sorting; positive correlation between reproductive isolation and ecological divergence;
397 and the genetic basis of reproductive isolation. The ABC analysis, as well as the egg and
398 larval size difference between FF and AF, indicated that our study system satisfies the former
399 two criteria. In addition, the latter two criteria were probably satisfied because of the clear
400 differences in life history and reproductive traits between the two forms (Yamasaki et al.,
401 2015), although they need to be confirmed by conducting mate choice and common garden
402 experiments in the future.

403 **4.3 | Correlation of ecosystem size and speciation probability**

404 Our logistic regression analysis showed that speciation probability was significantly
405 explained by island size, river length and catchment area in most of the data sets under the
406 various analysis conditions. This result suggests that some mechanisms promoting the
407 speciation process tend to be more effective in larger ecosystems. As the number of waterfalls
408 also explained speciation probability in some data sets, it also partly supported that waterfalls
409 promote speciation in the goby.

410 Why is the speciation probability of FF positively correlated with ecosystem size? To answer
411 this question, understanding the relationships between ecosystem size and the following three
412 components related to the mechanism of speciation is necessary: (a) establishment of
413 reproductive isolation, (b) persistence of diverged populations and (c) enhancement of the
414 former two components by population separation (Allmon, 1992; Mayer, 1963).

415 First, the establishment of reproductive isolation would be promoted in larger ecosystems,
416 because the strength of divergent selection is expected to be positively correlated with
417 ecosystem size. Divergent selection between flowing and staying larvae would become
418 stronger in longer rivers (i.e., larger ecosystems), as the spawning sites of AF are restricted
419 within a certain distance from the sea owing to the mortality of drifting larvae (see above;
420 Moriyama et al., 1998; Omori et al., 2012). Furthermore, divergent selection might have been
421 stronger during glacial periods, when sea level was lower, and hence rivers became longer
422 (Kondo et al., 2013). Therefore, the landscape in the period of low sea level might have
423 affected patterns of speciation. However, unfortunately, we could not test the correlation
424 between speciation probability and island size at the low sea level, represented by the Last
425 Glacial Maximum (i.e., when sea level was -122 m), because the 18 islands investigated were
426 estimated to have been merged into eight, and the statistical power was too low to detect the
427 correlation (data not shown). Although the correlation between the strength of divergent
428 selection and river length should be tested in the future, larger ecosystems might enhance
429 reproductive isolation via divergent selection, when the extent of the focal environmental
430 gradient is positively related to ecosystem size. This relationship might generally lead to the
431 development of a correlation between speciation probability and ecosystem size in other
432 systems as well.

433 Second, the amount of available space might increase with ecosystem size, which could result
434 in the greater persistence of freshwater-adapted populations through an increase in population
435 size. In longer rivers, the freshwater population could use larger areas in the upper reaches of
436 the rivers, because the optimal distance between the spawning area and sea for AF does not
437 increase with river length. A larger population size might reduce the risk of demographic
438 extinction. In addition, extinction risk due to fusion with the amphidromous population might
439 be alleviated in the larger freshwater populations. Furthermore, the reinforcement of
440 reproductive isolation might be more effectively developed when the population sizes of the
441 two secondarily contacted species are similar (Liou & Price, 1994). A larger population size is
442 also beneficial for retaining adaptive alleles (Gavrilets & Vose, 2005) and reducing
443 maladaptive gene flow (Hanski, Mononen, & Ovaskainen, 2011).

444 Third, the probability of habitat separation between rivers and seas might be positively
445 correlated with ecosystem size, promoting the previously described two processes during
446 speciation. Correlation between the number of waterfalls and river length suggests the more
447 frequent emergence of barriers to migration in longer rivers. Furthermore, greater river length
448 might hinder migration to the upper reaches. Such habitat separation in longer rivers might be
449 important for population divergence, because the separation would reduce gene flow between
450 FF and AF populations, and would impede the entrance of other predatory and competitive
451 fishes in the upper reaches.

452 **5 | CONCLUSIONS**

453 We have confirmed that speciation probability is explained by ecosystem size in our goby
454 system. This result suggests positive correlations between ecosystem size and several
455 components of the speciation process, including the strength of divergent selection,
456 population persistence and population separation frequency. These correlations could promote
457 the establishment of reproductive isolation and the persistence of new populations, although
458 further explicit tests for these relationships are necessary. These correlations could also be
459 important for understanding species diversification mechanisms at the phylogenetic level.
460 Simple systems, such as our parallel species pairs, could contribute to the development of
461 theories that link the mechanisms of each speciation to macroevolutionary patterns of
462 speciation.

463

464 **ACKNOWLEDGMENTS**

465 We thank K. Nakao for collecting samples; S. Matsuzaki, N. Muto, S. Chiba, K. Nakayama
466 and S. Hirase for helping with the molecular experiment; and A. Iwata, M. Hosono, T. Fujisawa,
467 J. Kitano and members of Laboratory of Animal Ecology of Kyoto University for providing
468 valuable comments. We thank C. Riginos and three anonymous reviewers for providing
469 helpful comments. Computations were partially performed on the NIG supercomputer at
470 ROIS National Institute of Genetics. This study was partially supported by the Sasakawa
471 Scientific Grant (No. 27-507) and JSPS KAKENHI (Nos. 26250044, 16J08762, 17H03720).

472 **AUTHOR CONTRIBUTIONS**

473 Y.Y.Y., M.N. and K.W. planned the project. Y.Y.Y., Y.K., N.O. and T.S. performed field
474 surveys. Y.Y.Y. and H.T. conducted molecular experiments. Y.Y.Y. conducted population
475 genetic and statistical analysis. Y.Y.Y. and K.W. wrote the first draft manuscript, and all
476 authors contributed to the improvement of the manuscript.

477 **DATA AVAILABILITY STATEMENT**

478 The microsatellite genotype data and scripts of ABC analysis have been deposited in Dryad
479 (<https://doi.org/10.5061/dryad.mpg4f4qvh>). NGS data for developing microsatellite makers
480 have been deposited in DDBJ (accession no. DRA006388).

481

482 **ORCID**

483 *Yo Y. Yamasaki* <https://orcid.org/0000-0002-7495-2712>

484 *Katsutoshi Watanabe* <https://orcid.org/0000-0003-2244-2902>

485

486 **REFERENCES**

487 Akihito, Sakamoto, K., Ikeda, Y., & Aizawa, M. (2013). Gobioidae. In T. Nakabo (Ed.),
488 *Fishes of Japan with pictorial keys to species*, 3rd ed. (pp. 1374–1608). Tokyo, Japan: Tokai
489 University Press. (in Japanese).

490 Allmon, W. D. (1992). A causal analysis of stages in allopatric speciation. *Oxford Surveys in*
491 *Evolutionary Biology*, 8, 219–257.

492 Beaumont, M. A. (2010). Approximate Bayesian computation in evolution and ecology.
493 *Annual Review of Ecology, Evolution, and Systematics*, 41, 379–406.
494 <https://doi.org/10.1146/annurev-ecolsys-102209-144621>

495 Bierne, N., Gagnaire, P. A., & David, P. (2013). The geography of introgression in a patchy
496 environment and the thorn in the side of ecological speciation. *Current Zoology*, 59, 72–86.
497 <https://doi.org/10.1093/czoolo/59.1.72>

498 Butlin, R. K., Saura, M., Charrier, G., Jackson, B., André, C., Caballero, A., ... Rolán-Alvarez,
499 E. (2014). Adaptation and reproductive isolation in the face of gene flow. *Evolution*, 68, 935–
500 949. <https://doi.org/10.1111/evo.12329>

501 Closs, G. P., Hicks, A. S., & Jellyman, P. G. (2013). Life histories of closely related
502 amphidromous and non-migratory fish species: A trade-off between egg size and fecundity.
503 *Freshwater Biology*, 58, 1162–1177. <https://doi.org/10.1111/fwb.12116>

504 Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sunderland, MA: Sinauer Associates.

505 Cutter, A. D., & Gray, J. C. (2016). Ephemeral ecological speciation and the latitudinal
506 biodiversity gradient. *Evolution*, 70, 2171–2185. <https://doi.org/10.1111/evo.13030>

507 Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program
508 for visualizing STRUCTURE output and implementing the Evanno method. *Conservation*
509 *Genetics Resources*, 4, 359–361. <https://doi.org/10.1007/s12686-011-9548-7>

510 Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals
511 using the software STRUCTURE: A simulation study. *Molecular Ecology*, 14, 2611–2620.
512 <https://doi.org/10.1111/j.1365-294X.2005.02553.x>

513 Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred
514 from metric distances among DNA haplotypes: application to human mitochondrial DNA
515 restriction data. *Genetics*, *131*, 479–491.

516 Gavrilets, S., & Vose, A. (2005). Dynamic patterns of adaptive radiation.
517 *Proceedings of the National Academy of Sciences of the United States of America*, *102*,
518 18040–18045. <https://doi.org/10.1073/pnas.0506330102>

519 Grabenstein, K. C., & Taylor, S. A. (2017). Breaking barriers: Causes, consequences, and
520 experimental utility of human-mediated hybridization. *Trends in Ecology & Evolution*, *33*,
521 198–212. <https://doi.org/10.1016/j.tree.2017.12.008>

522 Hanski, I., Mononen, T., & Ovaskainen, O. (2011). Eco-evolutionary metapopulation
523 dynamics and the spatial scale of adaptation. *American Naturalist*, *177*, 29–43.
524 <https://doi.org/10.1086/657625>

525 Harvey, M. G., Seeholzer, G. F., Smith, B. T., Rabosky, D. L., Cuervo, A. M., & Brumfield,
526 R. T. (2017). Positive association between population genetic differentiation and speciation
527 rates in New World birds. *Proceedings of the National Academy of Sciences of the United*
528 *States of America*, *114*, 6328–6333. <https://doi.org/10.1073/pnas.0506330102>

529 Hirashima, K. (2016). Variety of life history and strategy of *Rhinogobius* fishes. *Aquabiology*,
530 *38*, 363–369. (in Japanese with English abstract). Hirashima, K., & Tachihara, K. (2000).
531 Embryonic development and morphological changes in larvae and juveniles of two
532 land-locked gobies, *Rhinogobius* spp. (Gobiidae), on Okinawa Island. *Japanese Journal of*
533 *Ichthyology*, *47*, 29–41. (in Japanese with English abstract). [https://](https://doi.org/10.11369/jji1950.47.29)
534 doi.org/10.11369/jji1950.47.29

535 Iguchi, K., & Mizuno, N. (1999). Early starvation limits survival in amphidromous fishes.
536 *Journal of Fish Biology*, *54*, 705–712. <https://doi.org/10.1111/j.1095-8649.1999.tb02027.x>

537 Kano, Y., Nishida, S., & Nakajima, J. (2012). Waterfalls drive parallel evolution in a
538 freshwater goby. *Ecology and Evolution*, *2*, 1805–1817. <https://doi.org/10.1002/ece3.295>

- 539 Katoh, M. (1996). Seasonal variation in gonadal activity of females among four species of
540 freshwater gobies in the *Rhinogobius brunneus* species complex in Okinawa, Japan.
541 *Ichthyological Research*, *43*, 169–174.
- 542 Kisel, Y., & Barraclough, T. G. (2010). Speciation has a spatial scale that depends on levels
543 of gene flow. *American Naturalist*, *175*, 316–334. <https://doi.org/10.1086/650369>
- 544 Kisel, Y., McInnes, L., Toomey, N. H., & Orme, C. D. L. (2011). How diversification rates
545 and diversity limits combine to create large-scale species–area relationships. *Philosophical*
546 *Transactions of the Royal Society B*, *366*, 2514–2525. <https://doi.org/10.1098/rstb.2011.0022>
- 547 Kisel, Y., Moreno-Letelier, A. C., Bogarín, D., Powell, M. P., Chase, M. W., & Barraclough,
548 T. G. (2012). Testing the link between population genetic differentiation and clade
549 diversification in Costa Rican orchids. *Evolution*, *66*, 3035–3052. <https://doi.org/10.1111/j.1558-5646.2012.01663.x>
- 550
- 551 Kishi, Y. (1979). A geographical model of disruptive selection on offspring size and a
552 possible case of speciation in freshwater gobies characterized by egg-size difference.
553 *Researches on Population Ecology*, *20*, 211–215.
- 554 Kondo, M., Maeda, K., Hirashima, K., & Tachihara, K. (2013). Comparative larval
555 development of three amphidromous *Rhinogobius* species, making reference to their habitat
556 preferences and migration biology. *Marine and Freshwater Research*, *64*, 249–266.
557 <https://doi.org/10.1071/MF12234>
- 558 Lawson, D. J., van Dorp, L., & Falush, D. (2018). A tutorial on how not to over-interpret
559 STRUCTURE and ADMIXTURE bar plots. *Nature Communications*, *9*, 3258.
560 <https://doi.org/10.1038/s41467-018-05257-7>
- 561 Liou, L., & Price, T. D. (1994). Speciation by reinforcement of premating isolation. *Evolution*,
562 *48*, 1451–1459.
- 563 Losos, J. B., & Parent, C. E. (2009). The speciation–area relationship. In J. B. Losos, & R. E.
564 Ricklefs (Eds.), *The theory of island biogeography revisited* (pp. 415–438). Princeton, NJ,
565 USA: Princeton University Press.

- 566 Losos, J. B., & Schluter, D. (2000). Analysis of an evolutionary species-area relationship.
567 *Nature*, 408, 847–850. <https://doi.org/10.1038/35048558>
- 568 Lucek, K., Kristjánsson, B. K., Skúlason, S., & Seehausen, O. (2016). Ecosystem size
569 matters: The dimensionality of intralacustrine diversification in Icelandic stickleback is
570 predicted by lake size. *Ecology and Evolution*, 6, 5256–5272. [https://doi.org/10.1002/](https://doi.org/10.1002/ece3.2239)
571 [ece3.2239](https://doi.org/10.1002/ece3.2239)
- 572 MacArthur, R. H., & Wilson, E. O. (1967). *The theory of island biogeography*. Princeton, NJ:
573 Princeton University Press.
- 574 Machida, H., Ota, Y., Kawana, T., Moriwaki, H., & Nagaoka, N. (2001). *Regional*
575 *geomorphology of the Japanese Islands, vol. 7 Kyushu and Southern Islands*. Tokyo, Japan:
576 University of Tokyo Press. (in Japanese).
- 577 Matsui, M., Ito, H., Shimada, T., Ota, H., Saidapur, S. K., Khonsue, W., ... Wu, G.-F. (2005).
578 Taxonomic relationships within the pan-oriental narrow-mouth toad *Microhyla ornata* as
579 revealed by mtDNA analysis (Amphibia, Anura, Microhylidae). *Zoological Science*, 22, 489–
580 495. <https://doi.org/10.2108/zsj.22.489>
- 581 Matsui, M., Shimada, T., Ota, H., & Tanaka-Ueno, T. (2005). Multiple invasions of the
582 Ryukyu Archipelago by Oriental frogs of the subgenus *Odorrana* with phylogenetic
583 reassessment of the related subgenera of the genus *Rana*. *Molecular Phylogenetics and*
584 *Evolution*, 37, 733–742. <https://doi.org/10.1016/j.ympev.2005.04.030>
- 585 Mayer, E. (1963). *Animal species and evolution*. Cambridge, MA, USA: Belknap Press.
- 586 Mizuno, N. (1963). Distribution of *Cottus japonicus* Okada (Cottidae) and *Tukugobius*
587 *flumineus* Mizuno (Gobiidae), with special references to their peculiarities in both the
588 land-locking and the speciation from amphidromous to ancestors. *Bulletin of Osaka Gakugei*
589 *University*, 11, 129–161. (in Japanese with English abstract).
- 590 Mizuno, N. (1987). Yoshinobori rui (genus *Rhinogobius*). In N. Mizuno, & A. Goto (Eds.),
591 *Nihon no tansuigyo – sono bunpu, hen'i, shubunka wo megutte (Freshwater fishes in Japan –*
592 *Distribution, variation, and speciation)* (pp. 179–188). Tokyo, Japan: Tokai University Press.
593 (in Japanese).

- 594 Mizuno, N. (2001). *Rhinogobius*. In H. Kawanabe, N. Mizuno, & K. Hosoya (Eds.),
595 *Freshwater fishes of Japan*, 3rd ed. (p. 584) Tokyo, Japan: Yama-Kei Publishers. (in
596 Japanese).
- 597 Moriyama, A., Yanagisawa, Y., Mizuno, N., & Omori, K. (1998). Starvation of drifting goby
598 larvae due to retention of free embryos in upstream reaches. *Environmental Biology of Fishes*,
599 52, 321–329. <https://doi.org/10.1023/A:1007333302864>
- 600 Nei, M., Tajima, F., & Tateno, Y. (1983). Accuracy of estimated phylogenetic trees from
601 molecular data. *Journal of Molecular Evolution*, 19, 153–170.
- 602 Nishida, M. (2001). Dairanka eno purosusu: Yoshinobori rui (The process toward large egg
603 forms: *Rhinogobius*. In A. Goto, & K. Iguchi (Eds.), *Suisei doubutsu no ran-saizu:*
604 *Seikatsushi no hen'i, shubunka no seibutsugaku (Evolutionary biology of egg size in aquatic*
605 *animals)* (pp. 149– 170). Tokyo, Japan: Kaiyusha. (in Japanese).
- 606 Nosil, P. (2012). *Ecological speciation*. Oxford, UK: Oxford University Press.
- 607 Ohara, K., Takagi, M., Hashimoto, M., Miyazaki, K., & Hirashima, K. (2008). DNA markers
608 indicate low genetic diversity and high genetic divergence in the landlocked freshwater goby,
609 *Rhinogobius* sp. YB, in the Ryukyu Archipelago, Japan. *Zoological Science*, 25, 391–400.
610 <https://doi.org/10.2108/zsj.25.391>
- 611 Ohara, K., Takahashi, D., & Takagi, M. (2004). Isolation of microsatellite loci in the
612 freshwater goby, *Rhinogobius* sp. (Gobiidae). *Molecular Ecology Notes*, 4, 449–451.
613 <https://doi.org/10.1111/j.1471-8286.2004.00677.x>
- 614 Omori, K., Ohnishi, H., Hamaoka, H., Kunihiro, T., Ito, S., Kuwae, M., ... Iguchi, K. (2012).
615 Speciation of fluvial forms from amphidromous forms of migratory populations. *Ecological*
616 *Modelling*, 243, 89–94. <https://doi.org/10.1016/j.ecolmodel.2012.06.006>
- 617 Osozawa, S., Shinjo, R., Armid, A., Watanabe, Y., Horiguchi, T., & Wakabayashi, J. (2012).
618 Palaeogeographic reconstruction of the 1.55 Ma synchronous isolation of the Ryukyu Islands,
619 Japan, and Taiwan and inflow of the Kuroshio warm current. *International Geology Review*,
620 54, 1369–1388. <https://doi.org/10.1080/00206814.2011.639954>

- 621 Ota, H. (1998). Geographic patterns of endemism and speciation in amphibians and reptiles of
622 the Ryukyu Archipelago, Japan, with special reference to their paleogeographical implications.
623 *Researches on Population Ecology*, 40, 189–204. <https://doi.org/10.1007/BF02763404>
- 624 Parent, C. E., & Crespi, B. J. (2006). Sequential colonization and diversification of Galapagos
625 endemic land snail genus *Bulimulus* (Gastropoda, Stylommatophora). *Evolution*, 60, 2311–
626 2328. <https://doi.org/10.1111/j.0014-3820.2006.tb01867.x>
- 627 Pritchard, J. K., Stephen, M., & Donnelly, P. (2000). Inference of population genetic structure
628 using multilocus genotype data. *Genetics*, 155, 945–959.
- 629 R Core Team (2014). *R: A language and environment for statistical computing*. Vienna,
630 Austria: R Foundation for Statistical Computing. <http://www.R-priject.org/>
- 631 Rabosky, D. L. (2016). Reproductive isolation and the causes of speciation rate variation in
632 nature. *Biological Journal of the Linnean Society*, 118, 13–25.
633 <https://doi.org/10.1111/bij.12703>
- 634 Rabosky, D. L., & Matute, D. R. (2013). Macroevolutionary speciation rates are decoupled
635 from the evolution of intrinsic reproductive isolation in *Drosophila* and birds. *Proceedings of*
636 *the National Academy of Sciences of the United States of America*, 110, 15354–15359. <https://doi.org/10.1073/pnas.1305529110>
- 638 Reznick, D. N., & Ricklefs, R. E. (2009). Darwin's bridge between microevolution and
639 macroevolution. *Nature*, 457, 837–842. <https://doi.org/10.1038/nature07894>
- 640 Rosenblum, E. B., Sarver, B. A. J., Brown, J. W., Des Roches, S., Hardwick, K. M., Hether, T.
641 D., ... Harmon, L. J. (2012). Goldilocks meets Santa Rosalia: An ephemeral speciation model
642 explains patterns of diversification across time scales. *Evolutionary Biology*, 39, 255–261.
643 <https://doi.org/10.1007/s11692-012-9171-x>
- 644 Rosindell, J., & Phillimore, A. B. (2011). A unified model of island biogeography sheds light
645 on the zone of radiation. *Ecology Letters*, 14, 552–560.
646 <https://doi.org/10.1111/j.1461-0248.2011.01617.x>

- 647 Schluter, D., & Nagel, L. M. (1995). Parallel speciation by natural selection. *American*
648 *Naturalist*, 146, 292–301.
- 649 Seehausen, O., & Wagner, C. E. (2014). Speciation in freshwater fishes. *Annual Review of*
650 *Ecology, Evolution, and Systematics*, 45, 621–651.
651 <https://doi.org/10.1146/annurev-ecolsys-120213-091818>
- 652 Shinomiya, A., Sasabe, K., Sakurai, M., & Kishino, T. (2005). Morphological characters of
653 newly hatched larvae and their habitat selection of a fluvial goby, *Rhinogobius* sp. YB in
654 the upper stream of Sumiyo River, Amami-oshima Island, Japan. *Japanese Journal of*
655 *Ichthyology*, 52, 1–8. (in Japanese with English abstract). [https://doi.org/10.11369](https://doi.org/10.11369/jji1950.52.1)
656 [/jji1950.52.1](https://doi.org/10.11369/jji1950.52.1)
- 657 Suzuki, T., Shibukawa, K., Yano, K., & Senou, H. (2004). *A photographic guide to the*
658 *gobioid fishes in Japan*. Tokyo, Japan: Heibonsha. (in Japanese).
- 659 Tachihara, K. (2009). Two landlocked *Rhinogobius* species in the Ryukyu Archipelago:
660 Conservation and the future of gobies endemic to isolated rivers. *Japanese Journal of*
661 *Ichthyology*, 56, 70–74. (in Japanese). <https://doi.org/10.11369/jji.56.70>
- 662 Takahashi, D., & Kohda, M. (2004). Courtship in fast water currents by a male stream goby
663 (*Rhinogobius brunneus*) communicates the parental quality honestly. *Behavioral Ecology and*
664 *Sociobiology*, 55, 431–438. <https://doi.org/10.1007/s00265-003-0693-1>
- 665 Takezaki, N., Nei, M., & Tamura, K. (2010). POPTREE2: Software for constructing
666 population trees from allele frequency data and computing other population statistics with
667 Windows interface. *Molecular Biology and Evolution*, 27, 747–752.
668 <https://doi.org/10.1093/molbev/msp312>
- 669 Tamada, K. (2005). Clutch and egg size in *Rhinogobius* sp. DA inhabiting small rivers.
670 *Japanese Journal of Ichthyology*, 52, 17–20. (in Japanese with English abstract).
671 <https://doi.org/10.11369/jji1950.52.17>
- 672 Taylor, E. B., Boughman, J. W., Groenenboom, M., Sniatynski, M., Schluter, D., & Gow, J. L.
673 (2006). Speciation in reverse: Morphological and genetic evidence of the collapse of a

674 three-spined stickleback (*Gasterosteus aculeatus*) species pair. *Molecular Ecology*, 15, 343–
675 355. <https://doi.org/10.1111/j.1365-294X.2005.02794.x>

676 Tominaga, A., Matsui, M., Eto, K., & Ota, H. (2015). Phylogeny and differentiation of
677 wide-ranging Ryukyu Kajika frog *Buergeria japonica* (Amphibia: Rhacophoridae):
678 Geographic genetic pattern not simply explained by vicariance through strait formation.
679 *Zoological Science*, 32, 240–247. <https://doi.org/10.2108/zs140227>

680 Vähä, J.-P., & Primmer, C. R. (2005). Efficiency of model-based Bayesian methods for
681 detecting hybrid individuals under different hybridization scenarios and with different
682 numbers of loci. *Molecular Ecology*, 15, 63–72.
683 <https://doi.org/10.1111/j.1365-294X.2005.02773.x>

684 Vamosi, S. M. (2003). The presence of other fish species affects speciation in threespine
685 sticklebacks. *Evolutionary Ecology Research*, 5, 717–730.

686 Wagner, C. E., Harmon, L. J., & Seehausen, O. (2014). Cichlid species– area relationships are
687 shaped by adaptive radiations that scale with area. *Ecology Letters*, 17, 583–592.
688 <https://doi.org/10.1111/ele.12260>

689 Weigelt, P., Steinbauer, M. J., Cabral, J. S., & Kreft, H. (2016). Late Quaternary climate
690 change shapes island biodiversity. *Nature*, 532, 99–102. <https://doi.org/10.1038/nature17443>

691 Wellborn, G. A., & Langerhans, R. B. (2014). Ecological opportunity and the adaptive
692 diversification of lineages. *Ecology and Evolution*, 5, 176– 195.
693 <https://doi.org/10.1002/ece3.1347>

694 Yamasaki, Y. Y., Nishida, M., Suzuki, T., Mukai, T., & Watanabe, K. (2015). Phylogeny,
695 hybridization, and life history evolution of *Rhinogobius* gobies in Japan, inferred from
696 multiple nuclear gene sequences. *Molecular Phylogenetics and Evolution*, 90, 20–33.
697 <https://doi.org/10.1016/j.ympev.2015.04.012>

698 Yoder, J. B., Clancey, E., Des roches, S., Eastman, J. M., Gentry, L., Godsoe, W., ... Harmon,
699 L. J. (2010). Ecological opportunity and the origin of adaptive radiations. *Journal of*
700 *Evolutionary Biology*, 23, 1581–1596. <https://doi.org/10.1111/j.1420-9101.2010.02029.x>

701 Yoshigou, H. (2014). Annotated checklist and bibliographic records of inland water fishes of
702 the Ryukyu Archipelago, Japan. *Fauna Ryukyana*, 9, 1–153. (in Japanese with English
703 abstract).

704

705 **TABLE 1** Results of model selection by ABC analysis

Tested geographic scale	Origins of freshwater form in the best model	Gene flow	Posterior probability of the best model	Bayes factors of the best model against all models	Bayes factor of the best model against another origin models	Type 2 error rate (all models)	Type 2 error rate (against another origin models)
Within-islands							
Amami-Oshima Is.	Single origin	Partial	0.5087	1.668-37.70	12.04-37.70	0.05-0.17	0.09-0.12
Tokunoshima Is.	Single origin	All	0.3128	1.355-2069	133.5 - 2069	0.02-0.11	0.05-0.07
Okinawajima Is.	Single origin	All	0.6357	1.754 - 4300	544.4 - 4300	0-0.04	0
Kumejima Is.	Single origin	Partial	0.3254	1.047 - 18735	4368- 18735	0.02-0.17	0.02-0.09
Iriomotejima Is.	Single origin	Partial	0.4544	1.321 - 309807	3247- 309807	0.01-0.15	0.01-0.15
Among-islands within-regions							
Middle Ryukyu	Originated in respective islands	All	0.4471	1.090-519.2	23.97-24.37	0-0.35	0
South Ryukyu	Single origin	All	0.8949	13.63 - 191.0	36.98 - 191.0	0-0.08	0-0.08
Tokunoshima Is. – Okinoerabujima Is.	Single origin	All	0.7461	4.680 - 122.6	9.419 -122.6	0.06-0.1	0.06-0.1
Among-regions							
Middle Ryukyu – South Ryukyu	Robust result could not obtain.	-	-	-	-	-	-

706

707

708

709 **TABLE 2** Results of the logistic regression analysis

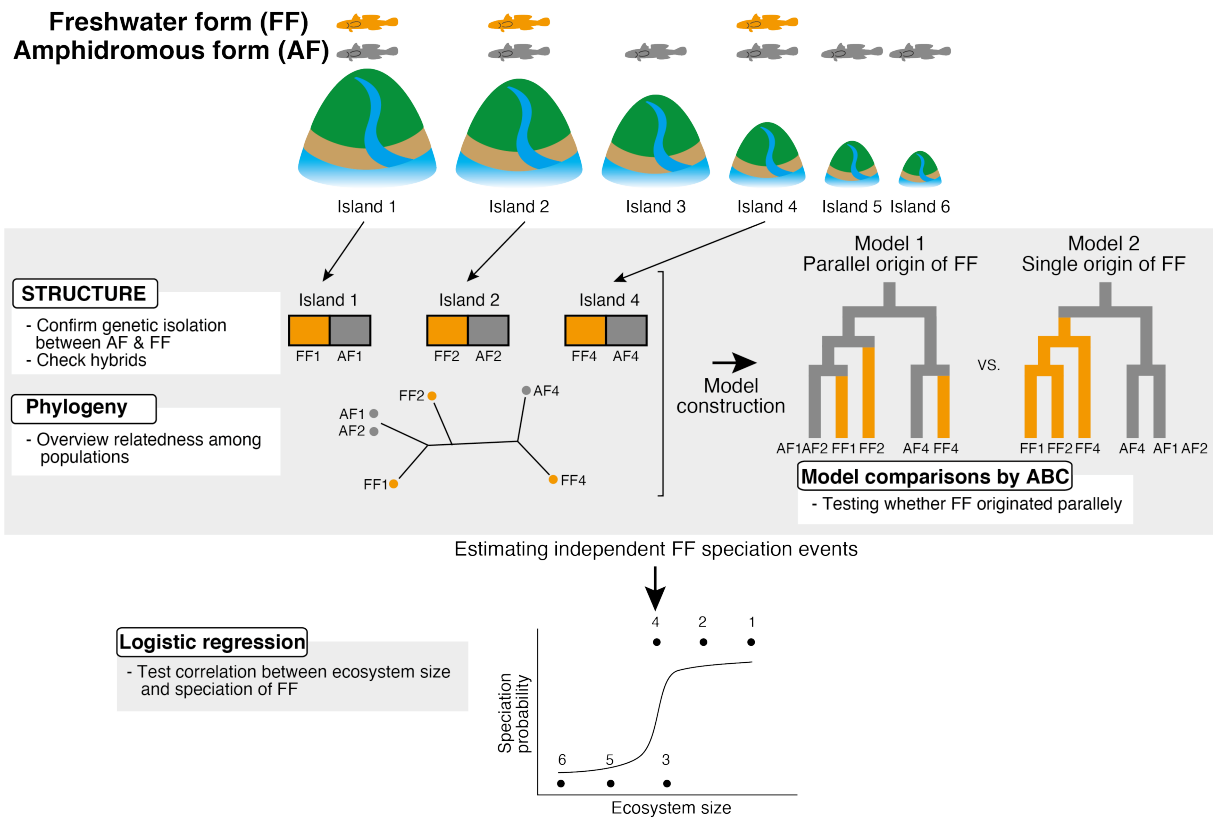
Explanatory variable	Data Set	<i>P</i> value	Difference of deviance	<i>R</i> ²
Island Area	Larger island	0.0016	12.265	0.577
	Smaller island	0.0061	8.650	0.407
Catchment Area	Larger island	0.0026	12.055	0.567
	Smaller island	0.0187	6.923	0.325
River Length	Larger island	0.0038	11.699	0.550
	Smaller island	0.0172	7.028	0.330
No. Waterfalls	Larger island	0.0013	13.712	0.645
	Smaller island	0.3277	1.407	0.066

710

711 *Note:* The results presented in this table are based on the assumption that the freshwater form
 712 (FF) has five independent origins. The data set column indicates the assumption of the islands
 713 where speciation occurred. Our ABC analysis indicated that FF in Tokunoshima Is. and
 714 Okinoerabujima Is., and Ishigakijima Is. and Iriomotejima Is. shared their origins. “Larger
 715 island” in the data set column assumed that the speciation of FF occurred in the larger island
 716 (i.e., Tokunoshima Is. and Iriomotejima Is.), whereas “smaller island” assumed that speciation
 717 occurred in the smaller island (i.e., Okinoerabujima Is. and Ishigakijima Is.). The *p* values for
 718 all combinations are presented in Table S9.

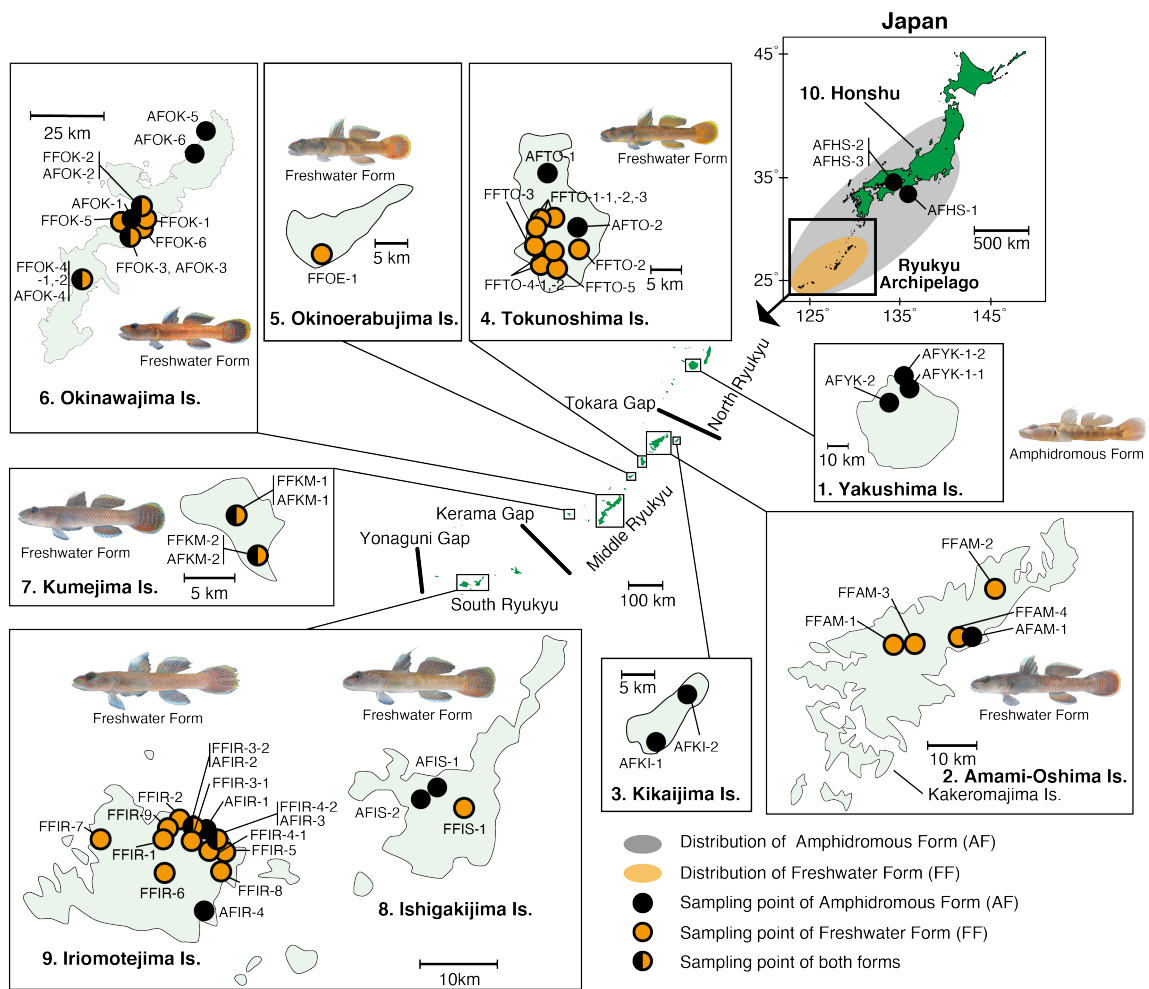
719

720



721

722 **FIGURE 1** Concept diagram of this study based on a hypothetical example. Our final goal
 723 was to determine whether ecosystem
 724 size explains the occurrence of speciation of the freshwater form (FF). To explain the study
 725 workflow, we postulate six islands with the amphidromous form (AF). FF was distributed in
 726 three islands among them. Four main analyses were conducted in this study. First, we
 727 confirmed genetic isolation between AF and FF by structure analysis for each island. Second,
 728 we estimated phylogenetic relationships among populations. Third, we tested independent
 729 speciation of FF by demographic comparison based on ABC analysis. The phylogenetic tree
 730 and geographical locations were used to construct demographic models. The occurrence of
 731 speciation in focal islands was identified by these three analyses. Finally, we tested the
 732 correlation between ecosystem size and occurrence of speciation by logistic regression
 733 analysis.

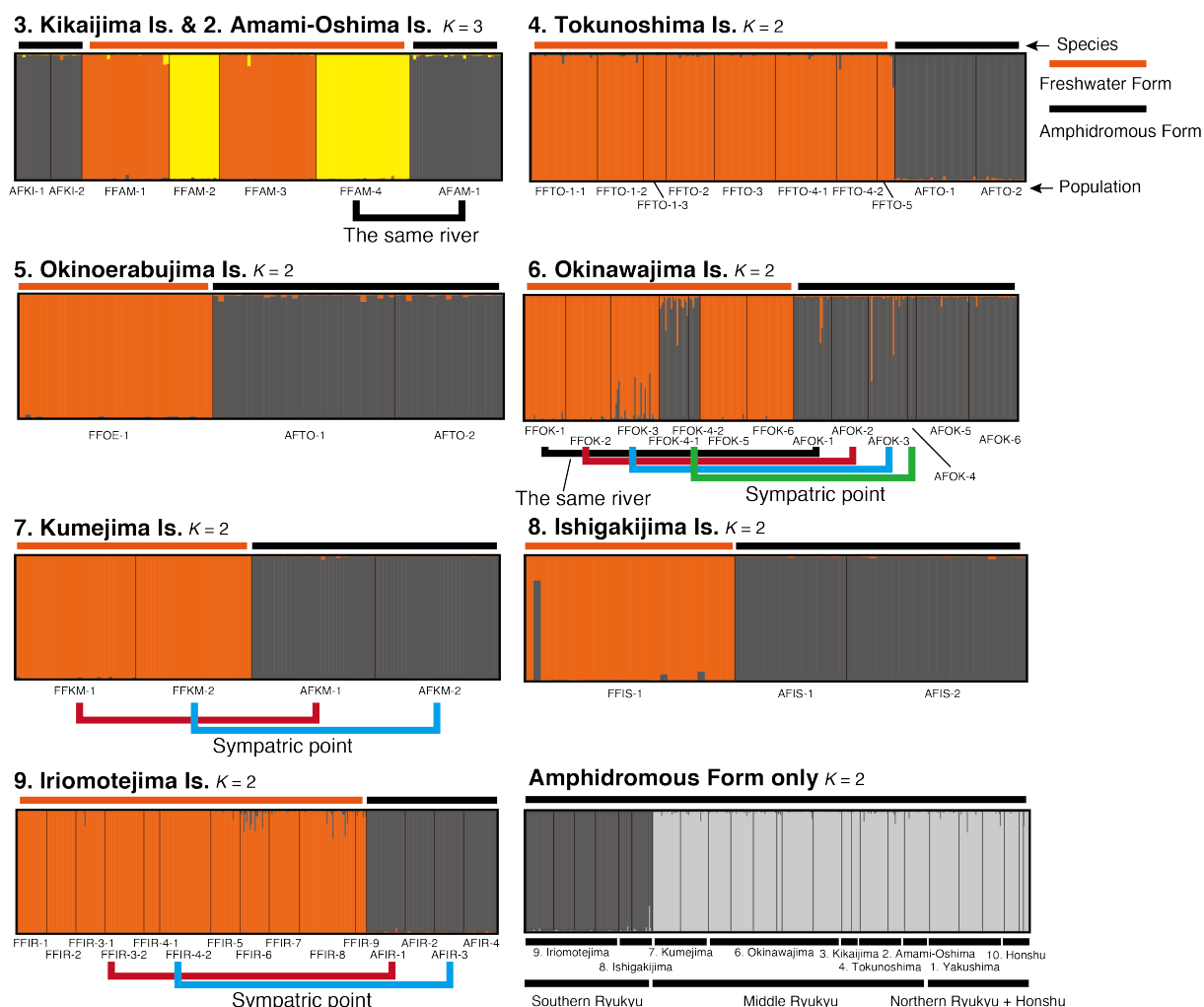


734

735 **FIGURE 2**

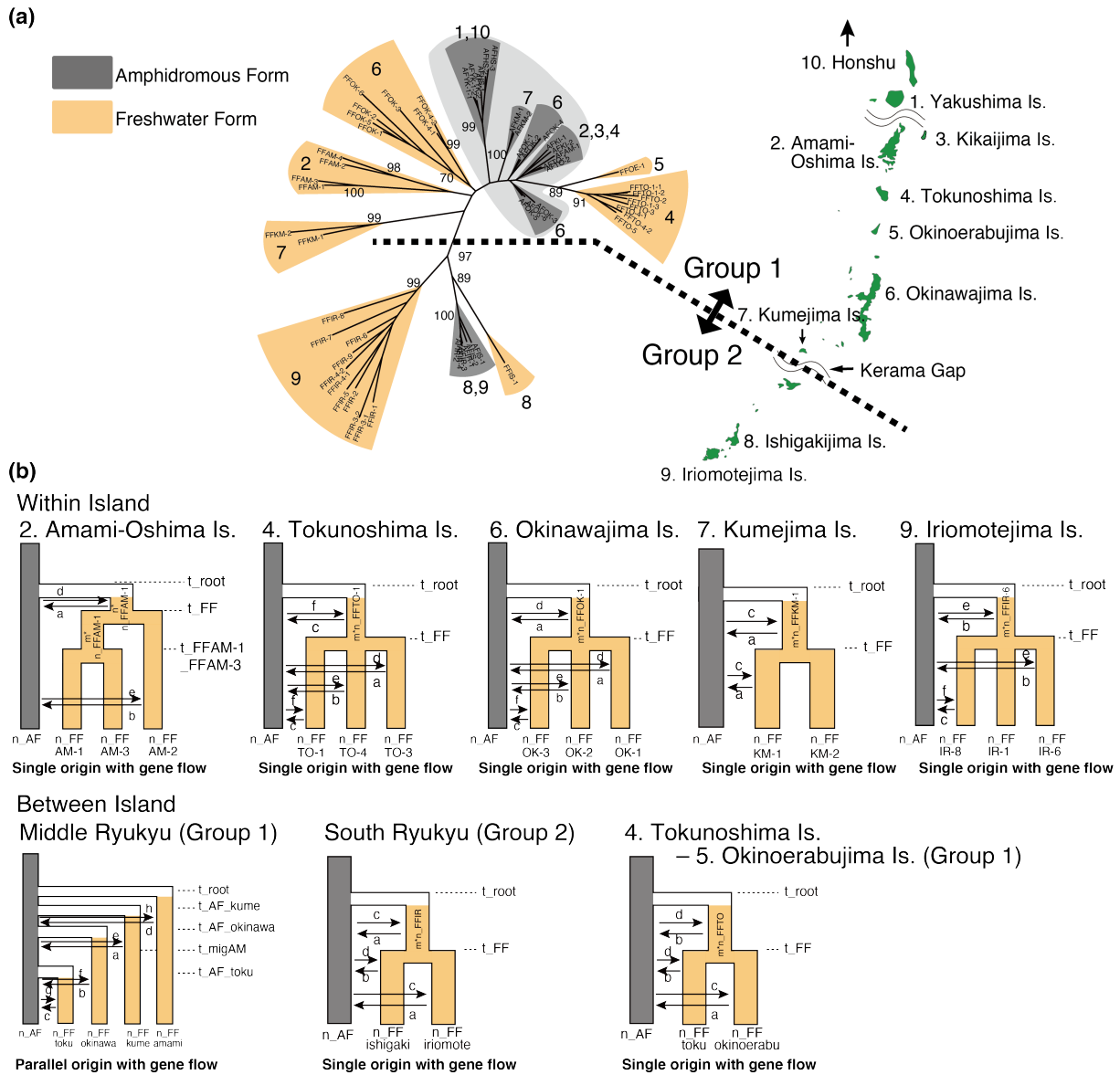
736 Sampling points in this study. The map in the upper right box indicates distribution ranges of the amphidromous form (AF; grey) and the freshwater form (FF; orange) in Japan as a whole. The central map is the whole area of the Ryukyu Archipelago. Black and orange circles indicate the points from where AF and FF were collected, respectively. Sympatrically collected points are indicated by both colours.

737
738
739
740



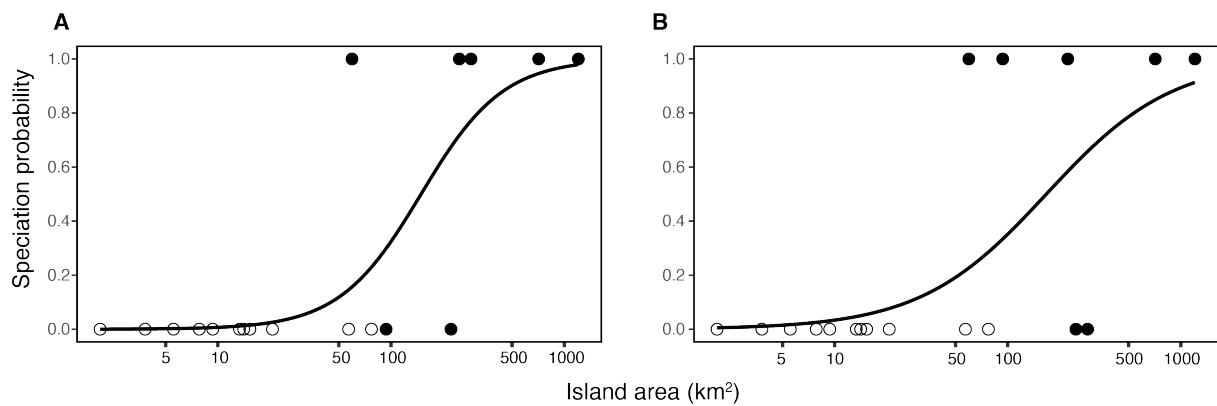
741

742 **FIGURE 3** Results of genetic clustering analysis performed in structure (Amami-Oshima
 743 Is., $K = 3$; others, $K = 2$). Sampling points are indicated by black vertical bars in the
 744 respective bar plots. Horizontal bars at the top of the plot indicate morphological
 745 identification of respective specimens. Sympatric points are indicated by an additional
 746 annotation. The specimens morphologically identified as the freshwater form (FF) are
 747 presented in orange and yellow colours, whereas those identified as the amphidromous form
 748 (AF) are presented in grey.



749

750 **FIGURE 4** Population relationships and evolutionary scenarios of the amphidromous
 751 form (AF) and the freshwater form (FF). (a) Neighbour-joining tree based on Nei's *Da*
 752 distance of 20 microsatellites. Numbers on tree nodes indicate bootstrap values. Only
 753 bootstrap values >70% are indicated. Island numbers are indicated at tips. (b) The models
 754 selected by ABC analysis. AF and FF are represented by black and orange colours,
 755 respectively. Arrows between populations indicate gene flow. Only representative populations
 756 were used for ABC analysis to reduce the amount of calculation (see Materials and methods
 757 and Supporting methods and results S1).



758

759 **FIGURE 5** Results of logistic regression analysis between island size and speciation
 760 probability. Islands with the freshwater form (FF) are represented by filled circles, whereas
 761 those without FF are represented by open circles. Filled circles plotted on the *x*-axis
 762 (speciation probability = 0) are for the islands with FF that was assumed to be of
 763 nonindependent origin. Our ABC analysis indicated that FF in Tokunoshima Is. and
 764 Okinoerabujima Is., and Ishigakijima Is. and Iriomotejima Is. shared their origins. (a)
 765 Speciation of FF is assumed to have occurred in the larger islands (Tokunoshima Is. and
 766 Iriomotejima Is.). (b) Speciation is assumed to have occurred in the smaller islands
 767 (Okinoerabujima Is. and Ishigakijima Is.).

768

769

770

Supplemental methods and results S1: Complete version of Materials and Methods and Results

Materials and Methods

1. Study system

1.1 Study species

The genus *Rhinogobius* (Gobiidae) is the most speciose group of gobies occurring in freshwater. This group contains more than 85 species in East and Southeast Asia, and 18 species in Japan (Oijen et al., 2011; Yamasaki et al., 2015). The freshwater resident form derived from the amphidromous species *R. brunneus* is recognized as an independent species but has not been described to date because of taxonomic confusion. Previous studies (e.g., Akihito et al., 2013) tentatively used the name “*Rhinogobius* sp. YB” for this freshwater form. However, until its evolutionary status is clarified, we will use the code ‘FF’ (freshwater form) for *Rhinogobius* sp. YB and ‘AF’ (amphidromous form) for *R. brunneus* based on their life history modes.

The FF completes its life in freshwater; conversely, AF larvae hatch in rivers and immediately drift down to the sea and return to rivers after 2–3 months (Mizuno, 2001). The eggs and larvae of FF are larger than those of AF. The sizes of eggs and hatched larvae of respective forms and islands have already reported by some studies. As follows, the both egg and larval sizes are greater in FF than in AF. The mean and standard deviation of egg size of FF varies across islands; namely, long axis 3.07 and 3.36 mm, short axis 1.55 and 1.52 mm on Amami-Oshima Island (Kon et al., 2003); long axis 5.06 ± 0.39 mm, short axis 1.77 ± 0.10 mm on Amami-Oshima Is. (Shinomiya et al., 2005); long axis 3.4 and 4.4 mm, short axis 1.4 and 1.7 mm on Tokunoshima Island (measured from the Figure 2 of Oshiro, 1981), long axis 4.3 ± 0.2 mm, short axis 1.4 ± 0.1 mm on Okinawajima Island (Hirashima & Tachihara, 2000); long axis 5.4 mm, short axis 1.7 mm on Ishigakijima Island (measured from the Figure 2 of Oshiro, 1981). We also measured the egg size of FF spawned in a tank by one individual collected from Shirase River on Kumejima Island (long axis 3.52 ± 0.24 mm, short axis 1.31 ± 0.07 mm, $n=26$). The mean and standard deviation of egg size of AF was recorded on Okinawajima Island; namely, long axis 2.8 ± 0.1 mm, short axis 0.9 ± 0.04 mm (Kondo et al., 2013). The mean and standard deviation for notochord length of hatched larvae of FF varies across islands; namely, 5.0 mm on Amami-Oshima Is. (Kon et al., 2013);

34 5.96–6.60 mm on Amami-Oshima Is. (Shinomiya et al., 2004); 5.3 mm on Okinawajima
35 Is. (Hirashima & Tachihara 2000). The mean and standard deviation in notochord length
36 of hatched larvae of AF on Okinawajima Is. was 3.7 ± 0.2 mm (Kondo et al., 2013);
37 3.59 ± 0.087 mm on Shikoku Island in the Japanese archipelago (Moriyama et al., 1998).
38 Fluvial species of this genus tend to produce larger eggs and larvae (Yamasaki et al.,
39 2015). Such traits are considered to be an adaptation to river environments, where small
40 food items (i.e., plankton) are scarce and higher level of swimming ability in river
41 current is needed during the larval period (Nishida, 2001; Shinomiya et al., 2005;
42 Hirashima & Tachihara, 2000; Yamasaki et al., 2015).

43 Although the distribution of the two species sometimes overlapped in a single river,
44 the tendencies of their distributions were different. AF mainly lived in the middle to
45 lower reaches of rivers. In fact, AF populations are actually only found in rivers shorter
46 than 18 km in the central Honshu region of Japan (Tamada, 2005). The larvae of AF
47 must rapidly reach the sea to avoid starvation because of the scarcity of small food in
48 rivers (Moriyama, Yanagisawa, Mizuno, & Omori, 1998; Iguchi & Mizuno, 1999).
49 Thus, they might not be able to reproduce effectively in areas far from the sea (Iguchi &
50 Mizuno, 1999). On the other hand, FF mainly live in the upper reaches of rivers, and are
51 often found in the upper areas of waterfalls where no other fishes are distributed (Kano
52 et al., 2012).

53

54 **1.2 Study area**

55 The Ryukyu Archipelago is a series of islands located in the southwestern part of
56 Japan (Figure 2). The islands separated simultaneously from the continent about 1.5
57 million years ago (Osozawa et al., 2012). Three deep straits (Tokara Gap, Kerama Gap,
58 and Yonaguni Gap) divide this archipelago into three regions: North Ryukyu, Middle
59 Ryukyu, and South Ryukyu. These gaps act as biogeographic barriers for terrestrial
60 organisms (Ota 1998). On the islands of North Ryukyu, the biota was destroyed by
61 pyroclastic flow 7300 years ago (Machida et al., 2001). AF is distributed in all three
62 regions, as well as mainland Japan, whereas FF is only distributed in Middle and South
63 Ryukyu (Yoshigou, 2014). In the Ryukyu Archipelago, primary freshwater fishes are
64 almost absent, and most fishes are diadromous. Therefore, fish assemblages in
65 freshwater areas are simple and similar among islands (Yoshigou, 2014).

66

67 **2. Sampling**

68 All animal experiments followed the rules, regulations, and guidelines relating to animal
69 experimentation of Kyoto University. Sampling of FF and AF was conducted from June
70 2012 to March 2015 on seven islands where former populations have been recorded
71 previously (Locality code 2. Amami-Oshima Is., 4. Tokunoshima Is., 5. Okinoerabujima
72 Is., 6. Okinawajima Is., and 7. Kumejima Is. in Middle Ryukyu; 8. Ishigakijima Is. and
73 9. Iriomotejima Is. in South Ryukyu; Figure 2; Table S1; Ohara et al. 2008; Yoshigo,
74 2014). In addition, AF specimens were collected from Kikaijima Is. (Locality code 3) in
75 Middle Ryukyu, Yakushima Is. (1) in North Ryukyu, and mainland Japan (10). The
76 specimens were identified by body colour patterns under fresh conditions (Suzuki et al.,
77 2004; Kano et al., 2012). Identification based on body colour was validated by the
78 results of the genetic analysis; however, some samples were difficult to identify or
79 showed inconsistent genetic characteristics (3.2% of specimens in total, including
80 potential hybrids). When genetic data suggested the focal specimens were not judged as
81 hybrids based on the criteria described below and that wrong identification based on
82 morphology in the field was likely, these samples were removed from further analysis.
83 In the case that a sample was judged to be a hybrid (see 4.2), we tentatively named it
84 based on the morphology. It should be noted that we found that morphologically
85 identified FFOK-4-1 and FFOK-4-2 populations of FF largely reflected the genetic
86 features of AF (Figure 3 and Supporting Information Figure S3). We judged them as
87 hybrid populations and did not remove any specimens from the analysis even if they
88 were not judged as hybrid (see Discussion). Specimens were anesthetized with iced
89 water or 2-phenoxy-ethanol, and then a fin was clipped and preserved in 99% ethanol.
90 Most specimens were fixed in 10% formalin or 99% ethanol. As a result, 842 FF and
91 578 AF specimens from 52 points in 42 river systems were included in the population
92 genetic analysis. The average numbers of localities and specimens per island were 5.2
93 and 142, respectively. Both forms were collected from nine river systems; they were
94 sympatrically collected at seven points.

95

96

97 **3. Molecular experiment**

98 **3.1 Microsatellite marker development**

99 We developed new microsatellite markers following the methods of Takeshima et al.

100 (2017). Total genomic DNA was extracted from one AF specimen collected from
101 DAIR-4 in Iriomotejima Island by using a Genomic DNA Purification Kit (Promega,
102 Madison, Wisconsin, USA). We concentrated CA repeat regions by using the target
103 capture method and sequenced them using the Roche 454 GS Junior system (Accession
104 No. DRA006388). Primers and multiplex systems were designed using QDD ver. 2
105 (Meglecz et al., 2010) and Multiplex Manager v. 1.2 (Holleley & Geerts 2009),
106 respectively. One out of four universal tail sequences was added to each forward primer
107 for fluorescent labelling during PCR (Blacket et al., 2012).

108

109 **3.2 DNA extraction, PCR, and genotyping**

110 Total genomic DNA was extracted from the fin clip by using the Genomic DNA
111 Purification Kit. Before PCR amplification, a droplet of 1.25 μ L extracted DNA was
112 placed on a 96 well plate and dried. Total 21 microsatellite markers (18 originally
113 developed loci and three loci reported by Ohara et al. (2004)) were analyzed. We
114 developed four multiplex sets and simultaneously amplified 4–10 loci (Table S2). For
115 the markers that were developed, the final concentration of each forward and reverse
116 primer was 1 and 2 μ M, respectively. The 1 μ M tailed fluorescent primer for each
117 forward primer was also added to the multiplex reaction mixture. For the three loci
118 reported by Ohara et al. (2004), 5'-fluorescent-labelled forward primers were used, and
119 the final concentration of each forward and reverse primers was adjusted to 1 μ M. PCR
120 amplification was performed in a 4 μ L volume containing 2.43 μ L of ultrapure water,
121 0.066 μ L of primer mix, and 1.5 μ L of Type-IT Microsatellite PCR Kit (Qiagen, Hilden,
122 Germany) with the dried DNA. The PCR settings for the originally developed markers
123 included a first step of denaturation at 95°C for 5 min, followed by 35 cycles of the
124 second step (denaturation: 94°C, 30 s; annealing: 63°C, 90 s; and extension: 72°C, 75 s),
125 and the final extension at 72°C for 30 min. For the PCR for the loci of Ohara et al.
126 (2004), the second step was changed as follows: 28 cycles of denaturation at 94°C for
127 30 s; annealing at 53°C for 90 s; and extension at 72°C for 60 s. We directly added 10
128 μ L HiDi Formamide (Thermo Fisher Scientific, Waltham, MA, USA) and 0.2 μ L of
129 GeneScan 500Liz dye Size Standard or GeneScan 400HD Rox dye Size Standard
130 (Thermo Fisher Scientific) for the respective PCR products. Next, we subjected the
131 products to heat shock treatment (95°C, 5 min; 0°C, 5 min). The peak size data were
132 acquired using ABI 3130xl sequencer (Applied Biosystems, Foster City, CA, USA) and

133 were analyzed using GeneMapper v.4.0 (Applied Biosystems). The acquired allele size
134 data were transformed for integral number by using tandem v.1.09 (Matschiner &
135 Salzburger 2009).

136 We genotyped a total of 21 microsatellite loci for population genetic analyses
137 (Table S2). Null alleles were checked using software Micro-checker ver.2.2.3 (van
138 Oosterhout et al., 2004) under 100% interval condition for the 25 populations of *R.*
139 *brunneus*, which were assumed to not be affected by the bottleneck effect. Null alleles
140 were detected in several *R. brunneus* populations. We excluded locus br_b_04 from the
141 analysis because null alleles were detected at this locus in both North Ryukyu and
142 Middle Ryukyu populations. Null alleles in other loci occurred sporadically; hence, we
143 used the other 20 loci for the following analysis to avoid reduction of information.
144 Linkage disequilibrium among loci was tested using 10,000 times of permutation
145 implemented in Arlequin ver.3.5 (Excoffier & Lischer 2010), and Bonferroni correction
146 was applied to *P*-values to avoid type I errors. No significant linkage disequilibrium
147 between loci was detected (adjusted $P \geq 0.05$). The final genotype data were deposited
148 in dryad (doi: 10.5061/dryad.mpg4f4qvh).

149

150 **4. Data analysis**

151 **4.1 Calculation of summary statistics and estimation of phylogenetic trees**

152 We calculated the expected and observed heterozygosity (*He* and *Ho*, respectively), the
153 effective number of alleles (*Ae*) for the respective populations, and population-pairwise
154 *F_{ST}* and Jost's *D* by using GenoDive 2.0b27 (Meirmans & van Tienderen 2004). We
155 also calculated allelic richness (*Rs*) of the respective populations by using FSTAT
156 v2.9.3.2 (Goudet 1995). Phylogenetic relationships among populations were estimated
157 based on *Nei's Da* distance by using the neighbor-joining method implemented in
158 Poptree2 (Takezaki et al., 2010). The credibility of each clade was evaluated using
159 1,000 bootstrap resamplings. To assess hierarchical genetic differentiation, we
160 conducted AMOVA (Excoffier, Smouse, & Quattro, 1992) implemented in GenoDive
161 separately for AF and FF. The number of hierarchies to be tested was five. However,
162 AMOVA could include only four hierarchies in one analysis. To deal with this problem,
163 we conducted AMOVA separately in two geographic scales. In the among-islands
164 within-regions scale, we set following four hierarchies: within-individuals,
165 among-individuals, among-rivers, and among-islands. This analysis was separately

166 conducted for Middle Ryukyu and South Ryukyu. In the among-regions scale, we set
167 the following four hierarchies: within-individuals, among-individuals, among-islands,
168 and among-regions. F_{ST} -analog distance was applied. Significance was tested using 999
169 permutations.

170

171 **4.2 STRUCTURE analysis**

172 To infer the existence and extent of gene flow and genetic isolation, we analyzed
173 the genetic structure of the population using STRUCTURE v.2.3.4 (Pritchard, Stephen,
174 & Donnelly, 2000; Lawson, van Dorp, & Falush, 2018). We prepared two datasets: one
175 for both FF and AF populations from each island, and another for only AF populations
176 from all islands. Because AF was not found on Okinoerabujima Is., the FF on the island
177 were analyzed along with the AF from the nearest island, Tokunoshima Is. Admixture
178 model with correlated allele frequencies was assumed. Markov Chain Monte Carlo was
179 performed with 100,000 generations for burn-in and final 1,000,000 generations. Ten
180 independent runs for each K from 1 to 10 were evaluated. After a run finished, we
181 calculated Evanno's ΔK by using STRUCTURE Harvester (Evanno, Regnaut, & Goudet,
182 2005; Earl & von Holdt, 2012) was used for a reference for K value selection. Hybrid
183 individuals were evaluated using q -values; we arbitrarily judged specimens as hybrids
184 (or individuals with hybrid origin) in cases where minor genetic components accounted
185 for 10% or more of the aggregate (Vähä & Primmer 2006).

186

187 **4.3 Testing parallel speciation of FF by using ABC**

188 To statistically examine the single or parallel origins of FF, and to estimate which FF
189 populations shared their origins we conducted a demographic model selection under the
190 approximate Bayesian computation (ABC framework, Beaumont, 2010; Bertorelle et
191 al., 2010) implemented in ABCtoolbox v.1 (Wegmann et al., 2010) and R package 'abc'
192 v.2.1 (Csilléry et al., 2012). Coalescent simulation was conducted using fastsimcoal2
193 v.2.5.2.1.1 (Excoffier & Foll 2011; Excoffier et al., 2013). Summary statistics were
194 calculated using arlsumstat v.3.5 for each simulation (Excoffier & Lischer 2010).

195

196 **4.3.1 Prior distribution, summary statistics, and simulation**

197 Prior distribution and summary statistics were determined after several preliminary
198 analyses. When the range of the prior distribution of a parameter was across the digits,

199 the value was \log_{10} transformed. A uniform distribution for each parameter was assumed.
200 We set a larger prior range for population size of AF (10^2 – $10^{5.3}$) than for FF (10^2 – $10^{4.7}$).
201 The prior distribution of coalescent time (generation) was set to 10^3 – $10^{6.2}$, considering
202 that the Ryukyu Archipelago had separated from the continents 1.5 million years ago,
203 and the generation time of the *Rhinogobius* species was one to two years (Yamasaki,
204 personal observation). The prior distribution of the migration rate was set to 10^{-9} – 10^{-3} .
205 In the case that the models of two or more populations were merged, the ancestral
206 population size was set to the range of 0.5 to 2 times of either of the derived populations.
207 We applied a generalized stepwise mutation model for microsatellite evolution. Our data
208 did not match the simulated data produced using the models under a previously reported
209 average mutation rate of microsatellites in fish (common carp: $5.56 \times$
210 10^{-4} /locus/generation, 95% interval, 1.52×10^{-4} – 1.63×10^{-3} , Yue et al., 2007) from
211 principal component analysis (PCA) (see 4.3.3). Simulated data under several average
212 mutation rates were tested, and then the value of 1.0×10^{-4} was selected, under which
213 the generated data matched the acquired data. The prior distribution of the mutation rate
214 for each locus was set to a Gamma distribution. The maximum difference of repeat
215 number of alleles of each microsatellite in coalescent simulations was set to 45.

216 To capture information regarding population size and differentiation, we selected
217 summary statistics listed in Table S3. Detailed formulae of respective summary statistics
218 are described in the Arlequin 3.5 manual (Excoffier & Lischer 2010). A total of 46 to
219 132 summary statistics for each analysis were used according to the number of the
220 analyzed populations. We performed 1.0×10^6 simulations for each demographic model.

221

222 **4.3.2 General strategy for the construction of demographic models**

223 Our general model construction strategies for ABC analysis were as follows. First,
224 we generated separate models using three geographic scales, i.e. within-islands, regional
225 (Middle and South Ryukyu), and archipelago scales, and examined FF origins in each
226 geographic scale to reduce the number of models and populations examined in a single
227 analysis. Second, we examined both models with and without gene flow between AF
228 and FF populations. This is because a simulation study suggests that gene flow could
229 alter apparent phylogenetic relationships among populations when simple genetic
230 distance was used (Bierne, Gagnaire, & David, 2013). Demographic model comparisons
231 that consider gene flow can overcome this problem (Butlin et al., 2014).

232 In the within-islands analyses, we mainly examined two questions: whether FF
233 had single or parallel origin, and whether gene flow between AF and FF existed on each
234 island. We hypothesized several plausible divergence orders between populations based
235 on the result of STRUCTURE analysis, phylogeny, and geographic locations for
236 respective islands. Thereafter, we compared the models with single/parallel origins and
237 with/without gene flow. Analyses were conducted for the five islands where FF is
238 distributed in multiple rivers (i.e., Amami-Oshima Is., Tokunoshima Is., Okinawajima
239 Is., Kumejima Is., and Iriomotejima Is.). Ishigakijima Is. and Okinoerabujima Is. were
240 not included in the analyses because the FF population was only found in one river
241 system on each island. We selected three (or two) river populations for each island to
242 reduce the model complexities and calculation loading. FF specimens from a single
243 river system were merged to increase calculation accuracy. Populations were selected by
244 the following criteria to extract information of population history as much as possible;
245 geographically distant river systems, to avoid bias from shared population history
246 among geographically nearby populations; and lower rates of missing data. The
247 constructed models (a total of 34 models) are described in Supporting Information
248 Figure S1A–E.

249 For the among-islands within-region analyses, we constructed several plausible
250 models separately for the regions of Tokunoshima Is. and Okinoerabujima Is., Middle
251 Ryukyu, and South Ryukyu. Five to ten models were constructed on the basis of the
252 phylogenetic tree and geographic distance between islands. Gene flow settings followed
253 the results of within-islands scale analysis. For each island, we selected only one FF
254 population whose population size was the largest, as judged from *He* values. Finally, in
255 the between-region analyses, we compared ten models generated from the results of the
256 previous analyses. All constructed models are described in Figure S1F–I. Scripts that
257 described models were deposited in Dryad (doi: 10.5061/dryad.mpg4f4qvh).

258

259 **4.3.3 Prior check**

260 We performed PCA by using *gfitpca* function implemented in *abc* package to check
261 the validity of the prior distribution. We confirmed that the observed summary statistics
262 included of the 90% envelope of each simulation result (data not shown).

263

264 **4.3.4 Posterior probabilities of models**

265 The posterior probability of each model based on the observed value was estimated
266 using the *neuralnet* method, which is a non-linear regression method involving neural
267 networks. We used *postpr* function implemented in the *abc* package. The regression step
268 was performed using 30 neural networks and 10 hidden layer units. Five hidden layer
269 units were set in the analysis of the archipelago scale, because an error occurred when
270 10 hidden layers were set. We used 1% of simulations with summary statistic values
271 closest to the observed ones (tolerance rate = 1%). We conducted 10 calculations for
272 each model selection by changing the tolerance rates from 0.1% to 1.0% with
273 increments of 0.1% to evaluate the robustness of the results by using a tolerance rate of
274 1%.

275

276 **4.3.5 Confidence in model selection**

277 To evaluate the accuracy of the model selection, we calculated the misclassification
278 rate by using function *cv4postpr* implemented in *abc* package. We randomly generated
279 100 pseudo-observed data sets (pods) of each model from the prior distribution and
280 performed model selection for the respective pods with 10^5 simulations for each model.
281 Model selection was performed using *neuralnet* method under 10 neural networks and
282 10 hidden layer units. Five hidden layer units were set in the analysis at the archipelago
283 scale as in the calculation of posterior probabilities. Type 1 and type 2 errors were
284 evaluated for each model: type 1 error rate is the probability that the other model was
285 selected despite the focal model being true, and type 2 error rate is the probability that
286 the focal model was selected despite the other model being true. For evaluating the
287 accuracy of the model selection results, the type 2 error rate should be carefully
288 considered (Cornuet et al., 2010; Lombert et al., 2014).

289

290 **4.3.6 Estimating posterior distributions of parameters**

291 We calculated the posterior distribution of each parameter under the selected model
292 by using function *abc* implemented in the *abc* package. We retained 1% of simulations
293 close to the observed summary statistics, and we performed regression analysis by using
294 the *neuralnet* method under 30 neural networks and five hidden layer units. Parameters
295 of population size and event time were log transformed before regression analysis.

296

297 **4.3.7 Posterior predictive check**

298 We checked whether the selected model could produce the observed value under
299 the posterior distribution. We generated 1000 simulations under the posterior
300 distribution of the selected model and calculated summary statistics. In addition to the
301 originally used summary statistics, the following summary statistics were also
302 calculated: the mean allelic range over loci for each population (R_{pop}) and its standard
303 deviation (Rsd_{pop}), the mean allelic range over loci and populations ($MEAN_R$) and
304 its standard deviation ($MEAN_Rsd$), and the mean total allelic range over loci and
305 populations (TOT_R), and F_{IS} (FIS). Next, PCA using function *gfitpca* implemented in
306 *abc* package was performed to check whether the observed summary statistics were
307 included in the range of summary statistics generated from the posterior distribution.
308

309 **5. Correlation between island size and speciation probability**

310 We first identified the islands on which the speciation of FF occurred based on the
311 results of STRUCTURE (genetic isolation) and ABC model selection (independent
312 origin). The relationship between ecosystem size and speciation was analyzed for the
313 Middle and South Ryukyu Islands between the Tokara Gap and Yonaguni Gap. This
314 analysis was conducted because these islands are assumed to be similar in age, were
315 never connected to any continent after separation, and were not catastrophically affected
316 by volcanic activity (Osozawa et al., 2012). We acquired distribution information for the
317 two forms from published literature (Yoshigou 2014) and our own field survey, and we
318 only chose the islands where either or both forms were distributed. This is because we
319 assumed that speciation was not likely to have occurred on the island where either of
320 these species is not currently distributed. We coded the occurrence of speciation in each
321 island as follows: 1 for the island where speciation occurred; 0 for the island where
322 speciation did not occur. As a proxy of ecosystem size, we used island area (km^2 ; data
323 obtained from Geospatial Information Authority of Japan), the maximum catchment
324 area (km^2), the maximum river length (m), or the maximum number of waterfalls in
325 each island. The data on rivers on the islands where FF is distributed were only obtained
326 for rivers with FF. The latter three parameters were calculated using the 10 m mesh
327 digital elevation model (Geographic Information Authority of Japan) by using
328 ArcGIS10.3 (ESRI Japan, Tokyo). Because of strong positive correlations among the
329 four variables (Pearson's correlation coefficient: island size and catchment area, 0.959;
330 island size and river length, 0.951; catchment area and river length, 0.993; island size

331 and number of waterfalls, 0.592; catchment area and number of waterfalls, 0.865; river
332 length and number of waterfalls, 0.865), we primarily used island area as the
333 explanatory variable.

334 We performed one-parameter logistic regression analysis in the generalized linear
335 model framework with binomial error structure for the response variable (speciation: 1,
336 presence; 0, absence) using glm package of R v. 3.1.1 (R Core Team, 2014).
337 Explanatory variables, except the number of waterfalls, were \log_{10} transformed. To test
338 the significance of explanatory variables, we conducted a likelihood ratio test using
339 10,000 rounds of parametric bootstrap sampling by using a constant model as the null
340 model that assumes no relationship between ecosystem variables and speciation
341 probability. We calculated McFadden's pseudo R^2 by using the following formula: $r^2 = 1$
342 $- (SSE/SST)$. In this formula, SSE and SST refer to the deviances of the focal and null
343 models, respectively.

344 Furthermore, to examine the effect of the ecosystem size parameters other than
345 island size (i.e., maximum catchment area, river length, and number of waterfalls), we
346 constructed another generalized linear model for speciation probability. We used the
347 residuals from a regression between island size and each of the other ecosystem-size
348 parameters, together with the island size and their interaction, as explanatory variables.
349 Stepwise model selections by AIC scores were conducted using *step* function on R.

350
351

352 **Results**

353 **1. Basic summary statistics**

354 Heterozygosity and allelic richness were generally higher in the AF populations than in
355 the FF ones, possibly suggesting a larger current or historical population size in the
356 former (AF: He , 0.296–0.554 and R_s , 1.94–3.65; FF: He , 0.070–0.390 and R_s , 1.21–
357 3.08; Supporting Information Table S1, Figure S2A, B). The AF populations showed
358 relatively little pairwise genetic differentiation within-islands ($F_{ST} = -0.008$ –0.145;
359 Jost's $D = -0.009$ –0.102; Supporting Information Tables S4, 5, Figure S2C, D). In the
360 case of the among-islands within-regions scale, pairwise genetic differentiation was
361 moderate (F_{ST} , 0.022–0.168; Jost's D , 0.025–0.219). AMOVA showed significant
362 differentiation in the among-rivers hierarchy in both Middle Ryukyu and South Ryukyu,
363 and in the among-islands hierarchy in Middle Ryukyu (Supporting Information Table

364 S6). Most of the variation was explained by within-individuals hierarchy. Compared
365 with this, the genetic differentiation of AF populations between regions was large (F_{ST} ,
366 0.110–0.456; Jost's D , 0.124–0.613). AMOVA showed significant differentiation in both
367 among-islands and among-regions hierarchies (Supporting Information Table S6).
368 Variations were mainly explained by within-individuals, among-islands, and
369 among-regions hierarchies. The FF populations often showed large genetic
370 differentiation within-islands (F_{ST} , 0.003–0.804; Jost's D , 0.001–0.529; Supporting
371 Information Tables S4, 5, Figure S2C, D) and among-islands within-regions (F_{ST} ,
372 0.312–0.859; Jost's D , 0.146–0.778), as well as among-regions (F_{ST} , 0.507–0.914;
373 Jost's D , 0.580–0.890). AMOVA for FF showed significant genetic differentiation in the
374 hierarchies of among-individuals, among-rivers, among-islands, and among-regions,
375 except among-islands in South Ryukyu (Supporting Information Table S6). In the
376 among-islands within-regions scale analysis, genetic variation was mainly explained by
377 within-individuals, among-rivers, and among-islands hierarchies.

378

379 **2 Genetic isolation and hybrids**

380 In the AF dataset, Evanno's ΔK showed the maximum value at $K = 2$; the populations in
381 North plus Middle Ryukyu and South Ryukyu were clearly separated. In the dataset for
382 each island, including FF and AF populations, ΔK had the maximum value at $K = 2$ for
383 all islands, except one (Amami-Oshima Is.). The two genetic clusters corresponded to
384 the two forms, except on Okinawajima Is. (see below; Figure 3 and Supporting
385 Information Figure S3). ΔK on Amami-Oshima Is. showed the maximum value at $K = 3$,
386 in which FF was further divided into two clusters. Suspected hybrid individuals
387 (q -value, $\geq 10\%$) were rare (2.0% of total individuals). No suspected hybrid individuals
388 were detected from five of the seven sympatric collection points. Several suspected
389 hybrid individuals were detected in 11 FF and AF populations, including two FF
390 populations on Tokunoshima Is. (FFTO-4-2, FFTO-5), three FF on Okinawajima Is.
391 (FFOK-3, FFOK-4-1, and FFOK-4-2), one FF on Ishigakijima Is. (FFIS-1), three FF on
392 Iriomotejima Is. (FFIR-3-1, FFIR-6, and FFIR-8), and two AF on Okinawajima Is.
393 (AFOK-1, AFOK-3). Among these, all FF individuals that were identified based on
394 morphology in the FFOK-4-1 and FFOK-4-2 populations (Okinawajima Is.; see Figures
395 2, 3 and Supporting Information Figure S3) exhibited the genetic features of AF as the
396 dominant component. Except for these two populations, the proportion of suspected

397 hybrid individuals ranged from 3.6% to 21.2% of individuals in a population. Potential
398 hybrids found in the FF populations contained 10.1%–80.5% of the AF component,
399 whereas those of the AF populations contained 25.6%–69.0% of the FF component.

400

401 **3 Genetic relationships among populations**

402 The neighbor-joining tree showed two divergent groups separated by the Kerama
403 Gap in all of the FF and AF populations (bootstrap value, 97%; Figure 4). Each group
404 included both forms, which we present as Group 1 (Honshu, North Ryukyu, and Middle
405 Ryukyu) and Group 2 (South Ryukyu), hereafter.

406 In both groups, FF populations on each island formed a monophyletic group
407 (bootstrap value, 89%–100%), except those on two islands in Group 1 (Amami-Oshima
408 Is., 67%; Okinawajima Is., 33%). In addition, FF populations on different islands
409 (Tokunoshimajima Is. and Okinoerabujima Is.) formed a monophyletic group (89%).
410 The AF populations within each group were closely related to each other, and the
411 populations of each island in Group 1 tended to form a monophyletic group.

412

413

414 **4 Test for multiple origins of FF**

415 In the within-island scale analysis, the single origin model of FF was selected for all
416 islands (Table 1 and Supporting Information Table S7; Figure 4 and Supporting
417 Information Figure S4). Models including gene flow among all or partial populations of
418 the two forms showed higher posterior probability than those not considering gene flow
419 (Table 1 and Supporting Information Table S7; Figure 4 and Supporting Information
420 Figure S4). Type 2 error rates for the selected single origin model against alternative
421 parallel origin models were generally low (Table 1 and Supporting Information Table
422 S7; Figure 4 and S4).

423 At the among-island scale, Group 1 in Middle Ryukyu and Group 2 in South
424 Ryukyu were analyzed separately. For Group 1, we first examined the single origin of
425 FF in the adjoining islands, Tokunoshima Is. and Okinoerabujima Is., which was
426 suggested by the phylogenetic analysis. ABC analysis supported their single origin
427 (posterior probability (PP), 0.746; Bayes factor (BF), 9.42–122.57; Type 2 error rate,
428 0.06–0.1; BFs and Type 2 error rates are of the selected model against parallel origin
429 models; Table 1 and Supporting Information Table S7; Figure S4); thus, only FF on

430 Tokunoshima Is. was used for subsequent analyses. The model selection for Group 1
431 strongly supported the four independent origins of FF (PP, 0.447; BF, 23.97–519.23;
432 Type 2 error rate, 0; BFs and Type 2 error rates are of the selected model against the
433 other non-four origin models; Tables 1 and S7; Figure S4). The model with gene flow
434 between the two forms obtained higher BF than the model not considering gene flow
435 (BF, 4.72; Type 2 error rate, 0.125). In Group 2, the single origin of FF of Iriomotejima
436 Is. and Ishigakijima Is. was supported (PP, 0.895; BF, 36.98–191.0; Type 2 error rate, 0–
437 0.08; BFs and Type 2 error rates are of the selected model against parallel origin models
438 Tables 1 and S7; Figure S4). The model including gene flow between the AF and FF
439 populations was better supported than that not considering gene flow (BF, 13.63; Type 2
440 error rate, 0.06).

441 For the between-region scale, the neighbor-joining tree for populations suggested
442 independent origins of FF in Groups 1 and 2. However, we could not obtain a
443 conclusion at this scale from ABC, because no single model was successfully selected,
444 probably because of insufficient data (PP, 0.0024–0.14; Table S7; Figure S4).

445

446 **5 Correlation between speciation probability and ecosystem size**

447 We confirmed the occurrence of AF or FF populations on 18 islands from the published
448 literature and our field survey (Supporting Information Table S8). As we could not
449 conclude from the ABC analysis whether the FF population originated independently
450 between Groups 1 and 2, we analyzed independent and non-independent scenarios
451 between the Middle and South Ryukyus. In addition, because the single origin of FF
452 was supported for Tokunoshima Is. and Okinoerabujima Is. populations, as well as for
453 Ishigakijima Is. and Iriomotejima Is. populations, we could not conclude which island
454 FF originated in. We developed four datasets for the independent scenario between
455 Groups 1 and 2 with all possible combinations of islands in which speciation would
456 have occurred independently: dataset 1, speciation in Tokunoshima Is. and Iriomotejima
457 Is.; dataset 2, in Tokunoshima Is. and Ishigakijima Is.; dataset 3, Okinoerabujima Is. and
458 Iriomotejima Is.; and dataset 4, Okinoerabujima Is. and Ishigakijima Is. We also
459 developed 16 datasets for the non-independent scenario (Supporting Information Table
460 S9).

461 For all four datasets in the scenario of independent origins between the regions,
462 speciation probability was significantly explained by island area ($P = 0.0016–0.0061$,

463 difference of deviance = 8.650–12.265, $r^2 = 0.407–0.577$; Table 2 and Supporting
464 Information Table S9; Figure 5 and S5). The relationship was also significant for 12 of
465 the 16 datasets in the non-independent scenario (Supporting Information Table S9).
466 Almost identical results were obtained for the maximum catchment area and river
467 length (Table 2 and Supporting Information Table S9). Furthermore, two of the four
468 datasets for the independent origin scenario and five of the 16 datasets for the
469 non-independent scenario, showed significant relationships between the number of
470 waterfalls and speciation (Table 2 and Supporting Information Table S9). Model
471 selection by AIC scores showed that the best model included no variables, except for
472 island area, as the predictors for speciation probability in most datasets when the
473 variables were controlled by the island area (Supporting Information Table S9).

474
475

476 **References**

- 477 Akihito, Sakamoto, K., Ikeda, Y., & Aizawa, M. (2013). Gobioidae. In T. Nakabo (Ed.),
478 *Fishes of Japan with Pictorial Keys to Species (Third ed.)*. (pp. 1374–1608). Tokyo,
479 Japan: Tokai University Press. (in Japanese).
- 480 Beaumont, M. A. (2010). Approximate Bayesian computation in evolution and ecology.
481 *Annual Review of Ecology, Evolution, and Systematics*, 41, 379–406.
482 <https://doi.org/10.1146/annurev-ecolsys-102209-144621>
- 483 Bertorelle, G., Benazzo, A., & Mona, S. (2010). ABC as a flexible framework to
484 estimate demography over space and time: some cons, many pros. *Molecular*
485 *Ecology*, 19, 2609–2625. <https://doi.org/10.1111/j.1365-294X.2010.04690.x>
- 486 Blacket, M. J., Robin, C., Good, R. T., Lee, S. F., & Miller, A. D. (2012). Universal
487 primers for fluorescent labelling of PCR fragments—an efficient and cost-effective
488 approach to genotyping by fluorescence. *Molecular Ecology Resources*, 12, 456–
489 463. <https://doi.org/10.1111/j.1755-0998.2011.03104.x>
- 490 Cornuet, J.-M., Ravigné, V., & Estoup, A. (2010). Inference on population history and
491 model checking using DNA sequence and microsatellite data with the software
492 DIYABC (v1.0). *BMC Bioinformatics*, 11, 401.
493 <https://doi.org/10.1186/1471-2105-11-401>

494 Csilléry, K., François, O., & Blum, M. G. B. (2012). abc: an R package for approximate
 495 Bayesian computation (ABC). *Methods in Ecology and Evolution*, 3, 475–479.
 496 <https://doi.org/10.1111/j.2041-210X.2011.00179.x>

497 Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and
 498 program for visualizing STRUCTURE output and implementing the Evanno
 499 method. *Conservation Genetics Resources*, 4, 359–361.
 500 <https://doi.org/10.1007/s12686-011-9548-7>

501 Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of
 502 individuals using the software STRUCTURE: a simulation study. *Molecular
 503 Ecology*, 14, 2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>

504 Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: a new series of
 505 programs to perform population genetics analyses under Linux and Windows.
 506 *Molecular Ecology Resources*, 10, 564–567.
 507 <https://doi.org/10.1111/j.1755-0998.2010.02847.x>

508 Excoffier, L., & Foll, M. (2011). fastsimcoal: a continuous-time coalescent simulator of
 509 genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics*,
 510 27, 1332–1334. <https://doi.org/10.1093/bioinformatics/btr124>

511 Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013).
 512 Robust Demographic Inference from Genomic and SNP Data. *PLoS Genetics*, 9,
 513 e1003905–17. <https://doi.org/10.1371/journal.pgen.1003905>

514 Goudet, J. (1995). Fstat version 1.2: a computation program to calculate Fstatistics.
 515 *Journal of Heredity*, 86, 485–486.
 516 <https://doi.org/10.1093/oxfordjournals.jhered.a111627>

517 Hirashima, K., & Tachihara, K. (2000). Embryonic development and morphological
 518 changes in larvae and juveniles of two land-locked gobies, *Rhinogobius* spp.
 519 (Gobiidae), on Okinawa Island. *Japanese Journal of Ichthyology*, 47, 29–41 (in
 520 Japanese with English abstract). <https://doi.org/10.11369/jji1950.47.29>

521 Holleley, C., & Geerts, P. (2009). Multiplex Manager 1.0: a cross-platform computer
 522 program that plans and optimizes multiplex PCR. *Biotechniques*, 46, 511–517.
 523 <https://doi.org/10.2144/000113156>

524 Iguchi, K., & Mizuno, N. (1999). Early starvation limits survival in amphidromous
 525 fishes. *Journal of Fish Biology*, 54, 705–712.
 526 <https://doi.org/10.1111/j.1095-8649.1999.tb02027.x>

- 527 Kano, Y., Nishida, S., & Nakajima, J. (2012). Waterfalls drive parallel evolution in a
528 freshwater goby. *Ecology and Evolution*, 2, 1805–1817.
529 <https://doi.org/10.1002/ece3.295>
- 530 Kon, T., & Yoshino, T. (2003). Coloration and ontogenetic features of fluviatile species
531 of *Rhinogobius* (Gobioidei: Gobiidae in Amami-oshima Island, Ryukyu Islands,
532 Japan. *Ichthyological Research*, 50, 109–116.
533 <https://doi.org/10.1007/s10228-002-0145-3>
- 534 Kondo, M., Maeda, K., Hirashima, K., & Tachihara, K. (2013). Comparative larval
535 development of three amphidromous *Rhinogobius* species, making reference to
536 their habitat preferences and migration biology. *Marine and Freshwater Research*,
537 64, 249–266. <https://doi.org/10.1071/MF12234>
- 538 Lombert, E., Guillemaud, T., Lundgren, J., Koch, R., Facon, B., Grez, A., ...Estoup, A.
539 (2014). Complementarity of statistical treatments to reconstruct worldwide routes
540 of invasion: the case of the Asian ladybird *Harmonia axyridis*. *Molecular Ecology*,
541 23, 5979–5997. <https://doi.org/10.1111/mec.12989>
- 542 Machida, H., Ota, Y., Kawana, T., Moriwaki, H., & Nagaoka, N. (2001). *Regional*
543 *Geomorphology of the Japanese Island vol. 7 Kyushu and Southern Islands*. Tokyo,
544 Japan: University of Tokyo Press. (in Japanese)
- 545 Matschiner, M., & Salzburger, W. (2009). TANDEM: integrating automated allele
546 binning into genetics and genomics workflows. *Bioinformatics* 25:1982–1983.
547 <https://doi.org/10.1093/bioinformatics/btp303>
- 548 Meglec, E., Costedoat, C., Dubut, V., Gilles, A., Malausa, T., Pech, N., & Martin, J. F.
549 (2010). QDD: a user-friendly program to select microsatellite markers and design
550 primers from large sequencing projects. *Bioinformatics*, 26, 403–404.
551 <https://doi.org/10.1093/bioinformatics/btp670>
- 552 Meirmans, P. G., & van Tienderen, P. H. (2004). genotype and genodive: two programs
553 for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*,
554 4, 792–794. <https://doi.org/10.1111/j.1471-8286.2004.00770.x>
- 555 Mizuno, N. (2001). *Rhinogobius*. In H. Kawanabe, N. Mizuno, & K. Hosoya (Eds.),
556 *Freshwater fishes of Japan* (Third ed.). (p. 584) Tokyo, Japan: Yama-Kei
557 Publishers (in Japanese).

- 558 Moriyama, A., Yanagisawa, Y., Mizuno, N., & Omori, K. (1998). Starvation of drifting
559 goby larvae due to retention of free embryos in upstream reaches. *Environmental*
560 *Biology of Fishes*, 52, 321–329. <https://doi.org/10.1023/A:1007333302864>
- 561 Ohara, K., Takahashi, D., & Takagi, M. (2004). Isolation of microsatellite loci in the
562 freshwater goby, *Rhinogobius* sp. (Gobiidae). *Molecular Ecology Notes*, 4, 449–
563 451. <https://doi.org/10.1111/j.1471-8286.2004.00677.x>
- 564 Oijen, M. J. P., Suzuki, T., & Chen, I. S. (2011). On the earliest published species of
565 *Rhinogobius*. With a redescription of *Gobius brunneus* Temminck and Schlegel,
566 1845. *Journal of the National Taiwan Museum*, 64, 1–17.
- 567 Omori, K., Ohnishi, H., Hamaoka, H., Kunihiro, T., Ito, S., Kuwae, M., ... Iguchi K.
568 (2012). Speciation of fluvial forms from amphidromous forms of migratory
569 populations. *Ecological Modelling*, 243, 89–94.
570 <https://doi.org/10.1016/j.ecolmodel.2012.06.006>
- 571 Oshiro, N. (1981). Ryukyu rettousan Yoshinobori no yomoyamabanashi (Some topics
572 about *Rhinogobius* fishes in Ryukyu Island). *Tansuigyo*, 7, 14–18 (in Japanese).
- 573 Osozawa, S., Shinjo, R., & Armid, A. (2012). Palaeogeographic reconstruction of the
574 1.55 Ma synchronous isolation of the Ryukyu Islands, Japan, and Taiwan and
575 inflow of the Kuroshio warm current. *International Geology Review*, 54, 1369–
576 1388. <https://doi.org/10.1080/00206814.2011.639954>
- 577 Ota, H. (1998). Geographic patterns of endemism and speciation in amphibians and
578 reptiles of the Ryukyu Archipelago, Japan, with special reference to their
579 paleogeographical implications. *Researches on Population Ecology*, 40, 189–204.
580 <https://doi.org/10.1007/BF02763404>
- 581 Pritchard, J. K., Stephen, M., & Donnelly, P. (2000). Inference of population genetic
582 structure using multilocus genotype data. *Genetics*, 155, 945–959.
- 583 Shinomiya, A., Sasabe, K., Sakurai, M., & Kishino, T. (2005). Morphological
584 characters of newly hatched larvae and their habitat selection of a fluvial goby,
585 *Rhinogobius* sp. YB in the upper stream of Sumiyo River, Amami-oshima Island,
586 Japan. *Japanese Journal of Ichthyology*, 52, 1–8 (in Japanese with English abstract).
587 <https://doi.org/10.11369/jji1950.52.1>
- 588 Suzuki, T., Shibukawa, K., Yano, K., & Senou, H. (2004). A photographic guide to the
589 gobioid fishes in Japan. Tokyo, Japan: Heibonsha (in Japanese).
- 590 Takeshima, H., Muto, N., Sakai, Y., Ishiguro, N., Iguchi, K., Ishikawa, S., & Nishida, M.

591 (2017). Rapid and effective isolation of candidate sequences for development of
592 microsatellite markers in 30 fish species by using kit-based target capture and
593 multiplexed parallel sequencing. *Conservation Genetics Resources*, 9, 479–490.
594 <https://doi.org/10.1007/s12686-017-0730-4>

595 Takezaki, N., Nei, M., & Tamura, K. (2010). POPTREE2: Software for constructing
596 population trees from allele frequency data and computing other population
597 statistics with Windows interface. *Molecular Biology and Evolution*, 27, 747–752.
598 <https://doi.org/10.1093/molbev/msp312>

599 Tamada, K. (2005). Clutch and egg size in *Rhinogobius* sp. DA inhabiting small rivers.
600 *Japanese Journal of Ichthyology*, 52, 17–20 (in Japanese with English abstract).
601 <https://doi.org/10.11369/jji1950.52.17>

602 Vähä, J.-P., & Primmer, C. R. (2005). Efficiency of model-based Bayesian methods for
603 detecting hybrid individuals under different hybridization scenarios and with
604 different numbers of loci. *Molecular Ecology*, 15, 63–72.
605 <https://doi.org/10.1111/j.1365-294X.2005.02773.x>

606 van Oosterhout, C., Hutchinson, W. F., Willis, D. P. M., & Shipley, P. (2004). MICRO-
607 CHECKER; software for identifying and correcting genotyping errors in
608 microsatellite data. *Molecular Ecology Notes*, 4, 535–538.
609 <https://doi.org/10.1111/j.1471-8286.2004.00684.x>

610 Wegmann, D., Leuenberger, C., Neuenschwander, S., & Excoffier, L. (2010).
611 ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC*
612 *Bioinformatics*, 11, 116. <https://doi.org/10.1186/1471-2105-11-116>

613 Yamasaki, Y. Y., Nishida, M., Suzuki, T., Mukai, T., & Watanabe, K. (2015).
614 Phylogeny, hybridization, and life history evolution of *Rhinogobius* gobies in Japan,
615 inferred from multiple nuclear gene sequences. *Molecular Phylogenetics and*
616 *Evolution*, 90, 20–33. <https://doi.org/10.1016/j.ympev.2015.04.012>

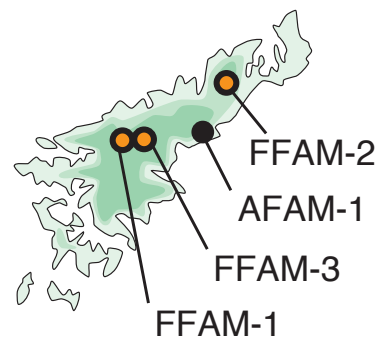
617 Yoshigou, H. (2014). Annotated checklist and bibliographic records of inland water
618 fishes of the Ryukyu Archipelago, Japan. *Fauna Ryukyana*, 9, 1–153.
619 http://w3.u-ryukyu.ac.jp/naruse/lab/Fauna_Ryukyuna.html

620 Yue, G. H., David, L., & Orban, L. (2007). Mutation rate and pattern of microsatellites
621 in common carp (*Cyprinus carpio* L.). *Genetica*, 129, 329–331
622 <https://doi.org/10.1007/s10709-006-0003-8>

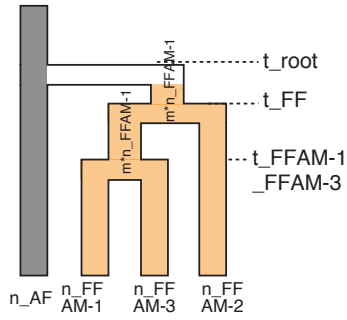
623

A

2. Amami-Oshima Is.

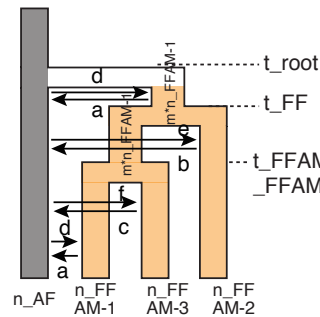


Single origin
No gene flow



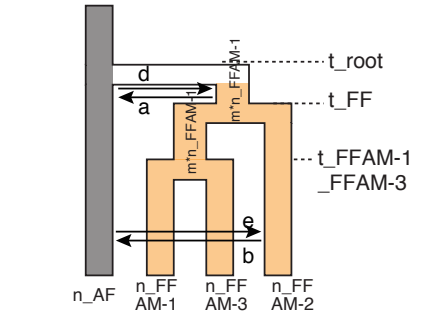
Model 3 (close relationship of FFAM-1&-3)

Single origin
Gene flow

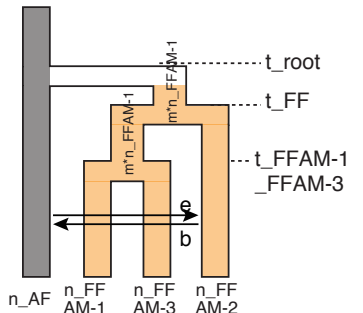


Model 4 (close relationship of FFAM-1&-3)

Different gene flow patterns

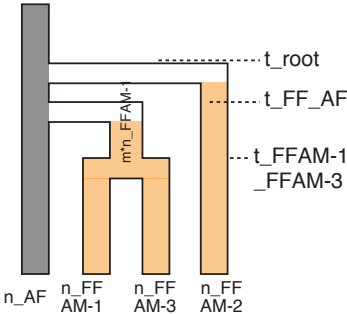


Model 6 (no gene flow for FFAM-1&-3)



Model 7 (no gene flow for FFAM-1&-3 and ancestor of FF)

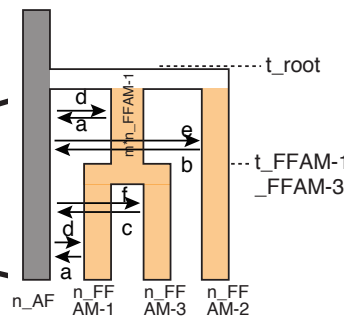
Parallel origin
No gene flow



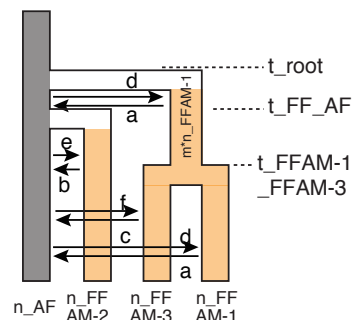
Model 8 (old divergence of FFAM-2)

Parallel origin
Gene flow

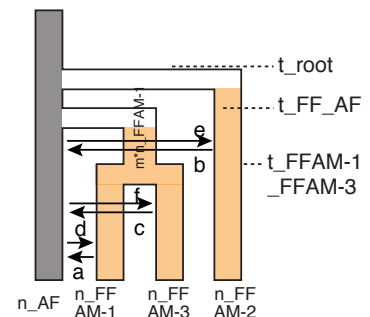
Different divergence order



Model 5 (simultaneous divergence)



Model 1 (recent divergence of FFAM-2)



Model 2 (old divergence of FFAM-2)

Figure S1 Tested demographic models and their relationships.

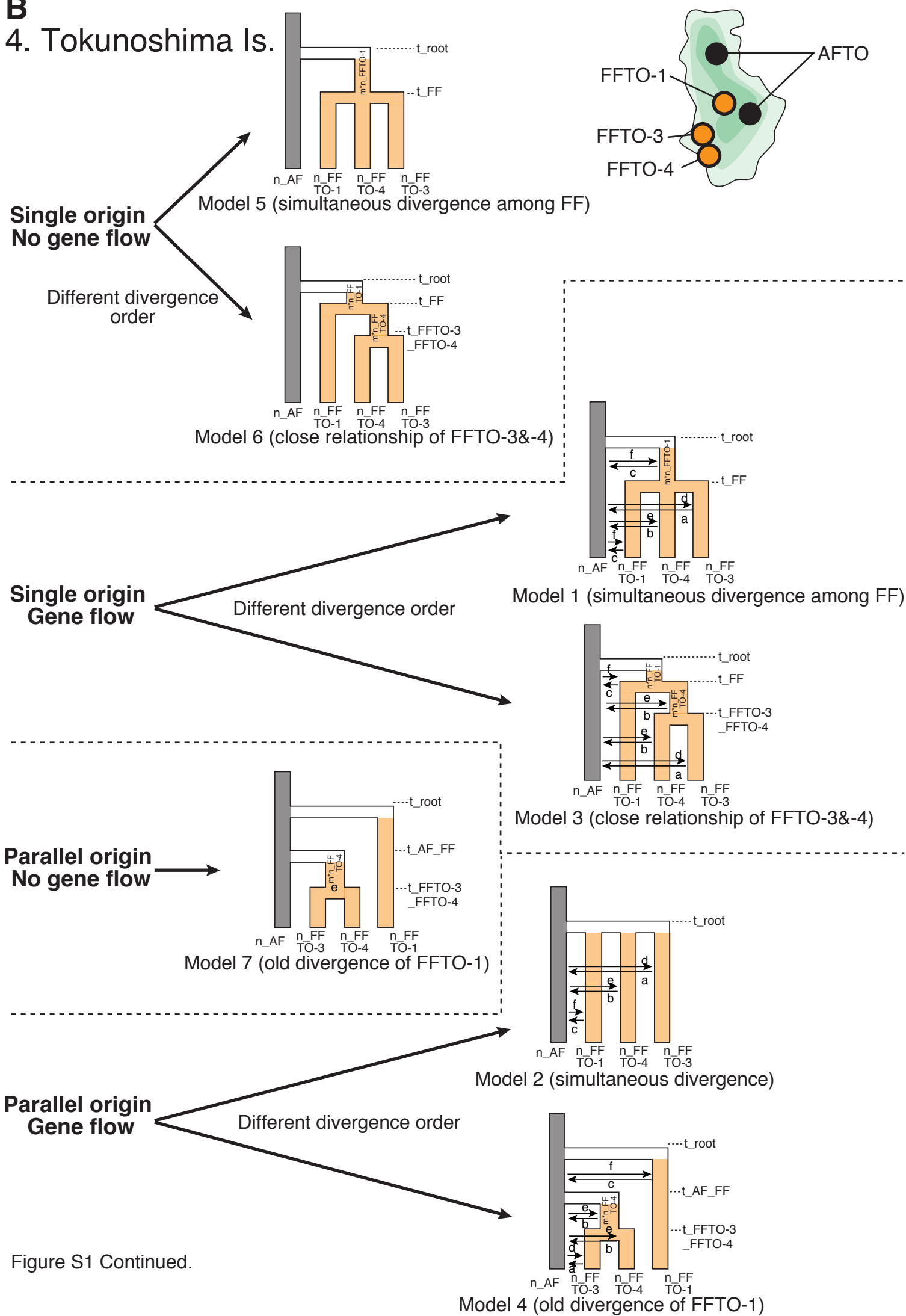
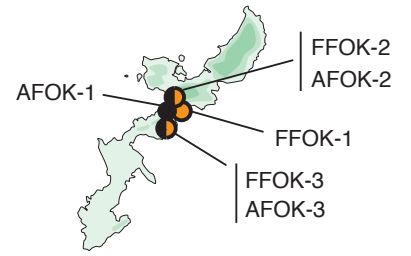
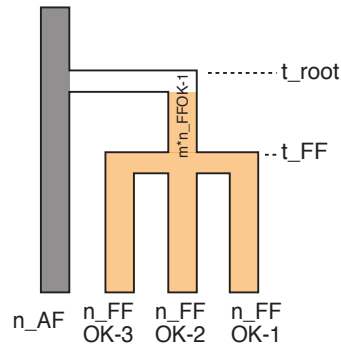
B**4. Tokunoshima Is.**

Figure S1 Continued.

C
6. Okinawajima Is.

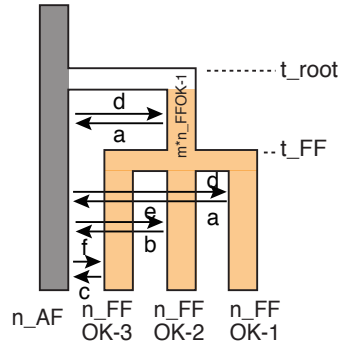


Single origin
No gene flow



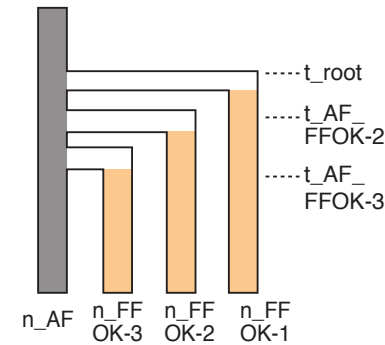
Model 5 (simultaneous divergence among FF)

Single origin
Gene flow



Model 1 (simultaneous divergence among FF)

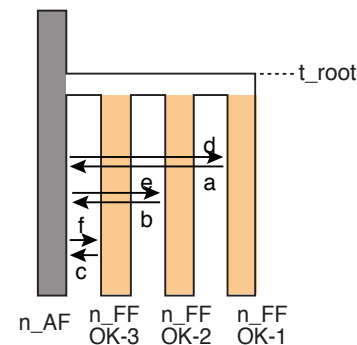
Parallel origin
No gene flow



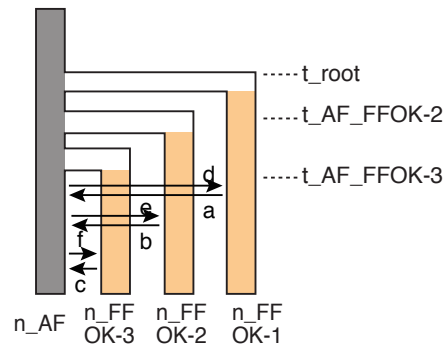
Model 3 (respective origins of FF)

Parallel origin
Gene flow

Different divergence order

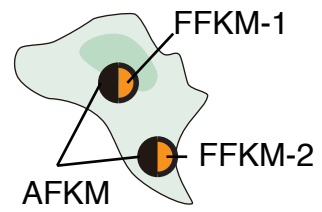


Model 2 (simultaneous divergence)

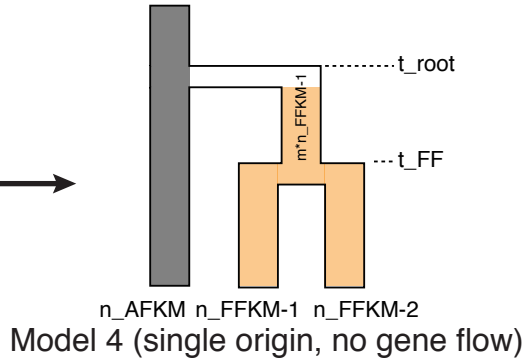


Model 4 (respective origins of FF)

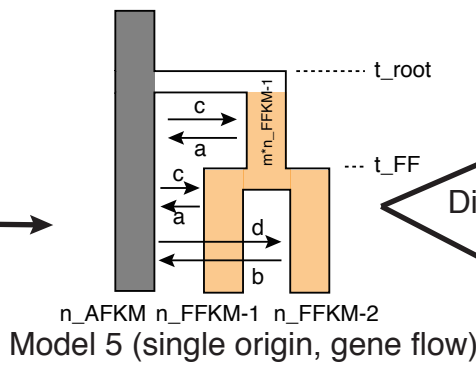
D
7. Kumejima Is.



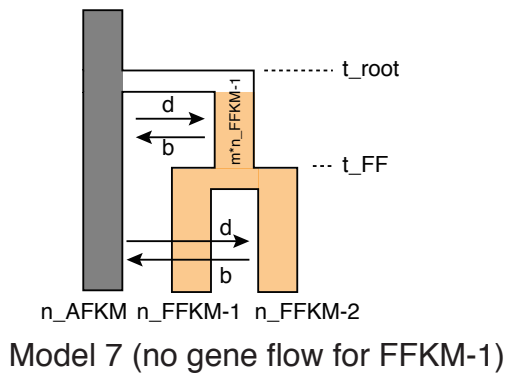
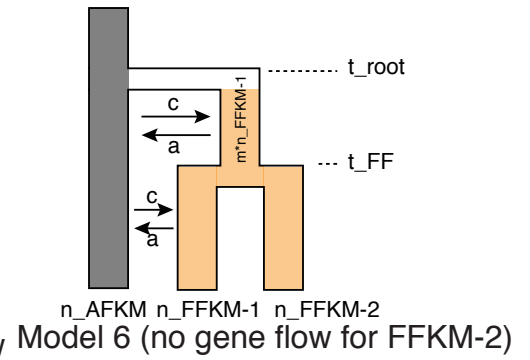
**Single origin
No gene flow**



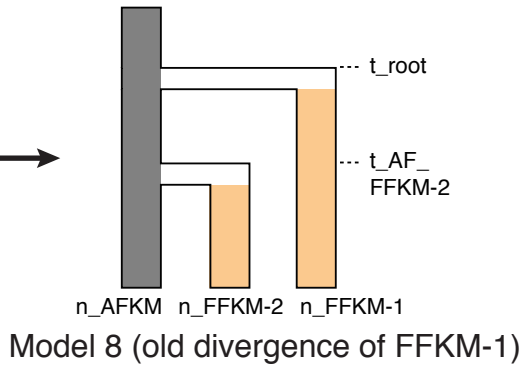
**Single origin
Gene flow**



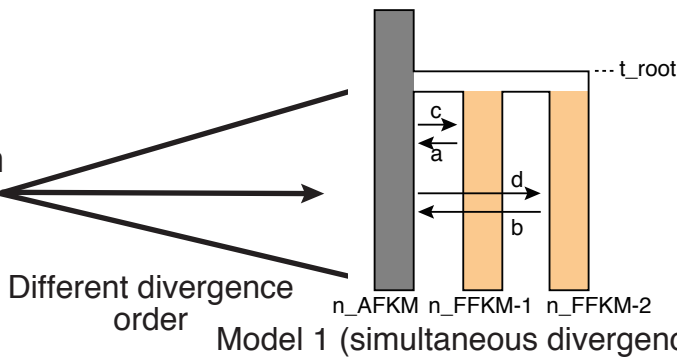
Different gene flow patterns



**Parallel origin
No gene flow**

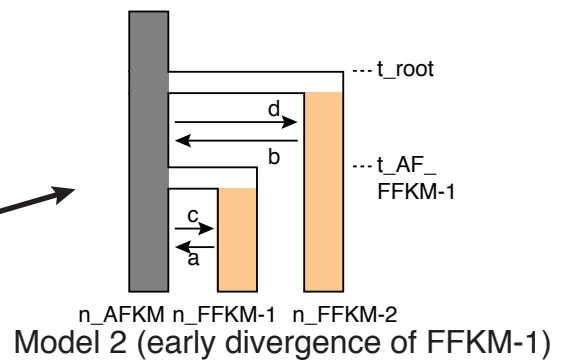


**Parallel origin
Gene flow**

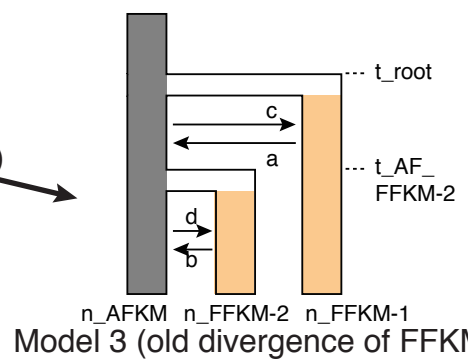


Different divergence order

Model 1 (simultaneous divergence)



Model 2 (early divergence of FFKM-1)

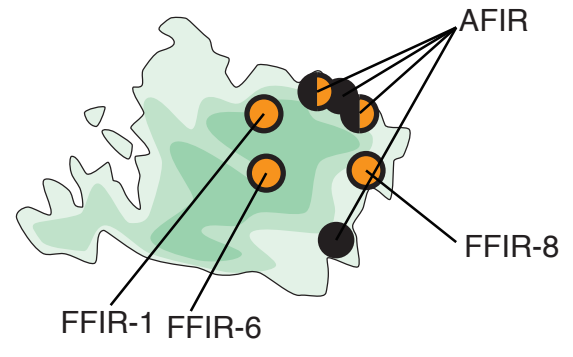


Model 3 (old divergence of FFKM-1)

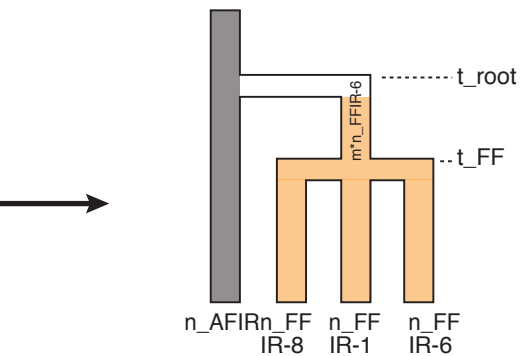
Figure S1 Continued.

E

9. Iriomotejima Is.

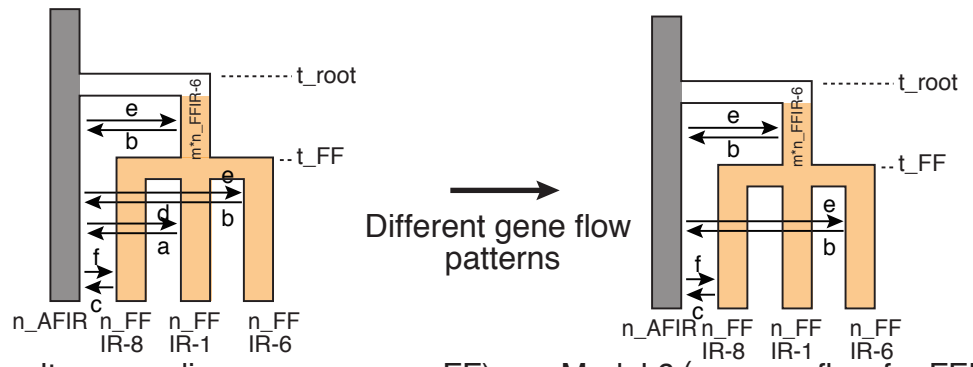


Single origin
No gene flow



Model 1 (simultaneous divergence among FF)

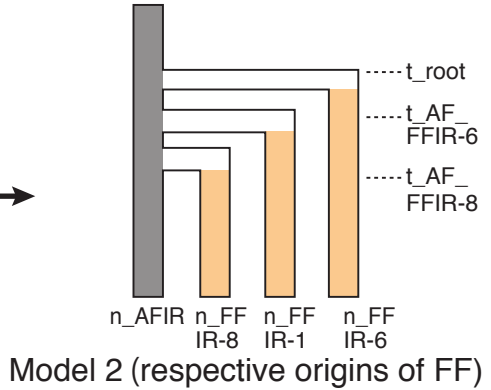
Single origin
Gene flow



Model 3 (simultaneous divergence among FF)

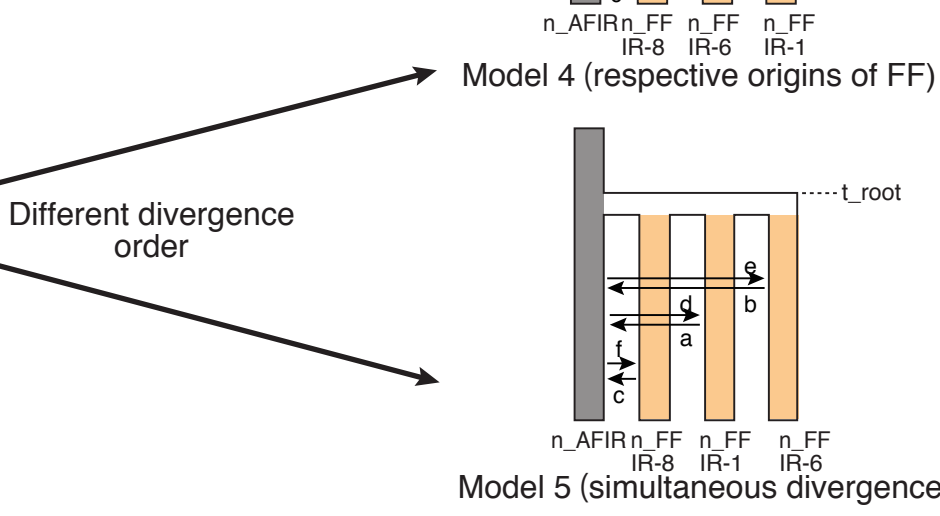
Model 6 (no gene flow for FFIR-1)

Parallel origin
No gene flow



Model 2 (respective origins of FF)

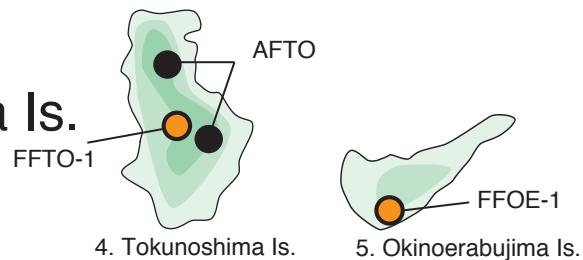
Parallel origin
Gene flow



Model 4 (respective origins of FF)

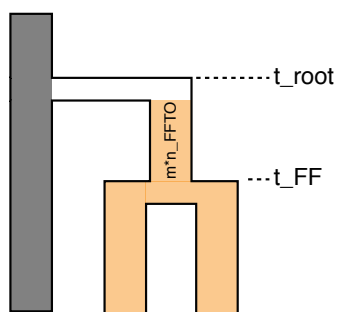
Model 5 (simultaneous divergence)

F
4. Tokunoshima Is.– 5. Okinoerabujima Is.

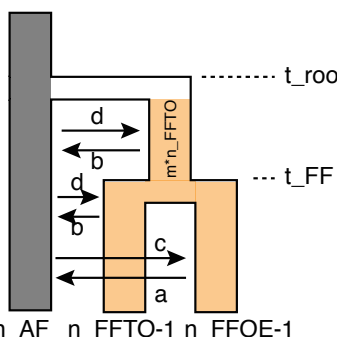


Single origin

Different gene flow patterns



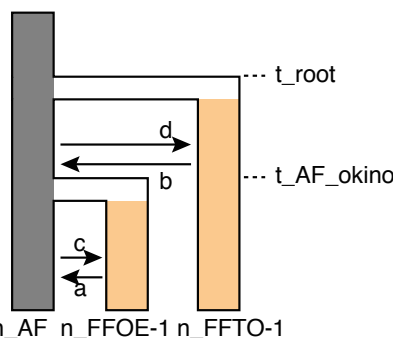
Model 1 (single origin, no gene flow)



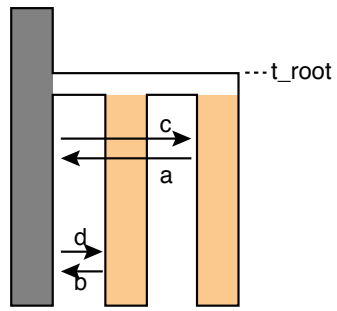
Model 2 (single origin, gene flow)

Parallel origin

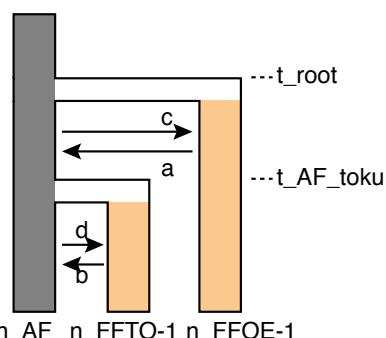
Different divergence order



Model 3 (old divergence of FFTO-1)

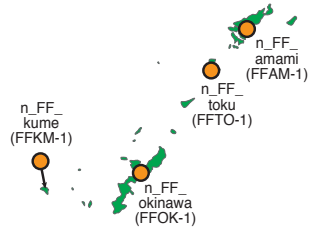


Model 4 (simultaneous divergence)

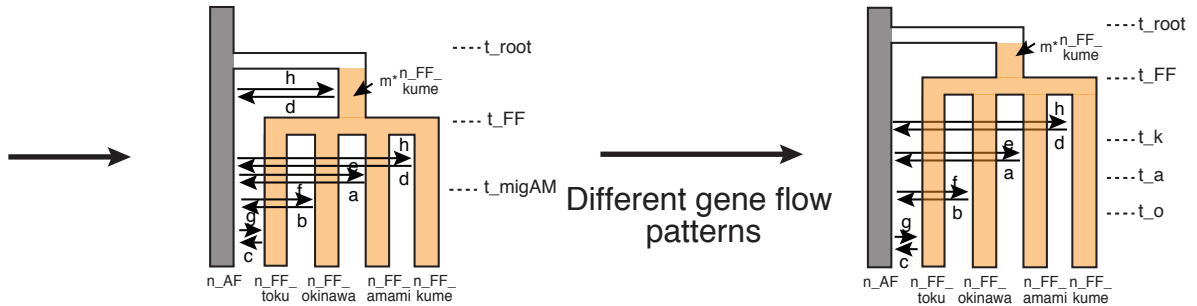


Model 5 (recent divergence of FFTO-1)

G Middle Ryukyu



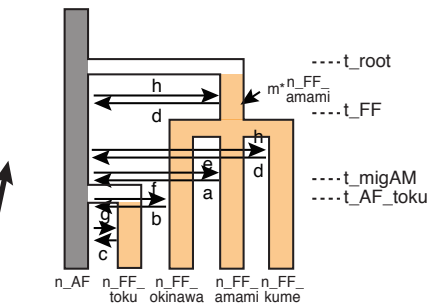
Single origin



Model 3 (simultaneous divergence among FF)

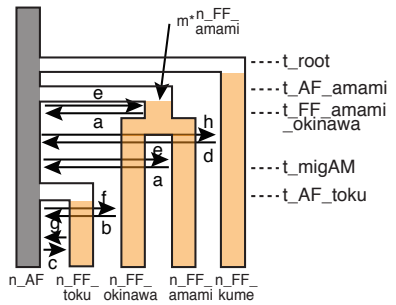
Model 7 (different gene flow timing among FF)

Two origins of FF



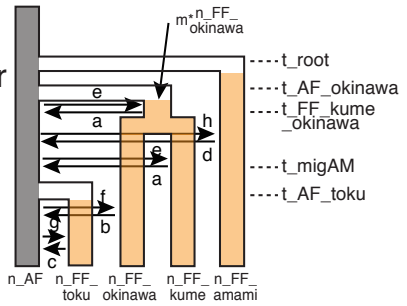
Model 1 (recent divergence of FF_toku and shared origin of others)

Three origins of FF



Model 5 (recent divergence of FF_toku and old divergence of FF_kume)

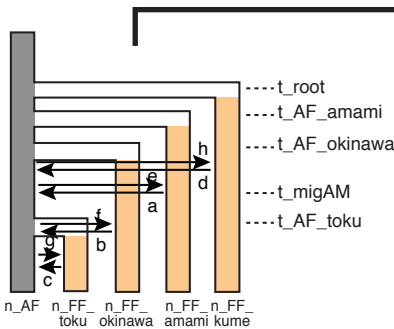
Different divergence order



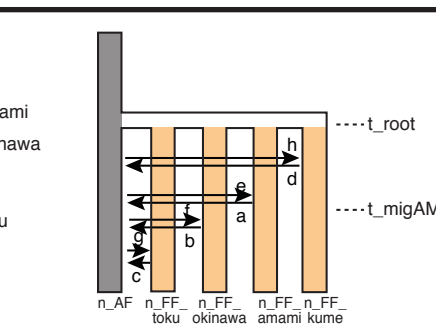
Model 10 (recent divergence of FF_toku and old divergence of FF_amami)

Parallel origin

Four origins of FF

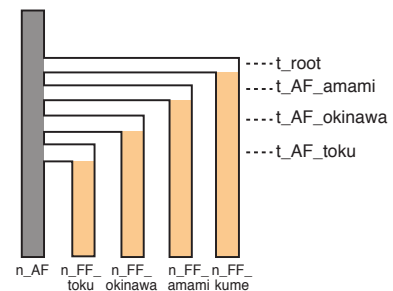


Model 2 (old divergence of FF_kume)



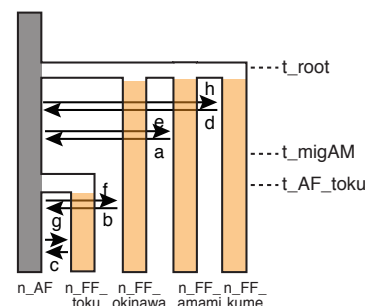
Model 4 (simultaneous divergence)

Different gene flow patterns

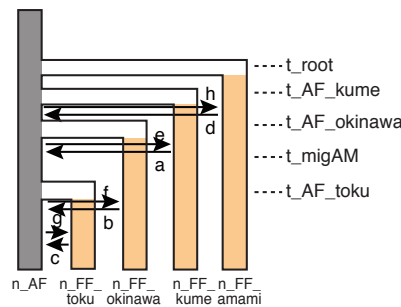


Model 8 (respective origins with no gene flow)

Different divergence order



Model 6 (simultaneous divergence of the three pops)

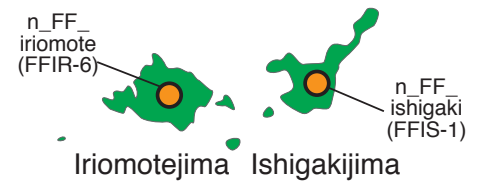


Model 9 (old divergence of FF_amami)

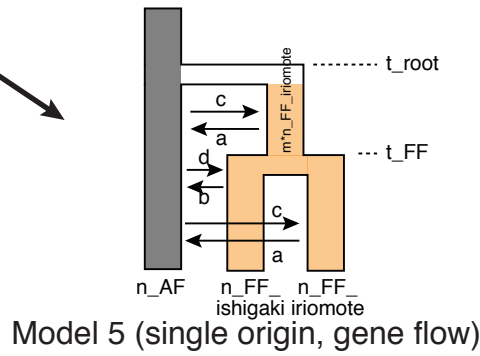
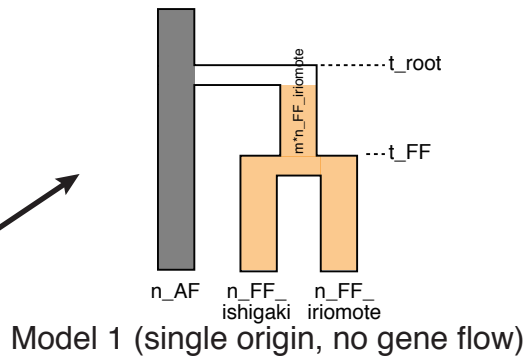
Figure S1 Continued.

H

South Ryukyu



Single origin



Parallel origin

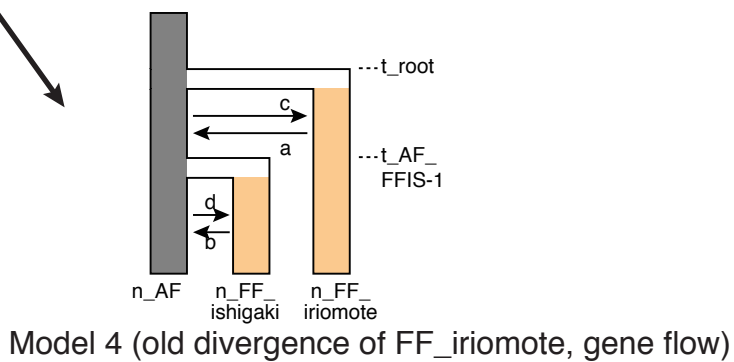
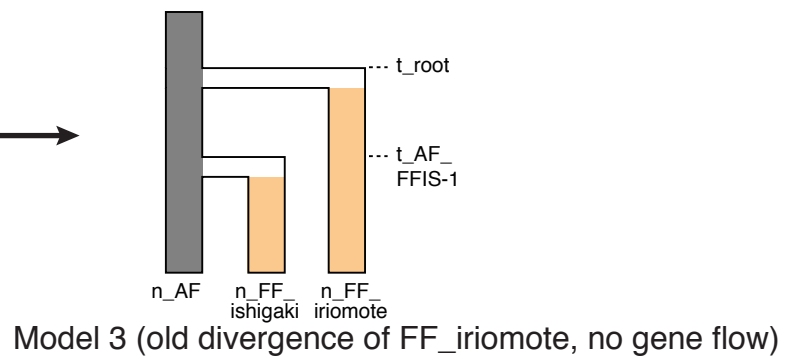
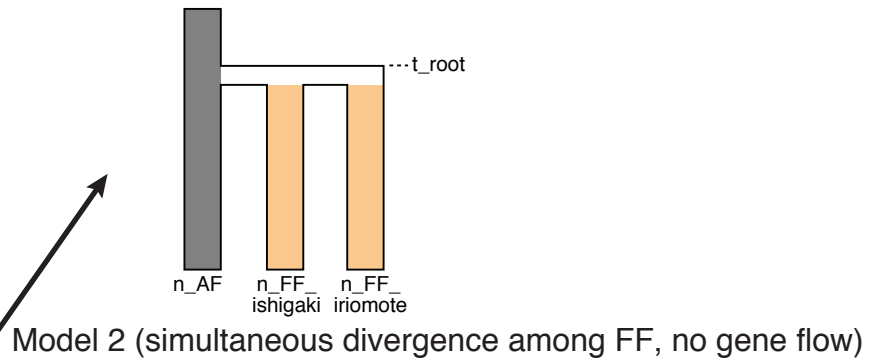


Figure S1 Continued.

Ryukyu

Single origin

Model 2 (shared origin between FFAM & FFIR, old divergence between AFs)

Model 6 (shared origin between FFAM & FFIR, recent divergence between AFs)

Different divergence order

Model 7 (shared origin between FFKM & FFIR, recent divergence between AFs)

Model 10 (no gene flow between AFs)

Different gene flow patterns

Model 1 (old divergence of AFs)

Model 3 (older divergence of FFIR & FFAM than that of AFs)

Parallel origin

Different divergence order

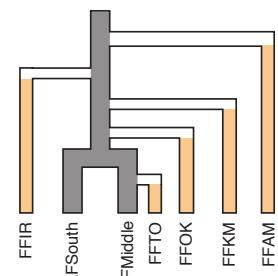
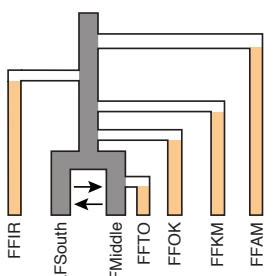
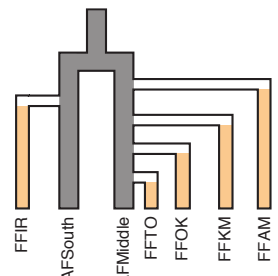
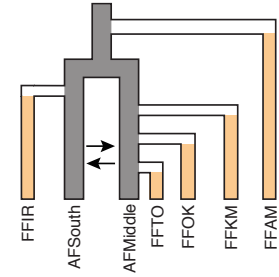
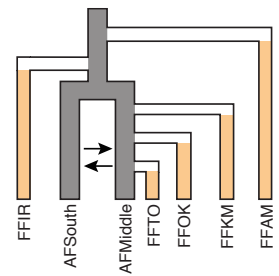
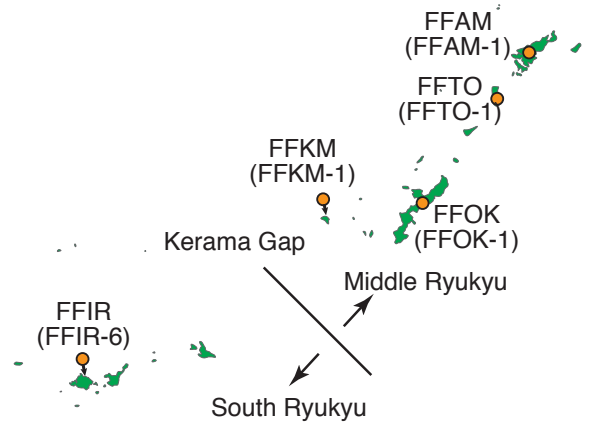
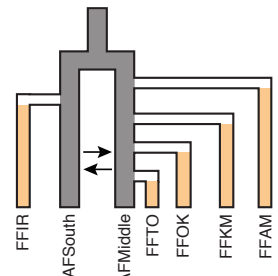
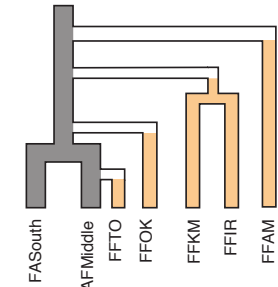
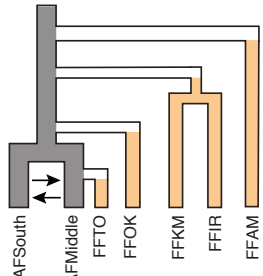
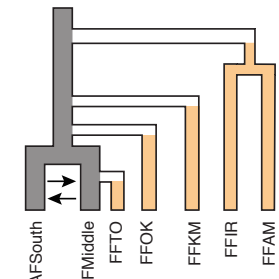
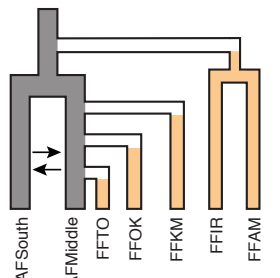
Model 4 (older divergence of FFAM than that of AFs)

Model 8 (no gene flow between AFs)

Figure S1 Continued.

Model 5 (recent divergence of AFs)

Model 9 (no gene flow between AFs)



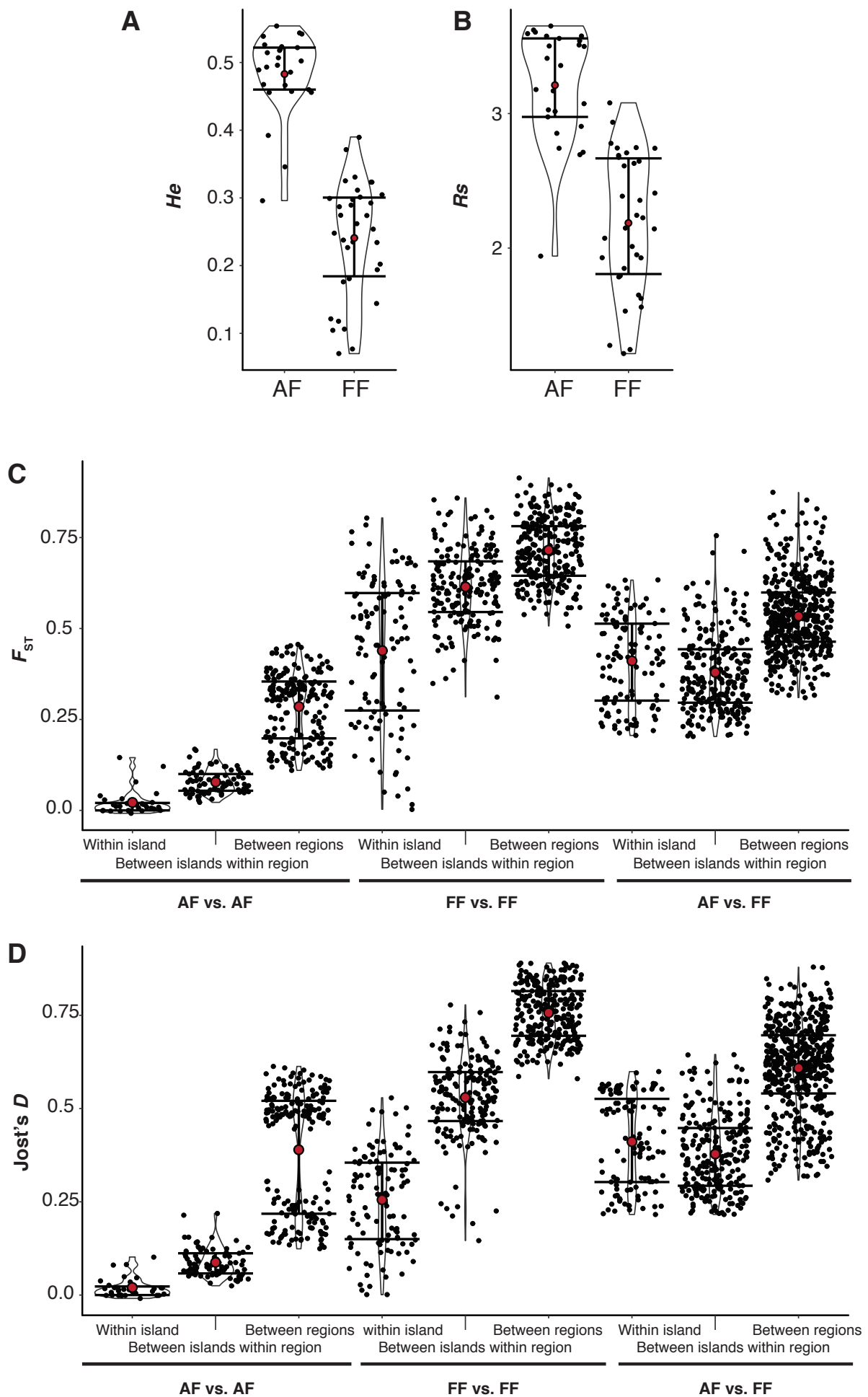


Figure S2 Distributions of summary statistics for genetic diversities within and between populations of the *Rhinogobius* species. Red dots and error bars indicate mean values and intermedian quantiles ranges, respectively. Genetic diversity values of populations were measured by expected heterozygosity (H_e , A) and allelic richness (R_s , B). Pairwise population differences were measured by F_{ST} (C) and Jost's D (D).

A**2. Amami-Oshima Is. & 3. Kikaijima Is.**

$$\Delta K = \frac{\text{mean}(L'(K))}{\text{sd}(L(K))}$$

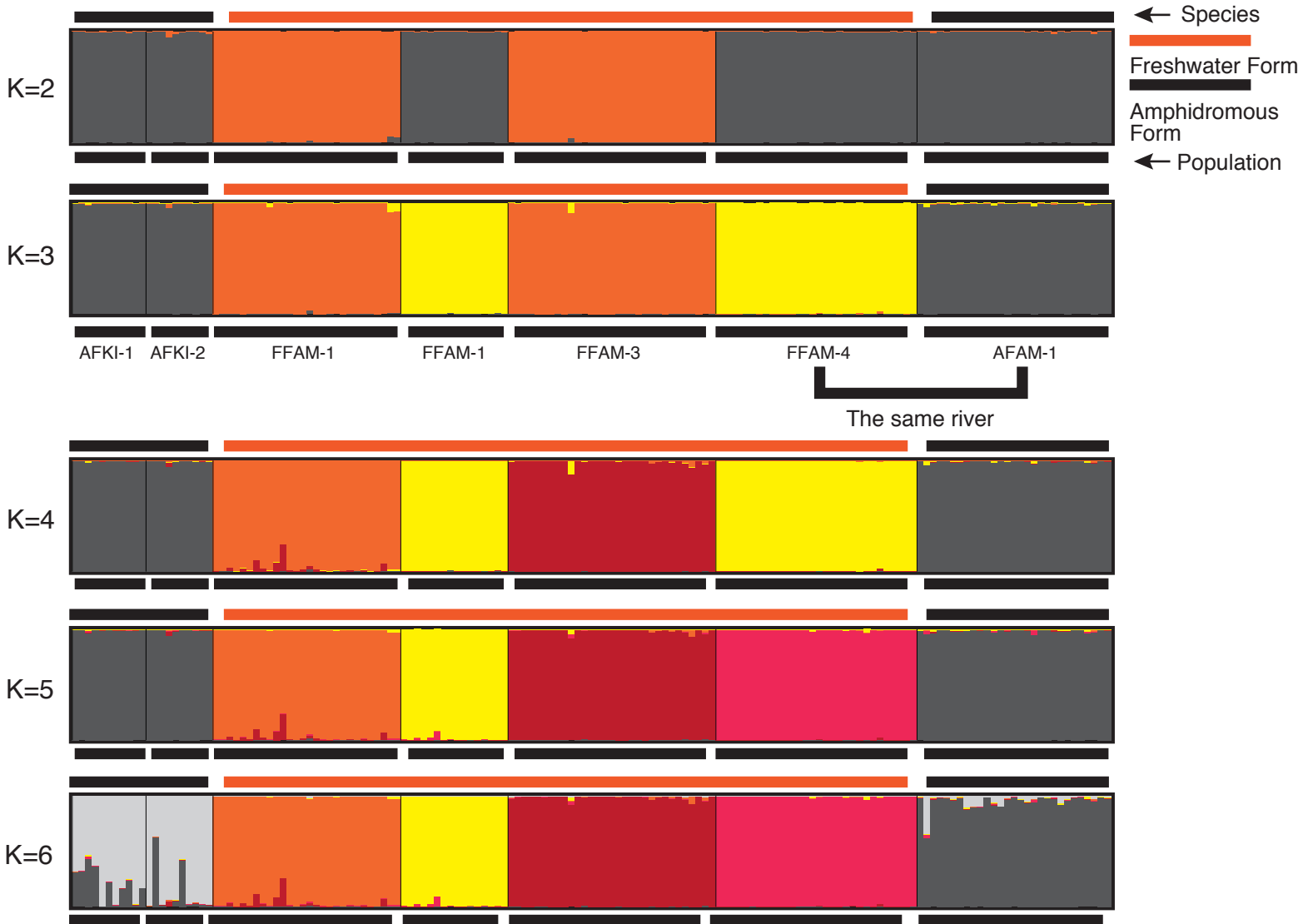
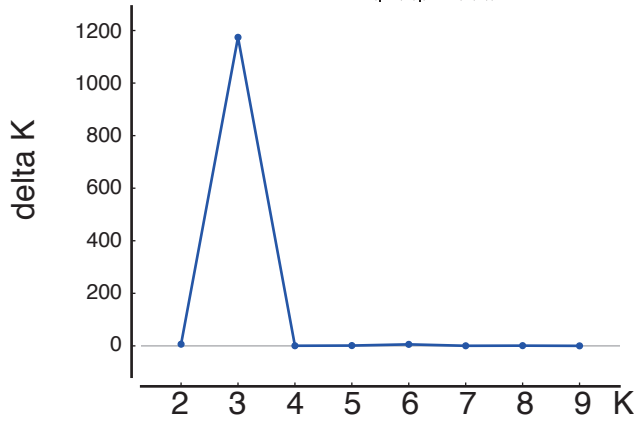


Figure S3 Results of STRUCTURE analysis for each island and AF. Sympatric points are indicated by additional annotation. Dominant genetic features in AF identified by morphology are shown in grey, whereas those in FF are shown in the other colours. Graphs of ΔK are also shown.

B

4. Tokunoshima Is.

$$\Delta K = \frac{\text{mean}(L'(K))}{\text{sd}(L(K))}$$

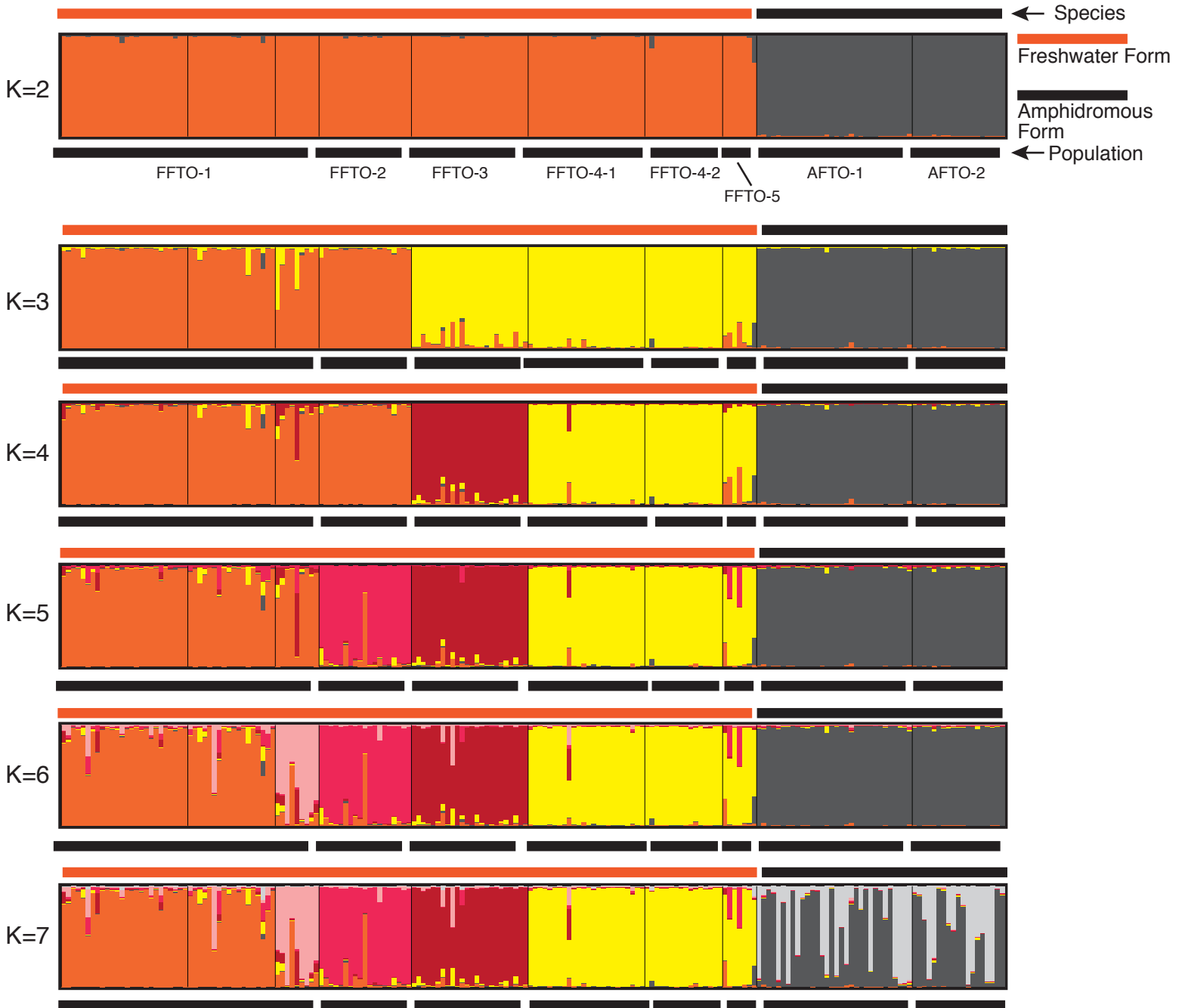
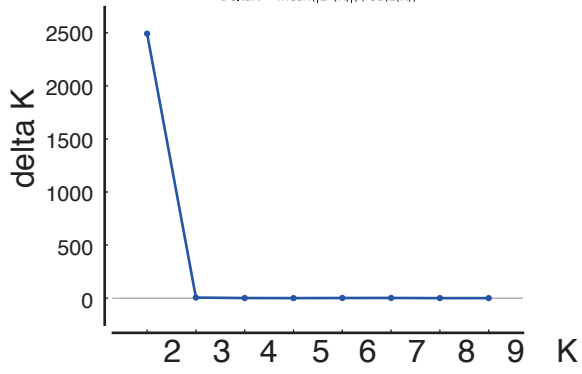


Figure S3 Continued.

C

5. Okinoerabujima Is. (but specimens of *R. brunneus* collected in Tokunoshima Is. were used)

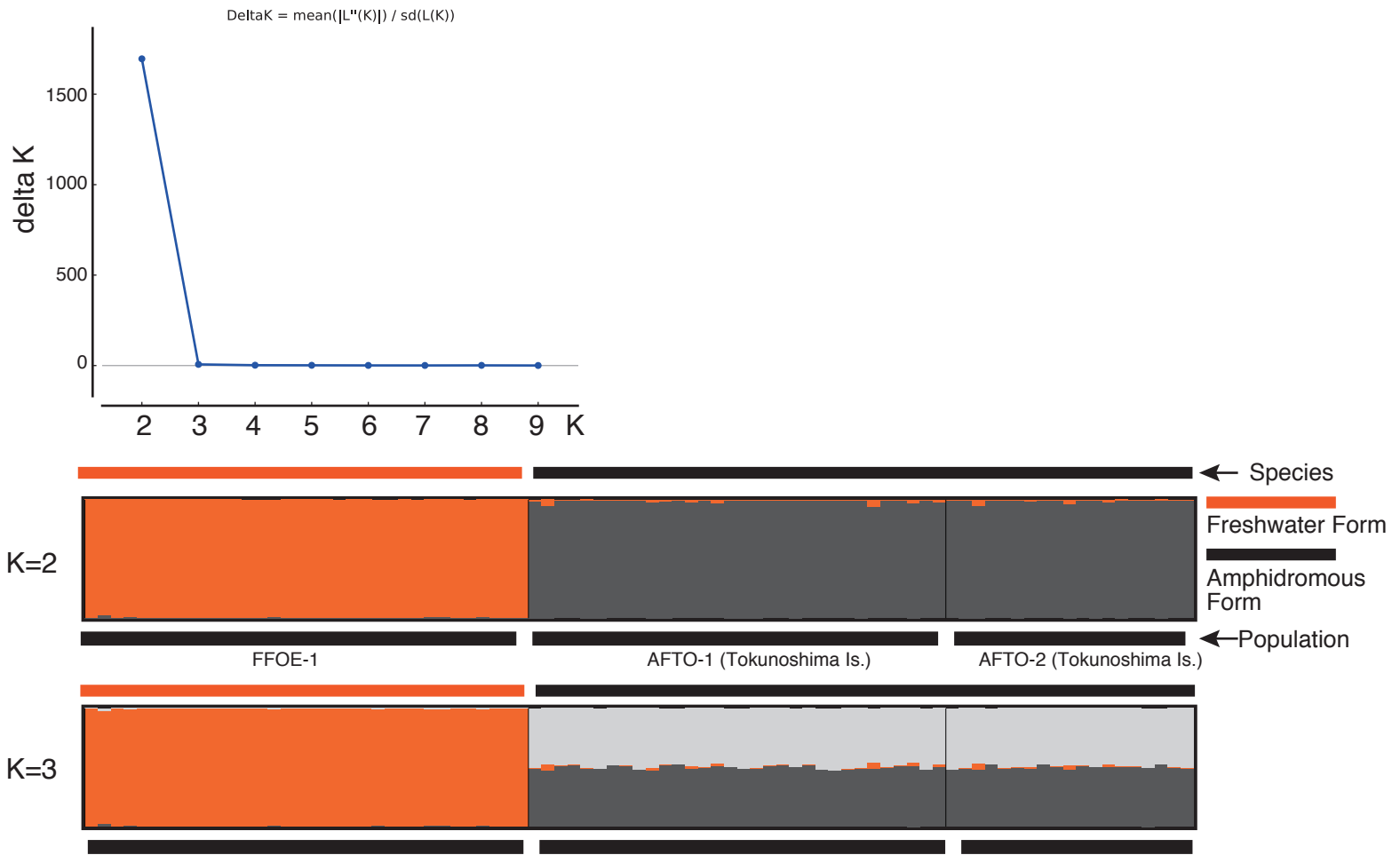


Figure S3 Continued.

D
6. Okinawajima Is.

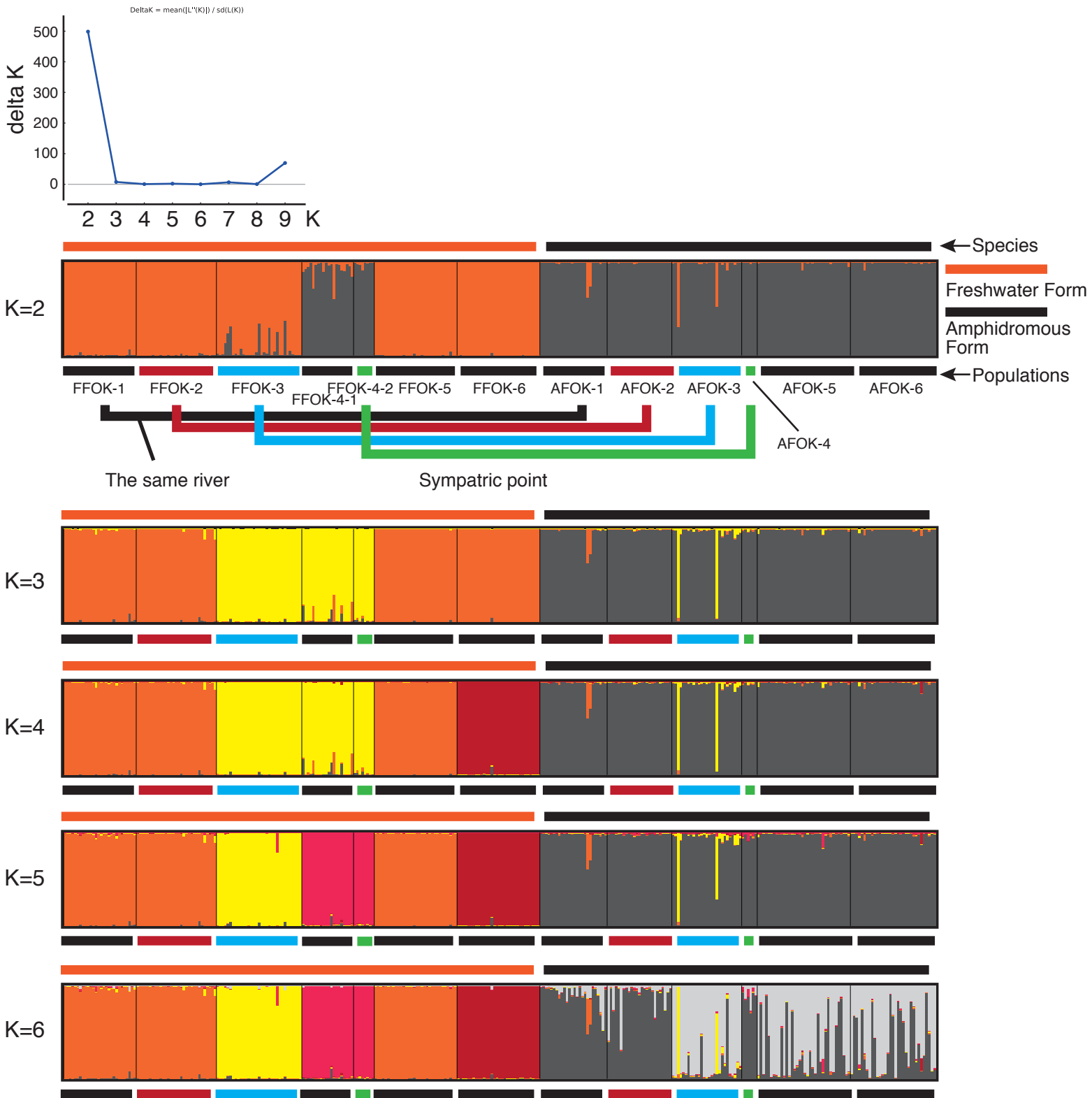


Figure S3 Continued.

E
7. Kumejima Is.

$$\Delta K = \frac{\text{mean}\{L(K)\}}{\text{sd}\{L(K)\}}$$

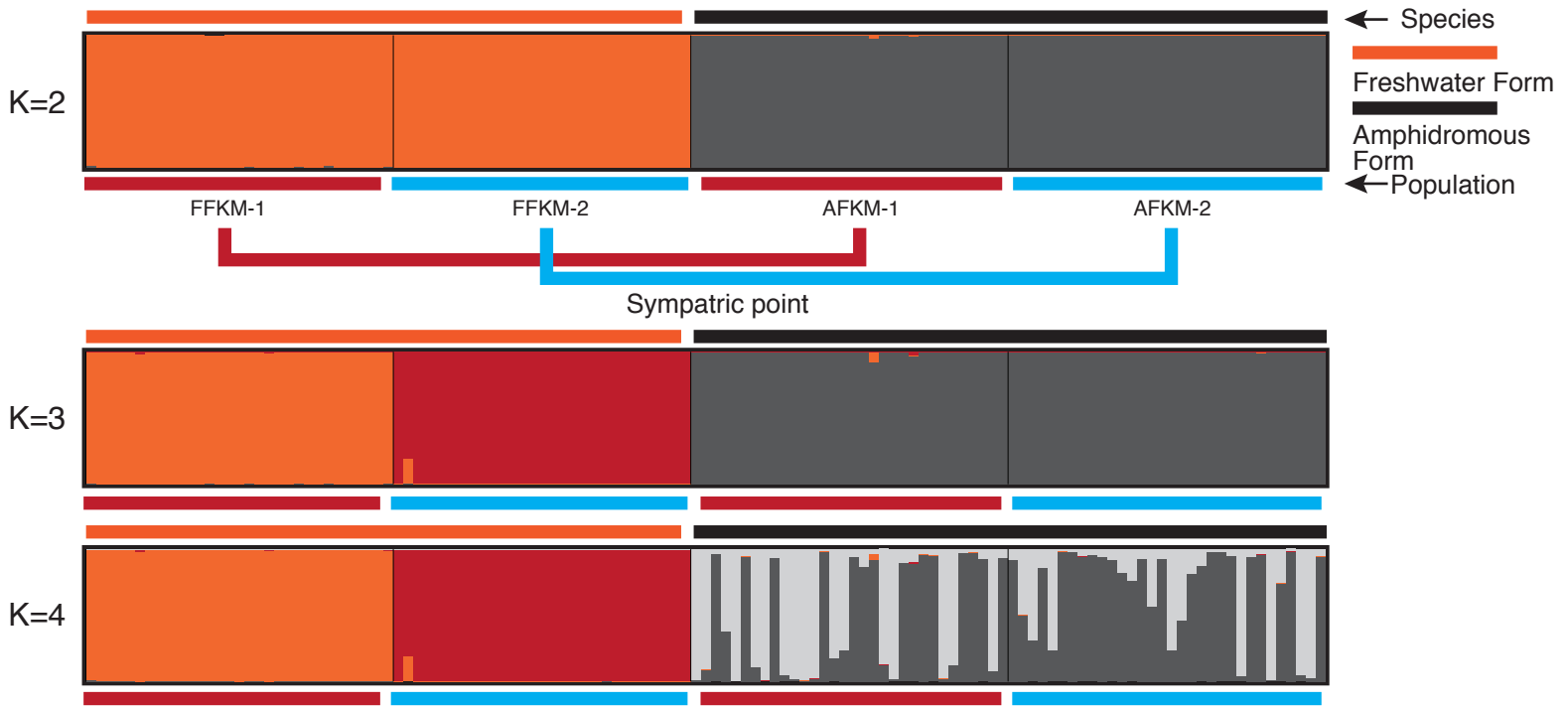
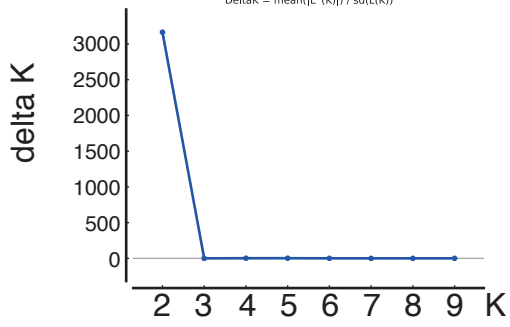


Figure S3 Continued.

F

8. Ishigakijima Is.

$$\text{DeltaK} = \text{mean}(|L''(K)|) / \text{sd}(L(K))$$

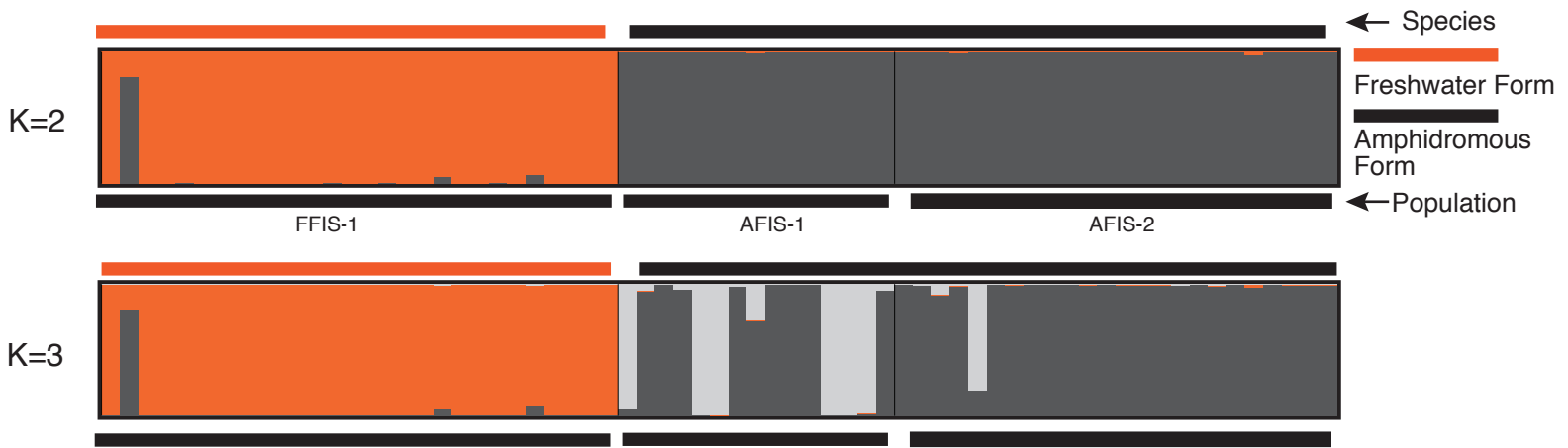
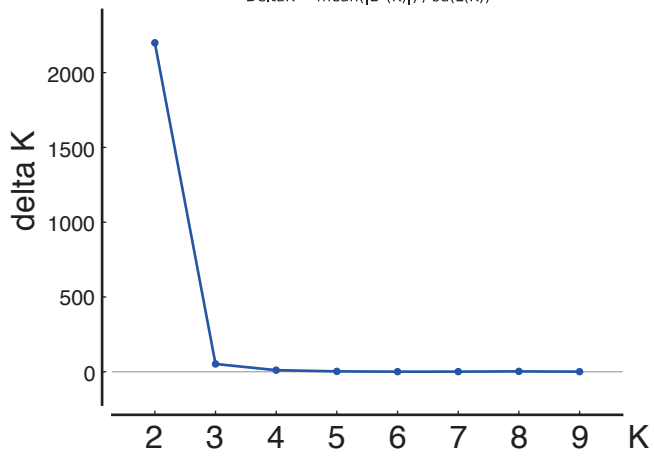


Figure S3 Continued.

G
9. Iriomotejima Is.

DeltaK = mean(|L''(K)|) / S(L(K))

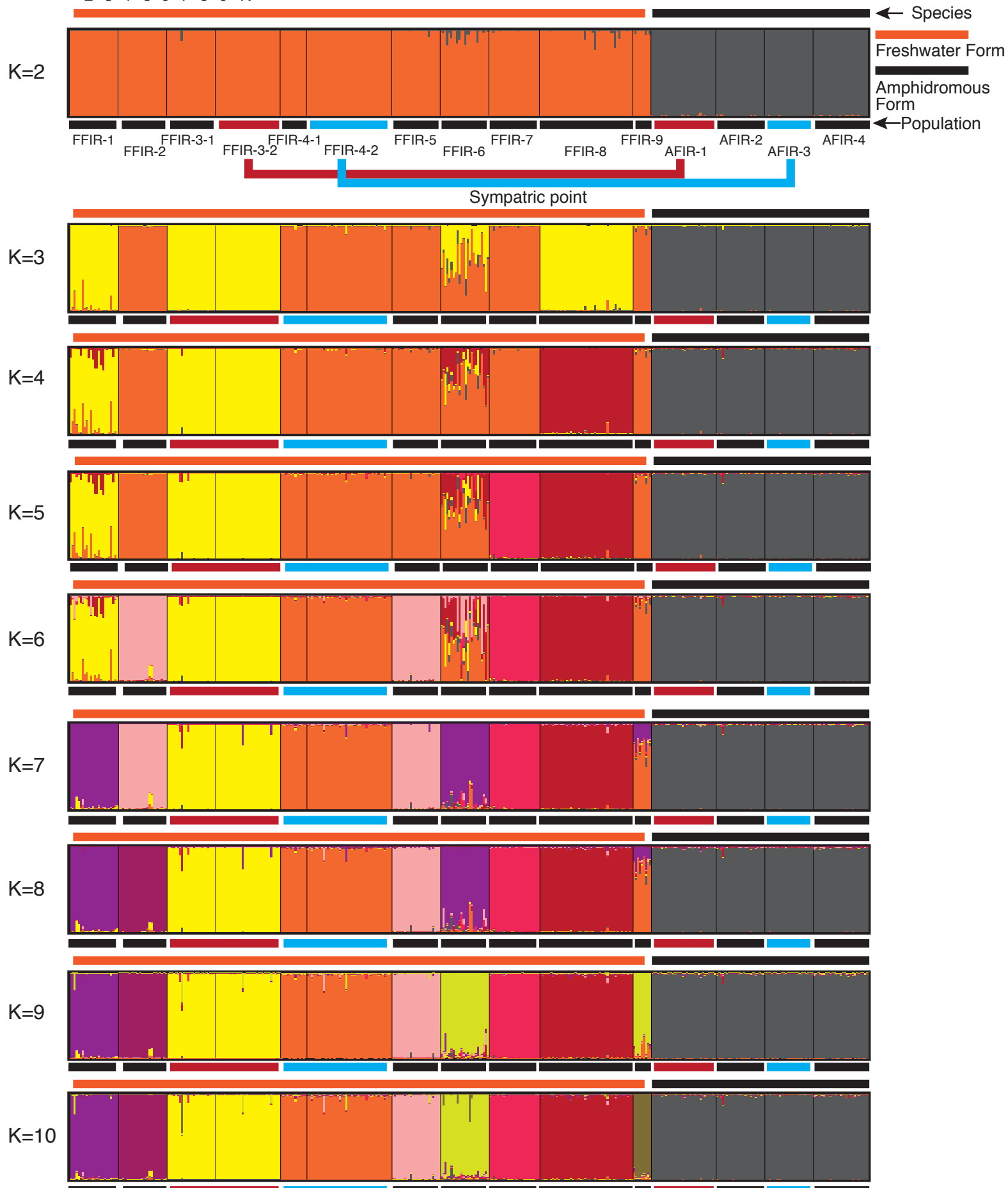
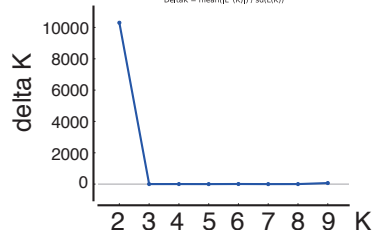


Figure S3 Continued.

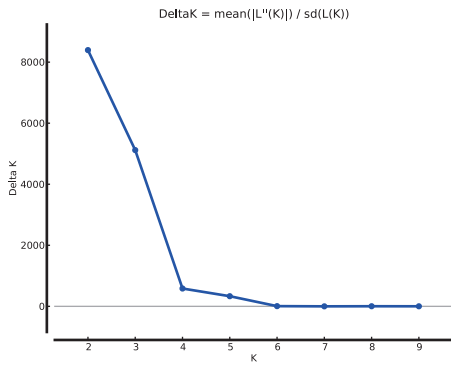
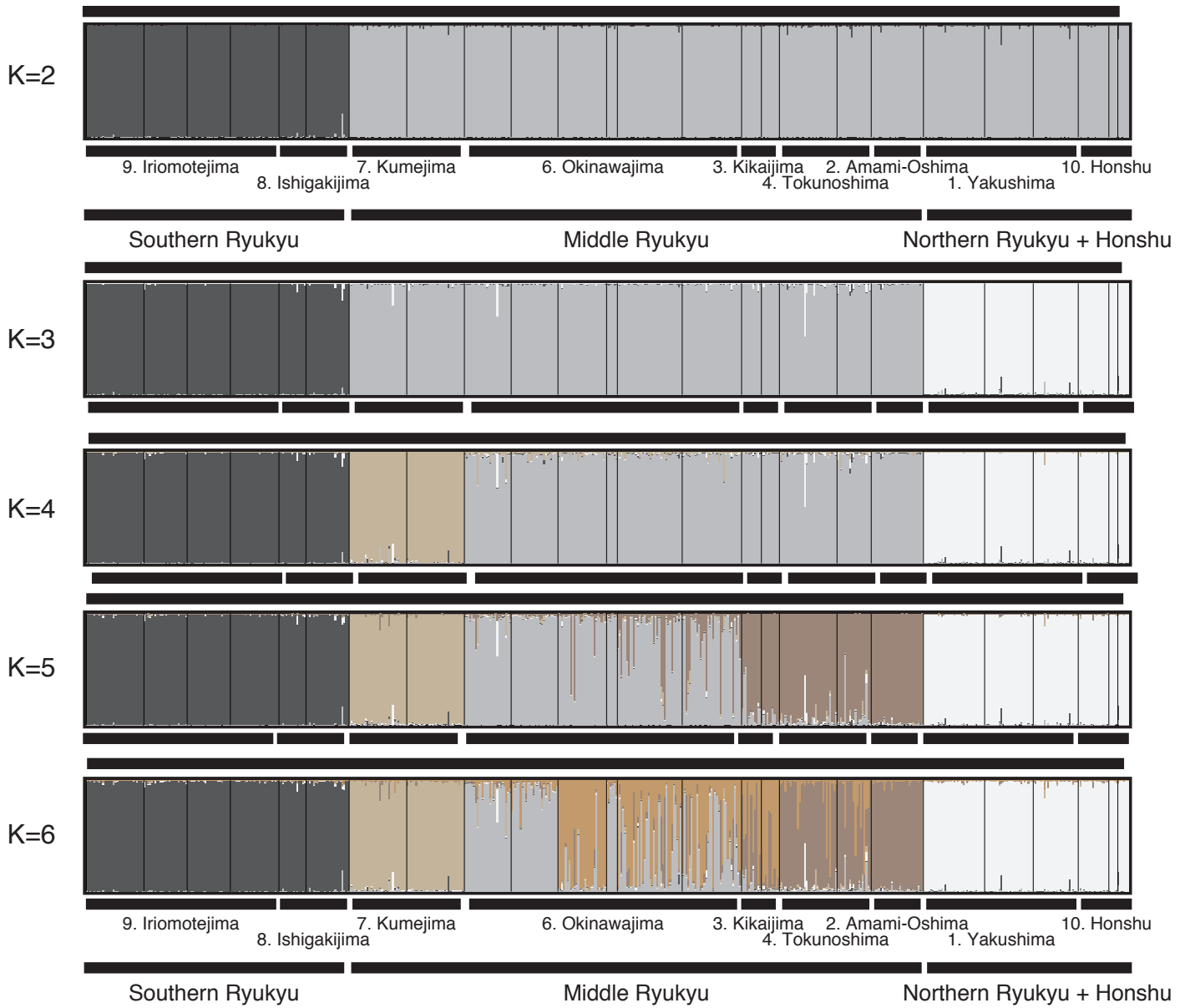
H*Rhinogobius brunneus**R. brunneus* only K = 2

Fig. S3 Continued.

A

2. Amami-Oshima Is.

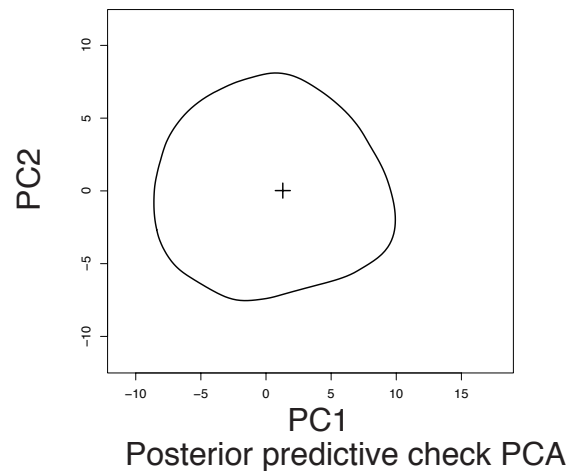
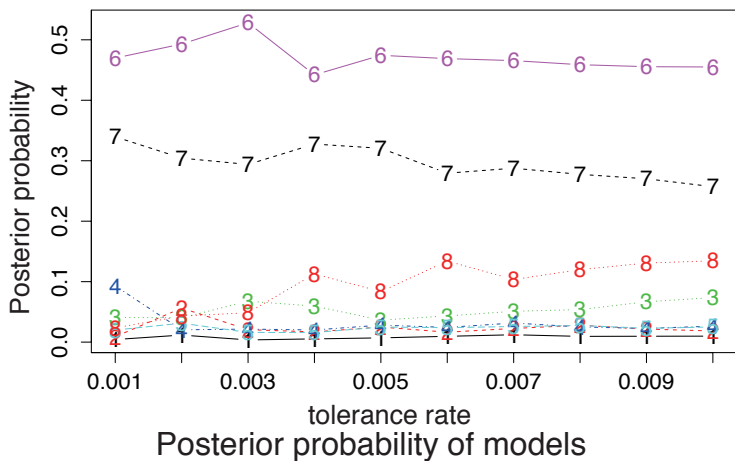
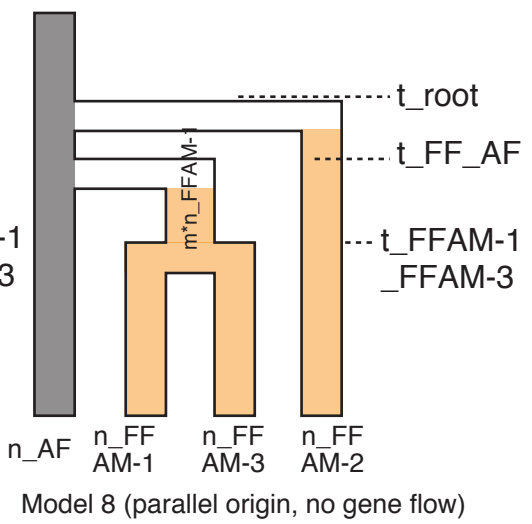
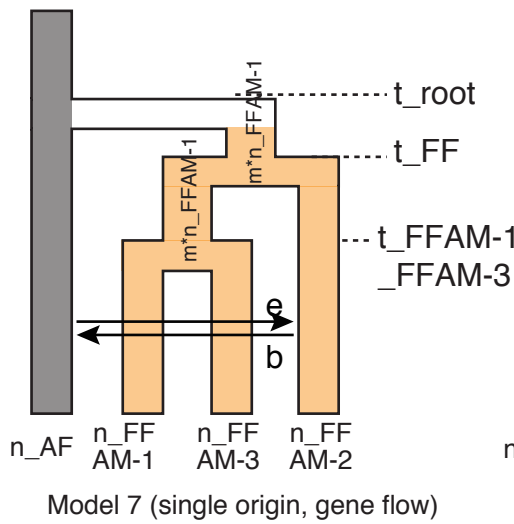
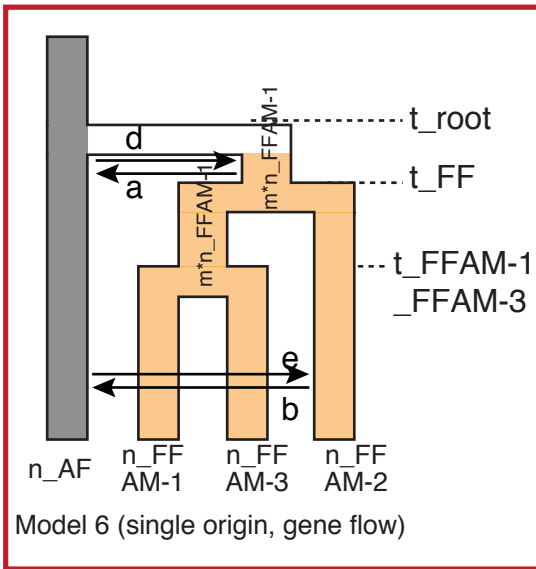
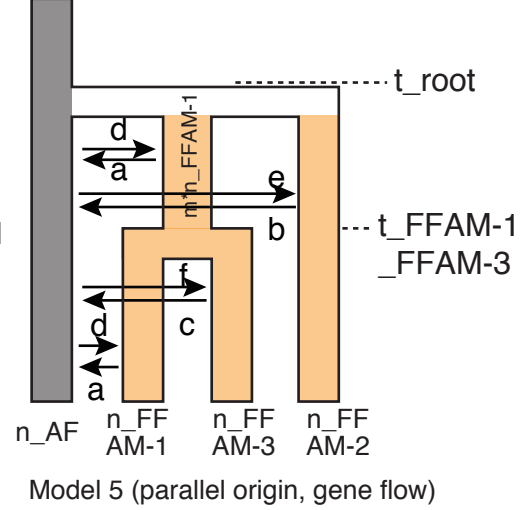
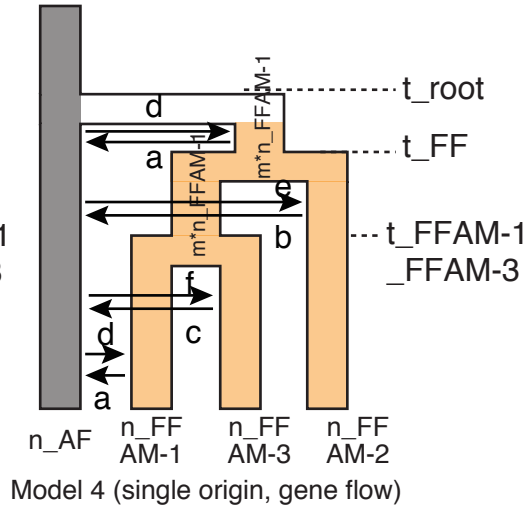
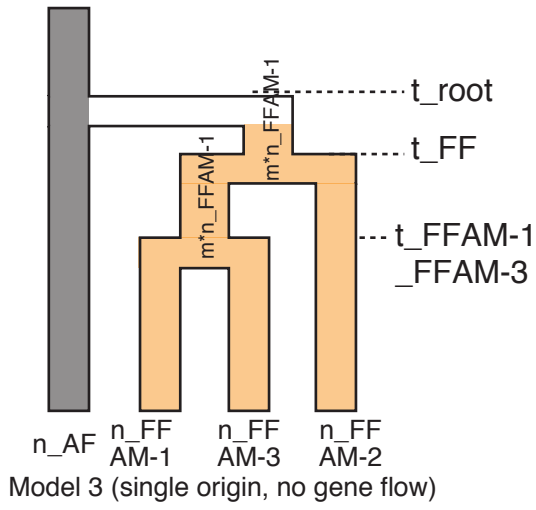
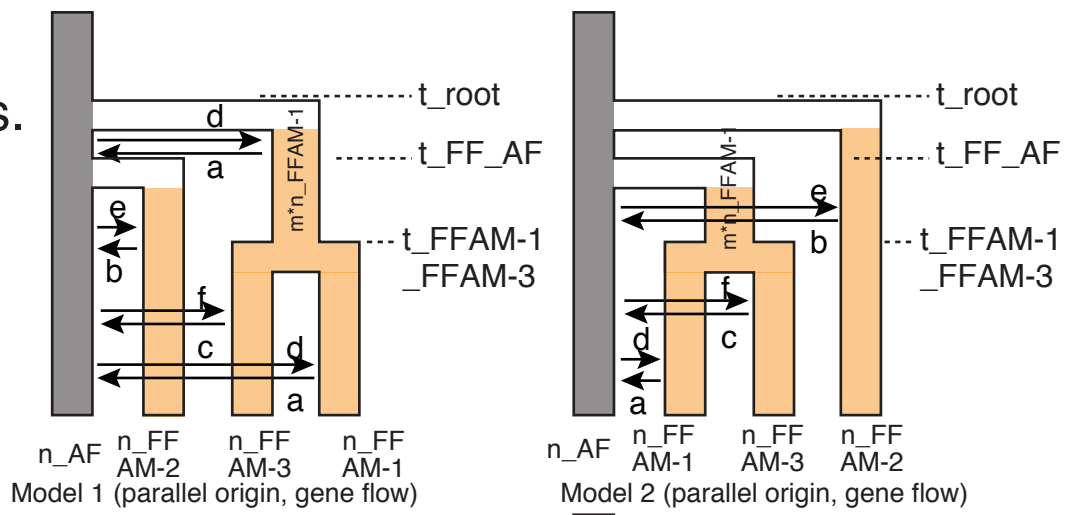
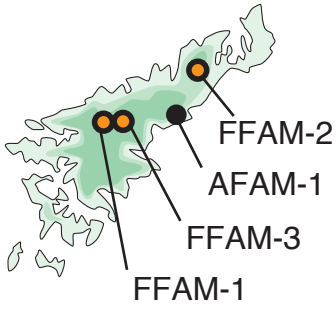
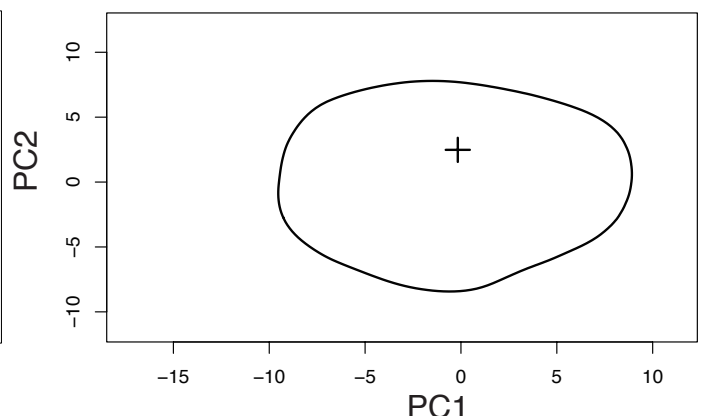
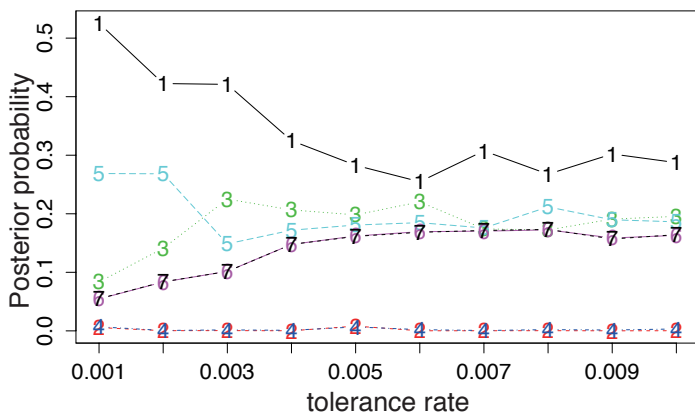
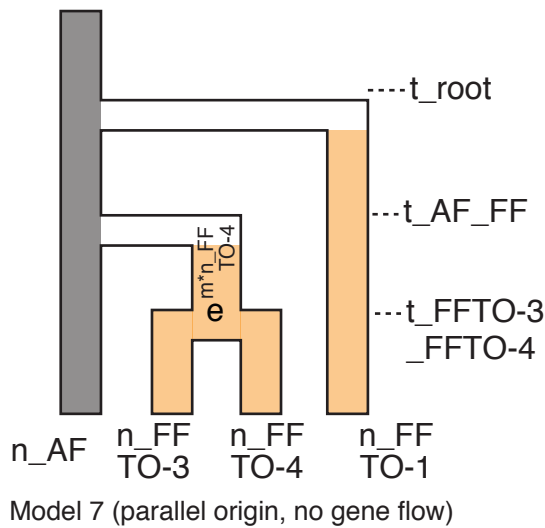
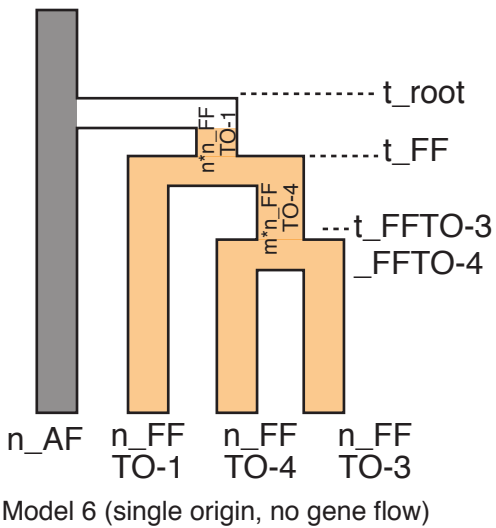
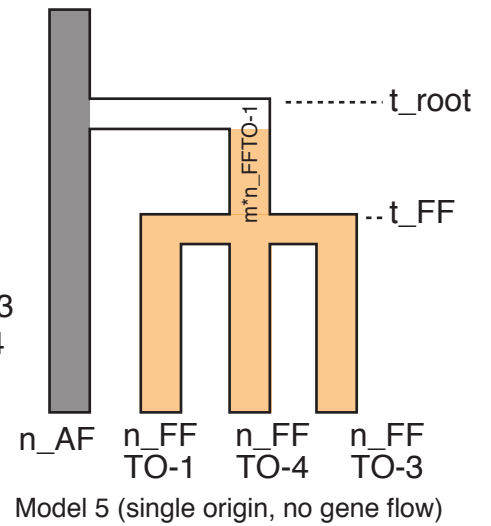
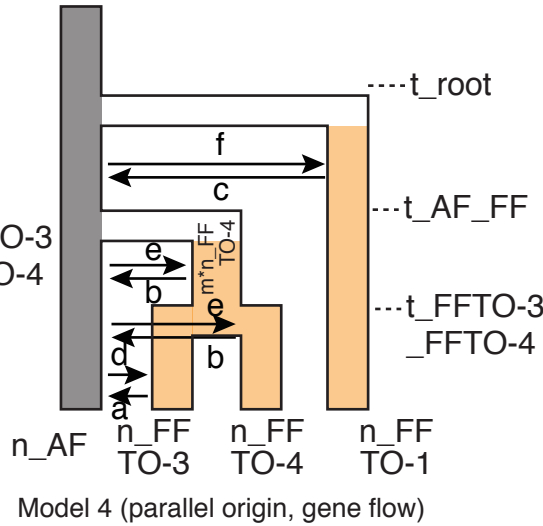
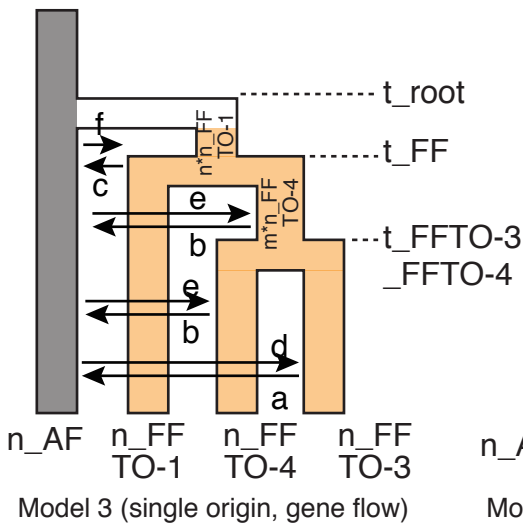
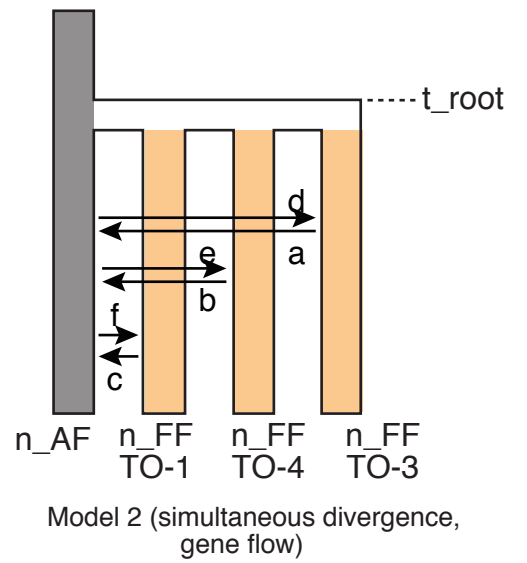
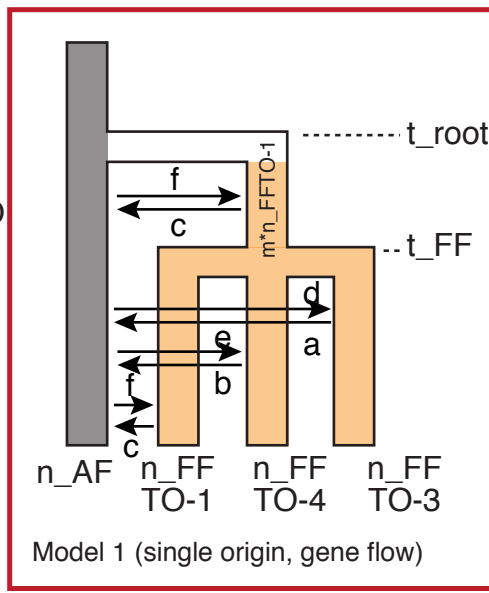
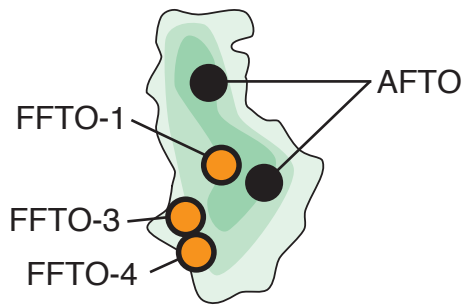


Figure S4 Tested demographic models and results of ABC model selection. Result of posterior predictive check conducted using principal component analysis for the selected model is also shown. Amphidromous and fluvial forms are indicated by black and orange, respectively. Arrows between populations indicate gene flow. The model that was selected is enclosed in a red square.

B

4. Tokunoshima Is.

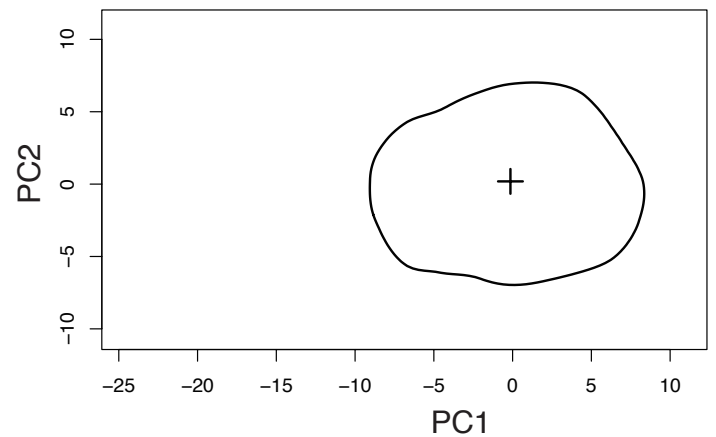
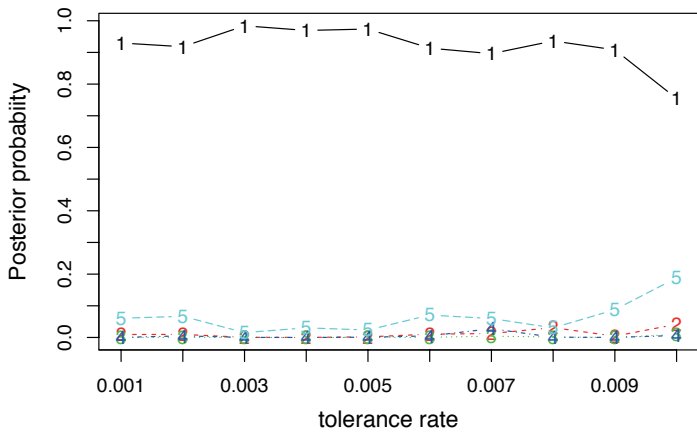
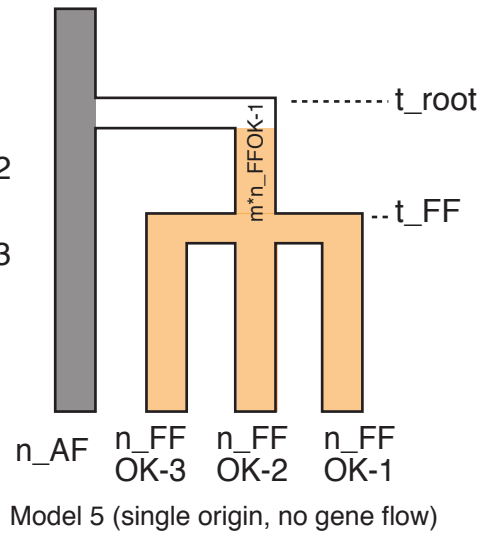
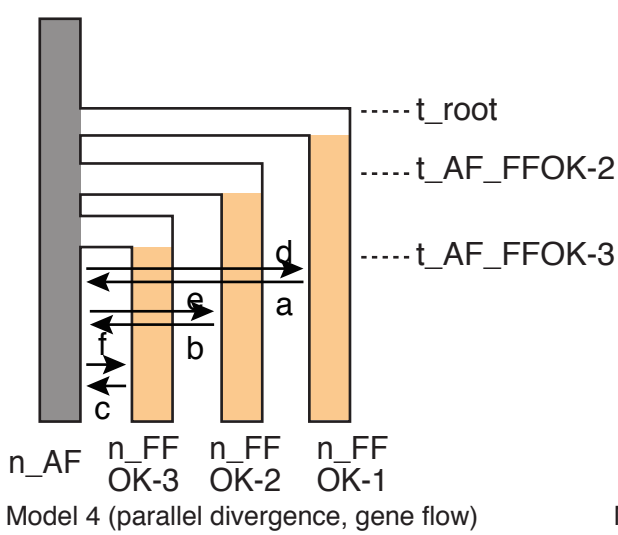
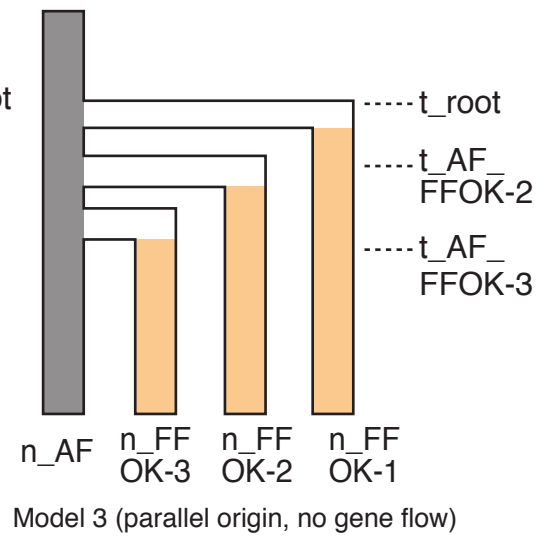
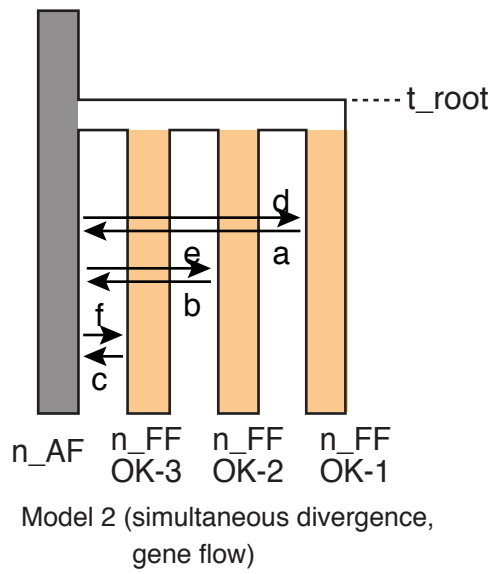
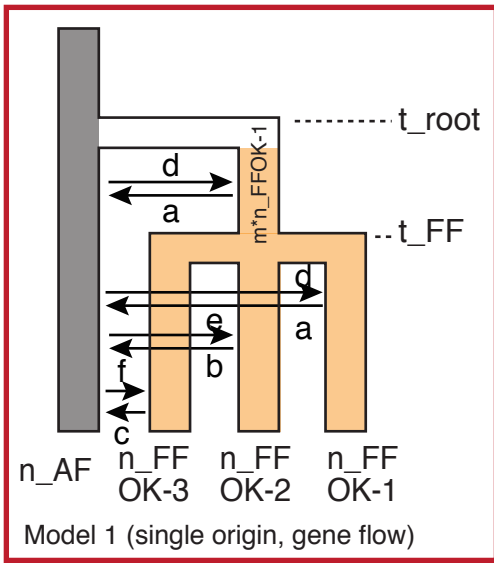
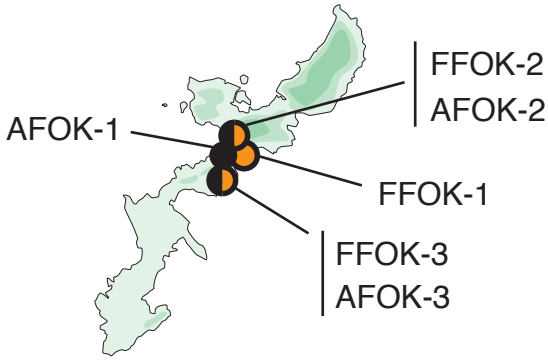


Posterior probability of models
Figure S4 Continued

Posterior predictive check PCA

C

6. Okinawajima Is.



Posterior probability of models
Figure S4 Continued

Posterior predictive check PCA

D

7. Kumejima Is.

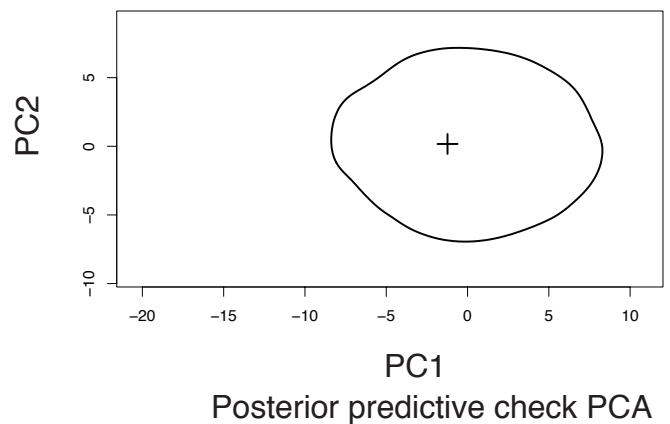
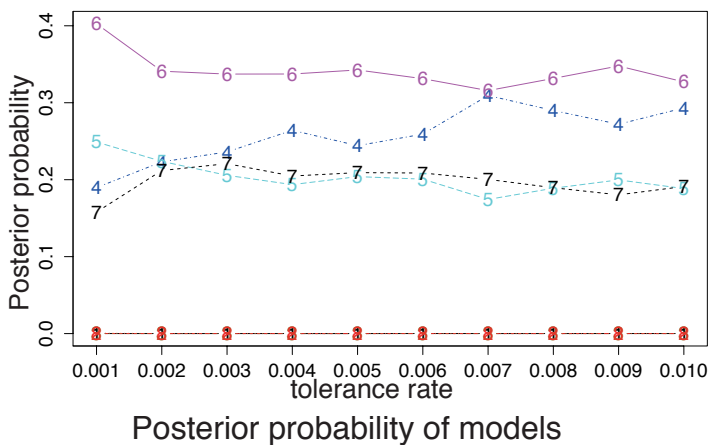
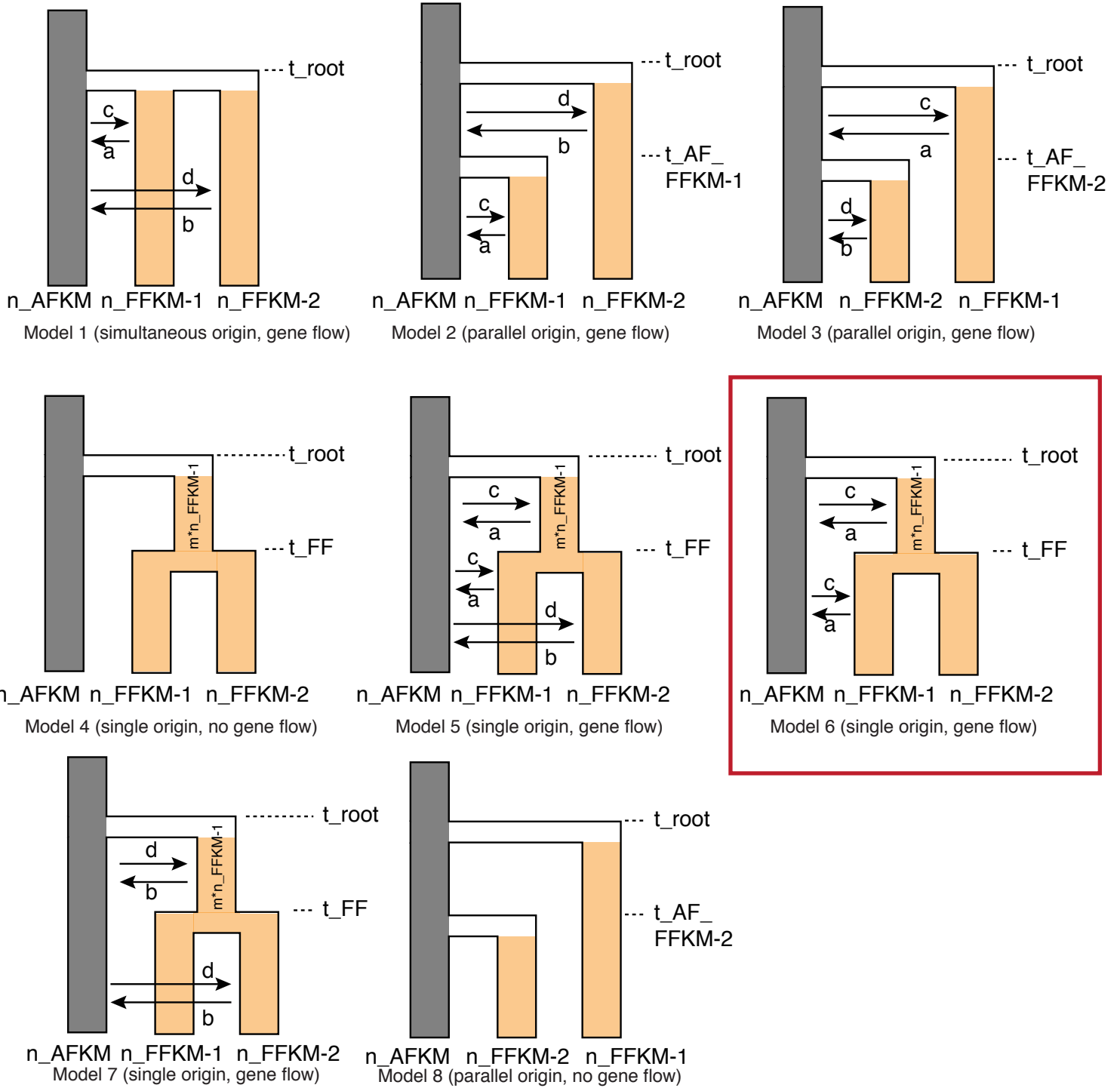
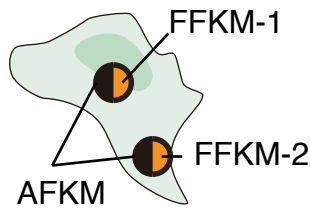
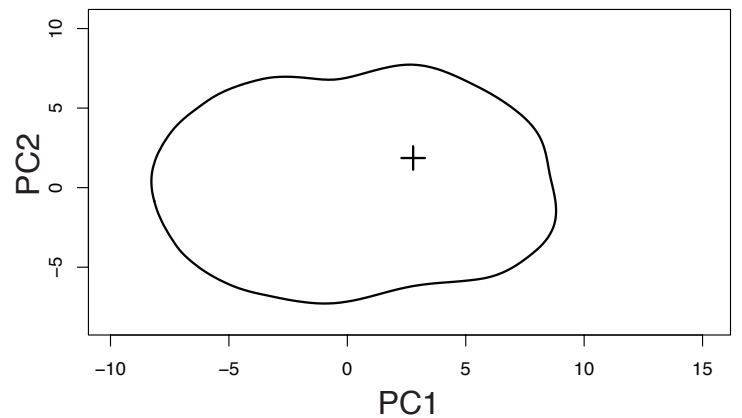
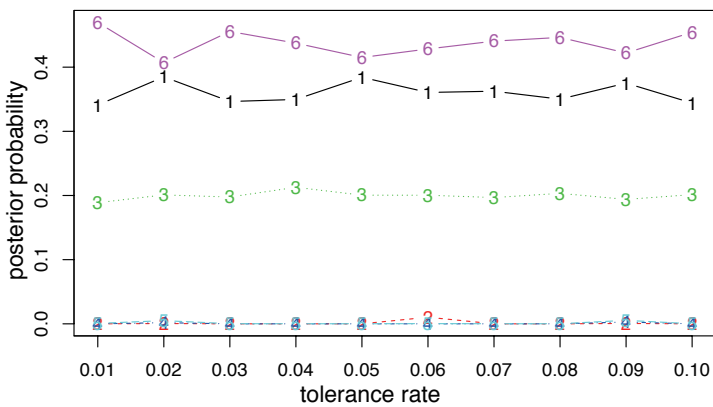
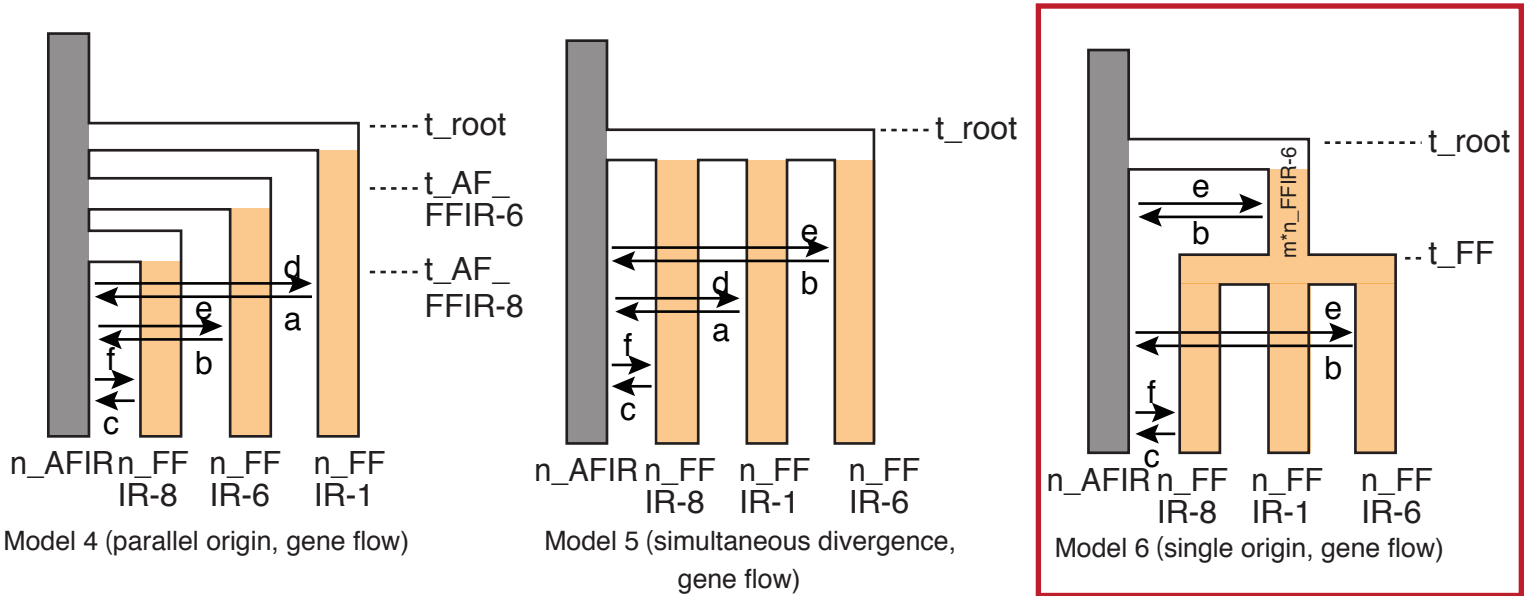
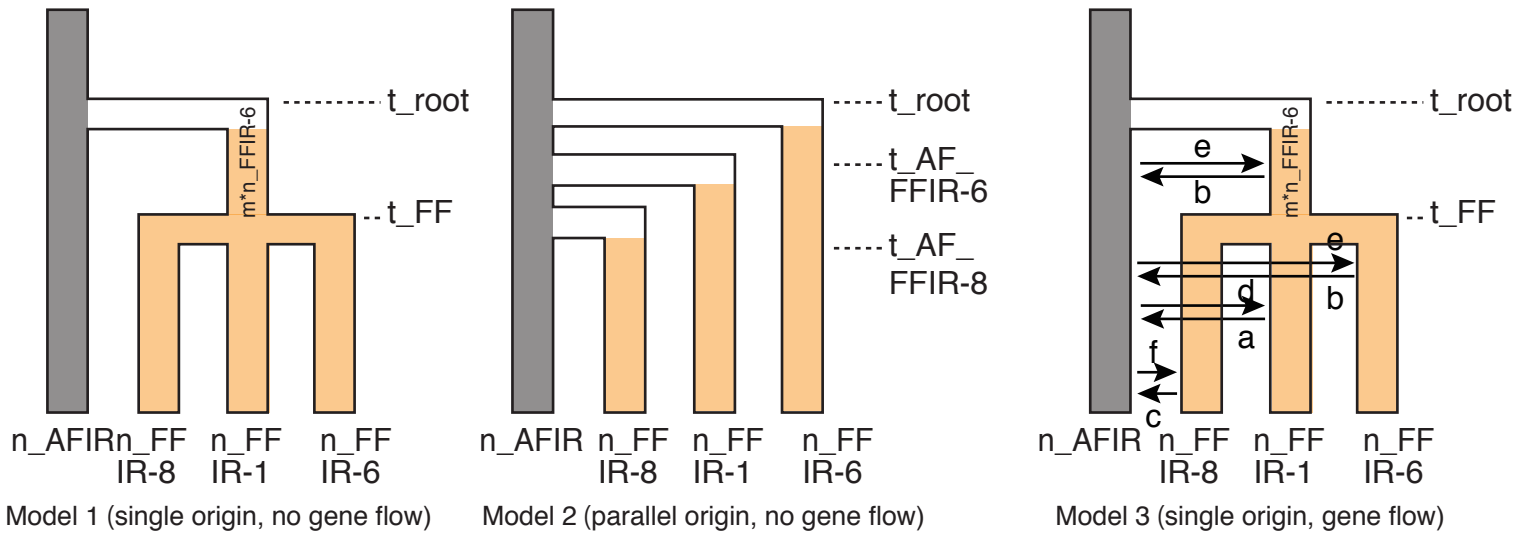
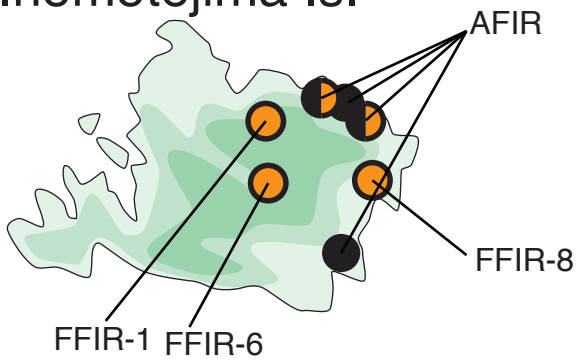


Figure S4 Continued

E

9. Iriomotejima Is.

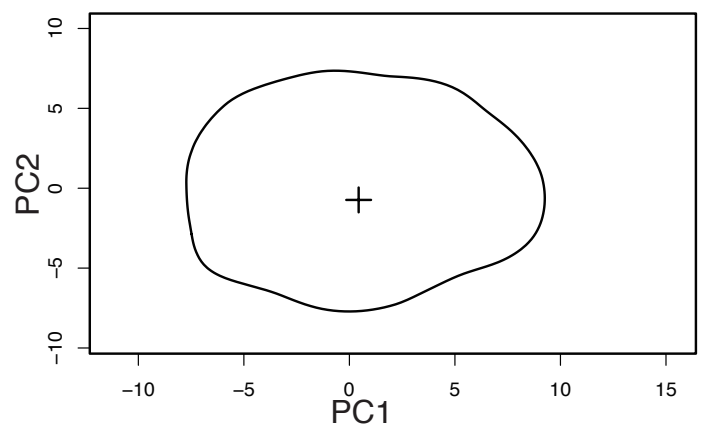
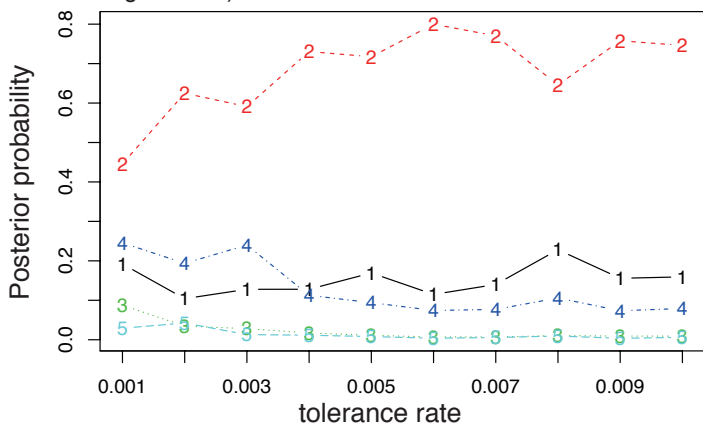
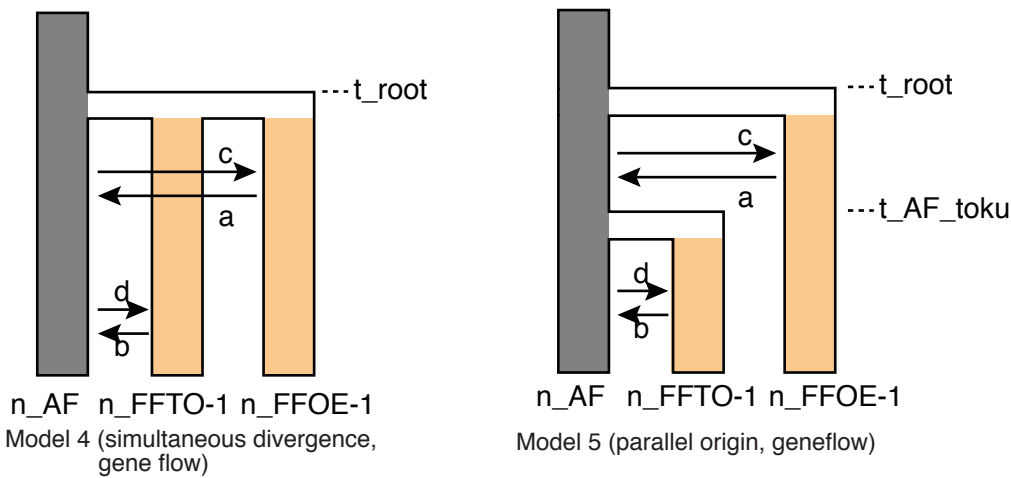
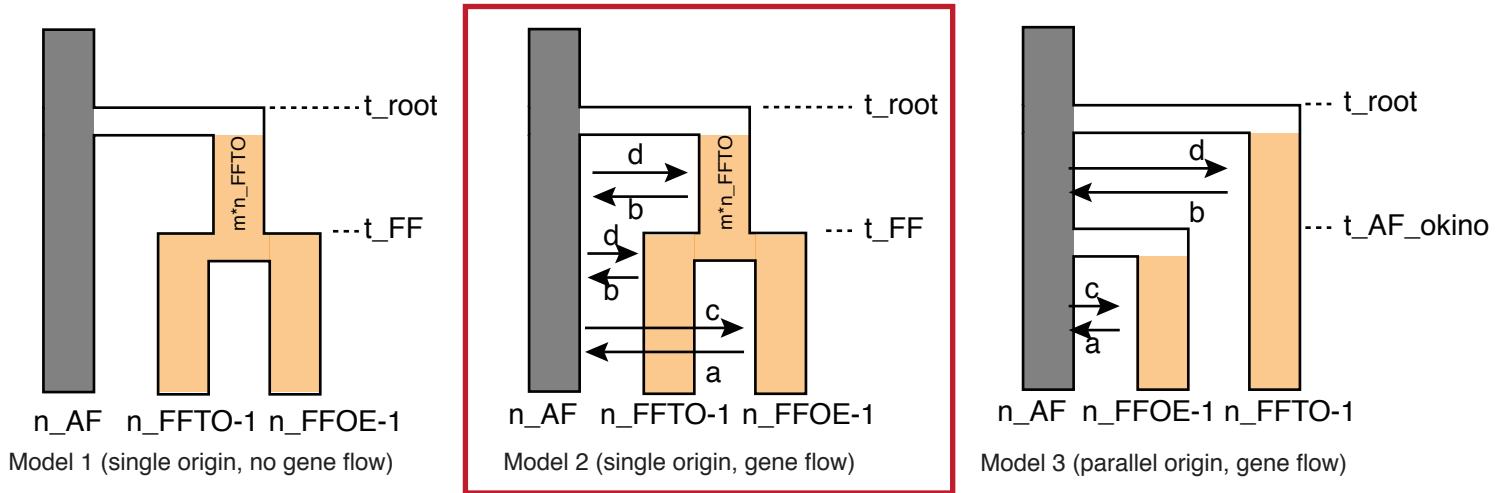
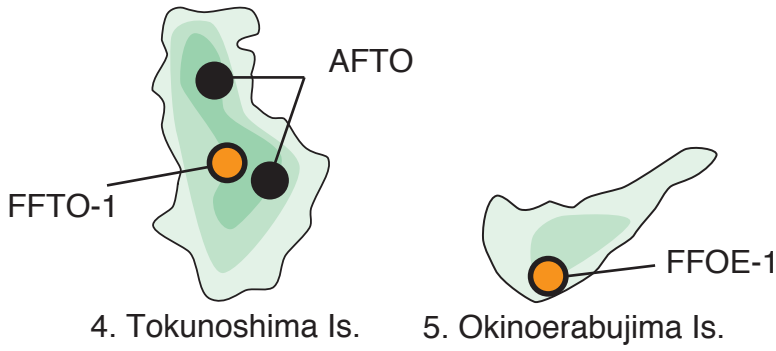


Posterior probability of models

Posterior predictive check PCA

F

4. Tokunoshima Is.– 5. Okinoerabujima Is.



Posterior probability of models

Posterior predictive check PCA

G Middle Ryukyu

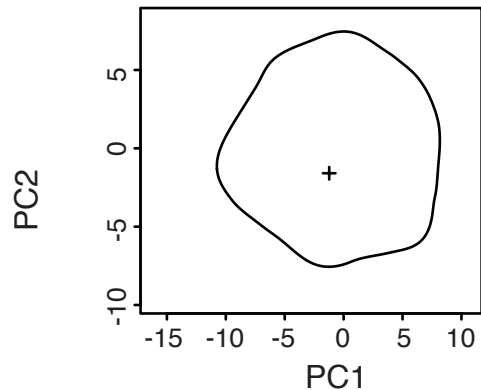
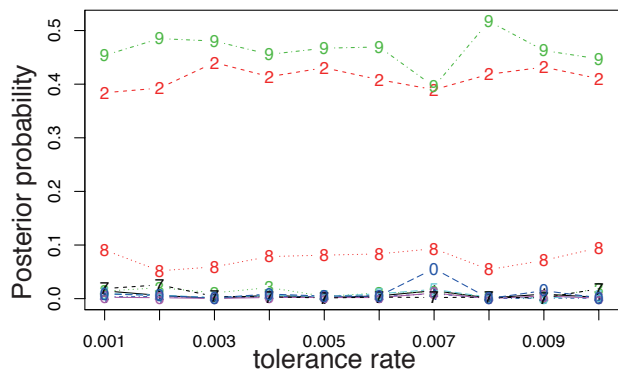
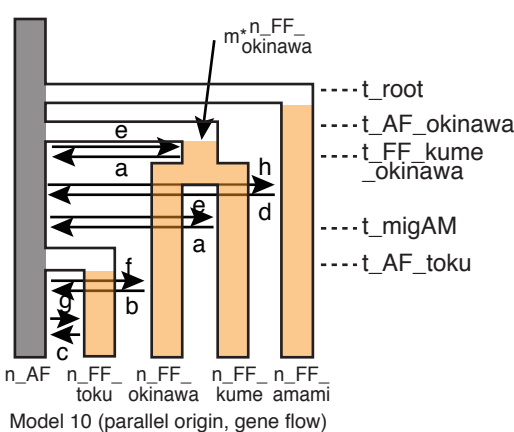
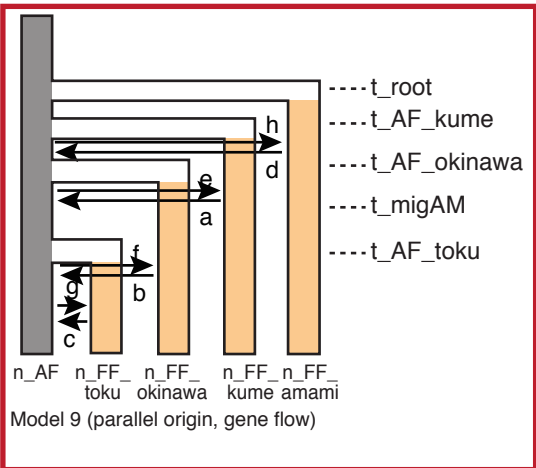
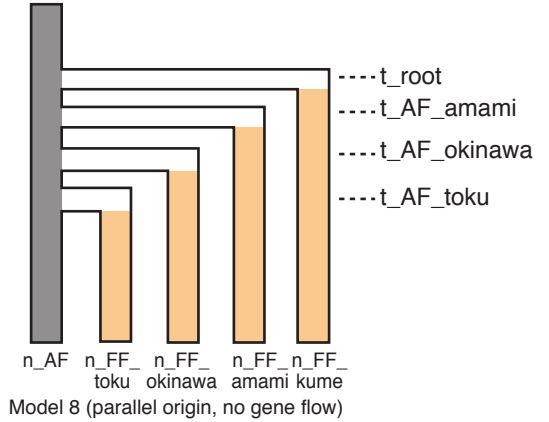
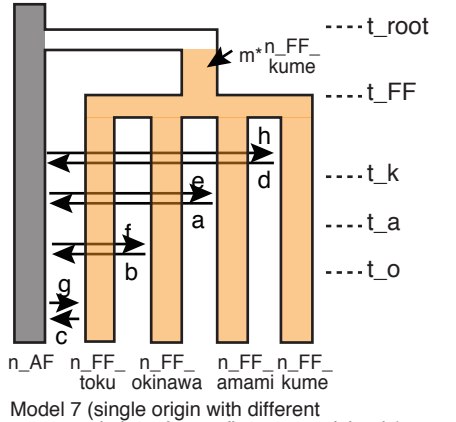
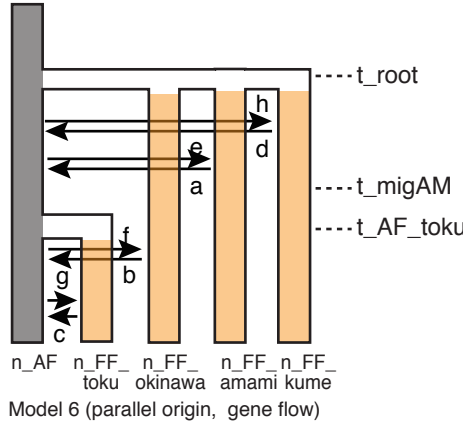
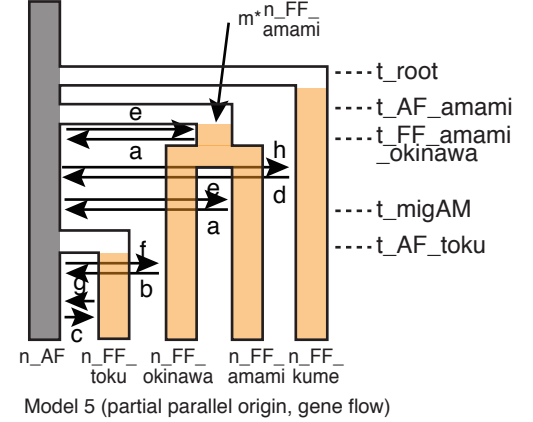
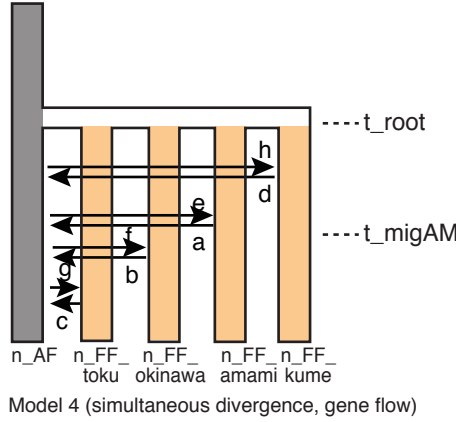
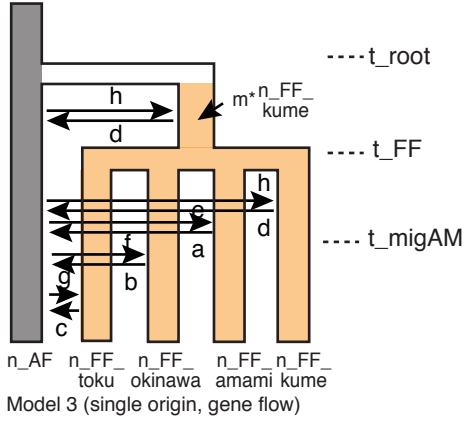
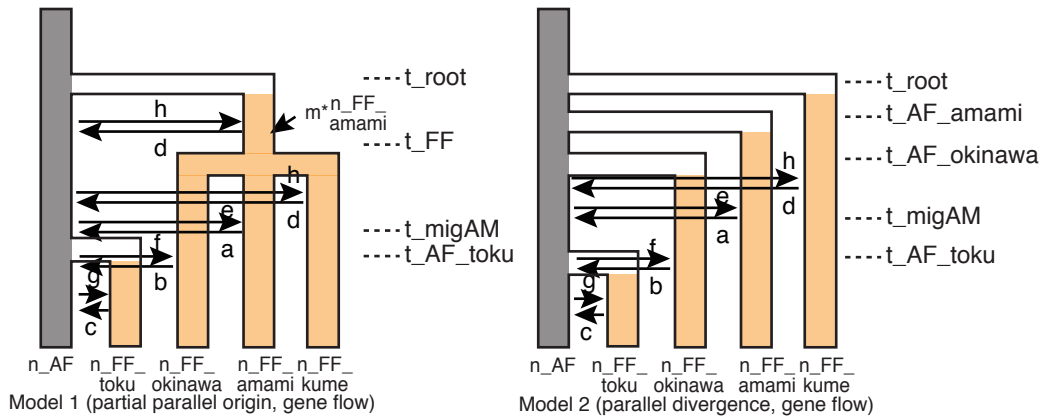
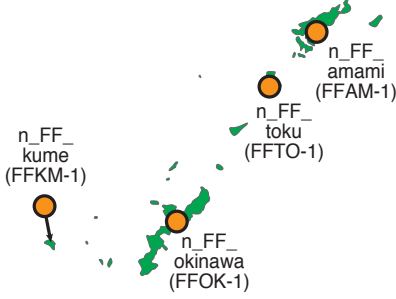


Figure S4 Continued

Posterior predictive check PCA

H

South Ryukyu

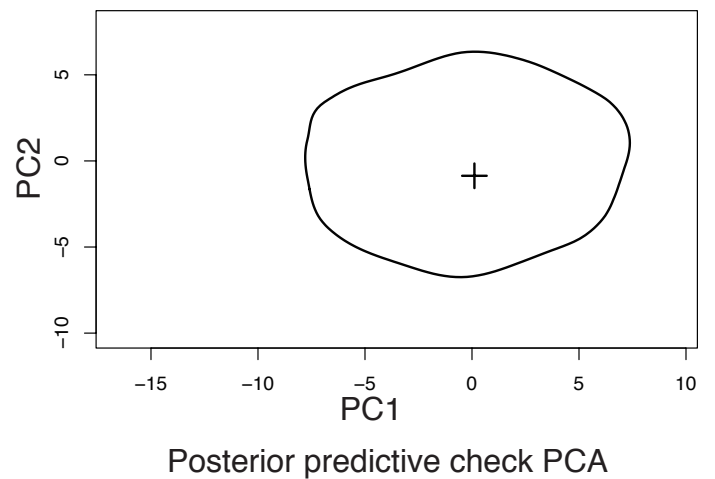
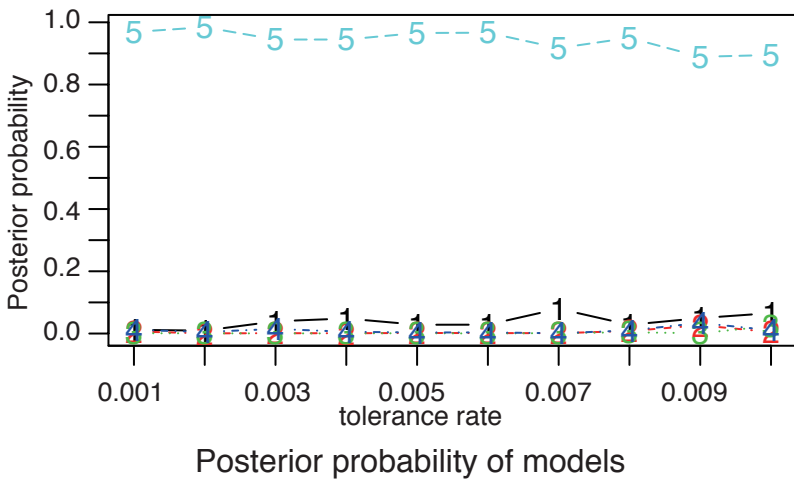
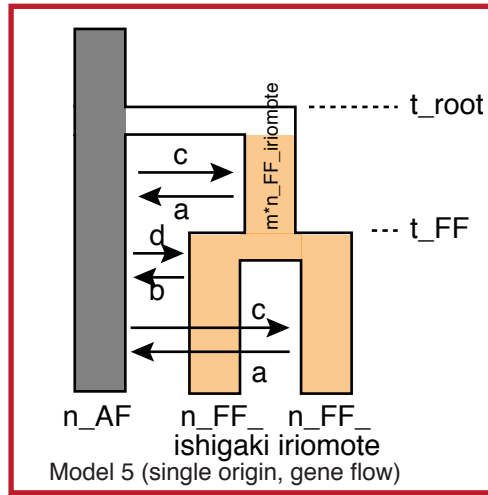
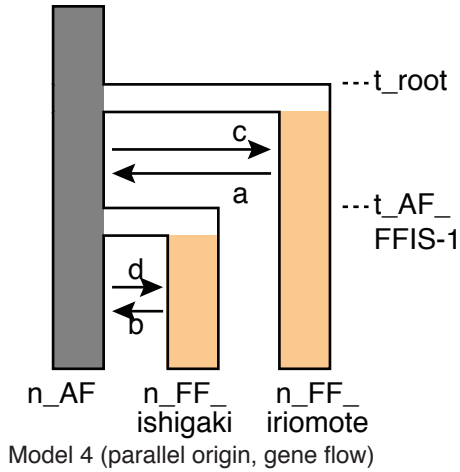
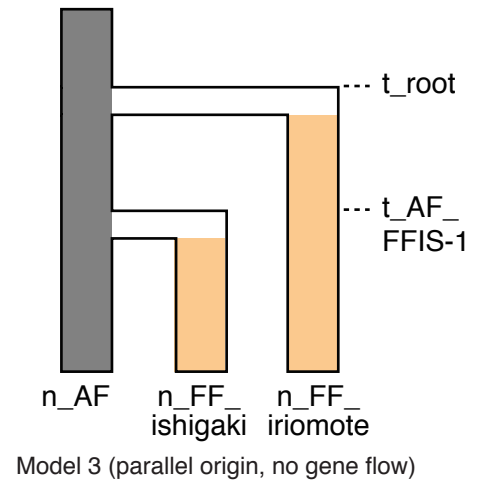
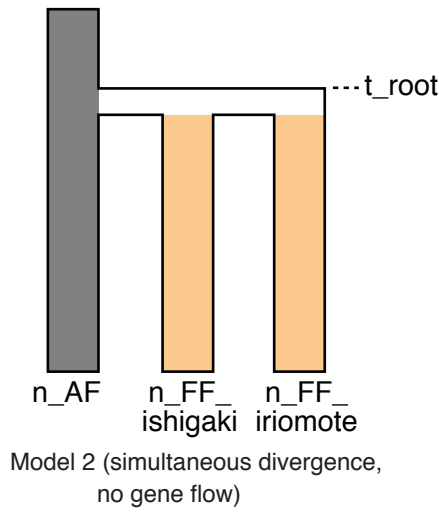
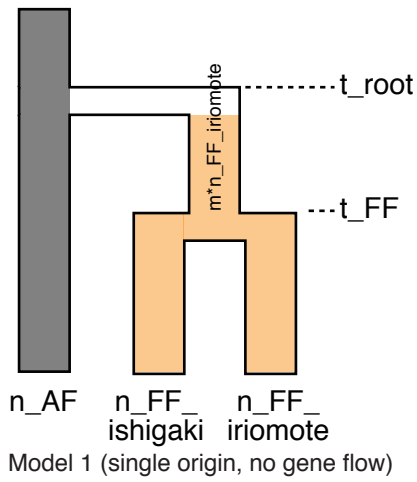
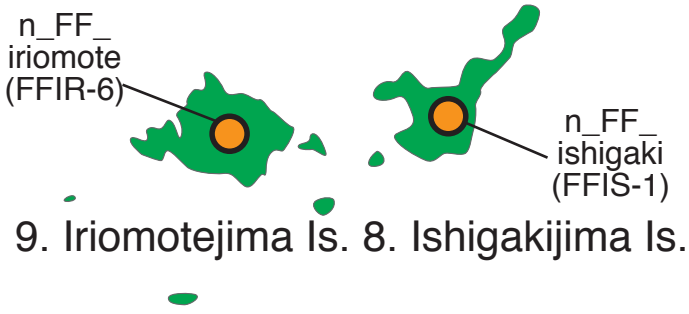
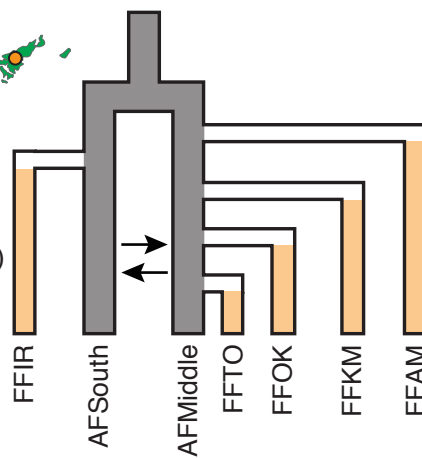
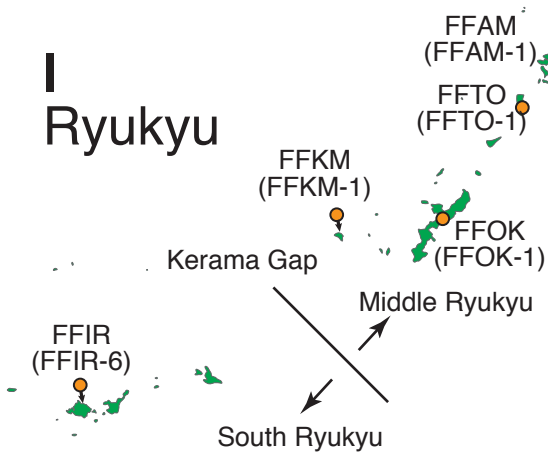
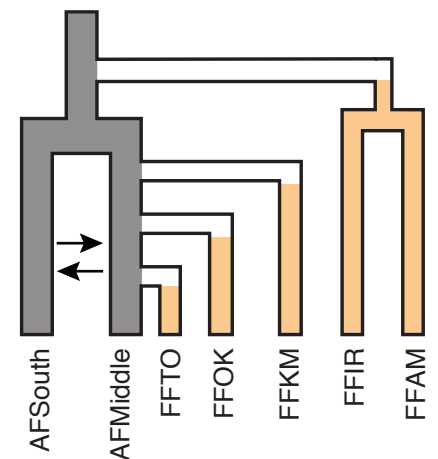


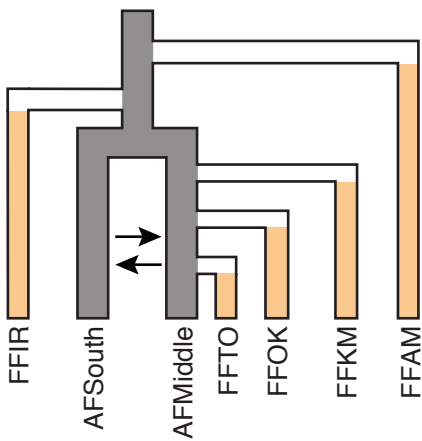
Figure S4 Continued



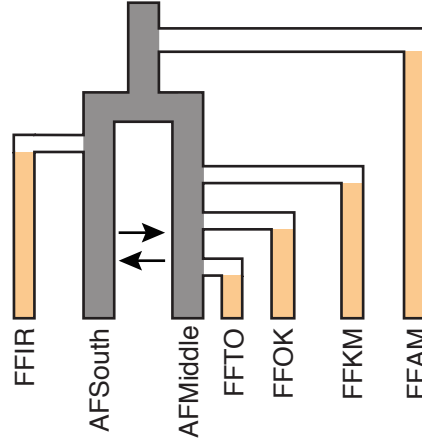
Model 1 (parallel origin between regions)



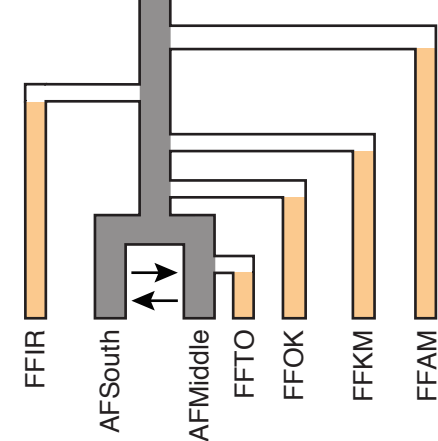
Model 2 (single origin between regions)



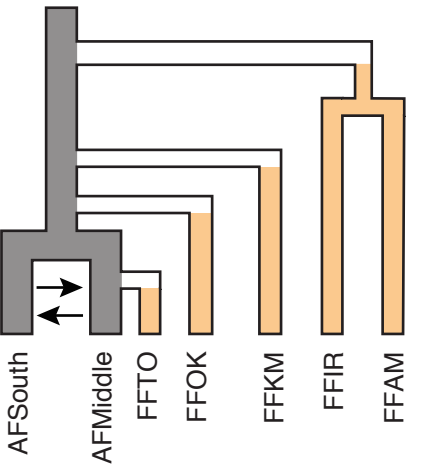
Model 3 (parallel origin between regions)



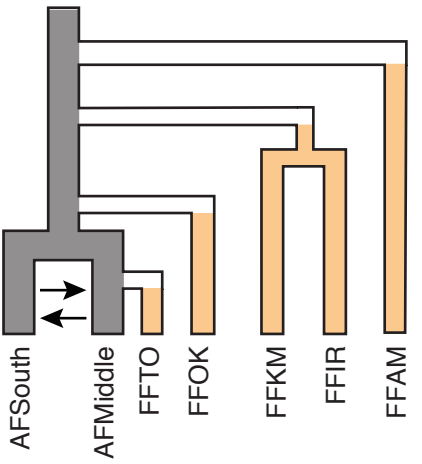
Model 4 (parallel origin between regions)



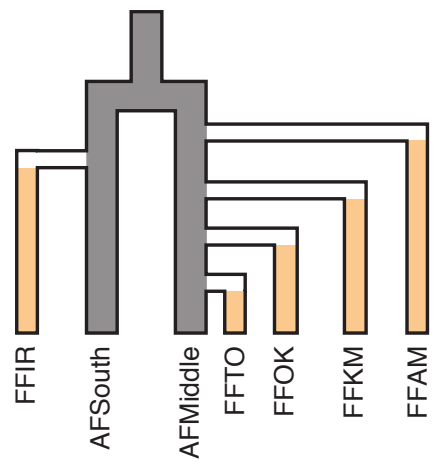
Model 5 (parallel origin between regions)



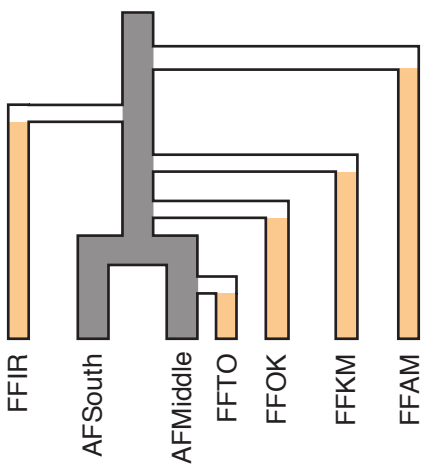
Model 6 (single origin between regions)



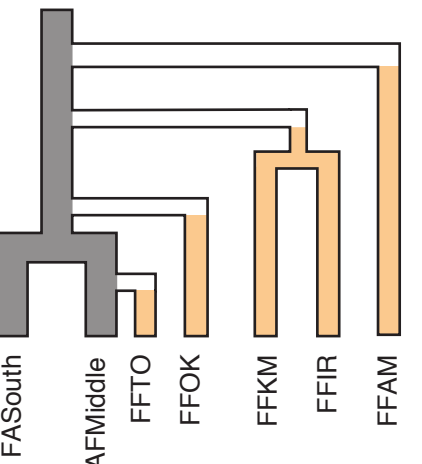
Model 7 (single origin between regions)



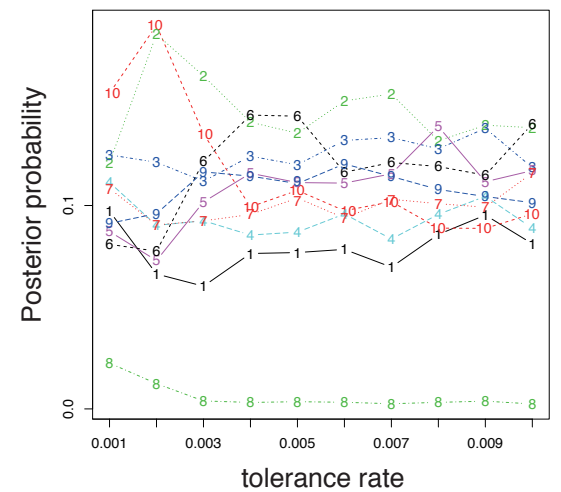
Model 8 (parallel origin between regions)



Model 9 (parallel origin between regions)



Model 10 (single origin between regions)



Parameters except gene flow between amphidromous form in Middle and South Ryukyu were omitted in this figure.

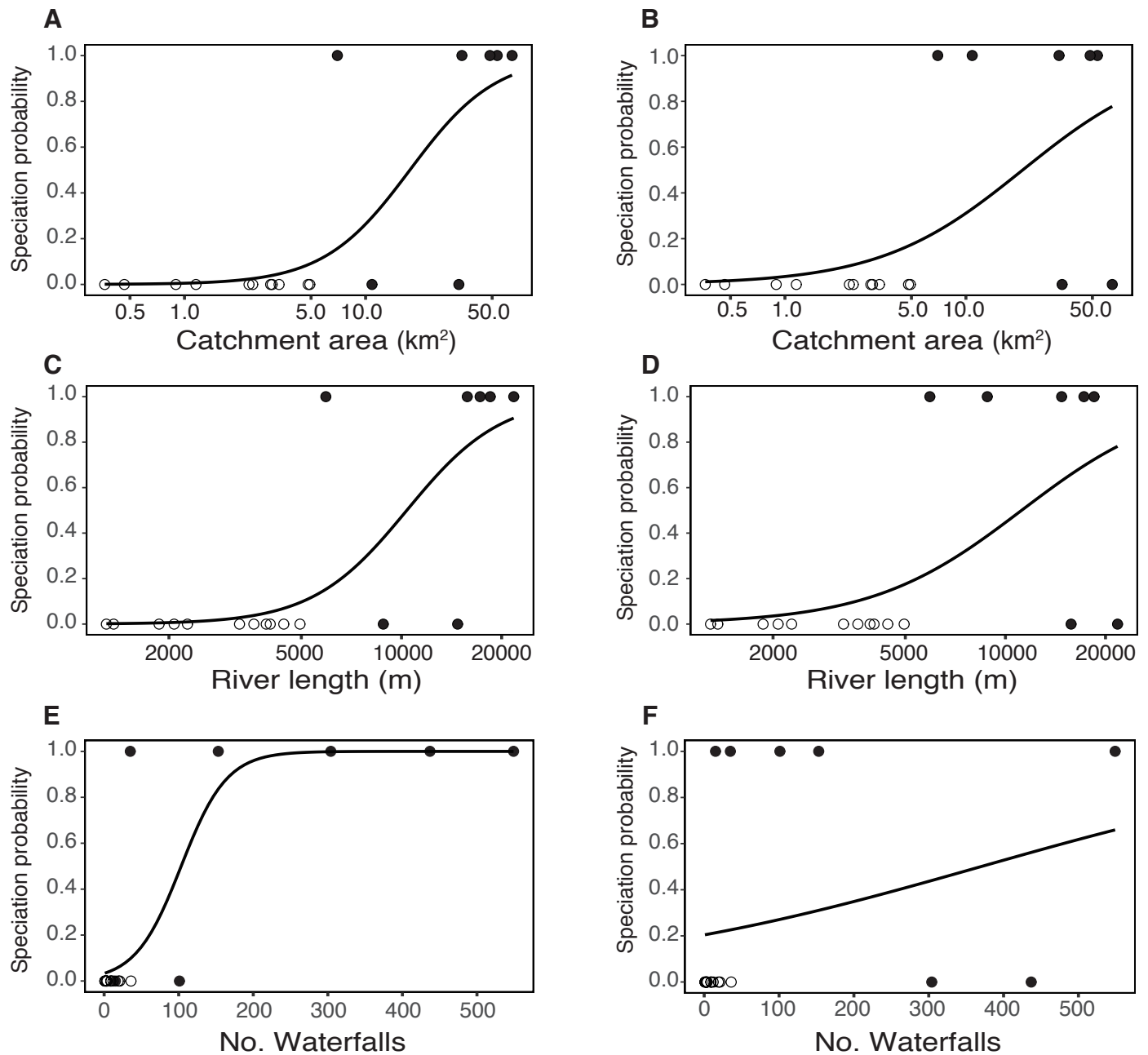


Figure S5 Results of logistic regression analysis between ecosystem size other than island size and speciation probability. Islands with the freshwater form (FF) are represented by filled circles, whereas those without FF are represented by open circles. Filled circles plotted on the x-axis (speciation probability = 0) are for the islands with FF that was assumed to be of non-independent origin. Our ABC analysis indicated that FF in Tokunoshima Is. and Okinoerabujima Is., and Ishigakijima Is. and Iriomotejima Is. shared their origins. Plots in the left column (A,C,E) assumed that the speciation of FF occurred in the larger island (Tokunoshima Is. and Iriomotejima Is.). Plots in the right column (B,D,F) assumed that speciation occurred in the smaller island (Okinoerabujima Is. and Ishigakijima Is.).

TableS1 Information about sampling point and summary statistics of each population.

Region	Island	Form	River system	Point	Population Name	River length m	Catchment area ha	N	Ho	He	Rs	ABC					
Middle Ryukyut	Amami-Oshima	FF	Sumiyo	Sumiyo	FFAM-1	18478	4858	28	0.337	0.371	2.38575	Amami-Oshima/Middle Ryukyu/Ryukyu					
				Akina	Akina	FFAM-2	9969	1135	16	0.228	0.235	2.6299	Amami-Oshima				
				Kawauchi	Kawauchi	FFAM-3	12484	2817	31	0.239	0.234	1.65095	Amami-Oshima				
				Asado	Asado-2	FFAM-4	13155	3238	30	0.128	0.181	2.77865	-				
		AF	Asado	Asado-1	AFAM-1	13155	3238	29	0.507	0.507	3.5093	Amami-Oshima/Middle Ryukyu/Ryukyu					
			Kikajijima	AF	Kikai-1	Kikai-1	AFKI-1	1859	69	11	0.495	0.514	3.3566	-			
		Kikai-2				AFKI-2	1825	56	10	0.438	0.489	3.5747	-				
		Tokunoshima	FF	Akirigami	Akirigami-1	FFTO-1-1	15754	3400	26	0.312	0.325	2.70765	Tokunoshima/Tokunoshima-Okinoerabu/Middle Ryukyu/Ryukyu				
						FFTO-1-2	18	0.304	0.323	2.74675	Tokunoshima/Tokunoshima-Okinoerabu/Middle Ryukyu/Ryukyu						
						FFTO-1-3	9	0.303	0.323	2.61145	-						
						Oose	Oose	FFTO-2	8262	886	19	0.25	0.254	2.74265	-		
						Uwanaru	Uwanaru	FFTO-3	6791	403	24	0.253	0.262	2.68755	Tokunoshima		
	Agon					Agon-1	FFTO-4-1	6958	492	24	0.3	0.311	2.0111	Tokunoshima			
	AF			Shikaura	Manda	Agon-2	FFTO-4-2	9038	1184	7	0.256	0.287	1.78625	Tokunoshima			
						FFTO-5	9038	1184	7	0.279	0.298	1.7933	-				
						AFTO-1	8817	1361	32	0.51	0.522	3.5405	Tokunoshima/Tokunoshima-Okinoerabujima/Middle Ryukyu/Ryukyu				
						AFTO-2	9034	969	19	0.526	0.539	3.49895	Tokunoshima/Tokunoshima-Okinoerabujima/Middle Ryukyu/Ryukyu				
						FFOE-1	8813	10.8	34	0.2	0.227	1.94955	Tokunoshima-Okinoerabujima				
						Okinoerabujima	FF	Sukuta	FFOK-1	5573	369	28	0.294	0.289	2.4092	Okinawajima/Middle Ryukyu/Ryukyu	
	FFOK-2			4442	403				31	0.098	0.121	2.2433	Okinawajima				
	FFOK-3			8967	1682				33	0.262	0.274	2.14875	Okinawajima				
	FFOK-4-1			17218	5330				20	0.307	0.331	2.14305	-				
				FFOK-4-2	8				0.296	0.293	2.07295	-					
	FFOK-5			6445	479				32	0.169	0.176	2.35485	-				
	Okinawajima			FF	Hija	Hija-2	FFOK-6	5299	662	32	0.133	0.144	1.92765	-			
		AFOK-1	26				0.506	0.554	3.62115	Okinawajima/Middle Ryukyu/Ryukyu							
		AFOK-2	25				0.533	0.542	3.55845	Okinawajima/Middle Ryukyu/Ryukyu							
		AFOK-3	27				0.477	0.523	3.6047	-							
		AFOK-4	6				0.475	0.493	3.16775	-							
		AFOK-5	7213				670	36	0.517	0.526	3.5023	-					
		AFOK-6	7700				1179	33	0.482	0.518	3.5941	-					
		Kumejima	FF				Shirase	Suhara	FFKM-1	5925	699	31	0.241	0.274	1.9285	Kumejima/Middle Ryukyu/Ryukyu	
									FFKM-2	3118	150	30	0.073	0.077	1.53105	Kumejima	
									AFKM-1	32	0.491	0.544	3.65115	Kumejima/Middle Ryukyu/Ryukyu			
									AFKM-2	32	0.494	0.52	3.5765	Kumejima			
									South Ryukyu	Ishigakijima	FF	Miyara	Sokohara	FFIS-1	14750	3275	28
	AFIS-1			2861	205	15							0.4	0.486	3.01515	South Ryukyu/Ryukyu	
	AFIS-2	8423	1604	24	0.505	0.496	3.02725	South Ryukyu/Ryukyu									
	Iriomotejima	FF	Pinai	FFIR-1	4705	361	24	0.152					0.202	1.5609	Iriomotejima		
				FFIR-2	2937	160	24	0.094					0.104	1.21535	-		
				FFIR-3-1	2885	131	24	0.083					0.106	1.2767	-		
				FFIR-3-2	32	0.078	0.07	1.24555	-								
FFIR-4-1				4266	498	13	0.233	0.248	2.6738	-							
FFIR-4-2				42	0.24	0.238	2.2249	-									
Iriomotejima	AF	Yuchin	YuchinR-1	FFIR-5	21736	6437	24	0.103	0.118	1.62565	-						
				FFIR-6	21736	6437	24	0.329	0.39	3.0796	Iriomotejima/South Ryukyu/Ryukyu						
				FFIR-7	3697	179	25	0.181	0.194	1.8484	-						
				FFIR-8	6130	557	46	0.289	0.305	2.9351	Iriomotejima						
				FFIR-9	4759	476	9	0.3	0.299	2.74405	-						
				AFIR-1	4266	498	32	0.46	0.456	2.8527	Iriomotejima/South Ryukyu/Ryukyu						
		AF	Geda	Geda-2	AFIR-2	3395	227	24	0.508	0.502	3.1789	Iriomotejima/South Ryukyu/Ryukyu					
					AFIR-3	4266	498	24	0.437	0.46	2.7118	-					
					AFIR-4	15379	3355	27	0.437	0.466	2.9053	-					
					North Ryukyu	Yakushima	AF	Takinokawa	AFYK-1-1	2000	70	34	0.426	0.457	2.97495	-	
									AFYK-1-2	27	0.417	0.456	3.07295	-			
									AFYK-2	10521	1471	25	0.424	0.468	3.4101	-	
Honshu	Honshu	AF	Esuno	AFHS-1					17	0.429	0.392	2.74215	-				
				AFHS-2					5	0.31	0.296	2.6938	-				
				AFHS-3					6	0.3	0.346	1.9406	-				
				Betto	6	0.3	0.346	1.9406	-								
				Yoshida	6	0.3	0.346	1.9406	-								
				Yoshida	6	0.3	0.346	1.9406	-								
Total			42 river system	50 points	59 populations			1420									

Table S2 Primer pairs used in this study.

Locus name	Motif	Forward primer sequence (5' →3') for microsatellite analysis.	Reverse primer sequence (5' →3')	Tm value	Fluorescent	Size range (bp)	No. of alleles
Ohara et al (2004)							
Rhi-3	(CA) ₅ (CA) ₄ (CA) ₁₀	GGATATTCTGTCTCTGTT	ATCTATTCCTTTCTGTTTGTCT	53	6-FAM	131-163	9
Rhi-8	(CA) ₁₈	ATACGCATAGTTTACCTTGA	CCTATGGTTTGAACCTGGGTGTG	53	HEX	155-195	8
Rhi-13	(TG) ₂ (TG) ₄ (GT) ₄	GACTCGCCATCAAATACAAAAA	AGTCTCCTCCTCACCCGCACACC	53	NED	99-132	7
New loci (This study)							
br_a_02	(AC) ₁₁	(GCCTTGCCAGCCCGC)ACTCCTAGCCTACAGCTCACTCG	GCGTCAATGCAGCACTATATTACC	63	VIC	82-126	22
br_a_05	(AC) ₁₀	(GCCTCCCTCGCGCCA)CCACTCAAGGCATTCTCCAGTTT	ATGTTTCTCTCACAAACAATCGC	63	6-FAM	195-243	14
br_a_06	(AC) ₉	(GCCTTGCCAGCCCGC)AACGTCATTATCAGATCCGCTCC	CTCCTAACTTGGCAATCACATGG	63	VIC	200-242	17
br_a_07	(AC) ₁₂	(CAGGACCAGGCTACCGTG)AGTTCATCGATCCATTACCAGA	CGTGGAGCTCTAAACAAGAGGTG	63	NED	174-226	14
br_a_08	(AC) ₁₃	(CGGAGAGCCGAGAGGTG)GTTCTGCTTCATCCATCACCAGT	TAAGATTTGTGCAGATGCGAAGG	63	PET	194-252	17
br_a_09	(AC) ₁₈	(GCCTCCCTCGCGCCA)CCAGCTGAACATGGTGTAGCTTT	TCTCAGCTGCCAGTGAACCTGAAC	63	6-FAM	278-334	29
br_a_10	(AC) ₁₁	(GCCTTGCCAGCCCGC)CGCTGTCAATCACGGTAAGAGTT	CCTATAAGCACGACCTCCATGTG	63	VIC	306-338	15
br_b_05	(AC) ₁₂	(GCCTCCCTCGCGCCA)GGAGCGGAGTTGTTGTGCTTAC	TGTCCCAAGATAGTGCACAAATG	63	6-FAM	182-280	45
br_b_06	(AC) ₁₁	(GCCTTGCCAGCCCGC)ATCCTTAGGCCAACACAAAGCTC	TTAACCATATGCAAACCTCTCCC	63	VIC	190-243	22
br_c_01	(AC) ₁₃	(GCCTCCCTCGCGCCA)TGCAGTGGTTGTGTTGAAAGGTA	CCTTGACAGATGTGCTTAGCAT	63	6-FAM	100-156	20
br_c_02	(AC) ₁₀	(GCCTTGCCAGCCCGC)GACAGCAGCACACTCCTAAGCTC	CAGATCCAGATCCTCTGTTGACAT	63	VIC	111-135	8
br_c_05	(AC) ₁₀	(GCCTCCCTCGCGCCA)TGGCTCTAGAACTCTTGATGATGG	GCACAGTATAGACGCTCTGCACAT	63	6-FAM	212-286	35
br_c_06	(AC) ₁₀	(GCCTTGCCAGCCCGC)CTGAGCAGGACAGGAAGGAAATC	TGAATTGTTTGGACCATGAGACAG	63	VIC	195-283	39
br_c_07	(AC) ₁₄	(CAGGACCAGGCTACCGTG)CCGTTTGACCGACTCTTAAGGT	TTATCCTCACGCCTCCTTTCTTC	63	NED	209-251	17
br_c_08	(AC) ₁₀	(CGGAGAGCCGAGAGGTG)TCTTGCATGACCACAATGTCAAC	AAGCTTGGTTTGTTCCTCTTG	63	PET	222-258	15
br_c_09	(AC) ₁₀	(GCCTCCCTCGCGCCA)TCCCACCAAATAAGCCACAATA	GCTCCTCTGATGGTACTTGCTCA	63	6-FAM	334-386	22
br_c_10	(AC) ₁₁	(GCCTTGCCAGCCCGC)GACAGGAGCTCTGATCATCTCCA	CCACCATCCACCTCTATAACTCTTTC	63	VIC	340-380	17

The sequence in parentheses in forward primers indicate in tail sequence for universal fluorescent primers (Blacket et al., 2012).

Table S3 List of summary statistics used in ABC analysis.

Summary statistics	Abbreviation	Analysis the summary statistics used
Mean number of alleles over loci for each population	K_pop	All analysis
Standard deviation of Mean number of alleles over loci for each population	Ksd_pop	All analysis
Mean number of alleles over loci and population	MEAN_K	All analysis
Standard deviation of mean number of alleles over loci and population	SD_K	All analysis
Mean total number of alleles over loci	TOT_K	All analysis
Mean heterozygosity over loci for each population	H_pop	All analysis
Standard deviation of mean heterozygosity over loci for each population	Hsd_pop	All analysis
Mean heterozygosity over loci and population	MEAN_H	All analysis
Standard deviation of Mean heterozygosity over loci and population	SD_H	All analysis
Mean total heterozygosity	ALL_H	All analysis
Mean Graza–Williamson statistic over loci for each population	GW_pop	All analysis
Standard deviation of mean Graza–Williamson statistic over loci for each population	GWSD_pop	All analysis
Mean Graza–Williamson statistic over loci and population	MEAN_GW	All analysis
Standard deviation of mean Graza–Williamson statistic over loci and population	SD_GW	All analysis
Mean Graza–Williamson statistic over all populations	TOT_GW	All analysis
Mean modified Graza–Williamson statistic over loci for each population	NGW_pop	All analysis
Standard deviation of mean modified Graza–Williamson statistic over loci for each population	NGWSD_pop	All analysis
Mean modified Graza–Williamson statistic over loci and population	MEAN_NGW	All analysis
Standard deviation of mean modified Graza–Williamson statistic over loci and population	SD_NGW	All analysis
Global Fst	FST	All analysis
Global Fit	FIT	All analysis
Pairwise Fst	FST_pop1_pop2	All analysis
Mean number of pairwise differences over populations	PI_pop1_pop2	All analysis
Mean delta mu-square	DMUSQ_pop1_pop2	All analysis
Mean allele range over loci for each population	R_pop	Posterior predictive check only
Standard deviation of mean allele range over loci for each population	Rsd_pop	Posterior predictive check only
Mean allelic range over loci and populations	MEAN_R	Posterior predictive check only
Standard deviation of mean allelic range over loci and populations	MEAN_Rsd	Posterior predictive check only
Mean total allelic range over loci and populations	TOT_R	Posterior predictive check only
Global Fis	FIS	Posterior predictive check only

Table with columns for Region, Island, species/river/system, point, Population Name, and various FFAM codes (FFAM1-FFAM4, FFO1-FFO4, FFK1-FFK4, etc.). The table is organized into blocks for Macizo Piyayyu, South Piyayyu, and North Piyayyu, with sub-sections for Anani/Oshina, Kujajina, Tonokoshina, and others. Each row contains a series of numerical values corresponding to the different FFAM codes.

Table S6. Results of AMOVA analysis.

Geographic scales	Form	Source of variation	%var	Statistics	F values	P values		
Among-islands within-regions	AF	Within-individuals	87.8	<i>FIT</i>	0.122	-		
		Among-individuals	4.4	<i>FIS</i>	0.048	0.001		
		Among-rivers in Middle Ryukyu	1.7	<i>FSC</i>	0.018	0.001		
		Among-islands in Middle Ryukyu	6.2	<i>FCT</i>	0.062	0.001		
			Within-individuals	91.8	<i>FIT</i>	0.082	-	
			Among-individuals	2.8	<i>FIS</i>	0.03	0.016	
			Among-rivers in South Ryukyu	1.7	<i>FSC</i>	0.018	0.001	
			Among-islands in South Ryukyu	3.7	<i>FCT</i>	0.037	0.057	
	FF		Within-individuals	34.9	<i>FIT</i>	0.651	-	
			Among-individuals	3.2	<i>FIS</i>	0.083	0.001	
			Among-rivers in Middle Ryukyu	30.9	<i>FSC</i>	0.448	0.001	
			Among-islands in Middle Ryukyu	31	<i>FCT</i>	0.31	0.001	
				Within-individuals	29.4	<i>FIT</i>	0.706	-
				Among-individuals	3.3	<i>FIS</i>	0.101	0.001
				Among-rivers in South Ryukyu	36.5	<i>FSC</i>	0.527	0.001
				Among-islands in South Ryukyu	30.8	<i>FCT</i>	0.308	0.1
Among-regions	AF	Within-individuals	68.1	<i>FIT</i>	0.319	-		
		Among-individuals	4.1	<i>FIS</i>	0.057	0.001		
		Among-islands	5.3	<i>FSC</i>	0.068	0.001		
		Among-regions	22.5	<i>FCT</i>	0.225	0.002		
	FF		Within-individuals	26.7	<i>FIT</i>	0.733	-	
			Among-individuals	22.5	<i>FIS</i>	0.457	0.001	
			Among-islands	34	<i>FSC</i>	0.409	0.001	
			Among-regions	16.8	<i>FCT</i>	0.168	0.038	

Table S7 (Continued)

Table with columns: parameter name, log10.n.FFOK-3, log10.n.FFOK-2, log10.n.FFOK-1, log10.L.FF, log10.L.root, log10.a, log10.b, log10.c, log10.d, log10.e, log10.f, m, p, a, b, c, d, e, f, n.AFOK, n.FFOK-3, n.FFOK-2, n.FFOK-1, L.FF, L.root

Table with columns: parameter name, log10.n.AFOK, log10.n.FFOK-3, log10.n.FFOK-2, log10.n.FFOK-1, log10.L.FF, log10.L.root, log10.a, log10.b, log10.c, log10.d, log10.e, log10.f, m, p, a, b, c, d, e, f, n.AFOK, n.FFOK-3, n.FFOK-2, n.FFOK-1, L.FF, L.root

Observed Summary Statistics Value

Table with columns: K.FFOK-1, K.FFOK-2, K.FFOK-3, K.AFOK, Kd.FFOK-1, Kd.FFOK-2, Kd.FFOK-3, Kd.AFOK, mean.K, sd.K, tot.K

Observed Value

Table with columns: H.FFOK-1, H.FFOK-2, H.FFOK-3, H.AFOK, Hd.FFOK-1, Hd.FFOK-2, Hd.FFOK-3, Hd.AFOK, mean.H, sd.H, tot.H

Observed Value

Table with columns: GW.FFOK-1, GW.FFOK-2, GW.FFOK-3, GW.AFOK, GWed.FFOK-1, GWed.FFOK-2, GWed.FFOK-3, GWed.AFOK, mean.GW, sd.GW, tot.GW

Observed Value

Table with columns: NGW.FFOK-1, NGW.FFOK-2, NGW.FFOK-3, NGW.AFOK, NGWed.FFOK-1, NGWed.FFOK-2, NGWed.FFOK-3, NGWed.AFOK, mean.NGW, sd.NGW, tot.NGW

Observed Value

Table with columns: R.FFOK-1, R.FFOK-2, R.FFOK-3, R.AFOK, Rd.FFOK-1, Rd.FFOK-2, Rd.FFOK-3, Rd.AFOK, mean.R, sd.R, tot.R

Observed Value

Table with columns: FBT, FIT, FIT

Observed Value

Table with columns: FBL.FFOK-1, FBL.FFOK-2, FBL.FFOK-3, FBL.AFOK, FBT.FFOK-1, FBT.FFOK-2, FBT.FFOK-3, FBT.AFOK, FST.FFOK-1, FST.FFOK-2, FST.FFOK-3, FST.AFOK

Observed Value

Table with columns: PL.FFOK-2.FFOK-1, PL.FFOK-3.FFOK-1, PL.FFOK-3.FFOK-2, PL.AFOK.FFOK-1, PL.AFOK.FFOK-2, PL.AFOK.FFOK-3

Observed Value

Table with columns: DMUSQ.FFOK-2.FFOK-1, DMUSQ.FFOK-3.FFOK-1, DMUSQ.FFOK-3.FFOK-2, DMUSQ.AFOK.FFOK-1, DMUSQ.AFOK.FFOK-2, DMUSQ.AFOK.FFOK-3

Observed Value

Confusion matrix based on 100 samples for each model (neuralnet, tolerance rate=0.01).

Table with columns: model local, I (best model), 1, 2, 3, 4, 5

Table with columns: 1 (Single origin, gene flow), 2 (Parallel origin, gene flow), 3 (Parallel origin, no gene flow), 4 (Parallel origin, gene flow), 5 (Single origin, no gene flow)

Mean model posterior probabilities (neuralnet, tolerance rate=0.01)

Table with columns: model local, I (best model), 1, 2, 3, 4, 5

Table with columns: 1 (Single origin, gene flow), 2 (Parallel origin, gene flow), 3 (Parallel origin, no gene flow), 4 (Parallel origin, gene flow), 5 (Single origin, no gene flow)

Bayes Factor (neuralnet, tolerance rate=0.01)

Table with columns: 1, 2, 3, 4, 5

Table with columns: 1 (Single origin, gene flow), 2 (Parallel origin, gene flow), 3 (Parallel origin, no gene flow), 4 (Parallel origin, gene flow), 5 (Single origin, no gene flow)

Table S7 (Continued)

Phonotajima
Prior distribution (Model 6)
parameter name log10_n_AFIR log10_n_FFIR-8 log10_n_FFIR-1 log10_n_FFIR-6 log10_l_FF log10_l_root log10_b log10_c log10_e log10_f m p b c e f n_AFIR n_FFIR-8 n_FFIR-1 n_FFIR-6 L_FF l_root

posterior distribution (Model 6)

parameter name log10_n_AFIR log10_n_FFIR-8 log10_n_FFIR-1 log10_n_FFIR-6 log10_l_FF log10_l_root log10_b log10_c log10_e log10_f m p b c e f n_AFIR n_FFIR-8 n_FFIR-1 n_FFIR-6 L_FF l_root
Min. 1.915487029 2.15476963 2.41479208 2.47373259 -1.21508222 -9.62650014 -9.80126982 -9.46544838 0.29780842 0.00271058 5.72E-12 2.07E-10 1.47E-10 3.15E-10 84.25277843 159.0962715 131.1922672 276.0443047 565.6402479 314.4296006

Observed Summary Statistics Value

K_AFIR-1 K_FFIR-8 K_FFIR-1 K_FFIR-6 Kd_FFIR-1 Kd_FFIR-6 Ksd_FFIR-8 Ksd_FFIR-6 Ksd_FFIR-1 Ksd_FFIR-6 Ksd_FFIR-8 Ksd_FFIR-6 mean_K sd_K tot_K
Observed Value 2.35 3.75 3.75 6.85 1.81442 2.4034 4.35135 4.51051 3.925 1.44309 6.85

Confusion matrix based on 100 samples for each model (neuralnet, tolerance rate=0.01)

model \ actual 1 2 3 4 5 6(best model)
1 (Single origin, no gene flow) 0.9 0.03 0.01 0 0 0.06
2 (Parallel origin, no gene flow) 0.02 0.88 0 0.07 0.02 0.01

Mean model posterior probabilities (neuralnet, tolerance rate=0.01)

model \ actual 1 2 3 4 5 6(best model)
1 (Single origin, no gene flow) 0.6882 0.0294 0.0885 0.0044 0.0271 0.1763
2 (Parallel origin, no gene flow) 0.0128 0.8313 0.0299 0.1906 0.0067 0.0536

Bayes Factor (neuralnet, tolerance rate=0.01)

numerator \ denominator 1 2 3 4 5 6
1 (Single origin, no gene flow) 1 94024.0848 1.7085 234566.4108 2458.9742 0.7571

Table S7 (Continued)
Tokunoshima, Okinoerabujima
Prior distribution (Model2)

parameter name	log10_a	log10_b	log10_c	log10_d	log10_n_AF	log10_n_okino	log10_n_toku	log10_t_FF	log10_t_root	m	p	a	b	c	d	n_AF	n_okino	n_toku	t_FF	t_root
distribution	uniform	uniform	uniform	uniform	uniform	uniform	uniform	uniform	uniform	uniform	uniform	NA	NA	NA	NA	NA	NA	NA	NA	NA
Min.	-9	-9	-9	-9	2	2	2	3	3	0.5	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Max.	-3	-3	-3	-3	5.3	4.7	4.7	6.2	6.2	2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Complex parameter	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	10*(log10_a)	10*(log10_b)	10*(log10_c)	10*(log10_d)	10*(log10_n_AF)	10*(log10_n_okino)	10*(log10_n_toku)	10*(log10_t_FF)	10*(log10_t_root)

Posterior distribution (Model2)

parameter name	log10_a	log10_b	log10_c	log10_d	log10_n_AF	log10_n_okino	log10_n_toku	log10_t_FF	log10_t_root	m	p	a	b	c	d	n_AF	n_okino	n_toku	t_FF	t_root
Min.	-10.14958659	-11.11341166	-11.45255462	-11.16476095	2.776673139	2.061336903	2.360266768	2.074409034	2.401879147	0.404548429	0.003573449	7.15E-11	6.89E-12	3.48E-12	6.82E-12	591.8862858	114.8724309	227.2451764	118.0719184	250.5689687
Weighted 2.5% Perc.	-9.202913073	-9.023659958	-8.941294563	-8.899265843	3.443037722	2.558173948	3.069487412	2.585305711	3.074096215	0.514615452	0.242293218	6.31E-10	9.35E-10	1.13E-09	1.21E-09	2738.742395	360.0914213	1161.04389	382.3846214	1196.245616
Weighted Median	-5.961562108	-5.348512608	-5.909820868	-5.452163411	4.050707087	3.27214837	3.710535377	3.241885296	4.424051541	1.146281297	0.532173827	1.10E-06	4.49E-06	1.22E-06	3.45E-06	11172.29011	1869.355655	5143.961304	1746.746089	26421.55556
Weighted Mean	-6.116380114	-5.784742914	-6.09210827	-5.890588205	4.053608151	3.246556204	3.698837717	3.347693538	4.452163395	1.176857129	0.524192096	3.82E-05	4.56E-05	1.75E-05	3.99E-05	14562.28311	2244.978483	6264.689698	7467.92739	113792.1471
Weighted Mode	-4.3469019	-4.50111073	-5.950859573	-5.511585451	4.003350363	3.328487865	3.714023312	3.1567225	3.890017278	0.666153834	0.579110955	-2.36E-06	-2.27E-06	-8.21E-07	2.27E-06	7453.902105	1283.334464	3690.010605	661.3095564	8150.788215
Weighted 97.5% Perc.	-3.500079707	-3.449034485	-3.847886318	-3.449370638	4.67254187	3.822723853	4.251151813	4.79650326	5.847947266	1.924979731	0.757162432	0.000315004	0.000348579	0.000140908	0.000347709	48575.89176	6619.335656	17901.73924	62485.42911	701379.1444
Max.	-2.542248289	-2.359737163	-2.920104543	-2.606485662	5.102251631	4.405683442	4.688525192	5.832934866	6.368778078	2.070120686	0.920796071	0.002853665	0.004411571	0.001196782	0.002424363	124846.5407	24946.85325	47142.42174	675162.8522	2326504.085

Observed Summary Statistics Value

Summary statistics	K_okino	K_toku	K_AF	Ksd_okino	Ksd_toku	Ksd_AF	mean_K	sd_K	tot_K
Observed Value	2.35	4.4	8.45	1.53125	4.41767	7.08947	5.06667	3.10416	9.4
Summary statistics	H_okino	H_toku	H_AF	Had_okino	Had_toku	Had_AF	mean_H	sd_H	tot_H
Observed Value	0.226868	0.324758	0.526077	0.268593	0.305237	0.338558	0.359234	0.152555	0.4863
Summary statistics	GW_okino	GW_toku	GW_AF	GWsd_okino	GWsd_toku	GWsd_AF	mean_GW	sd_GW	tot_GW
Observed Value	0.585317	0.607993	0.736569	0.241392	0.26451	0.163251	0.650293	0.0926739	0.722554
Summary statistics	NGW_okino	NGW_toku	NGW_AF	NGWsd_okino	NGWsd_toku	NGWsd_AF	mean_NGW	sd_NGW	
Observed Value	0.246524	0.382195	0.649745	0.197962	0.202091	0.224729	0.426155	0.205173	
Summary statistics	R_okino	R_toku	R_AF	Rad_okino	Rad_toku	Rad_AF	mean_R	sd_R	tot_R
Observed Value	6.38462	8.27778	12.444	7.98476	8.25908	11.4029	9.03561	3.10018	12.45
Summary statistics	FIS	FST	FIT						
Observed Value	0.0437566	0.299879	0.330513						
Summary statistics	FST_toku_okino	FST_AF_okino	FST_AF_toku						
Observed Value	0.343439	0.303634	0.271697						
Summary statistics	PI_toku_okino	PI_AF_okino	PI_AF_toku						
Observed Value	8.33941	10.9974	11.7025						
Summary statistics	DMUSQ_toku_okino	DMUSQ_AF_okino	DMUSQ_AF_toku						
Observed Value	1.69281	3.63525	6.54198						

Confusion matrix based on 100 samples for each model (neuralnet, tolerance rate=0.01).

model focal \ selected	1	2 (best model)	3	4	5
1 (Single origin, no gene flow)	0.85	0.08	0.01	0.05	0.01
2 (Single origin, gene flow)	0.23	0.29	0.16	0.17	0.15
3 (Parallel origin, gene flow)	0.02	0.07	0.37	0.24	0.3
4 (Parallel origin, gene flow)	0.08	0.1	0.22	0.37	0.23
5 (Parallel origin, gene flow)	0.04	0.06	0.17	0.15	0.58

Yellow mark indicates type2 error rate.

Mean model posterior probabilities (neuralnet, tolerance rate=0.01)

model focal \ selected	1	2 (best model)	3	4	5
1 (Single origin, no gene flow)	0.7214	0.1837	0.0197	0.0531	0.022
2 (Single origin, gene flow)	0.2064	0.3135	0.1605	0.184	0.1356
3 (Parallel origin, gene flow)	0.0195	0.1454	0.3236	0.2543	0.2572
4 (Parallel origin, gene flow)	0.08	0.1616	0.2381	0.2938	0.2265
5 (Parallel origin, gene flow)	0.0314	0.1381	0.2532	0.2236	0.3537

Bayes Factor (neuralnet, tolerance rate=0.01)

numerator \ denominator	1	2	3	4	5
1 (Single origin, no gene flow)	1.00	0.21	17.36	2.01	26.19
2 (Single origin, gene flow)	4.68	1.00	81.23	9.42	122.57
3 (Parallel origin, gene flow)	0.06	0.01	1.00	0.12	1.51
4 (Parallel origin, gene flow)	0.50	0.11	8.62	1.00	13.01
5 (Parallel origin, gene flow)	0.04	0.01	0.66	0.08	1.00

Table S8 Geographic data of respective islands.

Island Name	Freshwater form distribution	Latitude	Longitude	Area (km ²)	Area log10	Catchment area (km ²)	Catchment area log10	Length (m)	Length log	No. Steep Slopes
Okinawajima Is.	1	26°35'	128°2'	1206.93	3.082	53.30	1.727	17218	4.236	152
Amami-Oshima Is.	1	28°16'	129°20'	712.5	2.853	48.58	1.686	18478	4.267	548
Iriomotejima Is.	1	24°20'	123°50'	289.6	2.462	64.37	1.809	21736	4.337	435
Tokunoshima Is.	1	27°45'	128°57'	247.9	2.394	34.00	1.531	15754	4.197	304
Ishigakijima Is.	1	24°24'	124°11'	222.2	2.347	32.75	1.515	14750	4.169	101
Okinoerabujima Is.	1	27°22'	128°35'	93.7	1.972	10.84	1.035	8813	3.945	15
Kakeromajima Is.	0	28°7'	129°14'	77.3	1.888	2.99	0.475	3601	3.556	26
Kumejima Is.	1	26°22'	126°46'	59.5	1.775	6.99	0.844	5925	3.773	35
Kikajima Is.	0	28°18'	129°57'	57.0	1.755	4.81	0.683	4957	3.695	5
Iheyajima Is.	0	27°3'	127°58'	20.7	1.315	4.93	0.693	4428	3.646	2
Tokashikijima Is.	0	26°11'	127°21'	15.3	1.185	3.33	0.522	4029	3.605	19
Izenajima Is.	0	26°56'	127°56'	14.1	1.150	3.05	0.484	3918	3.593	3
Ukejima Is.	0	28°1'	129°15'	13.3	1.125	2.38	0.377	2275	3.357	21
Yorojima Is.	0	28°2'	129°9'	9.4	0.971	2.27	0.356	3257	3.513	12
Yagajijima Is.	0	26°40'	128°0'	7.8	0.893	1.16	0.063	2072	3.316	0
Miyagijima Is.	0	26°22'	128°0'	5.5	0.744	0.90	-0.048	1867	3.271	2
Akajima Is.	0	26°12'	127°17'	3.8	0.580	0.46	-0.333	1364	3.135	4
Hamahigajima Is.	0	26°19'	127°57'30"	2.1	0.320	0.36	-0.441	1297	3.113	0

Table S9 The results of statistical test and AIC model selection of logistic regression analysis for respective combinations of islands where speciation observed. 1 indicated speciation occurred, 0 indicated speciation did not occur.

Combinations of islands	Originate in respective regions				Share the origin in respective regions																
	Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5	Dataset 6	Dataset 7	Dataset 8	Dataset 9	Dataset 10	Dataset 11	Dataset 12	Dataset 13	Dataset 14	Dataset 15	Dataset 16	Dataset 17	Dataset 18	Dataset 19	Dataset 20	
Okinawajima Is.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
Amami-Oshima Is.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Tokunoshima Is.	1	1	0	0	1	1	0	0	0	0	1	0	1	1	1	0	0	1	1	0	0
Kumejima Is.	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1
Okinoerabujima Is.	0	0	1	1	0	0	1	1	0	0	0	1	0	1	0	1	1	0	0	1	1
Iriomotejima Is.	1	0	1	0	1	0	1	0	1	0	0	0	1	0	1	0	1	0	1	0	1
Ishigakijima Is.	0	1	0	1	0	1	0	1	0	1	0	0	0	1	0	1	0	1	0	1	1
Kakeromajima Is.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kikajima Is.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Tokashikijima Is.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Iheyajima Is.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Izenajima Is.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ukejima Is.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Yorojima Is.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Yagajima Is.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Miyajima Is.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Akajima Is.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Hamahigajima Is.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Maximum likelihood ratio	Explanatory variables	Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5	Dataset 6	Dataset 7	Dataset 8	Dataset 9	Dataset 10	Dataset 11	Dataset 12	Dataset 13	Dataset 14	Dataset 15	Dataset 16	Dataset 17	Dataset 18	Dataset 19	Dataset 20
<i>P</i> values	Area	0.0016	0.0026	0.0054	0.0061	0.0001	0.002	0.005	0.0064	0.0068	0.0093	0.0073	0.0253	0.024	0.0317	0.0534	0.0606	0.0342	0.046	0.0733	0.0868
	Catchment Area	0.0026	0.0069	0.008	0.0187	0.0003	0.0041	0.0046	0.0113	0.0105	0.025	0.0242	0.0637	0.0191	0.0322	0.0479	0.082	0.0214	0.0346	0.0518	0.0899
	River Length	0.0038	0.0046	0.0077	0.0172	0.0006	0.0031	0.0057	0.0115	0.0101	0.0294	0.0239	0.0596	0.0212	0.0307	0.0502	0.0789	0.0191	0.0295	0.0464	0.0729
	No. Waterfall	0.0013	0.0697	0.0449	0.3277	2.00E-04	0.0347	0.0203	0.2627	0.0177	0.2163	0.0628	0.3241	0.1269	0.587	0.5181	0.8131	0.0055	0.0879	0.0614	0.4257

Model Selection	Explanatory variables	Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5	Dataset 6	Dataset 7	Dataset 8	Dataset 9	Dataset 10	Dataset 11	Dataset 12	Dataset 13	Dataset 14	Dataset 15	Dataset 16	Dataset 17	Dataset 18	Dataset 19	Dataset 20
AIC scores	Island Area only	19.478	13.005	13.922	15.965	8.953	4	9.145	12.471	13.366	13.939	14.678	14.382	16.682	16.532	17.063	18.315	18.729	17.553	18.016	19.113
	Residuals of Catchment Area only	22.943	25.059	24.389	24.723	11.391	22.913	22.251	22.589	21.623	22.548	21.551	21.362	20.472	23.069	22.803	22.977	22.407	22.835	23.067	23.042
	Island Area + Residuals of Catchment Area	21.405	14.835	15.868	17.96	6	6	6	14.454	15.092	15.926	16.417	16.015	17.809	17.646	18.991	20.131	20.701	17.371	19.46	20.316
	Island Area + Residuals of Catchment Area	22.78	16.82	16.004	19.9	8	8	8	16.357	15.368	17.847	17.585	16.982	18.61	19.578	20.991	22.084	22.69	17.905	20.579	21.091
AIC scores	Island Area only	19.478	13.005	13.922	15.965	8.953	4	9.145	12.471	13.366	13.939	14.678	14.382	16.682	16.532	17.063	18.315	18.729	17.553	18.016	19.113
	Residual of River Length only	23.067	25.018	24.386	24.88	11.646	22.888	22.274	22.76	22.026	22.34	21.326	21.368	21.006	23.058	22.734	23.012	22.567	22.726	23.062	22.873
	Island Area + Residual of River Length	21.107	14.918	15.832	17.932	6	6	6	14.351	15.276	15.935	16.272	15.99	18.098	17.902	19.025	20.064	20.727	17.084	19.201	19.669
	Island Area + Residual of River Length	22.611	16.908	15.996	19.764	8	8	8	15.937	14.921	17.852	17.32	17.138	18.757	19.893	21.019	22.051	22.615	17.652	20.199	20.856
AIC scores	Island Area only	19.478	13.005	13.922	15.965	8.953	4	9.145	12.471	13.366	13.939	14.678	14.382	16.682	16.532	17.063	18.315	18.729	17.553	18.016	19.113
	Residuals of No. Waterfall only	20.894	24.196	20.881	22.692	9.13	22.551	19.758	21.339	16.818	21.889	18.335	20.969	18.718	20.977	15.981	18.731	9.806	23.03	22.305	22.913
	Island Area + Residuals of No. of Waterfall	21.288	13.13	14.906	17.963	10.952	6	6	14.442	6	15.538	15.159	16.38	17.679	18.52	15.828	19.161	11.775	14.038	19.719	20.019
	Island Area + Residuals of No. of Waterfall	15.168	15.081	16.756	19.49	8	8	8	16.087	8	17.538	16.986	18.377	18.832	20.438	16.985	21.052	13.672	8	18.523	14.982