

Population genetic structure and hybrid zone analyses for species delimitation in the Japanese toad (*Bufo japonicus*) (#87292)

1

First revision

Guidance from your Editor

Please submit by **14 Sep 2023** for the benefit of the authors .



Structure and Criteria

Please read the 'Structure and Criteria' page for general guidance.



Custom checks

Make sure you include the custom checks shown below, in your review.



Author notes

Have you read the author notes on the [guidance page](#)?



Raw data check

Review the raw data.



Image check

Check that figures and images have not been inappropriately manipulated.

If this article is published your review will be made public. You can choose whether to sign your review. If uploading a PDF please remove any identifiable information (if you want to remain anonymous).

Files

Download and review all files from the [materials page](#).

- 1 Tracked changes manuscript(s)
- 1 Rebuttal letter(s)
- 9 Figure file(s)
- 2 Table file(s)

! Custom checks

DNA data checks

- ! Have you checked the authors [data deposition statement](#)?
- ! Can you access the deposited data?
- ! Has the data been deposited correctly?
- ! Is the deposition information noted in the manuscript?

Vertebrate animal usage checks

- ! Have you checked the authors [ethical approval statement](#)?
- ! Were the experiments necessary and ethical?



Have you checked our [animal research policies](#)?


For assistance email peer.review@peerj.com



Structure your review

The review form is divided into 5 sections. Please consider these when composing your review:

1. BASIC REPORTING
2. EXPERIMENTAL DESIGN
3. VALIDITY OF THE FINDINGS
4. General comments
5. Confidential notes to the editor






 You can also annotate this PDF and upload it as part of your review

When ready [submit online](#).





Editorial Criteria

Use these criteria points to structure your review. The full detailed editorial criteria is on your [guidance page](#).




BASIC REPORTING

-  Clear, unambiguous, professional English language used throughout.
-  Intro & background to show context. Literature well referenced & relevant.
-  Structure conforms to [PeerJ standards](#), discipline norm, or improved for clarity.
-  Figures are relevant, high quality, well labelled & described.
-  Raw data supplied (see [PeerJ policy](#)).

EXPERIMENTAL DESIGN

-  Original primary research within [Scope of the journal](#).
-  Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
-  Rigorous investigation performed to a high technical & ethical standard.
-  Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

-  Impact and novelty not assessed. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
-  All underlying data have been provided; they are robust, statistically sound, & controlled.
-  Conclusions are well stated, linked to original research question & limited to supporting results.



The best reviewers use these techniques

Tip

Example

Support criticisms with evidence from the text or from other sources

Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.

Give specific suggestions on how to improve the manuscript

Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).

Comment on language and grammar issues

The English language should be improved to ensure that an international audience can clearly understand your text. Some examples where the language could be improved include lines 23, 77, 121, 128 - the current phrasing makes comprehension difficult. I suggest you have a colleague who is proficient in English and familiar with the subject matter review your manuscript, or contact a professional editing service.

Organize by importance of the issues, and number your points

1. Your most important issue
2. The next most important item
3. ...
4. The least important points

Please provide constructive criticism, and avoid personal opinions

I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC

Comment on strengths (as well as weaknesses) of the manuscript

I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.

Population genetic structure and hybrid zone analyses for species delimitation in the Japanese toad (*Bufo japonicus*)

Kazumi Fukutani ^{Corresp., 1}, Masafumi Matsui ¹, Kanto Nishikawa ^{1, 2}

¹ Graduate School of Human and Environmental Studies, Kyoto University, Kyoto, Japan

² Graduate School of Global Environmental Studies, Kyoto University, Kyoto, Japan

Corresponding Author: Kazumi Fukutani

Email address: fukutani.kazumi.55a@kyoto-u.jp

Hybridization following secondary contact may produce different outcomes depending on the extent to which genetic diversity and reproductive barriers have accumulated during isolation. The Japanese toad, *Bufo japonicus*, is distributed on the main islands of Japan. In the present study, we applied multiplexed inter-simple sequence repeat genotyping by sequencing to achieve the fine-scale resolution of the genetic cluster in *B. j. japonicus* and *B. j. formosus*. We also elucidated hybridization patterns and gene flow degrees across contact zones between the clusters identified. Using SNP data, we found four genetic clusters in *B. j. japonicus* and *B. j. formosus* and three contact zones of the cluster pairs among these four clusters. The two oldest diverged lineages, *B. j. japonicus* and *B. j. formosus*, formed a narrow contact zone consistent with species distinctiveness. Therefore, we recommend that these two subspecies be elevated to the species level. In contrast, the less diverged pairs of two clusters in *B. j. japonicus* and *B. j. formosus*, respectively, admixed over a hundred kilometers, suggesting that they have not yet developed strong reproductive isolation and need to be treated as conspecifics. These results will contribute to resolving taxonomic confusion in Japanese toads.

1 Population genetic structure and hybrid zone analyses for species delimitation in the Japanese toad (*Bufo*
2 *japonicus*)

3

4 Kazumi Fukutani¹, Masafumi Matsui¹, and Kanto Nishikawa^{1,2}

5

6 ¹Graduate School of Human and Environmental Studies, Kyoto University, Kyoto, Japan

7 ²Graduate School of Global Environmental Studies, Kyoto University, Kyoto, Japan

8

9 Corresponding Author:

10 Kazumi Fukutani¹

11 E-mail address: fukutani.kazumi.55a@kyoto-u.jp

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47 Abstract

48 Hybridization following secondary contact may produce different outcomes depending on the
49 extent to which genetic diversity and reproductive barriers have accumulated during isolation.
50 The Japanese toad, *Bufo japonicus*, is distributed on the main islands of Japan. In the present
51 study, we applied multiplexed inter-simple sequence repeat genotyping by sequencing to achieve
52 the fine-scale resolution of the genetic cluster in *B. j. japonicus* and *B. j. formosus*. We also
53 elucidated hybridization patterns and gene flow degrees across contact zones between the
54 clusters identified. Using SNP data, we found four genetic clusters in *B. j. japonicus* and *B. j.*
55 *formosus* and three contact zones of the cluster pairs among these four clusters. The two oldest
56 diverged lineages, *B. j. japonicus* and *B. j. formosus*, formed a narrow contact zone consistent
57 with species distinctiveness. Therefore, we recommend that these two subspecies be elevated to
58 the species level. In contrast, the less diverged pairs of two clusters in *B. j. japonicus* and *B. j.*
59 *formosus*, respectively, admixed over a hundred kilometers, suggesting that they have not yet
60 developed strong reproductive isolation and need to be treated as conspecifics. These results will
61 contribute to resolving taxonomic confusion in Japanese toads.

62

63 Introduction

64 Hybrid zones are natural laboratories that offer many insights into speciation processes, thereby
65 contributing to a more detailed understanding of evolution (Barton & Hewitt, 1985; Hewitt,
66 1988; Abbott et al., 2013). Hybridization following secondary contact may produce different
67 outcomes depending on the extent to which genetic diversity and reproductive barriers have
68 accumulated during isolation. This results in a reduction in differentiation as well as the fusion of
69 gene pools. Alternatively, an increase in the strength of the genetic barrier may lead to complete
70 reproductive isolation (Barton & Hewitt, 1985; Wu, 2001).

71 Hybridization is frequent and evolutionarily significant in amphibians (Burbrink and
72 Ruane, 2021). There are well-described studies on amphibians for the hybrid zone of fire-bellied
73 toads (*Bombina bombina* and *B. variegata*; e.g., Szymura & Barton, 1986, 1991; Yanchukov et
74 al., 2006), green toads (*Bufo viridis* subgroups; e.g., Stöck et al., 2006; Colliard et al., 2010;
75 Dufresnes et al., 2014), and, more recently, European common toads (*Bufo bufo* and *B. spinosus*;
76 e.g., Arntzen et al., 2016; Dufresnes et al., 2020; Riemsdijk et al., 2023). In contrast to many
77 other anuran species, the hybrid zone of Japanese toads has not yet been examined in detail (*Bufo*
78 *japonicus* subspecies; Miura, 1995). However, they have the advantage of comprising distinct
79 genetic lineages representing different stages of the speciation process because several contact
80 zones of the different genetic lineages have been recognized (Fukutani et al., 2022). Regarding
81 amphibian cases, the extent of natural hybridization in contact zones has been correlated with
82 divergence times (Hickerson, Meyer & Moritz, 2006; Dufresnes et al., 2021).

83 *Bufo japonicus* Temminck and Schlegel, 1838 is widely distributed in the Japanese
84 archipelago, Honshu, Shikoku, Kyushu, and some adjacent islands. This species is divided into
85 two subspecies, *B. j. japonicus* from western Japan and *B. j. formosus* Boulenger, 1883 from
86 eastern Japan. These two subspecies are parapatrically distributed with the boundary in the Kinki
87 region of central Japan (Matsui & Maeda, 2018). Matsui (1984) concluded that *B. j. japonicus*
88 and *B. j. formosus* showed a climatic cline in their morphometric characteristics, which was
89 insufficient to distinguish them as different species because of their identity in the fundamental
90 patterns of modes of life. However, Dufresnes & Litvinchuk (2021) recently proposed elevating
91 *B. j. japonicus* and *B. j. formosus* to the species level based on the Miocene split estimated by
92 mtDNA markers. However, they refrained from taxonomic changes because mitochondrial

93 distances may not reflect actual species distances. Other studies proposed the Kinki region as a
94 hybrid zone of *B. j. japonicus* and *B. j. formosus* by a C-banding analysis of chromosomes
95 (Miura, 1995).

96 The sympatric distribution of mitochondrial haplotypes of *B. j. japonicus* and *B. j.*
97 *formosus* was also found in the Kinki region (Fukutani et al., 2022). Furthermore, several contact
98 distributions of the genetic lineages in the two subspecies were identified. These findings
99 indicate the necessity of analyzing the degree of hybridization between the two subspecies and
100 other genetic lineages for taxonomic revision.

101 The delimitation of species must be connected to a species concept. We used the
102 integrative species concept (de Queiroz 2007, 2020) that considers both aspects, phylogeny and
103 the reproductive isolation mechanism.

104 In this study, we applied multiplexed ISSR genotyping by sequencing (MIG-seq; Suyama
105 & Matsuki, 2015) to achieve the fine-scale resolution of genetic clusters in *B. j. japonicus* and *B.*
106 *j. formosus*. MIG-seq has been effectively used to study molecular phylogenetic taxonomy for
107 various taxa (see Suyama et al., 2022).

108 We performed cline analyses to elucidate the degree of gene flows. The results of cline
109 analyses explained the transition between the characteristics of interbreeding species across the
110 hybrid zone and will contribute to a more detailed understanding of the mechanisms maintaining
111 species boundaries (Barton & Hewitt, 1985). Valid species need to exhibit significant divergence
112 and narrow transition zones. In contrast, insufficient diverged lineages that remained conspecific
113 need to admix freely across broad genetic areas. We revised the taxonomic status of *B. j.*
114 *japonicus* and *B. f. formosus* based on phylogenetic and hybrid zone analyses.

115

116 **Materials & Methods**

117 **Sampling and MIG-seq**

118 A total of 155 individuals of *B. japonicus* and 13 of *B. torrenticola* Matsui, 1976 were collected,
119 covering the complete distribution range (Table S1). The Animal Experimentation Ethics
120 Committee in the Graduate School of Human and Environmental Studies, Kyoto University
121 approved this research (20-A-5, 20-A-7, 22-A-2). DNA was extracted from frozen or ethanol-
122 preserved tissue samples (e.g., muscle, liver, or skin) with the Qiagen DNeasy Blood and Tissue
123 Kit following the manufacturer's instructions.

124 We prepared two genomic libraries and sequenced them separately for the convenience of
125 the experiment, and the data obtained were analyzed together as described below. Library 1
126 included 121 DNA samples of *B. japonicus* and 13 of *B. torrenticola*, while library 2 had 40
127 DNA samples of *B. japonicus*, with six of *B. japonicus* overlapping in both libraries (Table S1).
128 The two genomic libraries were prepared following the protocol described by Matsui et al.
129 (2019) for library 1 and that described by Watanabe et al. (2020) for library 2. Amplicons in
130 libraries 1 and 2 were purified and sequenced on the Illumina MiSeq Sequencer (Illumina San
131 Diego, CA, USA) using the MiSeq Reagent Kit v3 (150 cycles, Illumina). Two libraries were
132 prepared and sequenced separately for the convenience of the molecular experiment, and the raw
133 sequence data obtained were combined for subsequent data analyses.

134 The raw sequence reads of MIG-seq data were deposited in the DNA Data Bank of Japan
135 (DDBJ) Sequence Read Archive (DRA) under accession number DRA016475 (BioProject ID;
136 PRJDB15971: BioSample ID; SAMD00622809–SAMD00622982).

137 Raw paired-end sequences (reads 1 and 2) were filtered by fastp version 0.23.2 (Chen et
138 al., 2018) to trim the first 14 base sequences of read 2 and the primer regions of reads 1 and 2

139 and to discard reads shorter than 80 bp and low-quality sequences with phred quality $Q < 30$
140 according to Suyama and Matsuki (2015). We then mapped the filtered reads to reference
141 sequences because mapping obtains more loci than a *de novo* analysis of MIG-seq data (Takata
142 et al., 2021). As the reference genome sequence for Japanese toads, we used the genome
143 assembly of their closely related species, *B. gargarizans* (RefSeq assembly accession number:
144 GCF_014858855.1; https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_014858855.1/). The
145 assembly contained 11 chromosome-level contigs and unplaced scaffolds. We ultimately mapped
146 the filtered reads to the indexed reference sequences by bwa-mem2 version 2.2.1 (Vasimuddin et
147 al., 2019) to make SAMfiles, which were then converted to BAM files and sorted with a
148 minimum mapping quality of 20 using *samtools* version 1.15 (Li et al., 2009).

149

150 **Genotyping**

151 We prepared the following datasets: dataset I, data from samples of *B. japonicus* and *B.*
152 *torrenticola* to examine the genetic structure of Japanese toads, and dataset II, data from samples
153 of *B. j. japonicus* and *B. j. formosus* to investigate the degree of reproductive isolation between
154 the two subspecies. We excluded the 11 samples from these two datasets that were considered to
155 be from artificially introduced populations based on a previous study (Fukutani et al., 2022). We
156 instead prepared dataset III, which included these 11 samples with dataset II to verify their
157 genetic assignment in the population.

158 The reference-based analysis pipeline with the *gstacks* program followed by the
159 *populations* program in Stacks v2.60 (Rochette, Rivera-Colón & Catchen, 2019) was applied to
160 the mapped reads of all datasets to call SNPs and genotypes. The following filters were used for
161 the *populations* program in Stacks. We initially kept variant sites with a minimum allele count of
162 three (--min-mac 3) to ensure that an allele was in at least two diploid samples (Rochette, Rivera-
163 Colón & Catchen, 2019). We then set up the maximum observed heterozygosity at 0.5 (--max-
164 obs-het 0.50) because heterozygosity for a biallelic SNP was expected to be < 0.5 , and SNPs with
165 values above this threshold may belong to paralogous loci or multilocus contigs (Hohenlohe et
166 al., 2011; Willis et al., 2017). Subsequently, only one random SNP per locus was extracted (--
167 write-random-snp) to avoid any effect of linkages among SNPs on the multivariate analysis
168 (Gargiulo, Kull & Fay, 2020). In the population designation in a *population map*, we set two
169 populations corresponding to *B. japonicus* and *B. torrenticola* for dataset I. In datasets II and III,
170 we set two populations based on the admixture proportion (q -value, with q -value = 0.5 as a
171 boundary) at the optimal number of clusters (K) = 2 in the Structure analysis (see below:
172 Pritchard, Stephens & Donnelly, 2000) of dataset I. We ultimately only processed the loci
173 present in at least 80% of samples in a population ($-r = 0.80$) and those present in two
174 populations for all datasets ($-p = 2$). In the following stacks program, the two parameters, $-r$ and
175 p , were varied, and the others were common for each analysis.

176

177 **Estimation of genetic structures**

178 To estimate the population genetic structures of *B. japonicus* and *B. torrenticola*, we performed
179 three different methods using SNP genotyping information and compared grouping among these
180 methods: a discriminate analysis of principal components (DAPC; Jombart, Devillard & Balloux,
181 2010), Structure 2.3.4 (Pritchard, Stephens & Donnelly, 2000), and a principal component
182 analysis (PCA; Cavalli-Sforza, 1966). DAPC was used for the inference of the number of
183 clusters. We used Structure analyses to perform a Bayesian clustering analysis. In addition,
184 complementary to Structure analyses, we performed PCA.

185 DAPC was performed on dataset I in the R package *adegenet* 2.1.8 (Jombart, 2008;
186 Jombart, Devillard & Balloux, 2010; Jombart & Ahmed, 2011). This method maximizes the
187 variance among groups while minimizing variations within groups without making assumptions
188 about the underlying population genetic model. This approach transforms multilocus genotype
189 data using PCA to derive uncorrelated variables that are input for a discriminate analysis. The
190 optimal groups were initially assessed using the *de novo* clustering method, *find.cluster*, testing K
191 values from 1 to 8, and the best K value was selected with the Bayesian information criterion
192 (BIC) method. This *de novo* clustering method and initial DAPC using the *dapc* function were
193 run. The *optim.a.score* was then used to assess the optimal number of principal components
194 (PCs) to retain. Once the optimal number of PCs was selected, a second DAPC analysis was
195 conducted using this value.

196 The program Structure 2.3.4 (Pritchard, Stephens & Donnelly, 2000) performed the
197 analysis by an admixture model with correlated allele frequencies based on the Bayesian
198 clustering method to infer the population structure. Since excessive uneven sampling may
199 increase bias on admixture proportions in the Structure analysis (Toyama, Crochet & Leblois,
200 2020), we reduced the sample size in Yakushima and Tanegashima from dataset I, called dataset
201 I-2, and conducted Structure analyses. Structure analyses were performed for the number of
202 clusters K from 1 to 8, with ten runs for each K value. Markov chain Monte Carlo (MCMC;
203 Metropolis et al., 1953; Hastings, 1970) iterations of 100,000 were implemented for each run
204 after an initial burn-in of 100,000. The parallelization of Structure 2.3.4 calculations was
205 achieved using EasyParallel (Zhao et al., 2020) to reduce the computational time. The optimal
206 number of clusters was inferred in StructureSelector (Li & Liu, 2018) with the Delta K (ΔK ;
207 Evanno, Regnaut & Goudet, 2005), MedMeaK, MaxMeaK, MedMedK, and MaxMedK
208 (Puechmaille 2016). StructureSelector integrated the CLUMPAK program (Kopelman et al.,
209 2015) to cluster and merge data from independent runs and generate graphical representations of
210 the results. In a Structure analysis, an admixed ancestry is modeled by assuming that an
211 individual has inherited some proportion of its genome from its ancestors in the population
212 (Pritchard, Stephens & Donnelly, 2000).

213 PCA was performed on dataset I using the R package *adegenet* 2.1.8 (Jombart, 2008;
214 Jombart & Ahmed, 2011), and the first two eigenvectors were plotted in two dimensions.

215 Moreover, we conducted a Structure analysis of dataset III to identify the assignment of
216 genomic clusters for samples from introduced populations, reducing the sample size in
217 Yakushima and Tanegashima for the above reason as dataset III-2. A Structure analysis was
218 performed on the number of clusters K from 1 to 6, and other parameters were the same as
219 above.

220

221 **Phylogenetic estimations**

222 We used SNAPP 1.5.2 (Bryant et al., 2012) implemented in Beast v 2.6.7 (Bouckaert et al.,
223 2019) to estimate phylogenetic relationships among population groups identified by our
224 clustering. We selected four individuals for each population group and applied them to the stacks
225 program ($-r = 1.0$ and $-p = 5$). We ran SNAPP for 10,000,000 iterations with mutation rates u
226 and $v = 1.0$, a gamma distribution with $alpha = 2$ and $beta = 200$ for the lambda prior, and $alpha$
227 $= 1$, $beta = 250$, $kappa = 1$, and $lambda = 10$ for *snapprior*, sampling every 1,000 steps.
228 Convergence was examined using Tracer 1.7.2 (Rambaut et al., 2018), and the results obtained
229 were visualized by Densitree 2.2.7 with a burn-in of 10%. The maximum clade credibility tree
230 with posterior probability was calculated using TreeAnnotator version 2.6.7 (Bouckaert et al.,

231 2019). To perform comparisons, we reconstructed a mitochondrial phylogenetic tree using the
232 mitochondrial cytochrome *b* sequences from Fukutani et al. (2021) of the same individuals used
233 to construct the SNP tree adding the sequence of *B. g. gargarizans* as the outgroup. RaxML
234 version 8.2.12 (Stamatakis, 2014) was employed for 1,000 bootstrap iterations with the
235 GTRGAMMA model to infer a maximum likelihood phylogenetic tree based on mitochondrial
236 sequences.

237

238 **Effective estimates of migration surfaces**

239 We visualized the spatial patterns of gene flow using Fast Estimation of Effective Migration
240 Surfaces (FEEMS; Marcus et al., 2021) to assess the genomic context and geographic location
241 of any historical barriers to migration in *B. japonicus*. FEEMS is an improvement of Estimated
242 Effective Migration Surfaces (Petkova, Novembre & Stephens, 2016) and uses a Gaussian
243 Markov Random Field model in a penalized likelihood framework. This method uses locality
244 information and pairwise dissimilarity matrices calculated from SNP data to identify regions
245 where genetic similarity decays more quickly than expected under isolation by distance
246 (Petkova, Novembre & Stephens, 2016). To estimate effective migration parameters, we used the
247 genotype data of dataset II as well as the coordinate information of each individual as inputs. A
248 polygon grid was prepared using QGIS 3.28. Cross-validation was performed and the lambda
249 with the lowest cross-validation value was used to generate the final plot.

250

251 **Hybrid zone analyses**

252 To estimate the geographic gradient of genomic differences between adjacent clusters of *B.*
253 *japonicus*, we calculated the steepness of the cline of genetic differences. Assuming similar
254 dispersal abilities among the individuals of each cluster and no geographic barriers to gene flow
255 at their transitions, wide hybrid zones will be present for the younger pairs if they did not yet
256 evolve significant reproductive isolation. In contrast, narrow transitions will be present for the
257 older pairs if they represent distinct species.

258 We fit clines to the Structure *q*-value across the geographic transition between genetic
259 clusters using the R package *hzar* version 0.2-7 (Derryberry et al., 2014). The admixture
260 proportions inferred by the Structure program (Pritchard, Stephens & Donnelly, 2000) have
261 frequently been used to fit a geographic consensus cline, from which the width of the hybrid
262 zone is estimated (e.g., Tominaga et al., 2018; Dufresnes et al., 2020b). To avoid bias on the
263 admixture proportions of Structure, we also reduced the sample size in Yakushima and
264 Tanegashima from dataset II as dataset II-2. We also fit clines to mitochondrial haplogroup
265 frequency data from our previous study (Table S1; Fukutani et al., 2022) for comparison with
266 nuclear ancestry clines.

267 We performed the stacks program on this subset, setting four populations based on the
268 results of DAPC on dataset II, with $-r = 0.80$ and $-p = 4$, and conducted a Structure analysis
269 using the same parameters as above. This subset was divided into sub-datasets I, II, and III,
270 based on the *q*-value at $K = 4$ with some samples overlapping. Each sub-dataset contained
271 individuals of two pure clusters, considering a *q*-value > 0.90 as pure individuals and admixed
272 individuals between pure clusters. We applied the stacks program for each sub-dataset, setting
273 three populations (two pure and one admixed population) with $-r = 0.80$ and $-p = 3$, and
274 conducted a Structure analysis. The *q*-values on $K = 2$ for each sub-dataset were used to perform
275 *hzar*. In addition to the three sub-datasets, we prepared sub-dataset III-2, which is data excluding

276 samples in Shikoku and Seto Inland Sea (see discussion) and performed a similar analysis to that
277 for the other sub-datasets.

278 We reduced the two-dimensional space (latitude and longitude) into a single-dimensional
279 distance from the center line of the hybrid zone. The probable center line of the admixture was
280 estimated using R package *tess3r* version 1.1 (Caye, 2016, 2018) and considered to be the
281 baseline for *hzar*. The minimum distances from the baseline to individuals were calculated in
282 QGIS 3.28. We assigned a positive or negative sign to these distances depending on individual
283 orientations to the baseline.

284 The shape of a cline is modeled by combining three equations (Szymura & Barton 1986,
285 1991) that describe a sigmoid shape at the center of a cline (maximum slope) and two
286 exponential decay curves on either side of the central cline (tails). We tested 15 different models,
287 which combined three trait intervals and five fitting tails, for each cline plus a null model with no
288 cline. The three possible combinations of trait intervals were used to scale clines by the
289 minimum (p_{min}) and maximum (p_{max}) values in the cline: no scale (fixed to $p_{min} = 0$ and p_{max}
290 $= 1$), observed values (fixed to $p_{min} =$ minimum observed mean values, $p_{max} =$ maximum
291 observed mean values), and estimated values (p_{min} and p_{max} as the free parameter). The five
292 possible combinations of fitting tails represent the cline shapes: no tails, right tail only, left tail
293 only, symmetrical tails, mirror tails, and both tails estimated separately.

294 MCMC was performed for each model with the default values of 100,000 generations,
295 each with a randomly selected seed and 10% of steps discarded as a burn-in. After each run, we
296 compared the model performance using the Akaike information criterion score corrected for a
297 small sample size (AICc; Anderson & Burnham, 2002). The model with the lowest AICc score
298 was selected as the best-fit model to infer cline widths and centers along with a 95% confidence
299 interval (CI). The stability and convergence of the cline parameters of the best-fit model were
300 assessed by visualizing MCMC traces. We plotted the maximum-likelihood clines and 95%
301 credible cline region for the best-fit model.

302

303 **Introgression**

304 We assigned individuals in each contact zone to hybrid classes to estimate whether gene flow is
305 an ongoing or historic admixture. We temporarily designated individuals with q -values >0.98 for
306 $K = 2$ in the Structure analysis of sub-datasets I, II, and III as parental individuals for each cluster
307 following Scordato et al. (2017). We identified ancestry-informative markers by calculating
308 AMOVA F_{ST} for SNPs between pairs of parental clusters using the stacks program on the sub-
309 datasets, setting three populations (two parental and one admixed population) and $-r = 0.80$ and
310 $-p = 3$. The diagnostic loci, $F_{ST} = 1$, were selected as ancestry-informative markers segregating
311 between each pair of parental clusters.

312 We used the R package INTROGRESS version 1.2.3 (Gompert & Buerkle, 2010) to
313 calculate the maximum-likelihood estimates of the hybrid index for each individual and the
314 average heterozygosity of each individual across informative loci. We compared genomic hybrid
315 indices with heterozygosity to identify the individual hybrid classes. Pure individuals were
316 defined by a hybrid index of 0 or 1 because only loci fixed in parental individuals with $F_{ST} = 1$
317 were used. First-generation hybrids (F1) have an expected hybrid index of 0.5 and
318 heterozygosity of 1.0. We regarded individuals with intermediate hybrid indices (>0.25 and
319 <0.75) and high heterozygosity (≥ 0.5) as recent-generation hybrids, those with intermediate
320 hybrid indices (>0.25 and <0.75) and low heterozygosity (< 0.5) as later-generation hybrids, and

321 those with low hybrid indices (≤ 0.25 or ≥ 0.75) as backcrossed to one or the other parental type
322 according to previous studies (Milne & Abbott, 2008; Scordato et al., 2017; Slager et al., 2020).

323

324 **Estimation of migration rates**

325 Recent migration rates between parental and hybrid populations were calculated using the
326 Bayesian inference approach by BayesAss3-SNPs v 1.1 (Wilson & Rannala, 2003; Mussmann et
327 al., 2019). Using sub-datasets I, II, and III after applying for the stacks program with each setting
328 for three populations (two parental and one admixed population) and $-r = 0.80$ and $-p = 3$,
329 BA3-SNPs -autotune v2.1.2 (Mussmann et al., 2019) was performed with the default parameters
330 to find mixing parameters for BA3-SNPs. BayesAss3-SNPs was conducted with ten million
331 generations sampling every 100 generations using predefined mixing parameters. The first one
332 million generations were discarded as a burn-in, and chain convergence was assessed in Tracer v
333 1.7.2 (Rambaut et al., 2018).

334 All analyses by R were conducted in R studio version 2022.07.2.576 (Rstudio Team
335 2022) using R version 4.2.2 (R Core Team 2022).

336

337 **Results**

338 **Analyses of MIG-seq data**

339 A total of 46,889,160 clean reads in 168 samples passed quality filtering, with the average
340 percentage of reads that passed filtering for each sample being 77.6%. Among them, 17,644,888
341 reads were successfully mapped to the reference genome of *B. gargarizans* in the reference-
342 mapping approach with an average mapping quality of 27.2%.

343

344 **Genetic structure and phylogeny**

345 A total of 839 variants were identified in dataset I of 157 samples of *B. japonicus* and *B.*
346 *torrenticola*.

347 We retained all information (157 PCs) for the initial DAPC on dataset I. After running
348 the initial steps, the first 21 PCs were retained following the result of the *optim.a.score* function
349 (Fig. S1A). The BIC plot in DAPC displayed the lowest value at $K = 4$ and 5 (Fig. S1B), and
350 both clearly identified three clusters corresponding to *B. j. formosus*, *B. j. japonicus*, and *B.*
351 *torrenticola*. The results of $K = 4$ identified two subclusters within *B. j. japonicus*. In addition,
352 two subclusters within *B. j. formosus* were recognized for $K = 5$. However, these defined
353 subclusters within *B. j. japonicus* and *B. j. formosus* had markedly overlapping plots between
354 subclusters (Fig. 1A).

355 A total of 570 variants were identified in dataset I-2 of 131 samples. A Structure analysis
356 of dataset I-2 supported two peaks for the ΔK estimation, $K = 2$ and 5 (Fig. S2A), and the
357 number of K estimated from MedMeaK, MaxMeaK, MedMedK, and MaxMedK values was 5
358 (Fig. S2B).

359 Therefore, $K = 5$ may be the valid cluster number in our results, leading to a similar grouping
360 pattern to the DAPC (Fig. 1B). The five genetic clusters identified by DAPC and Structure
361 analyses corresponded to northern *B. j. formosus* (NF), southern *B. j. formosus* (SF), eastern *B. j.*
362 *japonicus* (EJ), western *B. j. japonicus* (WJ), and *B. torrenticola*. Cluster assignments for
363 individuals by DAPC are shown in Fig. 1A-2 and Table S1.

364 The Structure bar plot revealed that *B. torrenticola* has rare admixtures with *B. japonicus*,
365 three samples of *B. torrenticola* had q -values from 0.85 to 0.9, and one sample (Sample ID:
366 ALC8; Table S1) of *B. j. formosus* had a q -value of 0.09 admixed with *B. torrenticola*. These

367 samples appeared to be hybrid individuals based on the q -value threshold following Vähä &
368 Primmer (2006). Therefore, the admixed sample of *B. j. formosus* was excluded from the
369 subsequent analysis of *B. japonicus* (datasets II, II-2, III, and III-2 and all sub-datasets).

370 Structure assignments also revealed hybridization between each adjacent cluster of *B.*
371 *japonicus* (Fig. 1B). The admixture proportion assignment for each cluster of *B. japonicus*
372 changed in steps. High levels of continuous admixtures were indicated across the geographic
373 transition between NF and SF and between EJ and WJ. In contrast, hybrid individuals were
374 limited to the boundary between SF and EJ.

375 The first PC axis explained 25.1% of the genomic covariance in PCA. It separated the
376 two subspecies, *B. j. formosus* and *B. j. japonicus* (Fig. 1C). By the second PC axis, *B.*
377 *torrenticola* had clearly split from *B. japonicus*. In addition, the second axis separated two
378 continuous clusters within *B. j. japonicus*.

379 Based on SNAPP (290 SNPs), nuclear phylogeny confirmed deep splits between the five
380 main clades (Fig. 2A). Mitochondrial cytochrome *b* phylogeny (1,071 bp) recovered the splits of
381 the main clades confirmed in Fukutani et al. (2022; Fig. 2B).

382

383 **Artificially introduced population**

384 A total of 718 variants were identified in the 128 samples of dataset III-2, *B. japonicus*, including
385 the 11 samples from the artificially introduced populations in Hokkaido, Izu Islands, and the
386 Kanto region. Two individuals in Hokkaido (Asahikawa and Hakodate) had an admixture,
387 mainly two clusters of NF and SF, similar to those in Niijima and Kouzushima (Fig. S3).

388 Individuals in Tokyo and Kanagawa prefectures had four admixed clusters of NF, SF, EJ, and
389 WJ. The individual in Oshima had three clusters of SF, EJ, and WJ, and one in Hachiojima had
390 clusters of SF and EJ.

391

392 **Effective estimates of migration surfaces**

393 A total of 783 variants were identified in dataset II, 143 samples of *B. j. japonicus* and *B. j.*
394 *formosus*. The estimated effective migration rates confirmed low migration rates between *B. j.*
395 *formosus* and *B. j. japonicus* despite the absence of any geographic barrier that limits gene flow
396 between subspecies (Fig. 3). Among *B. j. japonicus*, low migration rates were detected between
397 Chugoku and Shikoku vs. Kyushu, and Kyushu vs. Yakushima, which appeared to be due to the
398 presence of straits. In contrast, high migration rates were detected within them. On the other
399 hand, among *B. j. formosus*, low migration rates were widely identified from Tohoku to Chubu,
400 likely due to fewer interactions between regions than among *B. j. japonicus*.

401

402 **Hybrid zone analyses**

403 Each sub-dataset consisted of cluster pairs, sub-dataset I (NF–SF) of 47 samples, sub-dataset II
404 (SF–EJ) of 47 samples, sub-dataset III (EJ–WJ) of 59 samples, and sub-dataset III-2 (EJ–WJ
405 excluding samples in Shikoku and Seto Inland Sea) of 48 samples. The geographic distribution
406 of each cluster detected by *tess3r* on each sub-dataset ($K = 2$) did not significantly differ from
407 that of Structure analyses by SNP data. The baselines across the three contact zones are shown in
408 Fig. 4. Regarding the SNP data of sub-dataset SF–EJ and the mtDNA data of EJ–WJ, the best-
409 supported model in *hzar* with the lowest AICc was that in which scaling was fixed to the
410 minimum value of 0 and maximum value of 1, and no exponential tails were desired. In sub-
411 datasets NF–SF and EJ–WJ and the mtDNA data of SF–EJ and NF–SF, the model selected was

412 that in which scaling was fixed to the minimum and maximum observed mean values, and no
413 exponential tails were desired.

414 Based on SNP data, the cline width decreased from NF–SF 170 (CI: 82–362) km to EJ–
415 WJ 162 (CI: 63–330) km and SF–EJ 29 (CI: 24–76) km (Fig. 5). The estimated centers based on
416 SNP data as the distance from the baseline were 0.6 (CI: -9.5–12) km for SF–EJ, 5.4 (CI: -42–
417 58) km for NF–SF, and 7.0 (CI: -40–56) km for EJ–WJ.

418 Based on mtDNA data, the cline width decreased from NF–SF 86 (CI: 35–223) km to
419 EJ–WJ 75 (CI: 31–212) km and SF–EJ 39 (CI: 18–106) km (Fig. 5). The estimated centers
420 based on mtDNA data as distances from the baseline were 0.3 (CI: -12–15) km for SF–EJ, 23
421 (CI: -12–64) km for NF–SF, and -33 (CI: -75–6.2) km for EJ–WJ.

422 In addition, in the sub-dataset EJ–WJ excluding samples in Shikoku and Seto Inland Sea
423 (48 samples), the model selected for SNP and mtDNA data was that in which scaling was fixed
424 to the minimum and maximum observed mean values, and no exponential tails were desired.
425 Based on SNP data, the width was 99 (CI: 33–301) km and the distance from the baseline was -
426 1.2 (CI: -38–53) km. Based on mtDNA data, the width was 79 (CI: 32–245) km and the distance
427 from the baseline was -33 (CI: -76–2.0) km (Fig. 5).

428

429 **Introgression**

430 We identified loci that were informative for assigning hybrid classes for each sub-dataset. There
431 were 40 loci with $F_{ST} > 1.0$ between parental SF and EJ, and six loci for the NF and SF pair and
432 EJ and WJ pair. Comparisons of individual hybrid indices and average heterozygosity using
433 these differentiated loci revealed that none of the pairs contained F1 individuals (Fig. 6). Recent-
434 generation hybrids with high heterozygosity were detected in the NF–SF contact zone only,
435 confirming ongoing gene flow. Later-generation hybrids were detected in all contact zones, and
436 hybrid individuals with intermediate hybrid index values and heterozygosity of zero were
437 identified in NF–SF and EJ–WJ contact zones, suggesting that old-origin hybrids survived.
438 Backcrossed individuals with both parental populations were identified in the NF–SF and SF–EJ
439 contact zones, while those with one parental population were detected in the EJ–WJ contact
440 zone.

441

442 **BayesAss directional migration**

443 The mixing parameters for migration rates ($-m$), allele frequencies ($-a$), and inbreeding
444 coefficients ($-f$) were selected using BA3-SNPs-autotune for each sub-dataset: sub-dataset NF–
445 SF, $-m = 0.2125$, $-a = 0.55$, $-f = 0.1$; sub-dataset SF–EJ, $-m = 0.2125$, $-a = 0.55$, $-f = 0.1281$; sub-
446 dataset EJ–WJ, $-m = 0.1563$, $-a = 0.325$, $-f = 0.1$.

447 All estimated migration rates between populations are shown in Table 1. In the sub-
448 dataset NF–SF, the self-recruitment estimate of the parental population of SF was high at >95%,
449 while those of the parental population of NF and the hybrid population were slightly lower (90–
450 95%). Northward migration rates through the hybrid zone, from the parental SF to the hybrid
451 (5.9%) and from the hybrid to the parental NF (3.6%), were higher than migration rates in the
452 opposite direction, from the parental NF to the hybrid (1.5%) and from the hybrid to the parental
453 SF (1.7%).

454 In the sub-dataset SF–EJ, self-recruitments within both parental populations were
455 estimated to be high at >95%. In contrast, the hybrid population had low self-recruitment rates at
456 76.2%. Correspondingly, outward migration rates from the hybrid population into parental
457 populations were low (2.0% to parental SF and 1.3% to parental EJ efflux). In contrast,

458 migration rates into hybrid populations were high (16.7% from parental SF and 7.1% from
459 parental EJ influx).

460 In the sub-dataset EJ–WJ, the self-recruitment of both parental populations was high at
461 >95%, while that of the hybrid population was intermediate at 80.1%. The estimations of
462 migration rates from hybrids into both parental populations were low (1.3% to parental EJ and
463 1.4% to parental WJ efflux). In contrast, migration rates into hybrid populations were high (7.3%
464 from parental EJ and 12.6% from parental WJ influx). The migration rates among each parental
465 population were estimated to be very low, ranging between 1.3 and 2.6%.

466

467 Discussion

468 Genetic clustering and phylogeny

469 Previous studies reported that Japanese toads diverged into six mitochondrial lineages from the
470 late Miocene to the middle Pleistocene (Igawa et al., 2006; Fukutani et al., 2022). The two
471 subspecies, *B. j. japonicus* and *B. j. formosus*, were recommended for elevation to the species
472 level given their Miocene split. However, the findings of these studies were insufficient for the
473 taxonomic conclusion because they were based solely on mitochondrial analyses (Dufresnes &
474 Litvinchuk, 2021). Given the contact between the distribution zones of the two subspecies
475 (Fukutani et al., 2022) and the possible presence of a hybrid zone between them (Miura, 1995),
476 identifying the status of the zone is necessary for the study of the taxonomic status of Japanese
477 toads because we followed the integrative species concept that considers phylogeny and
478 reproductive isolation.

479 We used SNP markers of samples covering virtually the complete distribution ranges of
480 *B. j. japonicus*, *B. j. formosus*, and *B. torrenticola* and presented the clustering and phylogenetic
481 relationship between the identified clusters. We then showed the results of a fine-scale analysis
482 of gene flow across the secondary contact zones of *B. j. formosus* and *B. j. japonicus*.

483 The consensus across independent methods suggested that $K = 5$ most accurately
484 described the population structure of *B. japonicus* and *B. torrenticola*. This SNP clustering was
485 roughly concordant with the five main mitochondrial clades in Fukutani et al. (2022), except for
486 the lesser diverged mitochondrial clade in the Tohoku region (NF). Based on SNP data,
487 phylogeny confirmed the splits between the five main clades. However, the topology was
488 discordant with the mitochondrial phylogenetic topology for the clades in western Japan (Fig. 2).
489 The SNP phylogenetic tree showed EJ and WJ as sister clades and supported the monophyly of
490 *B. j. japonicus*. However, in the mitochondrial phylogenetic tree, *B. j. japonicus* was paraphyletic
491 because *B. torrenticola* and WJ were identified as sister clades with a high node support (Fig. 2).
492 One explanation is that *B. torrenticola* and the ancestor of EJ and WJ may all simultaneously
493 diverge. Alternatively, discordance may stem from ancestral mitochondrial introgression
494 between *B. torrenticola* and WJ after they diverged. These hypotheses need to be tested
495 explicitly in future phylogenetic studies.

496

497 The hybrid zone between *B. j. japonicus* and *B. j. formosus*

498 We found that mitochondrial and SNP marker cline positions and shapes varied for the three
499 contact zones between four clusters of *B. j. japonicus* and *B. j. formosus* and showed different
500 patterns of gene flow.

501 The hybrid zone between *B. j. japonicus* and *B. j. formosus* showed a sharp genetic
502 transition, with concordant and coinciding clines between mtDNA and SNP (Fig. 5B). The cline
503 width depended, in part, on whether the hybrid zone was structured primarily by selection or by a

504 neutral process (Mallet et al., 1990). The cline width without any form of selection may be
505 calculated using the following diffusion approximation from Barton & Gale (1993): $w = 2.51\sigma$
506 \sqrt{T} , where w , cline width, T , number of generations since secondary contact, and σ , average
507 lifetime dispersal. While the lifetime dispersal distance for *B. japonicus* currently remains
508 unknown, the maximum dispersal distance recorded for native *B. j. formosus* between the
509 breeding pond and the summer home range is 0.26 km and the generation time is three years
510 (Kusano, Maruyama & Kaneko, 1995). At a dispersal of 0.78 km per generation, cline width
511 exceeds the 29.4 km width of the hybrid zone in *ca.* 677 years of unrestricted diffusion. Based on
512 their paleo distribution, these toads came into contact with expansion after the last glacial period
513 at the latest (Fukutani et al., 2022). Therefore, contact between the subspecies is arguably
514 markedly older than 677 years. The cline width may have been kept narrow over a long time
515 despite the absence of geographic barriers to dispersal, presumably through selection against
516 hybrids, suggesting that the two subspecies formed a tension zone (Key, 1968) in the Kinki
517 region. In addition, all hybrid individuals were classified by INTROGRESS as layer-generation
518 hybrids or backcrosses, suggesting the relatively ancient origin of their contact.

519


520 **The hybrid zone within *B. j. japonicus***

521 Based on the refugia distributions proposed by Fukutani et al. (2022), the mitochondrial
522 boundary of EJ and WJ may have been maintained at the western edge of the Chugoku region
523 from the last glacial period to the present. Therefore, EJ and WJ likely shared refugia during the
524 glacial period, resulting in admixture. Admixed individuals may have spread to the Shikoku
525 region and surrounding islands through the Seto Inland Sea, which covered a terrestrial and
526 freshwater environment due to the lower sea level during the glacial period until 13,000 years
527 ago between the western part of Chugoku and Shikoku regions (Yashima, 1994).

528 While the strait between the Chugoku and Kyushu regions formed 8,000 years ago
529 (Yashima, 1994), which was later than that between the Chugoku and Shikoku regions, admixed
530 individuals were identified in the Shikoku region, but not in the Kyushu region, suggesting
531 asymmetric introgression. Furthermore, this asymmetric introgression may have resulted in
532 discordance in mtDNA and nuclear cline positions between EJ and WJ (Fig 5C), where the
533 mitochondrial cline center shifted approximately 40 km west from the nuclear cline center, with
534 partially overlapping CI. The incongruity of clines inferred from different sets of molecular
535 markers is a common phenomenon of terrestrial vertebrate hybrid zones, including amphibians
536 (e.g., Dufresnes et al., 2014; Arntzen et al., 2017; Sequeira et al., 2020). Prezygotic or
537 postzygotic effects may explain the discordance in mtDNA and nuclear cline position. Sex-
538 biased asymmetries (Toews & Brelsford, 2012) and an environmental gradient acting on mtDNA
539 (Cheviron & Brumfield, 2009) as prezygotic effects and Haldane's rule (Haldane, 1922; Orr,
540 1997) and Dobzhansky–Muller incompatibilities (Dobzhansky, 1936; Muller, 1942) as
541 postzygotic effects may have produced discordance in mtDNA and nuclear clines. Future field
542 and genomic studies are needed to test these hypotheses and identify the factors that caused
543 admixed individuals to spread mainly to the east of the mtDNA boundary at the time of
544 secondary contact during the glacial period.

545 Regarding the width of the cline, including Shikoku, 20,490 years are needed to reach
546 161.8 km using the above formula, and the width of the cline, not including Shikoku, is 88.6 km
547 which requires 6,144 years to reach, suggesting that selection may not act specifically in the
548 Shikoku region. Furthermore, the range of present suitable habitats for EJ and WJ in Fukutani et
549 al. (2022) was consistent with the actual distribution boundaries within the Chugoku region,

550 indicating exogenous environmental factors. However, Matsui (1984) did not identify
551 morphological differences between EJ and WJ. Moreover, the distribution of hybrid individuals
552 in the Shikoku region suggests that EJ and WJ are the same species, despite the different degrees
553 of admixture on the transect in the Chugoku and Shikoku regions.

554 The toad population in Yakushima was once considered to be a different subspecies of *B.*
555 *japonicus* (as vulg:  Okada, 1928), but is now recognized as the same species based on
556 morphology (Matsui 1984). Based on mitochondrial phylogeny in a previous study (Fukutani et
557 al., 2022), morphologically defined groups were not monophyletic and did not form a single
558 cluster in this study. There may have been interbreeding between the Kyushu, Yakushima, and
559 Tanegashima populations when the straits between Yakushima, Tanegashima and Kagoshima
560 were terrestrial during the glacial period (Ikehara, 1992). Geographic isolation after the last
561 glacial period may have led to the deviation from isolation by distance (Fig. 2).

562 563 **The hybrid zone within *B. j. formosus***

564 We identified the hybrid zone between NF and SF as the widest in the present study (Fig. 5A),
565 which was an expected result because of their recent evolutionary histories (Fukutani et al.,
566 2022). Widespread gene flow and recent hybridization indicate the absence of endogenous
567 reproductive barriers between NF and SF. Furthermore, the mtDNA and SNP clines between NF
568 and SF had an almost concordant center (Fig. 5A), suggesting the absence of selection (Toews &
569 Brelsford, 2012). In contrast, the SNP cline was wider than the mitochondrial cline across the
570 transition between NF and SF due to the lower effective population size of mitochondrial DNA
571 than nuclear markers (Toews & Brelsford, 2012).

572 The time needed to reach the 170-km width of the SNP cline between NF and SF without
573 selection was calculated to be 22,619 years, suggesting a prominent role for neutral processes.
574 According to our previously predicted distributions during the glacial period, NF and SF may
575 have shared their refugia around the southern Tohoku to northern Kanto regions (Fukutani et al.,
576 2022). The expansion of distribution after the last glacial period may have led to widespread
577 hybridization. An expansive hybrid zone consisting of late-generation hybrids and backcrosses is
578 consistent overall with a prolonged period of neutral expansion. Although we did not find any
579 asymmetry for the hybrid class assignment in the triangle plots (Fig. 6), the results obtained on
580 the direction of migration were predominantly from SF to NF through the hybrid population
581 (Table 1), indicating that this hybrid zone will lead to the formation of a hybrid swarm in the
582 future.

583 584 **Taxonomic revision of *B. japonicus***

585 Based on the above discussion, we reviewed the taxonomy of *B. japonicus*. SNP clustering based
586 on DAPC supported four cluster numbers for *B. japonicus*, and nuclear phylogeny according to
587 SNAPP confirmed deep splits between the five main clades.

588 However, based on PCA, these defined subclusters had markedly overlapping plots
589 between NF and SF and between EJ and WJ. Additionally, hybrid zone analyses between NF and
590 SF and between EJ and WJ indicated weak or no selection against hybrids that was insufficient
591 for them to be regarded as different species.

592 In contrast, at the hybrid zone between *B. j. japonicus* and *B. j. formosus*, there was
593 sufficient selection against hybrids for them to be regarded as different species. Hybridization
594 persisted over time as parentals moved into the hybrid zone (Table 1). In contrast, introgression
595 was limited by negative selection against hybrids (Table 1), allowing species to maintain their

596 genetic distinctiveness (Barton & Hewitt 1985). These results call for a taxonomic revision of *B.*
597 *j. japonicus* and *B. j. formosus*. Therefore, we consider the eastern Japanese common toad *B.*
598 *formosus* as a distinct species as originally described (Boulenger, 1883) and not a subspecies of
599 the western Japanese common toad *B. japonicus* as currently considered (e.g., Matsui & Maeda,
600 2018).

601 We validated the two Japanese common toads, the western Japanese toad *Bufo japonicus*
602 Temminck and Schlegel, 1838 (type locality: Japan) distributed in south-western Japan, and the
603 eastern Japanese toad *Bufo formosus* Boulenger, 1883 (type locality: Yokohama, Japan)
604 distributed in north-eastern Japan.

605 Morphometric variation analyses of these two species were conducted by Matsui (1984).
606 However intermediate forms were not detected in the Kinki region, and the morphological
607 boundary extended more westerly to the Chugoku region (Matsui, 1984). The discordant patterns
608 in morphological and genetic markers warrant further study.

609 Speciation with gene flow is common in anurans (Dufresnes et al. 2021). For example, a
610 previous study on two European *Bufo* species, *B. bufo* and *B. spinosus*, which diverged in the
611 Late Miocene, showed limited gene flow across a narrow hybrid zone (width of approximately
612 30 km) in the northwest of France even with the absence of barriers to dispersal (Arntzen et al.,
613 2016). Despite the presence of a hybrid zone for *B. formosus* and *B. japonicus*, the identity of the
614 parental species is distinctive and appears to have been unaffected. These two species were
615 considered to remain in partial reproductive isolation over a long period (Servedio and
616 Hermisson, 2020). Cline coupling may have progressed further towards reproductive isolation
617 after secondary contact, and it may still be ongoing throughout the hybrid zone (Harrison &
618 Larson, 2014; Butlin & Smadja, 2018).

619 We also found that the geographic location of the hybrid zone between the two species
620 appeared to be independent of the environment. Ecological niche modeling in Fukutani et al.
621 (2022) showed that environmental conditions were suitable for both species across the hybrid
622 zone identified in this study, suggesting that environment-associated selection may not act
623 directly to keep the hybrid zone. Many anuran speciation processes are initiated through the
624 gradual accumulation of multiple barrier loci scattered across the genome, which reduces hybrid
625 fitness by intrinsic postzygotic isolation (Dufresnes et al., 2021). Similarly, for *B. formosus* and
626 *B. japonicus*, many genomic regions may represent local barriers to gene flow. We will attempt
627 to elucidate the genomic mechanism that induces speciation in future studies.

628

629 **Conclusion**

630 In summary, we presented three hybrid zones with different cline shapes. Populations with
631 greater divergence had a sharper hybrid zone cline. These results on Japanese toads are
632 consistent with other findings on anuran species (e.g., Dufresnes et al., 2018, 2020c). They are
633 also very applicable to the most deeply diverged populations, *B. japonicus* and *B. formosus*,
634 which had a sharp cline, suggesting the presence of strong selection (Mallet et al., 1990). Our
635 results will contribute to resolving taxonomic confusion in Japanese toads.

636

637 **Acknowledgments**

638 We acknowledge K. Eto, I. Fukuyama, R. Fukuyama, S. Ikeda, K. Kimura, Y. Misawa, Y.
639 Miyagata, T. Shimada, T. Sugahara, T. Sugihara, Y. Tahara, S. Tanabe, A. Tominaga, N.
640 Yoshikawa, and many more collaborators for collecting samples. We thank N. Yoshikawa and
641 Y. Fuke for helping to conduct MIG-seq and analyses. We also thank our laboratory members

642 for helping with specimen processing and molecular experiments. Finally, we thank the
643 reviewers for their valuable comments.

644

645 References

- 646 Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N, Boughman J, Brelsford A,
647 Buerkle CA, Buggs R, Butlin RK, Dieckmann U, Eroukhmanoff F, Grill A, Cahan SH,
648 Hermansen JS, Hewitt G, Hudson AG, Jiggins C, Jones J, Keller B, Marczewski T,
649 Mallet J, Martinez-Rodriguez P, Möst M, Mullen S, Nichols R, Nolte AW, Parisod C,
650 Pfennig K, Rice AM, Ritchie MG, Seifert B, Smadja CM, Stelkens R, Szymura JM,
651 Väinölä R, Wolf JBW, Zinner D. 2013. Hybridization and speciation. *Journal of*
652 *Evolutionary Biology* 26:229–246. DOI: 10.1111/j.1420-9101.2012.02599.x.
- 653 Anderson DR, Burnham KP. 2002. Avoiding pitfalls when using information-theoretic methods.
654 *The Journal of Wildlife Management* 66:912. DOI: 10.2307/3803155.
- 655 Arntzen JW, Trujillo T, Butôt R, Vrieling K, Schaap O, Gutiérrez-Rodríguez J, Martínez-Solano
656 I. 2016. Concordant morphological and molecular clines in a contact zone of the
657 Common and Spined toad (*Bufo bufo* and *B. spinosus*) in the northwest of France.
658 *Frontiers in Zoology* 13:52. DOI: 10.1186/s12983-016-0184-7.
- 659 Arntzen JW, Vries W, Canestrelli D, Martínez-Solano I. 2017. Hybrid zone formation and
660 contrasting outcomes of secondary contact over transects in common toads. *Molecular*
661 *Ecology* 26:5663–5675. DOI: 10.1111/mec.14273.
- 662 Barton NH, Gale K. 1993. Genetic analysis of hybrid zones. In: *Hybrid zones and the*
663 *evolutionary process*. New York: Oxford University Press, 13–45.
- 664 Barton NH, Hewitt GM. 1985. Analysis of hybrid zones. *Annual Review of Ecology and*
665 *Systematics* 16:113–148. DOI: 10.1146/annurev.es.16.110185.000553.
- 666 Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, Heled J,
667 Jones G, Kühnert D, Maio ND, Matschiner M, Mendes FK, Müller NF, Ogilvie HA,
668 Plessis L du, Poppinga A, Rambaut A, Rasmussen D, Siveroni I, Suchard MA, Wu C-H,
669 Xie D, Zhang C, Stadler T, Drummond AJ. 2019. BEAST 2.5: an advanced software
670 platform for Bayesian evolutionary analysis. *PLoS Computational Biology* 15:e1006650.
671 DOI: 10.1371/journal.pcbi.1006650.
- 672 Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A. 2012. Inferring species
673 trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent
674 analysis. *Molecular Biology and Evolution* 29:1917–1932. DOI:
675 10.1093/molbev/mss086.
- 676 Burbrink, F. T., & Ruane, S. 2021. Contemporary Philosophy and Methods for Studying
677 Speciation and Delimiting Species. *Ichthyology & Herpetology* 109:874–894. DOI:
678 10.1643/h2020073.
- 679 Butlin RK, Smadja CM. 2018. Coupling, reinforcement, and speciation. *The American Naturalist*
680 191:155–172. DOI: 10.1086/695136.
- 681 Cavalli-Sforza LL. 1966. Population structure and human evolution. *Proceedings of the Royal*
682 *Society of London. Series B. Biological Sciences* 164:362–379. DOI:
683 10.1098/rspb.1966.0038.
- 684 Caye K, Deist TM, Martins H, Michel O, François O. 2016. TESS3: fast inference of spatial
685 population structure and genome scans for selection. *Molecular Ecology Resources*
686 16:540–548. DOI: 10.1111/1755-0998.12471.

- 687 Caye K, Jay F, Michel O, François O. 2018. Fast inference of individual admixture coefficients
688 using geographic data. *The Annals of Applied Statistics* 12:586–608. DOI: 10.1214/17-
689 aoas1106.
- 690 Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor.
691 *Bioinformatics* 34:i884–i890. DOI: 10.1093/bioinformatics/bty560.
- 692 Cheviron ZA, Brumfield RT. 2009. Migration-selection balance and local adaptation of
693 mitochondrial haplotypes in Rufous-Collared Sparrows (*Zonotrichia capensis*) along an
694 elevational gradient. *Evolution* 63:1593–1605. DOI: 10.1111/j.1558-5646.2009.00644.x.
- 695 Colliard C, Sicilia A, Turrisi GF, Arculeo M, Perrin N, Stöck M. 2010. Strong reproductive
696 barriers in a narrow hybrid zone of West-Mediterranean green toads (*Bufo viridis*
697 subgroup) with Plio-Pleistocene divergence. *BMC Evolutionary Biology* 10:232–232.
698 DOI: 10.1186/1471-2148-10-232.
- 699 Derryberry EP, Derryberry GE, Maley JM, Brumfield RT. 2014. hzar: hybrid zone analysis using
700 an R software package. *Molecular Ecology Resources* 14:652–663. DOI: 10.1111/1755-
701 0998.12209.
- 702 Dobzhansky T. 1936. Studies on hybrid sterility. II. Localization of sterility factors in
703 *Drosophila pseudoobscura* hybrids. *Genetics* 21:113–135. DOI:
704 10.1093/genetics/21.2.113.
- 705 Dufresnes C, Bonato L, Novarini N, Betto-Colliard C, Perrin N, Stöck M. 2014. Inferring the
706 degree of incipient speciation in secondary contact zones of closely related lineages of
707 Palearctic green toads (*Bufo viridis* subgroup). *Heredity* 113:9. DOI:
708 10.1038/hdy.2014.26.
- 709 Dufresnes C, Brelsford A, Jeffries DL, Mazepa G, Suchan T, Canestrelli D, Nicieza A,
710 Fumagalli L, Dubey S, Martínez-Solano I, Litvinchuk SN, Vences M, Perrin N, Crochet
711 P-A. 2021. Mass of genes rather than master genes underlie the genomic architecture of
712 amphibian speciation. *Proceedings of the National Academy of Sciences*
713 118:e2103963118. DOI: 10.1073/pnas.2103963118.
- 714 Dufresnes C, Litvinchuk SN. 2021. Diversity, distribution and molecular species delimitation in
715 frogs and toads from the Eastern Palearctic. *Zoological Journal of the Linnean Society*
716 XX:1–66. DOI: 10.1093/zoolinlean/zlab083.
- 717 Dufresnes C, Litvinchuk SN, Rozenblut-Kościsty B, Rodrigues N, Perrin N, Crochet P, Jeffries
718 DL. 2020a. Hybridization and introgression between toads with different sex
719 chromosome systems. *Evolution Letters* 4:444–456. DOI: 10.1002/evl3.191.
- 720 Dufresnes C, Lymberakis P, Kornilios P, Savary R, Perrin N, Stöck M. 2018. Phylogeography of
721 Aegean green toads (*Bufo viridis* subgroup): continental hybrid swarm vs. insular
722 diversification with discovery of a new island endemic. *BMC Evolutionary Biology* 18:1–
723 12. DOI: 10.1186/s12862-018-1179-0.
- 724 Dufresnes C, Nicieza AG, Litvinchuk SN, Rodrigues N, Jeffries DL, Vences M, Perrin N,
725 Martínez-Solano I. 2020b. Are glacial refugia hotspots of speciation and cytonuclear
726 discordances? Answers from the genomic phylogeography of Spanish common frogs.
727 *Molecular Ecology* 29:986–1000. DOI: 10.1111/mec.15368.
- 728 Dufresnes C, Pribille M, Alard B, Gonçalves H, Amat F, Crochet P-A, Dubey S, Perrin N,
729 Fumagalli L, Vences M, Martínez-Solano I. 2020c. Integrating hybrid zone analyses in
730 species delimitation: lessons from two anuran radiations of the Western Mediterranean.
731 *Heredity* 124:423–438. DOI: 10.1038/s41437-020-0294-z.

- 732 Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the
733 software structure: a simulation study. *Molecular Ecology* 14:2611–2620. DOI:
734 10.1111/j.1365-294x.2005.02553.x.
- 735 Fukutani K, Matsui M, Tran DV, Nishikawa K. 2022. Genetic diversity and demography of *Bufo*
736 *japonicus* and *B. torrenticola* (Amphibia: Anura: Bufonidae) influenced by the
737 Quaternary climate. *PeerJ* 10:e13452. DOI: 10.7717/peerj.13452.
- 738 Gargiulo R, Kull T, Fay MF. 2021. Effective double-digest RAD sequencing and genotyping
739 despite large genome size. *Molecular Ecology Resources* 21:1037–1055. DOI:
740 10.1111/1755-0998.13314.
- 741 Gompert Z, Buerkle CA. 2010. INTROGRESS: a software package for mapping components of
742 isolation in hybrids. *Molecular Ecology Resources* 10:378–384. DOI: 10.1111/j.1755-
743 0998.2009.02733.x.
- 744 Haldane JBS. 1922. Sex ratio and unisexual sterility in hybrid animals. *Journal of Genetics*
745 12:101–109. DOI: 10.1007/bf02983075.
- 746 Harrison RG, Larson EL. 2014. Hybridization, introgression, and the nature of species
747 boundaries. *Journal of Heredity* 105:795–809. DOI: 10.1093/jhered/esu033.
- 748 Hastings WK. 1970. Monte Carlo sampling methods using Markov chains and their applications.
749 *Biometrika* 57:97. DOI: 10.2307/2334940.
- 750 Hewitt GM. 1988. Hybrid zones-natural laboratories for evolutionary studies. *Trends in Ecology*
751 *& Evolution* 3:158–167. DOI: 10.1016/0169-5347(88)90033-x.
- 752 Hickerson MJ, Meyer CP, Moritz C. 2006. DNA barcoding will often fail to discover new
753 animal species over broad parameter space. *Systematic Biology* 55:729–739. DOI:
754 10.1080/10635150600969898.
- 755 Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G. 2011. Next-generation RAD
756 sequencing identifies thousands of SNPs for assessing hybridization between rainbow
757 and westslope cutthroat trout. *Molecular Ecology Resources* 11:117–122. DOI:
758 10.1111/j.1755-0998.2010.02967.x.
- 759 Igawa T, Kurabayashi A, Nishioka M, Sumida M. 2006. Molecular phylogenetic relationship of
760 toads distributed in the Far East and Europe inferred from the nucleotide sequences of
761 mitochondrial DNA genes. *Molecular Phylogenetics and Evolution* 38:250–260. DOI:
762 10.1016/j.ympev.2005.09.003.
- 763 Ikehara K. 1992. Formation of duned sand bodies in the Osumi Strait, south of Kyushu, Japan.
764 *Journal of the Sedimentological Society of Japan* 36:37–45. DOI:
765 10.14860/jssj1972.36.37.
- 766 Jombart T. 2008. *adegenet*: a R package for the multivariate analysis of genetic markers.
767 *Bioinformatics* 24:1403–1405. DOI: 10.1093/bioinformatics/btn129.
- 768 Jombart T, Ahmed I. 2011. *adegenet 1.3-1*: new tools for the analysis of genome-wide SNP data.
769 *Bioinformatics* 27:3070–3071. DOI: 10.1093/bioinformatics/btr521.
- 770 Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new
771 method for the analysis of genetically structured populations. *BMC Genetics* 11:94–94.
772 DOI: 10.1186/1471-2156-11-94.
- 773 Key KHL. 1968. The concept of stasipatric speciation. *Systematic Biology* 17:14–22. DOI:
774 10.1093/sysbio/17.1.14.
- 775 Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I. 2015. Clumpak: a program
776 for identifying clustering modes and packaging population structure inferences across K.
777 *Molecular Ecology Resources* 15:1179–1191. DOI: 10.1111/1755-0998.12387.

- 778 Kusano T, Maruyama K, Kaneko S. 1995. Post-breeding dispersal of the Japanese toad, *Bufo*
779 *japonicus formosus*. *Journal of Herpetology* 29:633. DOI: 10.2307/1564755.
- 780 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R,
781 Subgroup 1000 Genome Project Data Processing. 2009. The sequence alignment/map
782 format and SAMtools. *Bioinformatics* 25:2078–2079. DOI:
783 10.1093/bioinformatics/btp352.
- 784 Li Y, Liu J. 2018. StructureSelector: A web-based software to select and visualize the optimal
785 number of clusters using multiple methods. *Molecular Ecology Resources* 18:176–177.
786 DOI: 10.1111/1755-0998.12719.
- 787 Mallet J, Barton N, Lamas G, Santisteban J, Muedas M, Eeley H. 1990. Estimates of selection
788 and gene flow from measures of cline width and linkage disequilibrium in *heliconius*
789 hybrid zones. *Genetics* 124:921–936. DOI: 10.1093/genetics/124.4.921.
- 790 Marcus J, Ha W, Barber RF, Novembre J. 2021. Fast and flexible estimation of effective
791 migration surfaces. *eLife* 10:e61927. DOI: 10.7554/elife.61927.
- 792 Matsui M. 1984. Morphometric variation analyses and revision of the Japanese toads (Genus
793 *Bufo*, Bufonidae). *Contributions from the Biological Laboratory, Kyoto University*
794 26:209–428.
- 795 Matsui M, Maeda N. 2018. *Encyclopedia of Japanese frogs*. Tokyo: Bun-ichi Sogo Shuppan.
- 796 Matsui M, Okawa H, Nishikawa K, Aoki G, Eto K, Yoshikawa N, Tanabe S, Misawa Y,
797 Tominaga A. 2019. Systematics of the widely distributed Japanese clouded salamander,
798 *Hynobius nebulosus* (Amphibia: Caudata: Hynobiidae), and its closest relatives. *Current*
799 *Herpetology* 38:32–90. DOI: 10.5358/hsj.38.32.
- 800 Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. 1953. Equation of state
801 calculations by fast computing machines. *The Journal of Chemical Physics* 21:1087–
802 1092. DOI: 10.1063/1.1699114.
- 803 Milne RI, Abbott RJ. 2008. Reproductive isolation among two interfertile *Rhododendron*
804 species: low frequency of post-F1 hybrid genotypes in alpine hybrid zones. *Molecular*
805 *Ecology* 17:1108–1121. DOI: 10.1111/j.1365-294x.2007.03643.x.
- 806 Miura I. 1995. The late replication banding patterns of chromosomes are highly conserved in the
807 genera *Rana*, *Hyla*, and *Bufo* (Amphibia: Anura). *Chromosoma* 103:567–574. DOI:
808 10.1007/bf00355322.
- 809 Muller HJ. 1942. Isolating mechanisms, evolution, and temperature. *Biological Symposia*:71–
810 125.
- 811 Musmann SM, Douglas MR, Chafin TK, Douglas ME. 2019. BA3-SNPs: Contemporary
812 migration reconfigured in BayesAss for next-generation sequence data. *Methods in*
813 *Ecology and Evolution* 10:1808–1813. DOI: 10.1111/2041-210x.13252.
- 814 Okada Y. 1928. Notes on Japanese frogs. *Annotationes Zoologicae Japonenses* 11:269–277.
- 815 Orr HA. 1997. Haldane's Rule. *Annual Review of Ecology and Systematics* 28:195–218. DOI:
816 10.1146/annurev.ecolsys.28.1.195.
- 817 Petkova D, Novembre J, Stephens M. 2016. Visualizing spatial population structure with
818 estimated effective migration surfaces. *Nature Genetics* 48:94–100. DOI:
819 10.1038/ng.3464.
- 820 Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus
821 genotype data. *Genetics* 155:945–959. DOI: 10.1093/genetics/155.2.945.

- 822 Puechmaille SJ. 2016. The program structure does not reliably recover the correct population
823 structure when sampling is uneven: subsampling and new estimators alleviate the
824 problem. *Molecular Ecology Resources* 16:608–627. DOI: 10.1111/1755-0998.12512.
- 825 Queiroz, K. D. 2007. Species concepts and species delimitation. *Systematic Biology* 56:879–886.
826 DOI: 10.1080/10635150701701083.
- 827 Queiroz, K. D. 2020. An updated concept of subspecies resolves a dispute about the taxonomy of
828 incompletely separated lineages. *Herpetological Review* 51:459–461.
- 829 Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarization in
830 Bayesian phylogenetics using Tracer 1.7. *Systematic Biology* 67:901–904. DOI:
831 10.1093/sysbio/syy032.
- 832 R Core Team (2022). R: A language and environment for statistical computing. *R Foundation for*
833 *Statistical Computing*, Vienna, Austria. URL <https://www.R-project.org/>.
- 834 Riemsdijk I van, Arntzen JW, Bucciarelli GM, McCartney-Melstad E, Rafajlović M, Scott PA,
835 Toffelmier E, Shaffer HB, Wielstra B. 2023. Two transects reveal remarkable variation in
836 gene flow on opposite ends of a European toad hybrid zone. *Heredity* 131:15–24. DOI:
837 10.1038/s41437-023-00617-6.
- 838 RStudio Team (2022). RStudio: Integrated Development Environment for R. *RStudio, PBC,*
839 *Boston, MA* URL <http://www.rstudio.com/>.
- 840 Rochette NC, Rivera-Colón AG, Catchen JM. 2019. Stacks 2: Analytical methods for paired-end
841 sequencing improve RADseq-based population genomics. *Molecular Ecology* 28:4737–
842 4754. DOI: 10.1111/mec.15253.
- 843 Scordato ESC, Wilkins MR, Semenov G, Rubtsov AS, Kane NC, Safran RJ. 2017. Genomic
844 variation across two barn swallow hybrid zones reveals traits associated with divergence
845 in sympatry and allopatry. *Molecular Ecology* 26:5676–5691. DOI: 10.1111/mec.14276.
- 846 Sequeira F, Bessa-Silva A, Tarroso P, Sousa-Neves T, Vallinoto M, Gonçalves H,
847 Martínez-Solano I. 2020. Discordant patterns of introgression across a narrow hybrid
848 zone between two cryptic lineages of an Iberian endemic newt. *Journal of Evolutionary*
849 *Biology* 33:202–216. DOI: 10.1111/jeb.13562.
- 850 Servedio MR, Hermisson J. 2020. The evolution of partial reproductive isolation as an adaptive
851 optimum. *Evolution* 74:4–14. DOI: 10.1111/evo.13880.
- 852 Slager DL, Epperly KL, Ha RR, Rohwer S, Wood C, Hemert C, Klicka J. 2020. Cryptic and
853 extensive hybridization between ancient lineages of American crows. *Molecular Ecology*
854 29:956–969. DOI: 10.1111/mec.15377.
- 855 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
856 large phylogenies. *Bioinformatics* 30:1312–1313. DOI: 10.1093/bioinformatics/btu033.
- 857 Stöck M, Moritz C, Hickerson M, Frynta D, Dujsebayeva T, Eremchenko V, Macey JR,
858 Papenfuss TJ, Wake DB. 2006. Evolution of mitochondrial relationships and
859 biogeography of Palearctic green toads (*Bufo viridis* subgroup) with insights in their
860 genomic plasticity. *Molecular Phylogenetics and Evolution* 41:663–689. DOI:
861 10.1016/j.ympev.2006.05.026.
- 862 Suyama Y, Hirota SK, Matsuo A, Tsunamoto Y, Mitsuyuki C, Shimura A, Okano K. 2022.
863 Complementary combination of multiplex high-throughput DNA sequencing for
864 molecular phylogeny. *Ecological Research* 37:171–181. DOI: 10.1111/1440-1703.12270.
- 865 Suyama Y, Matsuki Y. 2015. MIG-seq: an effective PCR-based method for genome-wide single-
866 nucleotide polymorphism genotyping using the next-generation sequencing platform.
867 *Scientific Reports* 5:srep16963. DOI: 10.1038/srep16963.

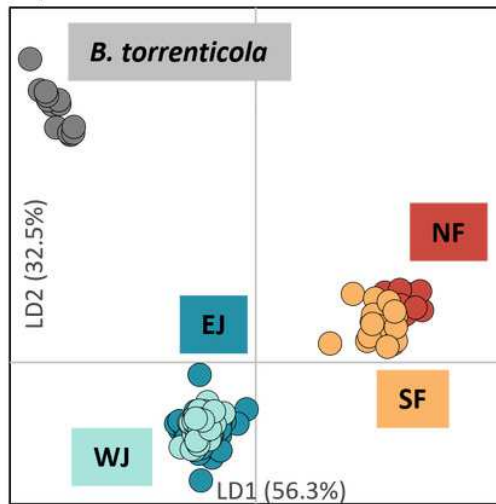
- 868 Szymura JM, Barton NH. 1986. Genetic analysis of a hybrid zone between the fire-bellied toads
869 *Bombina bombina* and *B. variegata*, near Cracow in Southern Poland. *Evolution*
870 40:1141–1159. DOI: 10.1111/j.1558-5646.1986.tb05740.x.
- 871 Szymura JM, Barton NH. 1991. The genetic structure of the hybrid zone between the fire-bellied
872 toads *Bombina bombina* and *B. variegata*: comparisons between transects and between
873 loci. *Evolution* 45:237–261. DOI: 10.1111/j.1558-5646.1991.tb04400.x.
- 874 Takata K, Iwase F, Iguchi A, Yuasa H, Taninaka H, Iwasaki N, Uda K, Suzuki T, Nonaka M,
875 Kikuchi T, Yasuda N. 2021. Genome-wide SNP data revealed notable spatial genetic
876 structure in the deep-sea precious coral *Corallium japonicum*. *Frontiers in Marine*
877 *Science* 8:667481. DOI: 10.3389/fmars.2021.667481.
- 878 Toews DPL, Brelsford A. 2012. The biogeography of mitochondrial and nuclear discordance in
879 animals. *Molecular Ecology* 21:3907–3930. DOI: 10.1111/j.1365-294x.2012.05664.x.
- 880 Tominaga A, Matsui M, Yoshikawa N, Eto K, Nishikawa K. 2018. Genomic displacement and
881 shift of the hybrid zone in the Japanese fire-bellied newt. *Journal of Heredity* 109:232–
882 242. DOI: 10.1093/jhered/esx085.
- 883 Toyama KS, Crochet P, Leblois R. 2020. Sampling schemes and drift can bias admixture
884 proportions inferred by structure. *Molecular Ecology Resources* 20:1769–1785. DOI:
885 10.1111/1755-0998.13234.
- 886 Vähä J-P, Primmer CR. 2006. Efficiency of model-based Bayesian methods for detecting hybrid
887 individuals under different hybridization scenarios and with different numbers of loci.
888 *Molecular Ecology* 15:63–72. DOI: 10.1111/j.1365-294x.2005.02773.x.
- 889 Vasimuddin M, Misra S, Li H, Aluru S. 2024. Efficient architecture-aware acceleration of BWA-
890 MEM for multicore systems. In: 2019 IEEE International Parallel and Distributed
891 Processing Symposium (IPDPS). Rio de Janeiro, Brazil, 314–324. DOI:
892 10.1109/ipdps.2019.00041.
- 893 Watanabe K, Tabata R, Nakajima J, Kobayakawa M, Matsuda M, Takaku K, Hosoya K, Ohara
894 K, Takagi M, Jang-Liaw N-H. 2020. Large-scale hybridization of Japanese populations
895 of Hinamoroko, *Aphyocypris chinensis*, with *A. kikuchii* introduced from Taiwan.
896 *Ichthyological Research* 67:361–374. DOI: 10.1007/s10228-019-00730-9.
- 897 Willis SC, Hollenbeck CM, Puritz JB, Gold JR, Portnoy DS. 2017. Haplotyping RAD loci: an
898 efficient method to filter paralogs and account for physical linkage. *Molecular Ecology*
899 *Resources* 17:955–965. DOI: 10.1111/1755-0998.12647.
- 900 Wilson GA, Rannala B. 2003. Bayesian inference of recent migration rates using multilocus
901 genotypes. *Genetics* 163:1177–1191. DOI: 10.1093/genetics/163.3.1177.
- 902 Wu C. 2001. The genic view of the process of speciation. *Journal of Evolutionary Biology*
903 14:851–865. DOI: 10.1046/j.1420-9101.2001.00335.x.
- 904 Yanchukov A, Hofman S, Szymura JM, Mezhzherin SV, Morozov-Leonov SY, Barton NH,
905 Nrnberger B. 2006. Hybridization of *Bombina bombina* and *B. variegata* (Anura,
906 Discoglossidae) at a sharp ecotone in western Ukraine: comparisons across transects and
907 over time. *Evolution* 60:583–600. DOI: 10.1554/04-739.1.
- 908 Yashima K. 1994. A geomorphological study of the caldrons in the Seto inland sea. *Report of*
909 *Hydrographic Researches*:237–327.
- 910 Zhao H, Beck B, Fuller A, Peatman E. 2020. EasyParallel: A GUI platform for parallelization of
911 STRUCTURE and NEWHYBRIDS analyses. *Plos One* 15:e0232110. DOI:
912 10.1371/journal.pone.0232110.

Figure 1

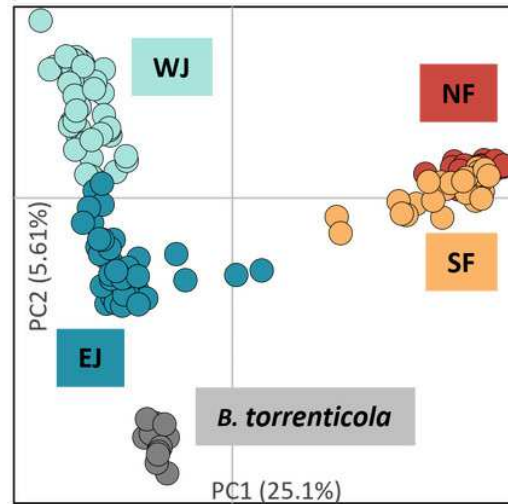
Population structure using (A) DAPC, (B) Structure, and (C) PCA based on SNPs datasets, dataset I for DAPC and PCA, and dataset I-2 for Structure.

The four different genetic clusters, northern *Bufo japonicus formosus* (NF), southern *B. j. formosus* (SF), eastern *B. j. japonicus* (EJ), western *B. j. japonicus* (WJ), are displayed with *B. torrenticola*. (A-1) DAPC plot shows the best fit for $K = 5$ clusters. The axes represent the first two linear discriminants (LD), and the dots represent individuals colored by their groups in DAPC. (A-2) The distribution map of individuals colored by the cluster assignments by DAPC. The map was created by QGIS 3.28 (<https://qgis.org>). The administrative areas dataset was obtained from the GADM database (www.gadm.org, version 3.4) and the inland water dataset from the Digital Chart of the World available at the DIVA-GIS online resource (www.diva-gis.org). (B) Structure bar plots show individual ancestry to the five clusters ($K = 5$). (C) PC1 and PC2 are plotted. Each dot corresponds to an individual colored according to their genetic cluster found in DAPC. The first axis distinguishes *B. j. formosus* and *B. j. japonicus*, and the second axis distinguishes *B. japonicus* and *B. torrenticola* and reflects intraspecific structure within *B. japonicus*.

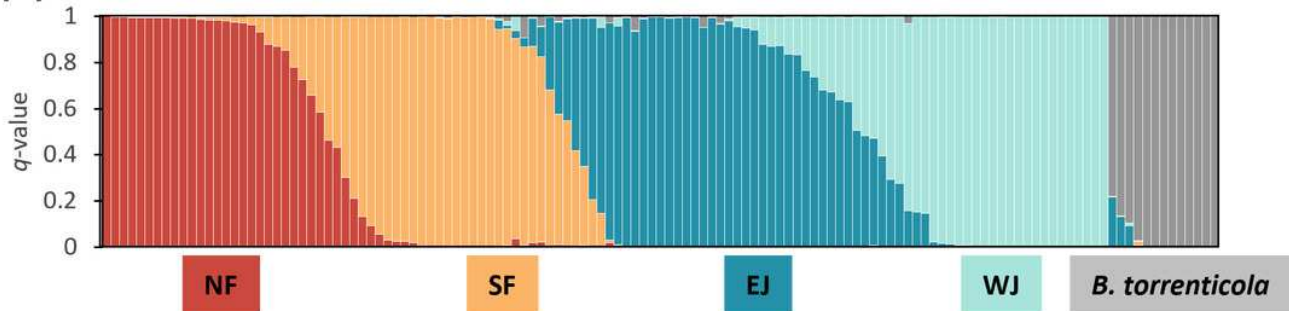
(A-1)



(C)



(B)



(A-2)

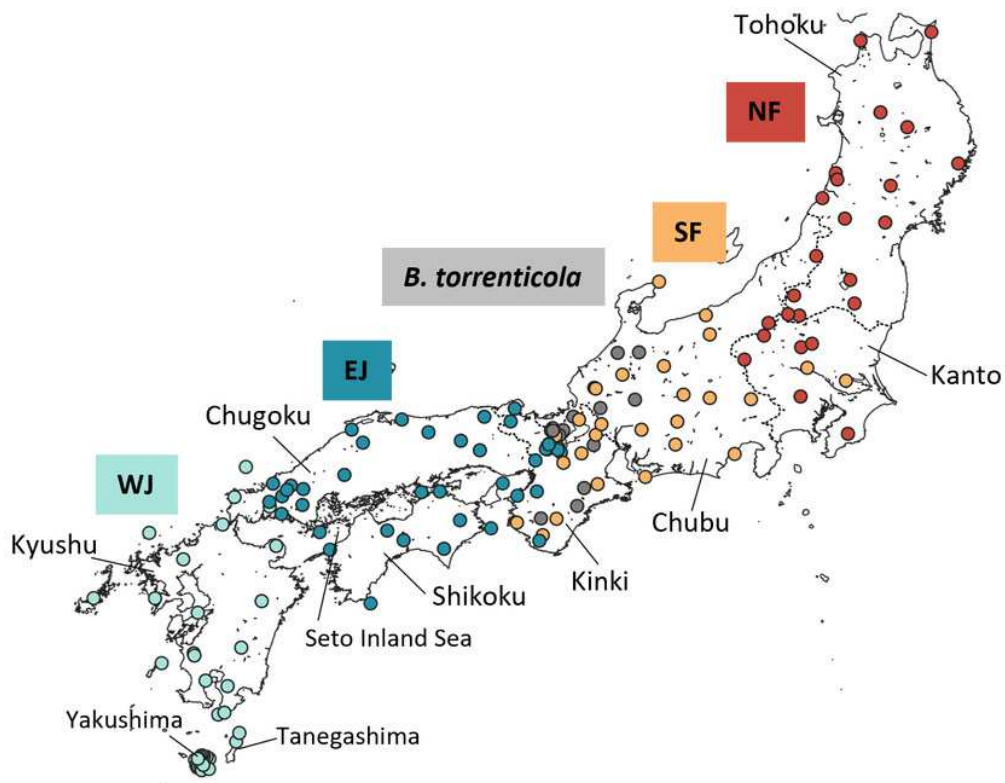


Figure 2

(A) Densitree diagram representing the species tree obtained from SNAPP using SNPs.
(B) The phylogenetic tree using mitochondrial cytochrome *b* sequences.

(A) All nodes were supported by posterior probabilities of 1.0. (B) Asterisks on each node indicate bootstrap supports are more than 85%.

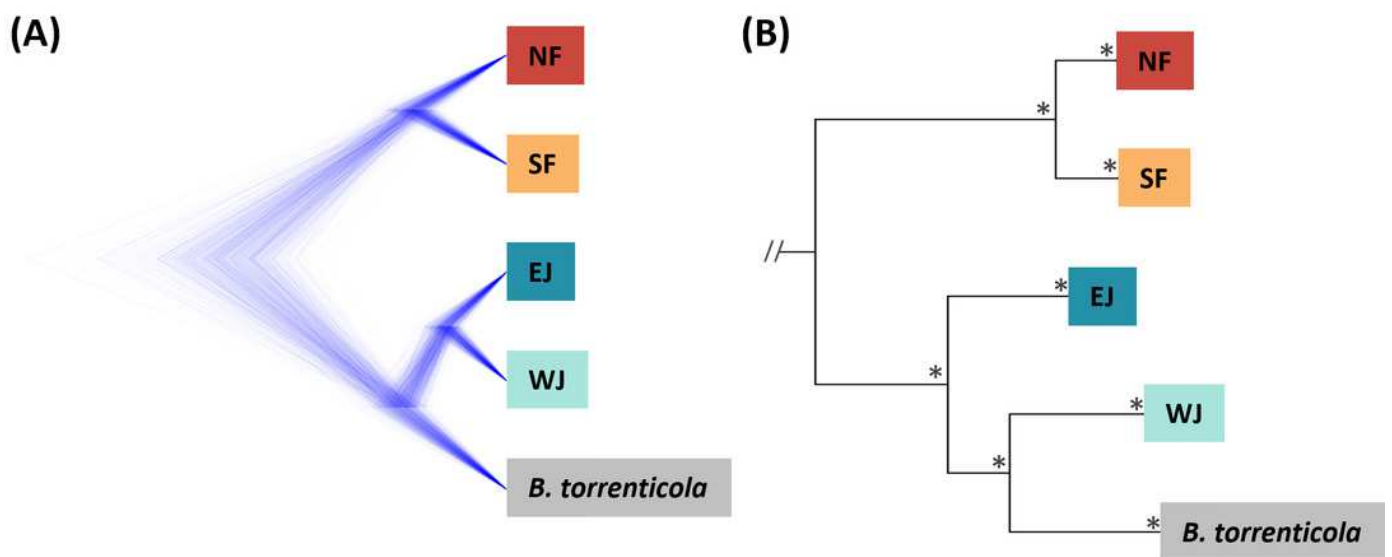


Figure 3

Effective migration rates for the lowest cross-validation lambda estimated by FEEMS (Fast Estimation of Effective Migration Surfaces) using dataset II.

The figure shows the fitted parameters in the log scale, with lower effective migration shown in orange and higher effective migration shown in blue. Dots represent individuals.

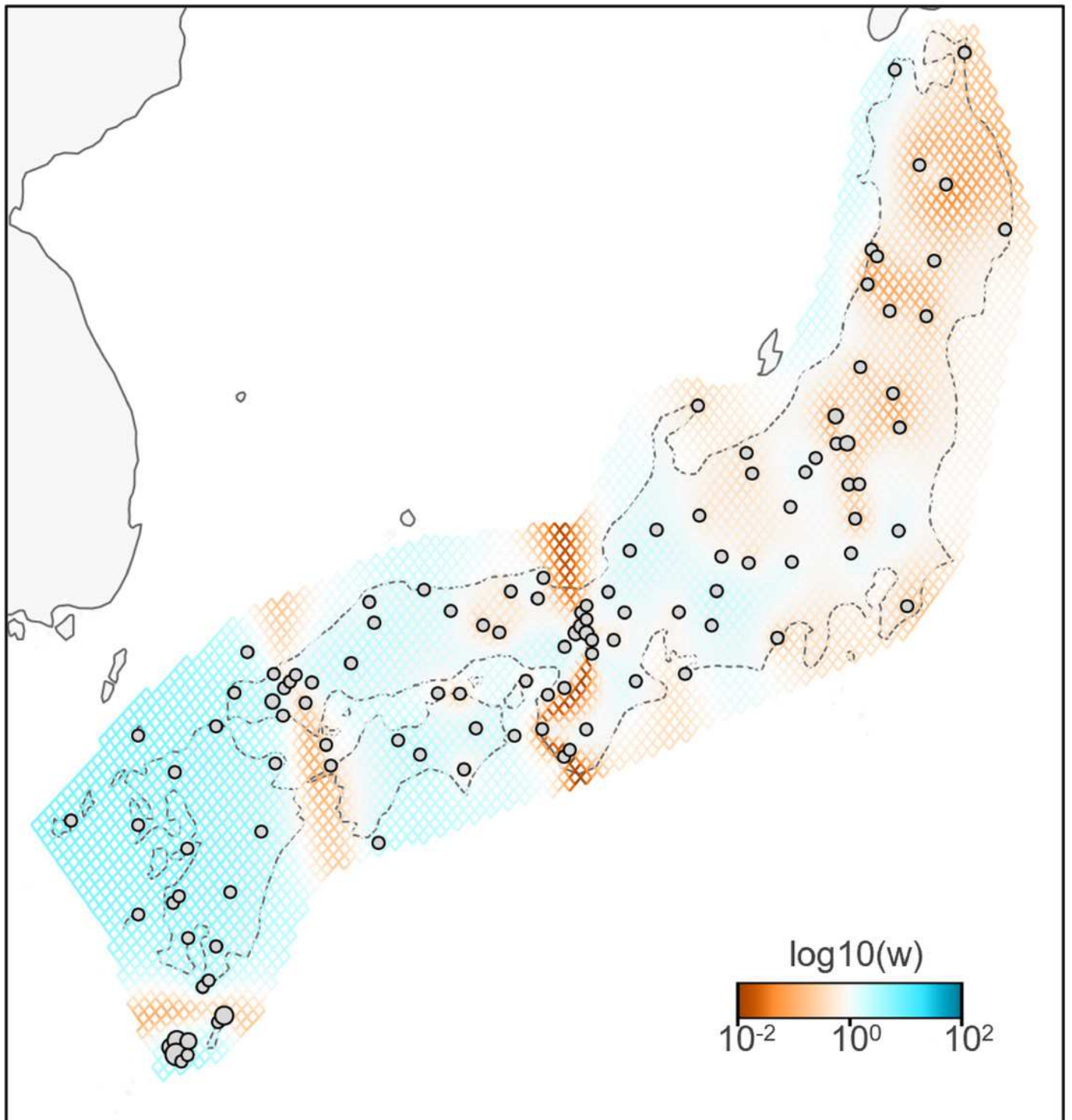


Figure 4

Maps showing sampling localities with pie charts for three different contact zone of sub-datasets, (A) SF-EJ, (B) NF-SF, and (C) EJ-WJ.

Pie charts show the q -values inferred by the Structure program for each individual. The dotted lines indicate the baselines used for *hzar*.

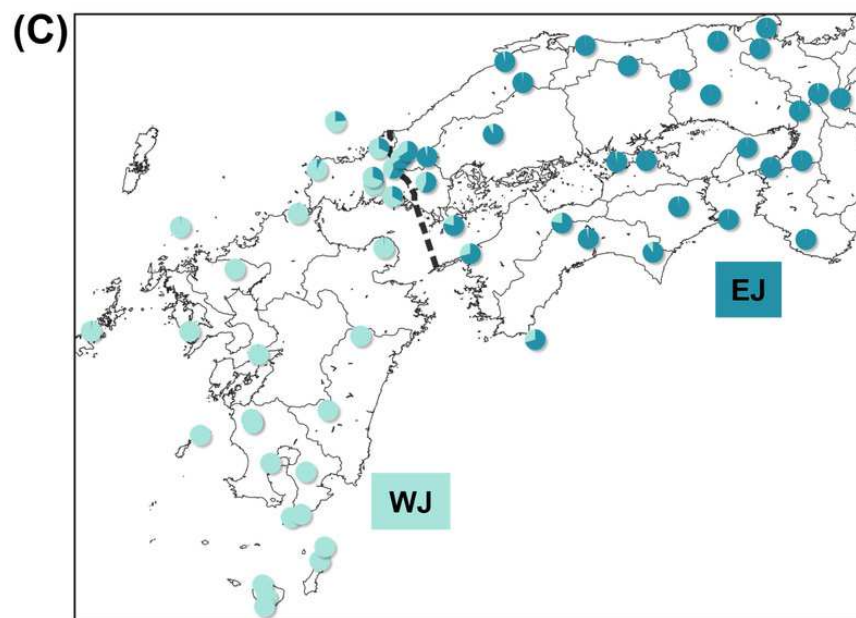
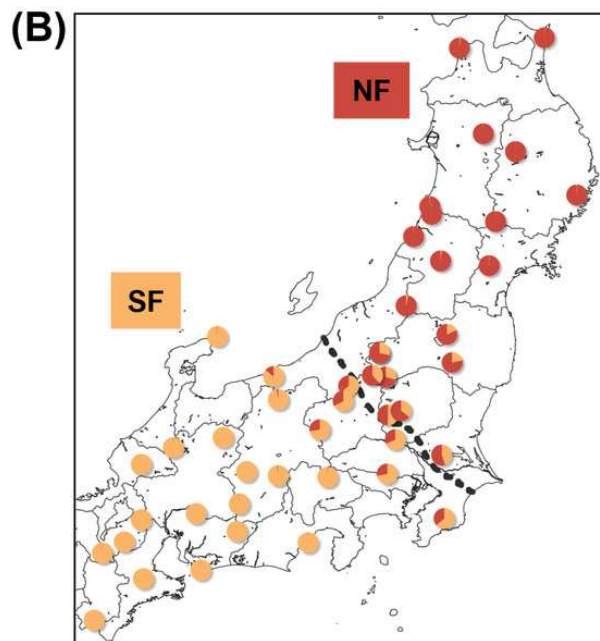
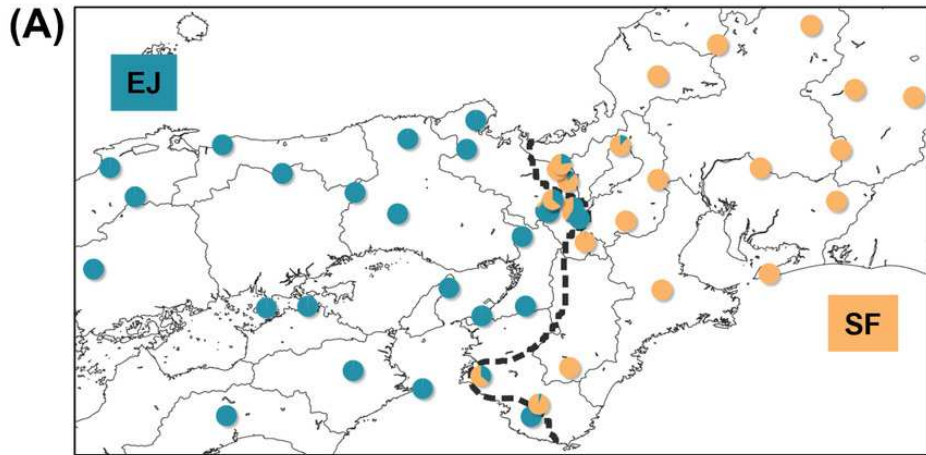


Figure 5

The maximum-likelihood clines fitted on nuclear genomic average ancestry and mitochondrial allele frequencies along three different transects of sub-datasets, (A) SF-EJ, (B) NF-SF, and (C) EJ-WJ.

The grey areas show the 95% credible cline region. The x-axis represents distances (km) from the baselines shown in Fig. 4. Crosses indicate the observed values for individuals.

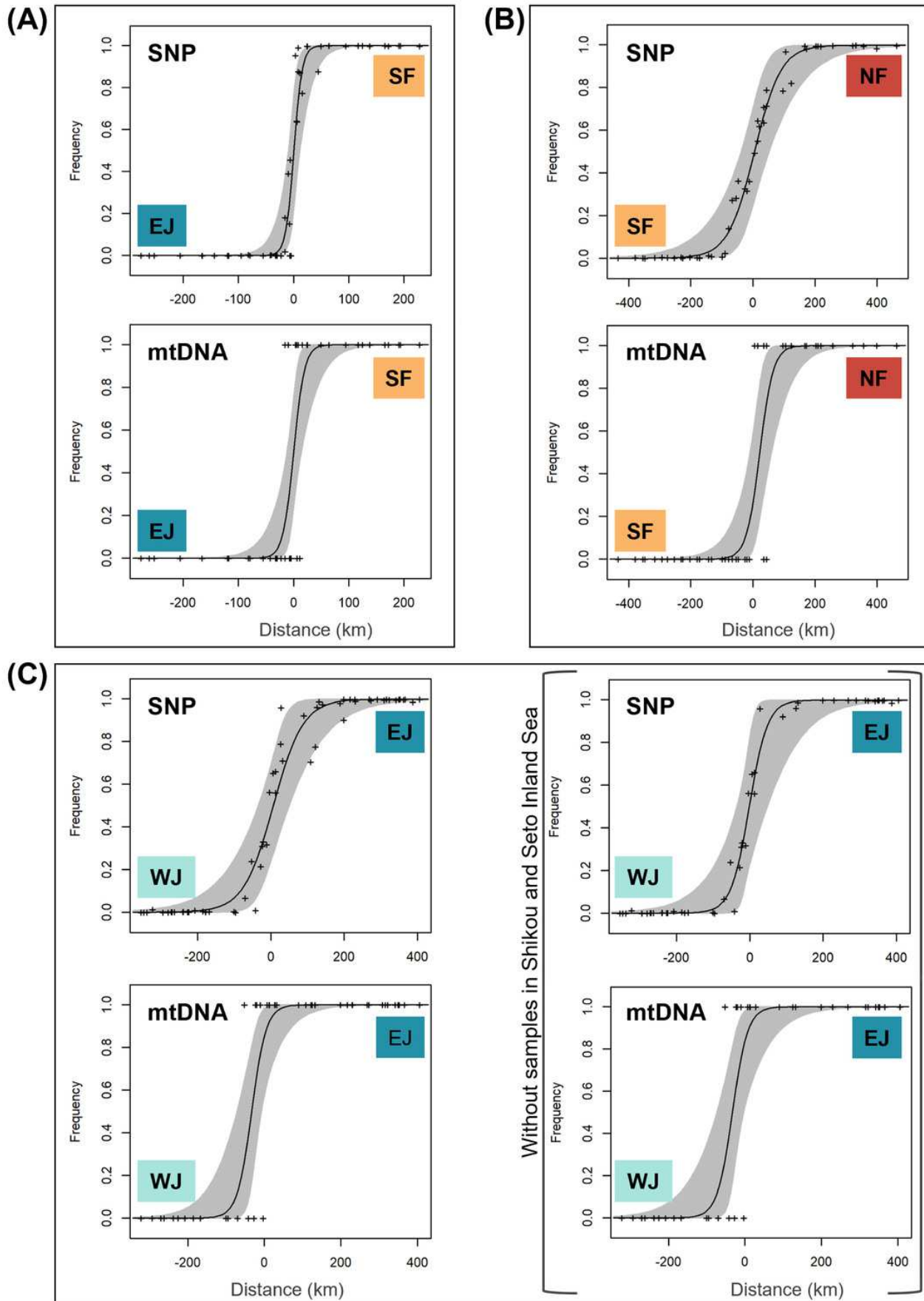


Figure 6

Triangle plots of the hybrid index versus heterozygosity of individuals based on selected ancestry-informative SNP markers ($F_{st} = 1$) for sub-datasets, (A) SF-EJ, (B) NF-SF, and (C) EJ-WJ.

Individual with intermediate hybrid indices (> 0.25 and < 0.75) and high heterozygosity (≥ 0.5) was considered as recent-generation hybrid (a gray square), and those with intermediate hybrid indices (> 0.25 and < 0.75) and low heterozygosity (< 0.5) as later-generation hybrids (gray triangles). Those with low hybrid indices (≤ 0.25 or ≥ 0.75) were considered as backcross to one or the other parental type (triangles colored by parental assignments). Each colored circle indicates the pure individuals.

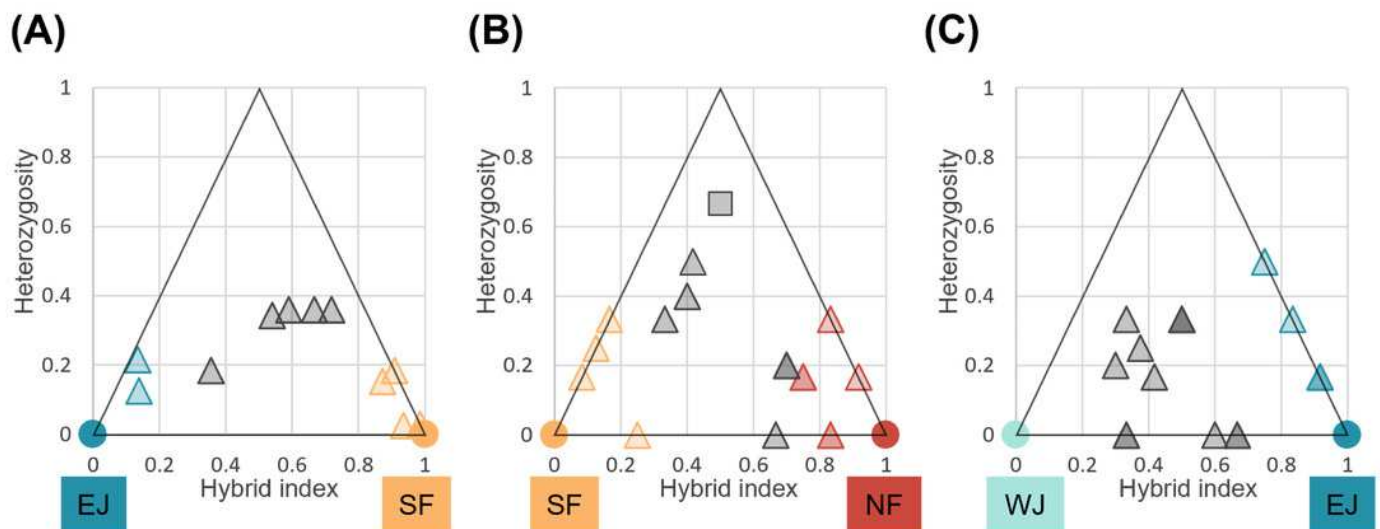


Table 1 (on next page)

Estimates of migrants from BayesAss3-SNPs analyses between population clusters, (A) SF-EJ, (B) NF-SF, and (C) EJ-WJ.

The row headers represent the populations into where the individuals migrated, and the column headers represent the populations from where the migrant derived. Standard deviations of the values are given in parentheses.

(A)

		Migration from		
		parental NF	Hybrid	parental SF
Migration into	parental NF	0.9407 (0.0349)	0.0355 (0.0290)	0.0238 (0.0221)
	Hybrid	0.0152 (0.0145)	0.9262 (0.0293)	0.0586 (0.0269)
	parental SF	0.0167 (0.0158)	0.0167 (0.0158)	0.9667 (0.0218)

(B)

		Migration from		
		parental SF	Hybrid	parental EJ
Migration into	parental SF	0.9607 (0.0253)	0.0196 (0.0185)	0.0196 (0.0185)
	Hybrid	0.1667 (0.0431)	0.7619 (0.0389)	0.0714 (0.0353)
	parental EJ	0.0133 (0.0128)	0.0133 (0.0128)	0.9733 (0.0178)

(C)

		Migration from		
		parental EJ	Hybrid	parental WJ
Migration into	parental EJ	0.9683 (0.0209)	0.0133 (0.0151)	0.0159 (0.0152)
	Hybrid	0.0725 (0.0280)	0.8013 (0.0340)	0.1262 (0.0337)
	parental WJ	0.0139 (0.0133)	0.0139 (0.0133)	0.9722 (0.0184)