Isolation and re-connection: the formation of a ring-shaped speciation continuum in an odorous frog (*Odorrana margaretae*)

Guannan Wen [1,2] and Jinzhong Fu [1,3]

[1] Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu, Sichuan 610041, China

[2] University of Chinese Academy of Sciences, Beijing 100049, China

[3] Department of Integrative Biology, University of Guelph, Ontario N1G 2W1, Canada

Running title: a ring-shaped speciation continuum

Keywords: speciation continuum, replicate hybrid zones, genetic coupling, frog, ddRAD

Corresponding author:

Jinzhong Fu

Phone: (519) 824 4120 ext. 52715

Fax: (519) 767 1656

Email: jfu@uoguelph.ca

Abstract

The Green Odorous Frog (*Odorrana margaretae*) around the Sichuan Basin of western China displays a ring-shaped distributional pattern and possesses multiple replicate contact zones between lineages at various levels of differentiation. To understand its unique speciation history and mechanisms, we obtained 1,540 SNPs from 29 populations and 227 individuals using ddRAD sequencing. Population structure analysis revealed three groups within the species: The West, the North & South, and the East groups. These groups were initially isolated at ~2.03 million years ago, and subsequent post-glacial expansion produced the current ring-shaped distribution around Sichuan Basin with three contact zones. Hybridization in those zones involved lineages with different levels of divergence and produced greatly different outcomes. Both the hybrid zones at southwest (S-W) and southeast (E-NS) of the Basin have extensive admixture and less barrier effect. Consequently, the southern region has the highest genetic diversity and becomes an 'evolutionary melting pot'. In contrast, the hybrid zone at northwestern corner (N-W), which resembles the overlap zone between two expansion terminals of a ring species, has limited admixture with a narrow geographic cline, suggesting partial reproductive isolation between the northern and western populations. The three hybrid zones likely resemble three time points along a speciation continuum; while both E-NS and S-W hybrid zones are merging, the N-W zone may have passed the 'tipping point' and is destined for a complete reproductive isolation over time.

Introduction

The formation of species has been a central issue in modern biology for over a century (Darwin, 1859; Dobzhansky, 1937; Mayr, 1942; Schluter, 2000; Gavrilets, 2004; Coyne & Orr, 2004; Nosil, 2012). Multiple forces that drive the speciation process, including geographic isolation, gene flow, and natural/sexual selection, have been identified and in some cases their relative importance have been quantified (Rice & Hostert, 1993; Sobel et al., 2010; Tazzyman & Iwasa, 2010; Maan & Seehausen, 2011). One challenge to the study of speciation process is its continuous nature. Although the speciation continuum has been recognized as early as Darwin (Darwin 1859; Walsh, 1861; Clausen, 1951), its complexity and impacts on our understanding of the speciation process have only been brought into broad view in the last two decades (Seehausen et al., 2008; Peccoud et al., 2009a; Merrill et al., 2011; Shaw & Mullen, 2014).

Theoretical work predicted that both gradual and sudden changes can occur in a speciation continuum. The process of speciation is driven simultaneously by multiple evolutionary forces (e.g. genetic drift, mutation, selection, gene flow), which often have opposite effects. For nearly a century, theoretical modelling has advanced our understanding of the speciation process from gradual to including sudden (Gavrilets, 2014). One of the major predictions is that there should be a critical threshold for the rapid transition in the degree of reproductive isolation from low to high ("threshold effect", Gavrilets, 2004; "tipping point", Nosil et al., 2017). Once passed the threshold, the speciation process will only move forward (Nosil et al., 2017). On the other hand, there are populations getting 'stuck

partway' at intermediate stages of speciation and may remain in relatively stable equilibrium levels of differentiation for an extended time (Berlocher & Feder, 2002; Seehausen et al., 2008). To understand the cause and consequence of speciation, it is essential to take an integrated view of all stages of the speciation process.

There are two common approaches to studying speciation continuum. One is comparing multiple species pairs after recent speciation, such as in the case of *Pundamilia* cichlids (Seehausen et al., 2008; Seehausen, 2009; Seehausen & Magalhaes, 2010). Alternatively, different population pairs within a single species that vary from recently diverged to strongly diverged (or near complete reproductive isolation) can be compared, and such approach would allow strong inferences about how transitions along the speciation continuum unfold from beginning to end within a single taxon (Nosil, 2012). Studies of the pea aphids (*Acyrthosiphon pisum*) probably best exemplified this approach (Peccoud et al., 2009a; Peccoud, et al., 2009b).

Species with a ring-shaped divergence, or ring species, represent a perfect demonstration of speciation continuum (Mayer, 1942; Cain, 1954; but see Coyne & Orr, 2004), with a wide variety of examples have been proposed and explicitly studied (e.g., *Ensatina* salamanders, Wake, 1997; Kuchta et al., 2009; Pereira & Wake, 2009; the Greenish Warbler, Irwin et al., 2001; Alcaide et al., 2014). The novelty of ring species is that they capture multiple aspects of divergence, from variation among populations, to isolation by distance, phenotypic divergence, and reproductive isolation (Kuchta & Wake, 2016). One can

sample several populations at different points in the continuum to reconstruct the speciation process indirectly instead of examining reproductive isolation build-up over real time.

The Green Odorous Frog (*Odorrana margaretae*) is a promising model system for studying speciation from a continuum perspective. Previous work showed that this stream dweller displays a ring-shaped divergence pattern around the Sichuan Basin (Fei et al., 1999; Qiao et al., 2018). Likely survived in two refugia during one glaciation episode, its populations formed at least two contact zones after subsequent post-glacial expansion. At the southern contact zone, extensive exchange occurred and the area formed a zone of admixture with high genetic diversity (Qiao et al., 2018). At the northwestern contact zone, the northern and western populations had limited gene exchange in a narrow hybrid zone and have likely developed partial reproductive isolation. These contrasting hybrid zone dynamics involving populations at different levels of divergence may simulate multiple points along a speciation continuum. These unique characteristics, multiple replicate hybrid zones with various levels of divergence and connected by continuous genetic variation along a ring, make *O. margaretae* an excellent model for studying speciation.

In this study, we examine the historical population dynamics of *O. margaretae* using data from reduced-representation genome sequencing (double digest RADseq or ddRADseq). Compared to microsatellite DNA data, this type of data (SNP) require smaller sample size from each site (Nazareno et al., 2017), which allowed us to include more sampling sites in this study. With a high-density sampling and a large amount of genomic data, we aim to 1) infer the evolutionary history of *O. margaretae*, and more importantly, 2) detect dynamics in

these replicate hybrid zones. By examining population dynamics at multiple points along a speciation continuum of a single species, we hope that this study will provide more insight into the speciation process.

Materials and Methods

Sampling and laboratory protocols

A total of 227 individuals of *O. margaretae* from 29 locations were collected around the Sichuan Basin (Figure 1a; Appendix S1). One or two toes were collected from most individuals and animals were released on-site. Two or three individuals from each site were euthanized and preserved as reference specimens, for which muscle or liver tissues were collected. Tissues were stored in 95% ethanol at -20℃. All reference specimens are deposited at the Herpetology Collection of the Chengdu Institute of Biology, Chinese Academy of Sciences. Total DNA was extracted using Qiagen Dneasy Blood & Tissue Kits (Qiagen Inc., Valencia, CA, USA). DNA quality, integrity, and quantity were checked using a NanoDrop (Thermo Scientific), 1% agarose gels, and a Qubit (Life Technologies).

The ddRAD libraries were prepared according to the protocols of Peterson et al. (2012). At least 200ng fresh extracted DNA was digested with SbfI and MseI (NEB, Hitchin, UK). We manually size-selected fragments between 300-500bp using 2% agarose gel electrophoresis (Biowest, Spain). All samples were randomly divided into four libraries as evenly as possible. To reduce batch effects, we included 1-2 biological replicates in each batch. Libraries were checked for fragment size range using capillary electrophoresis

(LabChip GX Touch, PerkinElmer) and for quantity using a fluorometer and real-time PCR. Libraries were sequenced on an Illumina HiSeq2500 platform as paired-end, 150 bp reads at the Novogene Corporation (Tianjin, China).

Data processing and quality control

Quality control for raw reads and genotyping were conducted using STACKS v2.53 (Catchen et al., 2013), and detailed parameters and settings are provided in Appendix S2. For the generated SNP data, we first excluded SNPs unique to each of the four libraries, and then conducted quality control using PLINK v1.90b3.46 (Purcell et al., 2007) with the following parameters: (1) include only SNPs with genotyping rates greater than 90%; (2) remove individuals with missing genotype rates greater than 5%; (3) exclude markers not in Hardy-Weinberg equilibrium (significance level 0.05). All sites were required to have depths of coverage of at least 30 reads and no more than twice the mean depth in each individual. We also removed loci that were potentially in physical linkage, because they would affect many population genetic analyses. If two loci were in linkage disequilibrium (LD; $r^2 > 0.8$ calculated by PLINK) in 60% or more of the populations, we removed one locus from the pair. The resulted loci constituted the *full* dataset for further analysis.

Population genetic structure analysis

A set of descriptive statistics, including average observed number of alleles, effective number of alleles, observed heterozygosity (H_O), and expected heterozygosity (H_E), were

obtained from STACKS for populations with sample sizes greater than four. Pairwise F_{ST} (Weir, 1996) was used to describe genetic differentiation between populations with minimum sample size of four, as n = 4 has been shown to provide accurate estimates (Willing, Dreyer & Van Oosterhout, 2012). F_{ST} was calculated using *populations* in STACKS (--fstats).

We identified global F_{ST} outlier SNP loci using BAYESCAN v2.1 (Foll & Gaggiotti, 2008). A total of 550,000 iterations were performed, including 50,000 as burn-in and 500,000 iterations with a thinning interval of 10. Loci with q-value below 0.05 for three replicates were considered as outliers. These loci are potentially under selection and we excluded them to obtain the *neutral* dataset for part of the downstream analyses.

We evaluated population genetic structure using four sets of analyses: network and tree, principle coordinate analysis (PCoA), genetic clustering, and isolation by distance (IBD). The *neutral* dataset was used for these analyses.

A neighbor-joining network was constructed based on uncorrected p-distances implemented in SPLITSTREE v4.13.1 (Huson & Bryant, 2005). A neighbor-joining tree with 100 bootstrap replicates was also constructed with R package 'ape' (Paradis, Claude, & Strimmer, 2004). A PCoA using Euclidean distance between individuals was carried out with the R package 'dartR' (v.1.0.5, Gruber et al., 2018). The R package 'ggplot2' (Wickham & Chang, 2008) was used for plotting the results.

The genetic clustering analysis was conducted using TESS3 in R (Caye et al., 2016). TESS3 provides a spatially explicit analysis for population structure among sampled individuals by modeling continuous geographic variation through space. Starting from K = 1

to K = 9, 100 independent runs were performed for each K with a tolerance of 10e-6 and 200 maximum iterations per run. We used equal weights to the loss function and to the penalty function (lambda = 1). For each converged K, we plotted one run with the lowest cross-validation error, and the result with the lowest value of the root-mean-squared errors (RMSE) was kept. Further analysis with subsets of individuals were also conducted, which would capture finer hierarchical population structure, although these K values might have no demographic or historical meaning (Meirmans, 2015).

An isolation-by-distance (IBD) pattern was examined using R package 'ade4' (Dray & Dufour, 2007). The genetic distance was represented by $F_{ST}/(1-F_{ST})$ values. All geographic distances were calculated using R package 'geosphere' (Hijmans, 2016). Since the populations form a ring-shaped distribution around the Sichuan Basin (except the East group; Figure 1), we used an approximate 'ring distance' between these populations (Qiao et al., 2018). Distance between two populations was estimated following suitable habitats around the edge of the Basin. For example, distances between a northern and a western site was estimated by the sum of distances from the northern site to site 4, from site 4 to site 14, from site 14 to site 22, and from site 22 to the western site (Figure 1). A Mantel test with 1,000 permutations was used to detect significant correlation between the genetic distances and geographic distances.

Historical demographic analysis

We used a model selection approach to test various demographic scenarios. To facilitate the analysis, we grouped all samples into four groups according to their genetic structure and geographic locations: North (sites 1-7), East (sites 8-10), South (sites 11-21), and West (sites 22-29) (Figure 1b). A total of 15 models with various levels of complexity were tested (Figure 2). We progressed from simple models to more complex models (see Appendix S2 for a detailed description of models). Finally, we used the best-fitting model to estimate parameters for migration, population sizes, and time of events.

We used a continuous-time coalescent framework FASTSIMCOAL2 V2.6.0.3 (Excoffier et al., 2013) to perform the model selection. Demographic modeling is based on the site frequency spectrum (SFS) of the neutral SNPs. The observed multi-dimensional SFS was computed with ARLEQUIN v3.5.2.2 (Excoffier et al., 2010), and the invariable sites in the SFS were excluded. We fixed the effective population size for the West group to enable the estimation of other parameters (Lanier et al., 2015), because the West group was the most cohesive group from our analysis. The East group was fixed when the West was not involved in the estimation. The effective population size was calculated from $Ne = \pi / 4\mu$. Nucleotide diversity (π) was estimated from all nucleotide sites using STACKS (for West group, average π = 0.00024; for East group, average π = 0.00049; which were estimated from pure populations only. More details below). The mutation rate per site per generation (μ = 0.776e-9) was inferred from *Nanorana parkeri* (Sun et al., 2015).

All models were compared using Akaike information criterion (AIC) from the approximated likelihood and number of parameters. We run 50 replicates for each model with

different group combinations to obtain the highest likelihood. Each run included 100,000-200,000 simulations for estimating the composite likelihood and 50 ECM cycles for estimating parameters.

Hybrid zone analysis

We investigated three hybrid zones: between the South group and the West group at southwest of the Basin (the S-W hybrid zone: sites 11-23), between the North group and the West group at northwest of the Basin (the N-W hybrid zone: sites 28-29, 1), and between the East group and North & South groups (the E-NS hybrid zone: sites 5-7, 10-21). The hybridization between the East and North & South groups were treated as one zone because their boundary was difficult to define (Figure 1). The remaining sites constituted the corresponding 'pure zones' (the N&S pure zone: site 2-4; the E pure zone: site 8-9; the W pure zone: site 24-27). We demarcated the 'pure zones' and 'hybrid zones' of different lineages based on the results of genetic clustering analysis (Figure 1b, Appendix S6). If all individuals of a site had q score > 0.99 in the TESS3 analysis, the site was designated as 'pure'; results from TESS3 analysis with subsets of individuals from two interacting groups (e.g. South vs West) were used.

We examined the geographic cline of the hybrid zones using the R package 'HZAR' (Derryberry et al., 2014) and the *neutral* dataset. This method examines hybrid zone in a one-dimensional space, which is particular suitable for the S-W and N-W hybrid zones (but not for E-NS zone). For the S-W hybrid zone, the cline was defined by distances from site 14 to site

27 along the peripheral of the Basin in a clockwise direction (Figure 1). Similarly, for the N-W hybrid zone, the cline was defined by distance from site 22 to site 7. We used the mean q score from each population, estimated from TESS3 runs without the East group. Fifteen models varied in the number of cline shape parameters were tested, which represented all possible combinations of three trait interval [pMin, pMax] (fixed to 0 and 1; observed values; estimated values) and five fitting tails (none fitted; left only; right only; mirror tails; both tails estimated separately) (Derryberry et al., 2014). A null model assuming independent genetic variation was also established to ease model comparison. The Markov chain process was set to 100,000 in length with a burn-in of 10,000 and a separate seed for each model. We plotted the raw data from MCMC process to confirm convergence using a standard plot function of MCMC raw entry. The best model was selected based on Akaike information criterion score corrected for small sample size (AICc). Finally, the maximum-likelihood clines and summary statistics were extracted from the best-fit model.

We also examined the genomic cline using BGC v1.03 (Gompert & Buerkle, 2011; Gompert & Buerkle, 2012). This Bayesian genomic cline model detects movement of genetic materials from one genomic background to another within hybrid zones, and estimates introgression patterns based on two key parameters. Cline parameter α designates an increase (positive value) or decrease (negative value) in the probability of ancestry from one group to the other for a locus, whereas cline parameter β specifies an increase (positive value) or decrease (negative value) in the rate of transition from one group to the other (Gompert & Buerkle, 2012). Thus, α represents the direction of shifts in genomic clines, whereas β reflects

the strength of the barrier effect to gene flow between the two groups. We screened diagnostic SNP loci from the *full* dataset for the BGC analysis. A locus is diagnostic when its frequency difference from the two interacting groups at the opposite ends of each hybrid zone is equal to or greater than 0.75. We run five independent MCMC chains with 100,000 steps and discarded the first 75,000 steps as burn-in. Samples were recorded from the posterior distribution every 25 steps. We combined five chains after inspecting the convergence of MCMC outputs. Loci with α or β values with 95% CI significantly deviated from zero were designated as gene flow outlier loci with exceptional introgression. The input transformation and result plotting were accomplished with the help of R package 'genepopedit' (Stanley et al., 2017) and 'ClinePlotR' (available at https://github.com/btmartin721/ClinePlotR.git).

We examined the distribution of locus-specific F_{ST} for each hybrid zone. The pairwise F_{ST} was calculated between populations (with sample size 4) respectively from two sides of the cline centre and then took the average. These included sites 5-7, 11, 14, 16, 18, 20, 21 vs. 10 (n = 83) for the E-NS zone, sites 11, 14, 16, 18, 20, 21 vs. 22, 23 (n = 64) for the S-W zone, and sites 28 vs. 1 (n = 22) for the N-W zone. The pairwise F_{ST} estimated from STACKS was used and we plotted the distribution of these F_{ST} values. Furthermore, we compared the F_{ST} values for those gene flow outlier loci detected by BGC to the genome average to test if those loci showed elevated differentiation.

Finally, we examined potential coupling among the BGC outlier loci using linkage disequilibrium (LD) and locus-specific geographic cline. Significant coupling is a signature of barrier loci in a hybrid zone (Felsenstein, 1981; Barton 1983; Butlin & Smadja, 2018). We

used the squared correlation coefficient of LD (r^2), which is particularly suitable for bi-allelic loci and small or variable sample sizes (Weiss & Clark, 2002; Blackburn et al., 2017). The coefficient was calculated using VCFtools (Danecek et al., 2011). We estimated pairwise LD for the BGC outlier loci. Within each hybrid zone, the closest site to the zone centre was used to represent the hybrid zone: site 11 (n=10) for the E-NS hybrid zone, site 21 (n = 7) for the S-W hybrid zone, and site 28 (n=12) for the N-W hybrid zone. For each zone, the r^2 was calculated within every population and took the average, and the r^2 value was taken into account only if a locus was polymorphic in every population. Lastly, we compared locus-specific geographic clines for the BGC outlier loci from the N-W and S-W hybrid zones.

Results

Data

Approximately 2,767 million raw reads were obtained, and 75.16% reads were included in SNP calling. Three individuals were removed because of excessive missing data. The dataset generated from STACKS consisted of 227 individuals, with 4,979 unique loci shared by all populations. After quality control, 1,540 SNPs were retained (the *full* dataset). The final mean coverage depth was 64.11 (±25.28 SE) per individual. We further constructed a *neutral* dataset by excluding 255 F_{ST} outlier loci detected by BAYESSCAN (FDR-adjusted $p < 0.05$). The final *neutral* dataset included 1,285 SNP loci.

Population genetic structure

Descriptive statistics and pairwise F_{ST} were calculated based on the *full* dataset, and the results are provided in Appendix S3 and S4. The southern populations had relatively high heterozygosity (H_E). Furthermore, overall population differentiation was high and the pairwise F_{ST} varied between 0.037 and 0.704. The western populations had high pairwise F_{ST} with other regional populations, and varied between 0.244 and 0.704 with a median of 0.401 (SD = 0.090). Although geographically close, the western and the northern populations showed the highest differentiation (F_{ST} = 0.284-0.704). In particularly, site 28 (W) and site 1 (N) were 48.21 km apart, but the pairwise F_{ST} between them was 0.635. The eastern populations also showed high differentiation from other regional populations (F_{ST} = 0.207-0.552). The southern and the northern populations had the least population differentiation, although they were geographically far apart (Figure 1a).

Analyses of population structure were based on the *neutral* dataset. Both the network and the NJ tree produced three main groups: The West group (sites 22-29), the East group (sites 8-10), and the North & South group (sites 1-7, 11-21; Figure 3a, Appendix S5). Samples from the western side of the Sichuan Basin formed the West group (Figure 3a). Within this group, individuals from several populations were mixed together (sites 22-25, Appendix S5), but populations became more distinctive further north (sites 26-29, Appendix S5). The East group included populations from the eastern margin of the Basin (sites 8-10; Figure 3a). The North & South group included populations from both the northern and the southern sides of the Basin. Populations from the south largely mixed together while

populations from the north were more distinctive and were nested inside of the southern populations (Appendix S5).

The PCoA also revealed a three-group pattern (Figure 3b), which was consistent with the network and the NJ tree. The first PCoA axis (27.91% of the total variation) captured primary differences between the West group (sites 22-29) and the others, whereas the second axis (15.92%) discriminated the East group (sites 8-10) from the North & South group (Figure 3b). Since populations of the West and the North & South groups formed the ring distribution (more description below), we further tested patterns among them by excluding the East group. Indeed, the populations showed a clear transition pattern compatible with their geographic locations (Figure 3c).

The genetic clustering analysis from TESS3 provided a spatially explicit population structure of sampled individuals (Figure 1b, c, Appendix S6). Three clusters were detected, including a West cluster, an East cluster, and a North & South cluster (Figure 1b, c). This is consistent with results from the network and PCoA. The West cluster was the most distinctive cluster from others (Figure 1b, c; Appendix S6). The cluster remained intact even with K= 4, and had little mixing with other clusters. The East cluster was also distinctive and remained intact when K= 4, but had extensive mixing with the southern populations. Interestingly, it had little mixing with the northern populations (Figure 1b, c; Appendix S6). The northern and the southern populations shared a large proportion of their genetic makeup, although geographically they were far apart. This integration between the northern and southern populations was further demonstrated when the East cluster was excluded from the analysis

(Appendix S6). Populations from the West cluster and the North & South cluster constituted a ring-shaped distribution around the Sichuan Basin (Figure 1), which was previously detected with microsatellite loci (Qiao et al. 2018). Furthermore, the admixing between the West cluster and North & South cluster revealed two very different patterns at two contact zones: at S-W contact zone, admixture occurred across a broad geographic zone (sites 11-23), but at N-W contact zone, admixture occurred across a narrow zone (sites 1, 28-29; Figure 1b, c).

A strong IBD pattern along the ring was detected. The Mantel tests detected a significant strong correlation between the $F_{ST}/(1-F_{ST})$ values and the ring geographic distances around the Sichuan Basin ($P_{Mantel} < 0.001$, $r = 0.69$).

Population historical demography

Among the three splitting models (models 1-3), model 2 was the best, suggesting that the West group split-off first and followed by the split between the South and East groups (Figure 2a). Model 6, with recent gene flow or secondary contact, outperformed all other migration models (models 4-8) in all four pairs (Figure 2b). Model 8 with two migration matrices performed near as well as Model 6, and the recent migration rate was much larger than the ancient migration rate (Appendix S7), which was similar to the secondary contact model. Therefore, we selected Model 6 to carry over to the next steps. For the population size change modeling (models 6, 9-12), models with a population size increase for the East group (Figure 2c: E-N, E-S; Appendix S7) and a population size decrease for the North/South group (Figure 2c: W-N, W-S; Appendix S7) performed better; however, the inclusion of population

growth introduced a very large fluctuation in parameter estimates in the subsequent complex models. Subsequently, we selected a suboptimal model, model 6 with constant population size for all pairs (Figure 2c). Among the three-group interaction models (models 13-15), model 13 performed the best (Figure 2d), suggesting that the first secondary contact event between the West and North & South groups occurred earlier than the contact event between the East and South groups. Another secondary contact event between West and North & South groups occurred much later.

Figure 4 presents all parameter estimates from the final model (Model 13) for all samples. Consistent with the ring distribution, two secondary contacts events were included between the North & South and the West groups: one at the southwest (S-W), and one at the northwest (N-W). A third secondary contact event was included between the East and the North & South groups. Under Model 13 (Figure 4), the first splitting event between the South and the West groups occurred at ~1,015,213 generations ago. At ~17,579 generations later, the second split event separated the East group from the North & South group. At ~466,735 generations ago, migration between the West and South groups initiated, followed by migration between the East and South groups at ~26,796 generations ago. The migration between the North and West groups occurred only recently at ~3,059 generations ago.

Comparison of hybrid zones

The geographic cline analysis revealed very different clines between S-W and N-W hybrid zones (Figure 5). The model without a fixed tail represented the best fit model for both

hybrid zones. The centre in the S-W hybrid zone was between site 21 and 22 with a cline width of 105.40 km (Figure 5a). For the N-W hybrid zone, the cline was much narrower with a width of 24.39 km and centred at region between site 1 and 29 (Figure 5b).

The genomic cline analysis also revealed markedly different profiles for the three hybrid zones. A total of 205 diagnostic SNPs for the E-NS hybrid zone and 298 SNPs for S-W and N-W hybrid zones were detected and retained. All three showed asymmetric introgression at different extents with more positive α outlier loci than negative ones, and this was particularly true for the N-W hybrid zone (Table 1, $\chi 2$ test $P<0.00001$). In terms of introgression rate, the E-NS hybrid zone exhibited the most uniformity of the entire genome with only two outlier loci for β (Table 1, Appendix S8). Conversely, the hybrid zones of N-W and S-W showed a large number of outlier loci with restricted introgression ($\beta > 0$, Table 1, Appendix S8), and a large proportion of them (48.5%) were shared between these two zones. Furthermore, the N-W hybrid zone showed a distinct profile with the largest number of β outlier loci and the strongest asymmetric introgression. Also, none of the positive β loci showed any asymmetric movement, and only 13 loci with negative β were in asymmetric movement ($\alpha > 0$; Appendix S8). In the S-W hybrid zone, 2 positive β outlier loci were with high probability of ancestry from the West ($\alpha < 0$), and 8 negative β outlier loci were with high probability of ancestry from the North & South ($\alpha > 0$).

The F_{ST} distributions of the E-NS and S-W hybrid zones were similar with left-skewed distribution and dominated by low values (Figure 6). For the N-W hybrid zone, the F_{ST} had a bimodal distribution with peaks near $F_{ST} = 0$ and $F_{ST} = 1$ (Figure 6). Of the 1540 loci analysed,

191 loci had fixed differences within N-W zone. Furthermore, the differentiation across loci was linked with patterns of introgression. The genomic cline outlier loci demonstrated a higher differentiation than the average genomic level (Table 1), particularly the β outliers. Among all BGC outliers, the positive β loci had the highest F_{ST} (except in the E-NS hybrid zone, which had only one positive β locus).

As expected, the LD levels of BGC outliers were markedly higher than average genomic level (Table 1). Furthermore, we noticed that there were much fewer polymorphic outlier loci in the N-W zone (25/189) than in the S-W zone (95/167) and many loci were fixed. We identified three large linkage groups from the S-W hybrid zone involving a total of 20 loci with extreme LD (Appendix S9). All of those loci were BGC outliers shared by the S-W and N-W hybrid zones, with the majority being β positive outliers. Thirteen of these loci were fixed and 12 had F_{ST} of 1 in the N-W zone (Appendix S10).

The locus-specific geographic cline analyses further demonstrated strong coupling patterns among loci with reduced introgression ($\beta > 0$) in both S-W and N-W hybrid zones. The pattern was particularly strong in the N-W zone; all loci (except one) revealed nearly identical narrow and steep clines with a small variability for the centre and width (Table 1, Appendix S11). Furthermore, other outlier loci also showed pattern of coupling in the N-W zone, but not in the S-W zone (Appendix S11).

Discussion

1. Historical demography and the formation of the ring-shaped divergence

Our data and analysis confirmed the ring-shaped divergence of the Green Odorous Frog around the Sichuan Basin (Qiao et al., 2018), but revealed substantial discontinuity along the ring. The populations are divided into three distinctive groups, the East, the West, and the North & South group, and the latter two groups form the ring distribution (Figure 1). Demographic analysis suggested that the three groups remained in isolation for extended time and only recently became re-connected (Figure 4). The re-connection produced three contact zones, and substantial admixture occurred in two of them. At the third contact zone, the North and West groups met at the northwestern corner of the Basin but only limited hybridization occurred, and partial reproductive isolation may have developed.

Past climatic changes, such as glaciation, likely caused the initial isolation of the three groups (Qiao et al., 2018), and each group has a separate refugial history. The West group includes sites 22-29 (Figure 1a), which are distributed along the western side of the Basin. All analyses consistently showed that they are closely related and form a cohesive group (clustering analyses, Figure 1b, Appendix S6; PCoA, Figure 3b, c; network and tree, Figure 3a, Appendix S5). This is consistent with the early results from microsatellite DNA data (Qiao et al., 2018). Nevertheless, there are subtle variations within this group. Populations from southern locations (sites 22-26) have higher genetic diversity than populations from northern locations (site 27-28; Appendix S3). This is likely a consequence of a range expansion from south to north (Qiao et al., 2018). There is a dense and reticulate water network in this region, which may function as corridor and facilitate the dispersal (riverine corridor hypothesis; Ye et al., 2018). Furthermore, the southmost population (site 22) has the

highest diversity (Appendix S3), which are likely results from admixture with populations of the South group.

The populations from the south (sites 11-21) and north (sites 1-7) form the second group (the North & South group) and are probably originated in another refugium. The northern populations are likely derived from the south and are results of a northward range expansion (Qiao et al., 2018). The NJ tree most clearly demonstrates the pattern, with the northern individuals nesting inside the southern individuals (Appendix S5). Although the NJ tree was constructed without an outgroup, the demographic analysis clearly demonstrated that the West group split off first and hence the root was placed between the West group and the rest. The dual northward expansions of this group and the West group eventually produced the ring-shaped distribution (Figure 1). A strong IBD pattern ($r = 0.69$) around the Basin suggests that the genetic exchanges among the chain population are mostly gradual and continuous, and the PCoA also shows gradual transformation (Figure 3c). Our results are consistent with those from the early study by Qiao et al. (2018). The Huaying Mountains may act as an important dispersal corridor between the south and the north (Figure 1a). The species distribution model clearly demonstrated that there are ample suitable habitats in these mountains (Qiao et al., 2018). The current population status at the Huaying Mts. is unclear. We discovered one distribution record with a single specimen deposited in the China West Normal University (Hu & Deng, 1990), but we have not been able to recover any samples from this region. Another interesting aspect of this group is that the southern populations have extensive gene exchange with both the West and East group (Figures 1& 4), which resulted in

455 the southern populations having the highest genetic diversity and being an evolutionary
456 melting pot (Qiao et al., 2018).
457 Three populations from southeastern corner of the Basin (sites 8-10) form the East
458 group (Figure 1). This group was not detected by the microsatellite data (Qiao et al., 2018).
459 The current range of this group is restricted to the southeastern corner of the Basin; however,
460 the admixture between the South and East groups occurs most intensely at the southern sites
461 (sites 11, 12, 13, Figure 1b, c), suggesting that the contemporary or historical range of the
462 East group are likely much larger and further south and/or east. There are several distribution
463 records in central and southern China, which are further east and south of the Sichuan Basin
464 (Fei et al., 2009). Our expeditions unfortunately did not recover any samples; if confirmed,
465 these populations are possibly closely related to the East group. Another interesting aspect of
466 this group is that it has limited exchange with the North group even within close geographic
467 proximity (i.e., sites 5, 6, 7), despite its extensive genetic exchange with the South group
468 (Figure 1a). A likely cause is a natural contemporary genetic barrier, the Yangtze River at the
469 Three Gorges area. The water flow is extremely fast in this section of the river, which may
470 prevent frogs from crossing. The connection between the South and North groups, as well as
471 the historical south to north expansion, likely across the river at a more interior location of the
472 Basin, where the river is wider but flatter. The northern population may further expand
473 eastward to their current location at north of the Three Gorges area. The deep gorges can be a
474 major cause for habitat segmentation and act as dispersal barrier, and major river as dispersal

barrier for amphibians have been well documented (the riverine barrier hypothesis, e.g. Funk et al., 2007; Gehring et al., 2012).

Based on all evidence, we reconstructed a scenario that has produced the current ring-shaped distributional pattern. The ancestral populations of the Green Odorous Frog might have been widely distributed in mountains of southwestern China. The first splitting event between the western populations and the rest occurred at ~2.03 million years ago (MYA; 1,015,213 generations, estimated with 2 years/generation). Shortly after, the South lineage split from the East at 997,634 generations ago. The region experienced intense geological activities with tectonic uplift and climatic oscillations in the late Pleistocene (0.78-2.58 MYA), which might have produced the initial isolation. Nevertheless, many localized areas in the southwestern mountains of China experienced mild climatic oscillation throughout the Pleistocene, and the presence of stable habitats in the region is likely the primary cause of lineages isolation and diversity maintenance (Hewitt, 1996, 2000; Li & Fang, 1999). The admixture between the West and the South groups occurred towards the end of Pleistocene (~0.93 MYA, 466,735 generations). In the meantime, populations from the southern regions expanded northward, and some of them passed through the fold-and-valley of the eastern Sichuan Basin (i.e. the Huaying Mts.) and further spread to the northern rim of the Basin. The reconnection between the East and South groups occurred during the expansion of southern populations at ~54,000 years ago (26,796 generations). Finally, the expanding fronts of the West group and North group met at the northwestern corner of the Basin at approximately 6,000 years ago (3,059 generation). With a large differentiation generated from the initial

isolation and along the dispersal, the two groups produced only limited hybridization at the contact zone.

The extended isolation produced the initial divergence, which is probably essential in the formation of the current ring-species like pattern. The most recent common ancestor of the chain populations is estimated at ~2.03 Ma, which is younger than any known ring species (Alcaide et al., 2014; Kuchta et al., 2009). The spatial distribution of the species, with a perimeter of ~1000 km, is also much smaller than other known ring species (a 'micro-ring'). The early prolonged isolation probably is necessary to produce the observed levels of divergence that eventually developed into partial reproductive isolation.

2. Hybrid zone dynamics and speciation continuum

The Green Odorous Frog presents an intraspecific system with three replicate contact zones at different levels of differentiation. Furthermore, two of the zones are along its ring, and the belt-shaped distribution makes the hybrid zone analyses more tangible. This case provides excellent opportunities to examine what happened to diverged lineages after re-connection.

The E-NS hybrid zone exhibits extensive admixture, particularly at the south. Genetic elements from the East group is detected in all southern populations (Figure 1), and the East and South groups have the lowest level of divergence (pairwise F_{ST} = 0.207-0.305; Appendix S4), compared to other interacting groups. Furthermore, we detected only one barrier gene in this hybrid zone ($\beta > 0$; Table 1). The number of barrier gene is directly related to the overall

effective selection against hybrids (Barton, 1983; Barton & Gale, 1993), and thus the barrier effect is the least among the three hybrid zones. The admixture likely occurred over a large area, potentially through multiple contact points at the southern area of the species' distribution. As discussed earlier, the contemporary or historical range of the East group is potentially much larger and several distribution records from southern China may belong the East group.

The S-W and N-W hybrid zones are two zones along the ring distribution, and likely represent two stages along a speciation continuum. At the contact fronts, the F_{ST} ranges 0.245-0.364 in the S-W zone, and ranges 0.559-0.704 in the N-W zone (Appendix S4). Both the S-W and N-W hybrid zones possess large numbers of barrier loci ($\beta > 0$; Table 1), and 48.5% of these loci (136 in total) are shared between the two zones. Previous studies of multiple hybrid zones between two species have suggested the shared markers with restricted gene flow could be linked to processes important for reproduction, and likely play roles in the initial phase of speciation (e.g., sunflowers, Buerkle & Rieseberg, 2001; crikets, Larson et al., 2014). The barrier effect appears to be enhanced along the northward expansion of the ring as the N-W zone has more barrier loci than the S-W zone (Table 1). Furthermore, these barrier loci show strong coupling with high LD (Table 1). It has been widely accepted that association among different barriers to gene flow is a key to the speciation process (Mayr 1963, Butlin & Smadja, 2018), and the buildup of LD among selected loci increases the efficacy of selection (Barton 1983). The strong coupling process is also reflected in the near identical and steep geographic cline for these loci in the N-W hybrid zone (Appendix S11). Moreover, the

coupling may have led to elevated divergence. These barrier loci have higher F_{ST} than the genomic average (Table 1), and many highly linked barrier loci reached fixation in the N-W zone (F_{ST} = 1; Appendix S9, S10). Such accumulation of differentiation at linked sites very likely enhanced reproduction isolation in the N-W zone.

The fates of the S-W and N-W hybrid zones are likely quite different: while the S-W zone is merging, the N-W zone is developing reproductive isolation. The S-W hybrid zone shows extensive exchange of genetic material across a large area (Figure 1), and has a wide geographic cline (105.40 km; Figure 5). On the other hand, the N-W hybrid zone has a narrow and steep geographic cline (24.39 km), which suggests low introgression rate. Its distribution pattern of locus-specific F_{ST} suggests that it is likely at a late stage of speciation with a substantial collection of loci in fixation (Figure 6; Wu, 2001; Seehausen et al., 2014). We would like to argue that the level of divergence of the N-W hybrid zone may have crossed a critical threshold (Gavrilets, 2004; or 'tipping point', Nosil et al., 2017), and is destined for a complete reproductive isolation. The overall level of divergence between the North and West groups is high (average pairwise F_{ST} = 0.451 and as high as 0.704 at contact front; Appendix S4). Additionally, specific genes (barrier loci, 'speciation genes') may be more important than the overall divergence (Felsenstein, 1981; Kirkpatrick & Servedio, 1999; Coyne, Coyne, & Orr, 2004; Nosil & Schluter, 2011). A large proportion of strongly coupled barrier loci have reached fixed difference. These hybridizing populations are likely on a positive feedback loop that will further enhance the divergence in the hybrid zone (Nosil et al., 2017).

Another interesting observation of these hybrid zones is the asymmetric introgression. All three zones have asymmetric introgression, but the N-W zone is particularly so (α outlier loci, Table 1; Figure 4). Selection is the primary mechanisms in determining patterns of hybridization and introgression in many species (Ballard & Whitlock, 2004; Lexer & Widmer, 2008). However, we found very few overlaps between α and β outlier loci and none between β positive loci (barrier) and α outlier loci (asymmetric movement) (Appendix S8). The small geographic range of the ring distribution and our fieldwork did not suggest any obvious abrupt ecological change along the ring. Alternatively, asymmetric introgression may reflect the genomic footprint left by the receding species in the wake of a moving hybrid zone (Buggs 2007). However, determining the cause of asymmetric introgression is difficult (Barton & Hewitt, 1985; Buggs 2007), and repeated sampling over time will be needed.

3. Conclusion and future directions

The Green Odorous Frog system provides three replicate contact zones with various levels of divergence within a single species, including one zone with partial reproductive isolation. This is unique comparing to other replicate systems, such as the cichlid fishes that involve multiple species pairs or the pea aphid system that involves a single species but with limited divergence. Furthermore, its ring-shaped distribution around the Sichuan Basin makes the contact zone close to be one-dimensional. This makes hybrid zone analysis much more tangible. Additionally, it adds another intriguing factor, migration, to the system. Migration occurs not only across the hybrid zone, but also through the chain populations of the ring. If

gene flow indeed enhances the reinforcement process (Dobzhansky, 1940; Nosil et al., 2017), this added gene flow through the ring may provide a unique opportunity to examine reinforcement.

Several key questions remain to be answered. For example, what have caused the partial reproductive isolation at the N-W hybrid zone, mate choice or genetic incompatibility? What are the functions of the coupled barrier loci shared between the S-W and N-W hybrid zones? Which process is more important in speciation, divergence of speciation genes (or key genetic elements) or loci coupling? We hope this unique study system will help us to address many important questions regarding the speciation continuum.

Acknowledgements

We are grateful to Dr. Z. Guo and G. Yang of the Kunming Institute of Botany for lab assistance, and to Dr. W. Yang of the Lund University for helping with data analyses. Most of the analyses benefit from the computing platform funded by the Science and Technology Support Project of Sichuan Province, China (2016NZ0103). Drs. P. Guo, X. Chen, and B. Zhu kindly provided us tissue samples, and X. Qiu, X., Zhao, D. Zhou, Y. Wu assisted in fieldwork. This project is supported by the National Natural Science Foundation of China (grant 31172061 to JF) and NSERC Canada (Discovery grant to JF).

References

Alcaide, M., Scordato, E. S., Price, T. D., & Irwin, D. E. (2014). Genomic divergence in a ring species complex. *Nature, 511*(7507), 83.

Ballard, J. W. O., & Whitlock, M. C. (2004). The incomplete natural history of mitochondria. *Molecular Ecology*, *13*(4), 729–744.

Barton, N. H. (1983). Multilocus clines. *Evolution, 37*(3), 454-471.

Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, *16*(1), 113–148.

Barton, N. H., & Gale, K. S. (1993). Genetic analysis of hybrid zones. *Hybrid Zones and the Evolutionary Process*, 13–45.

Berlocher, S. H., & Feder, J. L. (2002). Sympatric speciation in phytophagous insects: Moving beyond controversy? *Annual Review of Entomology*, 47(1), 773–815. https://doi.org/10.1146/annurev.ento.47.091201.145312

Blackburn, G. S., Brunet, B. M., Muirhead, K., Cusson, M., Béliveau, C., Levesque, R. C., Lumley, L. M., & Sperling, F. A. (2017). Distinct sources of gene flow produce contrasting population genetic dynamics at different range boundaries of a Choristoneura budworm. *Molecular Ecology*, *26*(23), 6666–6684.

Buerkle, C. A., & Rieseberg, L. H. (2001). Low intraspecific variation for genomic isolation between hybridizing sunflower species. *Evolution*, *55*(4), 684–691.

Buggs, R. J. A. (2007). Empirical study of hybrid zone movement. *Heredity*, *99*(3), 301–312.
Butlin, R. (1987). Speciation by reinforcement. *Trends in Ecology & Evolution, 2*(1), 8-13.
Butlin, R., & Smadja C. (2018) Coupling, reinforcement, and speciation. *American Naturalist*, 191 (2), 155-172.
Cain, A. J. (1954). Animal species and their evolution (Hutchinson). *London, L954*.
Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Mol Ecol, 22*(11), 3124-3140. doi:10.1111/mec.12354
Caye, K., Deist, T. M., Martins, H., Michel, O., & François, O. (2016). TESS3: fast inference of spatial population structure and genome scans for selection. *Molecular Ecology Resources, 16*(2), 540-548.
Clausen, J. (1951). Stages in the evolution of plant species. *Stages in the Evolution of Plant Species.*, *6d*.
Coyne, J. A., Coyne, H. A., & Orr, H. A. (2004). *Speciation*. Oxford University Press, Incorporated.
Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection Or the Preservation of Favoured Races in the Struggle for Life*. H. Milford; Oxford University Press.
Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., …, Durbin R. (2011). 1000 Genomes Project Analysis Group. The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330
Derryberry, E. P., Derryberry, G. E., Maley, J. M., & Brumfield, R. T. (2014). HZAR: hybrid zone analysis using an R software package. *Molecular Ecology Resources, 14*(3), 652-663.
Dobzhansky, T. (1937). *Genetics and the Origin of Species*. Columbia University Press. New York.
Dobzhansky, T. (1940). Speciation as a stage in evolutionary divergence. *The American Naturalist, 74*(753), 312-321.
Dray, S., & Dufour, A.-B. (2007). The ade4 Package: Implementing the Duality Diagram for Ecologists. *Journal of Statistical Software*, *22*(4), 1–20. https://doi.org/10.18637/jss.v022.i04
Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10(3), 564-567.
Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust Demographic Inference from Genomic and SNP Data. *PLoS Genetics, 9*(10), e1003905. doi:10.1371/journal.pgen.1003905
Fei, L., Ye, C., Huang, Y., & Liu, M. (2009). *Atlas of amphibians of China*. Henan Science and Technology Press.

Felsenstein, J. (1981). Skepticism towards Santa Rosalia, or why are there so few kinds of animals? *Evolution*, *35*(1), 124–138.
Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics, 180*(2), 977-993. doi:10.1534/genetics.108.092221
Funk, W. C., Caldwell, J. P., Peden, C. E., Padial, J. M., De la Riva, I., & Cannatella, D. C. (2007). Tests of biogeographic hypotheses for diversification in the Amazonian forest frog, *Physalaemus petersi*. *Molecular Phylogenetics and Evolution, 44*(2), 825-837.
Gavrilets, S. (2004). *Fitness landscapes and the origin of species (MPB-41)* (Vol. 41). Princeton University Press.
Gavrilets, S. (2014). Models of speciation: Where are we now? *Journal of Heredity*, *105*(S1), 743–755.
Gehring, P.-S., Pabijan, M., Randrianirina, J. E., Glaw, F., & Vences, M. (2012). The influence of riverine barriers on phylogeographic patterns of Malagasy reed frogs (*Heterixalus*). *Molecular phylogenetics and evolution, 64*(3), 618-632.
Gompert, Z., & Buerkle, C. A. (2012). bgc: Software for Bayesian estimation of genomic clines. *Molecular Ecology Resources*, *12*(6), 1168–1176.
Gompert, Zachariah, & Buerkle, C. A. (2011). Bayesian estimation of genomic clines. *Molecular Ecology*, *20*(10), 2111–2127.
Gruber, B., Unmack, P. J., Berry, O. F., & Georges, A. (2018). dartr: An r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Molecular Ecology Resources, 18*(3), 691-699. doi:10.1111/1755-0998.12745
Hewitt, G. M. (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological journal of the Linnean Society, 58*(3), 247-276.
Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature, 405*(6789), 907.
Hijmans, R. J. (2016). geosphere: spherical trigonometry. R package version 1.5-5. *See https://cran. r-project. org/package= geosphere*.
Huson, D. H., & Bryant, D. (2005). Application of Phylogenetic Networks in Evolutionary Studies. *Molecular Biology and Evolution, 23*(2), 254-267. doi:10.1093/molbev/msj030
Hu, G., Deng, Q. X. (1990). Amphibian Survey Report of Huaying Mount, Jinyun Mount and Yucheng Mount. *Sichuan Journal of Zoology*, 9(1), 36-37.
Irwin, D. E., Bensch, S., & Price, T. D. (2001). Speciation in a ring. *Nature*, *409*(6818), 333.
Kirkpatrick, M., & Servedio, M. R. (1999). The reinforcement of mating preferences on an island. *Genetics, 151*(2), 865-884.
Kuchta, S. R., Parks, D. S., Mueller, R. L., & Wake, D. B. (2009). Closing the ring: Historical biogeography of the salamander ring species *Ensatina eschscholtzii*. *Journal of Biogeography*, *36*(5), 982–995.
Kuchta, S. R., & Wake, D. B. (2016). Wherefore and Whither the Ring Species? *Copeia*, *104*(1), 189–201. https://doi.org/10.1643/OT-14-176

Lanier, H. C., Massatti, R., He, Q., Olson, L. E., & Knowles, L. L. (2015). Colonization from divergent ancestors: glaciation signatures on contemporary patterns of genomic variation in Collared Pikas (*Ochotona collaris*). *Molecular ecology, 24*(14), 3688-3705.

Larson, E. L., White, T. A., Ross, C. L., & Harrison, R. G. (2014). Gene flow and the maintenance of species boundaries. *Molecular Ecology*, *23*(7), 1668–1678.

Lexer, C., & Widmer, A. (2008). The genic view of plant speciation: Recent progress and emerging questions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1506), 3023–3036.

Li, J., & Fang, X. (1999). Uplift of the Tibetan Plateau and environmental changes. *Chinese Science Bulletin, 44*(23), 2117-2124.

Maan, M. E., & Seehausen, O. (2011). Ecology, sexual selection and speciation. *Ecology letters*, 14(6), 591-602.

Mayr, E. (1942). *Systematics and the Origin of Species*. New York: Columbia University Press.

Meirmans, Patrick G. (2015). Seven common mistakes in population genetics and how to avoid them." *Molecular Ecology*, 24(13):3223–31. doi: 10.1111/mec.13243.

Merrill, R. M., Gompert, Z., Dembeck, L. M., Kronforst, M. R., McMillan, W. O., & Jiggins, C. D. (2011). Mate preference across the speciation continuum in a clade of mimetic butterflies. *Evolution: International Journal of Organic Evolution, 65*(5), 1489-1500.

Nazareno, A. G., Bemmels, J. B., Dick, C. W., & Lohmann, L. G. (2017). Minimum sample sizes for population genomics: An empirical study from an Amazonian plant species. *Molecular Ecology Resources*, *17*(6), 1136–1147. https://doi.org/10.1111/1755-0998.12654.

Nosil, P., & Schluter, D. (2011). The genes underlying the process of speciation. *Trends in Ecology & Evolution, 26*(4), 160-167.

Nosil, P. (2012). *Ecological speciation*. USA: Oxford University Press.

Nosil, P., Feder, J. L., Flaxman, S. M., & Gompert, Z. (2017). Tipping points in the dynamics of speciation. *Nature ecology & evolution, 1*(2), 0001.

Paradis E, Claude J, Strimmer K (2004) ape. Analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20(1), 289– 290.

Peccoud, J., Ollivier, A., Plantegenest, M., & Simon, J.-C. (2009a). A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. *Proceedings of the National Academy of Sciences, 106*(18), 7495-7500.

Peccoud, J., Simon, J.-C., McLaughlin, H. J., & Moran, N. A. (2009b). Post-Pleistocene radiation of the pea aphid complex revealed by rapidly evolving endosymbionts. *Proceedings of the National Academy of Sciences, 106*(38), 16315-16320.

Pereira, R. J., & Wake, D. B. (2009). Genetic leakage after adaptive and nonadaptive divergence in the *Ensatina eschscholtzii* ring species. *Evolution: International Journal of Organic Evolution*, *63*(9), 2288–2301.

739 Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double
740 digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in
741 model and non-model species. *PLoS ONE, 7*(5), e37135.
742 doi:10.1371/journal.pone.0037135
743 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., . . . Sham,
744 P. C. (2007). PLINK: a tool set for whole-genome association and population-based
745 linkage analyses. *American journal of human genetics, 81*(3), 559-575.
746 doi:10.1086/519795
747 Qiao, L., Wen, G., Qi, Y., Lu, B., Hu, J., Song, Z., & Fu, J. (2018). Evolutionary melting pots
748 and reproductive isolation: A ring−shaped diversification of an odorous frog
749 (*Odorrana margaretae*) around the Sichuan Basin. *Molecular ecology, 27*(23), 4888-
750 4900.
751 Rice, W. R., & Hostert, E. E. (1993). Laboratory experiments on speciation: what have we
752 learned in 40 years? *Evolution, 47*(6), 1637-1653.
753 Shaw, K. L., & Mullen, S. P. (2014). Speciation Continuum. *Journal of Heredity*, *105*(S1),
754 741–742. https://doi.org/10.1093/jhered/esu060
755 Schluter, D. (2000). *The ecology of adaptive radiation*. USA: OUP Oxford.
756 Seehausen, O., Takimoto, G., Roy, D., & Jokela, J. (2008). Speciation reversal and
757 biodiversity dynamics with hybridization in changing environments. *Molecular*
758 *ecology, 17*(1), 30-44.
759 Seehausen, O. (2009). Progressive levels of trait divergence along a'speciation transect'in the
760 Lake Victoria cichlid fish *Pundamilia*.
761 Seehausen, O., & Magalhaes, I. (2010). Geographical mode and evolutionary mechanism of
762 ecological speciation in cichlid fish. *In Search of the Causes of Evolution: From Field*
763 *Observations to Mechanisms* (P.R. Grant and B.R. Grant, eds.), Princeton,
764 NJ:Princeton University Press, 282–308.
765 Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A.,
766 … Brännström, Å. (2014). Genomics and the origin of species. *Nature Reviews*
767 *Genetics*, *15*(3), 176.
768 Sobel, J. M., Chen, G. F., Watt, L. R., & Schemske, D. W. (2010). The biology of speciation.
769 *Evolution: International Journal of organic evolution*, 64(2), 295-315.
770 Stanley, R. R., Jeffery, N. W., Wringe, B. F., DiBacco, C., & Bradbury, I. R. (2017).
771 genepopedit: A simple and flexible tool for manipulating multilocus molecular data in
772 R. *Molecular Ecology Resources*, *17*(1), 12–18.
773 Sun, Y.-B., X., Zi-Jun, Xiang, Xue-Yan, Liu, Shi-Ping, Zhou, Wei-Wei, Tu, Xiao-Long, …
774 Zhang, Ya-Ping. (2015). Whole-genome sequence of the Tibetan frog *Nanorana*
775 *parkeri* and the comparative evolution of tetrapod genomes. *Proceedings of the*
776 *National Academy of Sciences of The United States of America, 112*(11), 6.
777 doi:10.1073/pnas.150176411210.5524/100132
778 Tazzyman, S. J., & Iwasa, Y. (2010). Sexual selection can increase the effect of random
779 genetic drift—A quantitative genetic model of polymorphism in *Oophaga pumilio*, the

780 strawberry poison−dart frog. *Evolution: International Journal of Organic Evolution*,
781 64(6), 1719-1728.
782 Wake, D. B. (1997). Incipient species formation in salamanders of the Ensatina complex.
783 *Proceedings of the National Academy of Sciences*, *94*(15), 7761–7767.
784 Walsh, B. D. (1861). *On phytophagic varieties and phytophagic species*.
785 Weir, B. S. (1996). *Genetic Data Analysis II: Methods for Discrete Population Genetic Data*.
786 Sinauer Associates.
787 Weiss, K. M., & Clark, A. G. (2002). Linkage disequilibrium and the mapping of complex
788 human traits. *TRENDS in Genetics*, *18*(1), 19–24.
789 Wickham, H., & Chang, W. (2008). ggplot2: An implementation of the Grammar of Graphics.
790 *R package version 0.7, URL: http://CRAN. R-project. org/package= ggplot2*.
791 Willing, E.-M., Dreyer, C., & Van Oosterhout, C. (2012). Estimates of genetic differentiation
792 measured by FST do not necessarily require large sample sizes when using many SNP
793 markers. *PLoS one, 7*(8), e42649.
794 Wu, C.-I. (2001). The genic view of the process of speciation. *Journal of Evolutionary*
795 *Biology*, *14*(6), 851–865. https://doi.org/10.1046/j.1420-9101.2001.00335.x
796 Ye, Z., Yuan, J., Li, M., Damgaard, J., Chen, P., Zheng, C., Yu, H., Fu, S., & Bu, W. (2018).
797 Geological effects influence population genetic connectivity more than Pleistocene
798 glaciations in the water strider *Metrocoris sichuanensis* (Insecta: Hemiptera:
799 Gerridae). *Journal of Biogeography*, *45*, 1–12. https://doi.org/10.1111/jbi.13148
800

801 **Data Accessibility**

802 -Demultiplexed fastq files of ddRADseq sequences have deposited at NCBI SRA under

803 Bioproject PRJNA598960 (https://dataview.ncbi.nlm.nih.gov/object/PRJNA598960?

804 reviewer=hepb3q1isj17j9qohtgql7hhkf)

805 -Input files used for FASTSIMCOAL2.6 and observed site frequency spectra are available

806 on Figshare under doi: 10.6084/m9.figshare.14129987.

807

808

809 **Author Contributions**

810 GW collected the data, conducted most of the analysis, and drafted the manuscript. JF
811 conceived the project, participated data analysis, and finalized the manuscript.

812

Table 1. Results of the hybrid zone comparison analyses. Bayesian genomic cline (BGC) analysis is based on 205 diagnostic loci for the E-NS hybrid zone and 298 for the S-W/N-W hybrid zones. Number of BGC outlier loci and the range of values for cline parameters are displayed. F_{ST} values are pairwise F_{ST} between populations on both sides of the cline centre in each hybrid zone. F_{ST} of E-NS zone are from sites 5-7, 11, 14, 16, 18, 20, 21 vs. 10; F_{ST} of S-W zone are from sites 11, 14, 16, 18, 20, 21 vs. 22, 23; F_{ST} of N-W zone are from sites 28 vs. 1. The squared correlation coefficient of LD (r^2) values are locus average from a representative population near the cline centre of each hybrid zone. F_{ST} and r^2 values, geographic cline centre and width for genomic cline outliers in different hybrid zones are compared. For r^2 and F_{ST} values, the standard deviations are indicated in parentheses.

	Genome-level	$\alpha > 0$	$\alpha < 0$	$\beta > 0$	$\beta < 0$
E-NS zone					
# of BGC outlier loci	-	46 (E→S)	32 (S→E)	1	1
BGC parameter values	-	1.597 – 4.548	-5.586 – -2.031	4.061	-4.984
F_{ST}	0.182 (0.247)	0.474 (0.362)	0.835 (0.249)	0.758	0.593
r^2 *(site 11)*	0.115 (0.147)	0.377 (0.176)	0.365 (0.116)	-	-
S-W zone					
# of BGC outlier loci	-	40 (S→W)	21 (W→S)	90	26
BGC parameter values	-	1.120 – 4.314	-3.177 – -1.338	2.695 – 13.281	-10.861 – -4.687
F_{ST}	0.194 (0.254)	0.257 (0.168)	0.139 (0.159)	0.789 (0.129)	0.590 (0.234)
r^2 *(site 21)*	0.190 (0.220)	0.477 (0.208)	0.486 (0.159)	0.483 (0.274)	0.395 (0.163)
Geographic cline centre	354.0 – 464.4	261.1 – 487.0	319.7 – 443.2	369.2 – 459.7	329.4 – 447.8
Geographic cline width	16.5 – 207.3	80.2 – 495.8	56.6 – 292.2	30.5 – 175.7	58.2 – 285.8
N-W zone					
# of BGC outlier loci	-	32 (N→W)	0 (W→N)	112	58
BGC parameter values	-	1.176 – 3.525	-	4.724 – 11.158	-10.659 – -3.927
F_{ST}	0.308 (0.398)	0.722 (0.334)	-	0.920 (0.114)	0.835 (0.242)
r^2 *(site 28)*	0.230 (0.290)	0.504 (0.217)	-	0.659 (0.252)	0.603 (0.257)
Geographic cline centre	358.7 – 399.8	338.9 – 385.5	-	361.6 – 407.8	356.0 – 401.7

Geographic cline width	3.9 – 62.8	6.3 – 81.8	3.1 – 58.0	2.5 – 61.8

Figure captions

Figure 1. (a) Map of western China with sampling sites of *O. margaretae*. (b) Individual assignment probability bar plot at the best-fit (K = 3), inferred by TESS3 and based on the *neutral* dataset (1,285 SNPs). Numbers 1-29 are sampling sites, which are arranged in a ring shape by their relative locations around the Sichuan Basin. Individual samples are grouped by their genetic composition and geographic location (North & South, East, and West), or grouped by their hybrid status (pure zone or hybrid zone). (c) Spatial interpolation of population ancestry coefficients across the geographic distribution.

Figure 2. Demographic simulation analysis using FASTSIMCOAL26. Four sets of comparison of 15 models with different parameter combinations were conducted. Distribution of the top 10 minimum AIC values from 50 replicate simulations are presented. (a) Models of three alternative splitting scenarios among the West (W), North & South (NS) and East (E) groups. (b) Five pairwise interaction models with migration. (c) Five pairwise interaction models with recent migration and population size change. For (b-c), 'P1' and 'P2' represent the groups applied in pairwise modelling, and their corresponding group pairs are indicated at the right side of figure. (d) Splitting models with secondary contact events happened in different order. Black lines in dots represent the time nodes for different historical events (e.g., migration, bottleneck). Lines with arrow indicate directions of gene flow.

Figure 3. Population differentiation in *O. margaretae*. (a) A Neighbor-Net network of all individuals. Numbers 1-29 are sampling sites; (b) The first two axes of PCoA for all samples; (c) The first two axes of PCoA when the East group is excluded. All analyses are based on the *neutral* dataset (1,285 SNPs).

Figure 4. A schematic representation of the historical demography of *O. margaretae* around the Sichuan Basin (time is not to scale). Polygons represent populations, whereas the width indicates effective population size. Nanc is the effective population size of the most recent common ancestor to all populations; NE&NS refers to the size of the common ancestor of the East and North & South groups. 'Tmrca' refers to the time to the most common ancestor of three groups. 'TE&NS' refers to the time to the most common ancestor for East and North & South groups. 'Tsc1' refers to the time of the first migration event between the North & South and West groups. 'Tsc2' refers to the time of the reconnection event between the East and South groups. 'Tsc3' refers to the time of the second migration event between the North & South and West groups. Lines with arrow indicate the directions of gene flow, and the numbers above the lines are numbers of effective immigrants per generation. All parameters were estimated from the simulation run with the highest likelihood in 50 repeats.

Figure 5. Geographic clines of the S-W and N-W hybrid zones inferred using HZAR. Dashed vertical lines show maximum-likelihood estimates of cline centre and width. All distances are approximate estimates following the edge of the Sichuan Basin.

Figure 6. Distributions of locus-specific pairwise F_{ST} of three hybrid zones.