Two transects reveal remarkable variation in gene flow on opposite ends of a European toad hybrid zone

Riemsdijk, I. van; Arntzen, J.W.; Bucciarelli, G.M.; McCartney-Melstad, E.; Rafajlović, M.; Scott, P.A.; ... ; Wielstra, B.M.

the geneticssociety

www.nature.com/hdy

ARTICLE

Check for updates

Two transects reveal remarkable variation in gene flow on opposite ends of a European toad hybrid zone

I. van Riemsdijk[1,2,3]✉, J. W. Arntzen[1,2], G. M. Bucciarelli[4,5,6,7], E. McCartney-Melstad[4,5], M. Rafajlović[8,9], P. A. Scott[4,10], E. Toffelmier[4,5], H. B. Shaffer[4,5] and B. Wielstra[1,2]

Speciation entails a reduction in gene flow between lineages. The rates at which genomic regions become isolated varies across space and time. Barrier markers are linked to putative genes involved in (processes of) reproductive isolation, and, when observed over two transects, indicate species-wide processes. In contrast, transect-specific putative barrier markers suggest local processes. We studied two widely separated transects along the 900 km hybrid zone between *Bufo bufo* and *B. spinosus*, in northern and southern France, for ~1200 RADseq markers. We used genomic and geographic cline analyses to identify barrier markers based on their restricted introgression, and found that some markers are transect-specific, while others are shared between transects. Twenty-six barrier markers were shared across both transects, of which some are clustered in the same chromosomal region, suggesting that their associated genes are involved in reduced gene flow across the entire hybrid zone. Transect-specific barrier markers were twice as numerous in the southern than in the northern transect, suggesting that the overall barrier effect is weaker in northern France. We hypothesize that this is consistent with a longer period of secondary contact in southern France. The smaller number of introgressed genes in the northern transect shows considerably more gene flow towards the southern (*B. spinosus*) than the northern species (*B. bufo*). We hypothesize that hybrid zone movement in northern France and hybrid zone stability in southern France explain this pattern. The *Bufo* hybrid zone provides an excellent opportunity to separate a general barrier effect from localized gene flow-reducing conditions.

Heredity (2023) 131:15–24; https://doi.org/10.1038/s41437-023-00617-6

INTRODUCTION

Hybrid zones provide windows into the processes driving the outcome of speciation (Hewitt 1988). As the genomic toolkit available to evolutionary biologists has grown, it has become increasingly clear that different genes do not flow at equal speed across a hybrid zone. Some genomic regions introgress freely, some gene variants cannot traverse a hybrid zone at all, and the direction of gene flow between taxa may vary at different points along the same hybrid zone (Baack and Rieseberg 2007; Rafajlović et al. 2016). As we continue to develop our understanding of the conditions under which genomic regions stop being exchanged between incipient species, studies involving multiple transects across hybrid zones with a standardized genome-wide dataset can provide insights into the dynamics of interspecific admixture and the final phases of the speciation process.

Barrier genes are directly involved in reproductive isolation. They underlie divergent ecological adaptation, differences in mate preference, and/or genetic incompatibilities, preventing the merging of parental populations (Abbott et al. 2013; Barton 2013; Ravinet et al. 2017). The resulting *barrier effect* is a reduction in the migration rate of genetic material between populations relative to the dispersal potential of the individuals that carry those genes (Ravinet et al. 2017). When gene frequency clines of multiple barrier genes become geographically coincident in a hybrid zone, the barrier effect is reinforced, a phenomenon known as *cline coupling* (Butlin and Smadja 2018).

Barrier genes contribute to genomic differentiation between species, because of the relatively strong genetic differentiation of the genes themselves as well as the linked genomic regions that surround them, allowing genetic differentiation to build up, and be retained in the face of gene flow. Their reduced introgression compared to the rest of the genome facilitates the identification of the genomic regions that harbor barrier genes (Gompert et al. 2012; Butlin and Smadja 2018). For the vast majority of natural systems, we lack the genomic resources to identify barrier genes as opposed to regions linked to them, and in these cases, we refer to markers with restricted introgression as *barrier markers* (following e.g. Ravinet et al. 2017). Identifying barrier markers is the first step in identifying the actual barrier genes that underlie reproductive isolation.

[1]Naturalis Biodiversity Center, Leiden, the Netherlands. [2]Institute of Biology Leiden, Leiden University, Leiden, the Netherlands. [3]Institute for Evolution and Ecology, Plant Evolutionary Ecology, Tübingen University, Tübingen, Germany. [4]Department of Ecology and Evolutionary Biology, UCLA, Los Angeles, CA, USA. [5]La Kretz Center for California Conservation Science, Institute of the Environment and Sustainability, UCLA, Los Angeles, CA, USA. [6]Institute of the Environment, UC Davis, Davis, CA, USA. [7]Department of Wildlife, Fish, and Conservation Biology, UC Davis, Davis, CA, USA. [8]Department of Marine Sciences, University of Gothenburg, Gothenburg, Sweden. [9]The Linnaeus Centre for Marine Evolutionary Biology, University of Gothenburg, Gothenburg, Sweden. [10]Natural Sciences Collegium, Eckerd College, 4200 54 Ave S, St Petersburg, FL 33711, USA.
Associate editor Lounès Chikhi
✉email: isolde.vanriemsdijk@naturalis.nl

Received: 6 June 2022 Revised: 27 March 2023 Accepted: 12 April 2023
Published online: 27 April 2023

When the same barrier markers are identified in geographically distant transects, the most reasonable explanation is that the barrier effect acting on (regions near) these markers is consistent along the hybrid zone (Teeter et al. 2009; Larson et al. 2013; Harrison and Larson 2014; Larson et al. 2014). For example, in field crickets (*Gryllus*), the same barrier markers act in two sections of the same hybrid zone, and have since been linked to intrinsic factors associated with prezygotic isolation (Larson et al. 2013; Larson et al. 2013; Larson et al. 2014).

Barrier markers may also differ between hybrid zone transects (see Harrison and Larson 2016 for an overiew), indicating that different genomic regions contribute to the barrier effect, depending on the local environment. Such heterogeneity among transects involving the same species pairs of *Catastomus* suckers in different rivers suggests that the underpinnings of reproductive isolation between these fish are influenced by interactions of multiple evolutionary and ecological factors (Mandeville et al. 2015), rather than a uniform genetic mechanism that evolved a single time and spread.

Finally, it is important to recognize that these two possibilities are not mutually exclusive. For example, the well-studied house mouse hybrid zone (*Mus*) shows distinct differences in gene flow among transects for some barrier markers (Teeter et al. 2009), while other barrier markers known to be linked to hybrid sterility consistently show narrow transitions across multiple *Mus* transects (Janoušek et al. 2012). Ultimately, independent analyses of replicate transects are required to disentangle species-wide barrier effects from ones that are geographically or ecologically more restricted.

We studied two transects at opposite ends of the ~900 km long hybrid zone between two European anuran species, the common toad (*Bufo bufo*) and the spined toad (*B. spinosus*). The zone runs diagonally across France, from the Atlantic coast in the north to the Mediterranean Sea in the south (Fig. 1; Arntzen et al. 2018; Arntzen et al. 2020; Dufresnes et al. 2020; Dufresnes et al. 2021). We characterized genome-wide patterns of interspecific gene flow based on ~1200 nuclear markers obtained by restriction-site associated DNA (RAD) sequencing, and used both genomic and geographical cline analyses to identify barrier markers that are either shared between transects or that are transect-specific (Gompert and Buerkle 2012; Stankowski et al. 2016). We mapped all loci against the recently sequenced genome of *B. bufo* (Streicher et al. 2021) to determine their proximity and identify physically nearby genes. We discuss the striking differences in gene flow across transects, the insights into the history of the hybrid zone that they provide, and more broadly, the insights that multiple transects can provide for the identification of potential barrier genes.

MATERIAL AND METHODS

3RAD sequencing

DNA from 387 individual toads (Arntzen et al. 2016, 2017, 2018, van Riemsdijk et al. 2019) were analyzed in this study. These included five reference localities each of pure *B. bufo* and *B. spinosus* (assumed to be unaffected by interspecific gene flow), 11 localities along a transect in northern France (northern transect), and 12 along a transect in southern France (southern transect; Supplementary Table S1; Fig. 1). We used the 3RAD protocol (Graham et al. 2015; Glenn et al. 2016; Hoffberg et al. 2016; Bayona-Vásquez et al. 2019) to produce reduced representation genomic libraries. Two restriction enzymes (Cla-*I* and Sbf-*I*) were used to cut 50 ng of genomic DNA, while a third enzyme (Msp-*I*) was added to cleave and eliminate phosphorylated adapter-adapter dimers. Internal barcodes were ligated to the resulting sticky ends and external Illumina iTru5 and iTru7 primers, differing by at least three base pairs, were added to the internal barcodes via an indexing PCR reaction (Glenn et al. 2016; Hoffberg et al. 2016; Bayona-Vásquez et al. 2019), followed by a bead-based clean-up. Libraries were combined to achieve equimolar concentrations in the final pool, size-selected for a 340–440 bp range using a Pippin Prep (Sage Science Inc. Beverly, MA, USA), quantified using intercalating dye on a VICTOR multilabel plate reader (PerkinElmer), and sequenced on two PE100 lanes on an Illumina HiSeq 4000. For three reference samples of each species (from sample locations 2.Bb and 19.Bs), libraries were prepared in triplicate to assess genotyping error rates.

Data clean-up and assembly

Cutadapt v.1.14 (Martin 2011) was used in three steps to remove 5′ and 3′ primers for each internal barcode combination, remove the Illumina sequencing adapters, and carry out a read quality control of 10 (default). We used ipyrad v.0.7.3 (Eaton 2014) for de novo assembly of our RAD loci and to call individual haplotypes and assign genotypes with the following settings: clustering threshold of 0.85, minimum read depth of six, maximum of eight heterozygous bases allowed per consensus sequence, and polymorphic sites in a locus shared across a maximum of 50% of the samples (for additional details, see supplemental material on Dryad). These settings were chosen based on stability of the number heterozygote calls (indicating absence of clusters of paralogs) after testing various settings and following recommendations from the ipyrad manual. In addition, with the unfiltered dataset, we obtained low SNP-calling error rates (see the Appendix, 1. Library triplicates). We generated two datasets, with either a 10% or a 50% maximum per-locus missingness rate across individuals (Supplementary Fig. S1).

Population structure

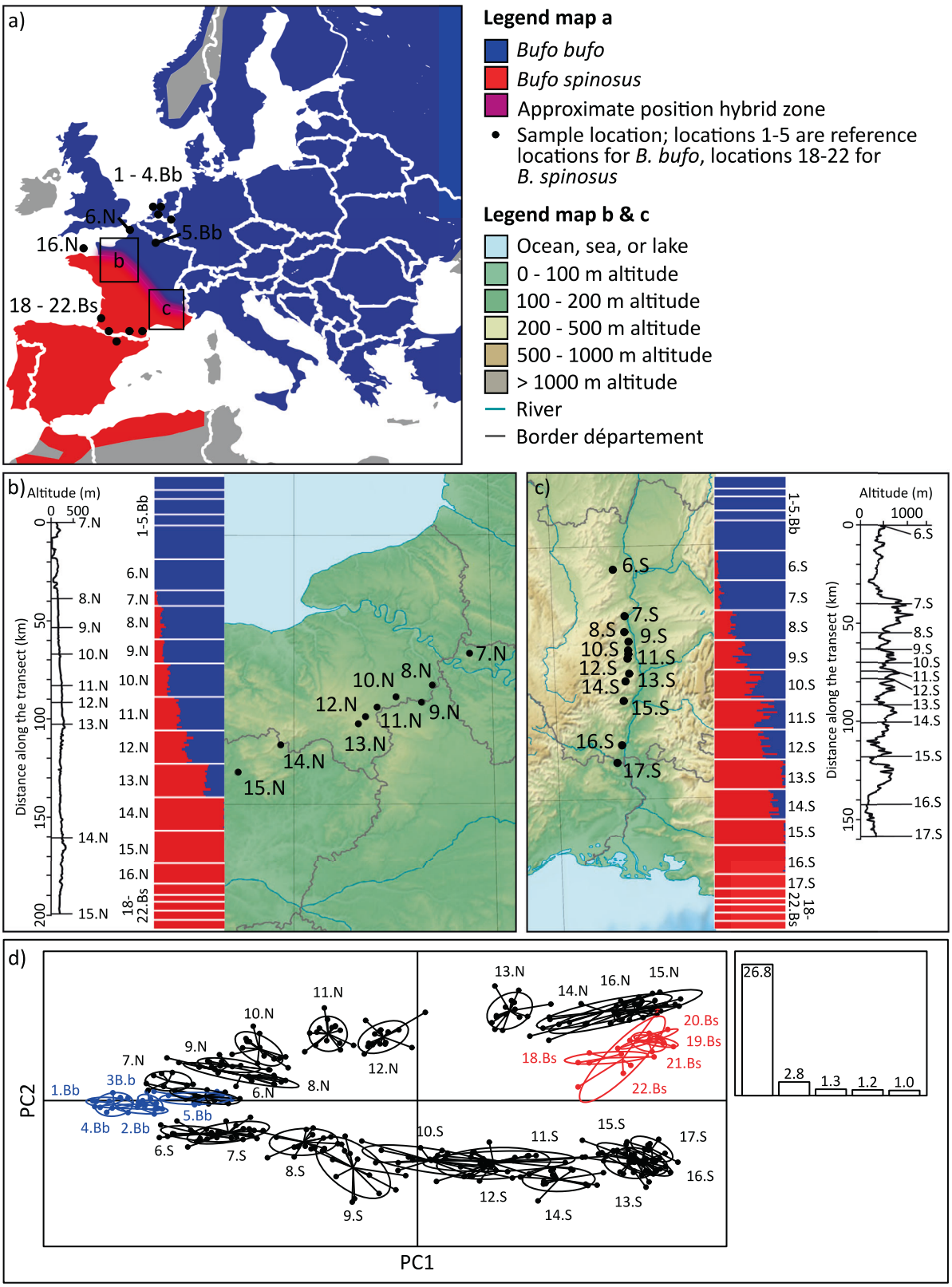
We randomly selected one single nucleotide polymorphism (SNP) per RAD locus from the 10 and 50% maximum per-locus missingness rate datasets (the actual missingness rates were 2.3 and 47.5%). We performed principal component analysis (PCA) on each dataset separately to visualize genomic variation across both transects. We used adegenet v.2.1.1 in R (Jombart 2008; Jombart and Ahmed 2011), and based the PCA on allele frequencies, replacing missing data with the mean allele frequency per population. We also quantified population structure with Structure v.2.3.4 (Pritchard et al. 2000) for the two datasets for each transect separately. We used StrAuto (Chhatre and Emerson 2017) to conduct ten independent runs for two genetic clusters ($K = 2$) representing the two parental species, with a burn in of 10 000 MCMC steps followed by 25,000 MCMC steps under the admixture model. Convergence was confirmed by visual examination of the consistency of the log likelihood and admixture proportion stability (Benestan et al. 2016) and summarized with CLUMPAK (Kopelman et al. 2015). We visualized these results with the R package POPHELPER (Francis 2017). The 10 and 50% maximum per-locus missingness rate datasets yielded nearly identical estimates of population structure (see Results), and we only used the larger (50%) missingness dataset for filtering and downstream analyses.

Diagnostic SNP selection and Hardy–Weinberg equilibrium

We identified species-diagnostic alleles (i.e. where one allele is fixed in one species and the other allele fixed in the other species) based on the reference samples (Fig. 1; Supplementary Table S1, custom R script available in supplemental material on Dryad), and randomly chose one species-diagnostic SNP from each of the 1189 species-diagnostic alleles if more than one SNP was available. We tested for signals of non-random mating within each locality by calculating deviations from Hardy–Weinberg equilibrium across species-diagnostic SNPs within populations with the R package 'genepop', based on the program GENEPOP v.1.0.5 (Rousset 2008), correcting for multiple comparisons using the sequential Bonferroni within-marker approach (P_c for $N = 1$ 189; Rice 1989; Narum 2006). Fourteen markers with a significant heterozygote deficit were excluded (Supplementary Table S2). This resulted in a dataset of 1 175 species-diagnostic SNP loci with 25.7% missing data that was used in downstream analyses.

Bayesian genomic cline analysis

To examine genome-wide variation of introgression among admixed individuals, we used the Bayesian genomic cline model as implemented in the software BGC (Gompert and Buerkle 2011, 2012; Gompert et al. 2012). This model is based on the probability that an individual with a certain hybrid index (HI) inherited a gene variant at a given locus from one species (φ; in this case *B. bufo*) or the other (1 – φ; *B. spinosus*). The cline parameter β measures the genomic cline rate based on ancestry for each locus, with a positive value indicating fewer heterozygotes at a locus than expected based on the HI of all loci, and a negative value indicating more



heterozygotes than expected (Gompert and Buerkle 2011; Parchman et al. 2013). Thus, positive β outliers can be regarded as barrier markers (Ravinet et al. 2017). In contrast, the cline parameter α describes the probability of *B. bufo* ancestry based on a single locus relative to the expectation based on the overall genetic composition (represented by the HI), with a positive value indicating an increase in the *B. bufo* ancestry probability, and a negative value a decrease. In summary, positive β outliers can identify potential barrier markers, while positive values of α can identify markers that have retained *B. bufo* ancestry in the face of higher background levels of *B. spinosus* genomic composition for any given population.

The input files for parental genotypes included only reference individuals and the input file for admixed genotypes included individuals

Fig. 1 Overview of the *Bufo* hybrid zone and genetic data presented in this study. Maps with (**a**) the distribution of *Bufo bufo* and *B. spinosus* in Europe and North Africa, with small labeled squares indicating the locations of insert map (**b**) for the northern transect (sample locations 6.N-16.N), and insert map (**c**) for the southern transect (sample locations 6.S-17.S). The reference populations for *B. bufo* are 1-5.Bb, and for *B. spinosus* are 18-22.Bs. The base map for panels **b** and **c** was downloaded from mapsland. Bar graphs in panels **b** and **c** are the result of Structure with K = 2 on the full 50% missingness dataset, taking a random SNP from each of the 4869 assembled RAD markers. Blue indicates *B. bufo* ancestry, red bars indicate *B. spinosus*. Panel **d** shows the principal component analysis of the same 50% dataset. The reference populations of *B. bufo* are colored blue and of *B. spinosus* red. Transect populations are black (".N" = northern transect, ".S" = southern transect). Ellipses represent the 95% inertia (based on the percentile) of the corresponding group. The bar graph on the right shows the eigenvalues for the first 5 principle components.

with an average admixture proportion (Structure Q score) between 0.05 and 0.95 (Supplementary Fig. S2). A single MCMC chain was run for 75,000 steps and samples were taken from the posterior distribution every 5th step, following a burn-in of 25,000 steps. Convergence was assessed (Supplementary Fig. S3) and we tested for outlier loci using 'estpost' to summarize parameter posterior distributions (Gompert and Buerkle 2011). Outlier loci were conservatively identified based on 99.9% confidence intervals. As a test for significance of outlier distributions, we performed: (1) a Chi-square test comparing the number of α outliers between the northern and southern transect to test if there is a significant difference in the number of outliers between the transects, and (2) a 2×2 contingency Chi-squared was conducted to test if the overlap between the transects of α or β outliers could be a coincidence, or if such overlap is unlikely to have occurred under a model of random resampling.

Geographic cline analysis

Classic geographic equilibrium cline models were fitted using the R package 'HZAR' (Derryberry et al. 2014) for all diagnostic SNPs. For the northern transect, sample localities 6.N and 16.N were removed because they disrupted the geographic linearity of the transect. Linear transects were drawn starting in *B. bufo* populations in the east and north, and ending in *B. spinosus* in the west and south, for the North and South transects respectively. The direction of the transect crosses the hybrid zone perpendicular, approximately west to east in the North transect, and north to south in the South transect. To determine the distance between sample localities along these transects, a custom R script was used (see supplementary material on Dryad), and the directions of the transect axes followed earlier analyses (Arntzen et al. 2016, 2017). Outliers identified in the Bayesian genomic cline analysis should also be noticeable in the geographical cline analysis: positive β outliers are expected to show relatively narrow geographic clines, with their center aligned with that of the hybrid zone, while negative and positive α outliers are expected to have shifted geographical cline centers and/or asymmetrical tails. The shapes and positions of many individual clines can be summarized in an expected cline based on the HI (Polechová and Barton 2011; Fitzpatrick 2012). We used the HI for all "non-outlier loci" in the BGC analysis as a neutral cline expectation for comparison with outlier loci. Thirty maximum likelihood searches were performed with random starting parameters, followed by a trace analysis of 60,000 generations on all models with a delta Akaike information criterion corrected for small-sample-size (dAICc < 10). Fifteen model variants were based on all possible combinations of trait intervals (allele frequency at the ends of the transects; three types) and tail shape (five types; Supplementary Fig. S4). Convergence was visually assessed in trace plots (for additional details, see supplementary material on Dryad).

Effective selection

Average effective selection on a locus (s*) is the selection pressure at the hybrid zone center due to both direct selection and indirect selection on linked loci. An estimate of s* can be obtained using $s^* = \left(\frac{2\sigma}{w}\right)^2$, where σ is the dispersal weighted for juvenile and adult life stages, and w is the width of the HI cline (Barton and Gale 1993). We calculated s* using scripts from van Riemsdijk et al. (2019) for each transect separately. We grouped those markers that were not indicated as outliers in BGC (832 for the northern transect and 652 loci for the southern transect; Table 1). We also excluded population 16.N, because it is isolated on the island of Jersey and could be inbred, which would violate model assumptions. We repeated the analysis for presumptive barrier markers identified as heterozygote deficiency outliers in BGC (β > 0; 56 loci for the northern and 121 loci for the southern transect). To calculate σ we used a recombination rate of 0.4997, calculated following equation (6) from Macholán et al. (2007) and based on the number of chiasmata per bivalent for *B. bufo* (1.95; Wickbom 1945) and the number of diploid chromosomes for *B. bufo* ($N = 22$), a generation time (sexual maturity) of 2.5 years for *Bufo* at the latitude of the hybrid zone (mean of 3 years in females and 2 years in males; Hemelaar 1988) and secondary contact established 8000 years ago (Arntzen et al. 2016). Admixture linkage disequilibrium (D') at the center of the hybrid zone was based on the variance in HI. The width of the hybrid zone (w) was derived from a general sigmoid cline model following HZAR (Derryberry et al. 2014), fitted to the HI (Barton and Gale 1993). The mean and 95% confidence interval (CI) were based on 1000 bootstrap replicates of the original genotype dataset sampled with replacement and maintaining the original sample sizes within sites.

Table 1. Bayesian genomic cline (BGC) results comparing significant outliers for the northern transect (N) and the southern transect (S), and the markers which were outliers in both transects (overlap), where significance of outliers is based on the 99.9% confidence interval (CI).

Outlier	Biological interpretation	N	S	Overlap
β < 0	Heterozygote excess	50	105	11[d]
β > 0	Heterozygote deficiency	56	123	26[c]
α < 0	Directional introgression from *B. bufo* into *B. spinosus*	151[a]	185[b]	92[c]
α > 0	Directional introgression from *B. spinosus* into *B. bufo*	110[a]	174[b]	49[c]

The total number of markers analyzed was 1175.
[a]Significant Chi-squared with 6.4406, $df = 1$, $P = 0.0112$.
[b]Not significant Chi-squared with 0.3371, $df = 1$, $P = 0.5615$.
[c]Significant 2 × 2 contingency Chi-squared with 285.05, 91.629, 81.288, $df = 1$, $P < 2.2e^{-16}$.
[d]Significant 2 × 2 contingency Chi-squared with 10.331, $df = 1$, $P = 0.001308$.

RAD marker genome position

The recently published genome of *B. bufo* is assembled into 11 chromosome-level scaffolds spanning 5045 Mbp, with a scaffold N50 of 636 Mbp (Streicher et al. 2021). Within this genome, 30,286 genes were identified. We aligned 1184 (diagnostic) RAD markers allowing gaps between the flanking regions of fragments to the *B. bufo* genome with bwa v.0.7.17 (Li and Durbin 2009) using the BWA-MEM algorithm and default settings. A chromosome map identifying RAD marker putative positions was constructed with the R packages GenomicRanges (Lawrence et al. 2013) and karyoploteR (Gel and Serra 2017). We used a cut-off of 750 kbp distance between SNPs and gene boundaries to link SNPs to gene IDs, which appears to be optimal for finding SNPs linked to causative genes (Brodie et al. 2016). This resulted in "closest gene" identifications for 1141 markers. We used a go-term enrichment analysis to identify "selection on biological processes" of the genes closest to the 26 markers identified as β > 0 outliers (selection against heterozygosity) in both transects, (PANTHER v.14; Mi et al. 2018), using *Xenopus tropicalis* and *Homo sapience* as references, with Fisher's exact test and Bonferroni correction for multiple testing. We further explored a 31 Mbp region on chromosome 3 to which seven out of 26 RAD markers identified as heterozygote deficient in the BGC analysis in both transects mapped. A STRING (Szklarczyk et al. 2021) analysis based on human genome annotation was used to identify the 33 genes in this region and infer function based on the human genome. Default settings were used, with medium confidence for a relation between genes to be reported (0.4). Higher confidence resulted in no links reported.

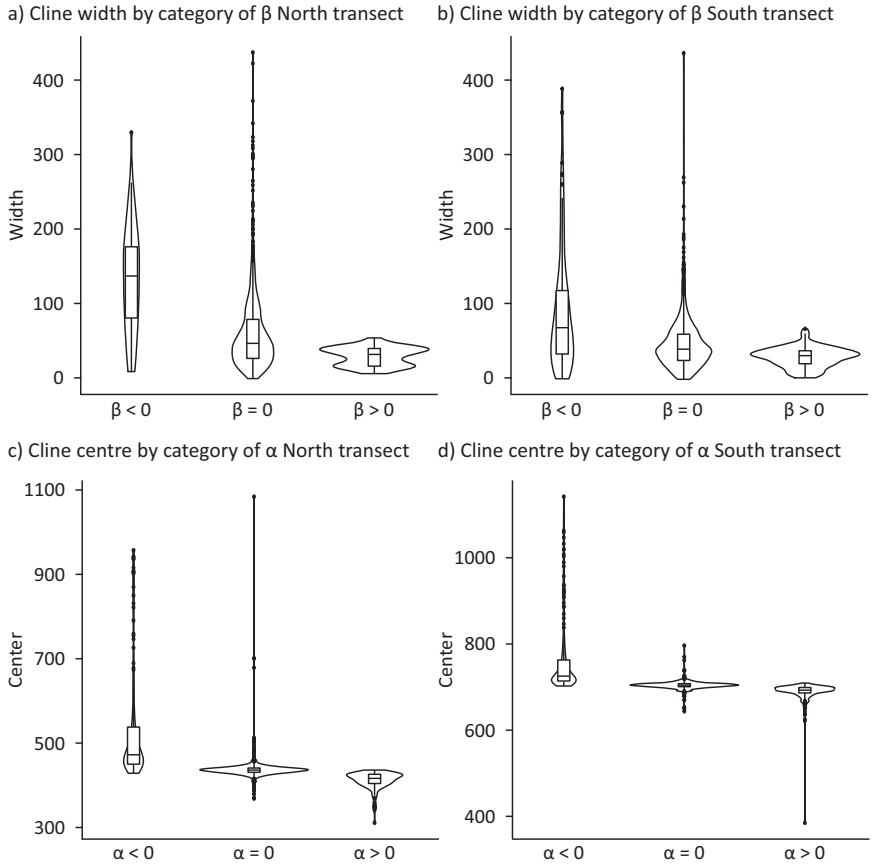


Fig. 2 Markers identified as β or α outlier by Bayesian genomic cline (BGC) analysis, also stand out in their geographic cline when comparing the parameters 'center' and 'width' (HZAR). Heterozygote deficiency outliers ($\beta > 0$) have steep geographic clines, both in the (**a**) northern transect and (**b**) southern transect. For markers showing directional introgression (α outliers in BGC), the geographical cline has a shifted center for both the (**c**) northern transect and (**d**) southern transect.

RESULTS

Data quality and population structure

The average read count per sample was ~1.5 million paired-end reads (maximum 4.2 million reads, Supplementary Table S3). Five samples had low raw read quantities (near or below 0.5 million reads) and were excluded. Low error rates in genotype calls for samples sequenced in triplicate (0.5%, Appendix, 1. Library triplicates) and near-identical population structure based on different datasets (see below) give us confidence that our de novo assembly parameters strike a reasonable balance between read depth and missingness. The 10% maximum per-locus missingness rate across individuals included loci for at least 349 individuals genotyped for 986 loci (10,535 SNPs), while the 50% data set included loci for at least 194 individuals genotyped for 4869 loci (39,750 SNPs). Because the PCA plots (Supplementary Fig. S5) and Structure analyses for both datasets were nearly identical we describe only the results of the larger 50% missingness dataset, and continued with this dataset in downstream analyses. The first axis (PC1, 26.8%) of the PCA reflects the genetic difference between pure *B. bufo* in the north (lower PC1 values) and pure *B. spinosus* in the south (higher PC1 values), with hybrids at an intermediate position (Fig. 1d, Supplementary Fig. S5). The second axis (PC2, 2.8%) separates the two transects at the *B. spinosus* end. Structure plots for $K = 2$ reflect differentiation between the two species, with a smooth transition between the two parental species via genetically admixed genotypes. Additionally, to verify the choice of species-diagnostic SNPs we ran additional analysis (Supplementary Appendix 2. Reference populations), further strengthening our confidence in the quality of the data we selected.

Bayesian genomic cline outlier detection

The BGC analysis, based on 1175 species-diagnostic SNPs, shows little bias in the distribution of observed hybrid indices for admixed individuals across both transects (Supplementary Fig. S6). The number of markers identified as heterozygote deficient or heterozygote excess outliers in the northern transect (deficient: $\beta > 0$, $n = 56$; excess: $\beta < 0$, $n = 42$) is less than half that observed in the southern transect ($\beta > 0$, $n = 123$; $\beta < 0$, $n = 105$). Considerably more markers were identified as outlier in either transect than would be expected by chance (last column, Table 1). Both transects shared a set of 26 heterozygote deficient markers indicating reduced introgression, a result that is extremely unlikely purely by chance (2×2 contingency Chi-squared, $df = 1$, $P < 2.2e^{-16}$; Table 1). The southern transect has a nearly equal number of markers with an increased or decreased probability of *B. bufo* ancestry ($\alpha > 0$, $n = 174$; $\alpha < 0$, $n = 185$; χ^2 test $P = 0.5615$). In contrast, the northern transect had fewer α outliers overall, and significantly more markers with a reduced than an increased probability of *B. bufo* ancestry relative to the overall HI (reduced $\alpha < 0$, $n = 151$; increased $\alpha > 0$, $n = 110$; χ^2 test $P = 0.0112$; Table 1). In other words, more genes are flowing from *B. bufo* into *B. spinosus* in the northern than the southern transect.

Geographic cline outlier detection

The same 1175 species-diagnostic SNPs were subjected to geographical cline analysis in HZAR. We verified that BGC outlier markers also stood out in the shape and/or position of their geographic cline by comparing them to the geographic cline

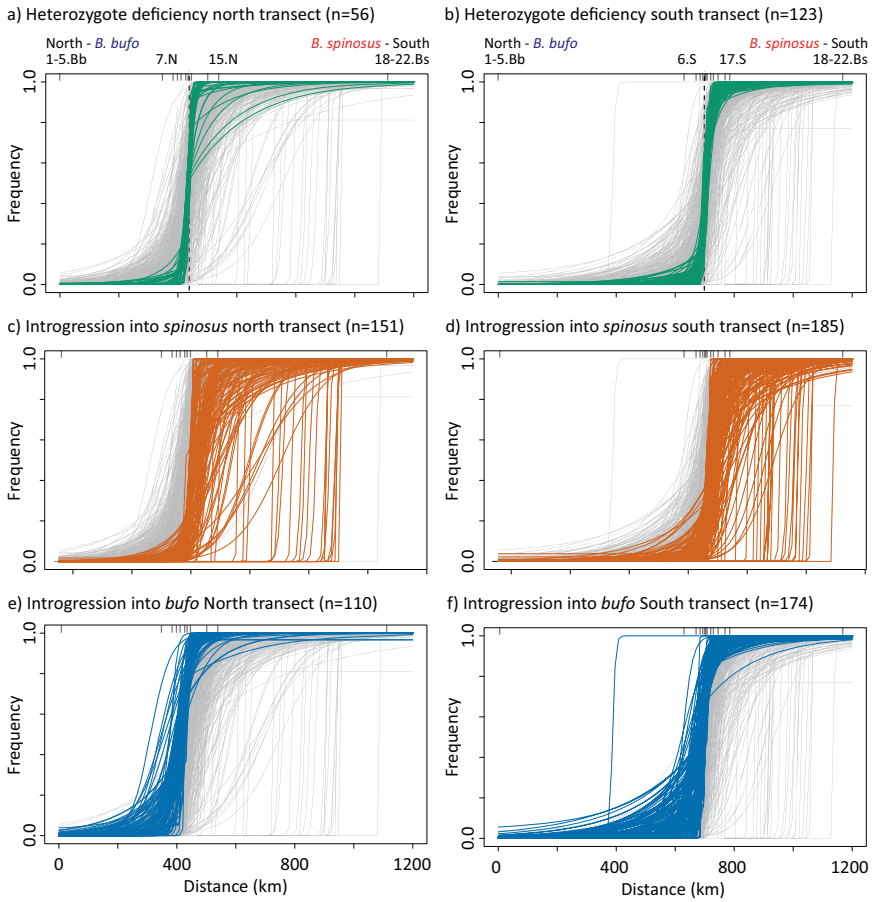


Fig. 3 Geographic cline graphs coloured by Bayesian Genomic Cline (BGC) analysis outlier identification. Geographic clines for barrier markers ($\beta > 0$, green) for the (**a**) northern transect and (**b**) southern transect; markers showing introgression into *B. spinosus* ($\alpha < 0$, orange) for the (**c**) northern transect and (**d**) southern transect; and markers showing introgression into *B. bufo* ($\alpha > 0$, blue) for the (**e**) northern transect, and (**f**) southern transect. In all panels, frequency of the *B. spinosus* allele on the y-axis and distance along the transect on the x-axis. The gray clines are all other markers (including all outliers of other types). Inward ticks on the top of each graph and notation near inward ticks on the top of the graph (panels a and b only) refers to locations in Fig. 1.

parameters 'center' and 'width' as determined with HZAR (Fig. 2, full analysis in online data repository). Heterozygote deficiency outliers according to BGC ($\beta > 0$) have steep geographic clines, and this was the case in both transects (green in Fig. 3, Supplementary Figs. S7 and S8). These are considered barrier markers because they show reduced introgression compared to background introgression (gray lines in Fig. 3). For α outliers in BGC ($\alpha < 0$ [dark orange] and $\alpha > 0$ [blue]), the geographical cline has a shifted center and/or shape (e.g., width), when compared to the genomic average, represented by the geographical cline of the HI based on all neutral markers (Fig. 3, Supplementary Fig. S9, Supplementary Table S4).

Effective selection

We observed lower estimates of effective selection against hybrids (s*) in the northern transect than in the southern transect. For the northern transect, s* based on markers that were not identified as outliers in the BGC analysis was 0.0022 (95% CI 0.0012–0.0034), while s* in the southern transect was almost an order of magnitude greater, at 0.0195 (95% CI 0.0143–0.0252; Supplementary Table S5, Supplementary Fig. S10). The s* based on barrier markers for the northern transect (0.0101; 95% CI 0.0054–0.0152) was also lower than the southern transect (0.0344; 95% CI 0.0220–0.0470), although the difference between transects was smaller. Notably, the confidence intervals for estimates of s* for neutral and for barrier markers do not overlap between the two transects.

RAD marker genome position

A total of 1140 RAD fragments were aligned to the *B. bufo* genome (Supplementary Table S6). Markers identified as genomic outliers by BGA were mapped on the chromosomes separately for each transect, showing clusters of especially beta outliers, and, sometimes, alpha outliers (Fig. 4). The go-term enrichment analysis of the genes closest to the 26 markers identified as heterozygote deficient ($\beta > 0$ in BGC, i.e. potential barrier markers) in both transects, gave no significant overrepresentation of any biological function (Supplementary Tables S7 and S8). Seven of these 26 markers were mapped in a region of ~ 31 Mbp on chromosome 3 (see Supplementary Fig. S11) and therefore we explored this region further. The SNPs in this region of the chromosome behave as a linkage block; if individuals are homozygote for one SNP, they are also homozygote for the neighboring SNPs, if they are heterozygote for one SNP, they are also heterozygote for the neighboring SNPs, etc., indicating there is a relatively low recombination rate in this region (Supplementary Fig. S12). In this region, 33 well-annotated genes are located (Supplementary Table S9). When we analyzed these 33 genes with STRING, we found these genes are reported (in domestic pigs, *Sus scrofa domesticus*) to be enriched for limb and ear growth (Supplementary Table S10; Ren et al. 2011; Zhang et al. 2014).

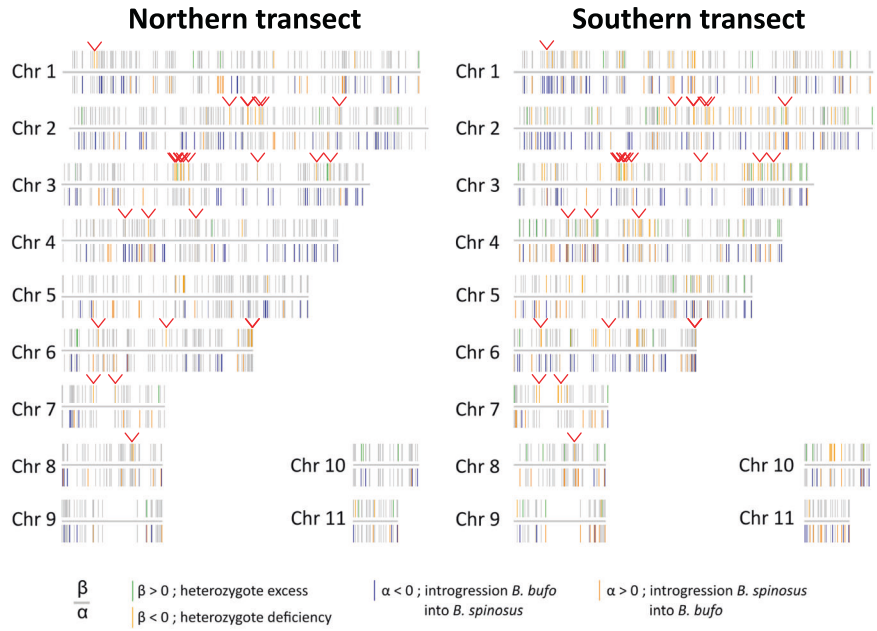


Fig. 4 Alignment of RAD markers to the *B. bufo* genome for each transect. Gray lines represent markers which were not identified as outlier by BGC analysis. Above the line are heterozygosity (β) outliers, below the line are directional gene flow outliers (α). Outlier identity is explained in the legend below the figure. Red arrows point to RAD markers which were heterozygote deficiency outliers in both transects.

DISCUSSION

A shared barrier to gene flow characterizes the Bufo hybrid zone

Although 900 km apart, our two transects share 26 barrier markers that all demonstrate a heterozygous deficit when compared to genome-wide expectations. It is highly unlikely that this pattern would be shared across these loci due to chance alone (2×2 contingency Chi-squared test, df = 1, $P < 2.2e^{-16}$; Table 1). Rather, we interpret this finding as indicating that a set of general genomic incompatibilities is consistently present between these two *Bufo* species and that these markers indicate a mechanism maintaining both species. Individual loci with restricted introgression ($\beta < 0$ in BGC) could imply local adaptation to e.g. hybridization, and therefore could be a response to existing barriers to gene flow, but they could also be evidence of other biological phenomena, such as heterozygote disadvantage. The corresponding general barrier effect is strong enough to have prevented collapse of the species boundary, despite initiation of secondary contact millennia ago (Arntzen et al. 2016; Arntzen 2019). Whilst there seems no enrichment of any biological function for these 26 barrier markers (Supplementary Tables S7 and S8), seven of them map to a 31 Mbp region of *B. bufo* chromosome 3 that contains 33 protein coding genes. Two RAD loci in this same region were 1) only found to be a barrier marker in the northern transect, and 2) found to be neutral or to have heterozygous excess. This suggests not all markers in this chromosomal region share an identical evolutionary history, and therefore, the following interpretation of the results is cautious.

The protein coding genes in the 31 Mbp region are enriched for limb and ear growth in domestic pigs (Ren et al. 2011; Zhang et al. 2014). It is possible this genomic region may control developmental pathways that are responsible for morphological or physiological differences between these two toad species. However, homology of this function in amphibians has not yet been confirmed (also not in the anuran model *Xenopus*) and it would require gene editing experiments to verify function. Although our reduced genome sequencing cannot pinpoint the specific causal variant(s), our finding suggests that one or more genes in this region are associated with reproductive isolation between these two *Bufo* species. Further exploration of this genomic region may help elucidate what causes reproductive isolation in this system.

An overlapping set of barrier markers identified in multiple transects across a sunflower (*Helianthus*) hybrid zone similarly indicated a common set of barrier genes; subsequently, genes in these regions were linked to pollen sterility and chromosomal rearrangements, and were thought to be responsible for the maintenance of lineage isolation (Feulner and De-Kayne 2017; Rieseberg et al. 1999; Buerkle and Rieseberg 2001). The *Helianthus* barrier genes were also found to restrict gene flow under laboratory conditions, excluding the possibility that environmental factors shared across all transects were playing a significant role (Buerkle and Rieseberg 2001). Future experimental studies and genomic analysis could test the hypothesis of such a functional barrier to gene flow across the *Bufo* hybrid zone. An alternative hypothesis, in which structural variation between the genomes of *B. bufo* and *B. spinosus* is responsible for reduced recombination between genomic regions, could be tested when a *B. spinosus* genome becomes available.

Spatial variation in barrier effect along the Bufo hybrid zone

The number of barrier markers identified in the genomic cline analysis is approximately twice as large in the southern transect ($n = 123$) than in the northern transect ($n = 56$). These barrier markers also act as expected under reduced introgression in the geographical cline analysis; their geographical clines are narrow and with centers confined to the midpoint of the hybrid zone. The estimated strength of selection against hybrids is significantly higher in the south ($s^* = 0.0195$) than in the north ($s^* = 0.0022$).

The simplest explanation for these results is a difference in timing of secondary contact establishment along the *Bufo* hybrid zone. During postglacial expansion, secondary contact between *B. bufo* and *B. spinosus* would have first been established in the south-east of France, where the geographic distance between refugia is smaller, and progressed over time towards the north-west, where the geographic distance between refugia is larger. Timing of this zipper-like establishment of secondary contact is well characterized using mtDNA data (Garcia-Porta et al. 2012; Recuero et al. 2012; Arntzen et al. 2017), and would have progressed relatively slow, as the effective dispersal in *Bufo* toads

is estimated at about max. 4.5 km per generation (Supplementary Table S.5). As a consequence of the longer secondary contact in the south compared to the north, there has been more opportunity for additional barrier genes to be recruited and converge geographically towards the hybrid zone -a process also known as "cline coupling" (Barton 1983; Barton and Gale 1993; Bierne et al. 2011; Harrison and Larson 2016; Vines et al. 2016; Butlin and Smadja 2018; Dagilis et al. 2019).

Differences in introgression along the Bufo hybrid zone

Another distinct difference between the two transects is that the direction of gene flow is symmetrical in the southern transect, but involves a greater number of *B. bufo* alleles flowing into *B. spinosus* than in the opposite direction in the northern transect (Table 1). Markers showing significant introgression are randomly distributed across the genome, and are consistent with the interpretation of hybrid zone movement, with random asymmetric introgression from the receding into the expanding species, reflecting a 'genomic footprint' of hybrid zone movement (Barton and Hewitt 1985; Buggs 2007; Currat et al. 2008; Wielstra 2019). Despite a longer history of introgression, the hybrid zone appears to be stable in the south, whereas the hybrid zone appears more mobile in the relatively recently established northern transect.

Reasons why the hybrid zone is more mobile in northern France remain uncertain. Barrier genes may become coupled in a steep gradient of locally adapted genes (Bierne et al. 2011). In such a situation the hybrid zone can become 'trapped' at an ecotone (Bierne et al. 2011). A suite of environmental variables was previously found to be associated with the position of the *Bufo* hybrid zone based on niche modeling and genetic data: *B. spinosus* prefers higher altitude, higher temperatures, and sandier soils than *B. bufo* (Arntzen et al. 2020). While the geography around the northern transect is relatively homogeneous, this is not the case for the southern transect, particularly with respect to elevational relief (Fig. 1). We propose that the hybrid zone has stabilized at an environmental gradient in southern, but not (yet) in northern France, perhaps reflecting the younger age of the northern transect. Previous studies provided evidence for both hybrid zone stability and movement in other systems (e.g. Visser et al. 2017; Sequeira et al. 2022), and further research could focus on time series to record further movement of the northern section of the hybrid zone and attempt to link genes to environmental or ecological processes.

CONCLUSIONS

Hybrid zones provide the unique opportunity to quantify the completeness (or lack thereof) of the speciation process. It is now clear that some alleles fare poorly in admixed populations, whereas others flow freely between species across a broad range of taxa. What is less well documented, is that these genomic incompatibilities may vary spatially across the same hybrid zone. Here, we show striking differences in interspecific gene flow between two replicate hybrid zone transects, presumably reflecting variation in secondary contact establishment and landscape features along the hybrid zone. We document a set of barrier markers shared between transects, reflecting genomic incompatibilities that characterize the two *Bufo* species as a whole and maintain the discreteness of the hybrid zone throughout its length. This tension zone apparently consists of species-level genomic regions maintaining the integrity of both species, and variable local landscapes interacting with additional barrier genes to define the placement and stability of the hybrid zone in different environments. Such complex patterns can only be documented by applying genomic hybrid zone analyses at multiple transect locations. The *Bufo* hybrid zone thus provides a system allowing us to separate general barriers to gene flow from those derived from localized environmental variation interacting, potentially, with specific genes.

DATA AVAILABILITY

Supplementary figures, tables, and text can be downloaded from the online article version. Raw data used in this study are available on GenBank under NCBI BioProject PRJNA954537. Data assembly, validation, and analysis code and input files can be found on Dryad: https://doi.org/10.5061/dryad.qv9s4mwkg.

REFERENCES

Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N, Zinner D (2013) Hybridization and speciation. J Evolut Biol 26:229–246. https://doi.org/10.1111/j.1420-9101.2012.02599.x

Arntzen JW (2019) An amphibian species pushed out of Britain by a moving hybrid zone. Mol Ecol 28:5145–5154

Arntzen JW, Canestrelli D, Martínez-Solano I (2020) Environmental correlates of the European common toad hybrid zone. Contrib Zool 89:1–12. https://doi.org/10.1163/18759866-bja10001

Arntzen JW, de Vries W, Canestrelli D, Martínez-Solano I (2017) Hybrid zone formation and contrasting outcomes of secondary contact over transects in common toads. Mol Ecol 26:5663–5675. https://doi.org/10.1111/mec.14273

Arntzen JW, de Vries W, Canestrelli D, Martínez-Solano I (2020) Genetic and morphological differentiation of common toads in the alps and the apennines. In Pontarotti P (ed), Evolutionary biology—a transdisciplinary approach. Springer International Publishing, pp. 1–13. https://doi.org/10.1007/978-3-030-57246-4_1

Arntzen JW, McAtear J, Butôt R, Martínez-Solano I (2018) A common toad hybrid zone that runs from the Atlantic to the Mediterranean. Amphib Reptilia 39:41–50. https://doi.org/10.1163/15685381-00003145

Arntzen JW, Trujillo T, Butot R, Vrieling K, Schaap OD, Gutiérrez-Rodriquez J, Martinez-Solano I (2016) Concordant morphological and molecular clines in a contact zone of the common and spined toad (*Bufo bufo* and *B. spinosus*) in the northwest of France. Front Zool 13:1–12. https://doi.org/10.1186/s12983-016-0184-7

Baack EJ, Rieseberg LH (2007) A genomic view of introgression and hybrid speciation. Curr Opin Genet Dev 17:513–8. https://doi.org/10.1016/j.gde.2007.09.001

Barton NH (1983) Multilocus clines. Evolution 37:454–471. https://doi.org/10.2307/2408260

Barton NH (2013) Does hybridization influence speciation? J Evolut Biol 26:267–269. https://doi.org/10.1111/jeb.12015

Barton NH, Gale KS (1993) Genetic analysis of hybrid zones. In: Harrison RG (Ed.) Hybrid zones and the evolutionary process. Oxford University Press, New York, p 13–45

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. Annu Rev Ecol Syst 16:113–148

Bayona-Vásquez NJ, Glenn TC, Kieran TJ, Pierson TW, Hoffberg SL, Scott PA, Faircloth BC (2019) Adapterama III: Quadruple-indexed, double/triple-enzyme RADseq libraries (2RAD/3RAD). PeerJ 7:e7724. https://doi.org/10.7717/peerj.7724

Benestan LM, Ferchaud A-L, Hohenlohe PA, Garner BA, Naylor GJP, Baums IB, Luikart G (2016) Conservation genomics of natural and managed populations: building a conceptual and practical framework. Mol Ecol 25:2967–77. https://doi.org/10.1111/mec.13647

Bierne N, Welch J, Loire E, Bonhomme F, David P (2011) The coupling hypothesis: why genome scans may fail to map local adaptation genes. Mol Ecol 20:2044–2072. https://doi.org/10.1111/j.1365-294X.2011.05080.x

Brodie A, Azaria JR, Ofran Y (2016) How far from the SNP may the causative genes be? Nucleic Acids Res 44:6046–6054. https://doi.org/10.1093/nar/gkw500

Buerkle CA, Rieseberg LH (2001) Low intraspecific variation for genomic isolation between hybridizing sunflower species. Evolution 55:684–691. https://doi.org/10.1554/0014-3820(2001)055[0684:livfgi]2.0.co;2

Buggs RJA (2007) Empirical study of hybrid zone movement. Heredity 99:301–312. https://doi.org/10.1038/sj.hdy.6800997

Butlin RK, Smadja CM (2018) Coupling, reinforcement, and speciation. Am Nat 191:155–172. https://doi.org/10.1086/695136

Chhatre VE, Emerson KJ (2017) StrAuto: automation and parallelization of STRUCTURE analysis. BMC Bioinforma 18:1–5. https://doi.org/10.1186/s12859-017-1593-0

Currat M, Ruedi M, Petit RJ, Excoffier L (2008) The hidden side of invasions: massive introgression by local genes. Evolution 62:1908–1920. https://doi.org/10.1111/j.1558-5646.2008.00413.x

Dagilis AJ, Kirkpatrick M, Bolnick DI (2019) The evolution of hybrid fitness during speciation. PLoS Genet 15:e1008125

Derryberry EP, Derryberry GE, Maley JM, Brumfield RT (2014) HZAR: hybrid zone analysis using an R software package. Mol Ecol Resour 14:652–663. https://doi.org/10.1111/1755-0998.12209

Dufresnes C, Litvinchuk SN, Rozenblut-Kościsty B, Rodrigues N, Perrin N, Crochet P, Jeffries DL (2020) Hybridization and introgression between toads with different sex chromosome systems. Evol Lett 4:444–456. https://doi.org/10.1002/evl3.191

Dufresnes C, Brelsford A, Jeffries DL, Mazepa G, Suchan T, Canestrelli D, Nicieza A, Fumagalli L, Dubey S, Martínez-Solano I, Litvinchuk SN, Vences M, Perrin N, Crochet P-A (2021) Mass of genes rather than master genes underlie the genomic architecture of amphibian speciation. Proc Natl Acad Sci 118(36):e2103963118. https://doi.org/10.1073/pnas.2103963118

Eaton DAR (2014) PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics 30:1844–1849. https://doi.org/10.1093/bioinformatics/btu121

Feulner PGD, De-Kayne R (2017) Genome evolution, structural rearrangements and speciation. J Evolut Biol 30(8):1488–1490. https://doi.org/10.1111/jeb.13101

Fitzpatrick BM (2012) Estimating ancestry and heterozygosity of hybrids using molecular markers. BMC Evolut Biol 12:131. https://doi.org/10.1186/1471-2148-12-131

Francis RM (2017) POPHELPER: an R package and web app to analyse and visualize population structure. Mol Ecol Resour 17:27–32. https://doi.org/10.1111/1755-0998.12509

Garcia-Porta J, Litvinchuk SN, Crochet PA, Romano A, Geniez PH, Lo-Valvo M, Carranza S (2012) Molecular phylogenetics and historical biogeography of the west-palearctic common toads (*Bufo bufo* species complex). Mol Phylogenetics Evol 63:113–130. https://doi.org/10.1016/j.ympev.2011.12.019

Gel B, Serra E (2017) KaryoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. Bioinformatics 33:3088–3090. https://doi.org/10.1093/bioinformatics/btx346

Glenn TC, Nilsen RA, Kieran TJ, Finger Jr JW, Pierson TW, Bentley KE, … Faircloth BC (2016) Adapterama I: universal stubs and primers for thousands of dual-indexed Illumina libraries (iTru & iNext). BioRxiv https://www.biorxiv.org/content/10.1101/049114v1.

Gompert Z, Buerkle CA (2011) Bayesian estimation of genomic clines. Mol Ecol 20:2111–2127. https://doi.org/10.1111/j.1365-294X.2011.05074.x

Gompert Z, Buerkle CA (2012) bgc: software for Bayesian estimation of genomic clines. Mol Ecol Resour 12:1168–1176. https://doi.org/10.1111/1755-0998.12009.x

Gompert Z, Parchman TL, Buerkle CA (2012) Genomics of isolation in hybrids. Philos Trans R Soc B Biol Sci 367:439–450. https://doi.org/10.1098/rstb.2011.0196

Graham CF, Glenn TC, McArthur AG, Boreham DR, Kieran T, Lance S, Somers CM (2015) Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). Mol Ecol Resour 15:1304–15. https://doi.org/10.1111/1755-0998.12404

Harrison RG, Larson EL (2014) Hybridization, introgression, and the nature of species boundaries. J Heredity 105:795–809. https://doi.org/10.1093/jhered/esu033

Harrison RG, Larson EL (2016) Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. Mol Ecol 25:2454–2466. https://doi.org/10.1111/mec.13582

Hemelaar A (1988) Age, growth and other population characteristics of *Bufo bufo* from different latitudes and altitudes. J Herpetol 22:369–388

Hewitt GM (1988) Hybrid zones - natural laboratories for evolutionary studies. Trends Ecol Evol 3:158–167. https://doi.org/10.1016/0169-5347(88)90033-X

Hoffberg S, Kieran T, Catchen J, Devault A, Faircloth BC, Mauricio R, Glenn TC (2016) RADcap: sequence capture of dual-digest RADseq libraries with identifiable duplicates and reduced missing data. Mol Ecol Resour 16:1264–1278. https://doi.org/10.1111/jnc.13494

Janoušek V, Wang L, Luzynski K, Dufková P, Vyskočilová MM, Nachman MW, Tucker PK (2012) Genome-wide architecture of reproductive isolation in a naturally occurring hybrid zone between *Mus musculus musculus* and *M. m. domesticus*. Mol Ecol 21:3032–3047. https://doi.org/10.1111/j.1365-294X.2012.05583.x

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27:3070–3071. https://doi.org/10.1093/bioinformatics/btr521

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Mol Ecol Resour 15:1179–1191. https://doi.org/10.1111/1755-0998.12387

Larson EL, Andrés JA, Bogdanowicz SM, Harrison RG (2013) Differential introgression in a mosaic hybrid zone reveals candidate barrier genes. Evolution 67:3653–3661. https://doi.org/10.1111/evo.12205

Larson EL, Guilherme Becker C, Bondra ER, Harrison RG (2013) Structure of a mosaic hybrid zone between the field crickets *Gryllus firmus* and *G. pennsylvanicus*. Ecol Evol 3:985–1002. https://doi.org/10.1002/ece3.514

Larson EL, White TA, Ross CL, Harrison RG (2014) Gene flow and the maintenance of species boundaries. Mol Ecol 23:1668–1678. https://doi.org/10.1111/mec.12601

Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Carey VJ (2013) Software for Computing and Annotating Genomic Ranges. PLoS Comput Biol 9:1–10. https://doi.org/10.1371/journal.pcbi.1003118

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Macholán M, Munclinger P, Šugerková M, Dufková P, Bímová B, Božíková E, Piálek J (2007) Genetic analysis of autosomal and X-linked markers across a mouse hybrid zone. Evolution 61:746–771. https://doi.org/10.1111/j.1558-5646.2007.00065.x

Mandeville EG, Parchman TL, McDonald DB, Buerkle CA (2015) Highly variable reproductive isolation among pairs of Catostomus species. Mol Ecol 24(8):1856–1872. https://doi.org/10.1111/mec.13118

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17:10. https://doi.org/10.14806/ej.17.1.200

Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD (2018) PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Res 47:D4190–D426. https://doi.org/10.1093/nar/gky1038

Narum SR (2006) Beyond Bonferroni: less conservative analyses for conservation genetics. Conserv Genet 7:783–787. https://doi.org/10.1007/s10592-005-9056-y

Parchman TL, Gompert Z, Braun MJ, Brumfield RT, McDonald DB, Uy JAC, Buerkle CA (2013) The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. Mol Ecol 22:3304–3317. https://doi.org/10.1111/mec.12201

Polechová J, Barton N (2011) Genetic drift widens the expected cline but narrows the expected cline width. Genetics 189:227–235. https://doi.org/10.1534/genetics.111.129817

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959. https://doi.org/10.1111/j.1471-8286.2007.01758.x

Rafajlović M, Emanuelsson A, Johannesson K, Butlin RK, Mehlig B (2016) A universal mechanism generating clusters of differentiated loci during divergence-with-migration. Evolution 70:1609–1621. https://doi.org/10.1111/evo.12957

Ravinet M, Faria R, Butlin RK, Galindo J, Bierne N, Rafajlović M, Westram AM (2017) Interpreting the genomic landscape of speciation: finding barriers to gene flow. J Evolut Biol 30:1450–1477. https://doi.org/10.1111/jeb.13047

Recuero E, Canestrelli D, Vörös J, Szabó K, Poyarkov NA, Arntzen JW, Martínez-Solano I (2012) Multilocus species tree analyses resolve the radiation of the widespread *Bufo bufo* species group (Anura, Bufonidae). Mol Phylogenetics Evol 62:71–86. https://doi.org/10.1016/j.ympev.2011.09.008

Ren J, Duan Y, Qiao R, Yao F, Zhang Z, Yang B, … Huang L (2011) A missense mutation in PPARD causes a major QTL effect on ear size in pigs. PLoS Genet *7* https://doi.org/10.1371/journal.pgen.1002043

Rice WR (1989) Analyzing tables of statistical tests. Evolution 43:223–225

Rieseberg LH, Whitton J, Gardner K (1999) Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. Genetics 152:713–727

Rousset F (2008) GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. Mol Ecol Resour 8:103–106. https://doi.org/10.1111/j.1471-8286.2007.01931.x

Sequeira F, Arntzen JW, van Gulik D, Hajema S, Diaz RL, Wagt M, & van Riemsdijk I (2022) Genetic traces of hybrid zone movement across a fragmented habitat. J Evol Biol https://doi.org/10.1111/jeb.13982

Stankowski S, Sobel JM, Streisfeld MA (2016) Geographic cline analysis as a tool for studying genome-wide variation: a case study of pollinator-mediated divergence in a monkeyflower. Mol Ecol https://doi.org/10.1111/mec.13645

Streicher JW, Wellcome Sanger Institute Tree of Life programme, Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective, Tree of Life Core Informatics collective, D. T. of L. C (2021) The genome sequence of the common toad, Bufo bufo (Linnaeus, 1758). Wellcome Open Res 6:281. https://doi.org/10.12688/wellcomeopenres.17298.1

Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, von Mering C (2021) The STRING database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res 49:D605–D612. https://doi.org/10.1093/nar/gkaa1074

Teeter K, Thibodeau LM, Gompert Z, Buerkle CA, Nachman C, M W, Tucker PK (2009) The variable genomic architecture of isolation between hybridizing species of house mice. Evolution 64:472–485. https://doi.org/10.1111/j.1558-5646.2009.00846.x

van Riemsdijk I, Butlin RK, Wielstra B, Arntzen JW (2019) Testing an hypothesis of hybrid zone movement for toads in France. Mol Ecol 28(5):1070–1083. https://doi.org/10.1111/mec.15005

Vines TH, Dalziel AC, Albert AYK, Veen T, Schulte PM, Schluter D (2016) Cline coupling and uncoupling in a stickleback hybrid zone. Evolution 70:1023–1038. https://doi.org/10.1111/evo.12917

Visser M, Leeuw M, De, Zuiderwijk A, Arntzen JW (2017) Stabilization of a salamander moving hybrid zone. Ecol Evol 7:689–696. https://doi.org/10.1002/ece3.2676

Wickbom T (1945) Cytological studies on Dipnoi, Urodela, Anura, and Emys. Hereditas 31:241–346. https://doi.org/10.1111/j.1601-5223.1945.tb02756.x

Wielstra B (2019) Historical hybrid zone movement: more pervasive than appreciated. J Biogeogr, 1–6. https://doi.org/10.1111/jbi.13600

Zhang LC, Li N, Liu X, Liang J, Yan H, Zhao K, Bin, Wang LX (2014) A genome-wide association study of limb bone length using a Large White × Minzhu intercross population. Genet Sel Evol 46:1–8. https://doi.org/10.1186/s12711-014-0056-6

ACKNOWLEDGEMENTS

We thank the reviewers and editor for their comments, which helped improve the manuscript. We thank Frido Welker, Tara Luckau, and the members of the Butlin, Allentoft, Richards, and Bossdorf laboratories for support and discussion. IvR was supported by the 'Nederlandse Organisatie voor Wetenschappelijk Onderzoek' (NWO Open Programme 824.14.014). This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 655487. Part of this project was carried out by IvR at the Shaffer laboratory in Los Angeles, at the University of California. This study trip has been sponsored by the Leiden University Fund / Swaantje Mondt Fonds (D7102). MR was funded by the Hasselblad Foundation Grant to Female Scientists, a grant from the Swedish Research Council Formas (grant number: 2019-00882), from the Swedish Research Council Vetenskapsrådet (grant number: 2021-05243), and by additional grants from Swedish Research Councils (Formas and VR) to the Centre for Marine Evolutionary Biology at the University of Gothenburg. This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 OD018174 Instrumentation Grant.

AUTHOR CONTRIBUTIONS

IvR, JWA, BS, and BW designed the study. IvR and JWA collected samples. IvR performed the laboratory work and data assembly with contributions from GB, EMM, PS, and ET. IvR analyzed and interpreted the data with contributions from BS, BW, JWA, MR, and PS. IvR, BW, and BS wrote the manuscript with input from all authors.

Compliance with ethical standards

COMPETING INTERESTS

The authors declare no competing interest.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41437-023-00617-6.

Correspondence and requests for materials should be addressed to I. van Riemsdijk.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.