MS. ISOLDE VAN RIEMSDIJK (Orcid ID : 0000-0001-9739-6512)

DR. BEN WIELSTRA (Orcid ID : 0000-0002-7112-5965)

DR. J. W. (PIM) ARNTZEN (Orcid ID : 0000-0003-3229-5993)

Article type : Original Article

Journal: Molecular Ecology

Title: Testing an hypothesis of hybrid zone movement for toads in France

Authors: I. van Riemsdijk[1,2 *], R. K. Butlin[3,4], B. Wielstra[1,2,3,5], J. W. Arntzen[1]

*Corresponding author

Author affiliations:

1. Taxonomy and Systematics, Naturalis Biodiversity Center, Leiden, The Netherlands
2. Institute of Biology Leiden, Leiden University, Leiden, The Netherlands
3. Department of Animal and Plant Sciences, University of Sheffield, Sheffield, The United Kingdom
4. Department of Marine Sciences, University of Gothenburg, Gothenburg, Sweden
5. Department of Ecology and Evolutionary Biology, University of California, Los Angeles, USA

Corresponding author: I. van Riemsdijk isolde.vanriemsdijk@naturalis.nl

Abstract

Hybrid zone movement may result in substantial unidirectional introgression of selectively neutral material from the local to the advancing species, leaving a genetic footprint. This genetic footprint is represented by a trail of asymmetric tails and displaced cline centres in the wake of the moving hybrid zone. A peak of admixture linkage disequilibrium is predicted to exist ahead of the centre of the moving hybrid zone. We test these predictions of the movement hypothesis in a hybrid zone between common (*Bufo bufo*) and spined toads (*B. spinosus*), using 31 nuclear and one mtDNA SNPs along a transect in the northwest of France. Average effective selection in *Bufo* hybrids is low and clines vary in shape and centre. A weak pattern of asymmetric introgression is inferred from cline discordance of seven nuclear markers. The dominant direction of gene flow is from *B. spinosus* to *B. bufo* and is in support of southward movement of the hybrid zone. Conversely, a peak of admixture linkage disequilibrium north of the hybrid zone suggests northward movement. These contrasting results can be explained by reproductive isolation of the *B. spinosus* and *B. bufo* gene pools at the southern (*B. spinosus*) side of the hybrid zone. The joint occurrence of asymmetric introgression and admixture linkage disequilibrium can also be explained by the combination of low dispersal and random genetic drift due to low effective population sizes.

Key words: Admixture linkage disequilibrium, asymmetric reproductive isolation, *Bufo bufo, Bufo spinosus,* cline coupling, hybrid zone movement

Introduction

During allopatric speciation, a taxon's range is split by a physical barrier, and the vicariant populations gradually diverge through processes such as mutation, natural selection, and genetic drift (Mayr, 1942; Lande, 1980; Wu & Ting, 2004). Diverged taxa may later meet in secondary contact and form a hybrid zone, for example when taxa expand their ranges from glacial refugia in the postglacial era (Barton & Hewitt, 1985; Hewitt, 1988, 2011). Upon secondary contact the two

populations may have established reproductive isolation, inhibiting exchange of genetic material (Rice, 1998). Conversely, when the fitness of hybrid populations is equal or increased compared to the fitness of parental populations, gene flow is extensive, and the two populations will merge (Anderson & Stebbins, 1954; Arnold, 2004; Seehausen, 2004; Hedrick, 2013). In between these extremes, a reduced hybrid fitness precludes merging and facilitates limited introgression (Mallet, 2005).

When hybrid fitness is reduced independent of the environment, a narrow tension zone exists between the two species (Moore, 1977; Barton & Hewitt, 1985). In the classical hybrid zone literature such tension zones are thought to stabilise where population density is low, or at an ecological barrier to dispersal (Endler, 1977; Barton & Hewitt, 1985). Recent studies suggest that hybrid zone movement is more common than previously thought (Arntzen & Wallis, 1991, 1999; Buggs, 2007; Roy, O'Connor, & Green, 2012; Taylor et al., 2014; Leaché, Grummer, Harris, & Breckheimer, 2017; Wielstra, Burke, Butlin, & Arntzen, 2017; Wielstra, Burke, Butlin, Avcı, et al., 2017; Ryan et al., 2018). Hybrid zone movement occurs when one parent species has a higher fitness than the other, such as through a competitive advantage, asymmetric hybrid fitness effects, or environmental change (Buggs, 2007). Introgression caused by hybrid zone movement is thought to affect selectively neutral markers distributed randomly across the genome, resulting in genetic traces of the displaced species in populations of the advancing species (Moran, 1981; Currat, Ruedi, Petit, & Excoffier, 2008). In contrast, under adaptive introgression, alleles from one species with a positive effect in the other species may introgress through the hybrid zone, regardless of whether the zone is stable or moving (Seehausen, 2004; Mallet, 2005; Hedrick, 2013; Schmickl, Marburger, Bray, & Yant, 2017). Adaptive introgression affects only the selected marker and markers that are nearby on the chromosome (physical linkage), or markers that interact functionally with the marker under selection (functional linkage; Barton & Hewitt, 1985; Baird, 2015; Sedghifar, Brandvain, & Ralph, 2016). Only a fraction of the genome is involved in adaptive introgression, whereas hybrid

zone movement results in genome-wide introgression (Currat et al., 2008; Wielstra, Burke, Butlin, Avcı, et al., 2017).

The transition from one species to the other through the hybrid zone can be described by a set of allele frequency gradients, i.e. geographical clines (Schmickl et al., 2017). The positions and shapes of geographical clines are indicative of evolutionary processes involved in hybrid zone formation (Barton & Hewitt, 1985). Two types of cline variation are distinguished based on cline position (coincident or displaced) and cline shape (concordant or discordant). Coincident clines share the same cline centre, whereas displaced clines have a cline centre away from the majority of the clines for other loci. Concordant clines are similar in shape (described by e.g. width and tail shape; Szymura & Barton, 1991), whereas discordant clines have a shape divergent from the majority of the clines. Coincident clines are considered evidence of a hybrid zone in a stable position over time (Abbott et al., 2013). Concordant and narrow clines (steep at the centre and with distinct tails) suggest that the hybrid zone is maintained by a balance between strong selection and dispersal, as opposed to displaced and wide clines indicating that selection against hybrids is low (shallow slope without distinct tails; Barton & Gale, 1993). Displaced and discordant clines exist in many species, both for nuclear and mitochondrial markers, and may be caused by sex-biased gene flow or by adaptive introgression (Mallet, 2005; Excoffier, Foll, & Petit, 2009; Toews & Brelsford, 2012; While et al., 2015; Sloan, Havird, & Sharbrough, 2016; Bonnet, Leblois, Rousset, & Crochet, 2017). However, displaced and discordant clines may also be caused by hybrid zone movement, when a trail of genetic material from the overtaken species remains present in the overtaking species (Rohwer, Bermingham, & Wood, 2001; Gay, Crochet, Bell, & Lenormand, 2008; Excoffier et al., 2009; Arntzen, de Vries, Canestrelli, & Martínez-Solano, 2017; Wielstra, Burke, Butlin, Avcı, et al., 2017).

Whether or not markers reflect past or ongoing demographic processes, including hybrid zone movement and adaptive introgression, is dependent on factors such as the number and genomic distribution of barrier loci, mutation rate, and recombination (reviewed in Ravinet et al., 2017). Barrier loci are genes which inhibit the merging of two diverged populations (Abbott et al., 2013;

Barton, 2013). A stronger overall barrier to gene flow is established when multiple barrier effects become coincident by processes summarised as coupling (reviewed in Butlin & Smadja, 2017). If genome-wide coupling of barrier effects occurs, the geographic clines for markers are expected to be coincident even if they are not physically or functionally linked to a barrier gene (Barton, 1983; Barton & Gale, 1993; Bierne, Welch, Loire, Bonhomme, & David, 2011; Vines et al., 2016). To overcome the coupling effect of multiple barrier genes and so to cause gene flow across the hybrid zone, the positive selection on a single gene will have to be disproportionally strong (e.g. Vines et al., 2016).

In addition to physical and functional linkage, associations between markers in hybrid zones can be caused by common descent. This is reflected by an increase in admixture linkage disequilibrium in the hybrid zone as alleles originating from the same parent species tend to be found together within the genomes of early generations of hybrid offspring. Recombination during subsequent backcrossing breaks down admixture linkage disequilibrium (Barton & Gale, 1993; Baird, 2015). However, when the hybrid zone is moving, the peak of admixture linkage disequilibrium is predicted to be positioned just ahead of the hybrid zone centre, where individuals with little history of recombination are involved in reproduction (Gay et al., 2008; Wang et al., 2011). This allows us to make predictions based on the hypothesis of hybrid zone movement: there will be a tail of introgression in the wake of the moving zone combined with a shift of the peak in admixture linkage disequilibrium ahead of the movement. This contrasts with the predictions for a stable situation, where clines are symmetric and the peak of admixture linkage disequilibrium appears in the centre of the hybrid zone (Fig. 1).

We test for a genetic footprint of movement in a hybrid zone in northwest France between two morphologically similar but genetically distinct species of toad, the common, *Bufo bufo* (Linnaeus, 1758) and the spined toad *B. spinosus* Daudin, 1803 (Recuero et al., 2012; Arntzen, Recuero, Canestrelli, & Martínez-Solano, 2013; Trujillo, Gutiérrez-Rodríguez, Arntzen, & Martínez-Solano, 2017). While *B. bufo* survived the last glacial maximum in Italy and the Balkans, *B. spinosus* survived

in southern France and Spain (Arntzen et al., 2017). Based on four nuclear gene coding markers, twelve microsatellites, and two mitochondrial SNP markers, the width of the hybrid zone (30 km) compared to the dispersal distance (1.3 km generation^{-1}) suggested that selection against hybrids restricts gene flow through the hybrid zone (Arntzen et al., 2016). In the southeast of France, the *Bufo* hybrid zone was hypothesised to move southwards based on traces of introgression of one nuclear gene coding marker (of four tested), and discordance between nuclear, mtDNA, and morphologic clines (Arntzen et al., 2017). In the northwest of France, a similar dataset including microsatellite data, showed introgression towards the north in two nuclear markers out of four, but was not analysed from the perspective of hybrid zone movement at the time (Arntzen et al., 2016). This scenario provides the opportunity to test for the hypothesis of hybrid zone movement in the northwest of France. We improve the resolution using 31 gene coding markers and one mtDNA marker to interpret interspecific gene flow in the light of the average effective selection on a locus, and place our findings in the context of results from the literature.

Materials and Methods

Sample collection and preparation

All DNA extracts were available from previous studies (Arntzen, Wilkinson, Butôt, & Martínez-Solano, 2014; Arntzen et al., 2016, 2017; Arntzen, McAtear, Butôt, & Martínez-Solano, 2018). Eight individuals from four *B. bufo* reference populations and twelve individuals from seven *B. spinosus* reference populations, all positioned > 500 km away from the centre of the hybrid zone (blue and red circles in Fig. 2 a, b), served to identify species diagnostic single nucleotide polymorphisms (SNPs) in a test run. Another 268 individuals from 29 populations, including transect populations (grey circles in Fig. 2 a, b) and additional samples from the reference populations were genotyped and included in the dataset, to a total of 306 individuals. Distances between transect populations were measured in a straight line, which follows the direction of the transect in Arntzen et al. (2016),

using a custom R script. Distances for reference populations were measured with Google Earth Pro v.7.3.0 (Table 1 and Table S.1).

SNP design based on transcriptome data

Two transcriptomes from liver tissue of a *B. bufo* individual from population 6 (Audresselles, France), and a *B. spinosus* individual from population 18 (Jublains, France) were sequenced commercially by ZF Screens, Leiden, on the Illumina HiSeq 2000 platform. The data were filtered with Trimmomatic v.0.35 software (Bolger, Lohse, & Usadel, 2014), and assembled with Trinity v.2.1.1 (Grabherr et al., 2011; Haas et al., 2013). When using exons from transcriptome data for SNP design in expressed genes, paralogs (gene copies) should be excluded to avoid interpreting SNPs between paralogs as SNPs within orthologues (e.g. Ryynänen & Primmer, 2006). Exon boundaries need to be taken into account, as primers designed across exon boundaries are the main cause of genotyping assay failure (Wang et al., 2008). First, a BLAST search (Altschul, Gish, Miller, Myers, & Lipman, 1990) was used to identify paralogs within the *B. bufo* transcriptome assembly. Next, a pipeline for exon boundary identification (Niedzicka, Fijarczyk, Dudek, Stuglik, & Babik, 2016) was used, employing the well-annotated *Xenopus tropicalis* (Gray, 1864) genome (genome version JGI4.2). The *X. tropicalis* genome was also used to annotate markers (after testing for diagnostic differences between the two species, 30 of 31 markers eventually used were annotated, one remained unidentified and was named *exon_1*). Finally, a second BLAST of the selected exons against the transcriptome of *B. spinosus* was used to exclude potentially undetected paralogs and to identify SNPs. A detailed description of the SNP design from transcriptome data is presented in the Supplemental Information text.

SNP validation

Fluorescence-based genotyping (Semagn, Babu, Hearne, & Olsen, 2014) was used in the Kompetitive Allele-Specific PCR (KASP) genotyping system at the SNP genotyping facility of the Institute of

Biology, Leiden University. Primer design, PCR setup, and data visualisation followed Arntzen et al. (2016). For 96 exon sequences, primers were designed with the Kraken software (LGC genomics, UK). After a first run, three SNPs were excluded as they failed for all individuals of one species, presumably due to a mutation at a primer binding site. Markers were initially considered diagnostic if they showed species-specific alleles in the test dataset. This resulted in 32 nuclear DNA SNPs (Table S.2). Because of limitations due to plate dimensions in the KASP system, the remainder of the samples was analysed with 31 SNPs (by excluding the least well performing marker in the test dataset, *scaf4*, for which 6 out of 20 reference individuals had missing data) and we added one previously published diagnostic mtDNA SNP (16S; Arntzen et al., 2016). We excluded five individuals from the final dataset with more than ten SNPs missing (presumably because of low quality DNA). After assessment of the final dataset of 306 individuals, 29 of 31 nuclear SNPs were considered species-diagnostic, defined by minor allele frequencies of < 5.0 % in reference populations in the final dataset (Table S.3). Missing data amounted to 5.2% (Table S.4).

Hardy-Weinberg proportions and marker linkage

Because SNPs in genes are more prone to be under direct or indirect selection than non-coding DNA, we were cautious to exclude any markers showing outlier behaviour. Signals of non-random mating or survival were tested for by calculating the Hardy-Weinberg proportions (heterozygote deficit and excess) with the R package genepop based on the program GENEPOP v.1.0.5 (Rousset, 2008). As the Bonferroni correction (Rice, 1989) can be overly conservative (e.g. Narum, 2006), we chose to account for independence of tests within markers (P_c for N = 31). Deviations from Hardy-Weinberg proportions by heterozygote excess were not significant (P_c > 0.05; Table S.5). Deviations by heterozygote deficit were significant for the marker *egflam* in five populations (P_c < 0.05), hence this marker was excluded in the HZAR cline fitting analysis and admixture linkage disequilibrium calculations (see below).

To infer independence of introgression, close physical or functional linkage of markers should be investigated. As some of the 31 nuclear markers are bound to be on the same chromosome (*Bufo* has 22 chromosomes; Olmo, Gargiulo, & Morescalchi, 1970; Olmo, 1973), we used a test of pairwise linkage disequilibrium (LD), indicative of close physical linkage of markers, based on the log likelihood ratio statistic (*G*-test) with the R-package genepop (Rousset, 2008). One pair of markers, *exon_1* and *ttc37*, was found to be in significant pairwise LD in two populations ($P_c < 0.05$; Table S.6). As these populations were located in the hybrid zone, where we expect the effect of admixture linkage disequilibrium to increase the number of markers in pairwise LD, we kept the two markers in downstream analyses, but we checked that this marker pair did not distort the general patterns described below. To investigate functional linkage, a protein-interaction network was analysed with STRING v.10.5 (Szklarczyk et al., 2015) on the basis of the *X. tropicalis* genome, and 30 out of 31 annotated nuclear markers. A protein function description was recorded for each annotated marker (Table S.7). The markers *brca2* and *rfc1* were found to be functionally linked. As these markers were neither deviating from Hardy-Weinberg proportions nor in pairwise linkage disequilibrium within populations, they were included in all downstream analyses but, again, we checked that this marker pair did not distort the general patterns described below. Note that it remains possible that individual markers are under direct or indirect selection, and the current and future analyses using these markers need to take this into account.

Population structure

Allocation of individuals to genetically defined groups was done using all 31 nuclear markers in Structure v.2.3.4 (Pritchard, Stephens, & Donnelly, 2000) with ten independent chains of one up to ten genetic clusters (*K*), a burn in of 10,000, and a chain length of 25,000 under the admixture model following recommendations of Benestan et al. (2016). Other settings were left at default. Credibility intervals (95%) were estimated and convergence was checked comparing log likelihood and similarity of admixture proportion (Structure Q scores) between runs. The results were summarized

with CLUMPAK (Kopelman, Mayzel, Jakobsson, Rosenberg, & Mayrose, 2015). The best value of *K* was determined using the Evanno method (Evanno, Regnaut, & Goudet, 2005; Table S. 8), and results were visualised using the R package POPHELPER (Francis, 2017). The admixture proportions are used in the cline analysis (see below) to approach an overall cline shape and position that summarizes all nuclear clines.

Cline analyses

Classic equilibrium cline models, describing a sigmoid change in allele frequency or phenotype across the hybrid zone, were fitted with the R package HZAR (Derryberry, Derryberry, Maley, & Brumfield, 2014) for 30 nuclear markers, mtDNA, and admixture proportion (K = 2), using a set of custom R scripts provided by G. Derryberry. Marker *egflam* was not analysed with HZAR as it was not behaving according to Hardy-Weinberg proportions. First, 30 maximum likelihood estimation 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 (ΔAICc) below ten. Fifteen model variants can be fitted in HZAR, based on all possible combinations of trait interval (allele frequency at the transect ends; three types) and cline shape. The different cline shapes represent either a single sigmoid curve, or a combination of (1) the central sigmoid portion and (2) shallower exponential decay curves at one or both ends (tails) with slopes shallower than expected from the central portion. We refer to these alternatives as 'tail types' (five types). Tail types were fitted as an exponential tail to the left (L), exponential tail to the right (R), both tails exponential with independent parameters (B), both tails exponential but with the same parameters mirrored on the cline centre (M), and no exponential tails fitted (N). For cline tails estimated separately (B), we assessed tail slope (τ; shallow tails reflected by low parameter values) and distance from the cline centre where a tail started (δ; short distances reflected by low parameter values). When τ and δ are both low on one side and both high at the other side of the hybrid zone, introgression is asymmetric. Convergence of the HZAR analysis was visually assessed in trace plots. Significance of cline model fits

was determined by calculating ΔAICc for each best model (ΔAICc > 2 compared to the second best model; Table S.9, S.10, Figure S.1). Deviation from an equal frequency of left and right tails (as expected in a stable hybrid zone) was tested for all markers with a coincident cline centre with the chi-square test for equal probabilities.

Admixture linkage disequilibrium

Using variance in hybrid index (HI, see also Table 1), admixture linkage disequilibrium (*D'*) can be calculated, and subsequently lifetime dispersal distance (σ) weighted for individuals in aquatic (pre-) and terrestrial (post-metamorphosis) life stages, expected cline width under neutrality, and average effective selection on a locus (s^*) can be calculated following Barton & Gale (1993). This was done using 29 nuclear markers, excluding marker *egflam*, which was not within Hardy-Weinberg proportions and marker banp, for which the cline centre was displaced compared to other clines (Figure S.2). Average effective selection is the selection pressure on a locus at the zone centre due to direct selection or association with other loci under selection. A few input parameters were needed for these calculations (Table S.11). A mean recombination rate between marker pairs of 0.4997 was calculated following formula (6) from Macholán et al. (2007), using the number of chiasmata per bivalent for *Bufo bufo* (1.95; Wickbom, 1945), and the number of chromosomes for *Bufo bufo* (N = 22). A generation time of 2.5 years for *Bufo* toads at the latitude of the hybrid zone (mean of three years in females and two years in males; Hemelaar, 1988), and initial secondary contact at 8,000 years ago following Arntzen et al. (2016) were used as input parameters. The width of the hybrid zone was derived from a general sigmoid cline model following the 'no-tails' formula in HZAR (Derryberry et al., 2014) fitted to the *HI*. At the cline centre, *D'* was estimated from its regression on p*(1-p), where p is the average over loci of the frequency of the *B. bufo* allele at a sample location. Means and 95% confidence intervals (CI) for cline width, *D'* at the centre, dispersal distance, expected cline width under neutrality, and effective selection were based on 1,000 bootstrap replicates of the original genotype dataset (with replacement, maintaining original sample size;

Table S.11), using custom R scripts. The amplitude and position of the peak in *D'* were estimated by fitting a Gaussian curve following Gay et al. (2008).

Hybrid index analysis

To visualise potential unequal contribution of both species to the genomic composition of admixed individuals, we used the R package HIest (Fitzpatrick, 2012). This analysis scales the number of markers derived from each parental species (*S*) to heterozygosity, which is calculated as the fraction of heterozygous markers with variants inherited from both parent species, within an individual (H_I). We used 29 markers with an allele frequency difference between species more than 0.8 (Larson, Andrés, Bogdanowicz, & Harrison, 2013). To assess unequal inheritance of mtDNA in hybrid offspring, we coloured the data points by mtDNA type.

Results

The number of genetic clusters (*K*) supported by Structure was two, concordant with the two toad species (Table S.8). Species admixture was recorded in populations 7-15. Individuals of populations 7-9 north of the hybrid zone centre were hybrid populations with a relatively high frequency of *B. spinosus* nuclear alleles when compared to the number of *B. spinosus* mtDNA haplotypes (Fig. 2c). Credibility intervals of the admixture proportion estimates are shown in Figure S.3.

HZAR clines were partitioned according to tail type (Fig. 3, for a plot including all clines see Figure S.2, Tables S.9, S.10). Eight nuclear marker clines best fitted a left tail of introgression (northwards into *B. bufo*) and for seven markers the fit was significantly better than alternative models (Fig. 3a). Two nuclear marker clines had a right tail of introgression (southwards into *B. spinosus*), of which one had a significantly better fit than alternative models (Fig. 3b). For four nuclear marker clines, both tails were exponential with independent parameters, of which three had a significantly better fit than alternative models (Fig. 3c). For these clines, tail slope (τ) and distance from the cline centre where the tail started (δ) showed an increase in one parameter (e.g. δ) and a decrease in the other

(e.g. τ) in all cases. Such results are not straightforward to interpret in terms of directional introgression and these clines were not taken into account when assessing unidirectional introgression. For eleven nuclear markers, clines had mirrored tails of which six had a significantly better fit than alternative models (Fig. 3d). For three nuclear markers, clines had no tails, and one marker had a significantly better fit than alternative models (Fig. 3e). All nuclear markers had cline centres within 30 km of the admixture proportion cline centre (positioned at 539 km) except for *banp*, with a cline centre > 300 km to the south, inside the range of *B. spinosus* (Figure S.2). This cline had a non-significant left tail into *B. bufo* (Fig. 3f). The mitochondrial marker had a significant right tail into *B. spinosus* (Fig. 3f). The functionally linked markers *brca2* and *rfc1* had similar cline models with mirrored tails and overlapping 95% credible cline regions for centre and width, and showed no pattern of introgression deviating from the admixture proportion cline. In summary, seven nuclear markers showed significant introgression of *B. spinosus* into *B. bufo* by cline discordance, whereas two nuclear markers showed significant introgression the other way around, one by cline discordance and one by displacement. The inequality of introgression was non-significant (Chi-square test for equal probabilities, $Chi^2 = 2.78$, df = 1, P = 0.096).

Populations 9-15 within the hybrid zone had higher values of *D'* than populations outside the hybrid zone (Fig. 4). This is reflected by an increase in pairwise linkage disequilibrium among loci (Table S.6). Populations 7, 8, and 10 north of the hybrid zone centre, and beyond the area of rapid allele frequency change, had larger *D'* with larger positive deviations from zero than other populations. However, it was not possible to fit a Gaussian curve through these data points, probably because of the large variation in *D'*, and the position and amplitude of the peak of *D'* remain undetermined.

Estimated lifetime dispersal (σ), based on *D'* at the cline centre, was 2.2 km generation$^{-1/2}$ (95% CI 0.6-3.7). The observed cline width based on *HI* was 114 km (95% CI 103-132). Assuming no selection against hybrids, the expected cline width would be 312 km (95% CI 91-532; Table S.11). The average effective selection on a locus (s*) was 0.0017 (95% CI 0.0001-0.0040). Despite variation among clines

suggesting weak coupling, the presence of an admixture LD peak supports the use of *s** as an average value to describe the hybrid zone. HIest showed a continuum of *B. bufo* to *B. spinosus* genotypes in which hybrids possessing mtDNA haplotypes typical of either *B. bufo* or *B. spinosus* were about equally frequent (Figure S.4). The peak of ancestry is at 50% heterozygosity, indicating no asymmetry in the contribution of both species to the hybrid population.

Discussion

The features describing the *Bufo* hybrid zone in northwest France in the current study are broadly in line with previous descriptions, but provide more resolution and estimates of key parameters. We found that individual marker cline positions and shapes vary (e.g. Figure S.2), widening the admixture proportion cline. We inferred weak effective selection per locus against hybrids (s*=0.0017), but with low precision of the estimate. Even when effective selection is low and clines are wide, a moving hybrid zone may stagnate at an ecotone, limiting further movement of the zone (Endler, 1977; Moore, 1977; Barton & Hewitt, 1985; Buggs, 2007; Gompert, Mandeville, & Buerkle, 2017). The section of the *Bufo* hybrid zone studied here appears to be located at a weak ecotone in a hilly landscape formed by the 'Collines de Normandie', with *B. bufo* at a lower and *B. spinosus* at a higher altitude (Arntzen et al., 2016). The genetic estimate of lifetime dispersal of 2.2 km generation$^{-1/2}$ diverges from the average maximum observed in field studies, but the field estimate is within the CI for the genetic estimate (1.3 km; Smith & Green, 2005; Daversa, Muths, & Bosch, 2012; Trochet et al., 2014). In amphibians, dispersal distance based on genetic data is more often found to be higher compared to field studies, because mark-recapture, seasonal, or field studies covering a small area are at risk of underestimating (rare) long-distance dispersal (Smith & Green, 2005).

The hybrid zone was calculated to be 114 km wide, based on the hybrid index for 30 nuclear markers, and width was 63 km based on the cline of the admixture proportion for 31 nuclear markers. A narrower width (30 km) was earlier recorded based on the first PCA axis of 12 microsatellites, whilst in the same study cline width for four nuclear markers ranged from 27 to 91

km, and morphological cline widths exceeded 200 km (Arntzen et al., 2016). Microsatellites are noncoding, have higher mutation rates, and they are more suitable for investigating small scale evolutionary processes, therefore they may reflect more recent demographic processes (Ellegren, 2000). This poses the question what is causing a more spatially restricted transition in the microsatellite markers as opposed to a wider transition in nuclear gene markers, as one would expect stronger selection against hybrids and thus narrower clines in markers from coding regions than in microsatellites. Including more markers would cause the hybrid index or admixture proportion clines to get wider, because drift causes the centres of individual marker clines to vary. The larger number of markers in this paper may thus explain the wider zone currently reported compared to the previously reported width based on less markers. Confidence intervals overlap for many of these width estimates and all widths are large relative to dispersal. This indicates the general conclusion of weak selection is robust.

We examined patterns of gene flow and admixture linkage disequilibrium to test the hypothesis of movement of the hybrid zone in two common toad species (*B. bufo* and *B. spinosus*) in northwest France. Introgression from *B. spinosus* into *B. bufo* was more frequent than introgression the other way around, a result that was found before with four nuclear markers (Arntzen et al., 2016, 2017). Disregarding whether cline model fits are significant, eleven markers show introgression from *B. spinosus* into *B. bufo* and two show introgression the other way around, a significant difference (Chi-square test for equal probabilities, $Chi^2 = 6.23$, $P = 0.013$). Cases where introgression is not explicit can also point out methodological difficulties, such as determining unidirectional introgression from cline shape parameters. To better compare the asymmetry of introgression between individual markers, future methods could compare e.g. the area underneath and above the geographic cline on both sides of the cline centre. Further simulations are needed to provide clear expectations for stable and moving hybrid zones.

Three populations with elevated admixture linkage disequilibrium (*D*') suggest a peak of early-generation hybrid offspring north of the hybrid zone. Fitting a Gaussian curve through *D'* proved impossible, as there was too much variance within the data, especially in populations 5, 9 and 11. The Gaussian curve is the shape of the peak expected under a symmetrical cline model (Gay et al., 2008). The absence of a clear peak can be caused by asymmetry in the cline model or insufficient sample size of individuals, populations, or markers, or a combination of both. More data should be generated to be able to determine the shape and position of the peak with higher precision. While the unidirectional introgression is in line with southward movement of the hybrid zone (Buggs, 2007), the northern position of a peak in *D'* is in line with movement in the opposite direction (Gay et al., 2008; Wang et al., 2011). Since our results do not fully support the original hypothesis of (southward) hybrid zone movement in *Bufo*, alternative explanations have to be considered, in particular for the position of the peak in admixture linkage disequilibrium, for which we compare our result with other case studies.

Concordant and narrow clines suggest that a hybrid zone is maintained by a balance between strong selection and dispersal, as opposed to wide and displaced clines indicating that selection against hybrids is low (Barton & Gale, 1993). By surveying the hybrid zone literature, we found eight studies with empirical data on effective selection (i.e. s*) and cline centre and shape in a wide variety of organisms (Table S.12; Mallet et al., 1990; Szymura & Barton, 1991; Phillips, Baird, & Moritz, 2004; Macholán et al., 2007; Kawakami, Butlin, Adams, Paull, & Cooper, 2009; Carneiro et al., 2013, Baldassarre, White, Karubian, & Webster, 2014; Hollander, Galindo, & Butlin, 2015; Wielstra, Burke, Butlin, & Arntzen, 2017). Six studies report substantial effective selection (0.19 – 0.37) with mostly concordant, coincident and steep clines, and three studies report low average effective selection (0.007 – 0.011) with multiple discordant and displaced clines. Our results for *Bufo*, with low effective selection and substantial cline variability, fall into the latter category.

We also consulted the literature on the direction of introgression and the location of elevated levels of admixture linkage disequilibrium in hybrid zones. Two studies document a peak in admixture linkage disequilibrium opposite to where introgression takes place (Fig. 5a), and one study documents a peak of admixture linkage disequilibrium and introgression on the same side (Fig. 5b). In the hybrid zone of the gull species *Larus glaucescens* (Naumann, 1840) and *Larus occidentalis* (Audubon, 1839), genotypic and phenotypic clines showed introgression towards the north, and a peak in admixture linkage disequilibrium towards the south, indicating a southward movement of the hybrid zone (Fig. 5a; Gay et al., 2008). In the hybrid zone of the house mouse subspecies *Mus musculus musculus* Linnaeus, 1758 and *M. m. domesticus* Schwarz & Schwartz, 1943 introgression to the east and a peak of admixture linkage disequilibrium to the west, supported a westward movement of the hybrid zone (Fig. 5a; Wang et al., 2011). Conversely, in the hybrid zone of the salamander subspecies *Ensatina eschscholtzii eschscholtzii* Gray, 1850 and *E. e. klauberi* Dunn, 1929, unidirectional introgression coincides with a peak of admixture linkage disequilibrium (Fig. 5b; Devitt et al., 2011). This result was explained by asymmetric pre- or postzygotic isolation. On the side of the hybrid zone where reproduction is reduced, introgression and admixture of genetically distinct individuals are absent, whereas at the other side where hybridisation succeeds, introgression and admixture of genetically distinct individuals occur. A deficit of hybrids with *E. e. eschscholtzii* mtDNA confirms that offspring of female *E. e. klauberi* and male *E. e. eschscholtzii* are successfully reproducing, whilst offspring of other parental combinations occur rarely.

When comparing patterns of introgression between hybrid zones of different biological systems, dispersal rate and effective population size need to be taken into account (Canestrelli et al., 2016; Ravinet et al., 2017). Introgression generally increases when dispersal rate is high, because more genetically distant populations interbreed (Nichols & Hewitt, 1994). The effect of a higher dispersal on introgression is greater for a narrow hybrid zone where selection against hybrids is high, than for a wide zone where selection against hybrids is low, because in a narrow zone the distance between genetically distinct individuals is smaller (Barton & Gale, 1993). Therefore, we chose to express

dispersal in terms of the number of generations needed to cross the hybrid zone, as this summarises width in relation to dispersal, and thus also in relation to selection against hybrids. To include the effect of genetic drift in the comparison of hybrid zones, effective population size can be used. When effective population size is low, an increase in the effect of random genetic drift obscures genetic patterns of demographic and evolutionary history, and clines are predicted to be more randomly distributed in the landscape and vary in shape, despite a high selection against hybrids (Polechová & Barton, 2011). The *Larus* hybrid zone would take five generations of unimpeded dispersal to cross (Fig. 5a; Gay et al., 2008). Populations consist of 10,000 – 100,000 individuals (Crochet et al., 2003). The *Mus* hybrid zone would take about 60 generations to cross (Fig. 5a; Raufaste et al., 2005; Macholán et al., 2007; Wang et al., 2011). The effective population size of mice is 500 to 5,000 individuals (Pocock, Hauffe, & Searle, 2005). The *Ensatina* hybrid zone would take seven generations to cross (Fig. 5b; Staub, Brown, & Wake, 1995; Devitt et al., 2011). In high quality habitat *Ensatina* population density may be 1,300 individuals ha^{-1} with effective population size unknown (Stebbins, 1954; Rosenberg, Noon, Megahan, & Meslow, 1998).

The *Bufo* hybrid zone would take about 60 generations to cross, and dispersal capacity seems comparable to *Mus*. However, the effective dispersal distance in *Mus* is thought to be underestimated by population structure and migration studies, because effective genetic dispersal in mice is inflated by (human mediated) long-distance movements and frequent extinction-recolonization effects (Barton & Hewitt, 1985; Slatkin, 1985; Macholán et al., 2007). Compared to *Bufo*, the effective population sizes in *Larus* and *Mus* are one to three orders of magnitude higher. Toads breed in large water bodies, and adult *Bufo* population sizes regularly consist of 2,500 – 5,000 individuals, but effective population sizes are about two orders of magnitude lower (Scribner, Arntzen, & Burke, 1997). Population sizes in *Bufo* may be more comparable to *Ensatina*, considering that effective population sizes are often much smaller than census population sizes in amphibians (Funk, Tallmon, & Allendorf, 1999; Zeisset & Beebee, 2003; Vences & Wake, 2007). A combination of

low dispersal and small population sizes may impede introgression and cause patterns of admixture linkage disequilibrium to be less pronounced in *Bufo* than in the other hybrid zones.

Drift is strong in populations with low dispersal and small population sizes, generating random variation in allele frequency in different populations. This variation causes clines to differ in shape and position more than would be predicted based on the width of clines of individual alleles (cline wobbling; Polechová & Barton, 2011). Cline wobbling increases the width of the hybrid index (or admixture proportion) cline. A low number of markers, sampling a small portion of the variation in cline shapes and position, may therefore be insufficient to assess patterns of admixture linkage disequilibrium. Further increasing the number of individuals sampled and the number of markers employed would increase our ability to test the hypothesis of hybrid zone movement. It is not currently known how sensitive the methods employed here are to the numbers of individuals or markers included in the dataset. Further investigation through simulations could help to improve our understanding of the limitations of these methods. It will be interesting to see if a higher number of markers generated randomly across the genome, such as RAD sequencing data (Baird et al., 2008), will behave similar or different to these gene-coding markers, as gene-coding markers may be more prone to local selection forces in small populations.

Asymmetric pre- or postzygotic isolation, such as displayed in *Ensatina*, seems most congruent with the co-occurrence of introgression and a peak of admixture linkage disequilibrium seen in *Bufo* (Fig. 5b). However, we find only slight evidence for pre- or postzygotic isolation in the *Bufo* hybrid zone. First, we find marginally higher introgression from *B. spinosus* into *B. bufo* than the other way around. Second, hybrids with both parental types of mtDNA are about equally frequent, and the peak of ancestry is symmetric and centred at 0.5 heterozygosity (Figure S.4), whereas in *Ensatina* almost all hybrid individuals carried mtDNA of only one parental type. Third, low effective selection within the *Bufo* hybrid zone allows the unimpeded exchange of genes and previous studies showed no indication of asymmetric incompatibility (Arntzen et al., 2016; Trujillo et al., 2017). The only evidence in favour of asymmetric isolation besides the position of the admixture linkage

disequilibrium peak, is the relatively high introgression of *B. bufo* mtDNA into *B. spinosus* (e.g. right tail of introgression in mtDNA). Therefore, asymmetric pre- or postzygotic isolation might be weak, with hybridisation inhibited slightly more on the *B. spinosus* side. But this result can also be explained by other processes, such as sex-biased dispersal or adaptive introgression (Currat & Excoffier, 2005; Toews & Brelsford, 2012).

In conclusion, we cannot reject the null hypothesis of a stable *Bufo* hybrid zone in northwest France, and low levels of dispersal and random genetic drift due to small population sizes are important in shaping the patterns of introgression. Based on earlier studies, this hybrid zone was predicted to move southwards (Arntzen et al., 2016, 2017). Our dataset shows similar characteristics to those of previous studies (e.g. northward introgression), but with additional analyses we showed asymmetric reproductive isolation may provide a more important driver of asymmetric introgression than hybrid zone movement. However, multiple processes appear to be at play in shaping the *Bufo* hybrid zone. One might imagine a situation where asymmetric reproductive isolation and hybrid zone movement co-occur where the strongest process may overrule any genetic patterns related to the other process. In addition, small population sizes and other demographic factors such as dispersal distance per generation may cause patterns to become obscured. With 31 nuclear SNPs, we sampled only a tiny portion of the 6 Gbp *B. bufo* genome (Vinogradov, 1998). Further increasing the number of individuals sampled and the number of markers employed can possibly increase our ability to test the hypothesis of hybrid zone movement. It is not currently known how sensitive the methods presented here are to the number of individuals or markers included in the dataset. Further investigation through simulations could help improve our understanding of the limitations of these methods. Looking back on past literature on hybrid zone movement in various organisms, it will be interesting to see if the hypotheses previously inferred still stand with new data and new analytical approaches. The *Bufo* hybrid zone exemplifies the complex influences of interspecific hybridization on genomic composition and can be used to test various hypotheses of introgression and speciation.

Acknowledgements

We thank four anonymous reviewers for their comments, which helped improve the manuscript. We thank Roland Butôt for maintenance of the DNA collection, Wieslaw Babik for advice on the bioinformatics in general, Marta Niedzicka for advice on the pipeline for extraction of exon sequences, Onno Schaap for running the SNP-line, and Graham Derryberry for providing the HZAR scripts. The PhD position of IvR is 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.

References

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., … Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, *26*, 229–246. https://doi.org/10.1111/j.1420-9101.2012.02599.x

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Anderson, E., & Stebbins, G. L. (1954). Hybridization as an evolutionary stimulus. *Evolution*, *8*, 378–388.

Arnold, M. L. (2004). Transfer and origin of adaptations through natural hybridization: were Anderson and Stebbins right? *The Plant Cell*, *16*, 562–570. https://doi.org/10.1105/tpc.HistPersp

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

Arntzen, J. W., McAtear, J., Butôt, R., & Martínez-Solano, I. (2018). A common toad hybrid zone that runs from the Atlantic to the Mediterranean. *Amphibia-Reptilia*, *in press*. https://doi.org/10.1163/15685381-00003145

Arntzen, J. W., Recuero, E., Canestrelli, D., & Martínez-Solano, I. (2013). How complex is the *Bufo bufo* species group? *Molecular Phylogenetics and Evolution*, *69*, 1203–1208. https://doi.org/10.1016/j.ympev.2013.07.012

Arntzen, J. W., Trujillo, T., Butot, R., Vrieling, K., Schaap, O. D., 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. *Frontiers in Zoology*, *13*, 1–12. https://doi.org/10.1186/s12983-016-0184-7

Arntzen, J. W., & Wallis, G. P. (1991). Restricted gene flow in a moving hybrid zone of the newts *Triturus cristatus* and *T. marmoratus* in western France. *Evolution*, *45*, 805–826. https://doi.org/10.2307/2409691

Arntzen, J. W., & Wallis, G. P. (1999). Geographic variation and taxonomy of crested newts (*Triturus cristatus* superspecies): morphological and mitochondrial DNA data. *Contributions to Zoology*, *68*, 181–203.

Arntzen, J. W., Wilkinson, J. W., Butôt, R., & Martínez-Solano, Í. (2014). A new vertebrate species native to the British Isles: *Bufo spinosus* Daudin, 1803 in Jersey. *Herpetological Journal*, *24*, 209–216.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., … Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, *3*, 1–7. https://doi.org/10.1371/journal.pone.0003376

Baird, S. J. E. (2015). Exploring linkage disequilibrium. *Molecular Ecology Resources*, *15*, 1017–1019.

https://doi.org/10.1111/1755-0998.12424
Baldassarre, D. T., White, T. A., Karubian, J., & Webster, M. S. (2014). Genomic and morphological analysis of a semipermeable avian hybrid zone suggests asymmetrical introgression of a sexual signal. *Evolution*, *68*, 2644–2657. https://doi.org/10.1111/evo.12457
Barton, N. H. (1983). Multilocus clines. *Evolution*, *37*, 454–471. https://doi.org/10.2307/2408260
Barton, N. H. (2013). Does hybridization influence speciation? *Journal of Evolutionary Biology*, *26*, 267–269. https://doi.org/10.1111/jeb.12015
Barton, N. H., & Gale, K. S. (1993). Genetic analysis of hybrid zones. In R. G. Harrison (Ed.), *Hybrid zones and the evolutionary process* (pp. 13–45). New York: Oxford University Press.
Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, *16*, 113–148.
Benestan, L. M., Ferchaud, A.-L., Hohenlohe, P. A., Garner, B. A., Naylor, G. J. P., Baums, I. B., … Luikart, G. (2016). Conservation genomics of natural and managed populations: building a conceptual and practical framework. *Molecular Ecology*, *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. *Molecular Ecology*, *20*, 2044–2072. https://doi.org/10.1111/j.1365-294X.2011.05080.x
Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)*, *30*, 2114–20. https://doi.org/10.1093/bioinformatics/btu170
Bonnet, T., Leblois, R., Rousset, F., & Crochet, P.-A. (2017). A reassessment of explanations for discordant introgressions of mitochondrial and nuclear genomes. *Evolution*, *71*, 2140–2158. https://doi.org/10.1111/evo.13296
Buggs, R. J. A. (2007). Empirical study of hybrid zone movement. *Heredity*, *99*, 301–312. https://doi.org/10.1038/sj.hdy.6800997
Butlin, R. K., & Smadja, C. M. (2017). Coupling, reinforcement, and speciation. *The American Naturalist*, *191*, 000–000. https://doi.org/10.1086/695136
Canestrelli, D., Porretta, D., Lowe, W. H., Bisconti, R., Carere, C., & Nascetti, G. (2016). The tangled evolutionary legacies of range expansion and hybridization. *Trends in Ecology and Evolution*. https://doi.org/10.1016/j.tree.2016.06.010
Carneiro, M., Baird, S. J. E., Afonso, S., Ramirez, E., Tarroso, P., Teotõnio, H., … Ferrand, N. (2013). Steep clines within a highly permeable genome across a hybrid zone between two subspecies of the European rabbit. Molecular Ecology, 22, 2511–2525. https://doi.org/10.1111/mec.12272
Crochet, P.-A., Chen, J. Z., Pons, J., Lebreton, J.-D., Hebert, P. D. N., & Bonhomme, F. (2003). Genetic differentiation at nuclear and mitochondrial loci among large white-headed gulls: sex-biased interspecific gene flow? *Evolution*, *57*, 2865–2878.
Currat, M., & Excoffier, L. (2005). The effect of the Neolithic expansion on European molecular diversity. *Proceedings of the Royal Society B: Biological Sciences*, *272*, 679–688. https://doi.org/10.1098/rspb.2004.2999
Currat, M., Ruedi, M., Petit, R. J., & 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
Daversa, D. R., Muths, E., & Bosch, J. (2012). Terrestrial movement patterns of the common toad (*Bufo bufo*) in Central Spain reveal habitat of conservation importance. *Journal of Herpetology*, *46*, 658–664. https://doi.org/10.1670/11-012
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*, 652–663. https://doi.org/10.1111/1755-0998.12209
Devitt, T. J., Baird, S. J. E., & Moritz, C. (2011). Asymmetric reproductive isolation between terminal forms of the salamander ring species *Ensatina eschscholtzii* revealed by fine-scale genetic analysis of a hybrid zone. *BMC Evolutionary Biology*, *11*. https://doi.org/10.1186/1471-2148-11-245
Ellegren, H. (2000). Microsatellite mutations in the germline: implications for evolutionary inference. *Trends in Genetics*, *16*, 551–558. https://doi.org/10.1016/S0168-9525(00)02139-9
Endler, J. A. (1977). *Geographic variation, speciation, and clines*. (R. M. May, Ed.) (2nd ed.). Princeton, New Jersey: Princeton University Press.
Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, *14*, 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x
Excoffier, L., Foll, M., & Petit, R. J. (2009). Genetic consequences of range expansions. *Annual Review of*

Ecology, Evolution, and Systematics, *40*, 481–501. https://doi.org/10.1146/annurev.ecolsys.39.110707.173414

Fitzpatrick, B. M. (2012). Estimating ancestry and heterozygosity of hybrids using molecular markers. *BMC Evolutionary Biology*, *12*, 131. https://doi.org/10.1186/1471-2148-12-131

Francis, R. M. (2017). POPHELPER: an R package and web app to analyse and visualize population structure. *Molecular Ecology Resources*, *17*, 27–32. https://doi.org/10.1111/1755-0998.12509

Funk, W. C., Tallmon, D. A., & Allendorf, F. W. (1999). Small effective population size in the long-toed salamander. *Molecular Ecology*, *8*, 1633–1640. https://doi.org/10.1046/j.1365-294X.1999.00748.x

Gay, L., Crochet, P. A., Bell, D. A., & Lenormand, T. (2008). Comparing clines on molecular and phenotypic traits in hybrid zones: a window on tension zone models. *Evolution*, *62*, 2789–2806. https://doi.org/10.1111/j.1558-5646.2008.00491.x

Gompert, Z., Mandeville, E. G., & Buerkle, C. A. (2017). Analysis of population genomic data from hybrid zones. *Annual Review of Ecology, Evolution, and Systematics*, *48*, 207–229. https://doi.org/10.1146/annurev-ecolsys-110316-022652

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*, 644–52. https://doi.org/10.1038/nbt.1883

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., … Regev, A. (2013). *De novo* transcript sequence reconstruction from RNA-seq: reference generation and analysis with Trinity. *Nature Protocols*, *8*, 1494–512. https://doi.org/10.1038/nprot.2013.084

Hedrick, P. W. (2013). Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular Ecology*, *22*, 4606–4618. https://doi.org/10.1111/mec.12415

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

Hewitt, G. M. (1988). Hybrid zones - natural laboratories for evolutionary studies. *Trends in Ecology and Evolution*, *3*, 158–167. https://doi.org/10.1016/0169-5347(88)90033-X

Hewitt, G. M. (2011). Quaternary phylogeography: the roots of hybrid zones. *Genetica*, *139*, 617–638. https://doi.org/10.1007/s10709-011-9547-3

Hollander, J., Galindo, J., & Butlin, R. K. (2015). Selection on outlier loci and their association with adaptive phenotypes in *Littorina saxatilis* contact zones. *Journal of Evolutionary Biology*, *28*, 328–337. https://doi.org/10.1111/jeb.12564

Kawakami, T., Butlin, R. K., Adams, M., Paull, D. J., & Cooper, S. J. B. (2009). Genetic analysis of a chromosomal hybrid zone in the Australian morabine grasshoppers (*Vandiemenella*, *viatica* species group). *Evolution*, *63*, 139–152. https://doi.org/10.1111/j.1558-5646.2008.00526.x

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, *15*, 1179–1191. https://doi.org/10.1111/1755-0998.12387

Lande, R. (1980). Genetic variation and phenotypic evolution during allopatric speciation. *The American Naturalist*, *116*, 463–479.

Larson, E. L., Andrés, J. A., Bogdanowicz, S. M., & Harrison, R. G. (2013). Differential introgression in a mosaic hybrid zone reveals candidate barrier genes. *Evolution*, *67*, 3653–3661. https://doi.org/10.1111/evo.12205

Leaché, A. D., Grummer, J. A., Harris, R. B., & Breckheimer, I. (2017). Evidence for concerted movement of nuclear and mitochondrial clines in a lizard hybrid zone, *38*, 42–49. https://doi.org/10.1111/ijlh.12426

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

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology and Evolution*, *20*, 229–237. https://doi.org/10.1016/j.tree.2005.02.010

Mallet, J., Barton, N., Gerardo Lamas, M., Jose Santisteban, C., Manuel Muedas, M., & Eeley, H. (1990). Estimates of selection and gene flow from measures of cline width and linkage disequilibrium in *Heliconius* hybrid zones. *Genetics*, *124*, 921–936.

Mayr, E. (1942). *Systematics and the Origin of Species from the viewpoint of a zoologist*. (L. C. Dunn, H. T. Clarke, M. M. Rhoades, S. R. Detwiler, F. Schrader, & S. Hecht, Eds.) (2nd ed.). New York: Columbia University Press.

Moore, W. S. (1977). An evaluation of narrow hybrid zones in vertebrates. *The Quarterly Review of Biology*, *52*,

263–277. https://doi.org/10.1086/409995
Moran, C. (1981). Genetic demarcation of geographical distribution by hybrid zones. *Proceedings of the Ecological Society Australia*, *11*, 67–73.
Narum, S. R. (2006). Beyond Bonferroni: less conservative analyses for conservation genetics. *Conservation Genetics*, *7*, 783–787. https://doi.org/10.1007/s10592-005-9056-y
Nichols, R. a, & Hewitt, G. M. (1994). The genetic consequences of long-distance dispersal during colonization. *Heredity*, *72*, 312–317. https://doi.org/10.1038/hdy.1994.41
Niedzicka, M., Fijarczyk, A., Dudek, K., Stuglik, M., & Babik, W. (2016). Molecular Inversion Probes for targeted resequencing in non-model organisms. *Scientific Reports*, *6*, 24051. https://doi.org/10.1038/srep24051
Olmo, E. (1973). Quantitative variations in the nuclear DNA and phylogenesis of the amphibia. *Caryologia*, *26*, 43–68. https://doi.org/10.1080/00087114.1973.10796525
Olmo, E., Gargiulo, G., & Morescalchi, A. (1970). Il contenuto di DNA nucleare di alcuni Anfibi. *Bollettina Di Zoologica*, *37*, 513–514.
Phillips, B. L., Baird, S. J. E., & Moritz, C. (2004). When vicars meet: a narrow contact zone between morphologically cryptic phylogeographic lineages of the rainforest skink, *Carlia rubrigularis*. *Evolution*, *58*, 1536. https://doi.org/10.1554/02-498
Pocock, M. J., Hauffe, H. C., & Searle, J. B. (2005). Dispersal in house mice. *Biological Journal of the Linnean Society*, *84*, 565–583. https://doi.org/10.1111/j.1095-8312.2005.00438.x
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, J. K., 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
Raufaste, N., Orth, A., Belkhir, K., Senet, D., Smadja, C., Baird, S. J. E., … Boursot, P. (2005). Inferences of selection and migration in the Danish house mouse hybrid zone. *Biological Journal of the Linnean Society*, *84*, 593–616. https://doi.org/10.1111/j.1095-8312.2005.00457.x
Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., … Westram, A. M. (2017). Interpreting the genomic landscape of speciation: finding barriers to gene flow. *Journal of Evolutionary Biology*, *in press*, 1450–1477. https://doi.org/10.1111/jeb.13047
Recuero, E., Canestrelli, D., Vörös, J., Szabó, K., Poyarkov, N. A., Arntzen, J. W., … Martínez-Solano, I. (2012). Multilocus species tree analyses resolve the radiation of the widespread *Bufo bufo* species group (Anura, Bufonidae). *Molecular Phylogenetics and Evolution*, *62*, 71–86. https://doi.org/10.1016/j.ympev.2011.09.008
Rice, W. R. (1989). Analyzing tables of statistical tests. *Evolution*, *43*, 223–225.
Rice, W. R. (1998). Intergenomic conflict, interlocus antagonistic coevolution and the evolution of reproductive isolation. In D. J. Howard & S. H. Berlocher (Eds.), *Endless Forms Species and Speciation* (pp. 261–270). Oxford University Press.
Rohwer, S., Bermingham, E., & Wood, C. (2001). Plumage and mitochondrial DNA haplotype variation across a moving hybrid zone. *Evolution*, *55*, 405–422. https://doi.org/10.1554/0014-3820(2001)055[0405:PAMDHV]2.0.CO;2
Rosenberg, D. K., Noon, B. R., Megahan, J. W., & Meslow, E. C. (1998). Compensatory behavior of *Ensatina eschscholtzii* in biological corridors: a field experiment. *Canadian Journal of Zoology*, *76*, 117–133. https://doi.org/10.1139/cjz-76-1-117
Rousset, F. (2008). GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, *8*, 103–106. https://doi.org/10.1111/j.1471-8286.2007.01931.x
Roy, J. S., O'Connor, D., & Green, D. M. (2012). Oscillation of an anuran hybrid zone: morphological evidence spanning 50 years. PLoS ONE, 7. https://doi.org/10.1371/journal.pone.0052819
Ryan, S. F., Deines, J. M., Scriber, J. M., Pfrender, M. E., Jones, S. E., Emrich, S. J., & Hellmann, J. J. (2018). Climate-mediated hybrid zone movement revealed with genomics, museum collection, and simulation modeling. Proceedings of the National Academy of Sciences of the United States of America, 201714950. https://doi.org/10.1073/pnas.1714950115
Ryynänen, H. J., & Primmer, C. R. (2006). Single nucleotide polymorphism (SNP) discovery in duplicated genomes: intron-primed exon-crossing (IPEC) as a strategy for avoiding amplification of duplicated loci in Atlantic salmon (*Salmo salar*) and other salmonid fishes. *BMC Genomics*, *7*, 192. https://doi.org/10.1186/1471-2164-7-192
Schmickl, R., Marburger, S., Bray, S., & Yant, L. (2017). Hybrids and horizontal transfer: introgression allows adaptive allele discovery. *Journal of Experimental Botany*, 10.1093/jxb/erx297. https://doi.org/10.1093/jxb/erx297

Scribner, K. T., Arntzen, J. W., & Burke, T. (1997). Effective number of breeding adults in *Bufo bufo* estimated from age-specific variation at minisatellite loci. *Molecular Ecology*, *6*, 701–712. https://doi.org/10.1046/j.1365-294X.1997.00238.x

Sedghifar, A., Brandvain, Y., & Ralph, P. (2016). Beyond clines: lineages and haplotype blocks in hybrid zones. *Molecular Ecology*, *25*, 2559–2576. https://doi.org/http://dx.doi.org/10.1101/043190

Seehausen, O. (2004). Hybridization and adaptive radiation. *Trends in Ecology & Evolution*, *19*, 198–207. https://doi.org/10.1016/j.tree.2004.01.003

Semagn, K., Babu, R., Hearne, S., & Olsen, M. (2014). Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. *Molecular Breeding*, *33*, 1–14. https://doi.org/10.1007/s11032-013-9917-x

Slatkin, M. (1985). Gene flow in natural populations. *Annual Review of Ecology and Systematics*, *16*, 393–430.

Sloan, D. B., Havird, J. C., & Sharbrough, J. (2016). The on-again-off-again relationship between mitochondrial genomes and species boundaries. *Molecular Ecology*, *26*, 2212–2236. https://doi.org/10.1111/mec.13959

Smith, M. A., & Green, D. M. (2005). Dispersal and the metapopulation in amphibian and paradigm ecology are all amphibian conservation: populations metapopulations? *Ecography*, *28*, 110–128. https://doi.org/10.1111/j.0906-7590.2005.04042.x

Staub, N. L., Brown, C. W., & Wake, D. B. (1995). Patterns of growth and movements in a population of *Ensatina eschscholtzii platensis* (Caudata: Plethodontidae) in the Sierra Nevada, California. *Journal of Herpetology*, *29*, 593–599. https://doi.org/10.2307/1564743

Stebbins, R. C. (1954). Natural history of the salamanders of the plethodontid genus *Ensatina*. *University of California Publications in Zoology*, *54*, 47–123.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., … Von Mering, C. (2015). STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research*, *43*, D447–D452. https://doi.org/10.1093/nar/gku1003

Szymura, J. M., & Barton, N. H. (1991). The genetic structure of the hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*: comparisons between transects and between loci. *Evolution*, *45*, 237–261.

Taylor, S. A., White, T. A., Hochachka, W. M., Ferretti, V., Curry, R. L., & Lovette, I. (2014). Climate-mediated movement of an avian hybrid zone. *Current Biology*, *24*, 671–676. https://doi.org/10.1016/j.cub.2014.01.069

Toews, D. P. L., & Brelsford, A. (2012). The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, *21*, 3907–3930. https://doi.org/10.1111/j.1365-294X.2012.05664.x

Trochet, A., Moulherat, S., Calvez, O., Stevens, V., Clobert, J., & Schmeller, D. (2014). A database of life-history traits of European amphibians. *Biodiversity Data Journal*, *2*, e4123. https://doi.org/10.3897/BDJ.2.e4123

Trujillo, T., Gutiérrez-Rodríguez, J., Arntzen, J. W., & Martínez-Solano, I. (2017). Morphological and molecular data to describe a hybrid population of the common toad (*Bufo bufo*) and the spined toad (*Bufo spinosus*) in western France. *Contributions to Zoology*, *86*, 1–10.

Vences, M., & Wake, D. B. (2007). Speciation, species boundaries and phylogeography of amphibians. In H. Heatwole & M. Tyler (Eds.), *Amphibian Biology, vol.6, Systematics* (pp. 2613–2660). Chipping Norton, Australia: Surrey Beatty & Sons.

Vines, T. H., Dalziel, A. C., Albert, A. Y. K., Veen, T., Schulte, P. M., & Schluter, D. (2016). Cline coupling and uncoupling in a stickleback hybrid zone. *Evolution*, *70*, 1023–1038. https://doi.org/10.1111/evo.12917

Vinogradov, A. E. (1998). Genome size and GC-percent in vertebrates as determined by flow cytometry: the triangular relationship. *Cytometry*, *31*, 100–109. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9482279

Wang, L., Luzynski, K., Pool, J. E., Janoušek, V., Dufkovú, P., Vyskočilovú, M. M., … Tucker, P. K. (2011). Measures of linkage disequilibrium among neighbouring SNPs indicate asymmetries across the house mouse hybrid zone. *Molecular Ecology*, *20*, 2985–3000. https://doi.org/10.1111/j.1365-294X.2011.05148.x

Wang, S., Sha, Z., Sonstegard, T. S., Liu, H., Xu, P., Somridhivej, B., … Liu, Z. (2008). Quality assessment parameters for EST-derived SNPs from catfish. *BMC Genomics*, *9*, 450. https://doi.org/10.1186/1471-2164-9-450

While, G. M., Michaelides, S., Heathcote, R. J. P., MacGregor, H. E. A., Zajac, N., Beninde, J., … Uller, T. (2015). Sexual selection drives asymmetric introgression in wall lizards. *Ecology Letters*, *18*, 1366–1375. https://doi.org/10.1111/ele.12531

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., Burke, T., Butlin, R. K., & Arntzen, J. W. (2017). A signature of dynamic biogeography: enclaves indicate past species replacement. *Proceedings of the Royal Society Biological Sciences*, *284*, 1–6.
Wielstra, B., Burke, T., Butlin, R. K., Avcı, A., Üzüm, N., Bozkurt, E., … Arntzen, J. W. (2017). A genomic footprint of hybrid zone movement in crested newts. *Evolution Letters*, 1–9. https://doi.org/10.1002/evl3.9
Wu, C.-I., & Ting, C.-T. (2004). Genes and speciation. *Nature Reviews Genetics*, *5*, 114–122. https://doi.org/10.1038/nrg1269
Zeisset, I., & Beebee, T. J. C. (2003). Population genetics of a successful invader: the marsh frog *Rana ridibunda* in Britain. *Molecular Ecology*, *12*, 639–646. https://doi.org/10.1046/j.1365-294X.2003.01775.x

Data Accessibility

Supplemental text contains details of the SNP design:

MEC_Supplemental_text_vanRiemsdijk _201x.pdf

Supplemental tables are available in:

MEC_Supplemental_tables_vanRiemsdijk _201x.xlsx

Supplemental figures are available in:

MEC_Supplemental_figures_vanRiemsdijk _201x.pdf

In- and output of analyses, and custom R scripts are freely available online:

Dryad link (sponsored by MEC): doi:10.5061/dryad.rg18034

Author contributions

IvR, BW, and JWA designed the study and collected data. IvR and RB analysed the data. IvR wrote the manuscript, with input from all co-authors.

Tables and Figures

Table 1: Overview table with the number of individuals sampled per population, the distance in kilometres (km), and the mean hybrid index, pooled for the reference populations (1-4 for B. bufo and 23-29 for B. spinosus), and for each transect population (5-22).

Figure 1: Schematic representation of an hypothesis of hybrid zone movement. In contrast to a stable hybrid zone (a) where the peak of admixture linkage disequilibrium is situated in the hybrid zone centre and clines are symmetric, a moving hybrid zone (b) is expected to have a peak of admixture linkage disequilibrium shifted ahead of the hybrid zone and a tail of introgression in the wake of the hybrid zone (t_0 is the previous hybrid zone position, t_1 the current position). Dashed lines indicate the hybrid zone centre.

Figure 2: Map of Western Europe (a) with reference populations to determine diagnostic nature of markers for *Bufo bufo* (blue circles) and *B. spinosus* (red circles), and transect populations (grey circles). Panel b details the transect orientation (dashed line) and the position of hybrid zone populations. Panel (c) shows bar plots with individual admixture proportion (Structure *Q* scores) on the left and mtDNA allele frequency per population on the right. The grey bar refers to one missing mtDNA data point.

Figure 3: Best fitting clines for 30 nuclear markers and the mtDNA marker determined with HZAR. Frequencies of 0 and 1 represent pure *Bufo bufo* and *B. spinosus* genotypes. Left tail clines are in panel (a), right tail clines are in (b), both tail clines in (c), mirror tail clines in (d), and no tail clines are in (e). Panel f shows the mtDNA cline with a significant right tail and the nuclear marker *banp* with a significantly displaced cline centre and a non-significant left tail. Cline models that fit significantly better than the next best model (ΔAICc > 2) are shown by green lines and the others by grey lines. Sample localities on the transect are shown on the x axis by inside ticks and the arrow (^) indicates the position of the hybrid zone centre based on the admixture proportion cline (dashed line).

Figure 4: Admixture linkage disequilibrium (*D'*) recorded over a transect from the north (*Bufo bufo*, populations 1-4) to the south (*B. spinosus*, populations 23-29). Note that the horizontal axis does not represent linear geographical distance. The dotted vertical line is the position of the hybrid zone centre inferred from the admixture proportion cline. Estimates for *D'* are obtained by bootstrap analysis and shown by box-and-whisker plots.

Figure 5: Schematic representation of inferred hybrid zone movement in (a) the gulls *Larus glaucescens* and *L. accidentalis* (Gay et al., 2008), and the house mice *Mus musculus musculus* and *M. m. domesticus* (Wang et al., 2011), and asymmetric reproductive isolation in (b) the salamanders *Ensatina eschscholtzii eschscholtzii* and *E. e. klauberi* (Devitt et al., 2011), and possibly also in the toads *Bufo bufo* and *B. spinosus* in the present study. In panel b, the tick mark shows where reproduction can take place and the cross shows where reproduction cannot take place. Cardinal directions are given by N, E, S and W. The question mark in panel b refers to the uncertainty of unidirectional introgression and of the magnitude and position of the peak of *D*' in the *Bufo* hybrid zone.

Population	Individuals sampled	Distance (*km*)	Hybrid index (HI)	
1	10	0	0.03	
2	10			
3	7			
4	6			
5	20	174	0.06	
6	18	335	0.03	
7	12	468	0.16	
8	10	469	0.17	
9	20	508	0.28	
10	10	530	0.38	
11	18	530	0.39	
12	20	541	0.50	
13	20	555	0.74	
14	20	561	0.79	
15	19	571	0.86	
16	12	648	0.96	
17	8	644	0.96	
18	10	689	0.97	
19	8	680	0.98	
20	8	695	0.96	
21	10	705	0.99	
22	18	791	1.00	
23	2	1222	1.00	
24	2			
25	2			
26	1			
27	2			
28	1			
29	2