**Discordant patterns of introgression across a narrow hybrid zone between two
cryptic lineages of an Iberian endemic newt**

Fernando Sequeira^{*1}, Adam Bessa-Silva⁴, Pedro Tarroso¹, Tiago Sousa-Neves^{1,2}, Marcelo
Vallinoto^{1,4}, Helena Gonçalves^{1,3}, Iñigo Martínez-Solano⁵

¹CIBIO-InBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO,
Laboratório Associado, Campus Agrário de Vairão, Universidade do Porto, 4485-661 Vairão,
Portugal

²Museu Paraense Emílio Goeldi, Coordenação de Zoologia, 399, 66040-170 Belém, Pará,
Brazil.

³Museu de História Natural e da Ciência, Universidade do Porto, Praça Gomes Teixeira, 4099-
002 Porto, Portugal

⁴Laboratório de Evolução (LEVO), Instituto de Estudos Costeiros (IECOS), Universidade
Federal do Pará, Campus de Bragança, 68600-000 Pará, Brasil

⁵Department of Biodiversity and Evolutionary Biology, Museo Nacional de Ciencias Naturales,
CSIC (MNCN-CSIC), Madrid, Spain

* Corresponding Author:

Fernando Sequeira, CIBIO-InBIO, Centro de Investigação em Biodiversidade e Recursos
Genéticos, InBIO, Laboratório Associado, Campus Agrário de Vairão, Universidade do Porto,
4485-661 Vairão. Portugal
fsequeira@cibio.up.pt

Short running head: Genetic structure of an Iberian endemic newt hybrid zone

ABSTRACT

The study of natural hybrid zones can illuminate aspects of lineage divergence and speciation in morphologically cryptic taxa. We studied a hybrid zone between two highly divergent but morphologically similar lineages (Southwestern and Southeastern) of the Iberian endemic Bosca's newt (*Lissotriton boscai*) in SW Iberia with a multilocus dataset (microsatellites, nuclear and mitochondrial genes). STRUCTURE and NEWHYBRIDS analyses retrieved few admixed individuals, which classified as backcrosses involving parental individuals of the Southwestern lineage. Our results show asymmetric introgression of mtDNA beyond the contact from this lineage into the Southeastern lineage. Analysis of non-geographic introgression patterns revealed asymmetries in the direction of introgression, but except for mtDNA, we did not find evidence for non-concordant introgression patterns across nuclear loci. Analysis of a 150-km transect across the hybrid zone showed broadly coincident cline widths (*ca.* 3.2-27.9 km wide), and concordant cline centres across all markers, except for mtDNA that is displaced *ca.* 60 km northward. Results from ecological niche modelling show that the hybrid zone is in a climatically homogenous area with suitable habitat for the species, suggesting that contact between the two lineages is unlikely to occur further south as their distributions are currently separated by an extensive area of unfavourable habitat. Taken together, our results suggest the genetic structure of this hybrid zone results from the interplay of historical (biogeographic) and population-level processes. The narrowness and coincidence of genetic clines can be explained by weak selection against hybrids and reflect a degree of reproductive isolation that is consistent with cryptic speciation.

Keywords: Contact zone, introgression, cline analysis, cryptic speciation, amphibians

INTRODUCTION

The study of natural hybrid zones has become an important avenue of research to understand the process of speciation. Although natural hybridization may occur in many different spatial and temporal contexts, hybrid zones originated by secondary contact between lineages that

previously diverged in allopatry, are among the most frequent and extensively studied systems (Hewitt, 1988; Barton & Hewitt, 1989; Abbot *et al.*, 2013). Hybridization after secondary contact can produce contrasting outcomes depending on the extent at which genetic differentiation and/or reproductive barriers (pre- and/or post-zygotic) accumulate during isolation. Theory predicts that hybridization at secondary contact can result in reduction of differentiation or even fusion of gene pools due to the rapid erosion of barriers to gene exchange, or, alternatively, in full reproductive isolation due to an increase in the strength of barriers that prevent gene exchange (Barton & Hewitt, 1985; Wu, 2001; Seehausen, 2006). An increasing number of studies have identified many stable, long-lasting natural hybrid zones, suggesting that varying degrees of reproductive isolation can be maintained despite the potential homogenizing effects of gene flow (Moore, 1977; Barton & Hewitt, 1985; Hewitt, 1988; Abbot *et al.*, 2013).

Assessing the potential outcomes of hybridization at secondary contact zones is relevant from evolutionary, taxonomic, and conservation perspectives, especially in the study of morphologically cryptic, genetically divergent lineages (Sequeira *et al.*, 2005; Singhal & Moritz, 2013, Gonçalves *et al.*, 2015). Detailed studies of secondary contact zones between morphologically cryptic lineages have shown that the nature and extent of reproductive isolation varies across species, ranging from substantial pre- and post-zygotic isolation (Philips *et al.*, 2004; Alexandrino *et al.*, 2005; Leache & Cole, 2007; Kindler *et al.*, 2017) to genetic merger with negligible barriers to the neutral diffusion of alleles (Wake & Schneider 1998; Sequeira *et al.*, 2005; Singhal & Moritz, 2013). Describing the structure of hybrid zones, namely the extent and nature of reproductive barriers, is key for predicting the evolutionary consequences of secondary contact (persistent or stable, replacement of one lineage by the other, fusion or further isolation), and shed light on the evolutionary processes influencing the spatial and temporal dynamics of these zones (Stuart-Fox *et al.*, 2009; Singhal & Moritz, 2012, 2013).

Among the several models that have been proposed to explain the structure and maintenance of hybrid zones (e.g. Moore 1977; Barton & Hewitt 1985, 1989; Barton & Gale, 1993), the tension zone and geographical selection-gradient models are those that have received more

support from empirical studies (Kruuk *et al.*, 1999; Alexandrino *et al.*, 2005; Singhal & Moritz, 2013). Both models are based on the analysis of allele frequency variation as a function of geographic distance along a transect spanning the hybrid zone (clines), assuming that shape (centre and width) of the curve at each individual locus is determined by a balance between dispersal from parental populations into the hybrid zone and the strength of selection acting on hybrids. These two models are mathematically similar, but differ in the type of selection acting against hybrids, which can be exogenous or environment-dependent (i.e. adaptive divergence), or endogenous or environment-independent (e.g. heterozygote breakdown) (Endler, 1977; Barton & Hewitt, 1985, 1989; Barton & Gale, 1993; Moore & Price, 1993; Kruuk *et al.*, 1999).

According to the predictions of these models, coincidence of cline centres and concordance of cline widths across loci are expected when a hybrid zone first forms after secondary contact. After several generations of hybridization with continued dispersal into the zone, coincidence and concordance among clines is expected if selection operating against hybrids is strong (Barton & Hewitt, 1985; Barton & Gale, 1993). In this case, the low fitness of hybrids causes them to be removed by selection, which reduces the extent of spatial mixing between parental lineages, resulting in a narrow hybrid zone. However, if selection is either negligible or acts only weakly against hybrids, each cline may exhibit different patterns of change in allele frequencies across the hybrid zone, as a result of different responses to locus-specific selection and/or drift, thus originating patterns of noncoincidence and nonconcordance of clines (Barton & Bengtsson, 1986; Harrison, 1990; Barton & Gale, 1993).

Studies of secondary contact zones using large numbers of markers have shown that introgression rate and its geographic extent can be highly variable among markers (Wu 2001; Gompert & Buerkle, 2009; Carneiro *et al.*, 2013; Harrison & Larson, 2016). This happens because many hybrid zones are porous boundaries that act as selective filters, allowing gene exchange for neutral or adaptive traits, whereas introgression of genes/genomic regions underlying local adaptation or resulting in hybrid incompatibilities are limited or impeded to cross boundaries (Barton & Hewitt, 1985). One pattern that has been frequently reported in hybrid zones is that of differential introgression of mtDNA relative to nuclear DNA (García-

104 Paris *et al.*, 2003; Sequeira *et al.*, 2005; Chatfield *et al.*, 2010; but see Toews & Brelsford,
105 2012). Cytonuclear discordances have been explained by a variety of evolutionary mechanisms,
106 including sex-biased life history differences, such as dispersal and mating preference, adaptive
107 introgression of mtDNA, or demographic processes (e.g. Funk & Omland, 2003; Toews &
108 Brelsford, 2012). Furthermore, although once thought to be rare, many recent studies suggest
109 that patterns of differential introgression among markers, including cytonuclear discordance,
110 may also result from hybrid zone movements caused by habitat/environmental changes through
111 time (Rohwer *et al.*, 2001; Buggs, 2007; Taylor *et al.*, 2015; Wielstra *et al.*, 2017). Climate and
112 concomitant habitat changes can modify the geographic range of species/lineages and their
113 demographic dynamics, affecting the degree to which hybridizing lineages interact (Currat *et*
114 *al.*, 2008; Ryan *et al.*, 2018). While inference of hybrid zone movement from genetic evidence
115 alone is very controversial (Buggs, 2007), it is well documented in several Holarctic organisms,
116 namely those distributed in southern refugial areas (Hewitt, 2000; Weiss & Ferrand, 2007),
117 which have experienced repeated episodes of population size changes (contraction and
118 expansion) related to decreased/increased levels of gene flow mediated by population
119 movements (Hofreiter & Stewart, 2009).

120 Here, we focused on a secondary contact zone between two morphologically cryptic
121 lineages of Bosca's newt, *Lissotriton boscai*. This species occurs throughout most of the
122 western half of the Iberian Peninsula, exhibiting deep levels of spatially structured genetic
123 variation (Martínez-Solano *et al.*, 2006; Teixeira *et al.*, 2015). In a recent work, Teixeira *et al.*
124 (2015) found notable congruence between nuclear and mtDNA datasets in uncovering deep and
125 consistent differentiation between two major lineages: 1) lineage A, distributed across most of
126 the species range; and 2) lineage B, restricted to the southwestern coastal region of Portugal.
127 According to published temporal estimates (Martínez-Solano *et al.*, 2006; Teixeira *et al.*, 2015),
128 the two major lineages of *L. boscai* would have originated during the Late Miocene (about 6 – 9
129 Myr.), with subsequent splits into differentiated population groups around the Pliocene-
130 Pleistocene boundary. These differentiated population groups have survived in different
131 allopatric refugia during the Quaternary ice ages, expanding their ranges from these refugia

once the climate ameliorated. Based on the spatial distribution of mtDNA and nuclear haplotypes, Martínez-Solano *et al.* (2006) and Teixeira *et al.* (2015) inferred the occurrence of several putative contact zones, including signs of nuclear DNA admixture between lineages A and B. Interestingly, the boundaries between these lineages correspond roughly to geographic areas where the species is either less abundant or even absent, suggesting that environmental/habitat conditions may have played an important role in constraining gene flow, perhaps contributing to reproductive isolation (Díaz-Paniagua, 2002; Teixeira, 2008, Teixeira *et al.*, 2015).

Discerning whether lineages A and B in *L. boscai* represent evolutionary ephemera or are under independent evolutionary trajectories and are unlikely to merge when meeting in secondary contact (thus representing incipient species) requires detailed assessment of the genetic cohesiveness of interacting lineages and the location, extension, nature and strength of reproductive barriers. Here we characterize a contact zone between lineages A and B in the south part of the range of *L. boscai* using multiple molecular markers (10 microsatellite loci, one mitochondrial and two single copy nuclear markers) in a cline analysis framework. We analysed patterns of introgression across markers along a 150 km-transect spanning the hybrid zone, using both geographic and genomic cline-fitting approaches, and inferred the strength of selection against hybrids by testing for the coincidence of cline centres and the concordance of cline widths. Finally, we investigated whether environmental factors might influence the position and structure of the hybrid zone using ecological niche modelling (ENM) to predict the geographic distribution of climatically suitable habitats for the species.

MATERIAL AND METHODS

Sampling and molecular markers

We gathered genetic data for 221 individuals from 28 localities throughout the species range, including 35 samples from 15 localities used in earlier studies (Martínez-Solano *et al.*, 2006; Teixeira *et al.*, 2015) (Fig. 1, Table 1, Table S1 and Table S2). Most samples (209 individuals, 23 sites) were collected in the southern part of the range, in the contact zone between lineages B

(hereafter LBSW), and A (hereafter LBSE). Sample sizes ranged from 1 to 26. Newts were collected between 2002 and 2012, but only once at each site. Tissue samples consisted of tail tips of larvae and either toe or tail clips of adults, which were released in the wild after tissue collection.

To characterize the genetic structure of the contact zone, we used different types of markers, including: i) 10 microsatellites specifically developed for *L. boscai* (Sequeira *et al.*, 2012; Peñalver-Alcázar *et al.*, 2017); ii) one mitochondrial gene (NADH dehydrogenase subunit 4, *ND4*); and iii) two single copy nuclear genes (Beta-fibrinogen intron 7 and incomplete exon 8, *β -fibint7*; and the Growth Hormone intron 4, *GH*).

For both mtDNA and nuclear genes, diagnostic restriction endonuclease (RE) assays were designed to distinguish between the LBSW and LBSE lineages through RFLP (Restriction fragment length polymorphism). For mtDNA and *β -fibint7* we used available sequence data from previous phylogeographic studies (Martínez-Solano *et al.*, 2006; Teixeira *et al.*, 2015), whereas for the nuclear gene GH we generated new sequences for this study (GenBank accession numbers: MK906820-MK906873; Table S1), from a subset of 28 individuals from different localities (including some far from the hybrid zone; Fig. 1; Table 1; Table S1; Fig. S1) across the species range. Description of diagnostic restriction endonuclease (RE) assays and detailed laboratory procedures are described in Supporting Information, and detailed sample size used for each molecular marker are described in Table S1.

We used 10 polymorphic microsatellite loci as described in Sequeira *et al.* (2012) (Ltb2, Ltb4, Ltb9, Ltb10, Ltb11, Ltb12, Ltb18, and Ltb20), and Peñalver-Alcázar *et al.* (2017) (Ltb31 and Ltb37) to genotype 209 individuals, distributed across 23 sites (Table S1 and Table S2). Measures of genetic diversity based on microsatellites, including unbiased estimates of expected (H_E) and observed (H_O) heterozygosities and number of alleles (N_a), were calculated with FSTAT v. 2.9.3 (Goudet, 2001) per locus and population. We also calculated pairwise F_{ST} (populations with $N \geq 5$) and F_{IS} and tested for deviations from Hardy-Weinberg equilibrium and for linkage disequilibrium in FSTAT. Significance levels were adjusted for multiple tests using sequential Bonferroni corrections ($\alpha = 0.05$; Rice, 1989). Pairwise linkage disequilibrium (LD),

measured by calculating the correlation coefficient between pairs of loci (R), was assessed at each site for each locus over the transect across the contact zone using the R package GENETICS v. 1.3.8. Confidence intervals were estimated by 1000 bootstrap iterations. To estimate LD independent of within-species variation, we reduced each microsatellite locus to a bi-allelic system using INTROGRESS, as described below.

Admixture analyses

Population structure and genetic admixture analyses were investigated without *a priori* assignment of individuals to populations using the software STRUCTURE v. 2.3.3 (Pritchard *et al.*, 2000). We ran STRUCTURE based on microsatellite loci (total 10 loci) and on a data set combining microsatellites and the two nuclear diagnostic markers (12 loci). Both STRUCTURE analyses were performed under the admixture model with ten independent runs for each K ranging from 1 to 10, using 500,000 MCMC iterations and discarding the first 50,000 as burn-in. To infer the number of groups (K) which best explains the structure in our data we applied the *ad hoc* statistic ΔK (Evanno *et al.*, 2005), as implemented in the pipeline CLUMPAK (Kopelman *et al.*, 2015). We also used the Bayesian clustering software NEWHYBRIDS v.1.1 (Anderson & Thompson, 2002) to estimate the posterior probability of each individual to be assigned into each of the six defined genotypic categories (parental, F1-hybrids, F2-hybrids and backcrosses in either direction), under default conditions. We used a posterior probability of assignment (Q) threshold of > 0.75 to be considered as part of a category (parental, F1-hybrids, F2-hybrids, or backcross) in NEWHYBRIDS, and 0.90 in STRUCTURE ($Q > 0.90$, parental; $0.10 < Q < 0.90$, mixed ancestry).

Genomic and geographic cline analysis

To reduce the complexity of multi-allelic variation in microsatellite loci, we first converted each locus to a two-allele system following the allele frequency differentials (δ) approach implemented in R package INTROGRESS (Gompert & Buerkle, 2009). As this procedure requires *a priori* definition of pure individuals of each parental lineage, both STRUCTURE and NEWHYBRIDS results were used to identify non-admixed individuals (Fig. 2). Following this

strategy, we decided to use populations 6 and 20 as representatives of parental lineages LBSW and LBSE, respectively (Fig. 1; Table S1). Then, we assessed differential introgression patterns by analyzing the relationship between individual loci (hybrid index for a focal locus) against the multilocus expectation (mean hybrid index across all loci) using the logit-logistic model of Fitzpatrick (Fitzpatrick, 2013), and Barton's concordance analysis (Szymura & Barton, 1986). Both models estimate two parameters; the first, μ or α (in Fitzpatrick and Barton's models, respectively) describes a shift of the cline position at a specific locus towards either the LBSW or the LBSE genomic background (positive values indicate a shift of the cline toward LBSW and negative values indicate a shift toward LBSE). The second parameter, v or β (Fitzpatrick/Barton), describes the relative abruptness of the slope for a focal locus relative to mean hybrid index (values >1 indicate an abrupt slope and stronger selection whereas values <1 indicate a more gradual slope and weaker selection). Model parameters were estimated using the R package H1est (CRAN; <http://cran.r-project.org/>).

Geographical cline analyses were performed using genetic data (all 13 loci) from eight populations ($N=119$) along a transect of approximately 150 km from Vila Nova de Mil Fontes (site 6) to Azaruja (site 15) (Fig. 1). As for genomic clines, geographical clines were fit for each microsatellite locus based on allelic classes (as calculated using INTROGRESS, see above for details). The transect was chosen based on the results of Bayesian assignment methods and haplotype genealogies of nuclear markers (see results below), and for two additional reasons: 1) signs of introgression were reported in previous phylogeographic studies (Martínez-Solano *et al.*, 2006; Teixeira *et al.*, 2015); and 2) populations in this region are continuously distributed, which contrasts with the disjunct distribution of LBSW and LBSE populations in the southern part of the species range (Fig. 1). Accordingly, populations away from this transect were excluded from the analysis for purposes of cline fitting because clines cannot be fitted in a transect representing a discontinuous dispersal route. Throughout the selected transect only locations with $N \geq 5$ (6, 9, 10, 12, 13, 14, 16, and 17) were analyzed (Table 1; Fig. 1). These eight localities were collapsed into a one-dimensional transect, and for ease of representation, distance (km) along the fitted transect was measured from locality 6 using Google Earth

(<https://earth.google.com>) (Fig. 1). Maximum likelihood sigmoid clines (Szymura & Barton, 1986, 1991; Barton & Gale, 1993) were fitted independently based on allele frequencies for each locus in each locality using the R package HZAR (Derryberry *et al.*, 2014) in R v. 3.1.1 (R Core Team, 2018). For each locus, we tested different models, all of which estimated cline center (c) and width (w), but varied in how the parameters $p_{\max}$ and $p_{\min}$ (maximum and minimum values of allelic frequencies in the tail ends of the cline) were fitted: model I, fixed to 0 and 1 (i.e., with $p_{\min}=0$ and $p_{\max}=1$); model II, $p_{\min}/p_{\max}$ set to observed values; and model III, $p_{\min}/p_{\max}$ set to estimated values (Szymura & Barton 1986; 1991). The best-fit model for each of the 13 loci was selected based on Akaike's information criterion scores corrected for small sample sizes (AICc) using the automated model selection function implemented in HZAR, and considered the model with the lowest AICc score as the best-fitting model.

We tested for cline coincidence and concordance by forcing each locus to have a common centre or width. Since STRUCTURE's Q -score is based on multilocus genotype information, it was chosen as representing a conservative genome-wide average estimate for cline centre and width (Wielstra *et al.*, 2017; Arntzen *et al.*, 2016). Accordingly, we fitted geographic clines based on STRUCTURE Q -scores (run with $K = 2$) for the total dataset (10 STRs plus 2 nuclear SNPs; *Structure* Q_{total}), and only for STRs (*Structure* Q_{STRs}) to estimate reference values for cline centre and width, which were then used to refit each locus cline using those reference values (constrained models). Statistical significance between constrained and free models were estimated using likelihood ratio tests, under the assumption that twice the difference in log likelihood ($LRT = -2\Delta LL$) between two models asymptotically follows a χ^2 distribution with the degrees of freedom being the difference in the number of parameters.

Environmental and ecological niche-based analyses

To explore whether environmental factors might influence the structure of the hybrid zone, we used ecological niche modelling (ENM) to predict the geographic distribution of climatically suitable habitats for *L. boscai*, and then examined the significance of correlations between a measure of genetic differentiation (F_{ST}) and both the geographic (Euclidean) and ecological

(resistance) distances. To this purpose, we initially downloaded 19 available bioclimatic variables from WorldClim (www.worldclim.org; Hijmans *et al.*, 2005) for the current climate with a 0.00833° spatial resolution. All bioclimatic variables were clipped to the extent of the study area, comprising the Iberian Peninsula, and upscaled to the study's spatial resolution (0.09° ≈ 10Km). As a surrogate for habitat, we used the Normalized Difference Vegetation Index (NDVI). The monthly NDVI (Didan, 2018) was downloaded through the Earth data Search at NASA's Land Processes Distributed Active Archive Center for the period between 2000 and 2015 with 1km spatial resolution. The NDVI time series was summarized with the median, maximum and standard deviation, clipped to the extent of the study area and upscaled to the study resolution. All variables (habitat and climate) were checked for pairwise correlations only retaining those that presented a Pearson correlation < 0.7.

Occurrence data for *L. boscai* were obtained for the study area from: 1) the Spanish (Pleguezuelos *et al.*, 2004) and Portuguese (Loureiro *et al.*, 2008) atlases of amphibians and reptiles, and 2) downloaded from the Global Biodiversity Information Facility (GBIF). The atlas data were converted to the study area grid with 10km spatial resolution. Presence in the study grid was only considered if the cell was covered by, at least, 50% of the UTM atlas presences. The data from GBIF (n=2208) were filtered by quality, with criteria including the presence of geographic coordinates, the precision of the location (locations with lower precision than the study resolution were discarded), and duplication of locations. The final data set included 1864 unique occurrence locations.

We built two sets of distance matrices (geographic and resistance) for eight populations with measured genetic differentiation. Geographic distances were calculated as simple Euclidean distances between the geographic coordinates of population locations. For the resistance distance matrix, we built an ecological niche model with the Maximum Entropy algorithm using software MaxEnt v.3.4.1 (Phillips *et al.*, 2017). We run MaxEnt for 50 replicates with cross-validate subsampling, with the remaining options set to the default values. The average model from the replicates was used as consensus model for resistance calculation with the package gdistance v.1.1-1 (van Etten, 2017) in R v.3.4.4 (R Core Team, 2018). The

model was used as a conductance layer from which we derived the resistance distance matrix for the populations with genetic information. We used the 'commuteDistance' function with the mean as transition function for eight neighboring cells. The transition matrix was corrected for the North-South direction with the function 'geoCorrection' due to the use of geographic coordinates. Model fitness was quantified by calculation of the area under the receiver operating characteristic curve (AUC).

We used a regression-based method to quantify the relationships of geographic and resistance distances with genetic structure. Using regression with matrices related to isolation models allows testing competing hypotheses underlying the isolation mechanism and quantifying them through the coefficients obtained in the regression equation (Shirk *et al.*, 2018). However, regression-based methods assume independence of the samples, which is often violated when using variables like distance matrices (Shirk *et al.*, 2018). We used a generalized least squares (GLS) regression with maximum-likelihood population effects (MLPE) covariance structure to account for the non-independence inherent to pairwise distance matrices (Van Strien *et al.*, 2012). The covariance structure from the MLPE identifies every two pairwise distances sharing a common location and allows controlling for the non-independence of the distances in the GLS regression (Clarke *et al.*, 2002).

RESULTS

Patterns of admixture

Analysis of the distribution patterns of LBSW and LBSE haplotypes based on RFLPs showed no overlap of the two mtDNA lineages in any of the 23 sites examined. All the southernmost sites from the hills of Serra de Monchique (south) and western populations to the Évora region were fixed for LBSW haplotypes, whereas sites at the northern edge, east and southeast were fixed for LBSE haplotypes (Figs. 1 and 2, Table 1). In general, the two nuclear diagnostic markers exhibited similar patterns of geographical distribution of LBSW and LBSE haplotypes, with LBSW haplotypes in the west and southwest, and LBSE haplotypes in the southeastern and eastern parts of the study area (Fig. 2; Table 1). However, contrary to mtDNA, both diagnostic

nuclear markers revealed mixed populations with the occurrence of intermediate haplotype frequencies in several localities, mainly in sites 7-17 (Fig. 2; Table 1).

Microsatellites were highly polymorphic with a mean number of alleles per locus of 27.8, ranging from 17 (Ltb10) to 38 (Ltb9) alleles. Most populations harbored high levels of genetic diversity (H_E), ranging from a minimum of 0.722 (site 12) to a maximum of 0.841 (site 15), while observed heterozygosity (H_O) ranged from 0.475 (site 2) to 0.838 (site 19) (Table S2). Only at four localities out of 13 we found loci out of Hardy-Weinberg equilibrium (Table S2). Significant linkage disequilibrium (LD) was detected in the combination Ltb18/Ltb2 in populations 10 and 12, but we found no evidence of linkage disequilibrium in other populations. Average estimates of inbreeding coefficients (F_{IS}) and pairwise linkage disequilibrium (LD), measured as correlation coefficient between pairs of loci (r), did not vary significantly across sites along the 150 km transect (Fig. S2). Population pairwise F_{ST} ranged from 0 to 0.210, with an overall $F_{ST} = 0.114$ between LBSE and LBSW populations (Table S3).

The Bayesian clustering analysis in STRUCTURE, under the admixture model, using either a dataset composed of only microsatellite genotypes (STRUCTURE Q_{STRs}) or a full dataset combining microsatellites and intron nuclear markers (full dataset; STRUCTURE Q_{total}) retrieved a best estimated $K = 2$ (Fig. 2; Fig. S3 and Fig. S4, for $L(K)$ and ΔK). STRUCTURE analysis showed that 96% of individuals were classified with a posterior probability $> 90\%$ as belonging to one of the two parental lineages. Most potentially admixed individuals were detected in population 10, which is the only one with an average Q -score $< 90\%$ for STRUCTURE Q_{STRs} (Fig. 2, Table 1). Interestingly, in the STRUCTURE Q_{total} analysis we did not recover populations with an average Q -score $< 90\%$ as belonging to one of the two parental lineages, and we only found admixed individuals in locality 10, suggesting that the combination of microsatellites and nuclear SNPs works well in delimiting the two lineages, but not in detecting admixed individuals.

NEWHYBRIDS analysis produced similar results, with very few individuals with mixed genome ancestry. Using a posterior probability cutoff > 0.75 , only $\approx 5\%$ (11-12) of the individuals were classified as hybrids (F2 or backcrosses). As in the STRUCTURE analysis,

most admixed individuals were detected in site 10 (Fig. 2). While the combined analysis of microsatellites and nuclear introns (full dataset) generated similar results to the analysis based only on microsatellite data, slight differences were observed regarding the hybrid class. The analysis with the full dataset only retrieved backcrosses (all in site 10), whereas the microsatellite-based analysis allowed the detection of three F2 hybrids (sites 10 and 14) (Fig. 2; Fig. S3). All of the 11-12 hybrid individuals detected possessed LBSW mtDNA, and all backcrosses resulted from crosses involving hybrids and pure LBSW individuals (Fig. 2).

Geographic and non-geographic analyses of introgression

Allele frequencies across the sampling area for each locus are shown in Table 1 and Fig. 2. Despite the observed variation among loci in the spatial distribution of allelic frequencies, all of them present a clinal pattern of variation (Fig. 3). There was considerable variation regarding the best-fitting model for each locus, but the most common model, chosen for 9 out of 13 loci, was model II ($p_{\min}/p_{\max}$ set to observed values without tail fitting) (Table 2). Maximum-likelihood centre estimates were highly variable among loci, ranging between ≈ 50 and 130 km (95% CI: 44.1-141 Km). Based on MLE estimates of cline centre for “multilocus” *Structure* Q_{Total} ($c = 63.9$; 95% CI = 62.0-83.9 km (Table 2; Fig. 3) as a proxy of the consensus cline centre, constraining all clines to a common centre resulted in a significant decrease in likelihood ($LRT_{\text{same-diff}} = 125.3$, $df=12$, $P < 0.05$). Considering only nuclear markers, 10 loci (7 microsatellites and 2 nuclear diagnostic markers) could be constrained to share this centre without a significant decrease in likelihood ($LRT_{\text{same-diff}} = 6.66$, $df=9$, $P > 0.05$; outliers: Ltb37, Ltb31 and Ltb20). However, when all microsatellite loci are combined (i.e. treated as a single locus; *Structure* Q_{STRs}), we found that all nuclear markers could be constrained to a common center of 64 Km ($LRT_{\text{same-diff}} = 3.43$, $df=3$, $P > 0.05$), suggesting cline coincidence across all nuclear loci (Table S4).

Cline widths were also highly variable (MLEs = 1.0–66.7 km; 95% CI: 0.001-144.4 km, with six of the 13 loci having relatively narrow cline widths < 17 km, and two loci with wider widths ≥ 60 km (Table 2). Constraining all clines to a common width ($w=3.2$ km; CI: 0.14 -

32.2 Km, as defined by *Structure* Q_{total}) resulted in a significant decrease in likelihood ($LRT_{same-diff} = 258.6$, $df=12$, $P << 0.05$). However, 8 out of 13 markers, including mtDNA and the two nuclear diagnostic loci, could be constrained to have the *Structure* Q_{Total} width without a significant decrease in likelihood ($LRT_{same-diff} = 6.46$, $df=7$, $P > 0.05$; outliers: Ltb10, Ltb18, Ltb2; Ltb11, and Ltb20). Furthermore, when all microsatellite loci are converted into a single multilocus cline, all markers could be constrained to a common width ($LRT_{same-diff} = 2.71$, $df=3$, $P > 0.05$), suggesting cline concordance across all loci. When all clines are constrained to the width defined by *Structure* Q_{STRs} ($w = 27.4$ CI: 17.0 - 38.8 Km) we also obtained a significant decrease in likelihood ($LRT_{same-diff} = 76.3$, $df=12$, $P << 0.05$), but in contrast with the Q_{total} analysis, the cline width for mtDNA and microsatellites Ltb10, Ltb20, and Ltb37 were significantly different from $w = 27.4$ (Table S4).

Analysis of non-geographic (genomic) introgression patterns comparing hybrid index (HI) at a focal locus versus HI based on multilocus information revealed asymmetries in the direction of introgression, with clines shifted either toward LBSW or LBSE. However, except for mtDNA, we did not find statistically significant deviations from neutral expectations (Fig.S5; Table S5).

Environmental and ecological niche-based analyses

Based on Pearson's pairwise correlation threshold (absolute values < 0.70), we selected: 1) five bioclimatic variables related to temperature (Mean Diurnal Range, Isothermality, Minimum Temperature of Coldest Month, Mean Temperature of Wettest Quarter, Mean Temperature of Driest Quarter); 2) three related to precipitation (Annual Precipitation, Precipitation of Driest Month, Precipitation Seasonality); and 3) two related to habitat (median and standard deviation of the NDVI time-series). The ecological niche models built with the occurrence data and these variables performed well ($AUC = 0.820 \pm 0.019$). Jackknife measures of variable importance and contribution suggest that Precipitation Seasonality, median NDVI and Annual Precipitation are the most important variables in the model. Potentially favorable areas for the species appear to be continuously distributed in most of LBSW range and in the northern part of the LBSE

range. However, in the southern part of the LBSE range, including the contact zone, areas of suitable habitat are interrupted by areas of unfavorable habitat with variable extension (Fig. 4).

The GLS with MLPE with standardized variables showed the model with geographical distances is the most appropriate to describe patterns of genetic differentiation, based on the highest log-Likelihood and lowest AIC (Table 3). When comparing the single variable model based on geographic distances with the model that includes both distance matrices (geographic and resistance), significance tests show that inclusion of the resistance distance does not explain more variation in the dependent variable. On the other hand, when comparing with models based on resistance distance alone, the variance in genetic distances explained was significantly higher in the model with both variables, suggesting geographic distance is the best variable explaining genetic differentiation.

DISCUSSION

Discordant patterns of introgression

Previous phylogeographic studies identified a putative contact zone between late Miocene lineages of *L. boscai* in southwestern Iberia (Martínez-Solano *et al.*, 2006; Teixeira *et al.*, 2015). Our fine-scale analysis of this contact zone over a 150 km-transect provides evidence for a narrow hybrid zone *ca.* 3 - 27 km wide, which is centered at *ca.* 64 km from the coast. Overall, we found broadly concordant and coincident clines for all markers across the hybrid zone, except for the mtDNA cline centre, that is displaced *ca.* 60 km northeast relative to nuclear cline centres.

Discrepancies between mtDNA and nuclear clines have been reported in many hybrid zones (e.g., Sequeira *et al.*, 2005; Toews & Brelsford, 2012; Stewart, 2016). A variety of processes (natural selection associated with the environment or adaptive introgression, sex-biased dispersal, asymmetric reproductive isolation, and stochasticity associated with the increased effects of drift on maternally inherited mtDNA), acting alone or in combination, can explain these discordances. Unlike other hybrid zones (e.g. Endler, 1977; Tarroso *et al.*, 2014), we found evidence that the geographical location of the hybrid zone between the two lineages of *L. boscai* is likely environment-independent. Indeed, ENM analysis showed that environmental

conditions are relatively homogenous and suitable for the species across the hybrid zone. Furthermore, the genetic structure across the hybrid zone depends mainly on geographic distance rather than on resistance distance (measured as a function of habitat suitability), suggesting that exogenous selection (environment-associated selection) may not be acting directly to promote the movement of either mtDNA or nuclear markers across the hybrid zone (Chatfield *et al.*, 2010).

Sex-biased dispersal has been invoked as a driver of asymmetric patterns of introgression in some amphibians (e.g. Joly & Grolet, 1996; Jockusch & Wake, 2002; García-París *et al.*, 2003). Available evidence on dispersal capacity in *L. boscai* suggests limited dispersal and high site-fidelity (Díaz Paniagua, 1980). Field-based studies monitoring dispersal while accounting for sex differences are required to test this hypothesis. Discordance in mtDNA and nuclear cline position could also arise from asymmetric reproductive isolating mechanisms. Given that (nearly) all admixed individuals detected possess LBSW mtDNA, it is possible that either: i) hybridization between LBSW and LBSE is essentially unidirectional, with hybrids resulting from the mating of LBSW females with LBSE males (asymmetric pre-zygotic isolation); or ii) hybridization is reciprocal, but offspring resulting from mating of LBSE females with LBSW males are inviable or have reduced hybrid viability (asymmetric post-zygotic isolation). While pre-zygotic reproductive isolating mechanisms (e.g. mating preference, gametic incompatibility) have been reported in several amphibian species (Lamb & Avise, 1986; Michalak & Rafinski, 1999; Wirtz, 1999; Pfennig, 2000), pre-zygotic effects are not known to occur in natural newt populations (Macgregor *et al.*, 1990; Arntzen, 2003). Conversely, several instances of postzygotic effects have been proposed to explain patterns of asymmetric introgression in different newt species (Arntzen *et al.*, 2009; Johnson *et al.*, 2015; Zielinski *et al.*, 2016), namely sex-biased offspring production or survival (*sensu* Haldane's rule; Haldane, 1922; Laurie, 1997), and Dobzhansky–Muller incompatibilities (DMIs) (Dobzhansky, 1937; Turelli & Orr, 2000; Coyne & Orr, 2004).

Although heterogamy in *L. boscai* was not specifically determined, all other closely related species karyotyped are male heterogametic (XX/XY; Hillis & Green, 1990). Following

expectations from Haldane's rule, we would expect selection against heterogametic XY individuals, and thus a higher proportion of inviable hybrid males. This scenario, however, would not explain neither the pattern of minimal mtDNA introgression from LBSE to LBSW nor the lack of backcrosses involving parental LBSE individuals, as revealed by NEWHYBRIDS results. Therefore, it is possible that the observed discordances between mtDNA and nuclear clines in *L. boscai* result from incompatibilities between the maternal genome of LBSE and the paternal genome of LBSW individuals, i.e. certain LBSW genes might be lethal in an LBSE cytoplasmic background, which is compatible with the Dobzhansky–Muller model of hybrid incompatibility (Turelli & Orr, 2000). For example, Arntzen et al. (2009) found clear signs of asymmetric viability of reciprocal-cross hybrids between *Triturus cristatus* (crested) and *T. marmoratus* (marbled) newts that likely resulted from an incompatibility between the crested X chromosome and marbled cytoplasm, as adult F1 hybrids resulting from marbled-female *versus* crested-male crosses consisted entirely of males. Herrero (1991) reported the formation of sterile hybrids of *L. boscai* resulting from crosses between two groups of populations differentiated by C-band chromosomal patterns. According to this author, the occurrence of sterile hybrids derived from abnormalities during meiosis. Considering that all *L. boscai* populations studied by Herrero (1991) belong to the LBSE lineage (Martínez-Solano et al., 2006; Teixeira et al., 2015), evidence of this intraspecific postzygotic reproductive barrier is somewhat intriguing in light of the successful experimental hybridization between *L. boscai* and other closely related newts (Bucci-Innocenti, 1983). While asymmetry in backcrosses and in the distribution of mtDNA lineages could be explained by the action of pre- and post-zygotic mechanisms, further field studies on population sex ratios and common garden experiments comparing the survival of parental lineages, hybrids, and backcrosses in both directions are required to test these alternative but not mutually exclusive hypotheses.

Lastly, it is not possible to rule out the possibility that demographic processes associated with the historical biogeography of the species have contributed to generate the mitochondrial nuclear discordance observed across the hybrid zone. Previous phylogeographic studies on *L. boscai* (Martínez-Solano et al., 2006; Teixeira et al., 2015) inferred an ancestral hypothetical

refugium in the westernmost end of the Iberian Sistema Central, from where the species first expanded latitudinally during the Plio-Pleistocene, and only later on (during the mid-late Pleistocene) colonized the southeastern regions of the Iberian Peninsula. According to those studies, several genetically differentiated lineages of *L. boscai* (including LBSE and LBSW) persisted in fragmented allopatric refugia during the Pleistocene glacial cycles, and subsequently underwent a relatively rapid demographic expansion at the onset of the Last Glacial Maximum (23–18 KBP). Accordingly, it is possible that following postglacial secondary contact the LBSE lineage expanded westward, displacing the LBSW nuclear genome but not its mtDNA. Many studies have related this higher stationarity of mtDNA with its fourfold lower effective population size compared to nuclear markers, which in turn are more prone to form waves of introgression and spread faster through populations in contact (e.g. Sequeira *et al.*, 2005; Toews & Brelsford, 2012, Johnson *et al.*, 2015). This hypothesis would imply that the LBSW lineage was distributed further east of the present-day centre of the hybrid zone (where the initial contact likely occurred), while introgression of LBSW mtDNA haplotypes into LBSE was likely amplified by the effect of drift at the low population density expansion front (Currat *et al.*, 2008, but see Wielstra & Arntzen, 2012).

Moving hybrid zones resulting from range shifts of hybridizing species/lineages due to environmental changes have recently received substantial attention (Buggs, 2007; Wielstra & Arntzen, 2012; Wielstra *et al.*, 2017; Ryan *et al.*, 2018). These hybrid zones have been associated with geographically asymmetric patterns of introgression, often involving unidirectional introgression of mtDNA (Buggs, 2007; Sequeira *et al.*, 2011; Wielstra *et al.*, 2017). In *L. boscai* hybrid zone, STRUCTURE results and geographical cline analysis of microsatellite loci showed alleles typically found in the LBSW lineage to be present in LBSE populations located away from the centre of the hybrid zone, but not the reverse. Similarly, we found *β -fibint7* (lineage-diagnostic marker) LBSW alleles in LBSE populations located away from the centre of the hybrid zone, which, as in the case of microsatellite loci, can be interpreted to represent genetic footprints left in the wake of the moving hybrid zone (Wielstra *et al.*, 2017).

Some possible limitations of our study should be discussed. First, care should be taken when interpreting data from natural populations with high genetic variation with techniques like RFLP analysis. RFLP is a cost-efficient technique providing robust information about the assignment of samples to species or intraspecific lineages, but may fail to detect rare haplotypes, which genealogical information and spatial distribution could be highly informative about population dynamics, especially on individual dispersal and gene flow (e.g., Gompert et al. 2014). Another limitation of RFLPs is related with the inference of ancestry from allelic states, which can be biased because genetic recombination can originate the same combination of alleles in hybrid and parental individuals (e.g, Gompert et al., 2016). On the other hand, accounting for variation in the genetic outcomes of secondary contact between the same hybridizing taxa over different spatial contexts can provide robust inferences about the spatiotemporal dynamics of hybridization and speciation (Teeter et al. 2010; Gompert et al. 2014; Harrison & Larson, 2016; Arntzen et al., 2017). High-throughput sequencing methods could be applied to generate genome-wide datasets across additional transects to further test and refine our inferences.

Cline structure, dispersal and strength of selection

An increasing number of studies have found discrepancies in cline structure among non-diagnostic/polymorphic markers contrasting with steep, concordant and coincident clines for diagnostic traits (Gay et al., 2008; Yuri et al., 2009; Teeter et al. 2010; Carneiro et al., 2013). These studies suggest that the proportion of the genome that is subject to introgression varies substantially between hybridizing taxa, and the exclusive use of highly differentiated markers often suggests reduced levels of introgression, thus offering a biased view of genetic structure and general patterns of admixture across hybrid zones. In the *L. boscai* hybrid zone, both cline centre position and width for some microsatellite loci were shifted (both wider and narrower clines) from those of the nuclear multilocus reference cline (“multilocus” Structure Q_{Total}). These results are in line with the general expectations for cline heterogeneity for putatively neutral microsatellite loci. The fact that we obtained largely coincident and concordant clines for most

markers thus suggests the existence of a barrier to gene flow between the two lineages of *L. boscai*. Moreover, considering sampling variability across the hybrid zone (which can make it difficult to accurately measure clinal allele frequency changes in highly polymorphic microsatellite loci) and that width and position of all microsatellite loci clines are within the CIs obtained for both diagnostic loci, we conservatively conclude that all nuclear markers across the hybrid zone are broadly concordant and coincident (Fig. 3).

Nevertheless, several lines of evidence suggest the genetic structure in the *L. boscai* hybrid zone cannot be simply explained by a tension zone model. Following predictions of this model, the hybrid zone should comprise mainly first-generation hybrids (Barton & Hewitt, 1985; Barton & Gale, 1993; Jiggins & Mallet, 2000), which is not the case in our system, as most admixed individuals apparently correspond to second generation hybrids and backcrosses, as revealed by both STRUCTURE and NEWHYBRIDS analyses. Besides, values of F_{IS} (heterozygote deficit, HWE deviations) and linkage disequilibrium (LD) do not vary substantially along the transect, and do not increase at the centre of the hybrid zone (localities 9 and 10), as expected under a tension zone model (Fig. S5). Assuming migration-selection equilibrium, we can calculate effective selection acting against hybrids (s^*) using the tension zone properties through the formula $s^* = 8\sigma^2/w^2$, where σ is the standard deviation of the distance between parents and offspring per generation, and w is the cline width (Bazykin, 1969). Little is known about dispersal rates in *L. boscai*, but Díaz-Paniagua (1980) registered 80 m as the maximal inter-pond migration distance for the species, and Trochet et al. (2014) reported *ca.* 180 m as the maximal dispersal distance for its sister species *L. italicus*. Considering a generation time of 3 years (Caetano & Leclair, 1999), and assuming a conservative value of $\sigma \approx 200$ m/generation^{-1/2}, maintaining this cline width observed at nuclear markers (Structure- Q_{total}) relative to dispersal would require a mean $s^* \approx 3\%$. Estimates of s^* would be even lower for each nuclear diagnostic locus and microsatellite Q_{STRs} widths separately (0.02%; 0.03%; 0.043%, for *GH*, *β -fibint7*, and Q_{STRs} , respectively). Overall, these estimates are lower than those reported in other systems, like *Bombina* toads (17-22%, Szymura & Barton, 1986, 1991), *Ensatina* salamanders (46-75%, Alexandrino et al., 2005),

Sceloporus lizards (32%; Marshall & Sites, 2001), or *Carlia* skinks (22–70%, Phillips *et al.*, 2004), but are more similar to estimates found in other systems in which hybridization presumably occurs under a neutral diffusion process, like in the Iberian salamander *Chioglossa lusitanica* (0.01–0.1%, Sequeira *et al.*, 2005) or tension zones where linkage disequilibrium is relatively low (0.01–0.6%; Baldassarre *et al.*, 2014; Wielstra *et al.*, 2017).

Despite the relatively low levels of selection against hybrids estimated in the *L. boscai* hybrid zone, it is worth mentioning that these quantitative values are heavily dependent on an accurate measure of dispersal rate, which may be underestimated in our case, as it did not consider the potential influence of juvenile dispersal, long-distance migration and extinction-recolonization events (Barton & Gale, 1993). Indeed, in several amphibian species, including newts, dispersal rates in juveniles are larger than those of conspecific adults (Müllner *et al.*, 2001; Sinsch, 2014). In the absence of selection, cline width (w) will be a function of the time at secondary contact (T , measured in generations) and dispersal rate (σ), and is given by the following equation: $w = \sigma 2.51 \sqrt{T}$ (Barton & Gale, 1993). Although we do not have information on the exact timing of hybridization between these *L. boscai* lineages, it is reasonable to assume that the contact must have been possible at least since the Holocene (10 Kya) following climatic amelioration in the Iberian Peninsula after the last glacial maximum (Hewitt, 2000; Sequeira *et al.*, 2005). Using our estimate of dispersal parameter $\sigma \approx 200$ m/generation^{-1/2} (Díaz-Paniagua, 1980), and a generation time of 3 years (Caetano & Leclair, 1999), the hybrid zone width would be *ca.* 29 km if maintained by neutral diffusion, which is within the CI interval of the widths obtained for the reference cline (Q_{total} ; CI 95% 0.14–32.2), the microsatellite reference cline (Q_{STRs} ; CI 95% 17.0–38.8), and the two diagnostic nuclear marker clines (CI 95% 0.25 - 35.2 for β -fibint7, and CI 95% 8.7 - 60.9 for *GH*).

This evidence, together with results from ecological niche modelling, which do not support a prominent role of the environment (i.e. exogenous selection) in shaping the *L. boscai* hybrid zone, suggest that the present-day genetic structure of this hybrid zone likely results from the interplay of complex demographic factors and population-level processes, albeit at least low levels of selection against hybrids must be invoked to explain the narrowness and coincidence

of clines observed. Indirect estimates of selection (s^*) are highly influenced by dispersal estimates and multilocus effects (i.e., differential selective strength among loci across the genome), and often represent only a fraction of the effective strength of selection on the zone or in the total fitness reduction in hybrids (Barton & Hewitt, 1983; Teeter *et al.*, 2010; Baldassarre *et al.*, 2014). Notwithstanding, the observed partial reproductive isolation between these two *L. boscai* lineages is somewhat intriguing given the potential for the evolution of genetic incompatibilities since their divergence in the late Miocene (Teixeira *et al.*, 2015). Although the degree of genetic divergence is usually regarded as a good predictor for the level of reproductive isolation in many vertebrates (Coyne & Orr, 2004; Pereira *et al.*, 2011; Dufresnes *et al.*, 2015), there are several well documented cases of weak correlation between hybridization potential or long lasting genetic compatibility and divergence time (Jiggins & Mallet, 2000; Harrison & Larson, 2016), including some amphibians (Arntzen *et al.*, 2014, 2016; Johnson *et al.*, 2015). For example, results of artificial hybridization experiments between *L. boscai* and its sister species *L. italicus* showed that despite the long time since divergence (early Miocene), they are still able to produce larvae and metamorphs, albeit with substantial post-zygotic barriers, as evidenced by the reduced percentage of cleaved eggs (50%) and the low number of larvae that reached the hatching stage (15%) (Bucci-Innocenti *et al.*, 1983).

Breeding experiments, including direct measures of hybrid fitness, could accurately quantify the effects of potential pre- and postzygotic barriers to hybridization and help tease apart the relative contributions of neutral and selective processes in shaping the current genetic structure of the hybrid zone.

Conclusions

Based on the deep mtDNA divergence of lineage LBSW found by Martínez-Solano *et al.* (2006), Dubois & Raffaëlli (2009) re-elected this lineage for species status (*L. maltzani*), removing it from the synonymy of *L. boscai*. However, this status has been questioned by other authors because of the lack of additional data, namely the analysis of nuclear data and the examination of areas of secondary contact (Speybroeck *et al.*, 2010). Our study, based on extensive sampling and the analysis of several nuclear markers, including mtDNA, single copy

nuclear genes and microsatellites, corroborates previous studies suggesting that lineages LBSE and LBSW (lineages A and B, respectively, following Martínez-Solano *et al.*, 2006) are clearly distinct evolutionary units that have followed long, independent trajectories. Our analysis of the hybrid zone showed that despite the lack of full reproductive isolation, gene flow is geographically restricted, producing a narrow genetic transition between the two *L. boscai* lineages, with evidence of (weak) selection against hybrids.

The results of ENM analyses provide evidence that the spatial overlap between these two lineages is prevented by an extensive area of unfavorable habitat south of the present-day hybrid zone. Thus, admixture following contact between the two lineages seems to be only possible in the northern part of the LBSW range. Under this scenario, it seems unlikely that genetic admixture between the two *L. boscai* lineages and the spread of hybrid individuals could revert the divergence process. While our results suggest LBSW may well qualify as a full, cryptic species, additional morphological and ecological studies are required to discern the relative roles of intrinsic and extrinsic factors in shaping the strength of reproductive isolation in this hybrid zone.

Acknowledgments

The ‘Consejerías de Medio Ambiente’ of Madrid, Extremadura, Castilla-La Mancha and Andalucía (Spain), and ICN (Portugal) provided the legal collecting permits. This study was funded by the Program Operacional Factores de Competitividade (COMPETE), and by national funds from Fundação para a Ciência e a Tecnologia (FCT), through the research Project PTDC/BIABEC/105083/2008 (PI: FS). FS was supported by FEDER (COMPETE, POCI-01-0145-FEDER-006821) and FCT (UID/BIA/50027/2013) funds. IMS was funded by Project “Biodiversity, Ecology and Global Change”, co-financed by ON. 2–O Novo Norte, under the NSRF, through the ERDF. HG was supported by a postdoctoral Grant from FCT (SFRH/BPD/102966/2014). This study was also financed by CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Finance Code 88881.132701/2016-01) and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) through PhD

655 scholarships to ABS and TSN, and research project and postdoctoral Grant (302892/2016-8;
656 232916/2013-2016) to MV. We thank the following people for sharing samples or helping with
657 field work: J.W. Arntzen; A.M. Barbosa; E. G. González; J. Gutiérrez; E. Recuero, Guilherme
658 Dias, and Coralie Nourisson.

659 **References**

- 660 Abbott, R., Albach, D., Ansell, S., Arntzen, J.W., Baird, S.J.E., Bierne, N. et al. 2013.
661 Hybridization and speciation. *J. Evol. Biol.* 26: 229–246.
- 662 Alexandrino, J., Baird, S.J.E., Lawson, L., Macey, J.R., Moritz, C. & Wake, D.B. 2005. Strong
663 selection against hybrids at a hybrid zone in the *Ensatina* ring species complex and its
664 evolutionary implications. *Evolution* 59: 1334–1347.
- 665 Anderson, E.C., Thompson, E.A. 2002. A model-based method for identifying species hybrids
666 using multilocus genetic data. *Genetics* 160:1217–1229.
- 667 Arntzen, J.W. 2003. *Triturus cristatus* Superspecies-Kammolch-Artenkreis (*Triturus cristatus*
668 (Laurenti, 1768)-Nordlicher Kammolch, *Triturus carnifex* (Laurenti, 1768)-Italienischer
669 Kammolch, *Triturus dobrogicus* (Kiritzescu, 1903)-Donau-Kammolch, *Triturus karelinii*
670 (Strauch, 1870)-Südlicher Kammolch). Pp. 421–514 in K. Grossenbacher and B. Thiesmeier,
671 eds. Handbuch der Reptilien und Amphibien Europas. Schwanzlurche IIA, Aula-Verlag,
672 Wiebelsheim.
- 673
674 Arntzen, J.W., Jehle, R., Bardakci, F., Burke, T. & Wallis, G. P. 2009. Asymmetric viability of
675 reciprocal-cross hybrids between crested and marbled newts (*Triturus cristatus* and *T.*
676 *marmoratus*). *Evolution* 63:1191–1202.
- 677 Arntzen, J.W., Wielstra, B. & Wallis, G.P. 2014. The modality of nine *Triturus* newt hybrid
678 zones assessed with nuclear, mitochondrial and morphological data. *Biol. J. Linn. Soc.* 113:
679 604–622.
- 680 Arntzen, J.W., Trujillo, T., Butôt, R., Vrieling, K., Schaap, O., Gutiérrez-Rodríguez, J. &
681 Martínez-Solano, I. 2016. Concordant morphological and molecular clines in a contact zone of
682 the Common and Spined toad (*Bufo bufo* and *B. spinosus*) in the northwest of France. *Front.*
683 *Zool.* 13: 52.
- 684 Arntzen, J. W., de Vries, W., Canestrelli, D. & Martínez-Solano, I. 2017. Hybrid zone
685 formation and contrasting outcomes of secondary contact over transects in common toads. *Mol.*
686 *Ecol.* 26: 5663–5675.
- 687 Baldassarre, D.T., White, T.A., Karubian, J. & Webster, M.S. 2014. Genomic and
688 morphological analysis of a semipermeable avian hybrid zone suggests asymmetrical
689 introgression of a sexual signal. *Evolution* 68: 2644–2657.
- 690 Barton, N.H. & Bengtsson, B.O. 1986. The barrier to genetic exchange between hybridising
691 populations. *Heredity* 57: 357–376.
- 692 Barton, N.H. & Gale, K.S. 1993. Genetic analysis of hybrid zones. In: Hybrid Zones and the
693 Evolutionary Process (R.G. Harrison, ed), pp. 13–45. Oxford University Press, New York.

- 694 Barton, N.H. & Hewitt, G.M. 1985. Analysis of hybrid zones. *Annu. Rev. Ecol. Syst.* 16: 113–
695 148.
- 696 Barton, N.H. & Hewitt, G.M. 1989. Adaptation, speciation and hybrid zones. *Nature* 341: 497–
697 503.
- 698 Bazykin, A.D. 1969. Hypothetical mechanism of speciation. *Evolution* 23: 685–687.
- 699 Bucci-Innocenti, S., Ragghianti, M. & Mancino, G. 1983. Investigation of karyology and
700 hybridization of *Triturus boscai* and *T. vittatus*, with a reinterpretation of the species groups
701 within the genus *Triturus* (Caudata, Salamandridae). *Copeia* 3: 662–672.
- 702 Buggs, R.J.A. 2007. Empirical study of hybrid zone movement. *Heredity* 99: 301–312.
- 703 Caetano, M.H. & Leclair, R. 1999. Comparative phenology and demography of *Triturus boscai*
704 from Portugal. *J Herpetol.* 33:192–202.
- 705 Carneiro, M., Baird, S.J., Afonso, S., Ramirez, E., Tarroso, P., Teotónio, H., Villafuerte, R,
706 Nachman, M.W. & Ferrand, N. 2013. Steep clines within a highly permeable genome across a
707 hybrid zone between two subspecies of the European rabbit. *Mol. Ecol.* 22:2511–2525
- 708 Chatfield, M.W.H., Kozak, K.H., Fitzpatrick, B.M. & Tucker, P.K. 2010. Patterns of differential
709 introgression in a salamander hybrid zone: inferences from genetic data and ecological niche
710 modelling. *Mol. Ecol.* 19: 4265–4282.
- 711 Clarke, R. T., Rothery, P. & Raybould, A. F. 2002. Confidence limits for regression
712 relationships between distance matrices: Estimating gene flow with distance. *J Agric Biol*
713 *Environ Stat*, 7: 361-372.
- 714 Coyne, J.A. & Orr, H.A. 2004. Speciation. Sinauer Associates, Sunderland, MA.
- 715 Currat M., Ruedi M., Petit R. J. & Excoffier L. 2008. The hidden side of invasions: massive
716 introgression by local genes. *Evolution* 62:1908–1920.
- 717 Derryberry, E.P., Derryberry, G.E., Maley J.M. & Brumfield, R.T. 2014. HZAR: hybrid zone
718 analysis using an R software package. *Mol. Ecol. Resour.* 14: 652–663.
- 719 Díaz-Paniagua, C. 1980. Actividad de *Triturus marmoratus* y *Triturus boscai* (Amphibia:
720 Caudata) durante su periodo de reproducción. II Reunión Iberoamericana. *Cons. Zool. Vert.*
721 *Cáceres*: 26-36.
- 722 Díaz-Paniagua, C. 2002. *Triturus boscai*. Pp. 61-63 in: Pleguezuelos JM, Márquez R, Lizana M,
723 eds., Atlas y Libro Rojo de los Anfibios y Reptiles de España. Madrid: Dirección General de
724 Conservación de la Naturaleza and Asociación Herpetológica Española.
- 725 Didan, K. 2018. Mod13a3 modis/terra vegetation indices monthly l3 global 1km sin grid v006.
726 NASA EOSDIS LP DAAC.
- 727 Dobzhansky, T., 1937. Genetics and the Origin of Species. Columbia University Press, New
728 York.
- 729 Dubois, A. & Raffaëlli, J. 2009. A new ergotaxonomy of the family Salamandridae Goldfuss,
730 1820 (Amphibia, Urodela). *Alytes* 26: 1–85.

- 731 Dufresnes, C., Brelsford, A., Crnobrnja-Isailović, J., Tzankov, N., Lymberakis, P. & Perrin, N.
732 2015. Timeframe of speciation inferred from secondary contact zones in the European tree frog
733 radiation (*Hyla arborea* group). *BMC Evol. Biol.* 15: 155.
- 734 Endler, J.A. 1977. Geographic Variation, Speciation and Clines. Princeton University Press,
735 Princeton.
- 736 Evanno, G., Regnaut, S., & Goudet, J. 2005. Detecting the number of clusters of individuals
737 using the software STRUCTURE: A simulation study. *Mol. Ecol.* 14: 2611–2620.
- 738 Fitzpatrick, B.M. 2013. Alternative forms for genomic clines. *Ecol. Evol.* 3:1951–66.
- 739 Funk, D.J. & Omland, K.E. 2003. Species level paraphyly and polyphyly: frequency, causes,
740 and consequences, with insights from animal mitochondrial DNA. *Annu. Rev. Ecol. Evol. Syst.*
741 34: 397–423.
- 742 García-París, M., Alcobendas, M., Buckley, D. & Wake, D.B. 2003. Dispersal of viviparity
743 across contact zones in Iberian populations of fire salamanders (*Salamandra*) inferred from
744 discordance of genetic and morphological traits. *Evolution* 57:129-143.
- 745 Gay, L., Crochet, P.A., Bell, D.A. & Lenormand, T. 2008. Comparing clines on molecular and
746 phenotypic traits in hybrid zones: a window on tension zone models. *Evolution*, 62, 2789–2806.
- 747 Gompert, Z., Lucas, L. K. Buerkle, C.A., Forister, M.L., Fordyce, J.A. & Nice, C.C. 2014.
748 Admixture and the organization of genetic diversity in a butterfly species complex revealed
749 through common and rare genetic variants. *Molecular Ecology* 23: 4555– 4573.
- 750 Gompert, Z. & Buerkle, C.A. 2009. A powerful regression-based method for admixture
751 mapping of isolation across the genome of hybrids. *Mol. Ecol.* 18: 1207–24.
- 752 Gompert, Z. & Buerkle, C. A. 2016. What, if anything, are hybrids: enduring truths and
753 challenges associated with population structure and gene flow. *Evol. Appl.* 9: 909–23.
- 754 Gonçalves, H., Maia-Carvalho, B., Sousa-Neves, T., García-París, M., Sequeira, F., Ferrand, N.
755 & Martinez-Solano, I. 2015. Multilocus phylogeography of the common midwife toad, *Alytes*
756 *obstetricans* (Anura, Alytidae): contrasting patterns of lineage diversification and genetic
757 structure in the Iberian refugium. *Mol Phylogenet Evol.* 93: 363 –379.
- 758 Goudet, J. 2001. FSTAT version 2.9.3, A program to estimate and test gene diversities and
759 fixation indices. Updated from Goudet (1995). <http://www2.unil.ch/popgen/softwares/fstat.htm>.
- 760 Haldane, J.B.S. 1922. Sex ratio and unisexual sterility in hybrid animals. *J. Genet.* 12: 101–109.
- 761 Harrison, R.G. 1990. Hybrid zones: windows on evolutionary processes. *Oxf. Surv. Evol. Biol.*
762 7: 69–128.
- 763 Harrison, R.G. & Larson, E.L. 2016. Heterogeneous genome divergence, differential
764 introgression, and the origin and structure of hybrid zones. *Mol. Ecol.* 25: 2454–2466.
- 765 Herrero, P. 1991. Polytypic chromosomal variation in *Triturus boscai* (Urodela:
766 Salamandridae). *Genet Sel Evol.* 23: 263-272

767 Hewitt, G.M. 1988. Hybrid zones: natural laboratories for evolutionary studies. *Trends Ecol.*
768 *Evol.* 3: 158–167.

769 Hewitt, G.M. 2000. The genetic legacy of the Quaternary ice ages. *Nature* 405: 907–913.

770 Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005). Very high resolution
771 interpolated climate surfaces for global land areas. *Int J Climatol.* 25: 1965–1978.

772 Hillis, D.M. & Green, D.M. 1990. Evolutionary phylogenetic changes of heterogametic history
773 of amphibians. *J. Evol. Biol.* 3: 49–64.

774 Hofreiter, M. & Stewart, J., 2009. Ecological change, range fluctuations and review population
775 dynamics during the Pleistocene. *Curr. Biol.* 19: R584–R594

776 Jiggins, C.D. & Mallet, J. 2000. Bimodal hybrid zones and speciation. *Trends Ecol. Evol.* 15:
777 250–255.

778 Jockusch, E.L. & Wake D.B. 2002. Falling apart and merging: diversification of slender
779 salamanders (Plethodontidae: *Batrachoseps*) in the American West. *Biol. J. Linn. Soc.* 76: 361–
780 391.

781 Johnson, B.B., White, T.A., Phillips, C.A. & Zamudio, K.R. 2015. Asymmetrical introgression
782 in a spotted salamander hybrid zone. *Heredity* 106: 608–617.

783 Joly, P. & Grolet O. 1996. Colonization dynamics of newly-created ponds in the Alpine newt,
784 *Triturus alpestris*: influence of landscape configuration and age structure of the colonizers. *Acta*
785 *Oecol* 17: 599–608.

786 Kindler, C., Chèvre, M., Ursenbacher, S., Böhme, W., Hille, A., Jablonski, D., Vamberger, M.
787 & Fritz, U. 2017. Hybridization patterns in two contact zones of grass snakes reveal a new
788 Central European snake species. *Sci. Rep.* 7: 7378.

789 Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A. & Mayrose, I. 2015. Clumpak: a
790 program for identifying clustering modes and packaging population structure inferences across
791 K. *Mol Ecol Res* 15:1179–1191

792 Kruuk, L.E.B., Baird, S.J.E., Gale, K.S. & Barton, N.H. 1999. A comparison of multilocus
793 clines maintained by environmental adaptation or by selection against hybrids. *Genetics* 153:
794 1959–1971.

795 Lamb, T. & Avise, J.C. 1986. Directional introgression of mitochondrial DNA in a hybrid
796 population of tree frogs — the influence of mating behavior. *Proc. Natl. Acad. Sci., USA* 83:
797 2526–2530.

798 Laurie, C.C. 1997. The weaker sex is heterogametic: 75 years of Haldane's rule. *Genetics* 147:
799 937–951

800 Leache, A.D. & Cole, C.J. 2007. Hybridization between multiple fence lizard lineages in an
801 ecotone: locally discordant variation in mitochondrial DNA, chromosomes, and morphology.
802 *Mol. Ecol.* 16: 1035–1054.

803 Loureiro, A., Ferrand, N., Carretero, M. A. & Paulo, O. (2008). Atlas dos Anfíbios e Répteis de
804 Portugal. Instituto de Conservação da Natureza e Biodiversidade, Lisboa.

805 Macgregor, H.C., Sessions, S.K. & Arntzen, J.W. 1990. An integrative analysis of phylogenetic
806 relationships among newts of the genus *Triturus* (family: Salamandridae), using comparative
807 biochemistry, cytogenetics and reproductive interactions. *J Evol Biol* 3: 329-374.

808 Marshall, J.C. & Sites J.W. 2001. A comparison of nuclear and mitochondrial cline shapes in a
809 hybrid zone in the *Sceloporus grammicus* complex (Squamata: Phrynosomatidae). *Mol Ecol*
810 10:435-449.

811 Martínez-Solano, I, Teixeira, J, Buckley, D. & García-París, M. 2006. Mitochondrial DNA
812 phylogeography of *Lissotriton boscai* (Caudata, Salamandridae): evidence for old, multiple
813 refugian in an Iberian endemic. *Mol Ecol* 15: 3375-3388.

814 Michalak, P. & Rafinski, J. 1999. Sexual isolation between two newt species, *Triturus vulgaris*
815 and *Triturus montandoni* (Amphibia, Urodela, Salamandridae). *Biol. J. Linn. Soc* 67: 343-352.

816 Moore, W.S. 1977. An evaluation of narrow hybrid zones in vertebrates. *Q. Rev. Biol.* 52, 263-
817 277.

818 Moore, W.S. & Price, J.T. 1993. Nature of selection in the northern flicker hybrid zone and its
819 implications for speciation theory. In: *Hybrid Zones and the Evolutionary Process* (R.G.
820 Harrison, ed), pp. 196-225. Oxford University Press, New York.

821 Müllner, A. 2001. Spatial patterns of migrating great crested newts and smooth newts: the
822 importance of the terrestrial habitat surrounding the breeding pond. *Rana* 4:279-293.

823 Peñalver-Alcázar, M., Martínez-Solano, I., Sequeira, F. & Aragón, P. 2017. Sex-dependent
824 implications of primary productivity and conspecific density on geographic body size variation
825 in a newt: Disentangling local, large scale and genetic factors. *J. Biogeogr* 44: 2096-2108.

826 Pereira, R., Monahan, W. & Wake, D. 2011. Predictors for reproductive isolation in a ring
827 species complex following genetic and ecological divergence. *BMC Evol. Biol.* 11: 194.

828 Pfennig, K.S. 2000. Female spadefoot toads compromise on mate quality to ensure conspecific
829 matings. *Behav. Ecol.* 11:220-227.

830 Phillips, B.L., Baird, S.J.E. & Moritz, C. 2004. When vicars meet: a narrow contact zone
831 between morphologically cryptic phylogeographic lineages of the rainforest skink, *Carlia*
832 *rubrigularis*. *Evolution* 58: 1536-1548.

833 Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E. & Blair, M. E. 2017. Opening the
834 black box: an open-source release of maxent. *Ecography* 40: 887-893.

835 Pleguezuelos, J. M., Márquez, R., & Lizana, M. 2004. Atlas y Libro Rojo de los Anfibios y
836 Reptiles de España. Madrid: Dirección General de Conservación de la Naturaleza and
837 Asociación Herpetológica Española.

838 Pritchard, J.K., Stephens, M. & Donnelly P. 2000. Inference of population structure using
839 multilocus genotype data. *Genetics* 155:945-959.

840 R Core Team. 2018. R: A Language and Environment for Statistical Computing [Computer
841 software manual]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from
842 <http://www.r-project.org>

- 843 Rice, W.R. 1989. Analysing tables of statistical tests. *Evolution* 43: 223–225.
- 844 Rohwer, S., Bermingham, E. & Wood, C. 2001. Plumage and mitochondrial DNA haplotype
845 variation across a moving hybrid zone. *Evolution* 55: 405-422.
- 846 Ryan, S., Deines, J.M., Scriber, J.M., Pfrender, M.E., Jones, S.E., Emrich, S.J. & Hellmann, J.J.
847 2018. Climate-mediated hybrid zone movement revealed with genomics, museum collection,
848 and simulation modeling. *Proc. Natl. Acad. Sci., USA* 115: E2284-E2291
- 849 Seehausen O. 2006 Conservation: losing biodiversity by reverse speciation. *Curr. Biol.* 16: 334–
850 347.
- 851 Sequeira, F., J. Alexandrino, S. Rocha, J. W. Arntzen, and N. Ferrand. 2005. Genetic exchange
852 across a hybrid zone within the Iberian endemic golden-striped salamander, *Chioglossa*
853 *lusitanica*. *Mol. Ecol.* 14: 245-254.
- 854 Sequeira, F., Sodr, D., Ferrand, N., Bernardi, J. A., Sampaio, I., Schneider, H., & Vallinoto, M.
855 2011. Hybridization and massive mtDNA unidirectional introgression between the closely
856 related Neotropical toads *Rhinella marina* and *R. schneideri* inferred from mtDNA and nuclear
857 markers. *BMC Evol. Biol.* 11: 264.
- 858 Sequeira, F., Silva-Ferreira, A. & Lopes, S. 2012. Microsatellite markers for the Iberian
859 endemic Bosca’s newt, *Lissotriton boscai* (Caudata, Salamandridae). *Conserv. Genet. Resour.*
860 4: 715–717.
- 861 Shirk, A.J., Landguth, E.L. & Cushman, S.A. 2018. A comparison of regression methods for
862 model selection in individual-based landscape genetic analysis. *Mol. Ecol. Resour.* 18: 55-67.
- 863 Singhal, S. & Moritz, C. 2012. Strong selection against hybrids maintains a narrow contact zone
864 between morphologically cryptic lineages in a rainforest lizard. *Evolution* 66: 1474–1489.
- 865 Singhal, S., and C. Moritz. 2013. Reproductive isolation between phylogeographic lineages
866 scales with divergence. *Proc. R. Soc. B Biol. Sci.* 280: 2013-2246.
- 867 Sinsch, U. 2014. Movement ecology of amphibians: from individual migratory behavior to
868 spatially structured populations in heterogeneous landscapes. *Can. J. Zool.* 92: 491–502.
- 869 Speybroeck, J., Beukema, W. and Crochet, P.-A. 2010. A tentative species list of the European
870 herpetofauna (Amphibia and Reptilia)—an update. *Zootaxa*, 2492: 1–27.
- 871 Stewart, K. A., Austin, J. D., Zamudio, K. R. & Loughheed, S.C. 2015. Contact zone dynamics
872 during early stages of speciation in a chorus frog (*Pseudacris crucifer*). *Heredity* 116: 239-47.
- 873 Stuart-Fox, D., Godinho, R., Bellocq J.G., Irwin, N., Brito, J., Moussalli, A., Siroky, P., Hugall,
874 A., & Baird, S. 2009. Variation in phenotype, parasite load and male competitive ability across
875 a cryptic hybrid zone. *PLoS ONE* 4: e5677
- 876 Szymura, J.M. & Barton, N.H. 1986. Genetic analysis of a hybrid zone between the fire-bellied
877 toads, *Bombina bombina* and *B. variegata*, near Cracow in southern Poland. *Evolution* 40:1141-
878 1159.
- 879 Szymura, J.M. & Barton, N.H. 1991. The genetic structure of the hybrid zone between the fire-
880 bellied toads *Bombina bombina* and *B. variegata*: Comparisons between transects and between
881 loci. *Evolution* 45: 237-261.

882 Tarroso, P., Pereira, R. J., Martinez-Freiria, F., Godinho, R. & Brito, J. C. Hybridization at an
883 ecotone: ecological and genetic barriers between three Iberian vipers. *Mol. Ecol.* 23: 1108–
884 1123.

885 Taylor S. A., Larson E. L. & Harrison R.G. 2015. Hybrid zones: windows on climate change.
886 *Trends Ecol. Evol.* 30:398–406.

887 Teeter, K.C., Thibodeau, L.M., Gompert, Z., Buerkle, C.A., Nachman, M.W. & Tucker, P.K.
888 2010. The variable genomic architecture of isolation between hybridizing species of house
889 mouse. *Evolution* 64:472–85

890 Teixeira, J. 2008. *Lissotriton boscai*. Pp. 98-99 in: Loureiro AM, Carretero MA, Ferrand N,
891 Paulo O., eds., Atlas dos Anfíbios e Répteis de Portugal. Lisboa: Instituto da Conservação da
892 Natureza.

893 Teixeira, J., Martínez-Solano, I., Buckley, D., Tarroso, P., García-París, M. & Ferrand, N. 2015.
894 Genealogy of the nuclear b-fibrinogen intron 7 in *Lissotriton boscai* (Caudata, Salamandridae):
895 concordance with mtDNA and implications for phylogeography and speciation. *Contributions*
896 *to Zool.* 84: 193–215.

897 Toews, D.P. & Brelsford, A. 2012. The biogeography of mitochondrial and nuclear discordance
898 in animals. *Mol. Ecol.* 21: 3907–3930.

899 Trochet, A., Moulherat, S., Calvez, O., Stevens, V., Clobert, J. & Schmeller, D. 2014. A
900 database of life-history traits of European amphibians. *Biodivers. Data J.* 2: e4123.

901 Turelli, M. & Orr, H.A. 2000. Dominance, epistasis and the genetics of postzygotic isolation.
902 *Genetics* 154: 1663-1679.

903 van Strien, M. J., Keller, D. & Holderegger, R. (2012). A new analytical approach to landscape
904 genetic modelling: Least-cost transect analysis and linear mixed models. *Mol. Ecol.* 21: 4010-
905 4023.

906 van Etten, J. 2017. R Package gdistance: Distances and Routes on Geographical Grids. *J. Stat.*
907 *Softw* 76: 21.

908 Yuri, T., Jernigan, R. W., Brumfield, R. T., Bhagabati, N. K. & Braun, M.J. 2009. The effect of
909 marker choice on estimated levels of introgression across an avian (Pipridae: Manacus) hybrid
910 zone. *Molecular Ecology*, 18: 4888-4903.

911 Wake, D. B. & Schneider C. J. 1998. Taxonomy of the Plethodontid salamander genus
912 *Ensatina*. *Herpetologica* 54:279–298.

913 Weiss, S. & Ferrand, N. 2007. Phylogeography of Southern European Refugia. Netherlands:
914 Springer

915 Wielstra, B. & Arntzen J.W. 2012. Postglacial species displacement in *Triturus* newts deduced
916 from asymmetrically introgressed mitochondrial DNA and ecological niche models. *BMC Evol.*
917 *Biol.* 12: 161

- 918 Wielstra, B., Burke, T., Butlin, R. K., Avcı, A., Üzümlü, N., Bozkurt, E., Olgun, K., Bozkurt, E.
919 & Arntzen, J. W. 2017. A genomic footprint of hybrid zone movement in crested newts. *Evol*
920 *Lett.* 1: 93-101.
- 921 Wirtz, P. 1999. Mother species-father species: unidirectional hybridization in animals with
922 female choice. *Animal Behav.* 58:1-12.
- 923 Wu, C. I. 2001. The genic view of the process of speciation. *J. Evol. Biol.* 14: 851–865.
- 924 Zieliński P., Nadachowska-Brzyska K., Dudek K. & Babik W. 2016. Divergence history of the
925 Carpathian and smooth newts modelled in space and time. *Mol. Ecol.* 25: 3912–3928

Legends:

Figure 1 – Maps showing: (A) Distribution of the two divergent lineages of *Lissotriton boscai* (LBSW, purple; LBSE, blue) in the Iberian Peninsula, and sampling localities (study area; square). Grey circles represent allopatric populations from which putatively pure individuals were used for the design of RFLP assays. Black circles represent localities within the hybrid zone; (B) sampling localities forming the transect (line) across the hybrid zone, and delimitation of the species' occurrence area (shading areas, adapted from Díaz-Paniagua, 2002 and Teixeira, 2008). Circles are sample locations indicating sites genotyped as having LBSW mtDNA (white) or LBSE mtDNA (black) haplotypes. Numbers of sampling localities as in Table 1.

Figure 2 - Identification and classification of the 129 individuals analysed over the transect: LBSW (purple) or LBSE (blue) according to the nuclear genes *GH* (A) and *Bfibint7* (B), and to mtDNA gene *ND4* (C). Bayesian assignment probabilities from STRUCTURE at $K=2$ based on 10 microsatellite loci (D). Individuals are represented as vertical bars partitioned into two segments, the length of which is proportional to the estimated membership probability to each cluster (LBSW, purple; LBSE, blue). NEWHYBRIDS results showing the posterior probability of each individual being pure LBSW (purple) or LBSE (blue), F2 hybrid (grey), or backcrossed with one of the parental species (black) based on 12 nuclear markers (E).

Figure 3 - Maximum likelihood fitted clines for 13 markers plotted as the frequency of LBSW alleles across a 150 km sampling transect. Multilocus “*Structure* Q_{Total} ” cline in orange, *Structure* Q_{STRs} in black, mitochondrial marker in red, diagnostic *GH* and *Bfibint7* in grey, and other nuclear markers in black. The dotted line represents the estimated cline centre and 95% credible interval for *Structure* Q_{total} (grey area). Distances are in km, starting from locality 6 (as in Table 1).

Figure 4 - Ecological niche modelling results showing present-day predicted geographic distribution for *L. boscai*. Sampling localities are labelled as Table 1. Colours indicate the probability of presence of the species. The inner box encompasses the location of the hybrid zone within the entire species range.

Supporting Information

Laboratory procedures

MtDNA *ND4* - DNA extraction and PCR amplification of a ≈ 700 bp fragment of ND4 were conducted as described in Martínez-Solano et al. (2006). A diagnostic restriction endonuclease (RE) assay was then designed to assign individuals to LBSW or LBSE through RFLP analysis. We digested amplified products in 10 μ l reactions, with 3 μ l template DNA, 0.3 μ l of enzyme FOK I (FastDigest), and 0.7 μ l of 10X FastDigest Buffer (Thermo Fisher Scientific, Massachusetts, USA) and incubated according to the manufacturer's protocol. RE fragment patterns were scored on 2 % agarose gels by electrophoresis.

Nuclear *β -fibint7* - Previous phylogeographic work based on β -fibint7 sequencing data showed several fixed differences across lineages (Teixeira et al., 2015). A diagnostic RE assay was designed to distinguish between the two lineages using a similar approach to that described above for mtDNA. Briefly, amplification of a ≈ 614 bp fragment of β -fibint7 was performed according to Teixeira et al. (2015), and PCR products were digested with the restriction enzyme Hpy8I in 10 μ l reactions, with 3 μ l template DNA, 0.3 μ l of enzyme, and 0.7 μ l of 10x Buffer Tango (Thermo Fisher Scientific, Massachusetts, USA), and incubated according to the manufacturer's protocol.

Nuclear *GH* - For GH nuclear gene, DNA extractions were performed using a standard proteinase K protocol (Sambrook et al., 1989). A fragment of ≈ 525 of GH was amplified via PCR using primers specifically designed for *L. italicus* (Caputo, 2012). PCR amplifications were performed using the MyTaq™ Mix (Bioline) following the manufacturer's instructions, with 0.5 μ M of each primer and approximately 10 ng of genomic DNA in a total reaction volume of 10 μ l. PCRs were performed over 37 cycles with the annealing temperature set to 56 °C. Successful amplifications were sequenced for both strands following the BigDye Terminator v3.1 Cycle sequencing protocol (Applied Biosystems) and sequenced in an ABI 3130 xl Genetic Analyzer. Sequences were aligned using Bioedit V. 7.1.9 (Hall, 1999), and checked manually.

In the case of heterozygous individuals, we inferred gametic phase by decoding size-heterozygous sequences directly interpreting the two allelic peaks superimposed onto each other, or in the case of two alleles that differed at one or more sites, we used software PHASE 2.1.1 (Stephens et al., 2001; Stephens & Donnelly, 2003) as implemented in DNAsp 5.0 (Librado & Rozas, 2009). We ran PHASE three times with different random seeds, and with haplotype probabilities assessed through 1000 replicates. Visual examination of the full GH alignment, with phased haplotypes, allowed the identification of fixed differences in alleles between the two

lineages (lineage-specific haplotypes), all of which were diagnosable with a RE assay. We digested amplified products in 10 µl reactions, with 3 µl template DNA, 0.3 µl of enzyme HpyF10VI, and 0.7 µl of 10x Buffer Tango and incubated according to the manufacturer's protocol.

Caputo, F.P. 2012. Frammentazione geografica, ibridazione introgressiva e nascita dell'hotspot di diversità genetica in Italia meridionale. PhD Thesis, Scienze Ecologiche e Biologiche – Università degli Studi della Tuscia.

Hall, T. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. 1999:95-98.

Librado, P. & Rozas, J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451– 1452.

Sambrook, J., Fritsch, E. & Maniatis, T. 1989. Molecular Cloning: A Laboratory Manual, 2nd edn. Cold Spring Harbor Press, New York.

Stephens, M., Smith, N. & Donnelly P. 2001. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet.* 68: 978-989.

Stephens, M. & Donnelly, P. 2003. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet.* 73:1162–1169.

TABLE S1. Individual allelic profiles for 10 microsatellites (SSR) and LTBSW ("1") and LTBSE ("2") alleles obtained through RFLP analysis for two nuclear loci (β -fibint7, and GH) and one mitochondrial DNA locus (ND4= across the hybrid zone within *Lissotriton boscai*. (please see Table S1.xlsx).

TABLE S2. Geographical location of sampled sites in the *Lissotriton boscai* hybrid zone, number of samples (N), mean number of alleles (N_A), expected and observed heterozygosity (H_E and H_O, respectively) and inbreeding coefficient (F_{IS}) obtained for the 23 sampled sites based on 10 microsatellites. The code numbers correspond to those shown in Figure 1.

Code	Locality	Longitude (°W)	Latitude (°N)	N	N _A	H _E	H _O	F _{IS}
1	Silves	-8.3529°	37.2511°	1 ₍₁₎	-	-	-	-
2	Monchique	-8.5207°	37.3015°	18 ₍₄₎	7.7	0.737	0.475	0.39* ^(2,3)
3	Arrifana	-8.8521°	37.2988°	1 ₍₁₎	-	-	-	-
4	Aljezur	-8.7354°	37.3983°	1 ₍₁₎	-	-	-	-
5	Cercal	-8.6587°	37.7447°	1 ₍₁₎	-	-	-	-
6	V.N. Milfontes	-8.7632°	37.8178°	25	10.4	0.810	0.653	0.22* ^(2,4)
7	Grândola	-8.5661°	38.1545°	4	-	-	-	-
8	Boticos	-8.4919°	38.0053°	1	-	-	-	-
9	Stª M. Serra	-8.6041°	38.1118°	20 ₍₃₎	9.1	0.741	0.564	0.27* ⁽¹⁾
10	Stª M. Sado	-8.3585°	38.1340°	20	8.6	0.810	0.702	0.16* ⁽¹⁾
11	Alcaçovas	-8.1377°	38.3893°	2	-	-	-	-
12	Vale de Brissos	-8.0719°	38.5154°	10	5.9	0.722	0.697	0.09
13	Guadalupe	-8.0203°	38.5601°	5	5.8	0.762	0.740	0.14
14	Évora	-8.1034°	38.5509°	8 ₍₇₎	6.8	0.762	0.680	0.19
15	S. B. Regadouro	-8.0442°	38.4714°	3	-	-	-	-
16	Mora	-8.1669°	38.8479°	23 ₍₂₎	11.5	0.841	0.761	0.12* ⁽⁵⁾
17	Azaruja	-7.7835°	38.7002°	7 ₍₃₎	8.1	0.832	0.731	0.20*
18	Portel	-7.6319°	38.3024°	22	10.7	0.836	0.749	0.13* ⁽⁵⁾
19	Aracena	-6.6236°	37.9307°	10 ₍₃₎	7.3	0.784	0.838	-0.02
20	Minas de Cala	-6.3448°	37.9587	16	7.4	0.755	0.786	-0.01
21	El Pedroso	-5.7773°	37.8119°	4 ₍₂₎	-	-	-	-
22	Calañas	-6.8828°	37.6515°	5 ₍₁₎	6.0	0.782	0.820	0.005
23	Huelva	-7.0175°	37.2672°	1 ₍₁₎	-	-	-	-
*24	Sintra	-9.4000	38.7972	3 ₍₃₎	-	-	-	-
*25	Tomar	-8.1077	39.6919	4	-	-	-	-
*26	Madrid	-4.3587	40.3379	2	-	-	-	-
*27	Valongo	-8.5056	41.2306	2	-	-	-	-
*28	Gerês	-8.0444	41.8139	2 ₍₂₎	-	-	-	-

$P < 0.05^*$; $p < 0.01^{**}$; $p < 0.005^{***}$

Microsatellite loci are: 1- *Ltb31*; 2- *Ltb2*; 3- *Ltb9*; 4 – *Ltb18*; 5-*Ltb20*

_(j) Number of samples used in previous studies (Martínez-Solano et al. 2006 ; Teixeira et al., 2016)

* Localities away from the hybrid zone

TABLE S4. Log- likelihood scores for the best-fitting unconstrained clines (Free), and clines constrained (Fixed) to have either the Structure Q_{total} or the Structure Q_{STRs} cline centre and width. Statistical significance between fixed and free models was estimated using likelihood ratio tests (2LL).

Loci	Centre			Width				
	Free	Fixed ($c = 64$ Km)	2LL	Free	Fixed ($w=27.4$ km)	2LL	Fixed ($w=3.2$ km)	2LL
Q_{STRs}	-1.23	-1.39	0.31	-1.34	-0.84	-1.094	-	-
ND4	-33.50	0	67.0*	0.000000	-4.75	9.51*	-0.000001	0.000001
β -fibint7	-10.78	-10.79	0.03	-10.79	-11.73	1.87	-10.79	0.01
GH	-5.06	-3.17	3.77	-3.49	-4.10	1.23	-5.389	3.81
LTB31	-17.34	-4.11	26.47*	-4.11	-4.43	0.65	-4.763	1.32
LTB37	-5.12	-2.39	5.46*	-2.39	-7.06	9.35*	-2.389	0.001
LTB10	-6.75	-5.26	2.98	-5.26	-7.62	4.73*	-11.32	12.11*
LTB18	-1.57	-0.72	1.70	-0.72	-1.09	0.75	-12.90	24.36*
LTB2	-13.29	-11.75	3.08	-11.75	-12.81	2.12	-20.50	17.52*
LTB9	-3.20	-2.77	0.86	-2.77	-2.74	0.06	-2.73	0.08
LTB11	-3.11	-3.11	0.01	-3.11	-4.53	2.84	-6.27	6.31*
LTB12	-5.25	-5.23	0.04	-5.23	-5.51	0.55	-5.94	1.42
LTB20	-7.02	-0.04	13.91*	-0.04	-20.92	41.77*	-95.96	191.86*
LTB4	-6.54	-1.39	0.31	-6.55	-6.73	0.36	-6.55	0.01

*Significantly different (cannot be constrained).

TABLE S5 - Logit-logistic model of Fitzpatrick and Barton's concordance for the 13 loci. For each individual per locus hybrid index is fitted against hybrid index calculated over all loci.

Loci	Logit-logistic			Barton		
	u	v	P	α	β	P
LTB31	2.321	1.587	0.449	-0.823	-0.472	0.694
LTB37	1.032	2.159	0.587	-0.139	0.492	0.878
LTB10	0.396	0.659	0.410	-0.249	-0.614	0.627
LTB18	0.914	3.065	0.091	-0.239	0.857	0.566
LTB2	-0.056	1.547	0.994	0.047	0.463	0.973
LTB9	0.994	1.411	0.846	-0.399	0.353	0.778
LTB11	-1.315	0.898	0.408	0.639	-0.362	0.319
LTB12	1.289	0.807	0.466	-0.638	-0.326	0.782
LTB20	1.564	1.802	0.641	-0.720	0.841	0.254
LTB4	-2.023	1.229	0.427	0.839	-0.489	0.161
β -fibint7	-1.235	1.338	0.709	0.476	0.051	0.726
GH	0.493	1.761	0.933	-0.150	0.549	0.841
ND4	-4.692	2.812	0.022	3.788	4.034	0.006

Figure S1 - Haplotype network for the 11 haplotypes determined by sequencing a fragment of the Growth Hormone intron 4 (GH) of 28 *Lissotriton boscai* individuals sampled at different localities (including some away from the hybrid zone; white circles). The size of circles is proportional to the number of individuals sharing the same haplotype belonging to lineages LBSW (purple) or LBSE (blue).

Figure S2 - Average estimates of inbreeding coefficients (FIS) and linkage disequilibria (R) for each site along the 150 km transect.

Figure S3 - Bayesian assignment probabilities from STRUCTURE at K=2 based on 10 microsatellite loci (A) and based on 12 markers (10 microsatellite loci and 2 nuclear genes) (B) for the entire dataset. Individuals are represented as vertical bars partitioned into two segments, the length of which is proportional to the estimated membership in the two clusters (LBSW, purple; and LBSE, blue). NewHybrids results based on 10 microsatellites, showing the posterior probability of each individual being pure LBSW (purple) or LBSE (blue), F2 hybrid (grey), or backcrossed with one of the parental species (black) (C).

Figure S4 - Best-fit K (clusters) based on ΔK method (Evanno et al., 2005) and values of the log posterior probability of the data [$\ln P(D)$] (Pritchard et al., 2000) as revealed by STRUCTURE analysis of the complete dataset (10 microsatellite loci + 2 nuclear genes) (a and b, respectively) and STRUCTURE analysis of 10 microsatellite loci (c and d, respectively).

Figure S5 - Analysis of introgression patterns comparing hybrid index for a focal locus (y axis) versus hybrid index for a multilocus expectation (x axis). Clines were compared using the logit-logistic model of Fitzpatrick (A), and Barton's concordance analysis (B). Lines show the individual per-locus hybrid index plotted against the hybrid index calculated over all loci. The equal-introgression expectation lies on the diagonal.

FIGURE S1

Figure S1. Haplotype network for the 11 haplotypes determined by sequencing a fragment of the Growth Hormone intron 4 (GH) of 28 *Lissotriton boscai* sampled at different localities (including some away from the hybrid zone; white circles). The size of circles is proportional to the number of individuals sharing the same haplotype belonging to LBSW (purple) or LBSE (blue) lineages.

FIGURE S2

Figure S2 - Average estimates of inbreeding coefficients (F_{is}) and linkage disequilibria (R) for each site along the 150 km transect.

FIGURE S3

Figure S3 - Bayesian assignment probabilities from STRUCTURE at $K = 2$ based on 10 microsatellite loci (a) and based on 12 (10 microsatellite loci and 2 SNPs) (b) for the entire dataset. Individuals are represented as vertical bars partitioned into two segments the length of which is proportional to the estimated membership in the two clusters (LBSW, purple; and LBSE, blue). NewHybrids results showing the posterior probability of each individual being pure LBSW (purple) or LBSE (blue), F2 hybrid (yellow), or backcrossed with one of the parental species (black) based on 10 microsatellite loci (c).

FIGURE S4

Figure S4 - Best-fit K (clusters) based on ΔK method (Evanno et al., 2005) and values of the log posterior probability of the data [$\ln P(D)$] (Pritchard et al., 2000) as revealed by Structure analysis of complete dataset (10 microsatellite loci + 2 SNPs) (a and b, respectively) and Structure analysis of 10 microsatellite loci (c and d, respectively).

FIGURE S5

Figure S5 - Analysis of introgression patterns comparing hybrid index for a focal locus (y axis) versus hybrid index for a multilocus expectation (x axis). Clines were compared using the logit-logistic model of Fitzpatrick (A), and Barton's concordance analysis (B). Lines show the individual per-locus hybrid index plotted against the hybrid index calculated over all loci. The equal-introgression expectation lies on the diagonal.