Hybridization and hybrid zone stability between two lizards explained by population genetics and niche quantification

IMANOL CABAÑA[1*,ⓞ], MARGARITA CHIARAVIGLIO[1], VALERIA DI COLA[2], ANTOINE GUISAN[2,3], OLIVIER BROENNIMANN[2,3], CRISTINA N. GARDENAL[1] and PAULA C. RIVERA[1,4,ⓞ]

[1]*Instituto de Diversidad y Ecología Animal (IDEA), CONICET-UNC and Facultad de Ciencias Exactas Físicas y Naturales, Universidad Nacional de Córdoba, Córdoba, Argentina*
[2]*Department of Ecology & Evolution, Biophore, University of Lausanne, Lausanne, Switzerland*
[3]*Institute of Earth Surface Dynamics, Geopolis, University of Lausanne, Lausanne, Switzerland*
[4]*Universidad Nacional de Chilecito, La Rioja, Argentina*

Understanding the factors that affect hybridization is an important issue in the study of species evolution. In this work, we analyse the genetic structure of two lizard species, *Salvator merianae* and *Salvator rufescens*, at a microscale within a climatic niche analysis framework, to reveal the main factors that contribute to the stability of their hybrid zone. We assess the effect of climate in hybridization by quantifying and decomposing the niche overlap of both species. Using a mitochondrial and a nuclear marker, we find that hybridization is frequent and is not restricted to the sympatric region. The gene flow is mainly from *S. rufescens* to *S. merianae*, with introgression into the range of *S. merianae*. Also, *S. merianae* would have long been present in the area, while *S. rufescens* appears to be a recent colonizer. The climate contributes to the population structure of *S. merianae*, but not to that of *S. rufescens*. The niches occupied by *S. rufescens* in the hybrid zone and the non-hybrid zone are similar, while the niches of *S. merianae* are different. Our results do not fit previous models of hybrid zone stability, suggesting the need to develop new models that consider the evolutionary factors that can differentially affect parental species and hybrids.

ADDITIONAL KEYWORDS: phylogeography – sibling species – natural hybridization – mtDNA – population genetics – South America – lizards.

INTRODUCTION

Identifying the main factors that structure genetic variation across a lineage distribution can lead to understanding the genetic basis of local adaptation and the patterns of speciation in a spatiotemporal context. Genetic differentiation is strongly influenced by two processes: (1) the trade-off between gene flow and genetic drift and (2) the local adaptation favoured by natural selection. When the first process prevails, the level of gene flow decreases as the geographical distances among populations increase, which results in a pattern of isolation by distance (IBD). On the other hand, when different adaptive optima occur in an area, the degree of genetic differentiation is expected to correlate with the difference between local environments rather than with geographical distance, a pattern called isolation by environment (IBE) (Lee & Mitchell-Olds, 2011; Sexton *et al.*, 2014). Studying the genetic structure of a species within an environmental niche analysis framework allows evaluation of the influences of environmental features (climatic gradients, changes in topography, etc.) on the distribution of genetic variability, aside from the geographic distances among populations. Thus, environmental barriers that favour genetic divergence or contact zones that promote gene flow can be identified (Rissler & Apodaca, 2007; Kozak *et al.*, 2008; Chan *et al.*, 2011). However, these studies become

*Corresponding Author. E-mail: imanolcabana@gmail.com

more complex when hybridization occurs, because this process may result in the transfer of alleles between species. In lineages that have the ability to hybridize only if their niches are similar, at least in a small portion, gene flow can take place, thus creating a hybrid zone (Cicero, 2004). In general, hybrid zones are regions of secondary contact in which each species presents a density gradient decreasing from its territory in allopatry towards the territory of the other species over relatively short geographic distances (Payseur, 2010). Furthermore, these regions often coincide with ecotones between major bioclimatic domains (Cicero, 2004; Tarroso *et al.*, 2014). Several models, which can be clustered in two groups, have been proposed to explain the stability of hybrid zones. One group includes the geographical selection-gradient model (Endler, 1977; Barton & Hewitt, 1985) and the bounded superiority model (originally named hybrid-superiority hypothesis) (Moore, 1977; Moore & Price, 1993), which assume that the position, size and maintenance of the hybrid zone is determined by environmental features (exogenous selection). In these models, the parental species are adapted to different environments in allopatry and give rise to a zone of secondary contact after the expansion of their ranges. Hence, the niche similarity between the parental species is the main factor shaping the hybrid zone. These models differ between them in how selection acts on hybrids. In the first one, selection acts against hybrids in transitional habitats; whereas in the second one, it favours these (Moore, 1977). The other group includes the model called tension zone that proposes an equilibrium between two opposing forces: dispersal of the parental species towards the hybrid zone and selection against hybrids (Key, 1968; Moore, 1977; Barton & Hewitt, 1985). In this, the hybrid zone is originated through secondary contact as well, but its maintenance depends solely on endogenous selection (hybrid inviability) and is not associated with environmental features. Hence, to quantify the environmental niches of both species and to decompose the overlapping between them allows an assessment as to how environmental features are associated with the hybrid zone and therefore which model better explains its stability. There are two types of methods used for this: one is based on species distribution models (SDMs) and the other is based on ordination techniques (Broennimann *et al.*, 2012). With simulated data, Broennimann *et al.* (2012) found that SDMs overestimate the levels of niche overlap and they suggested that ordination techniques are more appropriate than SDMs to assess niche similarity and their overlap in geographic space. Ordination techniques summarize the entire range of climatic variability of the study area in a multivariate space, where the occurrences of the species are then projected, allowing the direct comparison of species–environment

relationships in that environmental space. This makes it more precise and less prone to artificial maximization of ecologically irrelevant differences between species distributions (Broennimann *et al.*, 2012).

Natural hybridization between the lizards *Salvator merianae* (Duméril & Bibron, 1983) and *Salvator rufescens* (Günther, 1871) was demonstrated in a molecular phylogenetic study (Cabaña *et al.*, 2014). None of the hybrids found in that study were F1 hybrids, but they originated by backcrossing, thus demonstrating that introgression was occurring. Furthermore, all hybrids did not show the same mitochondrial haplotype, suggesting multiple hybridization events. These two lizards occur in different ecoregions (*S. merianae* in the humid Chaco, the Paranaense rainforest, the Espinal and Pampas, and *S. rufescens* in the dry Chaco), but overlap in a narrow area in central Argentina, in the ecotone between the arid Chaco and the Espinal (Cardozo *et al.*, 2012). In this ecotone, Cabaña *et al.* (2014) found hybrid specimens that are phenotypically identical to either one of the parental species. The species have a markedly different coloration: *S. merianae* is dark olive green, sometimes almost black; whereas *S. rufescens* is reddish.

To understand the main factors that contribute to the genetic variation of these two lizard species and the stability of their hybrid zone, we analysed the genetic structure of *S. merianae* and *S. rufescens* within a climatic niche analysis framework. Since hybridization is highly context-dependent (i.e. population size, sex ratio, intra- and interspecific communication, environmental condition, etc.; Abbott *et al.*, 2013) we worked at the microscale in a region that includes a climatic cline from east to west, covering two ecoregions (from Espinal to dry Chaco, respectively) and the ecotone between them. The region harbours populations of *S. merianae* alone, sympatric populations and populations of *S. rufescens* alone. Furthermore, we aim to (1) test the presence of an isolation by distance and/or an isolation by environment pattern, (2) infer the existence of barriers to gene flow within and between species and (3) assess which model (exogenous or endogenous selection) better explains the stability of the hybrid zone. Since the contact zone corresponds to the ecotone, we postulate that climatic niche divergence originates barriers to dispersal among populations and hence the hybrid zone is maintained by exogenous selection.

MATERIAL AND METHODS

SAMPLING AND LABORATORY PROTOCOLS

The study area is located in central Argentina, between 29° and 34° S (latitude) and 61° and 66° W (longitude). We obtained samples (muscle tissue or scales) of

75 *S. merianae* and 61 *S. rufescens* individuals, phenotypically classified on the basis of their coloration according to Cei (1993), who established that this is a valid character to differentiate them: *S. merianae* is dark olive green, sometimes almost black, and *S. rufescens* is reddish. Individuals were caught between October and March from 2008 to 2016 from wild populations. Commercial harvest of these species is permitted in Argentina (Naretto *et al.*, 2016). This research was approved by the Ethical Committee of the Instituto de Diversidad y Ecología Animal (IDEA, CONICET-UNC; resolution number: 6/2018) and authorized by government environmental agencies. All tissue samples were deposited in the Laboratorio de Biología del Comportamiento (IDEA, CONICET-UNC).

We extracted genomic DNA from ethanol (70–90%) preserved tissues using a saline extraction method (Bruford *et al.*, 1992). We sequenced a fragment of the nicotinamide adenine dinucleotide dehydrogenase subunit 4 gene (*ND4*) as mitochondrial DNA (mtDNA) marker and a fragment of the *α*-cardiac-actin intron 4 gene (*ACA4*) as nuclear DNA (nDNA) marker. Both markers show polymorphism at a local scale in these species (Cabaña *et al.*, 2014). We used the same amplification conditions as in that study, with the specific primers described by Forstner *et al.* (1995) for the *ND4* gene and by Waltari & Edwards (2002) for the *ACA4* gene.

All sequencing reactions were performed by Macrogen Inc. (Seoul, South Korea http://www.macrogen.com) in the 3' to 5' direction. Chromatograms were examined using Chromas Lite v.2.01 (http://technelysium.com.au/wp/chromas/). Sequences were aligned by MUSCLE using its default parameters (Edgar, 2004). For the *ND4* gene, sequences were translated into amino acids to confirm alignment. For the *ACA4* gene, we used PHASE v.2.1.1 (Stephens *et al.*, 2001) to determine the most probable pair of alleles for each sequence. The input file used in this software was obtained from SeqPHASE (Flot, 2010). The sequences of mtDNA and nDNA were separately analysed, since they can present incongruent phylogenetic signals between them due to the processes of hybridization and introgression.

POPULATION STRUCTURE ANALYSES

Haplotype networks were constructed for each gene following statistical parsimony (TCS) (Clement *et al.*, 2000), using PopArt v.1.7 (Leigh & Bryant, 2015). All sequences were analysed according to the cluster in which they grouped, independently from the phenotype of the individual the sequence belongs to. This was done to avoid problems posed by sequences of hybrids.

The number of polymorphic sites (S), number of haplotypes (K), haplotype diversity (Hd) and nucleotide diversity (π) were calculated using DnaSP

v.5.0 (Librado & Rozas, 2009). We identified as hybrid those individuals that presented inconsistencies among phenotype, mtDNA and/or nDNA, as in Cabaña *et al.* (2014). Since the delimitation of the hybrid zone (HZ) can influence the analyses of genetic structure and environmental divergence, we grouped populations in three scenarios following different criteria. The first scenario, called strict, included only sympatric populations in the HZ and only allopatric populations in the non-hybrid zone (NHZ). The second scenario, called genetic, was based on the genetic identity of individuals of each locality. Therefore, in the HZ of the first scenario, we added the allopatric localities where hybrids were found. The third scenario, called relaxed, considered possible gene flow among close localities. Hence, allopatric localities near the boundaries of the HZ of the second scenario were added to the HZ of this scenario. Thus, the HZ of each scenario was nested in the next one.

A hierarchical analysis of population structure (AMOVA) was made to partition the total genetic variation into variance between the HZ and the NHZ, among populations within zones and within populations. The AMOVA was performed using ARLEQUIN v.3.5 (Excoffier & Lischer, 2010), and its significance was tested with 10 000 permutations. Isolation by distance was tested by the correlation between the matrices of normalized genetic distance ($F_{ST}/(1–F_{ST})$) and that of the logarithm of geographical distance (Km) between populations (Rousset, 1997), using IBDWS (Jensen *et al.*, 2005). IBD was tested within each species and within the HZ and the NHZ in the three scenarios.

DEMOGRAPHIC INFERENCE

To detect the occurrence of events of population expansion, we performed Fu's F_s (Fu, 1997) and Tajima's *D* (Tajima, 1989) neutrality tests, using DnaSP v.5 (Librado & Rozas, 2009) and mismatch distribution analyses under a demographic expansion model (Rogers & Harpending, 1992; Harpending *et al.*, 1998). In the latter, both the deviations from the sum of squares between observed and expected mismatch distributions (SSD) and the Harpending inequality index (rg) were considered in order to test whether the data fit the demographic expansion model. A Bayesian skyline plot analysis (BSP) was performed with mtDNA using BEAST v.2.5.1 (Drummond & Rambaut, 2007). We used a strict molecular clock and the best-fitting model of molecular evolution for each lineage (HKY for *S. merianae* and HKY + I for *S. rufescens*). The mtDNA substitution rate used to calibrate the molecular clock was 0.00511 substitutions/site/Myr (Oliveira *et al.*, 2015). We ran 1×10^7 iterations with sampling every 1000 iterations and discarded the

initial 10% as burn-in. Results were visualized using TRACER v.1.6 (Rambaut *et al.*, 2018).

ENVIRONMENTAL NICHE ANALYSES

We performed the PCA-env analysis using the function 'ecospat.niche.dyn.index' in 'ecospat' R package to decompose each niche in an overlapped and non-overlapped proportion in relation to another niche (Broennimann *et al.*, 2012; Petitpierre *et al.*, 2012; Di Cola *et al.*, 2017). Hence, the value of the overlapped proportion will depend on which niche of the pair was analysed. We compared the niche of both species in the study area and the niche in the HZ and the NHZ of the three scenarios within each species. In the comparison of both species, we only used individuals genetically characterized here as non-hybrid, since hybrids could use different niches from parental species (Theodoridis *et al.*, 2013; Parisod & Broennimann, 2016). We measured niches along the two first axes of a Principal Component Analysis, previously calibrated using the climatic values from all the pixels in the study area. Calibration was made using the 19 bioclimatic variables from WorldClim v.1.4 (Supporting Information, Table S1; www.worldclim.org) at a resolution of 30 arc-seconds (~1 km) (Hijmans *et al.*, 2005). Differences in the position of individuals in the PCA-env reflect differences in their environmental niches. All the analyses were done in R v.3.3.1 (R Development Core Team, 2016).

ISOLATION BY ENVIRONMENT

To assess whether genetic variation across populations is associated with divergence in the climatic niche (IBE pattern), partial Mantel tests were conducted by comparing matrices of genetic distance (F_{ST}) vs. environmental distance, as implemented in the package 'Vegan' (Oksanen *et al.*, 2016). The latter was calculated as a Euclidean distance using the function 'dist' in the package 'Stat'. We also performed a distance-based redundancy analysis (db-RDA), since it is a more powerful and informative method of spatial analysis than Mantel tests (Legendre & Fortin, 2010; Legendre *et al.*, 2015; Wanderley *et al.*, 2018). The db-RDA was performed by using the function 'capscale' in the package 'Vegan'. We used Tamura & Nei (1993) as genetic distance. We performed Pearson correlation tests between the bioclimatic variables in the study area. For pairs of variables that were highly correlated ($r \geq 0.7$), the variable considered easier to interpret biologically was chosen to calculate the environmental distance. For each species, we ran a model conditioned by geographic coordinates. The significance of the model was obtained after 9999 permutations. All

analyses were performed in R v.3.3.1 (R Development Core Team, 2016).

RESULTS

We analysed 75 individuals phenotypically classified as *S. merianae* from 15 localities and 61 individuals classified as *S. rufescens* from 14 localities (Figs 1, 2). An 807-bp fragment of the *ND4* gene and a 413-bp fragment of the *ACA4* gene were sequenced for each specimen. Species, specimens, location and GenBank accession numbers are listed in the Supporting Information (Table S2). Some sampled individuals were previously analysed in Cabaña *et al.* (2014; Supporting Information, Table S2).

POPULATION STRUCTURE ANALYSES

For the mitochondrial gene, 63 sequences resulted in 12 haplotypes in *S. merianae* and 73 sequences in 15 haplotypes in *S. rufescens*. The nuclear sequences revealed five alleles in *S. merianae* for 146 sequences and four alleles in *S. rufescens* for 124 sequences. Because the latter species exhibited a deletion of one nucleotide in the *ACA4* gene, alleles can be assigned without ambiguity to one of the two parental lineages. The networks show two highly divergent clusters corresponding to *S. merianae* and *S. rufescens* (Figs 1, 2). Polymorphism indices for each species and each gene are shown in the Supporting Information (Table S3).

We found a total of 15 hybrids, of which 14 were phenotypically identified as *S. merianae*: 12 individuals with mtDNA of *S. rufescens* and nDNA of *S. merianae*; one with mtDNA and nDNA of *S. rufescens*; and the last one with mtDNA of *S. merianae* and nDNA of *S. rufescens*. The remaining individual was phenotypically identified as *S. rufescens* with mtDNA of *S. merianae* and nDNA of *S. rufescens*. Hybrids present the most frequent variant, in both mitochondrial and nuclear DNA (Figs 1, 2).

Based on the species distribution in the study area and on the individual genetic identification, we assigned localities to the HZ or the NHZ in the three scenarios described above (Fig. 3). In *S. rufescens*, HZ and NHZ for strict and genetic scenarios included the same localities, since hybrids are not found in localities where only this species occurs.

When the genetic structure was analysed considering the three scenarios posed, the mitochondrial networks showed few common haplotypes between regions and several exclusive ones. In the nuclear networks, exclusive alleles were observed in the NHZ for *S. rufescens*; whereas in *S. merianae*, all the alleles were present in both zones (Supporting Information, Fig. S1).
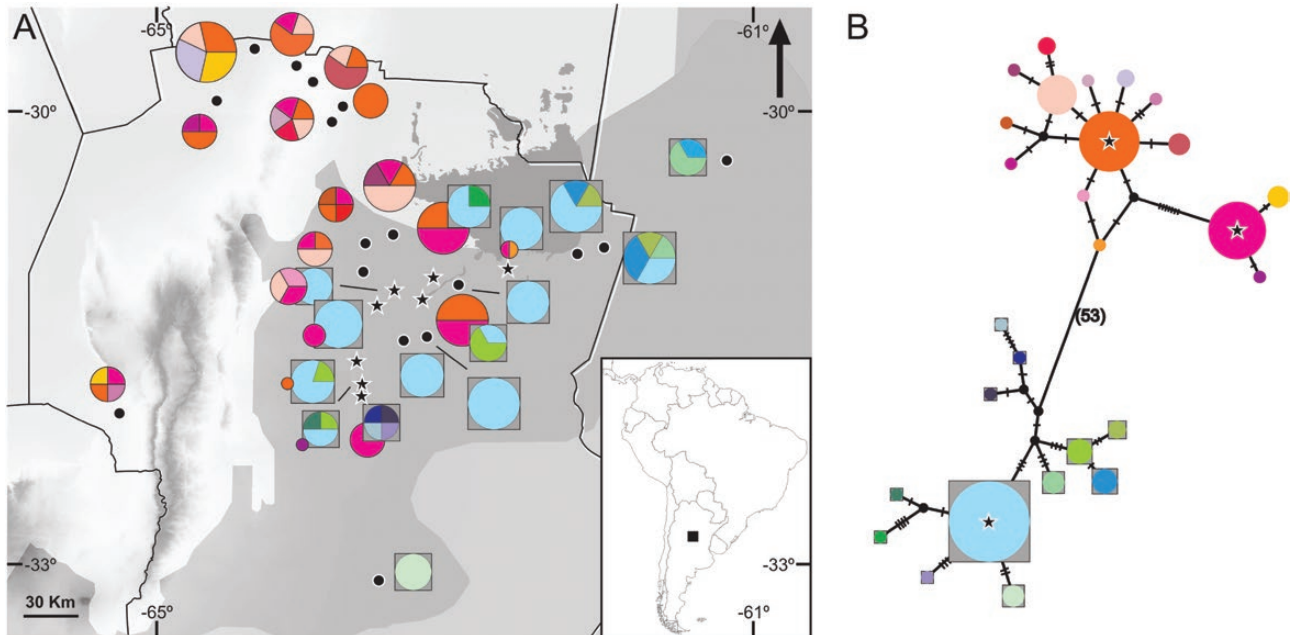
Figure 1. Mitochondrial haplotypes distribution and network. A, the circle size represents the number of individuals analysed in each locality (circles: individuals of *S. rufescens*; circles in boxes: individuals of *S. merianae*). The colours in the circles indicated the different haplotypes and their respective frequency in the locality. Those localities in which hybrids were found are represented by stars. The rest of them are marked by dots. The grey background represents the biogeographic regions (light grey: Arid Chaco; dark grey: Espinal). B, in the haplotype network, the circle size represents the frequency in the total sample (circles: haplotypes of *S. rufescens*; circles in boxes: haplotypes of *S. merianae*). Each bar between two circles represents one mutation step; the black dots mark the haplotypes that have not been found in the data set. The haplotypes of hybrid individuals are indicated by stars.

Since the *ACA4* gene presented low levels of variability, we did not perform the AMOVA analysis, mismatch distribution, neutrality tests and Mantel tests for this dataset. For the mtDNA data, the hierarchical AMOVA for the three scenarios showed a significant difference between the HZ and the NHZ in *S. rufescens* ($P = 0.006$ relaxed, $P = 0.001$ genetic, $P = 0.009$ strict $P = 0.009$). For *S. merianae*, we found a significant difference between zones only in the relaxed scenario ($P = 0.022$). In this scenario, the number of populations in each zone was unbalanced (11 populations from the HZ vs. 4 from the NHZ). To test if this difference was not the result of such unbalance, we compared through AMOVA the four populations from the NHZ with randomly subsampled four populations from the HZ. We found that, in most cases (90%), the difference between HZ and NHZ was not significant.

When all localities were considered, an IBD pattern was detected only in *S. merianae* ($R^2 = 0.177$, $P = 0.0129$; Supporting Information, Fig. S2A). We also tested for the presence of an IBD pattern in the HZ and the NHZ in the scenarios in which significant differences in the AMOVA analyses were found. Only the populations

of *S. merianae* in the NHZ of the relaxed scenario presented an IBD pattern ($R^2 = 0.698$, $P = 0.0330$; Supporting Information, Fig. S2B).

DEMOGRAPHIC INFERENCE

Neutrality tests were not significantly different from the null hypothesis of constant size for both species (Supporting Information, Table S4). The mismatch distributions analyses were clearly multimodal, but rg and SSD were not significant, thus the null hypothesis of population expansion is not rejected. Demographic analyses were performed separately for HZ and NHZ in the scenarios in which significant differences in the AMOVA analyses were observed. In general, none of them were significant. The mismatch distributions were multimodal in both zones, in all scenarios for both species (Supporting Information, Fig. S3); rg and/or SSD values were significant in the HZ of the three scenarios of *S. rufescens* and the NHZ of the relaxed scenario of *S. merianae*. The BSP analysis showed that *S. merianae* experienced a decrease in population size around 50 000 years ago
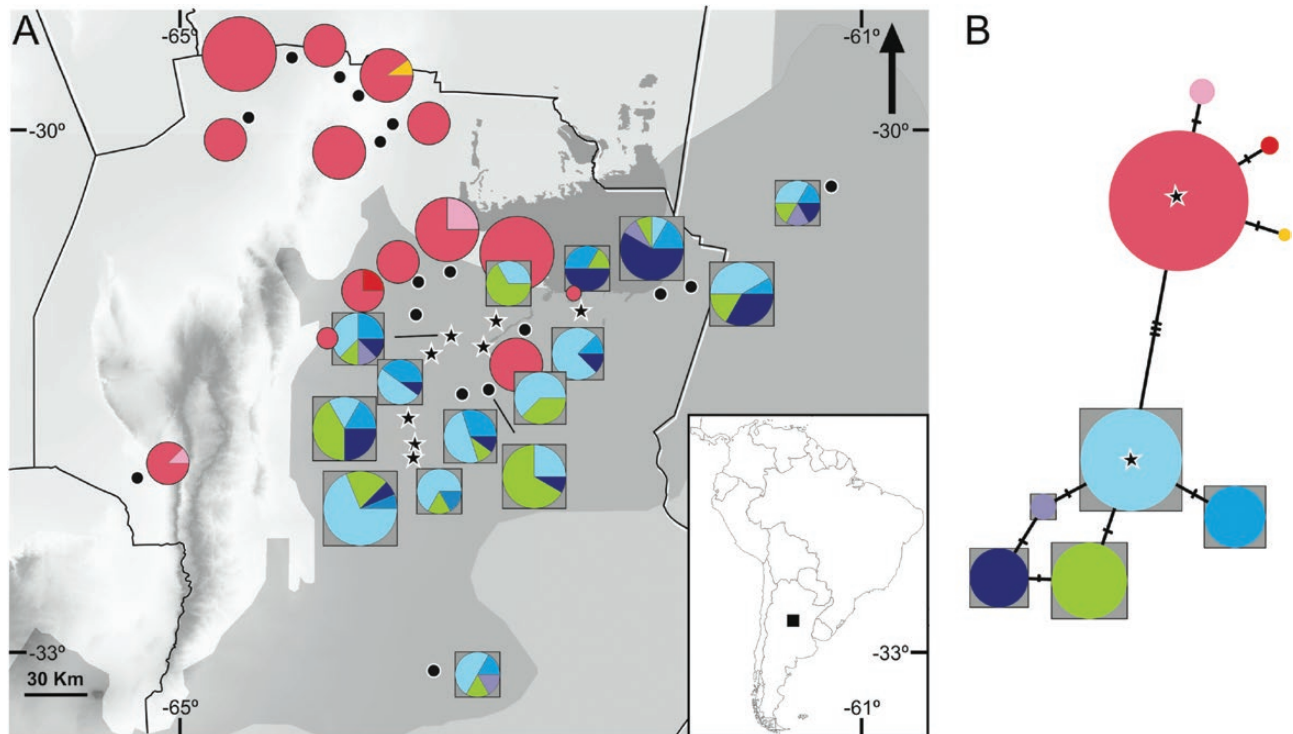
Figure 2. Nuclear haplotypes distribution and network. A, the circle size represents the number of individuals analysed in each locality (circles: individuals of *S. rufescens*; circles in boxes: individuals of *S. merianae*). The colours in the circles indicated the different haplotypes and their respective frequency in the locality. Those localities in which hybrids were found are represented by stars. The rest of them are marked by dots. The grey background represents the biogeographic regions (light grey: Arid Chaco; dark grey: Espinal). B, in the haplotype network, the circle size represents the frequency in the total sample (circles: haplotypes of *S. rufescens*; circles in boxes: haplotypes of *S. merianae*). Each bar between two circles represents one mutation step. The haplotypes of hybrid individuals are indicated by stars.
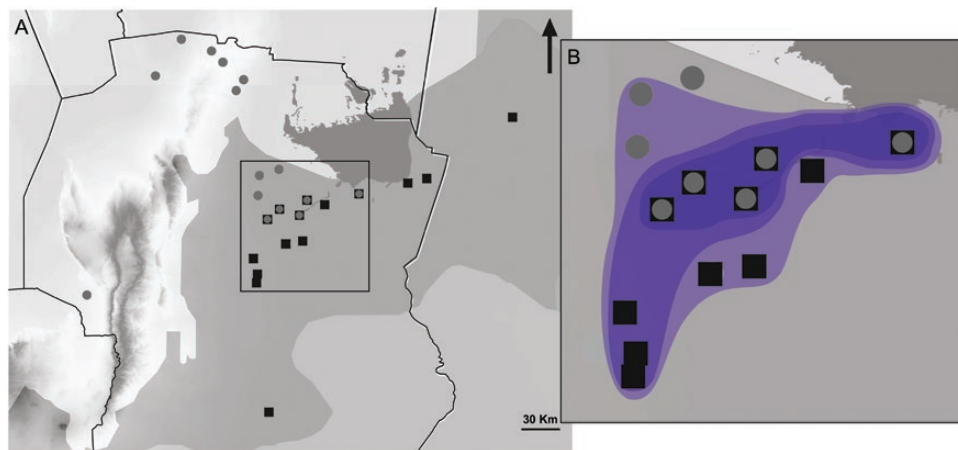

Figure 3. Sample distribution. A, population distribution of *S. merianae* (black squares) and *S. rufescens* (grey circles). B, zoom in the contact zone showing the three proposed scenarios (strict, genetic and relax from dark blue to light blue), each nested in the next one. Grey background represents the biogeographic regions (light grey: Arid Chaco; dark grey: Espinal).

(Fig. 4A), whereas *S. rufescens*, showed an increase in population size through time, started around 100 000 years ago (Fig. 4B).

Neutrality tests, mismatch distribution analysis and BSP for *S. rufescens* showed discordant results. According to the mitochondrial haplotype network
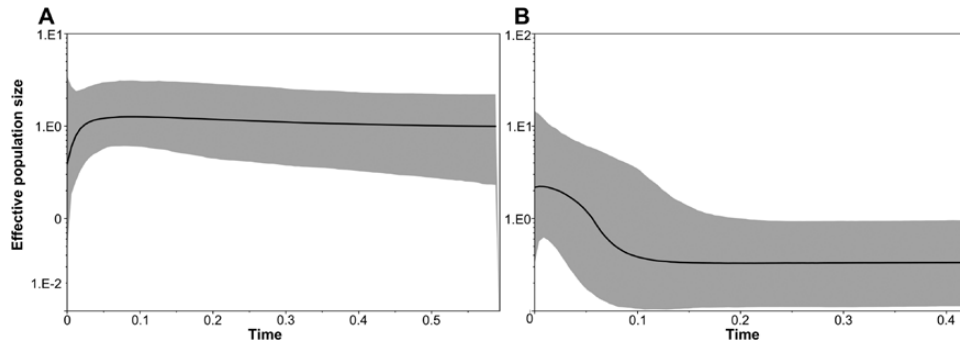
Figure 4. Bayesian skyline plots illustrating effective population sizes through time (Myr) of *S. merianae* (A) and of *S. rufescens* (B). The black line represents the median population size, and the grey area represents a 95% higher posterior probability.

(Fig. 1B), two haplogroups can be defined. Because the divergence between these two haplogroups may indicate different evolutive histories, we performed the analyses of demographic inferences for each of them. Mismatch distributions were unimodal in both cases (Supporting Information, Fig. S4), so the hypothesis of population expansion cannot be rejected. However, Tajima's *D* and Fu's F_S were negative and significant only for one haplogroup (D: –1.783; F_S: –7.144), thus suggesting demographic expansion. BSP analyses for each haplogroup could not be performed due to the small sample size.

ENVIRONMENTAL NICHE ANALYSES

When comparing the niches of the two species, the first two axes of the PCA explained 78% of the climatic variability (Fig. 5) in the study area. According to the environmental niche analyses, 0.93% of the niche of *S. merianae* overlaps with 33% of the niche of *S. rufescens* (OvSm and OvSr, respectively). The climatic niche space of *S. merianae* corresponded to higher values of precipitation than that of *S. rufescens*, whose environmental niche space is characterized by higher temperatures and a marked seasonality in temperature and precipitations than that of *S. merianae* (Fig. 5; Supporting Information, Table S1).

Regarding the comparison of the climate niche between zones within each species, in the first two scenarios of *S. merianae* (Fig. 6), the proportion of the HZ niche that overlapped with the NHZ (OvZH) is 1. Conversely, the overlap of the NHZ niche with the HZ (OvNHZ) is considerably lower (strict scenario: OvHZ = 100% and OvNHZ = 41%; genetic scenario: OvHZ = 100% and OvNHZ = 38%; relaxed scenario: OvHZ = 19% and OvNHZ = 6%; Fig. 6). In *S. rufescens* (Fig. 6), the niche indices showed that the HZ completely overlaps (OvHZ = 100%) with a small portion of the niche used in NHZ in all scenarios

(OvNHZ = 8%, OvNHZ = 13% and OvNHZ = 12% for strict, genetic and relaxed scenarios, respectively).

ISOLATION BY ENVIRONMENT

We did not find any significant IBE pattern in *S. rufescens* and *S. merianae* when performing a partial Mantel test. The bioclimatic variables selected to perform the db-RDA analysis were the annual mean temperature, the mean diurnal range of temperature, temperature seasonality, the minimum temperature of the coldest month and the annual precipitation. The dbRDA indicated that the model conditioned by the geographic coordinates was significant in *S. merianae* (*P* = 0.0032), and that the climatic variable that explains that model is the mean diurnal range of temperature (*P* = 0.0004). In *S. rufescens*, the model was not significant (*P* = 0.3).

DISCUSSION

In this study, we analysed the factors structuring the populations of two closely related species of lizards along an environmental cline, encompassing allopatric regions and a hybrid zone. The levels of genetic diversity were similar in both species (Supporting Information, Table S3). The mitochondrial haplotypes networks showed one or few widespread haplotypes, along with closely related haplotypes that were either exclusive to a population or shared by a few populations, generally nearby (Fig. 1). This pattern, called a star pattern, would indicate that *S. merianae* and *S. rufescens* have undergone an expansion of geographical range and that their populations subsequently began to differentiate themselves by the action of genetic drift, so they currently have low or moderate levels of gene flow. However, this pattern appeared more marked in *S. rufescens* than
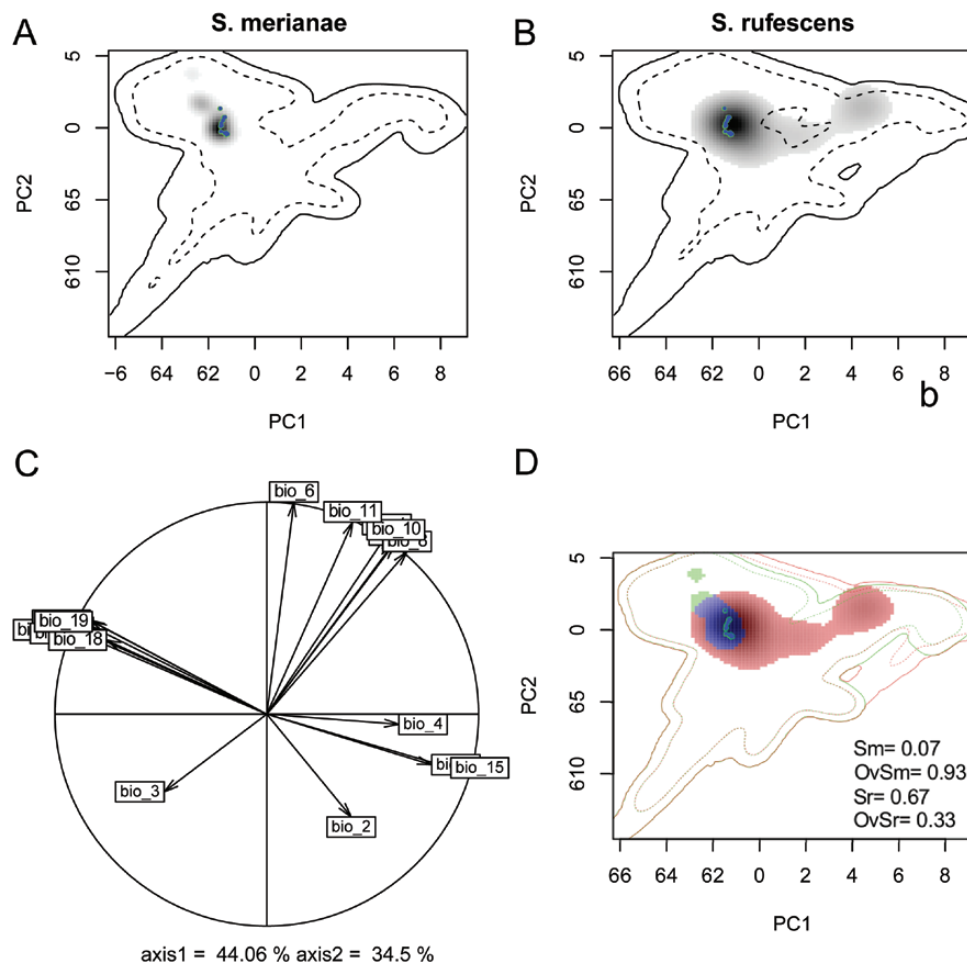
Figure 5. Niches of *S. merianae* (A) and *S. rufescens* (B) based on the PCA-env. Grey shading in the niches shows the density of individual occurrences by cell. The solid and dashed contour lines illustrate 100% and 50%, respectively, of the available environment (background). Blue dots represent the position of the hybrids in the environmental space. C, contribution of the environmental variables on the 1 and 2 axes of the PCA and percentage explained by those axes (variables are coded as in Supporting Information, Table S1). D, niche overlap (blue) of A (green) and B (red). Sm and OvSm: non-overlap and overlap proportion of the niche of *S. merianae*, respectively; Sr and OvSr: non-overlap and overlap proportion of the niche of *S. rufescens*, respectively.

in *S. merianae*. Nuclear networks showed haplotypes separated by fewer mutational steps than the mitochondrial ones, which is an expected outcome, given the lower mutation rate of nuclear markers (Palumbi *et al.*, 2001). Both the multimodal pattern of the mismatch analyses and the non-significant neutrality tests would indicate a constant population size for the two species (Supporting Information, Fig. S3, Table S3). However, the BSP analysis suggests that *S. rufescens* would have experienced an increase in population size. This inconsistency could be due to the fact that the populations of *S. rufescens* originated from two different lineages (haplogroups), and hence demographic analyses show the mix of their evolutionary histories (Malhi *et al.*, 2001). When the haplogroups were analysed separately, the results were consistent with a population expansion pattern (Supporting Information, Fig. S4). On the other hand, *S. merianae* would have experienced a slight population decrease according to the BSP analysis, a pattern consistent with the observed discontinuities among haplotypes in the network. As for whether geographic and/or environmental distances were influencing the population genetic structure of *S. merianae* and *S. rufescens*, we found that the former presents an IBD pattern, which indicates an equilibrium between gene flow and genetic drift, and an IBE pattern indicating that climate also plays a role in structuring their populations in the study area. On the other hand, *S. rufescens* did not present an IBD or an IBE pattern, and this could be due to its recent expansion and the fact that its populations originated from two haplogroups.
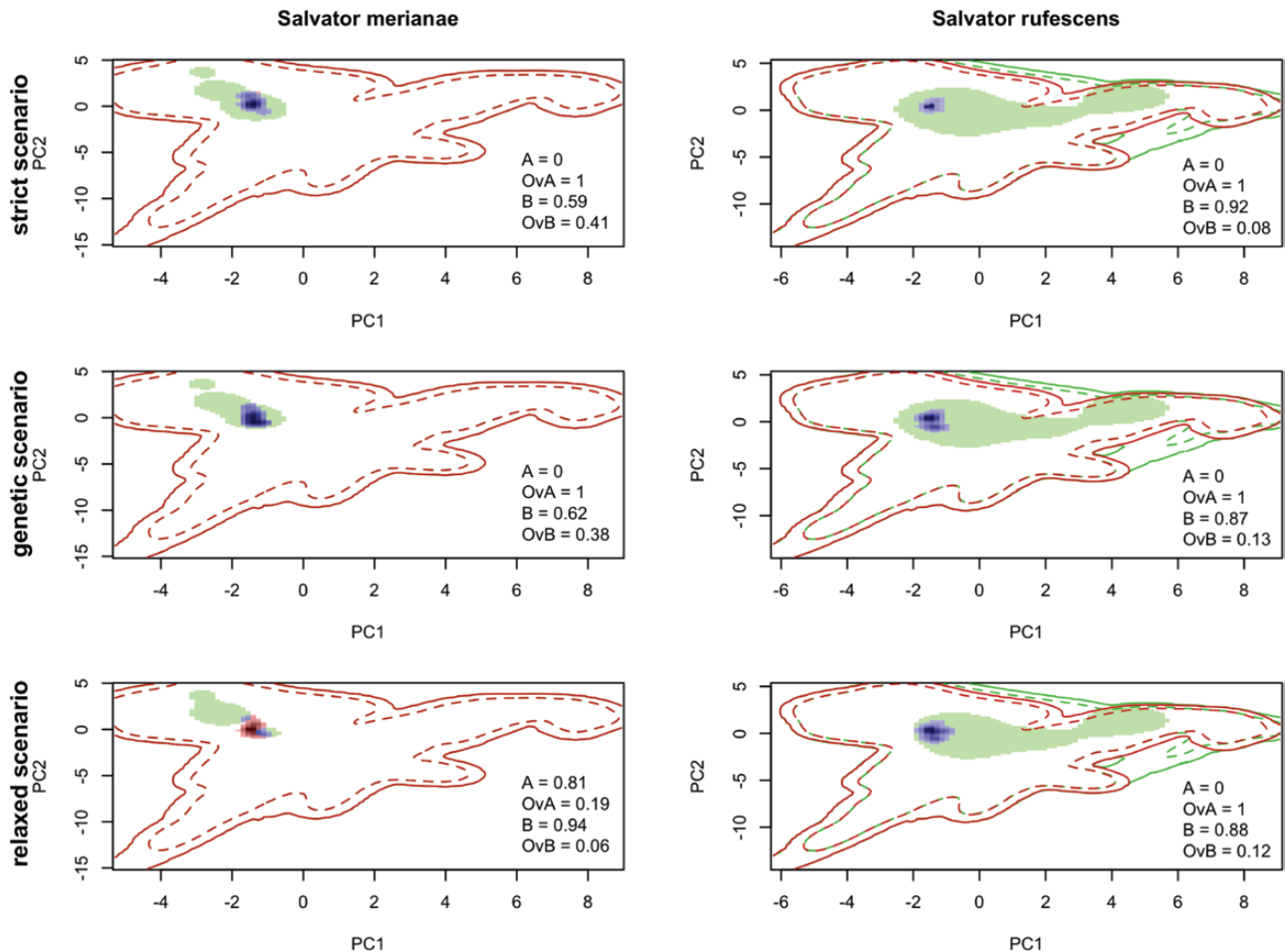
Figure 6. Niche overlap between hybrid and non-hybrid zone of *Salvator merianae* and *Salvator rufescens* based on the PCA-env. Niche overlap (blue) between the hybrid zone (HZ, red) and non-hybrid zone (NHZ, green) in three proposed scenarios for *S. merianae* and *S. rufescens* in the environmental space. Darker shading in niches shows the density of the occurrences of the species by cell. The solid and dashed contour lines illustrate, respectively, 100% and 50% of the available environment (background). HZ and OvHZ: non-overlap and overlap proportion of HZ, respectively; NHZ and OvNHZ: non-overlap and overlap proportion of NHZ, respectively.

The climatic niche analysis allowed us to quantify and decompose the niches of *S. rufescens* and *S. merianae*. This approach was designed and used in the context of invasive species (Guisan *et al.*, 2014; Di Cola *et al.*, 2017). However, in this study it is applied for the first time to compare species in a hybrid framework. When comparing the niches of both species, almost the entire niche of *S. merianae* overlaps with a small part of the niche of *S. rufescens*. This indicates that in the study area, *S. rufescens* occupies a broader range of climatic conditions than *S. merianae*, whose niche is nested in that of *S. rufescens*.

HYBRID ZONE

The three scenarios were thought to be a gradient from a restrictive view (strict scenario) to a relaxed view (relaxed scenario) of the hybrid zone. The haplotype networks and distributions (Supporting Information, Fig. S3) showed differences in haplotype frequencies and diversity between HZ and NHZ in both species. In *S. merianae*, the genetic difference between zones was only significant in the relaxed scenario, which proposes the idea of gene flow between the HZ and nearby populations. In this scenario, we also found different climatic niche spaces between the HZ and the NHZ. *Salvator rufescens* showed marked genetic differences in the three scenarios, but the climatic niche space in the HZ was nested in that of the NHZ in all scenarios, i.e. *S. rufescens* did not change its niche between zones and the species did not present an IBE pattern. Thus, climatic conditions do not explain the genetic differences between zones in *S. rufescens*.

The existence of a hybrid zone between these two lizard species was demonstrated in Cabaña *et al.* (2014), but its characteristics and the processes that maintain it were not analysed in that study. All the hybrids found in our study, as well as those detected in Cabaña *et al.* (2014), originated by backcrossing. The frequency of hybrids found here is high (15 out of 136 studied individuals) considering that the sampling included individuals from the NHZ, so they did not have the chance to hybridize. Furthermore, some backcrossed individuals may not have been detected, because they did not show inconsistencies among the phenotype and the two genes analysed. The hybrids were not exclusively restricted to the sympatry zone (Figs 1, 2; Supporting Information, Fig. S1), having most of them the mtDNA from *S. rufescens* and the nDNA from *S. merianae*, i.e. females of *S. rufescens* mated with males of *S. merianae*. Thus, despite the bidirectionality of hybridization in our study, this phenomenon was shown to be mainly asymmetric, in accordance with Cabaña *et al.* (2014). The difference in the parental species density could be the cause of different outputs in the process of hybridization. In a review, Wirtz (1999) found that in most of the analysed cases of bidirectional hybridization, the hybrids presented mainly mtDNA from one of the parental species, called the mother species, because females of the rare species tended to mate with males of the common species. In agreement with this, the two hybrids with *S. merianae* mtDNA were from the same locality, the only one where *S. rufescens* is predominant (I. Cabaña, personal communication).

Given the climatic characteristics of the study area, where the HZ corresponds with an ecotone between two ecoregions, we expected to find that the environment played a primary role in the maintenance of the hybrid zone. Although *S. merianae* used different environmental niches in the HZ and the NHZ, *S. rufescens* used the same niche space in both zones, indicating that, at least for the latter, environmental characteristics did not play the expected role under the geographical selection gradient model. In addition, the bounded superiority model did not fit, since both parental species and hybrids coexisted in the HZ, with parental species the most abundant. Gramlich & Hörandl (2016) discarded this model based on a similar result in their study. On the other hand, the hybrid zone stability can be associated with endogenous selection, i.e. with individual inherent characteristics. However, the fact that all the hybrids found originated by backcrossing, contradicts the idea of selection against them, as well as the prevention of introgression assumed by the tension zone model.

CONCLUSION

There are several possible outcomes of hybridization, depending on the genetic constitution, demography, ecology and spatial distribution of the species involved, all of which are highly variable (Abbott *et al.*, 2013). Working at a microscale allows us to reduce part of the variability given by these factors. It is important to disentangle the contribution of every single factor to hybridization dynamics. In this study, we aimed to understand the role played by the climate as a factor that structures populations and promotes hybridization. *Salvator merianae* and *S. rufescens* hybridization constitutes an excellent model to elucidate the ecological mechanisms that may contribute to the breakdown of the species recognition system and subsequently lead to hybridization and introgression in closely related taxa. Our results do not fit previous models of hybrid zone stability, suggesting the need for developing new models that consider that the evolutionary factors can differentially affect parental species and hybrids.

REFERENCES

Abbott R, **Albach D**, **Ansell S**, **Arntzen JW**, **Baird SJE**, **Bierne N**, **Boughman J**, **Brelsford A**, **Buerkle CA**, **Buggs R**, **Butlin RK**, **Dieckmann U**, **Eroukhmanoff F**, **Grill A**, **Cahan SH**, **Hermansen JS**, **Hewitt G**, **Hudson AG**, **Jiggins C**, **Jones J**, **Keller B**, **Marczewski T**, **Mallet J**, **Martinez-Rodriguez P**, **Möst M**, **Mullen S**, **Nichols R**, **Nolte AW**, **Parisod C**, **Pfennig K**, **Rice AM**, **Ritchie MG**, **Seifert B**, **Smadja CM**, **Stelkens R**, **Szymura JM**, **Väinölä R**, **Wolf JBW**, **Zinner D. 2013.** Hybridization and speciation. *Journal of Evolutionary Biology* **26:** 229–246.

Barton NH, **Hewitt GM. 1985.** Analysis of hybrid zones. *Annual Review of Ecology and Systematics* **16:** 113–148.

Broennimann O, **Fitzpatrick MC**, **Pearman PB**, **Petitpierre B**, **Pellissier L**, **Yoccoz NG**, **Thuiller W**, **Fortin MJ**, **Randin C**, **Zimmermann NE**, **Graham CH**, **Guisan A. 2012.** Measuring ecological niche overlap from occurrence and spatial environmental data. *Global Ecology and Biogeography* **21:** 481–497.

Bruford ME, **Hanotte O**, **Brookfield J**, **Burke T. 1992.** Single-locus and multilocus DNA fingerprinting. In: Hoelzel AR, ed. *Molecular genetic analysis of populations: a practical approach*. Oxford, New York: Oxford University Press.

Cabaña I, **Gardenal CN**, **Chiaraviglio M**, **Rivera PC. 2014.** Natural hybridization in lizards of the genus *Tupinambis* (Teiidae) in the southernmost contact zone of their distribution range. *Annales Zoologici Fennici* **51:** 340–348.

Cardozo G, **Naretto S**, **Zak M**, **Chiaraviglio M. 2012.** The role of landscape in contact zones of sister species of lizards. In: Tiefenbacher J, ed. *Perspectives on nature conservation – patterns, pressures and prospects*. London: IntechOpen, 161–176.

Cei JM. 1993. *Reptiles del noroeste, nordeste y este de la Argentina, herpetofauna de las selvas subtropicales, puna y pampas. Monografía XIV*. Turin: Museo Regionale di Scienze Naturali, 949.

Chan LM, **Brown JL**, **Yoder AD. 2011.** Integrating statistical genetic and geospatial methods brings new power to phylogeography. *Molecular Phylogenetics and Evolution* **59:** 523–537.

Cicero C. 2004. Barriers to sympatry between avian sibling species (Paridae: Baeolophus) in local secondary contact. *Evolution* **58:** 1573–1587.

Clement M, **Posada D**, **Crandall KA. 2000.** TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9:** 1657–1659.

Di Cola V, **Broennimann O**, **Petitpierre B**, **Breiner FT**, **Amen MD**, **Randin C**, **Engler R**, **Pottier J**, **Pio D**, **Dubuis A**, **Pellissier L**, **Mateo G**, **Hordijk W**, **Salamin N**, **Guisan A. 2017.** Ecospat: an R package to support spatial analyses and modeling of species niches and distributions. *Ecography* **40:** 1–14.

Drummond AJ, **Rambaut A. 2007.** BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7:** 214.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32:** 1792–1797.

Endler JA. 1977. *Geographic variation, speciation and clines*. Princeton: Princeton University Press.

Excoffier L, **Lischer HEL. 2010.** Arlequin suite v.3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10:** 564–567.

Flot JF. 2010. seqphase: a web tool for interconverting phase input/output files and fasta sequence alignments. *Molecular Ecology Resources* **10:** 162–166.

Forstner MR, **Davis SK**, **Arévalo E. 1995.** Support for the hypothesis of anguimorph ancestry for the suborder Serpentes from phylogenetic analysis of mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution* **4:** 93–102.

Fu YX. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147:** 915–925.

Gramlich S, **Hörandl E. 2016.** Fitness of natural willow hybrids in a pioneer mosaic hybrid zone. *Ecology and Evolution* **6:** 7645–7655.

Guisan A, **Petitpierre B**, **Broennimann O**, **Daehler C**, **Kueffer C. 2014.** Unifying niche shift studies: insights from biological invasions. *Trends in Ecology & Evolution* **29:** 260–269.

Harpending HC, **Batzer MA**, **Gurven M**, **Jorde LB**, **Rogers AR**, **Sherry ST. 1998.** Genetic traces of ancient demography. *Proceedings of the National Academy of Sciences of the United States of America* **95:** 1961–1967.

Hijmans RJ, **Cameron SE**, **Parra JL**, **Jones PG**, **Jarvis A. 2005.** Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25:** 1965–1978.

Jensen JL, **Bohonak AJ**, **Kelley ST. 2005.** Isolation by distance, web service. *BMC Genetics* **6:** 1–6.

Key KHL. 1968. The concept of stasipatric speciation. *Systematic Biology* **17:** 14–22.

Kozak KH, **Graham CH**, **Wiens JJ. 2008.** Integrating GIS-based environmental data into evolutionary biology. *Trends in Ecology & Evolution* **23:** 141–148.

Lee CR, **Mitchell-Olds T. 2011.** Quantifying effects of environmental and geographical factors on patterns of genetic differentiation. *Molecular Ecology* **20:** 4631–4642.

Legendre P, **Fortin MJ. 2010.** Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources* **10:** 831–844.

Legendre P, **Fortin MJ**, **Borcard D. 2015.** Should the Mantel test be used in spatial analysis? *Methods in Ecology and Evolution* **6:** 1239–1247.

Leigh JW, **Bryant D. 2015.** POPART: full-feature software for haplotype network construction. *Methods in Ecology and Evolution* **6:** 1110–1116.

Librado P, **Rozas J. 2009.** DnaSP v.5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25:** 1451–1452.

Malhi RS, **Schultz BA**, **Smith DG. 2001.** Distribution of mitochondrial DNA lineages among native American tribes of Northeastern North America. *Human Biology* **73:** 17–55.

Moore WS. 1977. An evaluation of narrow hybrid zones in vertebrates. *The Quarterly Review of Biology* **52:** 263–277.

Moore WS, **Price JT. 1993.** Nature of selection in the northern flicker hybrid zone and its implications for speciation theory. In: Harrison RG, ed. *Hybrid zones and the evolutionary process*. Oxford, New York: Oxford University Press, 196–225.

Naretto S, **Blengini CS**, **Cardozo G**, **Chiaraviglio M. 2016.** Pre- and postcopulatory traits of *Salvator* male lizards in allopatry and sympatry. *Scientifica* **2016:** 8176267.

Oksanen J, **Blanchet FG**, **Friendly M**, **Kindt R**, **Legendre P**, **McGlinn D**, **Minchin PR**, **O'Hara RB**, **Simpson GL**, **Solymos P**, **M. Henry HS**, **Szoecs E**, **Wagner H. 2016.** *Vegan: an introduction to ordination, R package version 2.4-1.* Available https://CRAN.R-project.org/package=vegan (date last accessed, 10 October 2016).

Oliveira EF, **Gehara M**, **São-Pedro VA**, **Chen X**, **Myers EA**, **Burbrink FT**, **Mesquita DO**, **Garda AA**, **Colli GR**, **Rodrigues MT**, **Arias FJ**, **Zaher H**, **Santos RML**, **Costa GC. 2015.** Speciation with gene flow in whiptail lizards from a Neotropical xeric biome. *Molecular Ecology* **24:** 5957–5975.

Palumbi SR, Cipriano F, Hare MP. 2001. Predicting nuclear gene coalescence from mitochondrial data: the three-times rules. *Evolution* **55:** 859–868.

Parisod C, **Broennimann O. 2016.** Towards unified hypotheses of the impact of polyploidy on ecological niches. *The New Phytologist* **212:** 540–542.

Payseur BA. 2010. Using differential introgression in hybrid zones to identify genomic regions involved in speciation. *Molecular Ecology Resources* **10:** 806–820.

Petitpierre B, **Kueffer C**, **Broennimann O**, **Randin C**, **Daehler C**, **Guisan A. 2012.** Climatic niche shifts are rare among terrestrial plant invaders. *Science* **335:** 1344–1348.

R Development Core Team. 2016. *R: a language and environment for statistical computing.* Vienna: R Foundation for Statistical Computing. Available at: https://www.R-project.org/ (date last accessed, 10 October 2016).

Rambaut A, **Drummond AJ**, **Xie D**, **Baele G**, **Suchard MA. 2018.** Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Systematic Biology* **67:** 901–904.

Rissler LJ, **Apodaca JJ. 2007.** Adding more ecology into species delimitation: ecological niche models and phylogeography help define cryptic species in the black salamander (*Aneides flavipunctatus*). *Systematic Biology* **56:** 924–942.

Rogers AR, **Harpending H. 1992.** Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* **9:** 552–569.

Rousset F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* **145:** 1219–1228.

Sexton JP, **Hangartner SB**, **Hoffmann AA. 2014.** Genetic isolation by environment or distance: which pattern of gene flow is most common? *Evolution* **68:** 1–15.

Stephens M, **Smith NJ**, **Donnelly P. 2001.** A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics* **68:** 978–989.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123:** 585–595.

Tamura K, **Nei M. 1993.** Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* **10:** 512–526.

Tarroso P, **Pereira RJ**, **Martínez-Freiría F**, **Godinho R**, **Brito JC. 2014.** Hybridization at an ecotone: ecological and genetic barriers between three Iberian vipers. *Molecular Ecology* **23:** 1108–1123.

Theodoridis S, **Randin C**, **Broennimann O**, **Patsiou T**, **Conti E. 2013.** Divergent and narrower climatic niches characterize polyploid species of European primroses in *Primula* sect. *Aleuritia*. *Journal of Biogeography* **40:** 1278–1289.

Waltari E, **Edwards SV. 2002.** Evolutionary dynamics of intron size, genome size, and physiological correlates in archosaurs. *The American Naturalist* **160:** 539–552.

Wanderley AM, **Machado ICS**, **de Almeida EM**, **Felix LP**, **Galetto L**, **Benko-Iseppon AM**, **Sork VL. 2018.** The roles of geography and environment in divergence within and between two closely related plant species inhabiting an island-like habitat. *Journal of Biogeography* **45:** 381–393.

Wirtz P. 1999. Mother species-father species: unidirectional hybridization in animals with female choice. *Animal Behaviour* **58:** 1–12.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.

Figure S1. Haplotype networks for each scenario. Green (*S. merianae*) and red (*S. rufescens*) indicate the proportion of each haplotype in the mitochondrial (*ND4*) and nuclear (*ACA4*) genes presents in the hybrid zone and in black the proportion presents in the non-hybridization zone.

Figure S2. Correlations between pairwise genetic distance as calculated by $F_{ST}/(1-F_{ST})$ and geographical distance (log km) in *S. merianae* with the mitochondrial gene (*ND4*) with all populations (A) and with populations from region NHZ in the relaxed scenario (B). R^2 = coefficient of determination; *P* = p-value.

Figure S3. Mismatch distributions of *S. merianae* and *S. rufescens* compared to the expected frequencies under the demographic expansion model with mtDNA. *Salvator merianae* (A); *S. rufescens* (B); strict scenario of *S. rufescens* in the hybrid zone (C) and the non-hybrid zone (D); genetic scenario of *S. rufescens* in the hybrid zone (E) and the non-hybrid zone (F); relaxed scenario of *S. rufescens* in the hybrid zone (G) and the non-hybrid zone (H); relaxed scenario of *S. merianae* in the hybrid zone (I) and the non-hybrid zone (J).

Figure S4. Mismatch analysis of *S. rufescens* haplogroups. Haplogroups of *S. rufescens* (A) and mismatch distribution compared to the expected frequencies under the demographic expansion model with mtDNA for haplogroup 1 (B) and haplogroup 2 (C).

Table S1. List of the 19 bioclimatic variables from Worldclim 1.4 (http://www.worldclim.org/). *: non-correlated variables in the study area.

Table S2. Sampling localities, specimen identification codes, coordinates, phenotypic identification (Sm: *S. merianae*, Sr: *S. rufescens*) and GenBank accession numbers of *ND4* and *ACA4* gene sequences. *: samples previously analysed in Cabaña *et al.* (2014). [1] Individuals identified as hybrids.

Table S3. Polymorphism indices calculated with mitochondrial (*ND4*) and nuclear (*ACA4*) genes in *S. merianae* and *S. rufescens*.

Table S4. Neutrality tests and statistics of the mismatch distribution analyses for the *ND4* gene in *S. merianae* and *S. rufescens* and for the hybrid zone (HZ) and the non-hybrid zone (NHZ) of the scenarios that presented significant differences in the AMOVA. rg: Harpending's raggedness index statistics; SSD: deviations from the sum of squares; D: Tajima's D; F_S: Fu's F_S.