ORIGINAL ARTICLE

Spatial variation in gene flow across a hybrid zone reveals causes of reproductive isolation and asymmetric introgression in wall lizards

Weizhao Yang,[1] (iD) **Nathalie Feiner,[1] Hanna Laakkonen,[1] Roberto Sacchi,[2] Marco A. L. Zuffi,[3] Stefano Scali,[4] Geoffrey M. While,[5] and Tobias Uller[1,6]**

[1]*Department of Biology, Lund University, Lund 223 62, Sweden*

[2]*Department of Earth and Environmental Sciences, University of Pavia, Pavia 27100, Italy*

[3]*Museum Natural History, University of Pisa, Pisa 56011, Italy*

[4]*Museum of Natural History of Milan, Milano 20121, Italy*

[5]*School of Biology, University of Tasmania, Hobart, Tasmania 7001, Australia*

[6]*E-mail: tobias.uller@biol.lu.se*

Hybrid zones provide insights into the evolution of reproductive isolation. Sexual selection can contribute to the evolution of reproductive barriers, but it remains poorly understood how sexual traits impact gene flow in secondary contact. Here, we show that a recently evolved suite of sexual traits that function in male-male competition mediates gene flow between two lineages of wall lizards (*Podarcis muralis*). Gene flow was relatively low and asymmetric in the presence of exaggerated male morphology and coloration compared to when the lineages share the ancestral phenotype. Putative barrier loci were enriched in genomic regions that were highly differentiated between the two lineages and showed low concordance between the transects. The exception was a consistently low genetic exchange around *ATXN1*, a gene that modulates social behavior. We suggest that this gene may contribute to the male mate preferences that are known to cause lineage-assortative mating in this species. Although female choice modulates the degree of reproductive isolation in a variety of taxa, wall lizards demonstrate that both male-male competition and male mate choice can contribute to the extent of gene flow between lineages.

KEY WORDS: barrier loci, genomic cline analysis, hybrid zone, introgression, *Podarcis muralis* , sexual selection.

Understanding how species form requires insight into both the processes that cause lineages to diverge and those that maintain lineage identity during secondary contact (Meyer 1993; Price 2008; Nosil and Feder 2012; Ravinet et al. 2017). Although it is often difficult to study these processes directly, it is possible to draw inference from patterns of gene flow across hybrid zones (Ravinet et al. 2017). Contrasting multiple contact zones can demonstrate if the phenotypes of hybridizing lineages influence genetic exchange, whereas variation in introgression across the genome can provide information about the genetic basis of re-

productive barriers (Janoušek et al. 2015; Riemsdijk et al. 2019). Because introgression partly is determined by the rate of hybridization, lineage divergence in sexual characters may be particularly likely to limit gene flow. Although this is reasonably well-established for female choice, male-male competition has received limited attention despite that it too can influence the direction and magnitude of gene flow (Lipschutz 2017; Tinghitella et al. 2018). Unfortunately, few systems are understood well enough to allow a priori predictions about geographic and genomic variation in the strength and direction of gene flow, which
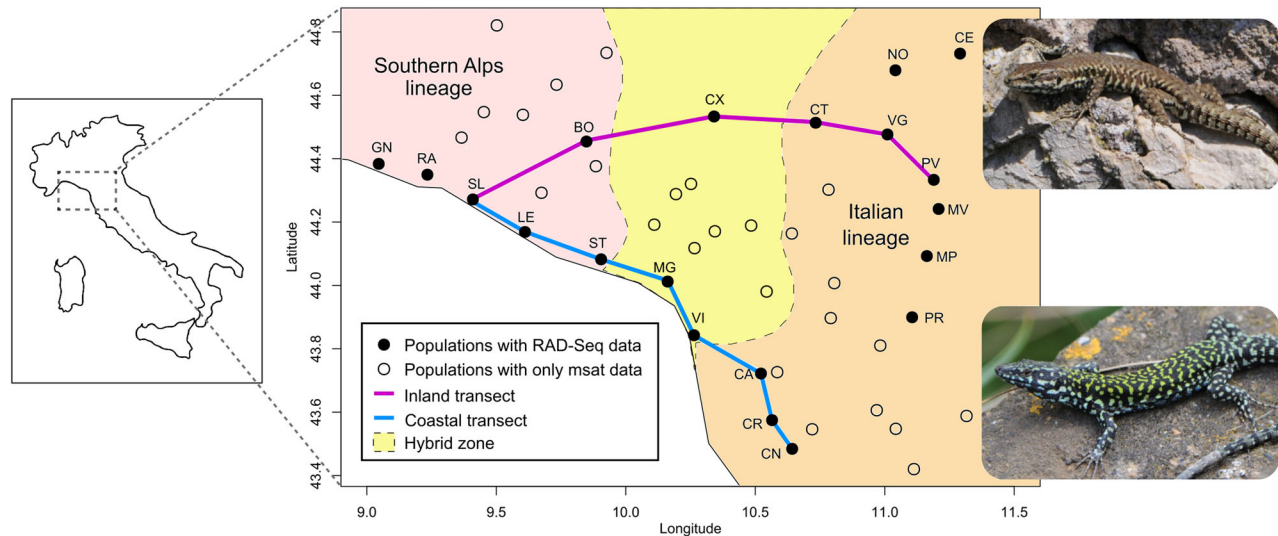
Figure 1. Geographic locations of all populations included in this study, and the inferred hybrid zone between the Italian (IT) and the Southern Alps (SA) lineage. At the location of the coastal transect, the IT lineage is characterized by a suite of sexually selected morphological traits and colors that are significantly more expressed than at the location of the inland transect (Table S1). On the right, typical male *Podarcis muralis* with (bottom) and without (top) the sexually selected phenotype. Note that population MG on the coastal transect is only represented by RAD-Seq data. Abbreviation: msat, microsatellite.

makes it difficult to know how robust the interpretations of many observed patterns really are.

The common wall lizard, *Podarcis muralis*, is a promising study system for unraveling the underlying causes of gene flow during the early stages of speciation. Previous work has revealed a contact zone in northern Italy between two lineages that diverged about 2.5 million years ago (Salvi et al. 2013): the Italian (IT) lineage in the south and the Southern Alps (SA) lineage in the north-west (Yang et al. 2018). Genomic introgression in the southernmost part of the contact zone (along the Ligurian coast) is highly asymmetric, from the IT into the SA lineage (Yang et al. 2018). This asymmetric introgression is accompanied by a suite of sexually selected morphological traits, colors, and behaviors that makes IT males competitively superior to males with the SA phenotype (Fig. 1; While et al. 2015). However, this sexual phenotype arose only recently within the IT lineage (Yang et al. 2018), and these characters are expressed to a very low degree, if at all, in the IT lineage further north (Fig. 1; Table S1). This provides a unique opportunity to test how secondary sexual traits influence the extent and direction of gene flow. We make two specific predictions: (1) if the suite of sexually selected traits in wall lizards is the cause of asymmetric introgression, as suggested by behavioral data (While et al. 2015), introgression should be asymmetric only when the IT lineage exhibits this phenotype (i.e., along the Ligurian coast) and (2) because divergence in secondary sexual characters can make lineages less prone to hybridize (e.g., Boughman 2001; Schield et al. 2017), the overall gene flow between lineages should be lower

in the presence of the sexually selected phenotype than in its absence.

Following a test of these predictions, we further compared genomic clines between the two regions of secondary contact to test if "barrier loci" were shared or unique, and associated with the sexual traits themselves. Finally, we assessed if barrier loci are located in genomic regions that are highly differentiated between lineages, and if those regions include any putative candidate genes.

Methods

The common wall lizard (*Podarcis muralis*) is a small diurnal lizard that is abundant in southern Europe. Previous studies have shown that there are two distinct genetic lineages of *P. muralis* in north-western Italy (Gassert et al. 2013; Salvi et al. 2013; While et al. 2015), which we here refer to as IT and SA. These lineages diverged ~2.5 million years ago but are now in secondary contact (Salvi et al. 2013; While et al. 2015; Yang et al. 2018). The contact zone has not previously been mapped out in detail but is known to extend from the coast around Pisa, across the Apennine mountains, and northward into Emilia-Romagna (e.g., Schulte et al. 2012; Salvi et al. 2013; While et al. 2015; Fig. 1).

SAMPLING STRATEGY
The lizards in this study were sampled from 47 locations in northern Italy (Fig. 1; Table S2), including 19 populations from previous studies (Michaelides et al. 2015; While et al. 2015;

Yang et al. 2018). We used 20 populations to generate restriction site-associated DNA sequencing (RAD-Seq) data, including eight novel populations and 12 populations previously included in the study by Yang et al. (2018). Forty-five populations were genotyped using microsatellite markers, including published data from 18 populations (Michaelides et al. 2015; While et al. 2015). Lizards were captured by noosing and tissue samples from the tail or toe (in case of complete tail loss) were collected for genetic analyses. We took standard morphological measures (snout-to-vent length, total length, head length and width, and body mass), scored the intensity of green dorsal coloration, and photographed the lizards to quantify the extent of black ventral coloration. For more details on the field protocol and the phenotyping of lizards, see While et al. (2015) and Yang et al. (2018). All lizards were released at the location of capture after processing. Sample collection was carried out in accordance with local laws and regulations in Italy, under the collection permit number Prot. PNM-2015-0009720.

LABORATORY WORK AND GENOTYPING

Genomic DNA for each sample was extracted using DNeasy blood and tissue kit (Qiagen, USA). We genotyped 547 individuals at 13 microsatellite loci developed by Richard et al. (2012) and Heathcote et al. (2015), and combined these data with published data that were generated using the same protocol (Michaelides et al. 2015; While et al. 2015). In brief, PCR products were labeled with the fluorescent dye 6-FAM, HEX, or NED, and, together with an internal ladder (red ROX-500), analyzed using an ABI 3730XL Genetic Analyser (Thermo Fisher Scientific). Alleles were scored in Geneious (version 6.1.7), and any ambiguous peaks were repeated to confirm the genotype (Table S3). The microsatellite data were examined for linkage disequilibrium (LD) and Hardy-Weinberg equilibrium (HWE) using Genepop (version 4.7.3; Raymond and Rousset 1995), and pegas (version 0.12; Paradis 2010) with 1000 replicates of Monte Carlo test. No locus showed LD, but four loci ("C38", "356", "109" and "Pm16") showed significant deviation from HWE, and were excluded from the analyses.

For RAD-Seq data, library preparation was conducted following the protocol in Peterson et al. (2012) and Yang et al. (2018) to generate the double-digest restriction site-associated DNA markers. A total of 95 individuals from eight locations were sequenced, and combined with previous data (compiled data deposited in NCBI Short Reads Archive [SRA] with accession number PRJNA486080). STACKS (version 2.2; Catchen et al. 2011; Rochette et al. 2019) was used to process the RAD-Seq reads and infer single nucleotide polymorphisms (SNPs) for each individual. At first, the "process_radtag" module was used to remove reads with low-quality score (Phred score <30), ambiguous base call, and incomplete barcode or restriction site. Clean reads were mapped to the genome of *P. muralis* (Andrade et al. 2019) using BWA (Li and Durbin 2009). Sorted bam files were used as input for the reference-based STACKS pipeline that contains modules "gstacks" and "populations" to estimate SNPs using a Marukilow model (Maruki and Lynch 2017). Stringent filtering was implemented to obtain a high-quality SNP dataset (see details in Text S1).

POPULATION STRUCTURE ANALYSIS

Although the genetic structure for populations along the coast was revealed by previous studies (While et al. 2015; Yang et al. 2018), the relationship among populations in the Apennines and further north remains unknown. To address this, we first estimated the population clustering from microsatellite data using a Bayesian method implemented in STRUCTURE (version 2.3.4; Pritchard et al. 2000) with an admixture model (Falush et al. 2003) and correlated allele frequencies. A total of 30 independent runs were performed with a burn-in of 10^5 iterations and a run length of 10^6 iterations for a number of genetic clusters (K) from 1 to 5. The best K was determined according to the method in Evanno et al. (2005). Each individual was assigned a hybrid index (Q) using the software STRUCTURE (version 2.3.4). tess3r (version 1.1.0; Caye et al. 2015) was used to estimate the spatial distribution of the hybrid zone (0.1 < Q < 0.9).

We also inferred the genetic relationship between the populations based on SNP genotypes from RAD-Seq data, including principal component analysis (PCA) in Plink (version 1.9; Chang et al. 2015), a neighbor-joining network based on pairwise genetic distance matrix in treemix version 1.13 (Pickrell and Pritchard 2012), and individual admixture assuming different numbers of clusters with co-ancestry clusters (K) from 1 to 15 in ADMIXTURE (version 1.3.0; Alexander et al. 2009).

MODELING OF GEOGRAPHIC CLINE AND DEMOGRAPHIC HISTORY

According to the population structure, we estimated geographic clines for two transects to establish the overall patterns of introgression across the hybrid zone based on the hybrid index Q for each population assigned by ADMIXTURE. The inland transect, north of the Apennines (Fig. 1), included six populations and stretched from Sestri Levante (SL), which belongs to the SA lineage, to Pian Di Venola (PV), which belongs to the IT lineage. To ensure consistency, we set the coastal transect south of the Apennines to also start at SL, and added eight locations eastward to Chianni (CN; Fig. 1). Note that SL is not the end-point of the transect of previous studies (While et al. 2015; Yang et al. 2018), which include populations further to the east. However, in this case, SL is an appropriate choice because it means that the two transects have a single population in common (i.e., the first population in the SA lineage). The hybrid indices for each transect

against their point-to-point distance were fitted to a series of equilibrium geographic cline models using the Metropolis-Hastings Markov chain Monte Carlo (MCMC) algorithm employed in the R package HZAR (version 0.2-5; Derryberry et al. 2014). We ran 15 separate models that varied in the number of cline shape parameters estimated for hybrid indices and selected the model with the lowest Akaike information criterion (AIC) as the best-fitting cline.

We tested which demographic scenarios best describe the pattern of secondary contact for the two transects based on the joint allele frequency spectrum (JAFS) of SNPs using the software δaδi (version 1.6.3; Tine et al. 2014; Rougeux et al. 2017). In δaδi, each demographic model consists of a series of population parameters, including effective population size (N), time scale (T), and migration rate (m). The same populations in geographic cline analysis with hybrid index $Q < 0.1$ and $Q > 0.9$ were used for demographic simulation (Fig. 1). Four models were fitted for each transect that represent alternative modes of divergence between the two genetic lineages, including Strict Isolation (SI), Isolation-with-Migration (IM), Ancient Migration (AM), and Secondary Contact (SC). Mutation rate per year was set to 2.1×10^{-10} according to data for a related species (Tollis and Boissinot 2014), and the generation time was set to 2.09 years/generation. The input JAFS was projected to a sample size of $N = 30$ for each lineage. Models were fitted independently using successively a hot and a cold simulated annealing procedure followed by "BFGS" optimization (Gutenkunst et al. 2009). We ran 20 independent optimizations for each model to achieve convergence and retained the best one for comparisons among models based on AIC. We also used an additional software package (fastsimcoal26; Excoffier et al. 2013) to infer the demographic history and confirm the consistency between best-fitting scenarios for these data. In fastsimcoal26, 1,000,000 maximum simulations were run for each parameter set with a maximum of 100 loops. We also ran 20 optimizations for each model starting from different random seeds. A total of 100 replicates of parametric bootstrap were used to estimate the parameter uncertainties and confidence intervals for both approaches.

BAYESIAN GENOMIC CLINE ANALYSIS

We used a total of 1029 diagnostic SNPs to assess patterns of introgression in the two hybrid zone transects based on Bayesian genomic cline analysis (BGCA) as implemented by Gompert and Buerkle (2011) and Gompert et al (2012). This analysis uses genomic cline models to describe patterns of introgression between two populations that are near fixed for the focal markers (i.e., "pure" populations). The genomic cline model has two basic parameters—α and β (Gompert and Buerkle 2011). The genomic cline parameter α measures the change in probability of ancestry relative to a null expectation. Increasing (positive val-

ues) or decreasing (negative values) α parameter values reflect shifts in genomic clines toward one of the two pure populations (IT → SA for positive values, and SA → IT for negative values in this study). The genomic cline parameter β reflects the rate of change in probability of ancestry from one pure population to the other. Positive values of β parameter denote steeper clines, whereas negative values denote wider clines. Thus, α parameter values represent directional movement of alleles from one lineage into another, whereas β parameter values reflect the strength of the barrier to gene flow between the two lineages.

The focal markers used in the BGCA analyses are diagnostic SNPs. We defined SNPs as diagnostic when the frequency of the SNP was >0.75 in the populations at the ends of each transect and with different alleles for the two lineages. In addition, we singled out a set of markers from the SNPs that were identified by Yang et al. (2018) as being associated with the sexually selected characters in the IT lineage ($F_{ST} > 0.066$ between populations with the characters and reference populations; for more information, see Table S2). Those loci were analyzed only for the coastal transect because this is where the sexually selected characters are found. To assess both α and β parameters, we ran five independent MCMC chains for each dataset (transect) for 50,000 steps with 25,000 steps as a burn-in period with random seeds. The output was recorded each 25th step to obtain 1000 samples. We merged the output of all five MCMC chains by averaging estimates over all chains for each marker. The 90% credible intervals (CIs) were merged over the five MCMC chains by choosing the most conservative (i.e., the widest) intervals. These merged 90% CIs were used to detect outliers. Pearson's χ^2 test with Yates's correction and Fisher's exact test were applied to evaluate the difference of α and β parameters between transects or datasets.

To identify candidate genes associated with outlier SNPs, we scanned the genes within 25-kb windows around SNPs that represented significant α and positive β outliers. Because the biological relevance of outlier SNPs with negative β values is unclear for our present aims, we excluded this category. We characterized the functional composition of candidate genes using the Gene Ontology (GO) term classification derived from the *P. muralis* genome annotation (Andrade et al. 2019). Fisher's exact test with false discovery correction rate correction was used to identify the overrepresented functional categories.

To further investigate the relationship between genetic divergence and degree of introgression, we calculated the pairwise F_{ST} between IT and SA for each transect using VCFTools (version 0.1.14; Danecek et al. 2011). An elevated pairwise F_{ST} can characterize genomic regions under lineage-specific selection (Chen et al. 2016; Lamichhaney et al. 2016), although it can also arise without selection (i.e., neutral divergence) or reflect genomic variation in, for example, recombination rates (Sodeland et al.

2016; Ochoa and Storey 2019). F_{ST} values were calculated across the genome over 100 kb fixed windows that contain at least two variants in the full SNP dataset. These windows were assigned to one of three categories: (i) containing at least one barrier locus (positive β outlier), (ii) containing at least one diagnostic SNP but no barrier loci, and (iii) containing only SNPs that are nondiagnostic and not barrier loci (i.e., "normal" SNPs). We assessed the significance levels of differences in F_{ST} values between categories (i) and both (ii) and (iii) by permutation tests with 1000 iterations. In addition to using pairwise F_{ST} as a relative measure of divergence, we also calculated D_{xy}, the average number of pairwise differences between alleles, as an absolute measure of divergence (Nei 1987; Cruickshank and Hahn 2014).

Results

In total, we obtained genotypes of nine microsatellite loci for 939 common wall lizards from 45 populations, and RAD-Seq data for 259 individuals from 20 populations, which included 21,305 SNPs with mean coverage of 25.01 per site, and average genotyping rate of 95.4%.

POPULATION STRUCTURE AND IDENTIFICATION OF HYBRID ZONES

For microsatellite data, the Bayesian clustering analysis indicated that two genetic clusters best describe the data, separating all populations into IT and SA lineages from west to east. Nine populations were identified with hybrid ancestry according to the hybrid index (i.e., 0.1 < Q < 0.9). The hybrid zone was estimated to be located between longitude 10.0°E and 10.7°E (Figs. 1 and S1).

For the RAD-Seq data, the first principal component (variance explained = 62.7%) also separated all populations into the two groups that correspond to the IT and SA lineages (Fig. 2A). The second principal component (variance explained = 14.7%) separated the IT lineage into a northern and a southern group, roughly separated by the Apennine mountains. Cedogno (CX), on the inland transect, exhibited clear evidence of genetic admixture, as did two populations Montignoso (MG) and Viareggio (VI) on the coastal transect (Fig. 2A). This result was consistent with results from the neighbor-joining network, where the IT and SA populations were separated, with hybrid populations in between (Fig. 2B). ADMIXTURE clustering analysis was also consistent with the separation between IT and SA when the number of presumed ancestral population (K) was set to two, and with the genetic divergence within the IT lineage when K was set to three (which was best supported by cross-validation). The inferred hybrid indices for CX, MG, and VI from clustering analysis also supported the hybridization between IT and SA at these locations (Fig. 2C).

MODELING OF GEOGRAPHIC CLINE AND DEMOGRAPHIC HISTORY

For the inland transect, the center of the best-fitting cline was 71.54 (95% CI: 63.70-79.38) km from the western population (SL), and the width of the cline was 49.53 (95% CI: 33.69-65.37) km (Fig. 3A). Similarly, the center and width of the cline for the coastal transect were 75.48 (95% CI: 71.76-79.20) km and 33.65 (95% CI: 24.89-42.41) km, respectively (Fig. 3B).

Model comparison in δaδi showed that the best-fitting model for both transects was the secondary contact model, with IT and SA lineages coming into secondary contact approximately 260 (CI: 141-423) thousand years ago for the inland transect and 358 (CI: 251-662) thousand years ago for the coastal transect (Table S4). The two transects showed significantly different patterns of gene flow between IT and SA. For the inland transect, the best-fitting model suggested that the gene flow measured as migration rate "m" between IT and SA was equal in magnitude in both directions (1.42×10^{-6} from IT into SA with CI: 1.15×10^{-6} to 1.45×10^{-6}; 1.11×10^{-6} from SA into IT with CI: 0.95×10^{-6} to 1.31×10^{-6}). In contrast, for the coastal transect, the genetic exchange was asymmetric, with about twice the gene flow from IT into SA than vice versa (1.41×10^{-6} with CI: 1.01×10^{-6} to 1.54×10^{-6} from IT into SA; 7.62×10^{-7} with CI: 5.79×10^{-7} to 7.80×10^{-7} from SA into IT; Figs. 3C and 3D; Table S5). The fastsimcoal26 simulations gave different absolute values, but also supported asymmetric gene flow for the coastal transect (1.44×10^{-7} with CI: 1.10×10^{-7} to 2.29×10^{-7} from IT into SA; 3.34×10^{-8} with CI: 1.07×10^{-9} to 5.89×10^{-8} from SA into IT) but not for the inland transect (1.09×10^{-7} with CI: 9.30×10^{-8} to 1.69×10^{-7} from IT into SA; 1.58×10^{-7} with CI: 1.34×10^{-7} to 2.05×10^{-7} from SA into IT; Table S5). From these results, it is also evident that the gene flow from SA into IT along the coast is significantly lower than any of the other estimates.

BAYESIAN GENOMIC CLINE ANALYSIS

Consistent with asymmetric introgression from IT to SA in the coastal transect but not in the inland transect, there were relatively more outliers with positive α values (i.e., IT → SA) in the coastal transect (199 and 134 outliers with positive and negative α in the coastal transect versus 112 and 108 outliers in the inland transect; $\chi^2 = 3.86$, $P = 0.049$; Fig. 4A). β values exhibited marginally more positive outliers (characteristic of steeper clines) than negative outliers (characteristic of wider clines) in both transects, with 142 and 82 for the coastal transect, and 39 and 27 for the inland transect ($\chi^2 = 0.24$, $P = 0.624$; Fig. 4B). The relationship between α and β values is best described by an inverted U-shaped curve ($P_{inland} = 0.014$; $P_{coastal} = 0.003$), where the positive β parameter values associated with α parameter values close to zero (Fig. S2). Thus, the barrier loci do not show any sign of directional movement (Table 1).
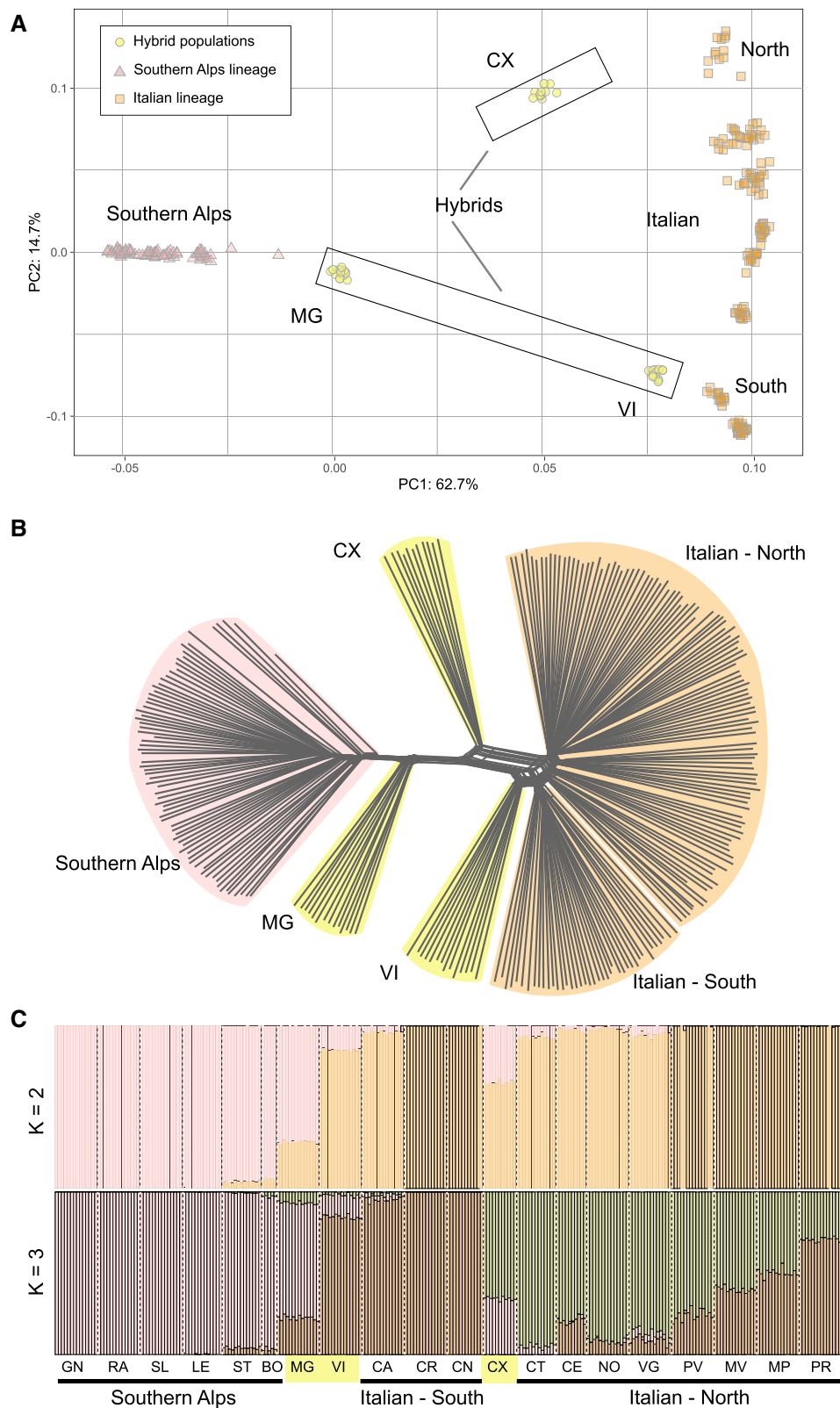
Figure 2. Summary of the results from the population genetic analyses. (A) Two dimensional PCA plots of all individuals based on 21,305 SNPs. The two genetic lineages (IT and SA) are separated on PC1 with hybrid populations falling in between. PC2 differentiates the Italian lineage following a North-South gradient. (B) Neighbor-joining network based on genetic distances measured by p-distance. (C) Admixture clustering of sampled individuals into two and three groups (clusters K). The proportion of each individual's genome assigned to each cluster is shown by the length of the colored segments.
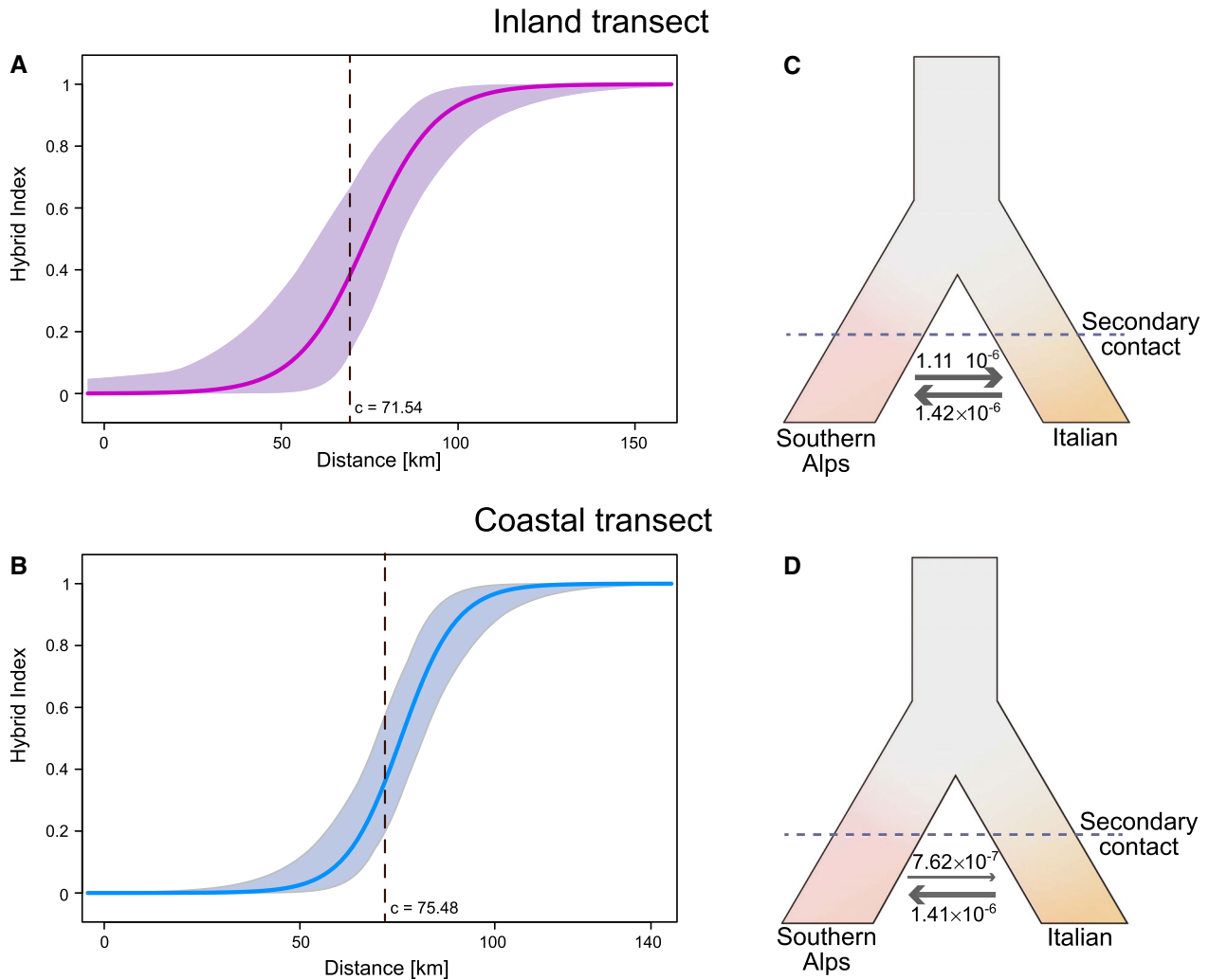
Inland transect

A

C

Coastal transect

B

D

Figure 3. Introgression patterns of the two transects. Maximum likelihood geographic clines (solid lines) for the inland (A) and coastal transect (B), respectively. Shadings represent the confidence intervals around the estimated clines, and the dashed lines indicate the center of each cline. Demographic simulations of the historical gene flow ("*m*") between the Southern Alps and Italian lineages in the inland (C) and coastal (D) transect, respectively. Arrows indicate the gene flow between the two lineages under the secondary contact scenario. The thickness of the arrows is proportional to the estimated gene flow.

Table 1. Summary of Bayesian genomic cline analysis (BGCA) results.

Outlier	Interpretation	Inland	Coastal[a]
$\alpha > 1$	Direction introgression from IT into SA	112	199 (7)
$\alpha < 1$	Direction introgression from SA into IT	108	134 (2)
$\beta > 1$	Increase the rate of change in ancestry	39	142 (4)
$\beta < 1$	Decrease the rate of change in ancestry	27	82 (2)

[a]Numbers in parentheses refer to 22 diagnostic loci that were previously shown to be associated with the sexually selected characters (Yang et al. 2018).

Of the 261 SNPs that were identified as being associated with the sexually selected characters in the IT lineage, 22 fulfilled the criteria for the genomic cline analysis (i.e., they are diagnostic SNPs for the focal clines). Amongst these, seven and

two outliers were identified with positive and negative α values, and four and two were identified with positive and negative β values, respectively. These frequencies are similar to those of the total SNP dataset (Fisher's exact $P_\alpha = 0.654$, $P_\beta = 0.532$; Table 1),
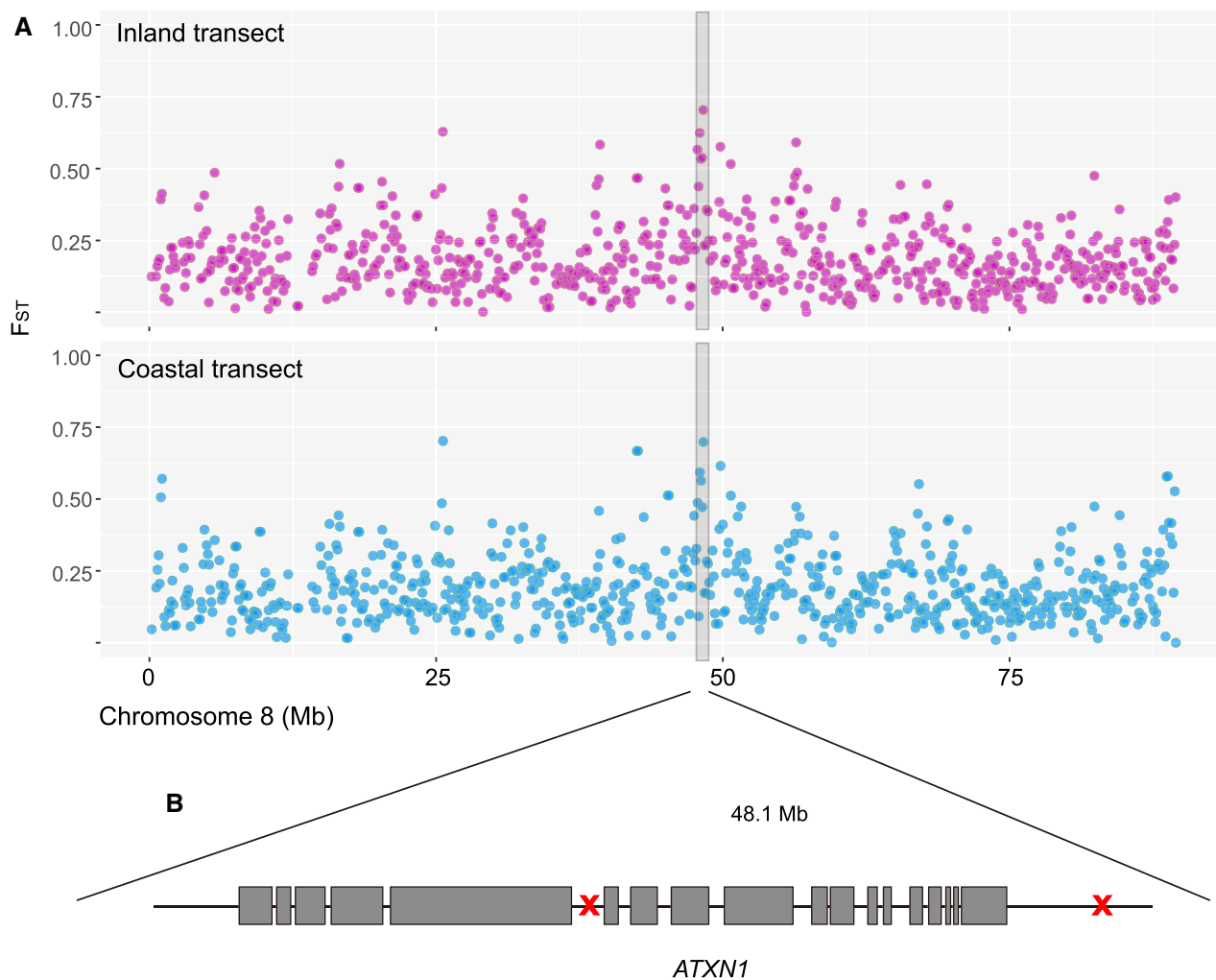
Figure 4. The genomic region with strongest signature of restricted introgression in both hybrid zones lies on chromosome 8. (A) F_{ST} values between the IT and SA lineage in both transects on chromosome 8. The shaded area marks the region with two shared positive β outliers between the two transects, which also shows elevated genetic divergence (high F_{ST}) between the two lineages. (B) The structure of the *ATXN1* gene in the candidate region. Gray boxes mark the positions of exons and red crosses indicate the position of the two outliers with positive β values.

which implies that highly lineage-divergent loci associated with sexually selected traits did not behave significantly different from genome-wide loci.

IDENTIFICATION OF GENES ASSOCIATED WITH BARRIER LOCI

For the inland transect, 57 genes were identified around outliers with positive β values (i.e., putative barrier genes between IT and SA), which were overrepresented in 10 GO categories (Fig. S3). For the coastal transect, the same outlier groups contained a total of 210 genes with positive β values, which were overrepresented in 10 GO categories (Fig. S3). The low number of diagnostic SNPs that are associated with sexually selected traits (22) precluded tests of functional enrichment for these genes (Table S5). A total of 12 candidate genes around positive β outlier

SNPs appeared in both transects. Among those genes, overrepresentation analysis identified only one GO category with positive β (Fig. S3). This was based on a total of 11 outlier SNPs with positive β in both transects (Table S6). For a summary of the corresponding analyses for positive and negative α values, see Text S2.

To further investigate the genetic divergence of barrier genes between IT and SA in both transects, we compared the pairwise F_{ST} and the absolute divergence D_{xy} between different regions of the genome (see Methods). We found for both transects that the F_{ST} values (inland: 0.5553; coastal: 0.5921) for regions containing barrier loci were significantly higher than the F_{ST} values for regions containing other diagnostic loci (inland: 0.4396; coastal: 0.4406; both $P < 0.001$) and only "normal" SNPs (inland: 0.1816; coastal: 0.1868; both $P < 0.001$). Similarly, the

D_{xy} values (inland: 0.0031; coastal: 0.0033) for regions containing barrier loci were significantly higher than the mean D_{xy} values for other diagnostic loci (inland: 0.0027; coastal: 0.0027; $P_{inland} = 0.001$; $P_{coastal} < 0.001$) and only "normal" SNPs (inland: 0.0018; coastal: 0.0019; both $P < 0.001$). Among all barrier loci, we found two neighboring loci (48,142,324 and 48,398,446) on chromosome 8 that were positive β outliers in both transects and that were most highly differentiated between IT and SA in both transects ($F_{ST} > 0.7$; Fig. 4A; $D_{xy} > 0.0044$; Fig. S4). The two loci are located in a region containing the gene *ataxin 1* (*ATXN1*; Fig. 4B), a gene that has been shown to play a role in the development of social behaviors (Celestino-Soper et al. 2012; Lu et al. 2017).

Discussion

Geographic variation in gene flow across a hybrid zone in common wall lizards suggests that sexual selection promotes asymmetric introgression in part by limiting gene flow from the less competitive lineage. Putative barrier loci showed limited overlap between the two transects, but were consistently highly differentiated between lineages and included one interesting candidate gene. Here, we discuss these results and how they fit with the role of male-male competition and mate choice in the evolution of reproductive barriers.

We have previously shown that a suite of sexually selected characters originated in the south of the IT lineage and later spread northwestward to reach the secondary contact zone between the IT and SA lineages (Yang et al. 2018). The sexually selected morphologies, colorations, and behaviors introgress into the SA lineage along the Ligurian coast, and this is accompanied by asymmetric gene flow in the same direction (While et al. 2015; Yang et al. 2018). Here, we show that this asymmetry in gene flow does not exist in the inland hybrid zone, where the two lineages both share the ancestral phenotype. The different pattern of genomic introgression for the two transects is further evidence that sexual selection caused asymmetric introgression between the two lineages along the Ligurian coast. Italian males, with stout bodies, large heads, and black and green color ornaments, outcompete other males for access to territories and females, which results in a bias toward hybridization between IT males and SA females in experimental contact zones (MacGregor et al. 2017a; While et al. 2015; Heathcote et al. 2016). Such differences in male competitive advantage may very well be an important cause of asymmetric introgression in other species too, a hypothesis that can be put to the test using multiple transects as we did here, or using comparative analysis across pairs of hybridizing species.

Although there is extensive phenotypic introgression along the coast (While et al. 2015), the geographic and genomic data

from the coastal and inland clines suggest that genetic material does not flow freely between lineages (see also Yang et al. 2018). Furthermore, the evolution of sexually selected characters appears to have limited this genetic exchange even further as gene flow from SA into IT was significantly lower on the coast than the symmetric gene flow between the two lineages in the inland. Thus, secondary sexual characters not only promote asymmetric introgression of particular alleles (Yang et al. 2018), but may also reduce the overall genetic exchange in the opposite direction.

Several studies have suggested that sexual selection can play a major role in preventing gene flow, an effect that is often attributed to female choice (Paterson 1985; Stein and Albert 2006). In contrast, female common wall lizards do not chose mates based on male morphology, coloration, or the composition of femoral secretions, and they therefore fail to discriminate between males of lineages that are very different genetically and phenotypically (Heathcote et al. 2016; MacGregor et al. 2017a,b). This appears to be common for the clade as a whole: for example, females do not discriminate between males from different lineages in the Iberian species complex (Martin and Lopez 2006; Font et al. 2012). We therefore interpret the reduced gene flow from the SA lineage into the IT lineage along the Ligurian coast as the result of IT males limiting the mating opportunities for males with a predominant SA phenotype and genotype. If so, wall lizards would be one of the few examples where male-male competition reduces gene flow in secondary contact (reviewed in Lipshutz 2017). Although sexual selection alone is very unlikely to complete speciation (Servedio and Boughman 2017), it can promote geographic exclusion and thereby increase the likelihood that lineages become reproductively isolated (Tinghitella et al. 2018). This scenario is plausible if selection on the sexual traits themselves vary across the landscape, because this could promote further genetic divergence between lineages and select against hybrids. Whether this is the case in common wall lizards is currently unknown. More generally, traits used in male-male competition appear highly evolutionarily labile within the wall lizard clade (e.g., Marshall et al. 2015; see Böhme 1986), which provides a useful setting for addressing the role of sexual selection at different stages in the speciation process, a topic that so far has been dominated by research on taxa with strong female choice (e.g., birds; Cooney et al. 2017).

Although male-male competition can explain the asymmetric introgression along the coast, the limited gene flow from SA into IT may be exaggerated by the males' preference for females from the same lineage. It is unclear what female traits that male lizards pay attention to, but male preferences reliably cause lineage-assortative mating in free-ranging common wall lizards (Heathcote et al. 2016; MacGregor et al. 2017a; see also Font et al. 2012). Male mate choice may therefore explain why

gene flow appears restricted across the length of the hybrid zone despite that there are no physical barriers, or evidence for low viability of hybrids (While et al. 2015). That the most consistent genomic signature of reproductive isolation in both transects mapped onto the *ATXN1* gene is interesting in this regard. This gene encodes a chromatin-binding factor that represses Notch signaling and is involved in neurogenesis, brain development, and social behavior (Celestino-Soper et al. 2012; Lu et al. 2017). For common wall lizards, mate recognition and assessment occur primarily through male premating courtship of females (Sacchi et al. 2007; Scali et al. 2013). Thus, we speculate that the *ATXN1* gene may contribute to the reproductive barrier between IT and SA via divergence in neurobehavioral functions. Such genetic differences could help explain why reproduction is assortative in populations of mixed origins and in the absence of opportunities for social learning of mate preferences from adults (While et al. 2015). However, more data are needed to substantiate that *ATXN1* is an important candidate gene and to identify other loci and biological functions that promote reproductive barriers between the two lineages.

Theory predicts that gene flow should be particularly restricted for genomic regions that have been under strong lineage-specific selection. In wall lizards, the putative barrier loci were indeed predominantly located in regions that were highly differentiated between lineages, which can be a signature of lineage-specific selection (Stern and Nielsen 2019). However, except for the association with *ATXN1*, very few candidate barrier loci were shared between the two transects. Similarly, the GO categories that were overrepresented by those candidate genes also showed limited overlap in the two hybrid zones. Thus, although the reproductive barriers consistently involve highly divergent genomic regions, the genomic basis of the barriers appears to show geographic differences. These differences are apparently not caused by the sexual phenotype because candidate SNPs for this phenotype did not feature among the putative barrier loci. Putative barrier loci differ between hybrid zones also in other taxa, such as *Heliconius* butterflies (Nadeau et al. 2014), mouse (*Mus musculus*; Janoušek et al. 2015), and *Bufo bufo/spinosus* (Riemsdijk et al. 2019). One possible explanation is local adaptation (e.g., to climate), creating a mosaic of genomic variation that results in local incompatibilities between lineages. However, high genomic differentiation could also arise from other population processes (e.g., founder effects, see Ochoa and Storey 2019), variation in recombination rates, or structural differences (e.g., chromosomal rearrangement; Sodeland et al. 2016). Thus, further studies of genome-wide patterns of differentiation between lineages and populations are needed to clarify the extent to which the putative barrier genomic regions identified here have been under divergent selection in the two lineages. It will also be important to assess if there is ongoing selection within the hybrid zone, and to what extent hybrids might be at a disadvantage.

In summary, geographic variation in gene flow across this zone of secondary contact suggests that secondary sexual traits that function in male-male competition can shape the extent of reproductive isolation between lineages and generate asymmetric patterns of introgression.

AUTHOR CONTRIBUTIONS

TU and WY conceived the study. All authors contributed to study design and data collection. WY conducted the laboratory work and statistical analyses. WY, NF, and TU wrote the manuscript. All authors read, edited, and approved the final submitted version of the manuscript.

ACKNOWLEDGMENTS

DATA ARCHIVING

Sequence data deposited in NCBI Short Reads Archive (SRA) with BioProject accession number PRJNA486080.

CONFLICT OF INTEREST

The authors declare no conflict of interest

LITERATURE CITED

Alexander, D. H., J. Novembre, and K. Lange. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19:1655–1664.

Andrade, P., C. Pinho, G. Perez, S. Afonso, J. Brejcha, C. J. Rubin, O. Wallerman, P. Pereira, and S. J. Sabatino. 2019. Regulatory changes in pterin and carotenoid genes underlie balanced color polymorphisms in the wall lizard. Proc. Nat. Acad. Sci. USA 116:5633–5642.

Böhme, W. 1986. Handbuch der Reptilien und Amphibien Europas. Aula-Verlag, Wiesbaden, Germany.

Boughman, J. W. 2001. Divergent sexual selection enhances reproductive isolation in sticklebacks. Nature 411:944–948.

Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko, and J. H. Postlethwait. 2011. Stacks: building and genotyping loci de novo from short-read sequences. G3-Genes Genom. Genet. 1:171–182.

Caye, K., T. M. Deist, H. Martins, O. Michel, and O. François. 2015. TESS3: fast inference of spatial population structure and genome scans for selection. Mol. Ecol. Resour. 16:540–548.

Celestino-Soper, P. B. S., C. Skinner, R. Schroer, P. Eng, J. Shenai, M. M. J. Nowaczyk, D. Terespolsky, D. Cushing, G. S. Patel, L. Immken, et al. 2012. Deletions in chromosome 6p22.3-p24.3, including ATXN1, are associated with developmental delay and autism spectrum disorders. Mol. Cytogenet. 5:17.

Chang, C. C., C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, and J. J. Lee. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4:7.

Chen, C., Z. Liu, Q. Pan, X. Chen, H. Wang, H. Guo, S. Liu, H. Lu, S. Tian, R. Li, et al. 2016. Genomic analyses reveal demographic history and temperate adaptation of the newly discovered honey bee subspecies *Apis mellifera sinisxinyuan* n. ssp. Mol. Biol. Evol. 33:1337–1348.

Cooney, C. R., J. A. Tobias, J. T. Weir, C. A. Botero, and N. Seddon. 2017. Sexual speciation and constraints on geographical range overlap in birds. Ecol. Lett. 20:863–871.

Cruickshank, T. E., and M. W. Hahn. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol. Ecol. 23:3133–3157.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, et al. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158.

Derryberry, E. P., G. E. Derryberry, J. M. Maley, and R. T. Brumfield. 2014. HZAR: hybrid zone analysis using an R software package. Mol. Ecol. Resour. 14:652–663.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14:2611–2620.

Excoffier, L., I. Dupanloup, E. Huerta-Sanchez, V. C. Sousa, and M. Foll. 2013. Robust demographic inference from genomic and SNP data. PLoS Genet. 9:e1003905.

Falush, D., M. Stephens, and J. K. Pritchard. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587.

Font, E., D. Barbosa, C. Sampedro, and P. Carazo. 2012. Social behavior, chemical communication, and adult neurogenesis: studies of scent mark function in Podarcis wall lizards. Gen. Comp. Endocr. 177:9–17.

Gassert, F. G., U. Schulte, M. Husemann, W. Ulrich, D. Rödder, A. Hochkirch, E. Engel, J. Meyer, and J. C. Habel. 2013. From southern refugia to the northern range margin: genetic population structure of the common wall lizard, *Podarcis muralis*. J. Biogeogr. 40:1475–1489.

Gutenkunst, R. N., R. D. Hernandez, S. H. Williamson, and C. D. Bustamante. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. 5:e1000695.

Gompert, Z., and C. A. Buerkle. 2011. Bayesian estimation of genomic clines. Mol. Ecol. 20:2111–2127.

Gompert, Z., T. L. Parchman, and C. A. Buerkle. 2012. Genomics of isolation in hybrids. Philos. T. Roy. Soc. B 367:439–450.

Heathcote, R. J. P., D. A. Dawson, and T. Uller. 2015. Characterisation of nine European wall lizard (*Podarcis muralis*) microsatellite loci of utility across sub-species. Conserv. Genet. Resour. 7:85–87.

Heathcote, R. J., G. M. While, H. E. MacGregor, J. Sciberras, C. Leroy, P. D'Ettorre, and T. Uller. 2016. Male behaviour drives assortative reproduction during the initial stage of secondary contact. J. Evol. Biol. 29:1003–1015.

Janoušek, V., P. Munclinger, L. Wang, K. C. Teeter, and P. K. Tucker. 2015. Functional organization of the genome may shape the species boundary in the house mouse. Mol. Biol. Evol. 32:1208–1220.

Lamichhaney, S., F. Han, J. Berglund, C. Wang, M. S. Almen, M. T. Webster, B. R. Grant, P. R. Grant, and L. Andersson. 2016. A beak size locus in Darwin's finches facilitated character displacement during drought. Science 352:470–474.

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics 25:1754–1760.

Lipshutz, S. E. 2017. Interspecific competition, hybridization, and reproductive isolation in secondary contact: missing perspectives on males and females. Curr. Zool. 64:75–88.

Lu, H., Q. Tan, M. W. C. Rousseaux, W. Wang, J. Y. Kim, R. Richman, Y. W. Wan, S. Y. Yeh, J. M. Patel, X. Liu, et al. 2017. Disruption of the ATXN1-CIC complex causes a spectrum of neurobehavioral phenotypes in mice and humans. Nat. Genet. 49:527–539.

MacGregor, H. E., G. M. While, J. Barrett, I. Perez, G. de Lanuza, P. Carazo, S. Michaelides, and T. Uller. 2017a. Experimental contact zones reveal causes and targets of sexual selection in hybridizing lizards. Funct. Ecol. 31:742–752.

MacGregor, H. E., R. A. Lewandowsky, P. d'Ettorre, C. Leroy, N. W. Davies, G. M. While, and T. Uller. 2017b. Chemical communication, sexual selection, and introgression in wall lizards. Evolution 71:2327–2343.

Marshall, K. L. A., K. E. Philpot, I. Damas-Moreira, and M. Stevens. 2015. Intraspecific colour variation among lizards in distinct island environments enhances local camouflage. PLoS One 10:e0135241.

Martin, J., and P. Lopez. 2006. Interpopulational differences in chemical composition and chemosensory recognition of femoral gland secretions of male lizards Podarcis hispanica: implications for sexual isolation in a species complex. Chemoecology 16:31–38.

Maruki, T., and M. Lynch. 2017. Genotype calling from population-genomic sequencing data. G3-Genes Genom. Genet. 7:1393–1404.

Meyer, A. 1993. Phylogenetic relationship and evolutionary processes in East-African cichlid fishes. Trends Ecol. Evol. 8:279–284.

Michaelides, S. N., G. M. While, N. Zajac, and T. Uller. 2015. Widespread primary, but geographically restricted secondary, human introductions of wall lizards, Podarcis muralis. Mol. Ecol. 24:2702–2714.

Nadeau, N. J., M. Ruiz, P. Salazar, B. Counterman, J. A. Medina, H. Ortiz-Zuazaga, A. Morrison, W. O. McMillan, C. D. Jiggins, and R. Papa. 2014. Population genomics of parallel hybrid zones in the mimetic butterflies, *H. melpomene* and *H. erato*. Genome Res. 24:1316–1333.

Nei, M. 1987. Molecular evolutionary genetics. Columbia Univ. Press, New York.

Nosil, P., and J. L. Feder. 2012. Genomic divergence during speciation: causes and consequences. Philos. T. Roy. Soc. B 367:332–342.

Ochoa, A., and J. D. Storey. 2019. New kinship and FST estimates reveal higher levels of differentiation in the global human population. bioRxiv. https://doi.org/10.1101/653279.

Paradis, E. 2010. Pegas: an R package for population genetics with an integrated - modular approach. Bioinformatics 26:419–420.

Paterson, H. E. H. 1985. The recognition concept of species. Pp. 21–29 *in* E. S. Vrba, ed. Species and speciation. Transvaal Museum, Pretoria, South Africa.

Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher, and H. E. Hoekstra. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One 7:e37135.

Pickrell, J. K., and J. K. Pritchard. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. 8:e1002967.

Price, T. 2008. Speciation in birds. Roberts & Company Publishers, Englewood, CO.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Ravinet, M., R. Faria, R. K. Butlin, J. Galindo, N. Bierne, M. Rafajlovic, M. A. F. Noor, B. Mehlig, and A. M. Westram. 2017. Interpreting the genomic landscape of speciation: finding barriers to gene flow. J. Evol. Biol. 30:1450–1477.

Raymond, M., and F. Rousset. 1995. An exact test for population differentiation. Evolution 49:1280–1283.

Richard, M., V. M. Stevens, M. L. Hénanff, and A. Coulon. 2012. Fourteen new polymorphic microsatellite loci for the wall lizard Podarcis muralis (Sauria: Lacertidae). Mol. Ecol. Resour. 12:374–376.

Riemsdijk, I., J. W. Arntzen, G. Bucciarelli, E. McCartney-Melstad, M. Rafajlovic, P. A. Scott, E. Toffelmier, H. B. Shaffer, and B. Wielstra. 2019. Spatial variation in introgression along a toad hybrid zone in France. bioRxiv. http://doi.org/10.1101/746073.

Rochette, N. C., A. G. Rivera-Colon, and J. M. Catchen. 2019. Stacks 2: analytical methods for paired-end sequencing improve RADseq-based population genomics. Mol. Ecol. 28:4737–4754.

Rougeux, C., L. Bernatchez, and P.-A. Gagnaire. 2017. Modeling the multiple facets of speciation-with-gene-flow toward inferring the divergence history of lake whitefish species pairs (*Coregonus clupeaformis*). Genome Biol. Evol. 9:2057–2074.

Sacchi, R., S. Scali, F. Pupin, A. Gentilli, P. Galeotti, and M. Fasola. 2007. Microgeographic variation of colour morph frequency and biometry of common wall lizards. J. Zool. 273:389–396.

Salvi, D., J. Harris, A. Kaliontzopoulou, M. A. Carretero, and C. Pinho. 2013. Persistence across Pleistocene ice ages in Mediterranean and extra-Mediterranean refugia: phylogeographic insights from the common wall lizard. BMC Evol. Biol. 13:147.

Scali, S., R. Sacchi, M. Azzusi, S. Daverio, T. Oppedisano, and M. Mangiacotti. 2013. Homeward bound: factors affecting homing ability in a polymorphic lizard. J. Zool. 289:196–203.

Schield, D. R., R. H. Adams, D. C. Card, B. W. Perry, G. M. Pasquesi, T. Jezkova, D. M. Portik, A. L. Andrew, C. L. Spencer, E. E. Sanchez, et al. 2017. Insight into the roles of selection in speciation from genomic patterns of divergence and introgression in secondary contact in venomous rattlesnakes. Ecol. Evol. 7:3951–3966.

Schulte, U., A. Hochkirch, S. Lotters, D. Rodder, S. Schweiger, T. Weimann, and M. Veith. 2012. Cryptic niche conservatism among evolutionary lineages of an invasive lizard. Global Ecol. Biogeogr. 21:198–211.

Servedio, M. R., and J. W. Boughman. 2017. The role of sexual selection in local adaptation and speciation. Annu. Rev. Ecol. Evol. Syst. 48:85–109.

Sodeland, M., P. E. Jorde, S. Lien, S. Jentoft, P. R. Berg, H. Grove, M. P. Kent, M. Arnyasi, E. M. Olsen, and H. Knutsen. 2016. "Islands of Divergence" in the Atlantic cod genome represent polymorphic chromosomal rearrangements. Genome Biol. Evol. 8:1012–1022.

Stein, A. C., and J. C. U. Albert. 2006. Unidirectional introgression of a sexually selected trait across an avian hybrid zone: a role for female choice? Evolution 60:1476–1485.

Stern, A. J., and R. Nielsen. 2019. Detecting natural selection. Pp. 397–440 *in* D. Balding, I. Moltke, and J. Marioni, eds. Handbook of statistical genomics. John Wiley & Sons, Hoboken, NJ.

Taylor, E., C. Foote, and C. Wood. 1996. Molecular genetic evidence for parallel life-history evolution within a Pacific salmon (sockeye salmon and kokanee, *Oncorhynchus nerka*). Evolution 50:401–416.

Tine, M., H. Kuhl, P.-A. Gagnaire, B. Louro, E. Desmarais, R. S. T. Martins, J. Hecht, F. Knaust, K. Belkhir, S. Klages, et al. 2014. European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. Nat. Commun. 5:5770.

Tinghitella, R. M., A. C. Lackey, M. Martin, P. D. Dijkstra, J. P. Drury, R. Heathcote, J. Keagy, E. S. Scordato, and A. M. Tyers. 2018. On the role of male competition in speciation: a review and research agenda. Behav. Ecol. 29:783–797.

Tollis, M., and S. Boissinot. 2014. Genetic variation in the green anole lizard (*Anolis carolinensis*) reveals island refugia and a fragmented Florida during the quaternary. Genetica 142:59–72.

Ulloa, M., J. N. Corgan, and M. Dunford. 1995. Evidence for nuclear cytoplasmic incompatibility between *Allium-Fistulosum* and *Allium-Cepa*. Theor. Appl. Genet. 90:746–754.

While, G. M., S. Michaelides, R. J. Heathcote, H. E. MacGregor, N. Zajac, J. Beninde, P. Carazo, I. Perez, G. de Lanuza, R. Sacchi, et al. 2015. Sexual selection drives asymmetric introgression in wall lizards. Ecol. Lett. 18:1366–1375.

Yang, W., G. M. While, H. Laakkonen, R. Sacchi, M. A. Zuffi, S. Scali, D. Salvi, and T. Uller. 2018. Genomic evidence for asymmetric introgression by sexual selection in the common wall lizard. Mol. Ecol. 27:4213–4224.

Associate Editor: A. Kaliontzopoulou
Handling Editor: D. W. Hall

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1.
Figure S2.
Figure S3.
Figure S4.
Table S1.
Table S2.
Table S3.
Table S4.
Table S5.
Table S6.
Supplementary Material
Supplementary Material