Differential introgression across newt hybrid zones: Evidencefrom replicated transects

Piotr Zieliński | Katarzyna Dudek | Jan W. Arntzen | Gemma Palomar |Marta Niedzicka | Anna Fijarczyk | Marcin Liana | Dan Cogălniceanu |Wiesław Babik

Received: 19 June 2019 | Revised: 16 September 2019 | Accepted: 19 September 2019

DOI: 10.1111/mec.15251

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Differential introgression across newt hybrid zones: Evidence from replicated transects

Piotr Zieliński[1] | Katarzyna Dudek[1] | Jan W. Arntzen[2] | Gemma Palomar[1] | Marta Niedzicka[1] | Anna Fijarczyk[3] | Marcin Liana[4] | Dan Cogălniceanu[5] | Wiesław Babik[1]

[1]Institute of Environmental Sciences, Faculty of Biology, Jagiellonian University, Kraków, Poland

[2]Naturalis Biodiversity Center, Leiden, The Netherlands

[3]Département de Biologie, Faculté des Sciences et de génie, Université Laval, Québec, QC, Canada

[4]Kraków, Poland

[5]Faculty of Natural Sciences and Agricultural Sciences, University Ovidius Constanţa, Constanţa, Romania

Correspondence
Piotr Zieliński, Institute of Environmental Sciences, Faculty of Biology, Jagiellonian University, Kraków, Poland.
Email: piotr.zielinski@uj.edu.pl

Wiesław Babik, Institute of Environmental Sciences, Jagiellonian University, Kraków, Poland.
Email: wieslaw.babik@uj.edu.pl

Funding information
Polish National Science Centre, Grant/Award Number: 2014/15/B/NZ8/00250; Jagiellonian University, Grant/Award Number: DS/WB/INoS/762/2018

Abstract

Genomic heterogeneity of divergence between hybridizing species may reflect heterogeneity of introgression, but also processes unrelated to hybridization. Heterogeneous introgression and its repeatability can be directly tested in natural hybrid zones by examining multiple transects. Here, we studied hybrid zones between the European newts *Lissotriton montandoni* and two lineages of *Lissotriton vulgaris*, with replicate transects within each zone. Over 1,000 nuclear genes located on a linkage map and mitochondrial DNA were investigated using geographical and genomic clines. Overall, the five transects were all similar, showing hallmarks of strong reproductive isolation: bimodal distribution of genotypes in central populations and narrow allele frequency clines. However, the extent of introgression differed between the zones, possibly as a consequence of their different ages, as suggested by the analysis of heterozygosity runs in diagnostic markers. In three transects genomic signatures of small-scale (~2 km) zone movements were detected. We found limited overlap of cline outliers between transects, and only weak evidence of stronger differentiation of introgression between zones than between transects within zones. Introgression was heterogeneous across linkage groups, with patterns of heterogeneity similar between transects and zones. Predefined candidates for increased or reduced introgression exhibited only a subtle tendency in the expected direction, suggesting that interspecific differentiation is not a reliable indicator for the strength of introgression. These hierarchically sampled hybrid zones of apparently different ages show how introgression unfolds with time and offer an excellent opportunity to dissect the dynamics of hybridization and architecture of reproductive isolation at advanced stages of speciation.

KEYWORDS
differential introgression, hybrid zone, newts, replicated transects

1 | INTRODUCTION

Understanding the effect of hybridization on interspecific divergence is at the heart of speciation research (Abbott et al., 2013; Seehausen et al., 2014). Introgression may differ between genomic regions, depending on the strength of selection counteracting or promoting genetic exchange (Baack & Rieseberg, 2007; Hedrick, 2013; Roux, Tsagkogeorga, Bierne, & Galtier, 2013). It was previously broadly accepted that genomic regions of elevated differentiation experience reduced introgression because they contain genes divergently selected between species and underlying reproductive isolation (Turner, Hahn, & Nuzhdin, 2005). To identify such genes, genomic differentiation has been characterized between many hybridizing taxa (Ellegren et al., 2012; Poelstra et al., 2014). More recently, however, it was pointed out that such islands of divergence are not necessarily due to reduced introgression. Instead, they may emerge under complete isolation in regions of low recombination, where selection at linked sites reduces diversity more than elsewhere in the genome (Burri et al., 2015; Cruickshank & Hahn, 2014; Vijay et al., 2016; Wolf & Ellegren, 2017). Therefore, to reveal the architecture of interactions between genomes of incipient species, it is crucial to distinguish introgression from other processes that produce similar patterns (Ravinet et al., 2017).

Hybrid zones are viewed as natural arenas where combinations of mutations from different species are functionally tested. Hybrid zones act on genomes as semipermeable membranes that prevent flow of the genes involved in reproductive isolation and genes linked to them, but allow relatively unimpeded introgression of advantageous alleles (Barton, 1979; Piálek & Barton, 1997). Regions that introgress more easily than the genomic background are likely to contain adaptive variants (Fijarczyk, Dudek, Niedzicka, & Babik, 2018; Hedrick, 2013). Thus, hybrid zones provide unique opportunity for direct investigation of genomic heterogeneity of introgression, while minimizing the effect of confounding processes (Harrison & Larson, 2016; Payseur, 2010).

Both the strength and the genomic patterns of introgression may vary along a contact zone (Harrison & Larson, 2016; Mandeville et al., 2017; Nolte, Gompert, & Buerkle, 2009; Teeter et al., 2010). Therefore, investigation of multiple transects through a zone is crucial to test repeatability of the genomic landscape of introgression, which in turn informs about the architecture of isolation across species' ranges (Harrison & Larson, 2016; Janoušek et al., 2012). However, to date the results of such studies have not been conclusive. In many cases patterns were not concordant between transects, with differences in extrinsic, environmental factors or variable genetic architecture of reproductive isolation commonly invoked as explanations (Harrison & Larson, 2016). It has also been suggested that differences between transects may arise from chance associations among incompatibilities (i.e., coupling) in various parts of the contact zone (Barton & De Cara, 2009; Bierne, Welch, Loire, Bonhomme, & David, 2011; Gagnaire, Normandeau, Pavey, & Bernatchez, 2013). If either species exhibits genetic substructure, different evolutionary lineages would hybridize in different parts of the contact zone. If some of them harbour fewer alleles incompatible with alleles of the other species, introgression could be easier, because of weaker intrinsic selection against hybrids. Thus, intraspecific variation in the genetic architecture of reproductive isolation may cause variable patterns of introgression between species across their contact zone (Cutter, 2012). Patterns of introgression may also be influenced by demographic processes, such as neutral introgression from the resident to the invading species (Currat, Ruedi, Petit, & Excoffier, 2008), which allows detection of hybrid zone movement (Wielstra et al., 2017). Differences in introgression between transects may also be due to differences in the age of secondary contact. The longer the time since contact, the higher the chance for neutral introgression but also for reinforcement to evolve. Whenever distinct intraspecific lineages hybridize in different regions of the zone, hierarchical sampling with replicated transects within regions is highly informative, as it allows the effect of lineage to be separated from other processes.

Here we investigate two hybrid zones formed by the Carpathian newt (*Lissotriton montandoni*, Lm) and the smooth newt (*Lissotriton vulgaris*, Lv), and assess differences between the zones and repeatability within the zones by investigating introgression in replicated transects (Figure 1a). The zones result from hybridization between Lm and two distinct lineages of Lv (divergence ~1 million years ago [Pabijan, Zielinski, Dudek, Stuglik, & Babik, 2017; Zieliński, Nadachowska-Brzyska, Dudek, & Babik, 2016]): Lm hybridizes with *L. v. ampelensis* within the Carpathian Basin (the IN zone) and with *L. v. vulgaris* outside the Carpathian basin (the OUT zone). Previous studies have suggested some spatial, temporal and genomic heterogeneity of introgression (Stuglik & Babik, 2016; Zieliński, Dudek, Stuglik, Liana, & Babik, 2014; Zieliński et al., 2016, 2013). Demographic models have suggested asymmetry of historical gene flow, predominantly into Lm, within the IN zone (Zieliński et al., 2016). Heterogeneity and asymmetry of introgression were also suggested by detection of two chromosomal blocks grouping genes exhibiting segregation distortion in F_2 interspecific hybrids (Niedzicka, Dudek, Fijarczyk, Zieliński, & Babik, 2017).

We use broad genomic and geographical sampling to address four specific objectives: (a) test for differences in recent/ongoing introgression between the two zones, (b) assess genomic clustering of genes with similar patterns of introgression, (c) test if introgression of genes highly differentiated between species or showing segregation distortion in F_2 hybrids is restricted and (d) test for increased introgression of previously identified putative targets of balancing selection (Fijarczyk, Dudek, & Babik, 2016; Fijarczyk et al., 2018). To address the first objective, we investigated genome-wide introgression in two transects through the IN zone and three transects through the OUT zone. Geographical heterogeneity of introgression was hypothesized to result from intraspecific polymorphism of genes underlying reproductive isolation within Lv (Zieliński et al., 2016). If the architecture of reproductive isolation between Lm and the two Lv lineages differs, then we would expect similar patterns of introgression between

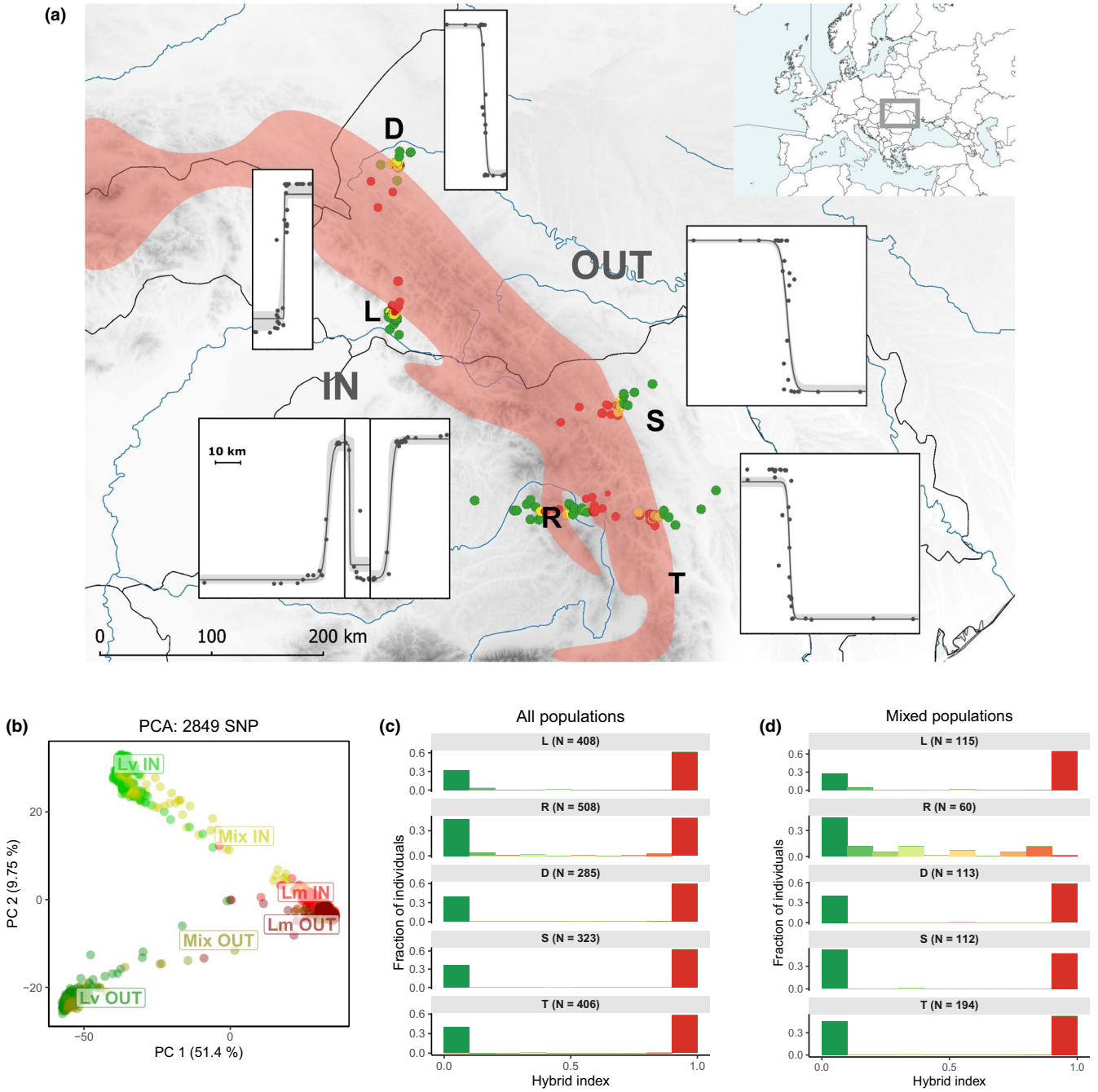

FIGURE 1 Sampling and general patterns of hybridization. (a) Transects and localities. Red dots—*Lissotriton montandoni* (Lm), green dots—*Lissotriton vulgaris* (Lv), yellow dots—mixed populations or populations where hybrids were found. Circle size corresponds to sample size. Genome-wide *Q* clines are shown for each transect; in the R transect, separate clines were fitted for each transition zone (R1–R3). (b) Ordination of all individuals in the space of the first two principal components, based on genotypes in 2,849 SNPs. Individuals were colour-coded according to the population type (Lm, Lv, mix) and the hybrid zone. The percentage of variation explained by principal components is given in parentheses. (c, d) Distribution of hybrid index in each transect in all (c) and mixed populations (d) [Colour figure can be viewed at wileyonlinelibrary.com]

transects within each zone but consistent differences between the zones (Cutter, 2012; Seehausen et al., 2014; Teeter et al., 2010). To address the second objective, we tested if genes with restricted or increased introgression cluster on the linkage map. For the third and fourth objectives, we compared introgression of preselected candidate genes to the genomic background. If introgression of highly differentiated genes is indeed restricted, they would mark genomic regions currently associated with reproductive isolation and/or those determining species-specific adaptations. A similar logic applied to genes showing segregation distortion in hybrids. Increased introgression of genes evolving under balancing selection would support the adaptive nature of their introgression.

2 | METHODS

2.1 | Samples and markers

The study area is located at lower mountain elevations in the Eastern Carpathians, where the ranges of the two species meet (Figure 1a). Newts were sampled along five transects through two hybrid zones. Lypcha (L) and Remetea (R) transects are located in the IN zone. Remetea is a compound transect that spans three consecutive transition zones between Lm and Lv: R1, R2, R3 (Figure 1a). Drohobych (D), Suceava (S) and Tazlău (T) are situated in the OUT zone. Within each zone, the transects were separated by a minimum distance of 100 km. Such a sampling design should allow us to disentangle random differences between transects from systematic differences in introgression between the two zones. Adult newts were sampled in water during the breeding season (March–May). The contact zone was searched thoroughly and all identified breeding sites were sampled. Individual ponds were treated as local populations. Samples from ponds where both species or morphologically identifiable hybrids occurred were labelled as mixed. Tailtip biopsies were taken and the animals were released afterwards. Tissues were stored in 96% ethanol and DNA was extracted using the Wizard Genomic DNA Purification Kit (Promega).

Fragments of 1,233 nuclear genes and one mitochondrial protein coding gene were resequenced with Molecular Inversion Probes (MIPs), following the protocol of Niedzicka, Fijarczyk, Dudek, Stuglik, and Babik (2016). A total of 1,485 MIPs, each targeting 112 bp of genomic sequence, were used (for details see Appendix S1), in most cases a single MIP per gene. For all nuclear markers, Mendelian segregation within families and a lack of departures from Hardy–Weinberg (HW) expectations in pure populations of each species were confirmed. Most of the genes (92%) were placed on the linkage map (Niedzicka et al., 2017). Single nucleotide polymorphism (SNP) calling was performed with GATK (details in Appendix S1). (Depristo et al., 2011; McKenna et al., 2010) Only biallelic (i.e., meeting the infinite sites mutation model) SNPs with <20% of missing data and global minor allele frequency (MAF) ≥5% were used in analyses. The analysis of mitochondrial DNA (mtDNA) was complicated by its extensive interspecific introgression, both historical and recent (Babik et al., 2005; Pabijan et al., 2017; Zieliński et al., 2013). As a consequence, all five distinct mtDNA lineages that occur in Lm and Lv in Central Europe, and to some extent also haplotypes within lineages, are shared between species. Depending on the geographical region, species may be locally diagnosable in mtDNA or not. Therefore, we applied transect-specific approaches for the analysis of mtDNA introgression and such analysis was not possible in the R transect due to negligible interspecific differentiation (details in Appendix S1).

2.2 | Genome-wide admixture, ancestry blocks and age of the zones

We used ADMIXTURE (Alexander, Novembre, & Lange, 2009) with $K = 2$ to estimate the proportion of Lv and Lm ancestry for each individual. Individual genome-wide admixture (Q) was treated as the hybrid index, a continuous variable taking values from 0 (Lv) to 1 (Lm). When discrete admixture categories were required, individuals with $Q \geq 0.9$ were classified as Lm, those with $0.9 > Q > 0.1$ as hybrids and those with $Q \leq 0.1$ as Lv (Aboim, Mavarez, Bernatchez, & Coelho, 2010; Taylor, Anderson, Zavalaga, & Friesen, 2012). The average Q was used as a proxy of the genomic composition of populations. To assess differentiation between species and between zones we used principal component analysis (PCA) in ADEGENET (Jombart, 2008).

The size of ancestry blocks in admixed genomes provides information about the age of the hybrid zones and also about the genomic architecture of reproductive isolation (Hvala, Frayer, & Payseur, 2018; Sedghifar, Brandvain, & Ralph, 2016; Sedghifar, Brandvain, Ralph, & Coop, 2015). Although our genome-wide data were not dense enough to confidently infer the length of ancestry blocks, they were still useful to compare the relative age of the zones. In each transect we estimated, for each individual with a minimum admixture of 0.01, the average length (in cM) of heterozygosity runs of mapped diagnostic ($\Delta p \geq 0.95$) markers. The length of runs was then compared between zones using a linear model in r (R Core Team, 2018), taking each individual's admixture (Q for $Q \leq 0.5$ or $1 - Q$ for $Q > 0.5$) as a covariate.

2.3 | Environmental correlates

Environmental correlates of the Lm–Lv contact zone were analysed over transects by logistic regression with Lv versus Lm as the dependent variable and three preselected environmental variables as candidate explanatory variables. A linear weighting factor was applied such that genetically pure populations, with Q-values at either zero (Lv) or unity (Lm), had maximum impact at the expense of fully admixed populations ($Q = 0.5$) that had no impact, while the total population sample size was unaffected. Environmental data considered were altitude, forestation and relief. These parameters were chosen on the basis of our expert knowledge on the ecological preferences of both *Lissotriton* species. Environmental data were extracted from Google Earth imagery over a 200-m-radius circular area around the sampling locality. Forestation was expressed as arcsin-transformed per cent cover, altitude was in metres above sea level (m a.s.l.) averaged over 25 regularly distributed point data and relief was the variance of the altitude data. Explanatory variables entering the model were selected on the basis of the Akaike information criterion (AIC) in SPSS 20 (IBM SPSS, 2011). Model fit was expressed as the area under the curve (AUC) statistic. The selected environmental models were compared with "geographical position at the transect (km)" that served as a null model.

2.4 | Measuring introgression in hybrid zones

Introgression of individual markers was characterized using both geographic and genomic clines, because these two methods provide complementary information. Geographic clines follow the transition of ancestry from one species to another along a spatial, usually

one-dimensional, transect, and cline parameters are expressed in kilometres (Barton & Gale, 1993). Genomic clines measure introgression of particular loci with respect to the genomic average, that is, the individual's hybrid index (Gompert, Parchman, & Buerkle, 2012). Both methods are powerful and have been used widely, but each has its advantages and limitations (Gompert, Mandeville, & Buerkle, 2017). Geographic clines assume monotonic change in allele frequency along the transect, which is often violated in mosaic hybrid zones. Geographic clines usually rely on diagnostic or nearly diagnostic markers, which has been criticized as the source of a potential bias (Payseur, 2010). In addition, the optimal scale for geographic clines is not clear. Genomic clines overcome these limitations, but they do not fully incorporate variation due to drift, which may increase the incidence of false positives (Gompert, Parchman, et al., 2012).

2.5 | Geographic clines—spatial patterns of introgression

2.5.1 | Transects

Because sampled localities did not fall on a straight line and transition between species in the centre of the zone departed from smooth clines, we used the approach of Yanchukov et al. (2006) to determine the position of localities along a one-dimensional transect. First, we obtained, using QGIS 2.8 (QGIS Development Team, 2016), the isoline representing the population average $Q = 0.5$ in two-dimensional space. Second, we determined the shortest distance between each locality and that line, and ordered the localities according to this distance, thus reducing the data to a one-dimensional transect. Finally, we selected reference populations, located at the ends of each transect, to obtain allele frequencies in the parental species.

2.5.2 | Data

Within each transect, we limited the analysis of geographical clines to genes with at least one SNP showing the allele frequency difference between the reference populations (Δp) ≥ 0.6. The single most differentiated SNP per gene was used. Nonrandom mating within populations and influx of parental individuals to the zone cause nonindependence between alleles within the locus. To take this effect into account, we adjusted sample sizes by calculating for each gene in each population the effective sample size as $N_{eff} = 2N/(1 + F_{IS})$.

2.5.3 | Cline models

Geographic clines were fitted with the R package HZAR (Derryberry, Derryberry, Maley, & Brumfield, 2014). In HZAR 15 cline models are available (plus a null model of no allele frequency cline). They differ in the number and combination of estimated parameters: there are three options for allele frequencies at the ends of the cline and five options to fit the cline tails. Allele frequencies at the ends of the cline can be: (a) estimated, (b) fixed to the observed values or (c) fixed to 0 and 1. Tail shape parameters can be as follows: (a) no tails fitted—a single smooth cline equation describes the entire cline, or tails modelled as exponential decay curves, (b) left or (c) right tail fitted (introgression into a single species), (d) both tails as mirrors (symmetric introgression into both species) or (e) each tail estimated separately (asymmetric introgression into both species). All models estimate the location of the cline centre (c) and the width (1/maximum slope, w) of the cline. The location of the cline centre informs about the shift towards either species and allows us to detect asymmetric introgression. Wider clines suggest increased introgression, and narrower clines reduced introgression.

2.5.4 | Fitting and comparing clines

Because our main goal was to compare cline centres and widths between genes and transects, we reduced the complexity that could result from selection of different cline models by fitting the same model for all genes in all transects: namely the estimated allele frequencies at the ends of the transect and no tails. This model was first compared to the null model of no cline. If the null model was rejected, c and w values were compared to the average genome-wide cline estimated from the population average Q (the Q cline). Individual clines were considered different from the Q cline if a model with the centre or width refitted with parameter values of the Q cline had an AIC score corrected for small sample size (ΔAICc) ≥ 5. Such genes were termed c- and w-outliers, respectively. Additionally, to test whether the zones differed in the extent of introgression beyond the centre of the transect, we compared (using the models with estimated allele frequencies at the ends of the transect) the fraction of genes for which clines with tails provided a better fit than the no tails model (Szymura & Barton, 1986). Within the R transect clines were fitted separately for the three transition zones.

2.6 | Genomic clines—genomic patterns of introgression

Genomic clines estimate the probability of a locus-specific ancestry with respect to the genome-wide admixture (Gompert & Buerkle, 2011; Szymura & Barton, 1986). A genomic cline has two parameters: α describes an excess (if positive) or a deficit (if negative) of a species (herein Lm) ancestry at a locus with respect to the genome-wide average (individual's hybrid index), while β describes an increased (if positive) or decreased (if negative) rate of transition in ancestry at a locus compared to the genome-wide average. Loci with unusual patterns of introgression can be identified as those that: (a) differ from the genome-wide average, and (b) are statistically unlikely given the genome-wide distribution of locus-specific introgression. If the 95% credibility interval of the posterior distribution of α or β for a given locus does not overlap zero, introgression different from the genome-wide average is inferred. If α or β estimates for a locus are unlikely given the genome-wide distribution of locus-specific introgression, we define such loci as α- and β-outliers, respectively. The locus is an outlier if the posterior estimates of α or β are not included

in q_n, which is the interval bounded by $n/2$ and $1 - n/2$ quantiles of the normal distribution $(0, \tau)$, where τ is the random-effect precision parameter, with $n = 0.05$ (Gompert, Parchman, et al., 2012). Introgression different from the genomic average (a) can be caused by both genetic drift and selection, whereas (b) outlier loci should be more common in regions where selection is strong but most of the genome evolves neutrally (Gompert, Parchman, et al., 2012). Thus, outliers are more conservative indicators of differential introgression caused by selection and we focused on these to identify genes with an unusual pattern of introgression.

In genomic cline analysis we used a model that accounts for linkage between genes, implemented in the program BGC (Gompert & Buerkle, 2011, 2012). Before the analysis populations needed to be assigned to one of three categories (two parental species and admixed). We made the assignment, for each transect separately, on the basis of the average *Q*-value for the populations. Populations with <3% admixture were classified as parental and the remaining populations were pooled as admixed. Such pooling was justified, as intraspecific genetic structure within transects was weak. The three transition zones (R1–R3) in transect R were pooled in the analysis of genomic clines.

2.7 | Comparison of geographic and genomic clines within transects

Genomic and geographic clines were compared within transects at the level of both outliers and parameter estimates. The overlap between the following categories was tested using chi-squared test with Yates's correction for continuity: (a) narrower *w*-outliers versus positive β-outliers, (b) broader *w*-outliers versus negative β-outliers, (c) *c*-outliers shifted to Lm versus negative α-outliers, and (d) *c*-outliers shifted to Lv versus positive α-outliers. For each transect we also calculated the Pearson correlation coefficient between *w* and β as well as between the geographic cline shift and α. In transect R a single set of genomic cline parameters calculated for the entire transect was compared to parameters of geographic clines estimated separately for each of the three transition zones.

2.8 | Comparison of introgression among transects and zones

To compare the overall introgression between the zones, we first tested for differences in the width of the genome-wide *Q* clines between transects. A single cline was fitted to data from all localities (or all localities within each zone) and we tested whether models with transect-specific cline widths fitted the data better than the model with width estimated for the overall cline. Second, we tested if the zones differed in the number of geographic and genomic cline outliers. Generalized linear mixed models (GLMMs) using the logit link function, with zone as a fixed factor, transect as a random factor nested within zone and gene as a crossed random factor, were fitted using glmer in LME4 (Bates et al., 2015). Third, to test repeatability of introgression, we checked for the overlap of geographic and genomic cline outliers across transects. Because the lists of genes differed slightly between transects, significance was tested with randomizations (10,000 replicates): in each replicate, the number of genes equal to the observed number of outliers were drawn at random from each transect and the number of shared genes between each pair of transects was recorded. These results formed the null distribution that we used to check whether sharing of outliers within and between zones was higher than by chance. Fourth, correlation between each pair of transects was tested for the parameters of geographic (*w*, shift) and genomic (α, β) clines for all genes.

2.9 | Genomic distribution of introgression patterns

Similarity between transects and zones in patterns of introgression can be manifested at the level of individual genes, but also at the level of genomic regions. To evaluate the latter, we tested whether any linkage group was enriched for geographical or genomic cline outliers. Significance was tested by shuffling (10^4 replicates) genes between linkage groups. To test whether the values of cline parameters were randomly distributed within linkage groups, we used a kernel smoothing and permutation test. This approach is similar to that of Bruneaux et al. (2013) and Ozerov et al. (2016), but we used the *locpoly* function within the KERNSMOOTH R package as in Palomar et al. (2017). With this approach, we took into account differences in marker density and increased the statistical power for detecting regions where several adjacent markers show exceptional patterns of introgression (for details see Bruneaux et al., 2013; Ozerov et al., 2016). For this analysis, we ran 10^6 permutations in R.

2.10 | Testing introgression of candidate genes

For four categories of candidate genes, we tested a priori expectations about their introgression compared to the genome-wide average (Fijarczyk et al., 2016, 2018; Payseur, 2010). Specifically, we tested for: (a) increased introgression of putative targets of balancing selection (BS genes), (b) reduced introgression of genes with elevated interspecific F_{ST} (F_{ST} genes), (c) reduced introgression of genes with elevated absolute interspecific sequence divergence, d*xy* (d*xy* genes), and (d) reduced and possibly asymmetric introgression of genes showing segregation distortion (highly distorted markers) in F_2 hybrids (HDM genes). Twelve single-copy BS genes were identified previously using two composite tests looking for an excess of high-frequency nonsynonymous polymorphisms and an excess of polymorphism over divergence (Fijarczyk et al., 2016). Forty-one F_{ST} genes and 33 d*xy* genes were identified from transcriptome resequencing of six Lm and six Lv individuals sampled to encompass the genetic diversity of the species (Stuglik & Babik, 2016). HDM genes identified in an F_2 hybrid family clustered in two blocks on linkage groups (LG) 4 and 10 (Niedzicka et al., 2017). The pattern of distortion was consistent with environment-dependent mortality of F_2 carrying Lm alleles, suggesting a role of these blocks in reproductive isolation.

First, we used the *t* test to check whether the estimates (average over all transects) of cline parameters *w* and β for candidate genes differed from those for the remaining genes; in the case of HDM, we also tested for the asymmetry of introgression (cline shift and α departing from zero). Second, the Cochran–Mantel–Haenszel test was used to check for an excess of candidate genes among *w*-outliers with significantly wider (BS) or narrower (F_{ST}/d*xy*/HDM) clines, among β-outliers with decreased (BS) or increased (F_{ST}/d*xy*/HDM) rate of ancestry transition, as well as among *c*- and α-outliers (HDM). Finally, also using the *t* test, we checked if Δp in candidate genes differed from that in other genes.

3 | RESULTS

3.1 | Strong genome-wide isolation but variable extents of introgression

We sampled and genotyped 1,930 individuals from 154 populations in five transects through two newt hybrid zones (Figure 1; Table S1). After quality filtering, 2,849 SNPs in 1,107 nuclear genes that met the criteria of MAF ≥0.05 and <20% missing genotypes were included in the analyses. Overall, 2.2% genotypes were missing. Repeatability of genotyping, estimated for 75 replicated samples and excluding concordant reference sites, was 99.2%.

Contact between the two species was narrow in all transects (Figure 1a; Figure S1), with the mean width of the genome-wide Q cline of 2.5 km, indicating an overall strong genome-wide isolation. However, patterns of hybridization differed between zones and transects. A single cline fitted to the combined data from all transects, although still narrow compared to dispersal abilities of newts (*w* = 4.5 km), was significantly wider than most transect-specific clines (Table 1). Also within zones transect-specific clines fitted the data better than a single zone-specific cline (Table 1).

The environmental parameters determining the location of transition, selected, alone or in combination, were altitude (for transects: R1, R3, S, T), forest cover (D, R2, R3) and relief (L). However, in four transects (R1–R3, S) a full species separation was achieved along the geographical axis, so that the null models outperformed or equalled the environmental models (Table S2). For all seven transects model fit was in excess of AUC = 0.83. *Lissotriton montandoni* (Lm) was found at higher altitudes than *Lissotriton vulgaris* (Lv) in environments with higher forest cover.

Although mixed populations were found in all transects, as much as 38% of populations OUT were mixed compared to only 18% IN (Figure 1a; Table S1), indicating rarer syntopy in the IN zone. Few individuals were strongly admixed (Figure 1b) and hybrids (i.e., individuals with 0.9 > *Q* > 0.1) were observed in a few localities. Although the majority (54%) of hybrids were found in mixed populations, they constituted only 1% (OUT) and 24% (IN) of newts there. The distribution of genotypes in mixed populations was highly bimodal in all transects except R, where the majority (53%) of newts in mixed populations were hybrids and "pure" Lm were rare (Figure 1d). A notable fraction (46%) of hybrids were detected in otherwise

TABLE 1 Parameters of geographical clines

		Q-cline		Clines for individual nuclear genes						mtDNA cline		r^2 Δ*p* with	
Zone	Transect	*c* (km) (L CI–H CI)	*w* (km) L CI–H CI)	*n* genes	Mean *w*	*SD*	CV	Mean *c*	*SD*	CV	*c* (km) (L CI–H CI)	*w* (km) (L CI–H CI)	*w*	β
IN	L	11.9 (11.27–11.95)	0.1 (0.00–3.49)[a]	638	0.7	1.2	1.7	11.8	0.3	0.02	10.9 (10.38–11.29)[a]	5.8 (4.57–7.23)[a]	0.01	.57
IN	R1	50.0 (49.30–51.26)	3.8 (0.52–6.89)[b]	617	3.4	3.1	0.9	50.2	1.1	0.02			0.02	.24
IN	R2	1.9 (1.39–3.34)	0.7 (0.00–2.61)[a]	584	1.1	2.0	1.9	2.2	0.8	0.34			0.02	
IN	R3	6.8 (5.91–7.58)	3.3 (1.10–5.23)[b]	612	3.6	3.7	1.1	6.9	1.0	0.14			0.03	
OUT	D	15.3 (15.18–15.45)	2.0 (1.65–2.50)[a,b]	861	1.8	0.6	0.3	15.4	0.2	0.01	15.4 (15.17–15.65)	2.2 (1.65–3.14)	0.00	.18
OUT	S	38.0 (37.65–38.27)	5.2 (4.37–6.26)	805	4.8	1.6	0.3	37.8	0.6	0.01	38.0 (37.39–38.52)	4.7 (3.42–6.64)	0.00	.19
OUT	T	17.2 (17.02–17.36)	2.1 (1.65–2.90)[a,b]	736	2.0	0.8	0.4	17.2	0.3	0.02	16.9 (14.22–21.23)	1.8 (0.03–48.05)	0.02	.33

Note: c, cline centre; CV, coefficient of variation; L and H CI, lower and upper limits of 95% credibility intervals, respectively; Q-cline, genome-wide cline estimated from population average Q provided by ADMIXTURE; *SD*, standard deviation; *w*, cline width; Δp, allele frequency difference between the reference populations (i.e., those at the ends of each transect).

[a]Parameter departs significantly from the cline fitted to combined data from all transects.

[b]Parameter departs significantly from the cline fitted to combined data from transects within the zone.

morphologically pure populations, but they constituted only 0.7% (OUT) and 5% (IN) of individuals found there. Taking all populations together the overall proportion of hybrids was low (5%), but they were more common IN (9%) than OUT (1%; Figure 1d), suggesting a more extensive hybridization in the IN zone. The zones differed also in the extent of introgression beyond the immediate contact (Figure S2). Along both transects of the IN zone, many individuals with low (0.03–0.1) genome-wide admixture were found: in the R transect almost all individuals had some admixture, while it was mostly limited to the Lv side of the L transect. In contrast, few individuals in the OUT zone showed any detectable admixture.

The per-individual average length of heterozygosity runs in diagnostic markers, a proxy for the introgression blocks, differed markedly between the zones—runs were longer in the OUT zone (GLM, $p = 1.8 \times 10^{-9}$), and their length increased faster with increased admixture (GLM, $p = 1.4 \times 10^{-7}$, Figure 2a). Runs of several cM or longer were observed in the OUT zone even in individuals with low genome-wide admixture (Figure 2b). These data suggest younger hybridization in the OUT zone.

3.2 | Geographical and genomic heterogeneity of introgression

Allele frequency differences between the reference populations (i.e., those at the ends of each transect) calculated for the most differentiated SNP per gene (Δp) were higher OUT than IN (GLMM, $p = 4 \times 10^{-4}$, Table S3). Δp values were also more strongly correlated between transects within than between zones (Table S3). The fraction of genes with at least one highly differentiated ($\Delta p \geq 0.6$) SNP was higher OUT (74%) than IN (57%). Although the difference could partly be a consequence of identification of markers, the randomly picked markers (see Appendix S1) followed the same pattern: the difference was lower (65% OUT vs. 53% IN), but highly significant (randomization test $p < 10^{-4}$). Thus, interspecific differentiation is notably lower in the IN zone.

For 97% of the genes (the average for all transects), the width and centre of geographic clines did not differ from these of the *Q* cline. The fraction of *w*-outliers did not differ between zones (GLMM, $p = .72$), while the fraction of *c*-outliers was higher in the OUT zone (4.7% vs. 1.9%, GLMM, $p = .009$). Numerous *c*-outliers were found in transects D, L and S (Figure S3) and in each case strong asymmetry was detected—in transects D and L most *c*-outliers were shifted towards Lv (chi-squared test, $p = 1.3 \times 10^{-7}$ and 2.0×10^{-5}), while in S a shift to Lm predominated ($p = 5.4 \times 10^{-13}$). Note, however, that in absolute terms the shift was small, generally <2 km (Figure S3). In three out of four transects where an mtDNA cline could be fitted, it did not depart from the *Q* cline—only in the L transect was the mtDNA cline wider than the *Q* cline and shifted towards Lv (Table 1). Variation between genes in estimates of cline parameters was higher in the IN zone (Table 1; Figure 3; Figure S3). Models with cline tails had better fit for 20% of genes in the IN zone and 25% in the OUT zone.

Differences between the zones were apparent in genomic clines (Table S4). Considering an excess of ancestry (α), measured as significant departure from zero, only 3% of genes differed from the genomic average in the OUT zone, versus 32% in the IN zone. Fewer genes were identified as α-outliers, but these also were more frequent in the IN zone (7.6% vs. 2.6%, GLMM, $p = 1.1 \times 10^{-15}$). No systematic excess of either species ancestry was found. The rate of transition in ancestry (β) differed from the genomic average for 63% of genes IN and 31% OUT, and the fraction of β-outliers was also

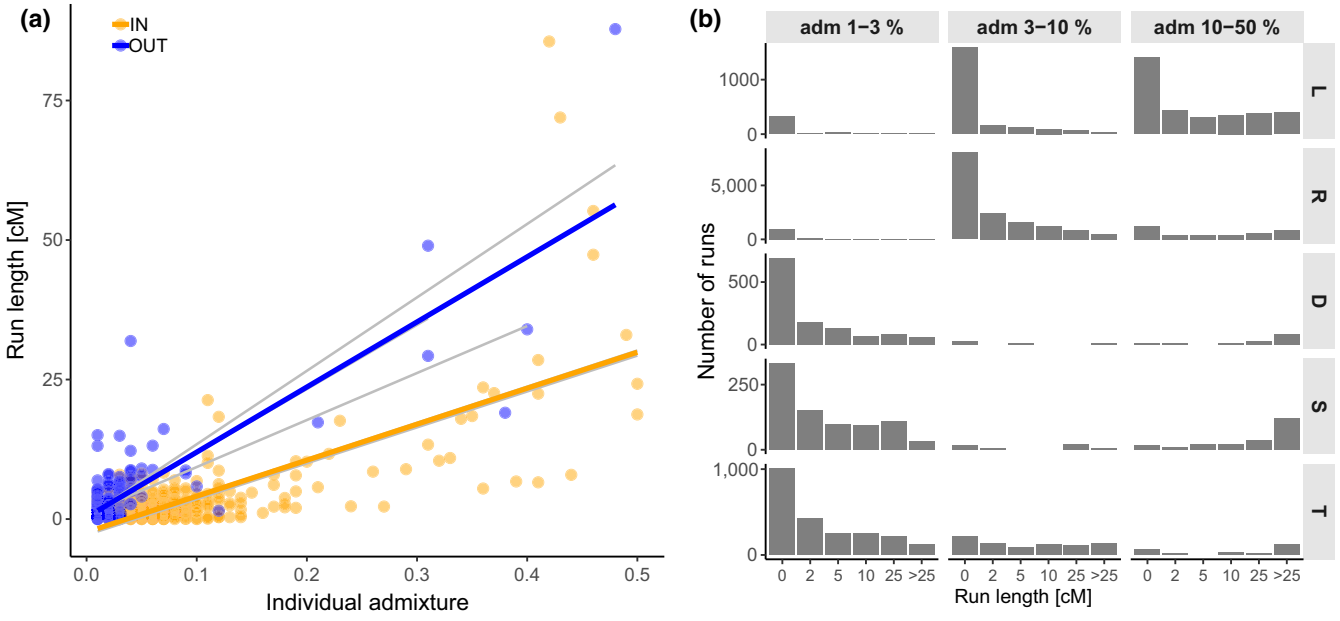

FIGURE 2 Length of heterozygosity runs in diagnostic markers. (a) Relationship between an individual's admixture and its average length of heterozygosity runs for all individuals with at least 1% admixture. The two zones were colour-coded; regression lines for individual transects are in grey. (b) Distribution of length of heterozygosity runs for various admixture classes; all runs, not per individual averages, were summarized [Colour figure can be viewed at wileyonlinelibrary.com]

higher in the IN zone (9.4% vs. 5.7%, GLMM, $p = 9.4 \times 10^{-7}$), indicating a more uniform rate of transition OUT. It is important to note that the results for genomic clines remained qualitatively the same for randomly picked markers (Table S4), suggesting that the analysis of genomic clines was not biased by marker selection. Within transects the overlap between geographic and genomic cline outliers was limited, especially between *w*- and β-outliers: after Bonferroni correction the overlap was significant only in transect R3 between the *w*-outliers with wider clines and β-outliers with reduced rate of transition. More overlap was detected between geographic and genomic clines shifted to Lm (transects R2 and R3) or to Lv (transects L and T). Estimates of the corresponding parameters of geographic and genomic clines were significantly correlated in most transects, although *r* did not exceed .4 in any case (Table S5).

Geographic cline outliers were generally transect-specific (Figure 3) and only genes shifted to Lv were shared within the IN zone more than expected by chance (Bonferroni-corrected $p < .0012$). In addition, parameters of geographic clines were weakly correlated between transects—most coefficients were not significant, the highest *r* reached .21 ($p = 1.8 \times 10^{-8}$) for cline width and .22 ($p = 6.3 \times 10^{-9}$) for shift, in both cases between transects S and T (Table S6). In genomic clines, β-outliers showed increased sharing within both zones, whereas α-outliers just within the OUT zone (Table S7). While sharing of geographic cline outliers between zones was not significant, genomic cline outliers (with increased and reduced rate of transition as well as an excess of Lv ancestry) were commonly shared between zones (all $p < .0012$). Considering all genes, the parameters of genomic clines were more strongly correlated between transects than those of geographic clines (Table S6). For α all correlations within zones were significant, although *r* reached only 0.25 ($p = 7.3 \times 10^{-15}$) between transects D and S, while for β most *r* values exceeded .4 and all were highly significant (Table S6). For all parameters of geographic and genomic clines, correlations between transects within zones exceeded those between transects from different zones (Table S6). The last observation suggests similarity in architecture of introgression within and differences between zones. However weighting the overall evidence, the support for the differences in the genomic architecture of introgression between the two zones was limited.

Cline outliers were not randomly distributed in the genome (Figures 3 and 4). Genes with narrower geographic clines were underrepresented, while genes with a reduced rate of ancestry transition ($\beta < 0$) were overrepresented on LG4 (permutation test, all Bonferroni-corrected $p \leq .03$). Genes with an increased rate of

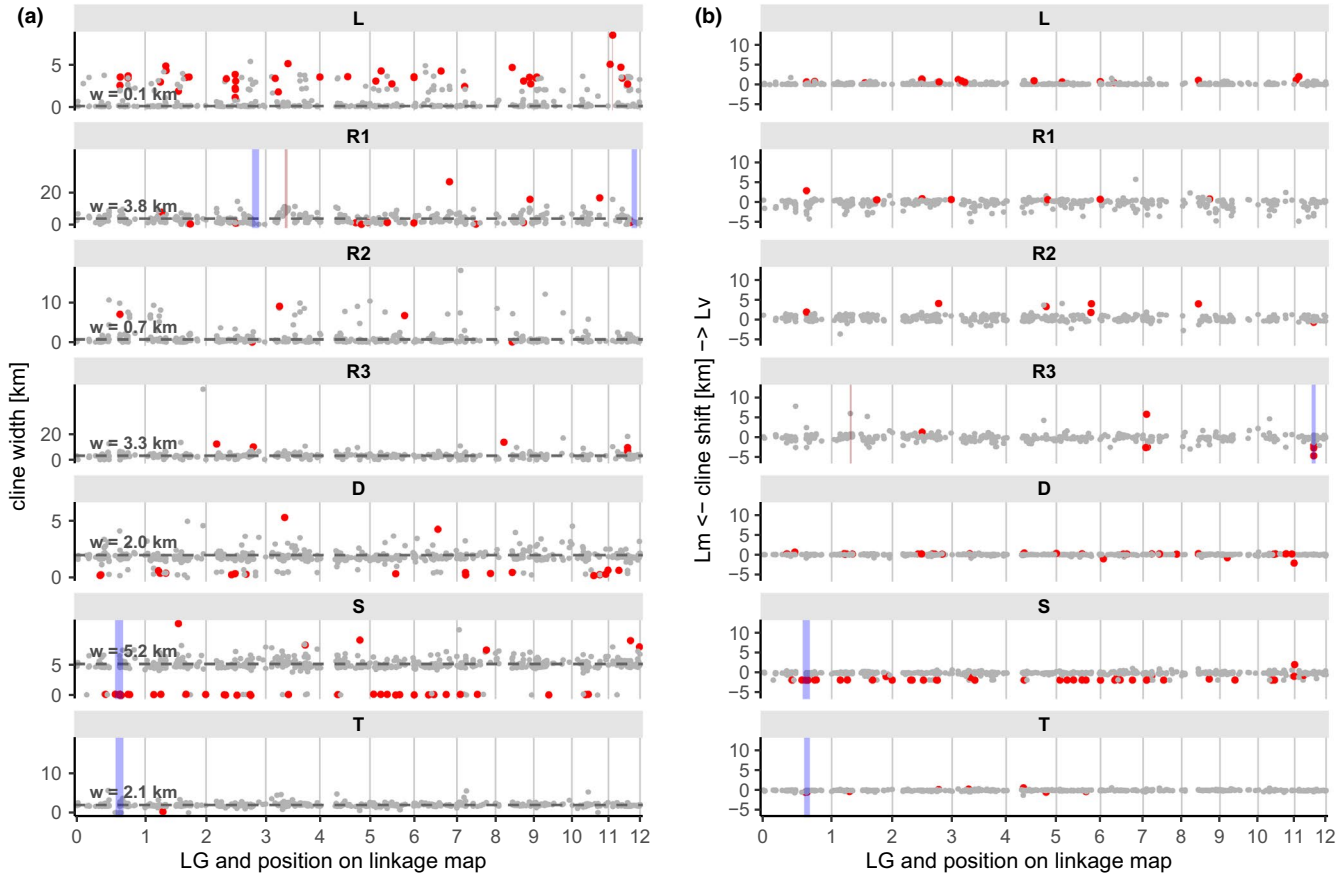

FIGURE 3 Genomic distribution of geographic cline parameters. Outliers from the genome-wide *Q* cline are marked in red. (a) Geographic cline width. (b) Geographic cline shift from the centre of the zone as defined by the location of the *Q* cline centre. Colour stripes indicate regions within linkage groups (LG) with an overrepresentation of: narrow (blue) or wide (pale red) clines in (a), and clines shifted towards *Lissotriton montandoni* (Lm, blue) or *Lissotriton vulgaris* (Lv, pale red) in (b) [Colour figure can be viewed at wileyonlinelibrary.com]

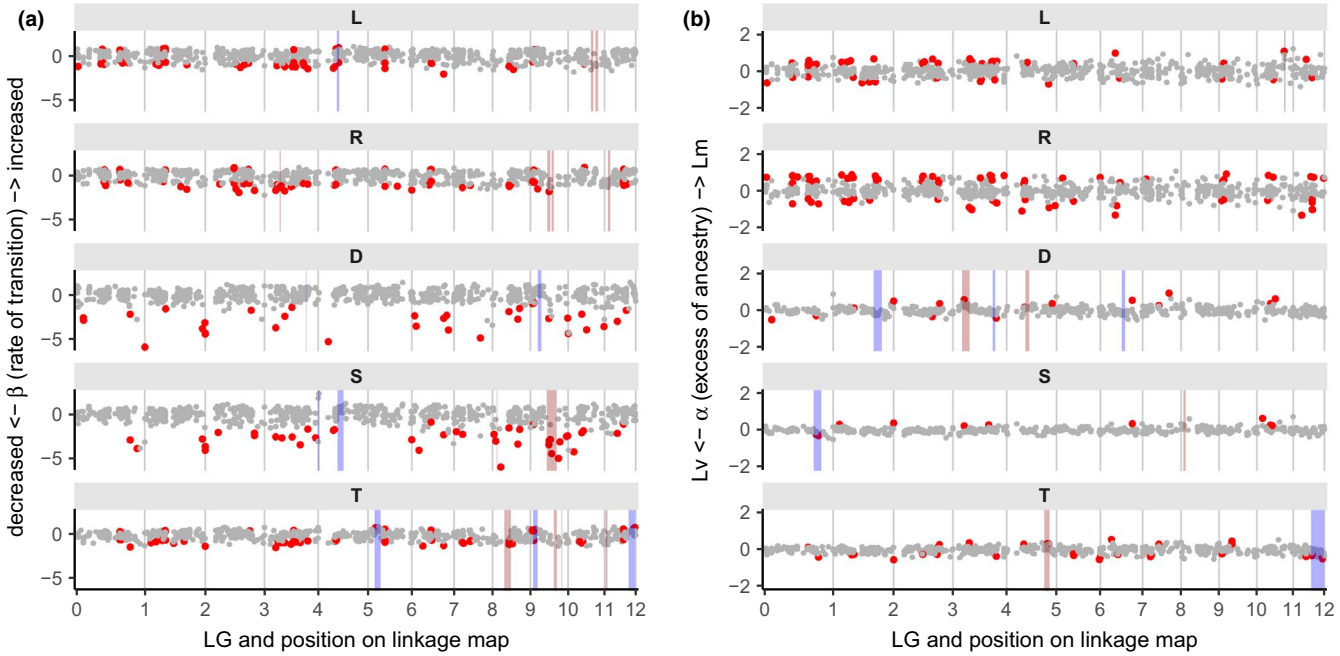

FIGURE 4 Genomic distribution of genomic cline parameters. Outliers are marked in red. (a) Rate of transition in ancestry (β). (b) Excess of ancestry with respect to the genome-wide average (α). Colour stripes indicate regions within linkage groups (LG) with an overrepresentation of genomic clines: with decreased (blue) or increased (pale red) rate of ancestry transition in (a), an excess of *Lissotriton montandoni* (Lm, pale red) or *Lissotriton vulgaris* (Lv, blue) ancestry in (b) [Colour figure can be viewed at wileyonlinelibrary.com]

ancestry transition ($\beta > 0$) were overrepresented on LG2 and LG5 and underrepresented on LG4 and LG9 (all $p \leq .04$). There was a strong excess of geographic clines shifted to Lm on LG1 and of genes with an excess of Lv ancestry on LG12 (both $p \leq .0024$). Finally, genes with an excess of Lm ancestry were underrepresented on LG9 ($p < .02$). Heterogeneity of introgression patterns was also detected within linkage groups, but there was little concordance between transects except for a single region on LG1 in transects S and T (Figures 3 and 4).

3.3 | Candidate genes are poor predictors of introgression

Genes with an increased interspecific F_{ST} showed a tendency for an increased rate of ancestry transition (*t* test, $p = .011$) and were overrepresented among $\beta > 0$ outliers (Mantel–Haenszel test, $p = .039$), while d*xy* genes tended to have narrower clines ($p = .020$, Table S8) and were also overrepresented among *w*-outliers with narrower clines (Mantel–Haenszel test, $p = .022$). F_{ST} genes had Δp higher and BS genes had Δp lower than the remaining genes (both $p = .0004$, Table S8). However, because Δp did not correlate with cline width (Table 1), such differences are difficult to interpret in the context of contemporary introgression. HDM genes had clines significantly shifted towards Lm (*t* test, $p = .0032$). BS genes showed similar patterns of introgression as other genes. Thus, although a tendency in the expected direction was observed in some cases, the overall differences in introgression between candidate genes and the rest of the genome were moderate and not consistent between measures.

4 | DISCUSSION

4.1 | Extent of introgression and the age of the zones

This study analysed hybrid zones between the newts *Lissotriton montandoni* and two distinct evolutionary lineages of *Lissotriton vulgaris*, with replicate transects within each zone, to assess repeatability of introgression between and within zones. Such hierarchical sampling should allow us to disentangle systematic differences between zones from random variation between transects (Janoušek et al., 2012). Overall, all transects were similar—allele frequency clines were narrow compared to the dispersal capabilities of newts (Smith & Green, 2005) and hybrids were uncommon, indicating strong reproductive isolation, due to assortative mating (Babik, Szymura, & Rafiński, 2003; Michalak & Rafiński, 1999) and possibly low fitness of hybrids. Local variation in environmental factors had some modulating effect on the characteristics of individual transects, but environmental associations were relatively weak compared to spatial position within the transect.

However, the zones did differ in the extent of admixture and introgression beyond syntopy, which were stronger in the IN zone. Such differences could reflect biological differences between the two lineages of Lv, affecting the strength or other characteristics of reproductive isolation with Lm. Some, albeit weak, support for such differences is provided by stronger correlation of cline parameters between transects within than between zones. However, the length of heterozygosity runs in diagnostic markers suggests that longer hybridization in the IN zone contributes to, or may even be sufficient to explain, the

differences in the extent of introgression. Our sampling of the genome was not dense enough to allow a formal analysis of the distribution of the length of introgression blocks to estimate the time of admixture (Racimo, Sankararaman, Nielsen, & Huerta-Sánchez, 2015; Sedghifar et al., 2016, 2015). Nevertheless, the differences between the zones were large enough to be detectable with the available density of markers. Ongoing hybridization complicates the use of the length of introgression tracts to compare the age of the contact, because the length of the tracts depends also on details of demographic history as well as the architecture of reproductive isolation (Hvala et al., 2018), information difficult to obtain in natural systems. As a consequence, we cannot rule out the possibility that demographic and biological differences between the zones affected observed differences in the length of introgression tracts. However, longer hybridization in the IN zone remains a plausible and probable explanation, especially in the light of biogeographical evidence—long-term persistence of Lv in the Carpathian Basin, in close proximity to the range of Lm (Babik et al., 2005; Pabijan et al., 2017).

A further aspect of the spatiotemporal dynamics of hybridization is the stability of the location of the hybrid zones. Hybrid zones can move surprisingly large distances and do so surprisingly often (Buggs, 2007; Wielstra, 2019). In this study we also found some indication of hybrid zone movement: in transect L, evidence from both nuclear markers and mtDNA suggests expansion of Lv into the area formerly inhabited by Lm, while in D and S nuclear markers suggest expansion of Lv and Lm, respectively. The spatial scale of these shifts, however, was small, on the order of ~2 km, not comparable to hybrid zone movement of several 100 km reported in the literature for *Triturus* newts, for example (Wielstra et al., 2017). Similar small-scale movements related to local de- and reforestation were previously suggested, based on patterns of mtDNA introgression, in an Lm–Lv hybrid zone in southern Poland (Babik et al., 2003). The relative stability of the location of the hybrid zone may be related to narrow ecological requirements of Lm, which is strongly associated with Carpathian-type beech forests (Zavadil, Pialek, & Dandova, 2003), determining the location of the zone within this ecotone.

Previous work on mtDNA phylogeography and nuclear gene genealogies (Babik et al., 2005; Pabijan et al., 2017; Zieliński et al., 2016, 2013) has testified to a long history of genetic exchange between Lm and Lv, the hybrid zones of which studied here appear to be a recent manifestation. Note that multiple lines of evidence suggest more extensive genetic exchange between species within the Carpathian Basin, where the IN zone is located. Whether this is due to biological features of the Lv lineage occurring there, environmental factors that have allowed more extensive, longer contact between the species there, or both is an interesting and open question that could be addressed using landscape genomics approaches.

4.2 | Geographical and genomic heterogeneity of introgression

The age of the zones may explain most observed differences in the extent of introgression, although biological differences between the evolutionary lineages of Lv that hybridize with Lm could still be manifested at the level of individual genes, genomic regions or entire chromosomes. Looking at cline outliers, we found generally more of these in the IN zone, which can be explained by a longer period of hybridization. However, the outliers showed limited overlap between transects and limited evidence for higher similarity between transects within a zone than between zones. Thus, either processes governing introgression of particular genomic regions are deterministic, but largely transect-specific, or geographical heterogeneity reflects random variation in the width and location of clines. Such random variation can be caused by drift and may lead to the detection of false outliers ("cline wobbling," Polechová & Barton, 2011).

Introgression was heterogeneous across the genome, with the patterns of heterogeneity broadly similar between transects and zones. Most notably, more extensive introgression was inferred on LG4. *Lissotriton* newts have poorly differentiated, recombining sex chromosomes (Zbożeń & Rafiński, 1993) so we were not able to test for reduced introgression of X-linked genes, a phenomenon observed in multiple studies (reviewed by Payseur & Rieseberg, 2016). Overall, we did not find much evidence for systematic differences in reproductive isolation between Lm and the two lineages of Lv involved in hybridization in the IN and OUT zones. The genomic barrier to introgression is strong but its strength appears somewhat variable throughout the genome and rather similar in various parts of the contact zone. Accordingly, our results did not confirm earlier suggestions of differences in the genomic architecture of reproductive isolation between Lm and the two lineages of Lv (Zieliński et al., 2014, 2016). Yet such differences may exist at finer genomic scales and their detection would then require sampling of the genome at higher resolution.

4.3 | Introgression of candidate genes differs little from the genomic average

Several categories of genes were selected a priori to compare introgression between them and the rest of the genome. Candidates with an expected increase and an expected reduction of introgression level were both included. Although a tendency in the expected direction was found in some cases, in general the differences were subtle and could not establish consistent differences between candidates and the remaining genes. We found a significant geographical cline shift towards Lm in HDM genes, which is in agreement with the direction of segregation distortion in F_2 hybrids (Niedzicka et al., 2017). In laboratory F_2 hybrids with HDM, Lm alleles showed increased, probably environment-dependent mortality, so introgression of Lv alleles into the Lm part of the zone could be expected.

We found much lower divergence but no evidence for increased introgression in BS genes compared to the genomic background. It is thus possible that these genes do not currently introgress more easily, but it is also possible that introgression extended beyond the scale of our sampling, making its detection difficult with the use of cline analysis. Indeed, we found in the IN zone pervasive introgression of major histocompatibility complex (MHC) genes, a prime target

of balancing selection, although cline parameters for MHC were not distinguishable from the genome-wide Q cline (Dudek, Gaczorek, Zieliński, Babik, 2019). Overall, our data indicate that genes at the tails of the genome-wide distribution of divergence between species only occasionally are found in the tails of the introgression—as predicted if processes other than introgression dominated the buildup of genomic differentiation between hybridizing species (Wolf & Ellegren, 2017). In addition, evidence from other hybrid zones is mixed with a tendency for reduced introgression of highly differentiated genes found in some studies (Gompert, Lucas, et al., 2012; Parchman et al., 2013) but little or none in others (Hamilton, Lexer, & Aitken, 2013; Larson, White, Ross, & Harrison, 2014; Nolte et al., 2009). Our study thus testifies to the power of hybrid zones as a tool for testing hypotheses about differential introgression derived from other sources.

5 | CONCLUSIONS

The analysis of multiple hierarchically sampled transects through *Lissotriton montandoni* × *Lissotriton vulgaris* hybrid zones revealed strong genome-wide isolation. However, despite this and despite an apparently long history of genetic exchange, there is still room for differential introgression and hybrid zones still act as selective filters. Thus, our work adds to the small but growing number of studies that demonstrate heterogeneous introgression at advanced stages of speciation. The differences between the zones in the extent of introgression may probably be explained by the differences in the age of current contacts. This allows us to study how introgression unfolds with time in a single system, but at the same time complicates the answer to the question about differences between the zones in the genomic architecture of reproductive isolation. Fully addressing this question would require the use of emerging powerful methods based on the length distribution of introgression blocks. The studied hybrid zones are an excellent system for such research as the time of the contact, although different for the zones, is in both cases suitable for such analyses.

ACKNOWLEDGEMENTS

We are grateful to Maciej Pabijan who helped with sample collection and provided valuable comments. Oleksandr Zinenko helped with sampling in Ukraine. We thank Ben Wielstra for his help with running the HZAR package. Many thanks to Mateusz Buczek for his help in DNA extraction. The work was supported by the Polish National Science Centre grant no. 2014/15/B/NZ8/00250 to P.Z. and by the Jagiellonian University grant no. DS/WB/INoS/762/2018.

AUTHOR CONTRIBUTIONS

P.Z. and W.B. designed the research. K.D. performed the laboratory procedures. P.Z., W.B., J.W.A., G.P., M.N., A.F., M.L. and D.C. performed research. P.Z. and W.B. contributed analytical tools. P.Z., W.B., J.W.A. and G.P. analysed the data. P.Z. and W.B. wrote the manuscript (with contribution from other authors).

DATA AVAILABILITY STATEMENT

Variant calling (VCF) files, mtDNA sequences, genotypes used for ADMIXTURE, per-gene population allele frequency data used for HZAR, genotype data used for genomic clines and a file with individual admixture proportions as well as assignment of individuals to populations are available on the Dryad Digital Repository entry https://doi.org/10.5061/dryad.918m0n2.

ORCID

Piotr Zieliński https://orcid.org/0000-0002-8533-6800
Jan W. Arntzen https://orcid.org/0000-0003-3229-5993
Dan Cogălniceanu https://orcid.org/0000-0003-2959-014X
Wiesław Babik https://orcid.org/0000-0002-1698-6615

REFERENCES

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., … Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, *26*(2), 229–246. https://doi.org/10.1111/j.1420-9101.2012.02599.x

Aboim, M., Mavarez, J., Bernatchez, L., & Coelho, M. (2010). Introgressive hybridization between two Iberian endemic cyprinid fish: A comparison between two independent hybrid zones. *Journal of Evolutionary Biology*, *23*(4), 817–828. https://doi.org/10.1111/j.1420-9101.2010.01953.x

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664. https://doi.org/10.1101/gr.094052.109

Baack, E. J., & Rieseberg, L. H. (2007). A genomic view of introgression and hybrid speciation. *Current Opinion in Genetics & Development*, *17*(6), 513–518. https://doi.org/10.1016/j.gde.2007.09.001

Babik, W., Branicki, W., Crnobrnja-isailovic, J., Cogalniceanu, D., Sas, I., Olgun, K., … Arntzen, J. W. (2005). Phylogeography of two European newt species - discordance between mtDNA and morphology. *Molecular Ecology*, *14*(8), 2475–2491. https://doi.org/10.1111/j.1365-294X.2005.02605.x

Babik, W., Szymura, J. M., & Rafiński, J. (2003). Nuclear markers, mitochondrial DNA and male secondary sexual traits variation in a newt hybrid zone (*Triturus vulgaris* × *T. montandoni*). *Molecular Ecology*, *12*(7), 1913–1930.

Barton, N. H. (1979). Gene flow past a cline. *Heredity*, *43*(3), 333–339. https://doi.org/10.1038/hdy.1979.86

Barton, N. H., & De Cara, M. A. R. (2009). The evolution of strong reproductive isolation. *Evolution*, *63*(5), 1171–1190. https://doi.org/10.1111/j.1558-5646.2009.00622.x

Barton, N. H., & Gale, K. S. (1993). Genetic analysis of hybrid zones. In R. G. Harrison (Ed.), *Hybrid zones and the evolutionary process* (pp. 13–45). New York, NY: Oxford University Press.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., … Rcpp, L. (2015). Package 'lme4'. *Convergence*, *12*, 1.

Bierne, N., Welch, J., Loire, E., Bonhomme, F., & David, P. (2011). The coupling hypothesis: Why genome scans may fail to map local

adaptation genes. *Molecular Ecology*, *20*(10), 2044–2072. https://doi.org/10.1111/j.1365-294X.2011.05080.x

Bruneaux, M., Johnston, S. E., Herczeg, G., Merilä, J., Primmer, C. R., & Vasemägi, A. (2013). Molecular evolutionary and population genomic analysis of the nine-spined stickleback using a modified restriction-site-associated DNA tag approach. *Molecular Ecology*, *22*(3), 565–582. https://doi.org/10.1111/j.1365-294X.2012.05749.x

Buggs, R. (2007). Empirical study of hybrid zone movement. *Heredity*, *99*(3), 301. https://doi.org/10.1038/sj.hdy.6800997

Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., … Garamszegi, L. Z. (2015). Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Research*, *25*(11), 1656–1665.

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, *23*(13), 3133–3157. https://doi.org/10.1111/mec.12796

Currat, M., Ruedi, M., Petit, R. J., & Excoffier, L. (2008). The hidden side of invasions: Massive introgression by local genes. *Evolution*, *62*(8), 1908–1920. https://doi.org/10.1111/j.1558-5646.2008.00413.x

Cutter, A. D. (2012). The polymorphic prelude to Bateson–Dobzhansky–Muller incompatibilities. *Trends in Ecology & Evolution*, *27*(4), 209–218. https://doi.org/10.1016/j.tree.2011.11.004

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., … Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics*, *43*(5), 491–498.

Derryberry, E. P., Derryberry, G. E., Maley, J. M., & Brumfield, R. T. (2014). HZAR: Hybrid zone analysis using an R software package. *Molecular Ecology Resources*, *14*(3), 652–663.

Dudek, K., Gaczorek, T. S., Zieliński, P., & Babik, W. (2019). Massive introgression of MHC genes in newt hybrid zones. *Molecular Ecology*, https://doi.org/10.1111/mec.15254.

Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., … Wolf, J. B. W. (2012). The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature*, *491*(7426), 756. https://doi.org/10.1038/nature11584

Fijarczyk, A., Dudek, K., & Babik, W. (2016). Selective landscapes in newt immune genes inferred from patterns of nucleotide variation. *Genome Biology and Evolution*, *8*(11), 3417–3432. https://doi.org/10.1093/gbe/evw236

Fijarczyk, A., Dudek, K., Niedzicka, M., & Babik, W. (2018). Balancing selection and introgression of newt immune-response genes. *Proceedings of the Royal Society of London B: Biological Sciences*, *285*(1884), 20180819.

Gagnaire, P. A., Normandeau, E., Pavey, S. A., & Bernatchez, L. (2013). Mapping phenotypic, expression and transmission ratio distortion QTL using RAD markers in the Lake Whitefish (*Coregonus clupeaformis*). *Molecular Ecology*, *22*(11), 3036–3048.

Gompert, Z., & Buerkle, C. A. (2011). Bayesian estimation of genomic clines. *Molecular Ecology*, *20*(10), 2111–2127. https://doi.org/10.1111/j.1365-294X.2011.05074.x

Gompert, Z., & Buerkle, C. (2012). bgc: Software for Bayesian estimation of genomic clines. *Molecular Ecology Resources*, *12*(6), 1168–1176.

Gompert, Z., Lucas, L. K., Nice, C. C., Fordyce, J. A., Forister, M. L., & Buerkle, C. A. (2012). Genomic regions with a history of divergent selection affect fitness of hybrids between two butterfly species. *Evolution: International Journal of Organic Evolution*, *66*(7), 2167–2181. https://doi.org/10.1111/j.1558-5646.2012.01587.x

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

Gompert, Z., Parchman, T. L., & Buerkle, C. A. (2012). Genomics of isolation in hybrids. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1587), 439–450. https://doi.org/10.1098/rstb.2011.0196

Hamilton, J. A., Lexer, C., & Aitken, S. N. (2013). Differential introgression reveals candidate genes for selection across a spruce (*Picea sitchensis × P. glauca*) hybrid zone. *New Phytologist*, *197*(3), 927–938.

Harrison, R. G., & Larson, E. L. (2016). Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. *Molecular Ecology*, *25*(11), 2454–2466.

Hedrick, P. W. (2013). Adaptive introgression in animals: Examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular Ecology*, *22*(18), 4606–4618. https://doi.org/10.1111/mec.12415

Hvala, J. A., Frayer, M. E., & Payseur, B. A. (2018). Signatures of hybridization and speciation in genomic patterns of ancestry. *Evolution*, *72*(8), 1540–1552. https://doi.org/10.1111/evo.13509

IBM Corp. (2011). IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM Corp.

Janoušek, V., Wang, L., Luzynski, K., Dufková, P., Vyskočilová, M. M., Nachman, M. W., … Tucker, P. K. (2012). Genome-wide architecture of reproductive isolation in a naturally occurring hybrid zone between *Mus musculus musculus* and *M. m. domesticus*. *Molecular Ecology*, *21*(12), 3032–3047.

Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Larson, E. L., White, T. A., Ross, C. L., & Harrison, R. G. (2014). Gene flow and the maintenance of species boundaries. *Molecular Ecology*, *23*(7), 1668–1678. https://doi.org/10.1111/mec.12601

Mandeville, E. G., Parchman, T. L., Thompson, K. G., Compton, R. I., Gelwicks, K. R., Song, S. J., & Buerkle, C. A. (2017). Inconsistent reproductive isolation revealed by interactions between *Catostomus* fish species. *Evolution Letters*, *1*(5), 255–268.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePristo, M. A. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*. 2010;*20*(9),1297–1303.

Michalak, P., & Rafiński, J. (1999). Sexual isolation between two newt species, *Triturus vulgaris* and *T. montandoni* (Amphibia, Urodela, Salamandridae). *Biological Journal of the Linnean Society*, *67*(3), 343–352.

Niedzicka, M., Dudek, K., Fijarczyk, A., Zieliński, P., & Babik, W. (2017). Linkage map of *Lissotriton* newts provides insight into the genetic basis of reproductive isolation. *G3: Genes, Genomes, Genetics*, *7*(7), 2115–2124.

Niedzicka, M., Fijarczyk, A., Dudek, K., Stuglik, M., & Babik, W. (2016). Molecular Inversion Probes for targeted resequencing in non-model organisms. *Scientific Reports*, *6*, 24051. https://doi.org/10.1038/srep24051

Nolte, A. W., Gompert, Z., & Buerkle, C. A. (2009). Variable patterns of introgression in two sculpin hybrid zones suggest that genomic isolation differs among populations. *Molecular Ecology*, *18*(12), 2615–2627. https://doi.org/10.1111/j.1365-294X.2009.04208.x

Ozerov, M., Gross, R., Bruneaux, M., Vähä, J. P., Burimski, O., Pukk, L., & Vasemägi, A. (2016). Genomewide introgressive hybridization patterns in wild Atlantic salmon influenced by inadvertent gene flow from hatchery releases. *Molecular Ecology*, *25*(6), 1275–1293. https://doi.org/10.1111/mec.13570

Pabijan, M., Zielinski, P., Dudek, K., Stuglik, M., & Babik, W. (2017). Isolation and gene flow in a speciation continuum in newts. *Molecular Phylogenetics and Evolution*, *116*, 1–12. https://doi.org/10.1016/j.ympev.2017.08.003

Palomar, G., Ahmad, F., Vasemägi, A., Matsuba, C., Nicieza, A. G., & Cano, J. M. (2017). Comparative high-density linkage mapping reveals conserved genome structure but variation in levels of heterochiasmy and location of recombination cold spots in the common frog. *G3: Genes, Genomes, Genetics*, *7*(2), 637–645. https://doi.org/10.1534/g3.116.036459

Parchman, T. L., Gompert, Z., Braun, M. J., Brumfield, R. T., McDonald, D. B., Uy, J. A. C., … Buerkle, C. A. (2013). The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. *Molecular Ecology*, *22*(12), 3304–3317. https://doi.org/10.1111/mec.12201

Payseur, B. A. (2010). Using differential introgression in hybrid zones to identify genomic regions involved in speciation. *Molecular Ecology Resources*, *10*(5), 806–820. https://doi.org/10.1111/j.1755-0998.2010.02883.x

Payseur, B. A., & Rieseberg, L. H. (2016). A genomic perspective on hybridization and speciation. *Molecular Ecology*, *25*(11), 2337–2360.

Piálek, J., & Barton, N. H. (1997). The spread of an advantageous allele across a barrier: The effects of random drift and selection against heterozygotes. *Genetics*, *145*(2), 493–504.

Poelstra, J. W., Vijay, N., Bossu, C. M., Lantz, H., Ryll, B., Müller, I., … Grabherr, M. G. (2014). The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science*, *344*(6190), 1410–1414.

Polechová, J., & Barton, N. (2011). Genetic drift widens the expected cline but narrows the expected cline width. *Genetics*, *189*(1), 227–235. https://doi.org/10.1534/genetics.111.129817

QGIS Development Team (2016). *QGIS geographic information system*. Open Source Geospatial Foundation Project.

R Core Team (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/

Racimo, F., Sankararaman, S., Nielsen, R., & Huerta-Sánchez, E. (2015). Evidence for archaic adaptive introgression in humans. *Nature Reviews Genetics*, *16*(6), 359. https://doi.org/10.1038/nrg3936

Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., … Westram, A. M. (2017). Interpreting the genomic landscape of speciation: A road map for finding barriers to gene flow. *Journal of Evolutionary Biology*, *30*(8), 1450–1477. https://doi.org/10.1111/jeb.13047

Roux, C., Tsagkogeorga, G., Bierne, N., & Galtier, N. (2013). Crossing the species barrier: Genomic hotspots of introgression between two highly divergent *Ciona intestinalis* species. *Molecular Biology and Evolution*, *30*(7), 1574–1587. https://doi.org/10.1093/molbev/mst066

Sedghifar, A., Brandvain, Y., & Ralph, P. (2016). Beyond clines: Lineages and haplotype blocks in hybrid zones. *Molecular Ecology*, *25*(11), 2559–2576.

Sedghifar, A., Brandvain, Y., Ralph, P., & Coop, G. (2015). The spatial mixing of genomes in secondary contact zones. *Genetics*, *201*(1), 243–261. https://doi.org/10.1534/genetics.115.179838

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., … Widmer, A. (2014). Genomics and the origin of species. *Nature Reviews Genetics*, *15*(3), 176. https://doi.org/10.1038/nrg3644

Smith, M. A., & Green, D. M. (2005). Dispersal and the metapopulation paradigm in amphibian ecology and conservation: Are all amphibian populations metapopulations? *Ecography*, *28*(1), 110–128. https://doi.org/10.1111/j.0906-7590.2005.04042.x

Stuglik, M. T., & Babik, W. (2016). Genomic heterogeneity of historical gene flow between two species of newts inferred from transcriptome data. *Ecology and Evolution*, *6*(13), 4513–4525. https://doi.org/10.1002/ece3.2152

Szymura, J. M., & Barton, N. H. (1986). Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina bombina* and *B. variegata*, near Cracow in southern Poland. *Evolution*, *40*(6), 1141–1159.

Taylor, S. A., Anderson, D. J., Zavalaga, C. B., & Friesen, V. L. (2012). Evidence for strong assortative mating, limited gene flow, and strong differentiation across the blue-footed/Peruvian booby hybrid zone in northern Peru. *Journal of Avian Biology*, *43*(4), 311–324. https://doi.org/10.1111/j.1600-048X.2012.05660.x

Teeter, K. C., Thibodeau, L. M., Gompert, Z., Buerkle, C. A., Nachman, M. W., & Tucker, P. K. (2010). The variable genomic architecture of isolation between hybridizing species of house mice. *Evolution*, *64*(2), 472–485. https://doi.org/10.1111/j.1558-5646.2009.00846.x

Turner, T. L., Hahn, M. W., & Nuzhdin, S. V. (2005). Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology*, *3*(9), 1572–1578. https://doi.org/10.1371/journal.pbio.0030285

Vijay, N., Bossu, C. M., Poelstra, J. W., Weissensteiner, M. H., Suh, A., Kryukov, A. P., & Wolf, J. B. (2016). Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. *Nature Communications*, *7*, 13195. https://doi.org/10.1038/ncomms13195

Wielstra, B. (2019). Historical hybrid zone movement: More pervasive than appreciated. *Journal of Biogeography*, *46*, 1300–1305. https://doi.org/10.1111/jbi.13600

Wielstra, B., Burke, T., Butlin, R. K., Avcı, A., Üzüm, N., Bozkurt, E., … Arntzen, J. W. (2017). A genomic footprint of hybrid zone movement in crested newts. *Evolution Letters*, *1*(2), 93–101. https://doi.org/10.1002/evl3.9

Wolf, J. B., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, *18*(2), 87. https://doi.org/10.1038/nrg.2016.133

Yanchukov, A., Hofman, S., Szymura, J. M., Mezhzherin, S. V., Morozov-Leonov, S. Y., Barton, N. H., & Nürnberger, B. (2006). Hybridization of *Bombina bombina* and *B. variegata* (Anura, Discoglossidae) at a sharp ecotone in western Ukraine: Comparisons across transects and over time. *Evolution*, *60*(3), 583–600.

Zavadil, V., Pialek, J., & Dandova, R. (2003). *Triturus montandoni* (Boulenger, 1880) – Karpatenmolch. In W. Bohme (Ed.), *Handbuch der Reptilen und Amphibien Europas* (pp. 657–706). Wiebelsheim, Germany: Aula.

Zbożeń, J., & Rafiński, J. (1993). A comparative study of mitotic and meiotic chromosomes of the Montandon's newt, *Triturus montandoni* (Urodela: Salamandridae) from Poland and Rumania. *Genetica*, *88*(1), 69–77. https://doi.org/10.1007/BF02424453

Zieliński, P., Dudek, K., Stuglik, M. T., Liana, M., & Babik, W. (2014). Single nucleotide polymorphisms reveal genetic structuring of the Carpathian newt and provide evidence of interspecific gene flow in the nuclear genome. *PLoS ONE*, *9*, e97431. https://doi.org/10.1371/journal.pone.0097431

Zieliński, P., Nadachowska-Brzyska, K., Dudek, K., & Babik, W. (2016). Divergence history of the Carpathian and smooth newts modelled in space and time. *Molecular Ecology*, *25*(8), 3912–3928. https://doi.org/10.1111/mec.13724

Zieliński, P., Nadachowska-Brzyska, K., Wielstra, B., Szkotak, R., Covaciu-Marcov, S. D., Cogălniceanu, D., & Babik, W. (2013). No evidence for nuclear introgression despite complete mtDNA replacement in the Carpathian newt (*Lissotriton montandoni*). *Molecular Ecology*, *22*(7), 1884–1903.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Zieliński P, Dudek K, Arntzen JW, et al. Differential introgression across newt hybrid zones: Evidence from replicated transects. *Mol Ecol*. 2019;28:4811–4824. https://doi.org/10.1111/mec.15251