Widespread Adaptive Introgression of Major Histocompatibility Complex Genes across Vertebrate Hybrid Zones

T. Gaczorek [1], K. Dudek [1], U. Fritz [2], L. Bahri-Sfar [3], S. J. E. Baird [4], F. Bonhomme [5], C. Dufresnes [6], V. Gvoždík [4,7], D. Irwin [8], P. Kotlík [9], S. Marková [9], P. McGinnity [10], M. Migalska [1], J. Moravec [7], L. Natola [8], M. Pabijan [11], K.P. Phillips [9,12], Y. Schöneberg [13,14], A. Souissi [3,15], J. Radwan [16], W. Babik [1,*]

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

[2]Museum of Zoology (Museum für Tierkunde), Senckenberg Dresden, Dresden, Germany

[3]Biodiversité, Parasitologie et Ecologie des Ecosystèmes Aquatiques, Faculté des Sciences de Tunis, Univ de Tunis El Manar, Tunis, Tunisia

[4]Institute of Vertebrate Biology of the Czech Academy of Sciences, Brno, Czech Republic

[5]Institut des Sciences de l'Evolution, Université de Montpellier, Montpellier, France

[6]Institut de Systématique, Evolution, Biodiversité (ISYEB), Muséum National d'Histoire Naturelle, CNRS, Sorbonne Université, EPHE, Université des Antilles, Paris, France

[7]Department of Zoology, National Museum of the Czech Republic, Prague, Czech Republic

[8]Biodiversity Research Centre and Department of Zoology, University of British Columbia, Vancouver, British Columbia, Canada

[9]Laboratory of Molecular Ecology, Institute of Animal Physiology and Genetics of the Czech Academy of Sciences, Liběchov, Czech Republic

[10]School of Biological, Earth and Environmental Sciences, University College Cork, Cork, Ireland

[11]Institute of Zoology and Biomedical Research, Faculty of Biology, Jagiellonian University, Kraków, Poland

[12]Canadian Rivers Institute, University of New Brunswick, Fredericton, New Brunswick, Canada

[13]Senckenberg Biodiversity and Climate Research Centre (BiK-F), Frankfurt am Main, Germany

[14]Institute for Ecology, Evolution and Diversity, Goethe University, Frankfurt am Main, Germany

[15]MARBEC, Univ Montpellier, 34000 Montpellier, France

[16]Institute of Environmental Biology, Faculty of Biology, Adam Mickiewicz University, Poznań, Poland

*Corresponding author: Email: wieslaw.babik@uj.edu.pl.

Associate editor: Sandro Bonatto

Abstract

Interspecific introgression is a potentially important source of novel variation of adaptive significance. Although multiple cases of adaptive introgression are well documented, broader generalizations about its targets and mechanisms are lacking. Multiallelic balancing selection, particularly when acting through rare allele advantage, is an evolutionary mechanism expected to favor adaptive introgression. This is because introgressed alleles are likely to confer an immediate selective advantage, facilitating their establishment in the recipient species even in the face of strong genomic barriers to introgression. Vertebrate major histocompatibility complex genes are well-established targets of long-term multiallelic balancing selection, so widespread adaptive major histocompatibility complex introgression is expected. Here, we evaluate this hypothesis using data from 29 hybrid zones formed by fish, amphibians, squamates, turtles, birds, and mammals at advanced stages of speciation. The key prediction of more extensive major histocompatibility complex introgression compared to genome-wide introgression was tested with three complementary statistical approaches. We found evidence for widespread adaptive introgression of major histocompatibility complex genes, providing a link between the process of adaptive introgression and an underlying mechanism. Our work identifies major histocompatibility complex introgression as a general mechanism by which species can acquire novel, and possibly regain previously lost, variation that may enhance defense against pathogens and increase adaptive potential.

Key words: adaptation, host–pathogen coevolution, hybridization, introgression, MHC.

Open Access

Introduction

The process of adaptation by natural selection requires genetic variation within populations. Consequently, the rate and extent of adaptation may be limited by the availability of suitable variation (Bell 2013; Rousselle et al. 2020). One potentially important source of such variation is introgression, which is the acquisition of new genetic variants through hybridization with closely related species (Edelman and Mallet 2021; Moran et al. 2021). Any introgressed variants will have been pretested by natural selection. Moreover, their initial frequencies in the recipient species are often higher than those of variants produced by de novo mutation, which may facilitate adaptation by reducing the risk of stochastic loss during the establishment phase (Hedrick 2013). While reproductive barriers and reduced fitness of hybrids limit introgression (Barton and Bengtsson 1986; Nosil et al. 2017; Butlin and Smadja 2018; Barton 2020), both theory (Barton 1979; Piálek and Barton 1997) and accumulating empirical evidence (Hedrick 2013; Enard and Petrov 2018; Edelman and Mallet 2021; Moran et al. 2021) indicate that genome-wide barriers to introgression can be readily overcome by genetic variants that confer a selective advantage in the recipient species. Nonetheless, empirically supported, process-based generalizations about the targets and mechanisms of adaptive introgression are lacking.

Certain functional categories of genes, such as those involved in immunity (Enard and Petrov 2018), pigmentation (Giska et al. 2019), or resistance to environmental (Whitney et al. 2006) and anthropogenic (Song et al. 2011) stressors, appear to be more prone to adaptive introgression, possibly due to similar selective pressures experienced by related species. One evolutionary process that should particularly favor introgression is balancing selection, especially when operating through the mechanism of rare allele advantage, which maintains multiple alleles in a population (multiallelic balancing selection) via negative frequency dependence (Castric et al. 2008; Fijarczyk and Babik 2015). Because introgressed alleles are initially rare, they would often confer a selective advantage, facilitating their establishment in the recipient species even in the face of strong genomic barriers to introgression, while negative frequency dependence will prevent their fixation (Castric et al. 2008; Fijarczyk et al. 2018).

A prime example of genes evolving under long-term balancing selection is the major histocompatibility complex (MHC) in jawed vertebrates (Spurgin and Richardson 2010). The MHC encodes two classes of molecules— Class I (MHC-I) and Class II (MHC-II), both of which present antigens to T cells, but differ in, among other features, the source of antigenic peptides and the subpopulations of T cells with which they interact (Murphy et al. 2022). Because of their key role in initiating the adaptive immune response, both MHC classes are involved in perpetual Red Queen coevolutionary dynamics with pathogens (Baird et al. 2012; Radwan et al. 2020). An advantage of MHC alleles from allopatric populations (novel to local parasites)

has been both predicted (Baird et al. 2012) and experimentally demonstrated (Phillips et al. 2018). There is also evidence for MHC introgression, possibly adaptive, in several systems, including humans (Abi-Rached et al. 2011), goats (Grossen et al. 2014), and newts (Dudek et al. 2019). If our understanding of the mechanisms promoting introgression and maintaining MHC diversity is correct, taxonomically widespread adaptive MHC introgression is expected, and this prediction can be tested in a comparative framework. However, testing for adaptive introgression of genes under balancing selection in general, and the MHC in particular, presents significant challenges. First, balancing selection itself may generate signatures similar to the signal of introgression in a way that confounds testing by, e.g. long-term maintenance of shared polymorphism (Fijarczyk and Babik 2015). Second, the genomic organization of the MHC region is often extremely complex (McConnell et al. 2016; Westerdahl et al. 2022), with multiple similar, highly variable genes exhibiting extensive intraspecific copy number variation. The latter makes the assembly of the MHC region difficult and its representation by a single reference sequence often impossible (Dilthey 2021), which in turn precludes the use of computational methods designed to detect adaptive introgression from genome resequencing data (Setter et al. 2020; Gower et al. 2021; Svedberg et al. 2021). Third, the dynamic evolution of the MHC region involves frequent duplication, pseudogenization and gene losses, as well as interlocus recombination. Combined with considerable allelic divergence within MHC loci, these processes produce patterns of sequence divergence that often preclude reliable assignment of MHC variants to loci. This is particularly evident for the data typically used to characterize MHC variation in natural populations—sequences of PCR-amplified fragments of variable exons. Given the nature of the data, it is necessary to employ analytical methods that do not rely on the assignment of MHC variants to loci.

Direct comparison of the extent of MHC introgression with that of genome-wide introgression in natural hybrid zones, applied to the multilocus MHC data, may help to overcome the challenges outlined above (Payseur 2010; Gompert et al. 2017; Dudek et al. 2019). For example, shared ancestral polymorphism would be expected to be uniformly distributed across species ranges, whereas introgressing variants should initially be concentrated near the hybrid zone, resulting in locally increased interspecific similarity. Recent or ongoing adaptive introgression would generate a signal of similarity extending further from the hybrid zone than that of genome-wide, neutral introgression. Here, we use data from 29 hybrid zones formed by fish, amphibians, squamates, turtles, birds, and mammals and employ three complementary statistical approaches to compare MHC introgression with genome-wide introgression. We synthesize the results in a comparative framework and provide evidence for widespread adaptive introgression of MHC genes across vertebrate hybrid zones, which leads to an important generalization about the process.
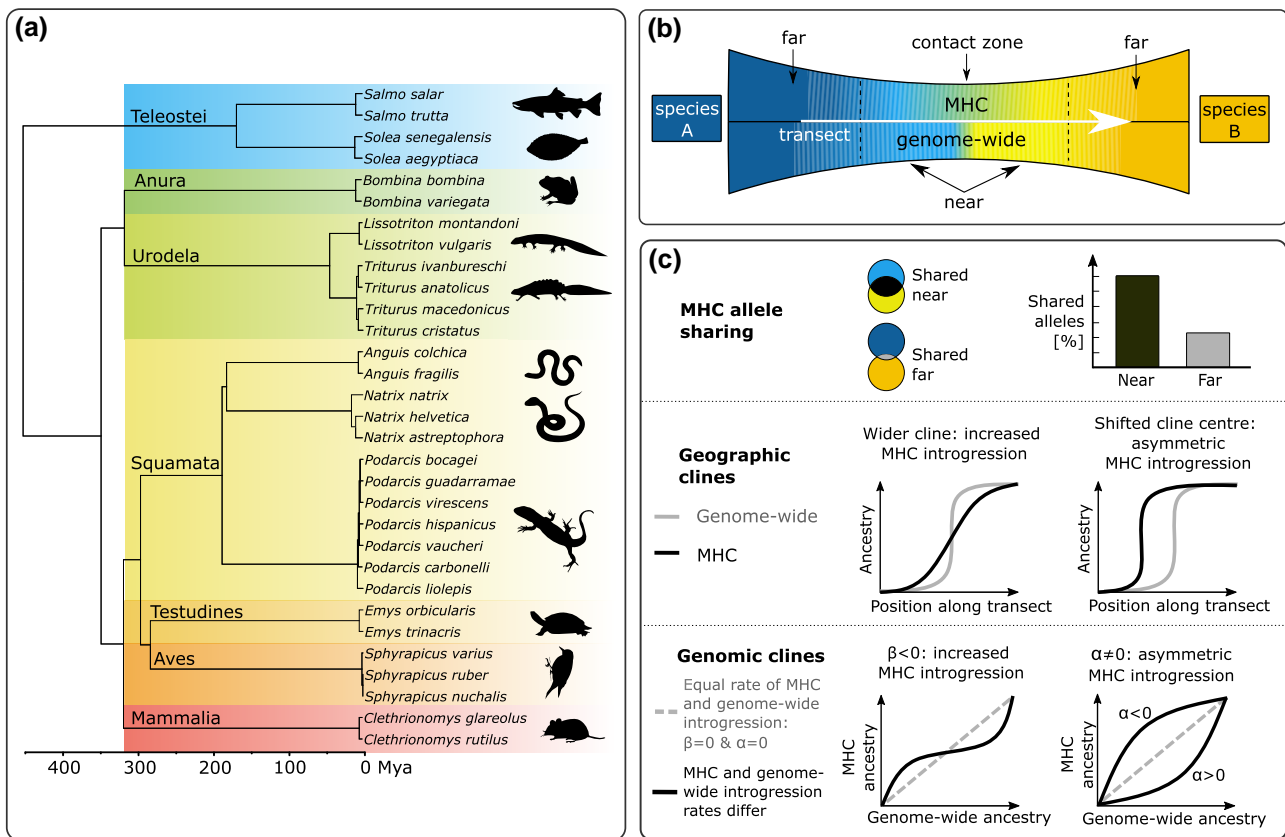
Fig. 1. Study design. a) Time-calibrated phylogeny of the species studied; b) the geographic extent of genome wide and MHC introgression expected under adaptive MHC introgression, along with the division of species ranges into areas classified as "near" and "far" from the contact zone; c) the approaches used to compare genome-wide and MHC introgression.

Results

We examined 29 hybrid zones formed by 31 species representing 11 genera of seven major vertebrate lineages (Fig. 1a); a total of 4,933 individuals were analyzed (supplementary table S1, Supplementary Material online). The selection of hybrid zones was based on the following criteria: (i) the taxa were known to have limited genome-wide introgression (see Genome-Wide Admixture)—which was crucial for our tests, which detect signals of MHC introgression extending beyond the geographic range of genome-wide introgression; (ii) reliable, individual-based genome-wide estimates of interspecific admixture were available; and (iii) DNA or tissue samples for MHC genotyping were accessible.

MHC Diversity

MHC data for eight genera were generated for the purposes of this study, while for the remaining three genera, data were derived entirely (Gaczorek et al. 2023a, 2023b) or partially (Dudek et al. 2019) from previous work. MHC diversity was assessed by Illumina sequencing of the PCR-amplified second exon of MHC-I and MHC-IIB genes. This exon is part of the peptide-binding region that determines the antigen specificity of MHC molecules and is highly variable. For simplicity, we refer to each MHC

sequence variant as an allele, although we are aware that these variants are derived from multiple loci and hence an individual typically carries more than two such defined MHC alleles. The repeatability of genotyping, estimated by running a fraction of the samples in duplicate, averaged across genera was 0.95 (range 0.88 to 1.00) for MHC-I and 0.91 (range 0.76 to 1.00) for MHC-II (supplementary table S2, Supplementary Material online).

The range of MHC variation, measured as both the total and sample size standardized number of alleles per species, i.e. allelic richness, spanned two orders of magnitude (supplementary table S3, Supplementary Material online). In all genera except *Natrix*, MHC-I was more diverse than MHC-II. The mean number of alleles per individual was similar for species within a genus but varied more than tenfold between genera, ranging from 3.7 (*Natrix*) to 33.7 (*Clethrionomys*) in MHC-I and from 1.8 (*Triturus*) to 34.6 (*Natrix*) in MHC-II (Fig. 2). The number of alleles per individual may be indicative of the number of MHC genes—the minimum number of genes, assuming all were heterozygous, would be half the number of alleles. As almost all taxa had more than one gene per each MHC class and unambiguous assignment of alleles to specific genes was not possible, MHC genotypes were coded as binary matrices of allele presence/absence. These matrices were used in subsequent analyses of introgression.
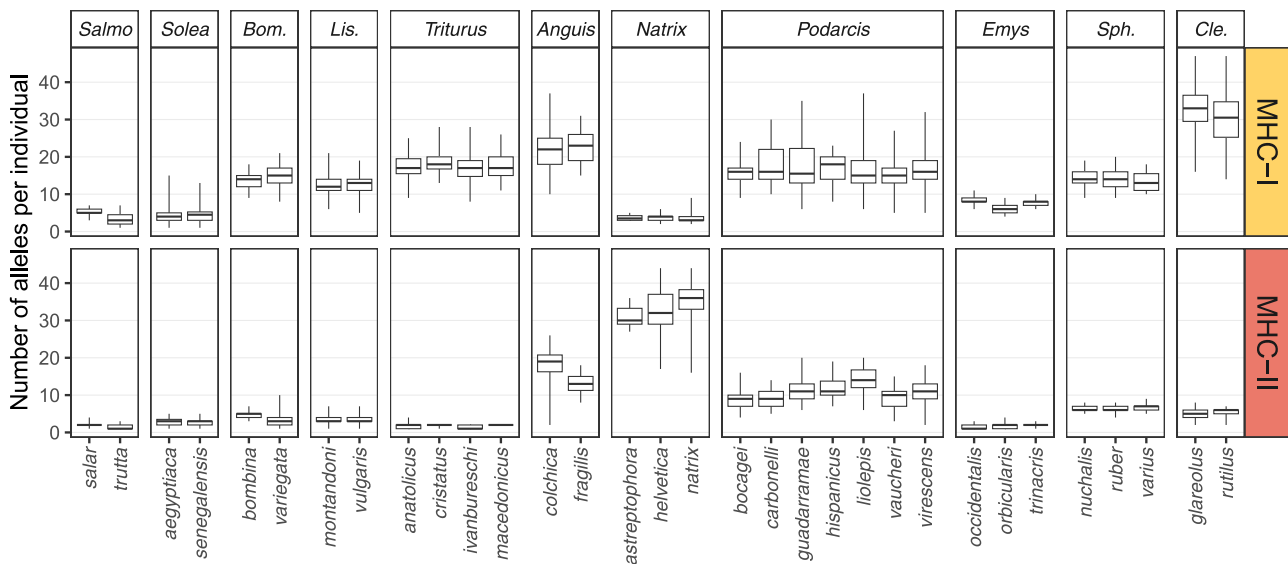
Fig. 2. The number of MHC alleles per individual in the studied species. These results show clear differences in the complexity between MHC classes and between taxa. Note that we use term "allele" for each unique sequence variant, although these variants may originate from different MHC genes. Boxplots show the median (horizontal line), the first and third quartile (box), and the range (whiskers). Different MHC markers were used for *Clethrionomys* (see Materials and Methods), so estimates of MHC variation for this genus are not directly comparable with those for other taxa. *Bom.*, *Bombina*; *Lis.*, *Lissotriton*; *Sph.*, *Sphyrapicus*; *Cle.*, *Clethrionomys*.

For each hybrid zone, the number of rows in the binary genotype matrix was equal to the number of individuals examined, and the number of columns was equal to the total number of detected alleles, with a "1" at the intersection of a column and a row indicating the presence of the allele in the individual and "0" indicating its absence.

Comparison of MHC and Genome-Wide Introgression

We tested two major patterns predicted by the adaptive MHC introgression hypothesis for hybrid zones—the theoretical expectations are summarized in Fig. 1b and c. First, MHC introgression should extend geographically further than genome-wide introgression. This prediction is based on two assumptions: (i) selective advantage allows introgressed MHC alleles to increase in frequency and rapidly spread outside the contact zone, and (ii) the perpetual nature of the host–pathogen arms race sustains an influx of foreign MHC alleles over extended periods of time, resulting in a broad MHC introgression gradient (Fig. 1b). We tested this prediction using two approaches (Fig. 1c): (i) randomizations to test for elevated interspecific MHC allele sharing in the vicinity of contact zones (near) but beyond the actual hybrid zone (defined as the area of genome-wide introgression detectable using single nucleotide polymorphism (SNP) or microsatellite markers), compared to the areas further from the contact (far), and (ii) comparison of the width and position of MHC and genome-wide geographic clines, which measure the transition in population-level ancestry of each species along a geographic transect. The second prediction is a relative excess of heterospecific ancestry in MHC compared to the

genome-wide average in admixed individuals, which would indicate a selective advantage of heterospecific MHC even in early generation hybrids. This prediction was tested using the genomic clines approach (Fig. 1c).

MHC Allele Sharing

The significance of differences in allele sharing between species near and far from the contact zone was tested with a permutation test, randomly assigning individuals within species into near/far categories. Randomization tests were performed in 25 hybrid zones representing ten genera (supplementary table S4, Supplementary Material online). In most cases, the proportion of alleles shared between species was higher near than far from the contact, the difference being significant for at least one MHC class in 17 out of 25 (68%) zones and for both MHC classes in 11 out of 25 (44%) zones (Fig. 3a). In addition, in *Salmo*, we detected significantly increased allele sharing within the contact zone for both MHC classes ($P < 0.001$), which is indicative of MHC introgression but is not informative about its adaptiveness due to the lack of the near category. The results were summarized using standardized effect size (SES; Gotelli and McCabe 2002; Botta-Dukát 2018), which allowed a meaningful comparison of the excess of MHC allele sharing near the contact zone across the investigated taxa. Phylogenetic generalized least squares (PGLS) was used to assess the significance of SES. The phylogenetic correlation matrix input to PGLS accounted for the fact that our data were from pairs of species rather than single species. The SES was significantly greater than 0 for both MHC classes (MHC-I $P = 0.0001$, MHC-II $P = 0.0004$; Fig. 3b), indicating elevated allele sharing near the contact, with no significant difference
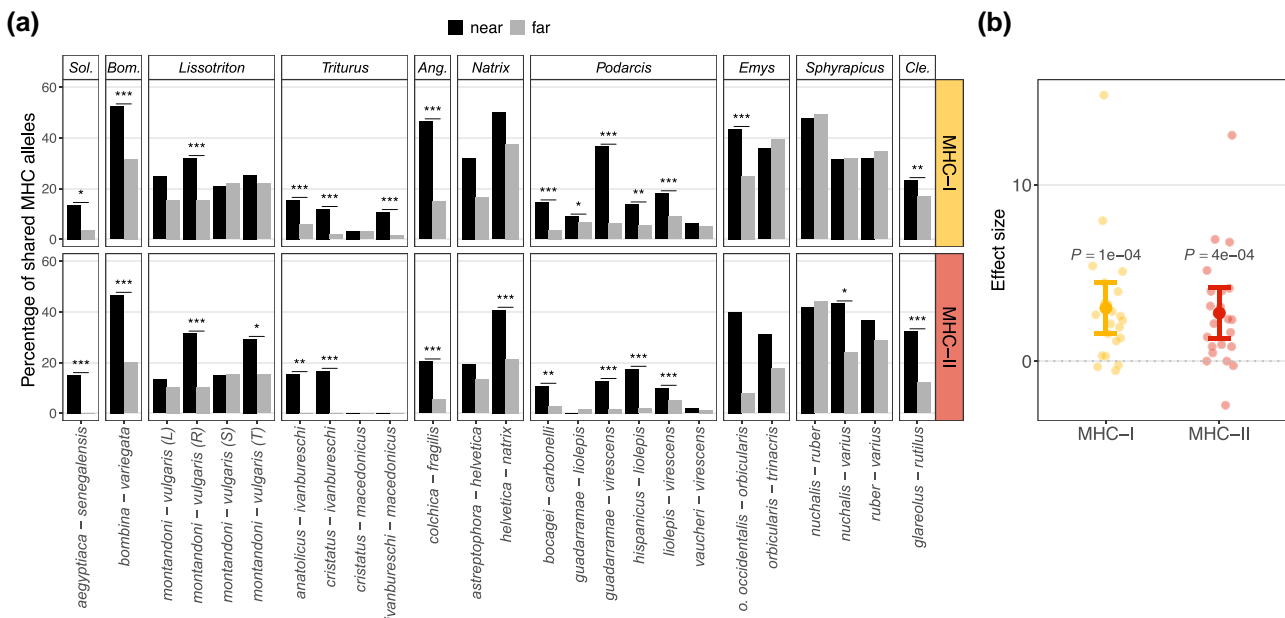
Fig. 3. Interspecific MHC allele sharing was elevated in the vicinity of contact but beyond the actual hybrid zone ("near") compared to the areas further from the contact ("far"). a) Allele sharing between species near and far from the contact zone; b) SESs (Cohen's d) of the difference in allele sharing between species near and far from the contact zone and the PGLS model estimates with 95% confidence intervals; *Sol., Solea; Bom., Bombina; Ang., Anguis; Cle., Clethrionomys.*

between MHC classes ($P = 0.78$). When time of divergence between hybridizing species was included in the PGLS model (supplementary fig. S1a, Supplementary Material online), its effect was not significant ($P = 0.42$, model without class × time interaction); including interaction did not improve model fit (interaction likelihood ratio test [LRT]; $P = 0.74$). There was no phylogenetic signal in either model (Pagel's $\lambda = 0.0$).

Geographic Clines

Transects were designated for 19 hybrid zones representing eight genera. Geographic clines could be fitted to all transects except those for *Natrix astreptophora* × *N. helvetica* and one of the *N. helvetica* × *N. natrix* transects (HN3). In these two cases, a null model of no cline (i.e. straight line) was supported for at least one MHC class and these transects were excluded from the comparative analysis of geographic clines, as the values of cline parameters would not be interpretable.

The use of binary-coded MHC genotypes to calculate the hybrid index (HI), which is then used to fit clines, has the potential to affect the accuracy of cline parameter estimates. To explore whether this was the case, we simulated five-locus haplotypes evolving under strong negative frequency-dependent selection in two species that evolved in isolation following divergence from a common ancestor. At three time points, 0.5, 1, and 4N generations after divergence, we created a "transect" through a hybrid zone by sampling different proportions of MHC haplotypes from each species along a predefined smooth ancestry cline. We then estimated the cline width using: (i) the actual, known ancestry of the haplotypes; (ii) the hybrid index

estimated from known genotypes at each MHC locus; and (iii) the hybrid index estimated considering each MHC allele as an independent binary dominant locus, reflecting the treatment of the observed MHC data. We found that binary coding did not affect the estimated cline width (supplementary fig. S2, Supplementary Material online), consistent with previous work that used a somewhat different simulation setup (Gaczorek et al. 2023b).

MHC clines tended to be wider than genome-wide clines. In all transects, a cline estimated for at least one MHC class was wider than the genome-wide cline, and in 9 out of 17 (53%) transects, clines for both MHC classes were wider (Figs. 4a and 5). In many cases, the 95% confidence intervals of SES were wide and overlapped 0, indicating no significant difference between MHC and genome-wide cline widths. This lack of significance was a direct consequence of the wide 2 log-likelihood intervals for the parameters of the MHC clines. The PGLS analysis of SES supported overall wider MHC clines (MHC-I $P = 0.007$, MHC-II $P = 0.004$; Fig. 4a) and no difference between classes ($P = 0.86$). In the model including time of divergence between hybridizing taxa (supplementary fig. S1b, Supplementary Material online), the effect of time was not significant (PGLS, $P = 0.08$, model without interaction, interaction LRT $P = 0.48$). There was no evidence of phylogenetic signal ($\lambda = 0.0$).

We found a significant MHC cline shift toward species with lower MHC diversity (PGLS, MHC-I $P = 0.037$, MHC-II $P = 0.011$; Fig. 4a) with no significant difference between classes ($P = 0.70$). In models including divergence time (supplementary fig. S1b, Supplementary Material online), its effect was not significant (PGLS, $P = 0.24$, model
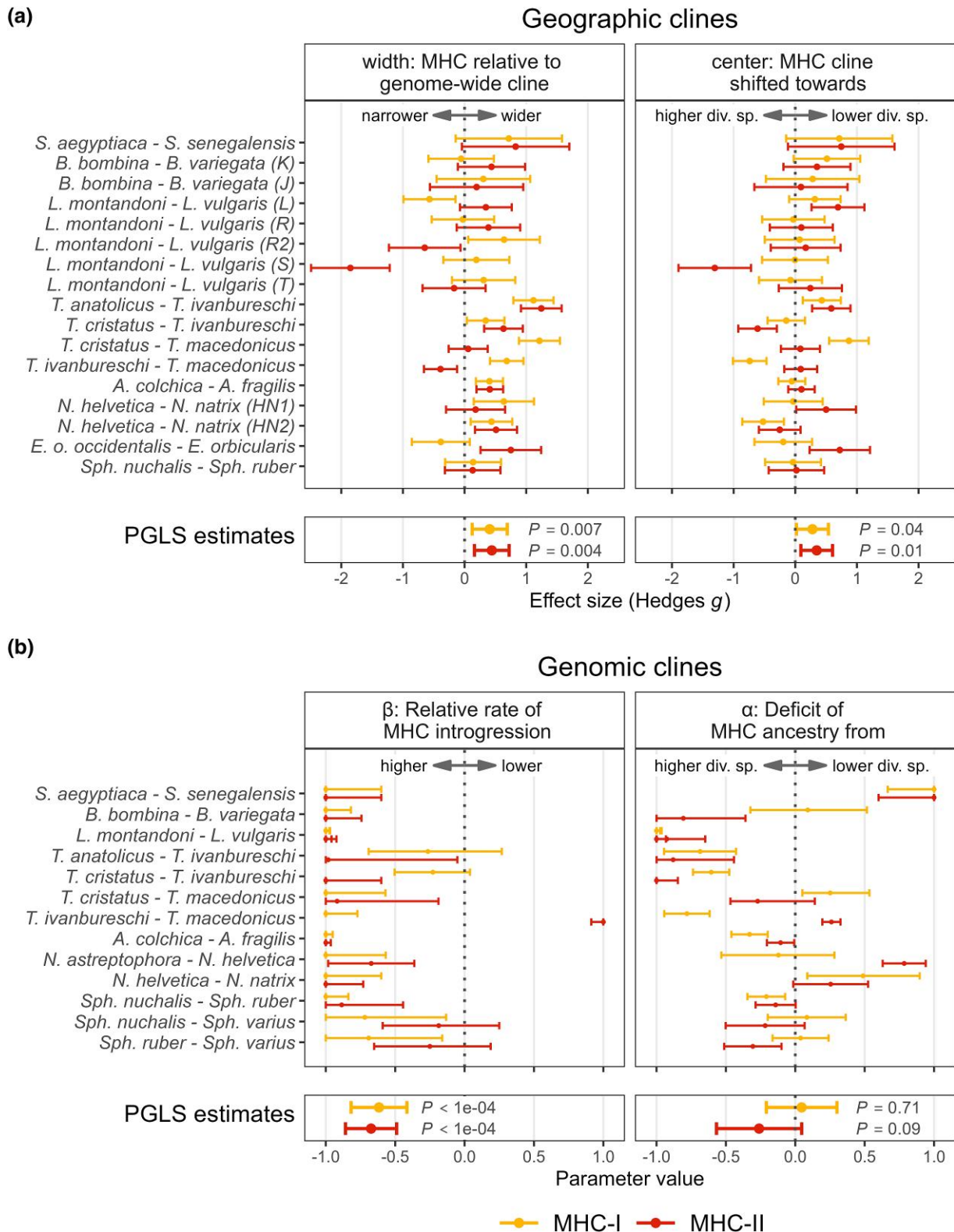
Fig. 4. Comparison of MHC and genome-wide a) geographic and b) genomic clines indicates more extensive introgression of both MHC classes compared to the genome-wide introgression. For each hybrid zone and MHC class, points show effect size estimates with 95% confidence intervals (a) or the ML parameter estimates with 2 log-likelihood ranges (b). *S., Solea; B., Bombina; L., Lissotriton; T., Triturus; A., Anguis; N., Natrix; Sph., Sphyrapicus*; PGLS estimates, phylogenetic generalized least squares estimates together with 95% confidence intervals; higher/lower div. sp., species in a pair exhibiting higher/lower MHC diversity. Note that uncertainty in parameter estimates for individual taxa is incorporated as weights in PGLS models, so confidence intervals of PGLS estimates should not be interpreted in the context of uncertainty in individual estimates.
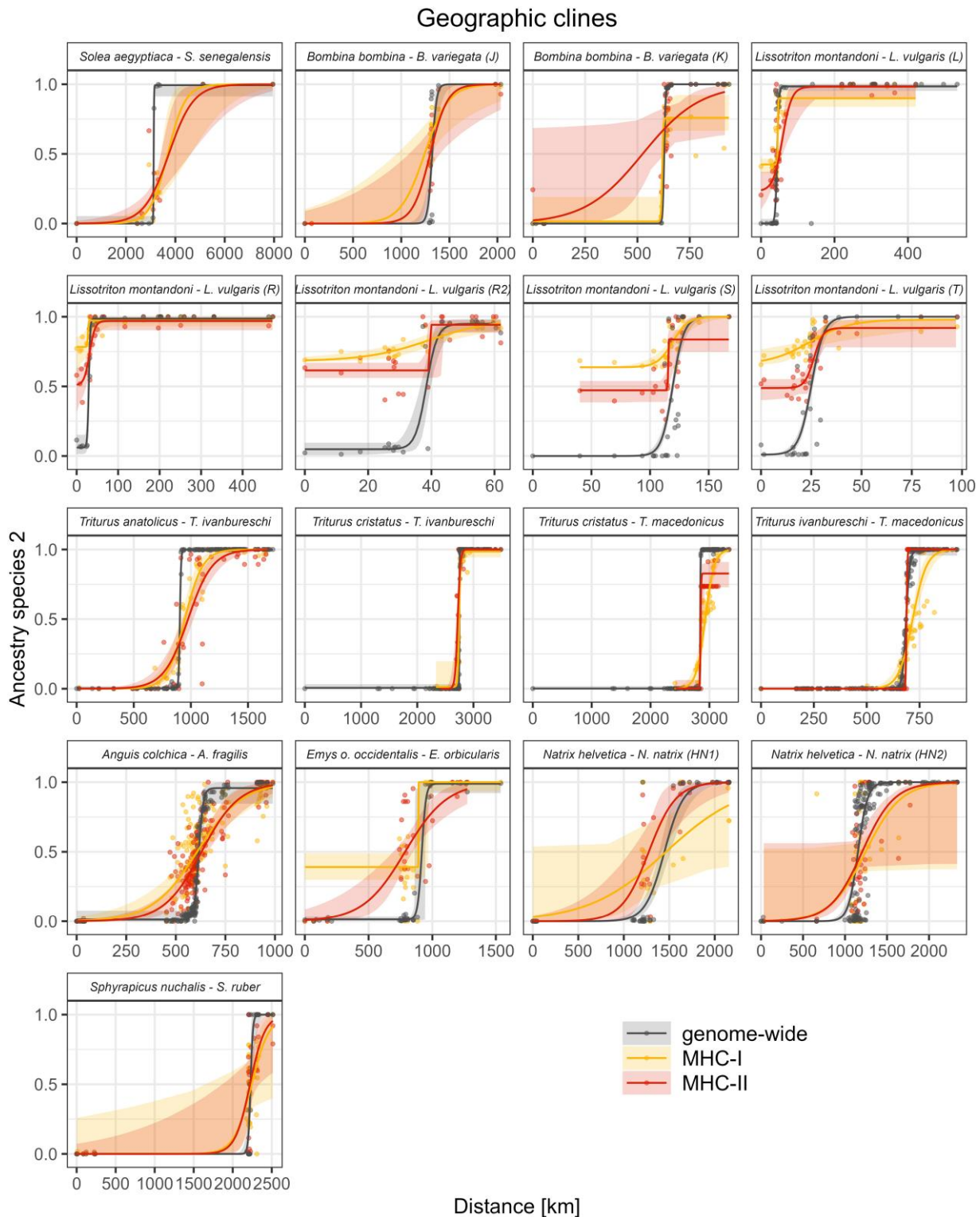
Fig. 5. Comparison of the MHC and genome-wide geographic clines. Light-colored bands show 2 log-likelihood intervals, while points correspond to populations.

without interaction, interaction LRT $P = 0.56$). Again, there was no evidence of a phylogenetic signal ($\lambda = 0.0$).

Genomic Clines

We fitted genomic clines to data from 14 hybrid zones representing seven genera. In the other systems, the distribution of genome-wide ancestry was strongly bimodal, indicating very limited genome-wide introgression and precluding the use of genomic clines, which rely on comparing locus-specific and genome-wide ancestry in admixed individuals (Gompert and Buerkle 2011; Fitzpatrick 2013). The rate of MHC introgression exceeded that of genome-wide introgression

7
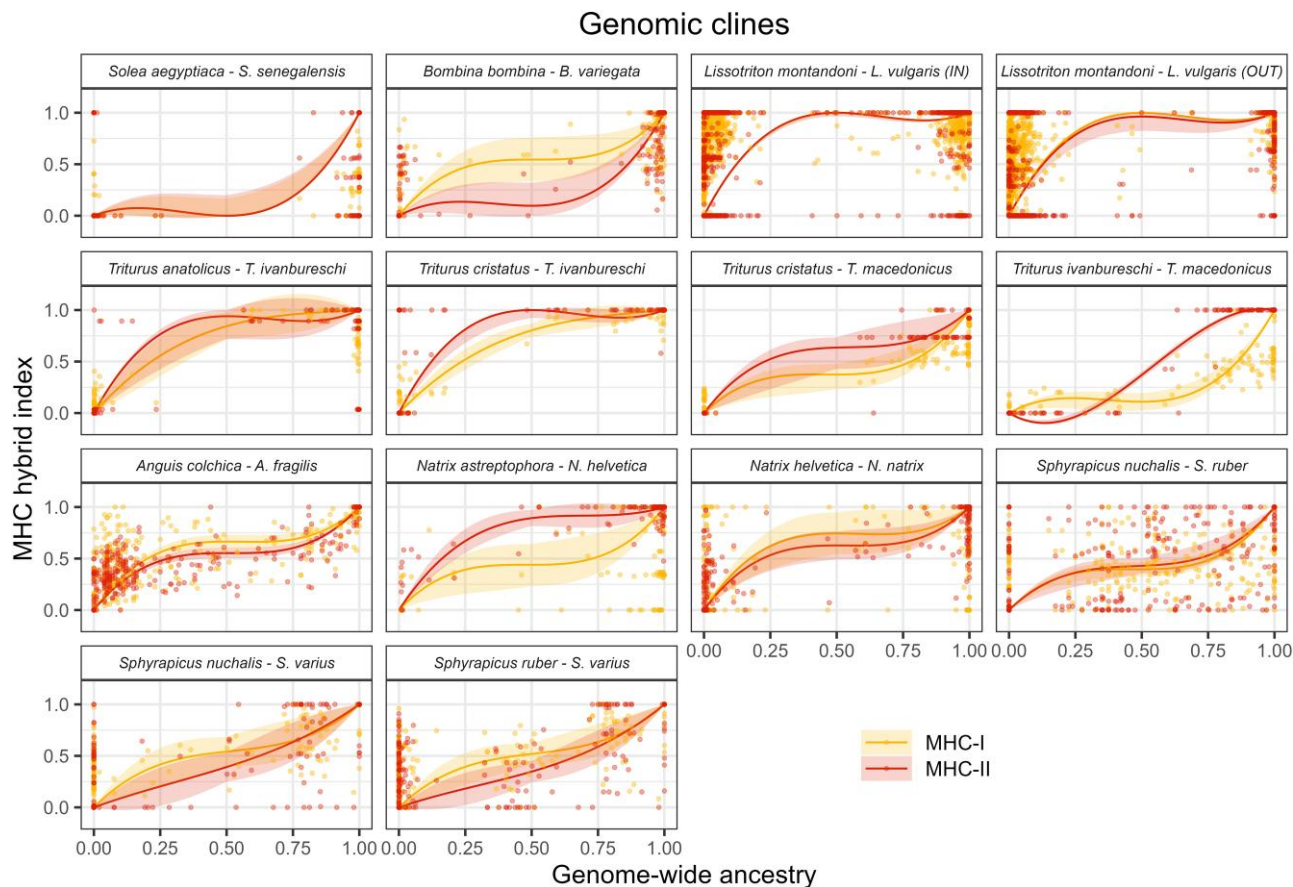
Genomic clines

Fig. 6. Genomic clines. Light-colored bands show 2 log-likelihood intervals, while points correspond to individuals.

(indicated by the negative value of the β parameter of genomic clines, reflecting the lower rate of transition in MHC ancestry compared to genome-wide ancestry) in all zones except MHC II in *Triturus ivanbureschi* × *Triturus macedonicus* (Figs. 4b and 6). The PGLS analysis supported a higher rate of MHC introgression compared to genome-wide introgression ($P < 0.0001$ for both MHC classes; supplementary fig. S1c, Supplementary Material online) and revealed no difference between MHC classes ($P = 0.67$). In the model including divergence time (supplementary fig. S1c, Supplementary Material online), there was a tendency for an increased rate of introgression between more distantly related species (PGLS, $P = 0.04$, model without interaction, interaction LRT $P = 0.09$), and no phylogenetic signal ($\lambda = 0.0$) was detected.

We did not observe an excess of MHC ancestry from the species with higher MHC variability (i.e. introgression toward species with lower MHC diversity, $\alpha > 0$; Fig. 4b). The PGLS analysis did not support asymmetry, as α was not significantly different from 0 for either MHC class (MHC-I $P = 0.71$, MHC-II $P = 0.09$; Fig. 4b), and there was no significant difference between classes ($P = 0.12$). In the model including divergence time (supplementary fig. S1c, Supplementary Material online), there was a tendency toward higher values of α for more distantly related species (PGLS, $P = 0.02$, model without interaction, interaction

LRT $P = 0.71$). There was no phylogenetic signal in models without time of divergence ($\lambda = 0.0$) while a substantial phylogenetic signal was detected in models including time of divergence ($\lambda = 0.54$).

Discussion

Our results indicate that MHC introgression across vertebrate hybrid zones is more extensive than genome-wide introgression, strongly suggesting its adaptive nature. This finding is consistent with the expectation that rare allele advantage, a mechanism of balancing selection often acting on MHC genes as a consequence of coevolution with pathogens, should initially favor novel introgressed MHC variants (Radwan et al. 2020). Alleles that have already been tested for functionality in another species, but to which local pathogens are not yet adapted, may be particularly advantageous and thus introgress easily. A similar magnitude of introgression observed for both MHC classes suggests that such an advantage applies to both groups of these antigen-presenting molecules, despite evident functional differences between them (Pishesha et al. 2022).

The empirical results presented here support theoretical predictions (Schierup et al. 2000) of widespread adaptive introgression of genes evolving under balancing selection,

offering a link between the process and the underlying mechanism. We expect adaptive introgression to be widespread also in other genetic systems evolving under long-term multiallelic balancing selection, such as self-incompatibility genes in plants, where it has indeed been documented in some taxa (Castric et al. 2008). Adaptive MHC introgression may be further facilitated by the fact that introgressed haplotypes would carry rare alleles of multiple MHC genes, thereby increasing the cumulative selective advantage. This effect could be amplified if the mechanism of divergent allele advantage, for which there is some empirical support (Pierini and Lenz 2018), was to operate. A strong fitness advantage of introgressed MHC haplotypes could help to overcome barriers to introgression due to negative epistatic interactions between the MHC region and other genomic regions in hybrids or by negative pleiotropic effects of MHC haplotypes. Such pleiotropic effects appear to be common within species, as evidenced by multiple associations between the genomic MHC region and disease, including autoimmune disorders in humans (Dendrou et al. 2018). The analysis of genomic clines within contact zones suggests that the selective advantage of introgressing variants may indeed be substantial, aligning with previous findings in mice (Baird et al. 2012). This is because an increased rate of MHC introgression is observed, even though recombination within the zone typically does not operate long enough to free MHC variants from the genetic background involved in negative epistatic interactions.

The extensive variation in the number of MHC alleles per individual indicates that the investigated taxa differ in the complexity of the genomic MHC region. How might such heterogeneity in genetic architecture affect our results? The notable consistency among the three approaches, each of which was able to use only a subset of the taxa, indicates the overall robustness of our results, suggesting that increased introgression is a general feature of the MHC, regardless of its genomic complexity. An unknown, but in some taxa substantial, portion of the variation in MHC complexity may result from differences in the number of neutrally evolving pseudogenes and nonclassical MHC genes, which are likely evolving under purifying selection—our molecular methods did not allow unambiguous identification of classical MHC sequences. The most likely effect of this limitation would simply be a reduced power to detect adaptive introgression, making our approach conservative. In most jawed vertebrates, both MHC classes are tightly linked, making an independent assessment of their introgression challenging. In this context, results from *Solea*, a representative of teleost fishes that have unlinked MHC-I and MHC-II (Sato et al. 2000), are particularly informative. In this genus, both MHC classes show similar patterns suggesting they are subject to similar mechanisms promoting introgression.

Our work identifies introgression as a general mechanism by which species may acquire novel—and possibly also regain previously lost—MHC variants that may facilitate fighting pathogen assault and boost adaptive potential. We did not find evidence that the time since divergence between species substantially influences the relative strength of MHC introgression, indicating that adaptive introgression can persist for prolonged periods following the initial divergence. The asymmetry detected in the analysis of geographic clines suggests that the selective advantage of introgressed variants may be larger in less MHC-variable species. This pattern may indicate that demographic fluctuations or pathogen-mediated selection can often reduce MHC variation below optimal levels. Of course, hybridization may also result in the transfer of pathogens, leading to an interesting dynamic feedback between disease and introgression, as suggested in the case of human–Neanderthal hybridization (Enard and Petrov 2018; Greenbaum et al. 2019). The relationships between pathogen transmission and hybridization are, however, complex and still incompletely understood (Baird and Goüy de Bellocq 2019; Theodosopoulos et al. 2019). Despite this controversy, these findings have important implications for practical conservation, supporting the view that translocations and artificial or assisted introgression may be effective conservation measures in the age of the global biodiversity crisis (Hamilton and Miller 2016). This is especially relevant as the risks related to outbreeding depression and the general negative fitness effects of hybridization may be exaggerated (Aitken and Whitlock 2013; Ralls et al. 2018).

Widespread MHC introgression may contribute to the phenomenon of trans-species polymorphism—the long-term retention of allelic lineages through multiple speciation events, often observed in MHC and other genes evolving under long-term balancing selection (Radwan et al. 2020). The maintenance of trans-species polymorphisms by balancing selection alone appears problematic given the evidence for rapid turnover of MHC allelic lineages (Ejsmond et al. 2018; Herdegen-Radwan et al. 2021), but introgression can reintroduce lost lineages and lead to sharing of identical alleles. Finally, our results are relevant for understanding and predicting genomic patterns of divergence and gene flow in advanced stages of speciation, a topic that has been extensively studied in recent years (Butlin and Smadja 2018; Barton 2020; Kulmuni et al. 2020). According to the evidence accumulated so far, divergence along the genome in the late stages of speciation is generally high, indicative of multiple strong and coupled barriers to introgression that result in genome congealing (Barton 1983; Nosil et al. 2017). This landscape is occasionally dotted with small regions of increased introgression, which appear hybrid zone specific even in cases of parallel adaptive speciation (Ravinet et al. 2018). Recent progress in genomics should enable the inclusion of the MHC region in such scans of genomic divergence. If, as our results suggest, the MHC would be among the last genomic regions to stop introgressing between diverging taxa before reproductive isolation is complete, then it should repeatedly emerge as an outlier in surveys of divergence and gene flow in advanced stages of speciation.

Two potential limitations of our approach need to be considered. First, the coding of MHC genotypes as allele presence/absence and the subsequent analyses of the binary matrices could lead to inaccurate estimation of the hybrid index and to an erroneous inference of more extensive MHC introgression. The effect of binary coding of multilocus MHC genotypes was addressed by simulations, which showed that the hybrid index we calculated accurately reflected individual and population MHC ancestry, resulting in the correct estimation of cline parameters. In addition, randomization tests of MHC allele sharing did not use the MHC hybrid index and were therefore unaffected by potential issues related to its estimation. Second, it is in principle possible that more extensive MHC introgression, even if correctly inferred, is not due to the adaptive advantage of the introgressed alleles. The observed signal of more extensive MHC introgression could be a result of a parallel increase in the frequency of the same alleles, maintained as ancestral polymorphisms in both species, in the vicinity of the contact, possibly due to shared pathogen pressure. However, this scenario would require not only shared selection pressure but also the maintenance of a shared pool of identical alleles that can be selected in parallel in both species, which is unlikely (Dudek et al. 2019). Thus, considering the weight of the overall evidence, we conclude that processes other than adaptive MHC introgression are unlikely to explain the patterns detected in our study.

Adaptive MHC introgression may actually be even more pronounced than our results suggest. First, all of our approaches relied on the geographical structuring of introgressed variants, assuming that they were overrepresented in the vicinity of the contact. This implies ongoing or recent introgression that has not yet spread across the entire species ranges, as such spread would make past introgression difficult or impossible to detect. In such cases, the signal of historical introgression can still be detected by comparing MHC allele sharing between species with adjacent and nonadjacent ranges. Indeed, comparative analyses in *Triturus* newts (Gaczorek et al. 2023b) and *Podarcis* lizards (Gaczorek et al. 2023a) showed increased MHC allele sharing between species with adjacent ranges, even when considering only core range areas, unlikely to have been affected by recent introgression. The second indication that our approach may have underestimated MHC introgression comes from the *Lissotriton montandoni × L. vulgaris* hybrid zone, where multiple transects have been examined in detail (Dudek et al. 2019). In this system, interspecific MHC similarity at local scales exceeded similarity between geographic regions within species, indicating massive MHC introgression. At the same time, there was no difference in the width of MHC and genome-wide clines in any transect (Dudek et al. 2019). Reanalysis of the *Lissotriton* system presented here, using longer transects and more distant allopatric populations, provided some support for wider geographic MHC clines. If MHC introgressed extensively across the entire length of a transect, the actual wide geographic clines could be impossible to capture, as might be

the case for two *Natrix* transects where the no-cline model fitted the observed MHC data best. However, in such cases, as previously suggested (Dudek et al. 2019), it may be possible to detect a shallow MHC cline forming de novo in the center of the hybrid zone due to restricted hybridization. This newly formed MHC cline may coincide with the genome-wide cline, giving a false impression of no increased MHC introgression.

A more nuanced and detailed picture of MHC introgression is likely to emerge in the future with the adoption of recent developments in long-read genomics (van Dijk et al. 2023) and associated analytical approaches (Eizenga et al. 2020; Dilthey 2021; Liao et al. 2023), which will provide direct access to long-range haplotype information (Ebert et al. 2021; Li et al. 2023). With the application of these methods at the population scales (De Coster et al. 2021), we should be able to directly trace the introgression of individual MHC haplotypes and their fragments, as they are progressively broken down by recombination. With such data, it will be possible to identify which haplotypes are introgressing adaptively and whether selection also prevents the introgression of some MHC haplotypes. The availability of population-level haplotype data will also allow us to study the effect of the genomic architecture of the MHC region, such as the number of genes, the size of the region, and the strength of linkage between MHC classes on introgression. While these advances will extend our understanding of introgression in general and MHC introgression in particular, they are unlikely to alter the overall conclusion of widespread adaptive MHC introgression.

In conclusion, we have found evidence supporting widespread adaptive introgression of MHC genes in numerous vertebrate hybrid zones. This finding establishes a much-needed generalization about the process of adaptive introgression, linking its many instances to a reasonably well-understood mechanism of rare allele advantage. We hypothesize that adaptive introgression is likely to be common in other genetic systems where balancing selection operates through rare allele advantage. Such genomic regions are expected to be among the last where gene flow persists during advanced stages of speciation. Our results indicate that adaptive MHC introgression is a major source of new adaptive MHC variation in natural systems, implying that assisted MHC introgression may be a valuable tool in practical conservation and management efforts.

Materials and Methods

Study Systems

We studied 29 hybrid zones (in cases where multiple geographically distant contacts between a pair of species were studied, they were considered separate zones) formed by 31 species representing 11 genera from major groups of jawed vertebrates: *Salmo* and *Solea* (teleost fishes); *Bombina* (anuran amphibians); *Lissotriton* and *Triturus* (urodele amphibians); *Emys* (turtles); *Anguis*, *Natrix*, and *Podarcis* (squamate reptiles); *Sphyrapicus* (birds); and *Myodes* (mammals)

(supplementary table S1, Supplementary Material online). MHC introgression has been previously studied in three of the genera analyzed here: *Lissotriton* (Dudek et al. 2019), *Podarcis* (Gaczorek et al. 2023a), and *Triturus* (Gaczorek et al. 2023b). We combined published data with those from the eight genera obtained in this study and additional newly generated data from *Lissotriton*.

The time-calibrated phylogeny of the taxa included in this study is shown in Fig. 1a. This tree was used to control for the effect of phylogenetic relationships in the comparative analyses described below. The backbone phylogeny is from Irisarri et al. (2017), and divergence times within some major lineages and within genera were taken from the literature: Teleostei (Rabosky et al. 2018), Urodela (Palomar et al. 2021), *Bombina* (Pabijan et al. 2013), *Natrix* (Schöneberg et al. 2023), *Podarcis* (Yang et al. 2021), and *Triturus* (Wielstra et al. 2019). For all other genera not included in Irisarri et al. (2017), we obtained divergence times between closely related species from TimeTree (Kumar et al. 2017).

Genome-Wide Admixture

For the majority of taxa, we used published SNP or microsatellite data (for references, see supplementary table S1, Supplementary Material online) that allowed an adequate estimation of the genome-wide interspecific admixture. For *Anguis* and *Clethrionomys*, admixture was estimated from SNP data for the purposes of this study (see the Supplementary Materials and Marková et al. 2020, 2023). Admixture estimates were expressed as Q-scores calculated using STRUCTURE (Pritchard et al. 2000), ADMIXTURE (Alexander et al. 2009), or NGSadmix (Skotte et al. 2013). For a part of *Sphyrapicus* samples, we used the genetic hybrid index estimates from Seneviratne et al. (2016), based on 180 SNPs, which we consider fully comparable to Q-scores. References, a summary of the markers and the proportion of admixed individuals, are given in supplementary table S1, Supplementary Material online.

To compare all hybrid zones under a similar and conservative regime, we treated all individuals with genome-wide admixture > 3% as admixed. This admixture threshold is often more stringent than the thresholds used in the original studies, which were tailored to the characteristics of the systems and the specific aims of the studies. Despite our best efforts, the sets of individuals genotyped for MHC and those with genome-wide admixture estimates did not always overlap completely, although overall genome-wide and MHC introgression were mostly estimated for the same individuals. In cases where genome-wide estimates were unavailable (e.g. *Salmo*), we treated an individual as nonadmixed if its population of origin was outside any contact zone or if the locality from which it originated contained only nonadmixed individuals of a single species. These cases are marked as "inferred" in supplementary table S5, Supplementary Material online. In addition, since the original studies often examined single hybrid zones involving only two species, we set the level of

admixture from all other species to 0 for multispecies systems. We consider this approach justified because hybrid zones between different species in a genus are either geographically isolated (e.g. *Natrix*) or the original studies sampled transects beyond the areas of potential introgression with other species (e.g. *Sphyrapicus*). The compiled estimates of genome-wide admixture are given in supplementary table S5, Supplementary Material online.

MHC Data
Samples

To genotype the MHC genes, we used previously obtained DNA samples for all taxa analyzed. We amplified the variable second exons of the MHC-I and MHC-IIB genes. The only exception was *Clethrionomys*, for which, due to primer availability, we amplified the third MHC-I exon and the second exon of one of the MHC-IIB genes (*DQB*). We aimed to maximize the number of populations while sampling several individuals from each locality to account for intrapopulation variation. The number of species, localities, and individuals for each taxon is shown in supplementary table S1, Supplementary Material online. Maps of sampled localities are provided in supplementary figs. S3 to S11, Supplementary Material online.

MHC Primers

The MHC data for *Lissotriton*, *Triturus*, and *Podarcis* were reported previously (Dudek et al. 2019; Gaczorek et al. 2023a, 2023b). For all other systems, including several additional allopatric *Lissotriton* populations located outside of previously analyzed transects, MHC genes were genotyped for the purpose of this study. Except for *Clethrionomys DQB* (Migalska et al. 2022), we used newly designed primers.

To design primers, we started by collecting available MHC sequences from as many species within a genus as possible to capture sequence variation in primer binding sites. For *Salmo*, this was straightforward as the sequences of *Salmo salar* MHC alleles used for primer design were retrieved from the IPD-MHC database (Maccari et al. 2017). For *Clethrionomys*, we used previously reported sequences of the third MHC-I exon (Migalska et al. 2017, 2019). In all other cases, we relied on existing genome or transcriptome assemblies (Benzekri et al. 2014; Nürnberger et al. 2021) or generated de novo transcriptome assemblies ourselves with Trinity (Grabherr et al. 2011) using publicly available RNAseq data. Due to the lack of genomic resources for *Anguis*, we sequenced transcriptomes of two *Anguis colchica* and two *Anguis fragilis* individuals (NCBI BioProject PRJNA1054985). With the assemblies in hand, we retrieved the MHC-I and MHC-IIB genes using tblastn or blastn from BLAST+ v2.9.0 (Camacho et al. 2009). As queries, we used sequences retrieved from Ensembl (Cunningham et al. 2022): human HLA-B (ENSG00000234745) and HLA-DRB1 (ENSG00000196126) proteins (*Solea*, *Bombina*, and *Anguis*), or sequences of the second exons of the MHC genes from *Chrysemys picta* (ENSCPBG00000005211, *Emys*) and

Gallus gallus (ENSGALG00000033932, *Sphyrapicus*). For *Natrix*, we used published sequences of the second MHC exon from *Podarcis* (Gaczorek et al. 2023a).

In the next step, to increase the sensitivity of our searches, we extracted the second exons from the MHC genes identified in the previous step and used them as queries for an additional examination of focal assemblies (blastn). Finally, we designed primers in regions of low variation. For *Clethrionomys* MHC-I, to achieve satisfactory amplification, we also sequenced flanking introns of the third exon and designed primers overlapping them. In all other cases, the primers were located within the second exon. The sequences of primers, expected product lengths, and accession numbers for the raw sequencing data are given in supplementary table S6, Supplementary Material online.

MHC Genotyping

MHC exons were amplified in 10-μl PCR reactions containing: 50- to 100-ng genomic DNA, 5 μl of Multiplex PCR Kit (Qiagen) and primers at concentrations of 0.5 to 1 μM. Individuals were barcoded with a combination of 6-bp indexes at the 5′ end of the forward and reverse primers. PCR conditions for MHC-I amplification were as follows: initial denaturation at 95 °C for 15 min, followed by 33 cycles: 95 °C for 30 s, 56 °C for 30 s, and 72 °C for 70 s and final elongation at 72 °C for 10 min. PCR conditions for MHC-II amplification were as follows: initial denaturation at 95 °C for 15 min, followed by 35 cycles: 95 °C for 30 s, 55 °C for 30 s, and 72 °C for 70 s and final elongation at 72 °C for 10 min. Amplicons were pooled approximately equimolar based on gel-band intensity, pools were gel purified, and Illumina adapters were ligated using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs) according to the manufacturer's protocol optimized for PCR-free workflow. Libraries were sequenced on an Illumina MiSeq using v2 500 cycle kits.

Genotyping was performed using the adjustable clustering method in AmpliSAS (Sebastian et al. 2016). Clustering and filtering parameters are detailed in supplementary table S2, Supplementary Material online. For *Emys*, we increased the *min_dominant_frequency_threshold* to 50% due to the overrepresentation of MHC clusters differing by a single base pair, indicating an elevated rate of sequencing errors. In the next step, we filtered out putatively nonfunctional alleles that differed by more than 6 bp from the expected product length or contained STOP codons or frameshift mutations. For each taxon and MHC class, we also adjusted the minimum per-amplicon frequency threshold to account for sequencing errors (supplementary table S2, Supplementary Material online) and filtered out all sequences that fell below this threshold. The threshold value was related to per-individual MHC diversity, with a higher threshold for taxa exhibiting lower diversity, indicating a lower number of MHC gene copies. Genotyping repeatability was estimated for multiple individuals (4 to 15) amplified and sequenced in two replicates by dividing the number of alleles detected in both replicates by the total number of

alleles identified. Finally, the data were transformed into a binary presence–absence matrix, with the number of rows equal to the number of individuals and the number of columns equal to the number of alleles.

We omitted additional filtering steps applied elsewhere (Fijarczyk et al. 2018; Dudek et al. 2019; Gaczorek et al. 2023a, 2023b) that aimed to enrich the datasets for functional alleles of classical MHC genes by filtering out putatively nonclassical/nonfunctional variants based on expression information and phylogenetic clustering. Although these approaches are useful, they are highly taxon specific and their consistent application across a multitaxon data set was not feasible. Therefore, for *Lissotriton*, *Triturus* MHC-I, and *Podarcis*, for which MHC sequences have already been published, we included in our analyses all sequences categorized as putatively nonfunctional, except those containing STOP codons or frameshift mutations. For *Triturus* MHC-II, due to strong evidence for a single functional locus (Babik et al. 2009; Gaczorek et al. 2023b), we retained only MHC alleles assigned to this locus. The procedures described above used analytical settings tailored to the MHC characteristics of each taxon to achieve a single overarching goal—complete information about the MHC variants possessed by each individual. In other words, the heterogeneity of genotyping settings was necessary to achieve uniformity of MHC genotyping.

Tests of MHC Introgression

To test for adaptive introgression of MHC genes, we applied three complementary analyses: MHC allele sharing, geographic clines, and genomic clines. However, due to the requirements of each approach, sample availability, and hybrid zone characteristics, we were not able to apply every test to each taxon. The list of taxa used for each test is shown in supplementary table S4, Supplementary Material online.

MHC Allele Sharing

If MHC introgression is adaptive, it should extend beyond areas of detectable genome-wide introgression. At the same time, the signal of ongoing or recent introgression should be stronger in the vicinity of the contact. Therefore, we expect to observe an excess of interspecific MHC allele sharing in the vicinity of the contact zone but beyond the actual hybrid zone, defined as the area of detectable genome-wide introgression (near) compared to areas further from the contact (far).

We only considered nonadmixed individuals from localities with an average admixture of <5%. Localities were classified as "near" or "far" based on their distance from the center of the hybrid zone, aiming for comparable sample sizes in both categories while ensuring that "far" populations were located away from areas of current and historical contact. A detailed description of the categorization for each taxon can be found in Supplementary Materials and figs. S3 to S11, Supplementary Material online. The significance of differences in allele sharing between species in

"near" and "far" populations was tested with a permutation test, randomizing individuals within species between "near" and "far" populations 1,000 times.

For each hybrid zone and MHC class, we calculated a SES of differences in MHC allele sharing between "near" and "far" populations (Gotelli and McCabe 2002; Botta-Dukát 2018), which controls for variation in the permutation runs and, indirectly, for MHC variation in the system analyzed. This SES is a special case of Cohen's d effect size (Cohen 1988) where the size of one group (observed) is equal to 1. The absolute value of the SES measures the strength of the signal, while its sign indicates increased MHC allele sharing in "near" (+) or "far" populations (−).

Estimation of MHC Hybrid Index

For each MHC class and hybrid zone, we calculated the hybrid index (HI), which estimates the proportion of an individual's ancestry derived from one of the species, ranging between 0 and 1. The HI was estimated using the method suitable for dominant markers (Buerkle 2005). To obtain the reference allele frequencies for each species, we used all individuals from localities classified as "far." For *Lissotriton*, reference allele frequencies were calculated separately for two zones (IN and OUT zones of Zieliński et al. 2019). The effect of binary coding of MHC genotypes on HI estimation was assessed with simulations (see below).

Geographic Clines

We designated transects, or pseudotransects in the case of *Triturus*, in which sampling deviated considerably from a 1D transect (Gaczorek et al. 2023b), through the hybrid zones (supplementary figs. S3 to S11, Supplementary Material online) and fitted clines to population means of (i) genome-wide admixture and (ii) the MHC HI, for each MHC class separately. Three models of smooth clines, i.e. without separate modeling the tails of the cline, were fitted with hzar (Derryberry et al. 2014): (i) ancestry fixed at 0 and 1 at the ends of the transect ("none"), (ii) estimated equal admixture at both ends of the transect ("fixed"), and (iii) admixture estimated separately at both ends of the transect ("free"). The best model was selected using the Akaike information criterion. We avoided using the single "free" model in all cases to prevent overfitting.

To summarize the difference in width and center between geographic MHC and genome-wide clines across multiple hybrid zones with a scale-free measure, we calculated Hedges g effect sizes (Hedges 1981; Rosenberg et al. 2013) for each MHC class in each hybrid zone. We conservatively assumed that the 2 log-likelihood intervals of the estimated cline parameters corresponded to 95% confidence intervals and used these to obtain the standard deviations necessary to calculate Hedges g. To test for the effect of MHC variation on cline shift, the difference between the centers of MHC and genome-wide clines was calculated such that its positive values indicated a shift of the cline center toward the less MHC diverse species.

The Effect of Binary Coding of Multilocus MHC Genotypes

Individual-based simulations were performed in SLiM 4 (Haller and Messer 2023). The ancestral population of $N = 500$ diploid individuals evolved for $20N = 10,000$ generations before splitting into two descendant species (each with $N = 500$) that evolved in isolation. The MHC region contained five completely linked loci, each evolving according to the infinite alleles model with either (i) identical ($\theta_{1-5} = 0.6$) or (ii) different ($\theta_1 = 0.2$, $\theta_2 = 0.4$, $\theta_3 = 0.6$, $\theta_4 = 0.8$, and $\theta_5 = 1.0$) population mutation rates under strong ($s = 0.5$) negative frequency dependence. Fitness of an allele was defined as $w = 1 + s/2 − fs$, where f is the allele frequency, giving the range of possible fitness values ($1 − s/2$; $1 + s/2$). Each scenario was simulated 20 times. Transects through a clinal hybrid zone were constructed at three time points: 0.5, 1, and 4N generations after the split. Note that we were not attempting to simulate the formation and evolution of a cline but were interested in the effect of binary coding of MHC genotypes on the estimation of parameters of a known cline. At each time point, we created populations with expected admixture determined by position along the smooth cline ("none" model in hzar) of width $w = 6$ km, with the center in the middle of the transect. Fifteen equally spaced populations were formed within the 6-km center of the zone and ten equally spaced populations were formed at each 97-km tail of the transect. Each population contained ten individuals, composed of MHC haplotypes randomly sampled from the parental species, with the probability of a haplotype coming from each species determined by the locality's expected admixture proportion. The origin of each haplotype was known without error, so the actual admixture level of an individual and locality was calculated from these data and taken to represent genome-wide admixture ("known-ancestry"). The HI for each individual was calculated from its complete MHC genotype, i.e. considering each locus separately and assigning alleles to their loci of origin ("MHC genotype"), as well as multilocus MHC genotypes binary coded as described above ("MHC binary"). Reference allele frequencies were estimated from 30 individuals randomly sampled from each parental species. Thus, some alleles present in individuals in the transect had unknown reference frequencies and were ignored in the calculation of HI, as was the case in the analysis of empirical data.

Genomic Clines

To examine MHC ancestry relative to genome-wide ancestry in admixed individuals, we used a genomic cline approach—Barton's concordance analysis (Szymura and Barton 1986). For each hybrid zone and MHC class, we fitted, using maximum likelihood in R v. 4.2.2, the equation modeling individual MHC ancestry as a function of genome-wide ancestry: $p_{MHC} = \bar{p} + 2\bar{p}\bar{q}(\alpha + (\bar{p} − \bar{q})\beta)$ (Szymura and Barton 1986), where p_{MHC} is the proportion of the individual's MHC ancestry derived from species 1, $\bar{p}$ is the proportion of genome-wide ancestry derived from species 1, $\bar{q} = (1 − \bar{p})$ is the proportion of genome-wide

ancestry derived from species 2, and α and β are the genomic cline parameters. The α parameter measures a relative excess of MHC ancestry from one of the two species (genomic cline "shift"); in our case, a positive α always indicated an excess of ancestry from the more MHC diverse species. The β parameter measures the relative rate of transition in ancestry (genomic cline "width"), with $\beta < 0$ indicating a slower rate of transition in MHC, i.e. more MHC admixture than expected from genome-wide admixture. The genomic cline curve is characterized by considerable flexibility and fixed ends at points (0, 0) and (1, 1), corresponding to individuals without genome-wide admixture, which do not contribute to the cline fitting in this model. Note that we could not use the more recently developed approaches to fit genomic clines (Gompert and Buerkle 2011) as they require allele counts at individual loci, which were not available. The 2 log-likelihood parameter intervals were taken as a proxy for 95% confidence intervals, which we used to calculate the variances of the estimates. As the parameters of genomic clines are always measured on the same scale, the estimates and their variances were used directly in subsequent modeling.

Comparative Analysis

The significance of the signal of increased MHC introgression across taxa was tested using PGLS with the ape (Paradis and Schliep 2019), matrix (Bates et al. 2024), and nlme (Pinheiro et al. 2023) R packages. The time-calibrated phylogeny (Fig. 1) was used to obtain the expected correlation matrix between pairs of hybridizing species (note that pairs of species, not individual species, are the units of measurement here). The matrix of cophenetic distances, which summed the branch lengths between nodes representing the most common ancestors of hybridizing species, was transformed into the correlation matrix by dividing the distance between nodes by the cophenetic distance between the most distant species. The strength of the phylogenetic signal was estimated using Pagel's λ.

All models included SESs or genomic cline parameters as response variables and MHC class as an explanatory variable. Where parameters were estimated for multiple transects for the same species pair, both estimates and their variances were averaged. In a subset of models, we also included time of divergence between hybridizing species and the MHC class × time of divergence interaction. Models with and without interaction were compared using LRT in R anova(); and if interaction did not significantly improve fit, it was omitted. We assumed no phylogenetic correlation between MHC classes, and the effect of phylogeny within each class was included as described above. The model also included weights corresponding to the inverse of the variation in the estimates, accounting for their uncertainty. For randomization tests, the weights were not applied because a large and constant number of permutations consistently resulted in nearly identical variances (≈ 1) of SES for any reasonable value of effect size.

Supplementary Material

Acknowledgments

Funding

Conflict of Interest

The authors declare no conflicts of interest.

Data Availability

The data underlying this article are available in the article, in its online supplementary material, at: https://github.com/TomekGa/MHC_comparative_analysis and in NCBI BioProject PRJNA1054985. The code is at: https://github.com/TomekGa/MHC_comparative_analysis.

References

Abi-Rached L, Jobin MJ, Kulkarni S, McWhinnie A, Dalva K, Gragert L, Babrzadeh F, Gharizadeh B, Luo M, Plummer FA, et al. The shaping of modern human immune systems by multiregional admixture with archaic humans. Science. 2011:**334**(6052):89–94. https://doi.org/10.1126/science.1209202.

Aitken SN, Whitlock MC. Assisted gene flow to facilitate local adaptation to climate change. Annu Rev Ecol Evol Syst. 2013:**44**(1): 367–388. https://doi.org/10.1146/annurev-ecolsys-110512-135747.

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

Babik W, Pabijan M, Arntzen JW, Cogalniceanu D, Durka W, Radwan J. Long-term survival of a urodele amphibian despite depleted major histocompatibility complex variation. Mol Ecol. 2009:**18**(5):769–781. https://doi.org/10.1111/j.1365-294X.2008.04057.x.

Baird SJ, Goüy de Bellocq J. Shifting paradigms for studying parasitism in hybridising hosts: response to Theodosopoulos, Hund, and Taylor. Trends Ecol Evol. 2019:**34**(5):387–389. https://doi.org/10.1016/j.tree.2019.01.011.

Baird SJ, Ribas A, Macholán M, Albrecht T, Piálek J, Goüy de Bellocq J. Where are the wormy mice? A reexamination of hybrid parasitism in the European house mouse hybrid zone. Evolution. 2012:**66**(9): 2757–2772. https://doi.org/10.1111/j.1558-5646.2012.01633.x.

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

Barton NH. Multilocus clines. Evolution. 1983:**37**(3):454–471. https://doi.org/10.2307/2408260.

Barton NH. On the completion of speciation. Philos Trans R Soc Lond B. 2020:**375**(1806):20190530. https://doi.org/10.1098/rstb.2019.0530.

Barton N, Bengtsson BO. The barrier to genetic exchange between hybridising populations. Heredity (Edinb). 1986:**57**(3):357–376. https://doi.org/10.1038/hdy.1986.135.

Bates D, Maechler M, Jagan M. Matrix: sparse and dense matrix classes and methods. R package version 1.7. 2024. https://CRAN.R-project.org/package=Matrix.

Bell G. Evolutionary rescue and the limits of adaptation. Philos Trans R Soc Lond B. 2013:**368**(1610):20120080. https://doi.org/10.1098/rstb.2012.0080.

Benzekri H, Armesto P, Cousin X, Rovira M, Crespo D, Merlo MA, Mazurais D, Bautista R, Guerrero-Fernández D, Fernandez-Pozo N, et al. De novo assembly, characterization and functional annotation of Senegalese sole (Solea senegalensis) and common sole (Solea solea) transcriptomes: integration in a database and design of a microarray. BMC Genomics. 2014:**15**(1):952. https://doi.org/10.1186/1471-2164-15-952.

Botta-Dukát Z. Cautionary note on calculating standardized effect size (SES) in randomization test. Commun Ecol. 2018:**19**(1): 77–83. https://doi.org/10.1556/168.2018.19.1.8.

Buerkle CA. Maximum-likelihood estimation of a hybrid index based on molecular markers. Mol Ecol Notes. 2005:**5**(3):684–687. https://doi.org/10.1111/j.1471-8286.2005.01011.x.

Butlin RK, Smadja CM. Coupling, reinforcement, and speciation. Am Nat. 2018:**191**(2):155–172. https://doi.org/10.1086/695136.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, BealerK, Madden TL. BLAST+: architecture and applications. BMC Bioinf. 2009:**10**(1):421–429. https://doi.org/10.1186/1471-2105-10-421.

Castric V, Bechsgaard J, Schierup MH, Vekemans X. Repeated adaptive introgression at a gene under multiallelic balancing selection. PLoS Genet. 2008:**4**(8):e1000168. https://doi.org/10.1371/journal.pgen.1000168.

Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale (NJ): Lawrence Erlbaum; 1988.

Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Austine-Orimoloye O, Azov AG, Barnes I, Bennett R, et al. Ensembl 2022. Nucl Acid Res. 2022:**50**(D1):D988–D995. https://doi.org/10.1093/nar/gkab1049.

De Coster W, Weissensteiner MH, Sedlazeck FJ. Towards population-scale long-read sequencing. Nat Rev Genet. 2021:**22**(9):572–587. https://doi.org/10.1038/s41576-021-00367-3.

Dendrou CA, Petersen J, Rossjohn J, Fugger L. HLA variation and disease. Nat Rev Immun. 2018:**18**(5):325–339. https://doi.org/10.1038/nri.2017.143.

Derryberry EP, Derryberry GE, Maley JM, Brumfield RT. HZAR: hybrid zone analysis using an R software package. Mol Ecol Res. 2014:**14**(3):652–663. https://doi.org/10.1111/1755-0998.12209.

Dilthey AT. State-of-the-art genome inference in the human MHC. Int J Biochem Cell Biol. 2021:**131**:105882. https://doi.org/10.1016/j.biocel.2020.105882.

Dudek K, Gaczorek T, Zieliński P, Babik W. Massive introgression of MHC genes in newt hybrid zones. Mol Ecol. 2019:**28**(21): 4798–4810. https://doi.org/10.1111/mec.15254.

Ebert P, Audano PA, Zhu Q, Rodriguez-Martin B, Porubsky D, BonderMJ, Sulovari A, Ebler J, Zhou W, Serra Mari R, et al. Haplotype-resolved diverse human genomes and integrated analysis of structural variation. Science. 2021:**372**(6537):eabf7117. https://doi.org/10.1126/science.abf7117.

Edelman NB, Mallet J. Prevalence and adaptive impact of introgression. Annu Rev Genet. 2021:**55**(1):265–283. https://doi.org/10.1146/annurev-genet-021821-020805.

Eizenga JM, Novak AM, Sibbesen JA, Heumos S, Ghaffaari A, Hickey G, Chang X, Seaman JD, Rounthwaite R, Ebler J, et al. Pangenome graphs. Annu Rev Genom Hum Genet. 2020:**21**(1):139–162. https://doi.org/10.1146/annurev-genom-120219-080406.

Ejsmond MJ, Phillips KP, Babik W, Radwan J. The role of MHC supertypes in promoting trans-species polymorphism remains an open question. Nat Comm. 2018:**9**(1):4362. https://doi.org/10.1038/s41467-018-06821-x.

Enard D, Petrov DA. Evidence that RNA viruses drove adaptive introgression between Neanderthals and modern humans. Cell. 2018:**175**(2):360–371.e13. https://doi.org/10.1016/j.cell.2018.08.034.

Fijarczyk A, Babik W. Detecting balancing selection in genomes: limits and prospects. Mol Ecol. 2015:**24**(14):3529–3545. https://doi.org/10.1111/mec.13226.

Fijarczyk A, Dudek K, Niedzicka M, Babik W. Balancing selection and introgression of newt immune-response genes. Proc Biol Sci. 2018:**285**(1884):20180819. https://doi.org/10.1098/rspb.2018.0819.

Fitzpatrick BM. Alternative forms for genomic clines. Ecol Evol. 2013:**3**(7):1951–1966. https://doi.org/10.1002/ece3.609.

Gaczorek TS, Chechetkin M, Dudek K, Caeiro-Dias G, Crochet PA, Geniez P, Pinho C, Babik W. Widespread introgression of MHC genes in Iberian Podarcis lizards. Mol Ecol. 2023a:**32**(14): 4003–4017. https://doi.org/10.1111/mec.16974.

Gaczorek TS, Marszałek M, Dudek K, Arntzen JW, Wielstra B, Babik W. Interspecific introgression of MHC genes in Triturus newts: evidence from multiple contact zones. Mol Ecol. 2023b:**32**(4): 867–880. https://doi.org/10.1111/mec.16804.

Giska I, Farelo L, Pimenta J, Seixas FA, Ferreira MS, Marques JP, Miranda I, Letty J, Jenny H, Hackländer K, et al. Introgression drives repeated evolution of winter coat color polymorphism in hares. Proc Natl Acad Sci U S A. 2019:**116**(48):24150–24156. https://doi.org/10.1073/pnas.1910471116.

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

Gompert Z, Mandeville EG, Buerkle CA. Analysis of population genomic data from hybrid zones. Annu Rev Ecol Evol Syst. 2017:**48**(1):207–229. https://doi.org/10.1146/annurev-ecolsys-110316-022652.

Gotelli NJ, McCabe DJ. Species co-occurrence: a meta-analysis of JM Diamond's assembly rules model. Ecology. 2002:**83**(8): 2091–2096. https://doi.org/10.1890/0012-9658(2002)083[2091:SCOAMA]2.0.CO;2.

Gower G, Picazo PI, Fumagalli M, Racimo F. Detecting adaptive introgression in human evolution using convolutional neural networks. eLife. 2021:**10**:e64669. https://doi.org/10.7554/eLife.64669.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotech. 2011:**29**(7):644–652. https://doi.org/10.1038/nbt.1883.

Greenbaum G, Getz WM, Rosenberg NA, Feldman MW, Hovers E, Kolodny O. Disease transmission and introgression can explain the long-lasting contact zone of modern humans and Neanderthals. Nat Commun. 2019:**10**(1):5003. https://doi.org/10.1038/s41467-019-12862-7.

Grossen C, Keller L, Biebach I, Croll D, Consortium IGG. Introgression from domestic goat generated variation at the major histocompatibility complex of Alpine ibex. PLoS Genet. 2014:**10**(6): e1004438. https://doi.org/10.1371/journal.pgen.1004438.

Haller BC, Messer PW. SLiM 4: multispecies eco-evolutionary modeling. Am Nat. 2023:**201**(5):E127–E139. https://doi.org/10.1086/723601.

Hamilton JA, Miller JM. Adaptive introgression as a resource for management and genetic conservation in a changing climate. Conserv Biol. 2016:**30**(1):33–41. https://doi.org/10.1111/cobi.12574.

Hedges LV. Distribution theory for Glass's estimator of effect size and related estimators. J Educ Stat. 1981:**6**(2):107–128. https://doi.org/10.3102/10769986006002107.

Hedrick PW. Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. Mol Ecol. 2013:22(18):4606–4618. https://doi.org/10.1111/mec.12415.

Herdegen-Radwan M, Phillips KP, Babik W, Mohammed RS, Radwan J. Balancing selection versus allele and supertype turnover in MHC class II genes in guppies. Heredity (Edinb). 2021:126(3):548–560. https://doi.org/10.1038/s41437-020-00369-7.

Irisarri I, Baurain D, Brinkmann H, Delsuc F, Sire J-Y, Kupfer A, Petersen J, Jarek M, Meyer A, Vences M, et al. Phylotranscriptomic consolidation of the jawed vertebrate timetree. Nat Ecol Evol. 2017:1(9):1370–1378. https://doi.org/10.1038/s41559-017-0240-5.

Kulmuni J, Butlin RK, Lucek K, Savolainen V, Westram AM. Towards the completion of speciation: the evolution of reproductive isolation beyond the first barriers. Philos Trans R Soc Lond B. 2020:375(1806):20190528. https://doi.org/10.1098/rstb.2019.0528.

Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: a resource for timelines, timetrees, and divergence times. Mol Biol Evol. 2017:34(7):1812–1819. https://doi.org/10.1093/molbev/msx116.

Li T, Du D, Zhang D, Lin Y, Ma J, Zhou M, Meng W, Jin Z, Chen Z, Yuan H, et al. CRISPR-based targeted haplotype-resolved assembly of a megabase region. Nat Commun. 2023:14(1):22. https://doi.org/10.1038/s41467-022-35389-w.

Liao W-W, Asri M, Ebler J, Doerr D, Haukness M, Hickey G, Lu S, Lucas JK, Monlong J, Abel HJ, et al. A draft human pangenome reference. Nature. 2023:617(7960):312–324. https://doi.org/10.1038/s41586-023-05896-x.

Maccari G, Robinson J, Ballingall K, Guethlein LA, Grimholt U, Kaufman J, Ho C-S, De Groot NG, Flicek P, Bontrop RE, et al. IPD-MHC 2.0: an improved inter-species database for the study of the major histocompatibility complex. Nucl Acid Res. 2017:45(D1):D860–D864. https://doi.org/10.1093/nar/gkw1050.

Marková S, Horníková M, Lanier HC, Henttonen H, Searle JB, Weider LJ, Kotlík P. High genomic diversity in the bank vole at the northern apex of a range expansion: the role of multiple colonizations and end-glacial refugia. Mol Ecol. 2020:29(9):1730–1744. https://doi.org/10.1111/mec.15427.

Marková S, Lanier HC, Escalante MA, da Cruz MO, Horníková M, Konczal M, Weider LJ, Searle JB, Kotlík P. Local adaptation and future climate vulnerability in a wild rodent. Nat Comm. 2023:14(1):7840. https://doi.org/10.1038/s41467-023-43383-z.

McConnell SC, Hernandez KM, Wcisel DJ, Kettleborough RN, Stemple DL, Yoder JA, Andrade J, de Jong JL. Alternative haplotypes of antigen processing genes in zebrafish diverged early in vertebrate evolution. Proc Natl Acad Sci U S A. 2016:113(34):E5014–E5023. https://doi.org/10.1073/pnas.1607602113.

Migalska M, Przesmycka K, Alsarraf M, Bajer A, Behnke-Borowczyk J, Grzybek M, Behnke JM, Radwan J. Long term patterns of association between MHC and helminth burdens in the bank vole support Red Queen dynamics. Mol Ecol. 2022:31(12):3400–3415. https://doi.org/10.1111/mec.16486.

Migalska M, Sebastian A, Konczal M, Kotlík P, Radwan J. De novo transcriptome assembly facilitates characterisation of fast-evolving gene families, MHC class I in the bank vole (Myodes glareolus). Heredity (Edinb). 2017:118(4):348–357. https://doi.org/10.1038/hdy.2016.105.

Migalska M, Sebastian A, Radwan J. Major histocompatibility complex class I diversity limits the repertoire of T cell receptors. Proc Natl Acad Sci U S A. 2019:116(11):5021–5026. https://doi.org/10.1073/pnas.1807864116.

Moran BM, Payne C, Langdon Q, Powell DL, Brandvain Y, Schumer M. The genomic consequences of hybridization. eLife. 2021:10:e69016. https://doi.org/10.7554/eLife.69016.

Murphy K, Weaver C, Berg L. Janeway's immunobiology. New York: Norton; 2022.

Nosil P, Feder JL, Flaxman SM, Gompert Z. Tipping points in the dynamics of speciation. Nat Ecol Evol. 2017:1(2):0001. https://doi.org/10.1038/s41559-016-0001.

Nürnberger B, Baird SJ, Čížková D, Bryjová A, Mudd AB, Blaxter ML, Szymura JM. A dense linkage map for a large repetitive genome: discovery of the sex-determining region in hybridizing fire-bellied toads (Bombina bombina and Bombina variegata). G3. 2021:11(12):jkab286. https://doi.org/10.1093/g3journal/jkab286.

Pabijan M, Wandycz A, Hofman S, Węcek K, Piwczyński M, Szymura JM. Complete mitochondrial genomes resolve phylogenetic relationships within Bombina (Anura: Bombinatoridae). Mol Phyl Evol. 2013:69(1):63–74. https://doi.org/10.1016/j.ympev.2013.05.007.

Palomar G, Dudek K, Migalska M, Arntzen J, Ficetola GF, Jelić D, Jockusch E, Martínez-Solano I, Matsunami M, Shaffer HB, et al. Coevolution between MHC class I and antigen-processing genes in salamanders. Mol Biol Evol. 2021:38(11):5092–5106. https://doi.org/10.1093/molbev/msab237.

Paradis E, Schliep K. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. 2019:35(3):526–528. https://doi.org/10.1093/bioinformatics/bty633.

Payseur BA. Using differential introgression in hybrid zones to identify genomic regions involved in speciation. Mol Ecol Res. 2010:10(5):806–820. https://doi.org/10.1111/j.1755-0998.2010.02883.x.

Phillips KP, Cable J, Mohammed RS, Herdegen-Radwan M, Raubic J, Przesmycka KJ, van Oosterhout C, Radwan J. Immunogenetic novelty confers a selective advantage in host–pathogen coevolution. Proc Natl Acad Sci U S A. 2018:115(7):1552–1557. https://doi.org/10.1073/pnas.1708597115.

Piálek J, Barton NH. The spread of an advantageous allele across a barrier: the effects of random drift and selection against heterozygotes. Genetics. 1997:145(2):493–504. https://doi.org/10.1093/genetics/145.2.493.

Pierini F, Lenz TL. Divergent allele advantage at human MHC genes: signatures of past and ongoing selection. Mol Biol Evol. 2018:35(9):2145–2158. https://doi.org/10.1093/molbev/msy116.

Pinheiro J, Bates D, R Core Team. 2023. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3. https://CRAN.R-project.org/package=nlme.

Pishesha N, Harmand TJ, Ploegh HL. A guide to antigen processing and presentation. Nat Rev Immun. 2022:22(12):751–764. https://doi.org/10.1038/s41577-022-00707-2.

Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000:155(2):945–959. https://doi.org/10.1093/genetics/155.2.945.

Rabosky DL, Chang J, Title PO, Cowman PF, Sallan L, Friedman M, Kaschner K, Garilao C, Near TJ, Coll M, et al. An inverse latitudinal gradient in speciation rate for marine fishes. Nature. 2018:559(7714):392–395. https://doi.org/10.1038/s41586-018-0273-1.

Radwan J, Babik W, Kaufman J, Lenz TL, Winternitz J. Advances in the evolutionary understanding of MHC polymorphism. Trends Genet. 2020:36(4):298–311. https://doi.org/10.1016/j.tig.2020.01.008.

Ralls K, Ballou JD, Dudash MR, Eldridge MD, Fenster CB, Lacy RC, Sunnucks P, Frankham R. Call for a paradigm shift in the genetic management of fragmented populations. Conserv Lett. 2018:11(2):e12412. https://doi.org/10.1111/conl.12412.

Ravinet M, Yoshida K, Shigenobu S, Toyoda A, Fujiyama A, Kitano J. The genomic landscape at a late stage of stickleback speciation: high genomic divergence interspersed by small localized regions of introgression. PLoS Genet. 2018:14(5):e1007358. https://doi.org/10.1371/journal.pgen.1007358.

Rosenberg MS, Rothstein HR, Gurevitch J. Effect sizes: conventional choices and calculations. In: Handbook of meta-analysis in ecology and evolution. Princeton: Princeton University Press; 2013. p. 61–71.

Rousselle M, Simion P, Tilak M-K, Figuet E, Nabholz B, Galtier N. Is adaptation limited by mutation? A timescale-dependent effect of genetic diversity on the adaptive substitution rate in animals. PLoS Genet. 2020:16(4):e1008668. https://doi.org/10.1371/journal.pgen.1008668.

MBE

Sato A, Figueroa F, Murray BW, Málaga-Trillo E, Zaleska-Rutczynska Z, Sültmann H, Toyosawa S, Wedekind C, Steck N, Klein J. Nonlinkage of major histocompatibility complex class I and class II loci in bony fishes. Immunogenetics. 2000:**51**(2):108–124. https://doi.org/10.1007/s002510050019.

Schierup MH, Vekemans X, Charlesworth D. The effect of subdivision on variation at multi-allelic loci under balancing selection. Genet Res. 2000:**76**(1):51–62. https://doi.org/10.1017/S0016672300004535.

Schöneberg Y, Winter S, Arribas O, Di Nicola MR, Master M, Owens JB, Rovatsos M, Wüster W, Janke A, Fritz U. Genomics reveals broad hybridization in deeply divergent Palearctic grass and water snakes (*Natrix* spp.). Mol Phyl Evol. 2023:**184**:107787. https://doi.org/10.1016/j.ympev.2023.107787.

Sebastian A, Herdegen M, Migalska M, Radwan J. Amplisas: a web server for multilocus genotyping using next-generation amplicon sequencing data. Mol Ecol Res. 2016:**16**(2):498–510. https://doi.org/10.1111/1755-0998.12453.

Seneviratne SS, Davidson P, Martin K, Irwin DE. Low levels of hybridization across two contact zones among three species of woodpeckers (*Sphyrapicus* sapsuckers). J Avian Biol. 2016:**47**(6):887–898. https://doi.org/10.1111/jav.00946.

Setter D, Mousset S, Cheng X, Nielsen R, DeGiorgio M, Hermisson J. VolcanoFinder: genomic scans for adaptive introgression. PLoS Genet. 2020:**16**(6):e1008867. https://doi.org/10.1371/journal.pgen.1008867.

Skotte L, Korneliussen TS, Albrechtsen A. Estimating individual admixture proportions from next generation sequencing data. Genetics. 2013:**195**(3):693–702. https://doi.org/10.1534/genetics.113.154138.

Song Y, Endepols S, Klemann N, Richter D, Matuschka FR, Shih CH, Nachman MW, Kohn MH. Adaptive introgression of anticoagulant rodent poison resistance by hybridization between old world mice. Curr Biol. 2011:**21**(15):1296–1301. https://doi.org/10.1016/j.cub.2011.06.043.

Spurgin LG, Richardson DS. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. Proc Biol Sci. 2010:**277**(1684):979–988. https://doi.org/10.1098/rspb.2009.2084.

Svedberg J, Shchur V, Reinman S, Nielsen R, Corbett-Detig R. Inferring adaptive introgression using hidden Markov models.

Mol Biol Evol. 2021:**38**(5):2152–2165. https://doi.org/10.1093/molbev/msab014.

Szymura JM, Barton NH. Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina bombina* and *B. variegata*, near Cracow in southern Poland. Evolution. 1986:**40**(6):1141–1159. https://doi.org/10.1111/j.1558-5646.1986.tb05740.x.

Theodosopoulos AN, Hund AK, Taylor SA. Parasites and host species barriers in animal hybrid zones. Trends Ecol Evol. 2019:**34**(1):19–30. https://doi.org/10.1016/j.tree.2018.09.011.

van Dijk EL, Naquin D, Gorrichon K, Jaszczyszyn Y, Ouazahrou R, Thermes C, Hernandez C. Genomics in the long-read sequencing era. Trends Genet. 2023:**39**(9):649–671. https://doi.org/10.1016/j.tig.2023.04.006.

Westerdahl H, Mellinger S, Sigeman H, Kutschera VE, Proux-Wéra E, Lundberg M, Weissensteiner M, Churcher A, Bunikis I, Hansson B, et al. The genomic architecture of the passerine MHC region: high repeat content and contrasting evolutionary histories of single copy and tandemly duplicated MHC genes. Mol Ecol Res. 2022:**22**(6):2379–2395. https://doi.org/10.1111/1755-0998.13614.

Whitney KD, Randell RA, Rieseberg LH. Adaptive introgression of herbivore resistance traits in the weedv sunflower *Helianthus annuus*. Am Nat. 2006:**167**(6):794–807. https://doi.org/10.1086/504606.

Wielstra B, McCartney-Melstad E, Arntzen J, Butlin RK, Shaffer HB. Phylogenomics of the adaptive radiation of *Triturus* newts supports gradual ecological niche expansion towards an incrementally aquatic lifestyle. Mol Phyl Evol. 2019:**133**:120–127. https://doi.org/10.1016/j.ympev.2018.12.032.

Yang W, Feiner N, Pinho C, While GM, Kaliontzopoulou A, Harris DJ, Salvi D, Uller T. Extensive introgression and mosaic genomes of Mediterranean endemic lizards. Nat Commun. 2021:**12**(1):2762. https://doi.org/10.1038/s41467-021-22949-9.

Zieliński P, Dudek K, Arntzen JW, Palomar G, Niedzicka M, Fijarczyk A, Liana M, Cogălniceanu D, Babik W. Differential introgression across newt hybrid zones—evidence from replicated transects. Mol Ecol. 2019:**28**(21):4811–4824. https://doi.org/10.1111/mec.15251.