ORIGINAL ARTICLE

ASSESSING MULTILOCUS INTROGRESSION PATTERNS: A CASE STUDY ON THE MOUSE X CHROMOSOME IN CENTRAL EUROPE

Miloš Macholán,[1,2] Stuart J. E. Baird,[3] Petra Dufková,[4,5] Pavel Munclinger,[6] Barbora Vošlajerová Bímová,[1,4] and Jaroslav Piálek[4,*]

[1]Laboratory of Mammalian Evolutionary Genetics, Institute of Animal Physiology and Genetics, Academy of Sciences of the Czech Republic, Brno, Czech Republic

[2]E-mail: macholan@iach.cz

[3]CIBIO, University of Porto, Campus Agrário de Vairão, Vairão, Portugal

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

[5]Department of Genetics, Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic

[6]Biodiversity Research Group, Department of Zoology, Faculty of Science, Charles University in Prague, Czech Republic

Multilocus hybrid zone (HZ) studies predate genomics by decades. The power of early methods is becoming apparent and now large datasets are commonplace. Relating introgression along a chromosome to evolutionary process is challenging: although reduced introgression regions may indicate speciation genes, this pattern may be obscured by asymmetric introgression of linked invasive genes. Further, HZ movement may form salients and leave islands in its wake. Barton's concordance was proposed 24 years ago for assessing introgression where geographic patterns are complex. The geographic axis of introgression is replaced with the hybrid index. We compare this, a recently proposed genomic clines approach, and two-dimensional (2D) geographic analyses, for 24 X chromosome loci of 2873 mice from the central-European house mouse HZ. In 2D, 14 loci show linear contact, seven precisely matching previous studies. Four show introgression islands to the east of the zone, suggesting past westward zone movement, two show westward salients. Barton's concordance both recovers and refines this information. A region of reduced introgression on the central X is supported, despite X centromere-proximal male-biased westward introgression matching a westward 2D geographic salient. Genomic clines results are consistent regarding introgression asymmetries, but otherwise more difficult to interpret. Evidence for genetic conflict is discussed.

KEY WORDS: Genetic conflict, hybrid zone movement, *Mus musculus musculus*, *Mus musculus domesticus*, sex biased introgression, X chromosome.

Elucidating the genetic basis of reproductive isolation is one of the crucial problems in the study of speciation. Much current research focuses on two areas: reproductive isolation between nascent species arising as a result of accumulation of incom-

*Contact address: Department of Population Biology, IVB, ASCR, Studenec, Czech Republic.

patibilities between interacting loci, the Bateson–Dobzhansky–Muller model (Bateson 1909; Dobzhansky 1936; Muller 1940, 1942; Orr 1996; though see Forsdyke 2011), and the role of genes involved in genetic conflicts (Tao et al. 2001; Orr and Irving 2005; Orr et al. 2007; Macholán et al. 2008; Phadnis and Orr 2009). Regarding the distribution of such genes across the genome, it has been shown that sex chromosomes harbor more genes

causing disruption of fertility and/or viability in hybrids than autosomes (Grula and Taylor 1980; Zouros et al. 1988; Coyne and Orr 1989, 2004; Prowell 1998; Jiggins et al. 2001; Tao et al. 2003; Counterman et al. 2004; Oka et al. 2004; Storchová et al. 2004, 2010; Harr 2006; Good et al. 2008) as predicted by the "large X effect" hypothesized to explain Haldane's rule (Haldane 1922; Orr 1997; Coyne and Orr 2004). One approach to furthering our understanding of reproductive barriers is to cross species displaying partial reproductive isolation under controlled laboratory conditions and then examine the basis of hybrid incompatibilities. An alternative approach is based on the study of naturally occurring hybrid zones (HZ).

Notwithstanding the precise causes and processes underlying the origin and evolution of reproductive barriers, gene flow at loci responsible for these barriers is expected to be impeded, resulting in abrupt changes of allele frequencies across a contact zone, that is, narrow clines that will tend to coincide geographically unless forcibly perturbed (Barton and Gale 1993). Conversely, alleles at neutral loci should introgress more or less freely, to distance^2 proportional with time elapsed since contact. Hence, the relationship between cline width and selection seems straightforward: the stronger the counterselection, the narrower the cline (Slatkin 1973, 1975; Nagylaki 1975, 1976; Endler 1977). This simple relationship has led to the suggestion that plotting introgression measures for markers of known position in the genome, we could potentially detect genomic areas responsible for reproductive isolation, or speciation genes (Rieseberg et al. 1999; Payseur et al. 2004; Buerkle and Lexer 2008; Gompert and Buerkle 2009).

Patterns of introgression will be more complicated if not all genes are simply neutral or responsible for barriers to gene flow. Payseur et al. (2004) in their seminal analysis of X chromosome introgression along the Bavarian transect across the European house mouse HZ, discuss in particular the effects of adaptive introgression. Here, we wish to take into account also the potential importance of (nonadaptive) genetic conflict in driving introgression. Deterministic introgression occurs when a gene has an advantage over alternative alleles. If the gene is also advantageous to the individuals in which it is expressed, there is no genetic conflict and introgression is indeed adaptive (i.e., increasing population fitness). However, a gene's advantage may allow it to increase despite lowering population fitness (for example by distorting sex ratio away from the Fisher equilibrium; Macholán et al. 2008). In this case its action is selfish, its spread is nonadaptive, and it is not advantageous to all the individuals in which it is expressed. It seems including the potential effects of nonadaptive introgression on HZ dynamics may be important, as gene advantages due, for example, to segregation distortion can be much greater than otherwise observed in nature (Galtier et al. 2009). Fortunately, the origin and fate of both types of gene can be described simul-

taneously by referring to them as "advantaged" genes. During isolation of populations, an advantaged gene may arise and increase in one population that would also have an advantage if it were introduced into other populations. On secondary contact such universally advantaged genes are predicted to quickly recombine away from their original background and introgress across HZs (Barton and Hewitt 1985; Barton and Bengtsson 1986; Barton and Gale 1993; Piálek and Barton 1997). Given time, they will spread far away from the original contact area, and so perhaps become of little relevance to HZ studies. However, if universally advantaged genes are tightly linked to speciation genes, they may be unable to quickly recombine away from their original background, making their action highly relevant to HZ studies even many generations after secondary contact. If linkage groups in a HZ include universally advantaged genes trapped within a background of speciation genes we may expect asymmetric, noncoincident clines and HZ movement (cf. Buggs 2007). In addition, if advantaged genes are in the process of escaping their background, descendants of particular background recombinants will become over-represented. Then, effective recombination (depending on the number of descendants inheriting a recombination event) may bear little relation to map distance (depending only on the frequency with which recombination events occur). This effect may give rise to a genomic equivalent of the geographic pattern of "footprints" left by a moving HZ.

The spread of advantaged genes, and by analogy genomes, need not be geographically smooth. Instead, advance may be through the formation of salients (Ibrahim et al. 1996), which then merge around invaginations. This geographic stochasticity may lead to scattered islands of one genome type, in their original territory, but now surrounded by the other genome type, a pattern first called the footprints of a moving HZ by Arntzen (Arntzen and Wallis 1991; Espregueira Themudo and Arntzen 2007). If HZ movement is driven by advantaged genes, then this stochastic geographic process, coupled with high effective recombination between genetic backgrounds (see above), will tend to lead also to genomic islands of introgression, scattered across the genome, each associated with different geographic islands. This passive stochastic pattern of introgression by inclusion is quite different from that expected of the deterministic introgression of advantaged genes.

All of these predicted introgression patterns arising from different processes are potentially overlaid across the geography of a HZ, and along recombining chromosomes. A HZ survey of the heterogeneity of introgression patterns along, for example, an X chromosome may therefore potentially reveal a great deal about the nature and recent history of sex chromosomes as barriers between taxa. A problem arises, however, in accessing this information: extremely intensive geographic sampling is necessary to capture and de-confound such complex geographic

patterns. In designing his multilocus analyses of the *Bombina* HZ (Szymura and Barton 1986), Barton saw this fundamental limitation, and produced what we call here "Barton's concordance analysis," which replaces the geographic axis of introgression with the hybrid index, an analysis designed specifically for the case in which geographic pattern is complex or difficult to sample.

The house mouse (*Mus musculus*) appears a very useful model for studies of differential introgression of various parts of the genome across HZs. The complete genome sequence is known (Mouse Genome Sequencing Consortium 2002) and a vast array of suitable molecular markers is available for this species (Dietrich et al. 1996; Lindblad-Toh et al. 2000; Abe et al. 2004; Pletcher et al. 2004; Shifman et al. 2006). Moreover, a HZ between two subspecies, *M. m. musculus* and *M. m. domesticus* (sometimes referred to as species, *M. musculus* and *M. domesticus*), exists in Europe, running across the Jutland peninsula in Denmark and from the Baltic coast across central Europe and the Balkans to the Black Sea (Boursot et al. 1993; Sage et al. 1993; Macholán et al. 2003; Fig. 1). Studies carried out on various portions of this zone have shown a remarkably consistent picture of wider autosomal clines versus more narrow sex linked clines spanning more than 2500 km of the contact front between the two subspecies (Vanlerberghe et al. 1988a; Tucker et al. 1992; Dod et al. 1993; Payseur et al. 2004; Macholán et al. 2007). Because of this geographic repeatability and the fact that a model organism is involved, the *musculus–domesticus* HZ can play a central role in understanding the nature of barriers to gene flow between taxa in general and the action and fate of speciation genes on sex chromosomes in particular.

Several recent findings have suggested that introgression patterns may be quite complex in the house mouse. First, a survey of the distribution of five autosomal and two sex-linked diagnostic markers across the Czech and Slovak Republics showed an unexpectedly gradual transition of the Y chromosome across the area of north-eastern Bavaria and western Czechia relative to other loci (Munclinger et al. 2002), in sharp contrast to results from other parts of the zone (Vanlerberghe et al. 1986, 1988b; Prager et al. 1997; Dod et al. 2005). A subsequent study revealed an introgression of the *musculus* Y chromosome up to 22 km into *domesticus* territory, covering a triangular area of about 330 km^2 and accompanied by a perturbation of the census sex ratio (Macholán et al. 2008). It was proposed that the unusual pattern has resulted from genetic conflict between loci on the X and Y chromosomes (and potentially autosomes) and that clines for genes involved in sex-ratio distortion have escaped from the zone centre, causing a decay in the barrier to gene flow between *M. m. musculus* and *M. m. domesticus*. Second, for several autosomal and X-chromosome loci a far reaching introgression of *domesticus* alleles into the *musculus* range has been reported from two

German transects (Payseur et al. 2004; Teeter et al. 2010) and Vošlajerová Bímová et al. (unpubl. ms.) found *domesticus* alleles more than 1700 km behind the centre of the Czech portion of the zone.

The mouse X chromosome is acrocentric, allowing simple description of chromosome regions as proximal or distal with respect to the centromere. To date, two surveys of X chromosome introgression heterogeneity across the European house mouse HZ (HMHZ) have been carried out. Working from a one-dimensional (1D) transect of samples that straddles this zone in Bavaria, to the south of the Czech portion, Payseur et al. (2004) fitted 1D geographic stepped (six-parameter) clines to data for 13 loci, identifying a central X-chromosome region of reduced introgression, and also proposing the X chromosome inactivation locus *Xist* as marking an adaptive introgression into *musculus* territory. Teeter et al. (2010), working with the same Bavarian dataset, and an additional 1D transect sampling across the zone in Saxony (to the north of the Czech zone), found the likelihood surfaces for the six-parameter geographic cline model impracticably complex, instead fitting 1D geographic sigmoid (two-parameter) clines to 41 SNPs (but only three X markers, a subset of Payseur et al. 2004: *Emd*, *Pola1* and *Xist*) drawn from both datasets. However, in recognition of the difficulty introduced by geographic complexity the authors also used a nongeographic introgression analysis (the genome clines approach, as implemented in the INTROGRESS program, Gompert and Buerkle 2010) that, as with Barton's concordance, assesses introgression relative to the hybrid index rather than geography. They concluded that "the differences between transects raise the possibility that there may not be a single genetic architecture of isolation between these species" (Teeter et al. 2010: p. 481), although two out of the three X markers did not show any difference between transects, but noted that "it is possible that stochastic variation, differences in sampling between transects, or a combination, could have contributed to these differences" (Teeter et al. 2010: p. 481). Clearly, to make further progress, geographic stochastic effects must be de-confounded from deterministic introgression patterns.

In this study we compare X chromosome introgression patterns across a 24 locus superset of the Payseur et al. (2004) study, on a sample of more than 2870 mice from a geographically intensively sampled 145 × 50 km rectangular section of the Czech portion of the HMHZ, which lies between the previously mentioned Saxony and Bavarian linear transects. Given the anticipated complexity of introgression, and the possibility of sex-specific effects, we compare results from analyses both with and without geographic information, and both with and without separating sexes. In our geographic analyses, we attempt to capture the full two-dimensional nature of introgression patterns and classify loci accordingly. We then analyze the same loci using both nongeographic introgression approaches: Barton's concordance
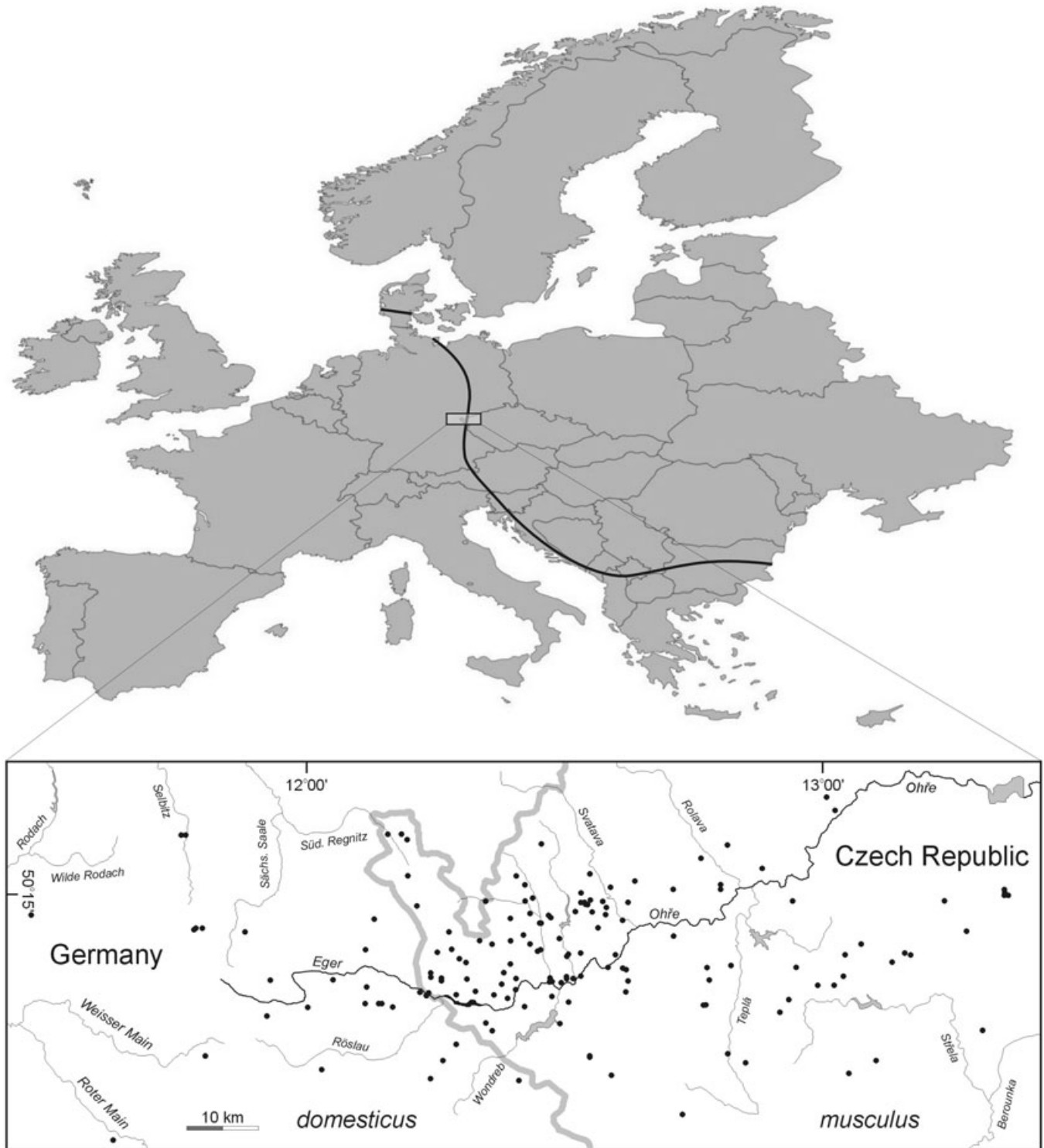
Figure 1. The course of the *M. m. musculus*/*M. m. domesticus* hybrid zone in Europe (shown as bold line) and the location of the study area as shaded rectangle. Below, the detail of the study area with sampling sites indicated.

and genomic clines. In this way, we both assess the utility of the nongeographic approaches and gain insight into the nature of the HMHZ in general and the role of X chromosome loci in particular.

Materials and Methods
SAMPLING
In total, 2873 mice (1351 males and 1522 females) were live-trapped at 166 sites scattered across a 145 × 50 km transect
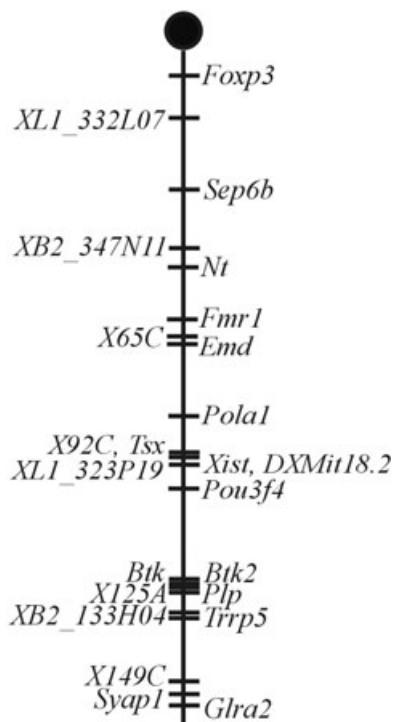
Figure 2. Markers on the X chromosome analyzed in this study. Insertions are displayed on the left side and SNPs on the right side.

stretched from western Bohemia (Czech Republic) to northeastern Bavaria (Germany) (Fig. 1 and Table S1). Most of the captured mice were euthanized and dissected directly in a field laboratory. The remaining mice were transported to a breeding facility at Department of Population Biology in Studenec and dissected after laboratory-based experiments. DNA was extracted from spleen and/or tail using DNeasy® 96 Tissue Kit (Qiagen GmbH, Hilden, Germany) following the manufacturer's instructions.

MARKERS

Altogether, 24 diagnostic X-chromosome markers were scored. Their positions on the physical map (in Mb) according to NCBI Build 37.1 are given, along with numbers of individuals scored for particular markers, in Table S2 (see also Fig. 2). Eleven of the markers were transposable elements, nine SINEs (B1 and B2) and two LINEs. Primers and reaction conditions for *Btk* and *Tsx* were described in Munclinger et al. (2003); for *Syap1*, see Macholán et al. (2007). Primer sequences, expected product sizes, and annealing temperatures for remaining insertions are given in Table S2.

Templates were amplified using the Mastercycler®ep (Eppendorf, Hamburg, Germany) in 10-μl volume under the following conditions: 2 mM MgCl$_2$, 0.2 μM primers (0.3 μM for *X65C*), 100 μM dNTPs and 0.5U Taq polymerase (Invitrogen Corp., Carlsbad, CA). PCR conditions were: a 3-min activation step at

94°C was followed by 35 cycles of 45 s at 94°C, 45 s at 60–65°C (Table S3), and 45 s at 72°C; a final elongation step lasted 5 min at 72°C. For *XL1_332L07*, we applied Multiplex PCR Kit (Qiagen; 0.2 μM concentration of primers) in 5-μl volume under conditions described in the producer's protocol; *XL1_323P19* marker was amplified in the presence of an (NH$_4$)$_2$SO$_4$ reaction buffer (Invitrogen). PCR products were run on 2% agarose gels and visualized either by Ethidium Bromide or GoldView™ Nucleic Acid Stain (SBS, Beijing, China). Genotypes were deduced from the size of the amplified DNA fragments (longer in the presence of an insertion). To verify PCR reactions, we analyzed samples from wild-derived inbred strains STRA and BUSNA representing both subspecies (Piálek et al. 2008) and F1 hybrid females as a control. The remaining 13 markers were SNPs used by Payseur et al. (2004) and analyzed as described in Dufková et al. (2011).

ANALYSES

To control for potential nonindependence of alleles due to relatedness and deviation from Hardy–Weinberg equilibrium, the effective number of alleles sampled at each locality was estimated following the procedure described in Macholán et al. (2008).

Geographic patterns in two dimensions

To allow for the potentially complicated spatial patterns outlined in the introduction, a spatially explicit Bayesian model-based clustering algorithm GENELAND 3.1.4 (Guillot et al. 2005) was employed. Using a Markov chain Monte Carlo (MCMC) approach, the program combines individual genotypic data and geographic coordinates of sampled individuals to infer and locate genetic discontinuities between populations in space. The number of clusters (*K*) can be inferred during the MCMC runs, but our analyses are simplified by fixing *K* at 2, that is, the number of subspecies present in the field region. Two independent MCMC runs were carried out for each locus with 10^7 iterations and saving every 100th iteration (i.e., 10^5 iterations saved in total). The spatial model was used with no uncertainty ascribed to spatial coordinates; the uncorrelated allele frequency model was applied and no null alleles were allowed. The post-MCMC processing phase included exclusion of the first 2000 iterations as burn-in and the computation of posterior probabilities of population membership for each individual and each pixel; the resolution of the final maps was set to 1400 pixels along the *x*-axis and 40 pixels along the *y*-axis, corresponding roughly to a pixel size of 100 m side. Results across corresponding runs were then visually checked for consistency; no inconsistent results were obtained.

Nongeographic analyses of introgression: Barton's concordance and genomic clines

We used two similar approaches based on plotting clines against genome distance: Barton's concordance analysis (Szymura and

Barton 1986) and the genomic clines analysis of Gompert and Buerkle (2009). The original description of Barton's approach is extremely clear, but here we briefly re-express it in terms of hybrid indices, just to clarify the similarity between the two approaches. For diagnostic loci (as in the current case) the proportion of *musculus* alleles in any set of alleles can be calculated, and called the hybrid index (*HI*) for that set of alleles. A hybrid index can thus be calculated for the set of alleles belonging to an individual (producing an individual *HI*). This set of alleles can be further subdivided: some of the individual's alleles come from locus 1, some from locus 2, and so on. A hybrid index can be calculated for each subset, producing individual × locus *HI*s. Barton's first key insight is that if all loci introgress in a similar fashion, then *HI* expectations do not change over loci and so the expectation for an individual × locus *HI* is the same as for an individual's *HI*. That is, plotting all individual's individual × locus *HI*s against their individual *HI*s, the expectation for equal introgression over loci lies on the diagonal (i.e., $x = y$). Barton's second key insight was to formulate a simple two-parameter model of deviations from this equal-introgression expectation, which can be expressed as

$$HI_{loc} = HI + H_e[\alpha + \beta(2HI - 1)].$$

That is, the locus expectation can deviate from $x = y$ as a function of the expected heterozygosity $H_e = 2HI(1 - HI)$. Parameter α (directionality) shifts introgression toward one or the other genome, relative to the equal-introgression expectation. Parameter β (abruptness) allows the introgressive change at the locus to be more or less abrupt than the equal-introgression expectation. As hybrid indices and expected heterozygosity can be calculated irrespective of ploidy, the approach can be applied equally to haploid, haplo–diploid, and diploid loci. The simplicity of the model allows the parameters to be estimated in the likelihood inference framework, and the significance of deviations from the average introgression pattern, or differences between introgression patterns in males and females, to be assessed directly using likelihood ratio tests. The concordance analysis was carried out using a Mathematica (Wolfram 1992) routine developed for ANALYSE 2.0beta by one of the authors (SJEB).

Gompert and Burkle (2009) have recently suggested a similar approach, calling the clines thus produced genomic clines. In this method clines are estimated using multinomial regression of the observed individual genotypes on the hybrid index. Neutral introgression expectations are simulated using permutations, accounting for stochastic variation among loci due to finite sampling. The method is developed to discriminate among a number of selection regimes (Gompert and Buerkle 2009; Nolte et al. 2009): deviations from neutrality are interpreted as either underdominance (heterozygotes less common than expected) or overdomi-

nance (heterozygotes more common than expected) or directional selection against one allele (clines for both homozygotes and heterozygotes shifted). Sharp homozygote and heterozygote clines are interpreted as an indication of epistatic selection between loci and the simple underdominance pattern is suggested to be consistent with the action of BDM incompatibilities (Gompert and Buerkle 2009). Genomic clines were estimated using the R program INTROGRESS 1.2.1 (R Development Core Team 2007; Gompert and Buerkle 2010) with 1000–5000 permutation steps. Significance levels ($\alpha = 0.05$) for genotype-specific deviations tests were adjusted using the false discovery rate procedure according to Benjamini and Hochberg (1995). Because females are diploid and males haploid for the X markers considered we analyzed the data separately for each sex. Although this leads to a decrease of sample size, the numbers of analyzed specimens were still rather high (794–1519, 1168 on average in females; 850–1347, average 1097 in males). Nevertheless, to avoid the risk of weak results due to insufficient sample size we also pooled females and males, treating males as diploid homozygotes. This procedure substantially reduces heterozygosity but as both expected and observed probabilities of heterozygote genotypes are estimated from the same data this reduction does not affect the expectation of the final results, only our confidence in them.

The essential difference between the two nongeographic approaches is that Barton's concordance inspects introgression patterns for allele frequencies, whereas genome clines inspect introgression patterns for diplotype frequencies—homozygotes from each source population and their heterozygotes. At Hardy–Weinberg equilibrium (HWE) the allele frequencies are sufficient information to predict the diplotype frequencies, and so we would not expect to gain any additional information from the genome clines approach at HWE relative to Barton's concordance. We therefore carry out an analysis of deviation from HWE that follows directly from the concordance approach. Recall (above) that for an individual with hybrid index *HI* we can calculate its expected heterozygosity $H_e = 2HI(1 - HI)$. It is straightforward to compare this with the individual's "observed" heterozygosity—the proportion of its loci that are in the heterozygous state. If the individual's observed numbers of homozygous and heterozygous loci are $\{N_{AA}, N_{AB}, N_{BB}\}$, the log likelihood of the individual-based heterozygote deficit F is

$$LL(F \mid N_{AA}, N_{AB}, N_{BB}) \propto N_{AA}Log(p + Fpq)$$
$$+ N_{AB}Log(2pq - F) + N_{BB}Log(q^2 + Fpq),$$
$$p = HI \quad q = (1 - HI).$$

Plotting maximum likelihood estimates (MLEs) of individual F against individual *HI* allows us to see whether hybrid individuals tend to have, for example, a greater deficit of heterozygous loci.
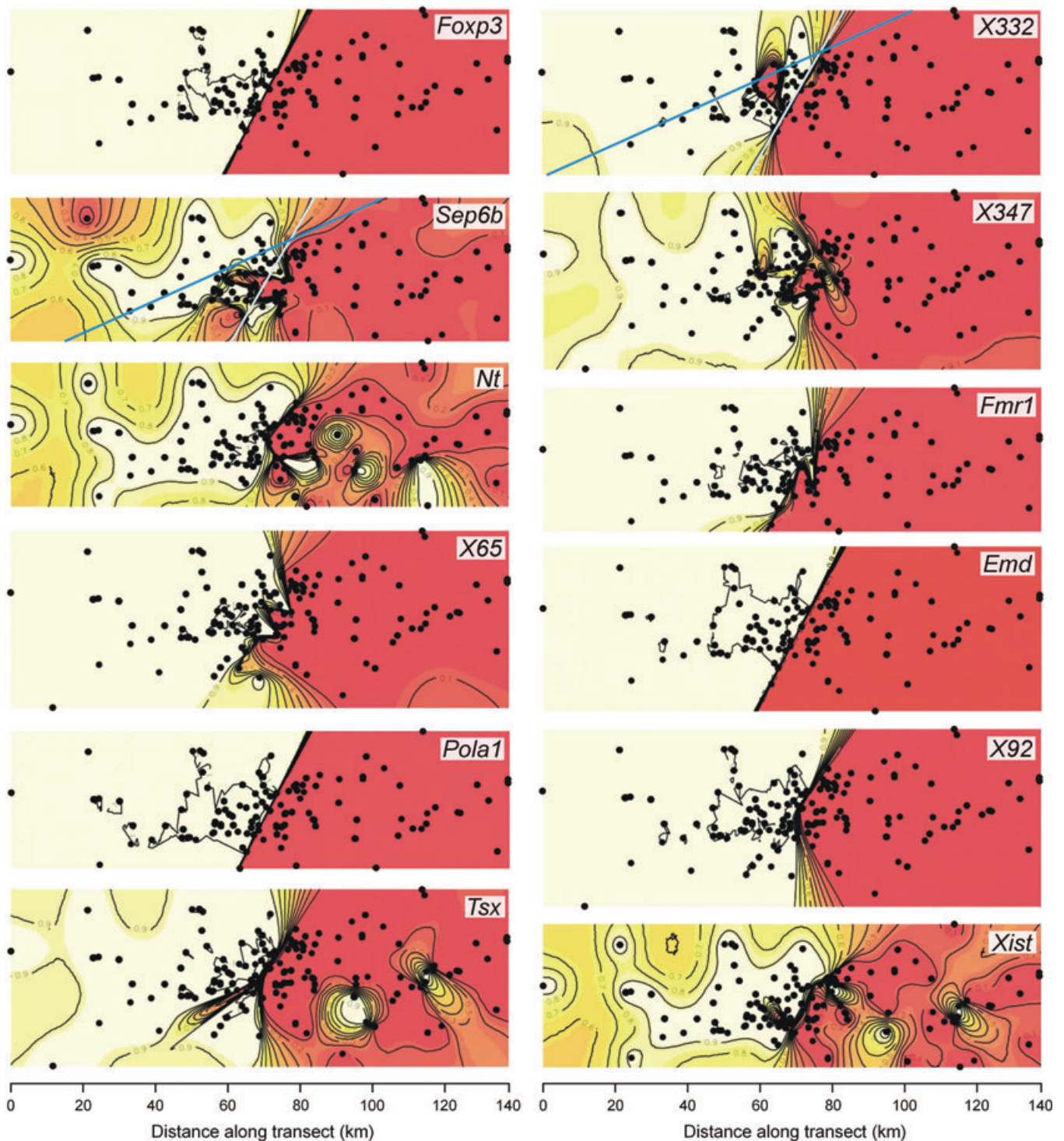
Figure 3. Maps of Geneland individual assignments to *M. m. domesticus* for all 24 loci scored. The highest membership values are in light yellow, the lowest in red. The level contours show the spatial changes in assignment values. The light blue line indicates orientation of the consensus cline based on seven autosomal loci (Macholán et al. 2008) whereas the dark blue line depicts the cline centre for the Y chromosome.

Results

GEOGRAPHIC PATTERNS IN TWO DIMENSIONS

The geographic patterns revealed by the Geneland analysis are highly heterogeneous across loci (Fig. 3). Of the 24 loci, seven

show clear and smooth transition congruent with simple 1D clinal predictions (Macholán et al 2008). These seven are indicated in Table 1 with symbol "/". A further six are passably consistent with these predictions (Table 1, symbols "(" and "/-"). Of the
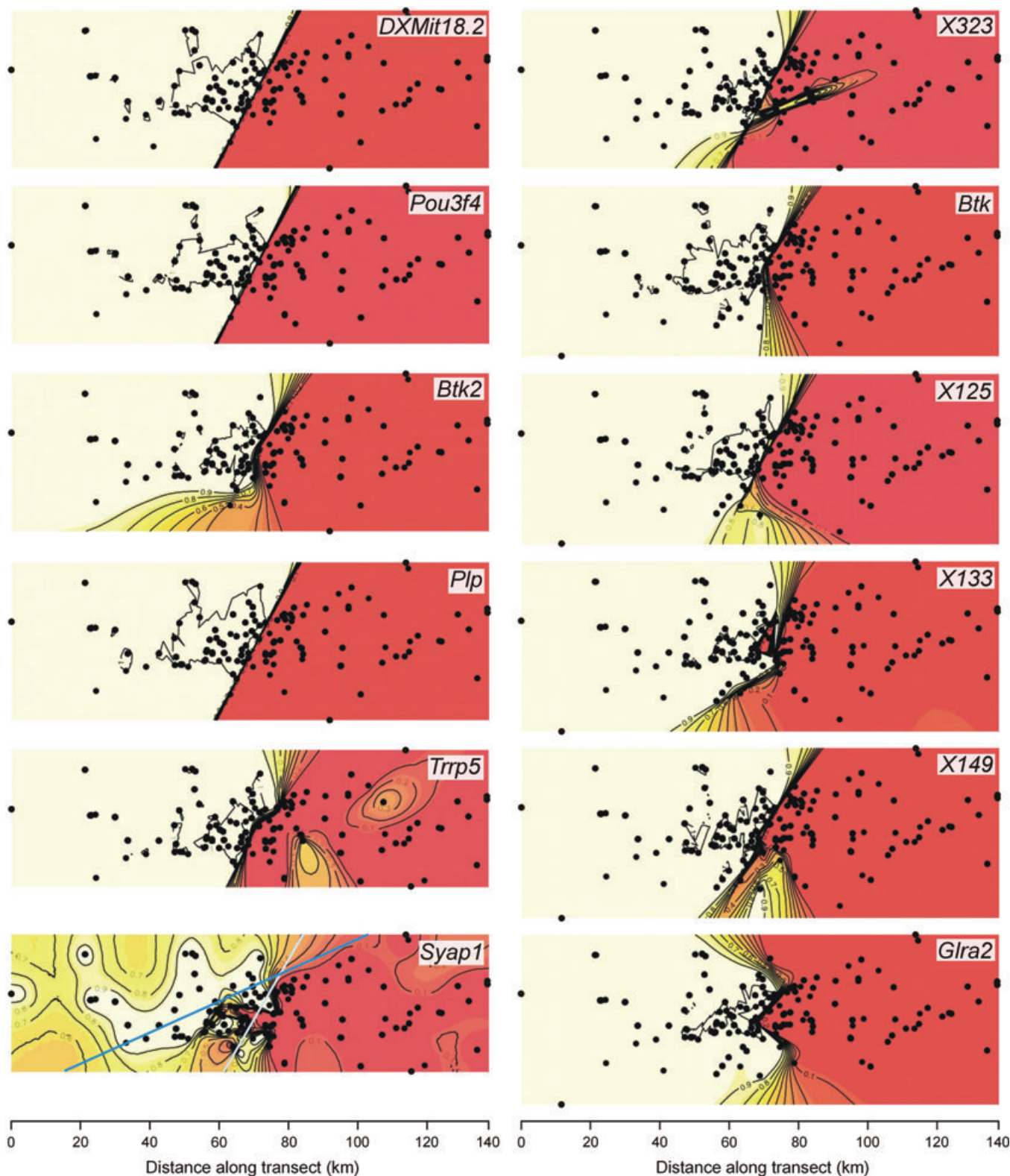
Figure 3. Continued

remaining 11 loci, five show complex interlocking patterns at the HZ centre (Table 1, symbol "§"), four display islets of *domesticus* alleles in *musculus* territory, and two show a salient of *musculus* alleles in *domesticus* territory next to an otherwise clearly defined zone centre (Table 1). Introgression of *musculus* alleles into the *domesticus* range is more or less limited to the triangular area of the introgressed *musculus* Y chromosome (indicated with blue lines in Fig. 3; cf. Macholán et al. 2008). In contrast, the islets

Table 1. Summary comparison of 2D geographic and nongeographic introgression patterns along the X chromosome. "Gap" refers to an interval between two adjacent loci (see also Table S2 for physical distances for the loci). Symbols: GENELAND – "/" clear and sharp zone centre consistent with 1D analysis; /-" clear and sharp zone centre consistent with 1D analysis, excepting one D inclusion in M territory; "(" clear and sharp zone centre, but convex, bulging toward M territory; "§" complex 2D geographic pattern at zone centre. Nongeographic introgression analyses: INTROGRESS—M>D: introgression of *musculus* alleles into *domesticus* background; M>>D: strong introgression of *musculus* alleles into *domesticus* background.

Locus	Gap (Mb)	2D Geneland	Asymmetry Concordance D	Asymmetry Concordance M	Asymmetry Introgress	Nature Concordance Gradual	Nature Concordance Abrupt	Nature Introgress
Foxp3	10	/	−0.27		M>D	−0.09		Epistasis
X332	18	Salient into D	−0.80		M>>D	−0.81		Epistasis
Sep6b	15	§	−0.61		M>>D	−1.17		Epistasis
X347	4	§	−0.21		M>D	−0.15		Epistasis
Nt	12	Islets in M		+0.63	D>>M	−0.67		
Fmr1	5	/		+0.09			+0.36	Weak epistasis
X65	0.368	§	+0.03				+0.41	Underdominance
Emd	20	/		+0.80	D>M		+1.36	Strong epistasis
Pola1	19	/		+0.36	D>M		+0.47	Strong epistasis
X92	8	(	−0.31		M>D		+0.43	Strong epistasis
Tsx	2	Islets in M		+0.56	D>>M	−0.62		
Xist	0.052	Islets in M		+0.65	D>>M	−0.90		
DXMit	0.168	/		+0.15			+0.28	Underdominance
X323	1	/−		+0.17			+0.23	Nearly neutral
Pou3f4	6	/	−0.10		M>D		+1.10	Strong epistasis
Btk	23	(	−0.30		M>D		+0.40	Epistasis
Btk2	0.020	(	−0.35		M>D		+0.66	Strong epistasis
X125	1	(		+0.13	D>M		+0.44	Epistasis
Plp	1	/	−0.21		M>D		+1.02	Strong epistasis
X133	6	Salient into D		+0.12		+0.06		Nearly neutral
Trrp5	1	Islets in M		+0.06		−0.32		Nearly neutral
X149	15	/−		+0.04			+0.60	Strong epistasis
Syap1	3	§	−0.14		M>D	+0.05		Epistasis
Glra2		§	−0.32		M>D		+0.13	Epistasis

of *domesticus* alleles in *musculus* territory occur throughout the sampled range, and their geographic locations are not correlated over loci (except for closely linked *Tsx* and *Xist*, which appear to share two islets, Fig. 3).

NONGEOGRAPHIC ANALYSES OF INTROGRESSION: BARTON'S CONCORDANCE AND GENOMIC CLINES

Simple informational considerations suggest we would not gain any additional insight, relative to Barton's concordance, by applying the genome clines approach to a system at HWE. Figure S1 shows that for the female mouse dataset, the individuals with hybrid indices lying in the central portion of the range have no significant heterozygote deficit. However, to either side of this central portion individuals do indeed show deviation from HWE

expectations, with a strong deficit in heterozygotes, potentially lending power to the genome clines approach.

Barton's concordance analysis provides MLEs and their support bounds for directionality and abruptness of introgression for each locus (Table S4). When males and females are pooled all loci differed significantly from the equal-introgression expectation, and this is also true for separate treatment of the sexes, except for locus *X133* in males (Table S4). Proximal locus *X332* (16.7 Mb; Table S2) is alone in showing significantly different introgression patterns between the sexes at the 95% level, although distal loci *X125* and *Syap1* showed sex differences significant at the 90% level. Table S4 indicates the nature of these sex differences: at *X332* there is greater introgression of *musculus* alleles onto the *domesticus* X background in males; at *Syap1* this pattern

Figure 4. Analysis of X introgression patterns in males and females along the chromosome using Barton's concordance. Nongeographic introgression relative to X chromosome hybrid index is summarized in two components: vertical axis (α, eq. 1), gauges whether introgression is *abrupt* or *gradual* relative to the X hybrid index; horizontal axis (β, eq. 1), whether introgression is asymmetric toward the *domesticus* or *musculus* genetic background. The analogous geographic introgression descriptors are shown in parentheses. Blue lines connect male X locus (α, β) maximum likelihood estimates in the order the loci occur along the chromosome. Where loci are less than 0.5 Mb apart connecting lines are bold. *Foxp3* (Gradual, *domesticus* quadrant) is the most proximal marker, *Glra2* (close above, Abrupt, *domesticus* quadrant) the most distal. Red lines connect the female introgression MLEs, again loci less than 0.5 Mb apart, in bold. Green lines connect male and female estimates that differ significantly.

is reversed; at *X125* the degree of introgression is similar across the sexes, but the transition is steeper in males.

Figure 4 plots the abruptness/directionality MLEs for each locus on two axes. Loci are joined in the order they occur along the X chromosome, one (blue) line for estimates from male samples, a second (red) line for female samples. Lines between physically close (<0.5Mb, Table 1) neighbors are bold. Green lines connect the previously mentioned MLEs that differ significantly between

sexes. The plot has four quadrants. The (Gradual, *domesticus*) introgression quadrant contains a proximal X linked wide asymmetric introgression toward *domesticus*. At the extreme of this proximal linked introgression, locus *X332*, males introgress significantly more than females, and there is a 2D geographic salient into *domesticus* territory (cf. Fig. 3, Table 1). Considering the remaining quadrants in anticlockwise order, the (Gradual, *musculus*) introgression quadrant contains all loci identified as having multiple islands of 2D introgression (cf. Fig. 3, Table 1). In contrast to the proximal X-linked introgression, loci in this quadrant come from all along the linkage ordering, the only neighboring loci being the very closely linked *Xist* and *Tsx*, separated by only 52 kb. The (Abrupt, *musculus*) introgression quadrant contains *Pola1*, previously proposed marker for a speciation gene (Payseur et al. 2004), but also a more extreme candidate neighboring it: *Emd*. Finally, the (Abrupt, *domesticus*) quadrant contains *Btk*, a locus previously used to mark the consensus centre of the HZ (Macholán et al. 2008), although *Plp* and *Pou3f4* are indicated to be even more abrupt in their introgression pattern. Of the four pairs of physically very close loci (Fig. 4, bold red and blue), two pairs show very different introgression patterns. *Emd*, the sharpest of all introgressions, and the most *musculus* biased, is proximally flanked at 368 kb by *X65*, unremarkable in bias and abruptness. *Xist*, showing 2D islands of introgression far into *musculus* territory, is distally flanked by *DXMit18.2* at 168 kb, showing unremarkable bias toward *musculus*.

The INTROGRESS approach relies more on user interpretation of graphical output than parameter estimates and formal hypothesis comparison, and Table 1 details interpretation formed independently of the results of Barton's concordance analysis. Figure 5 depicts the raw data ordered for each marker from *domesticus*-specific to *musculus*-specific alleles. The results are shown for females and males separately; for better clarity, we excluded from the graphs those individuals for which less than 75% of loci were typed ($N \geq 18$). In both sexes, there are loci characterized by very abrupt transition between the two subspecies (e.g., *Pola1*, *Pou3f4*, *Plp*), intermingled with loci displaying extensive introgression either to *M. m. musculus* (*Nt*, *Tsx*, *Xist*, *Trrp5*) or to *M. m. domesticus* (*X332*, *Sep6b*, to a lesser extent *X347*, *Syap1*). When averaged across all loci, the gradients of *musculus* alleles are apparently steeper on the *musculus* side (right diagrams for each sex in Fig. 5). More importantly, the positions of the two gradients are clearly different: as it appears, the proportion of *domesticus* alleles is higher in males than in females.

Results of the genomic cline analysis revealed significant deviations from neutral expectations in all loci for females, even after correction for false positive discovery, whereas in males only five loci (*X332*, *Sep6b*, *Nt*, *Tsx*, *Emd*) revealed significant differences. This sex difference is likely due to the previously mentioned deviations from HWE that, naturally, only apply to the female part

of the dataset. Examples of 12 female genomic clines displaying the most conspicuous deviations are shown in Figure 6A, clines for the remaining loci are presented along with all male clines in Figure S2. (For the purposes of comparison, Figure 6B shows the locus by locus Barton's concordance analyses for the same 12 loci.) Homozygote clines for males and females are mostly very similar; when different, the female clines are usually steeper, with the only exception *Emd* where the male cline is apparently more abrupt than the female one (cf. Figs. 6 and S1). A closer inspection of these figures shows most clines are shifted either to the *musculus* (10 loci) or *domesticus* side (12 loci), some of them markedly (*Nt*, *Tsx* to *musculus*; *X332*, *Sep6b* to *domesticus*), and these shifts are summarized in Table 1. A heterozygote deficiency with a cline lying outside the 95% confidence interval of the null model was found in *X65* and *DXMit18.2* only. This pattern could suggest underdominance (see Material and Methods) but more likely this is caused by BDM incompatibilities. At some loci (e.g., *Emd*, *Pola1*, *X92*, *Pou3f4*, *Btk2*, *Plp*) both homozygote and heterozygote clines are quite steep indicating strong epistatic interactions among loci but most of them are substantially shifted to one side suggesting more complex selection regimes and/or geographic situations. Finally, *Trrp5* and *Syap1* display homozygote genomic clines wider than expected without any apparent shift to either side (see Table 1).

Discussion
GEOGRAPHIC CLINES AND SELECTION AGAINST HYBRIDS IN THE MOUSE HZ

During the last few decades applications of cline theory to empirical data have been accumulating with an increasing pace (Szymura and Barton 1986, 1991; Porter et al. 1997; Kruuk et al. 1999; Bridle et al. 2001; Phillips et al. 2004; Raufaste et al. 2005; Yanchukov et al. 2006; Macholán et al. 2007; Teeter et al. 2008). In the house mouse, detailed analyses using rigorous statistical treatment have been carried out in Denmark (Raufaste et al. 2005; Dod et al. 2005), southern Bavaria (Tucker et al. 1992; Payseur et al. 2004, Teeter et al. 2008, 2010) and the Czech Republic (Božíková et al. 2005; Macholán et al. 2007). It was shown that the width of the mouse HZ in northern Europe (Denmark), which is suggested to be much younger than in southern parts of the continent (Cucchi et al. 2005), is not significantly different from the zone width in the Czech Republic (cf. Raufaste et al. 2004 and Macholán et al. 2007). Moreover, X-linked markers have been shown to be, in general, characterized by narrower clines than clines for autosomal loci (Tucker et al. 1992; Macholán et al. 2007), in agreement with predictions of the large X-effect theory of Haldane's rule (Haldane 1922; Orr 1997; Coyne and Orr 2004). Because theory predicts that narrower clines indicate areas of the genome under strong selection and hence potentially
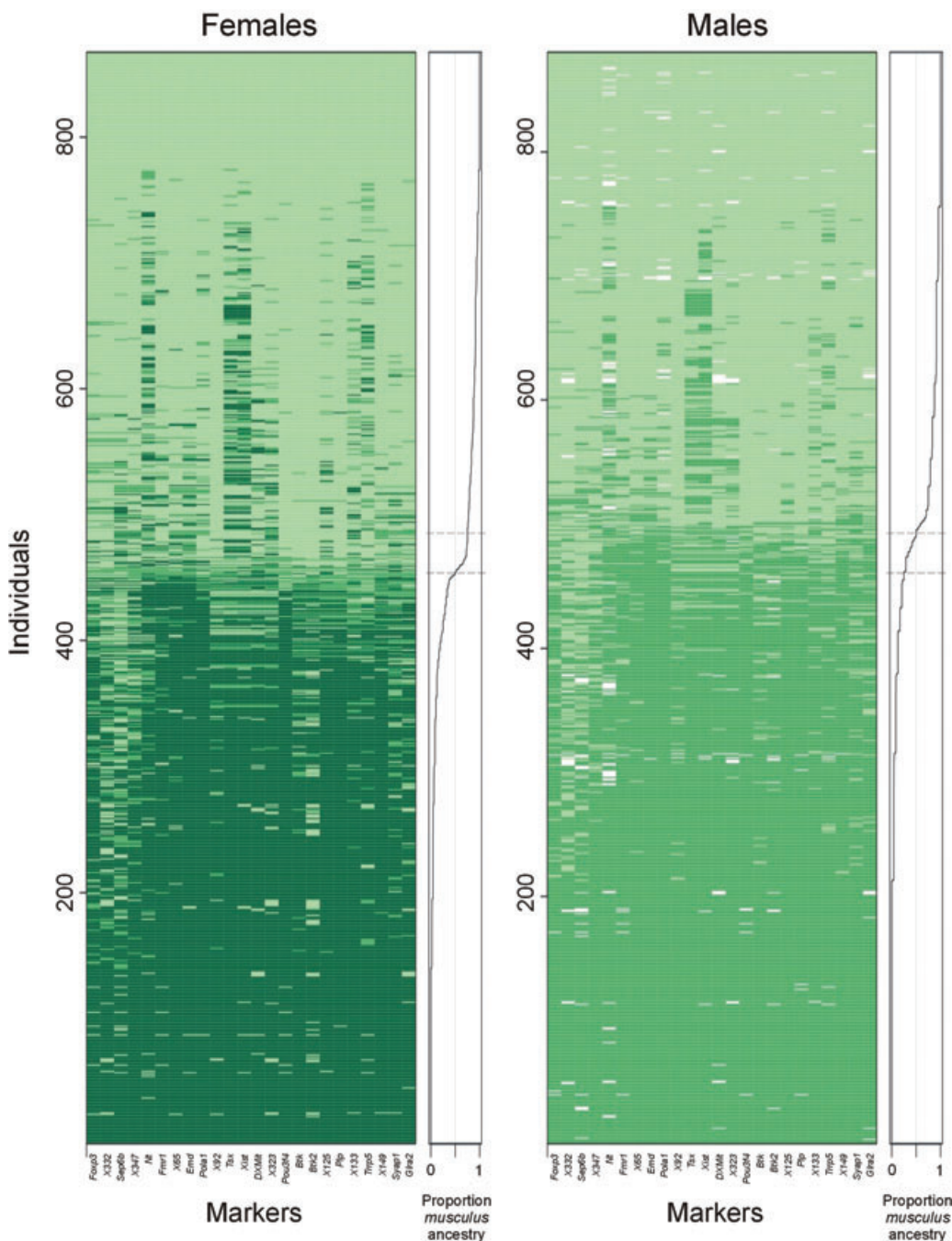
Figure 5. Genotypes of female (left) and male (right) mice across the Czech-Bavarian hybrid zone. Each rectangle corresponds to an individual's genotype at a given locus. Colors are arranged from darker green, indicating *M. m. domesticus* homozygotes, to light green, indicating *M. m. musculus*; in females, heterozygote genotypes are represented by intermediate green blocks. White blocks indicate missing data. The plots to the right of each panel depict the fraction of the genome of *M. m. musculus* ancestry. Horizontal dashed lines indicate equal proportion of both genomes. Note the higher number of males carrying predominantly *domesticus* genome (see text for explanation).

involved in reproductive isolation between taxa (Slatkin 1973, 1975; Nagylaki 1975, 1976; Endler 1977), if a relatively large set of molecular markers with known positions is analyzed along the transect, genomic regions under such selection might be easily detected (Rieseberg et al. 1999; Payseur et al. 2004; Buerkle and Lexer 2008; Gompert and Buerkle 2009; Payseur 2010). Advances in technology have now made such studies of many markers in a single linkage group practicable.
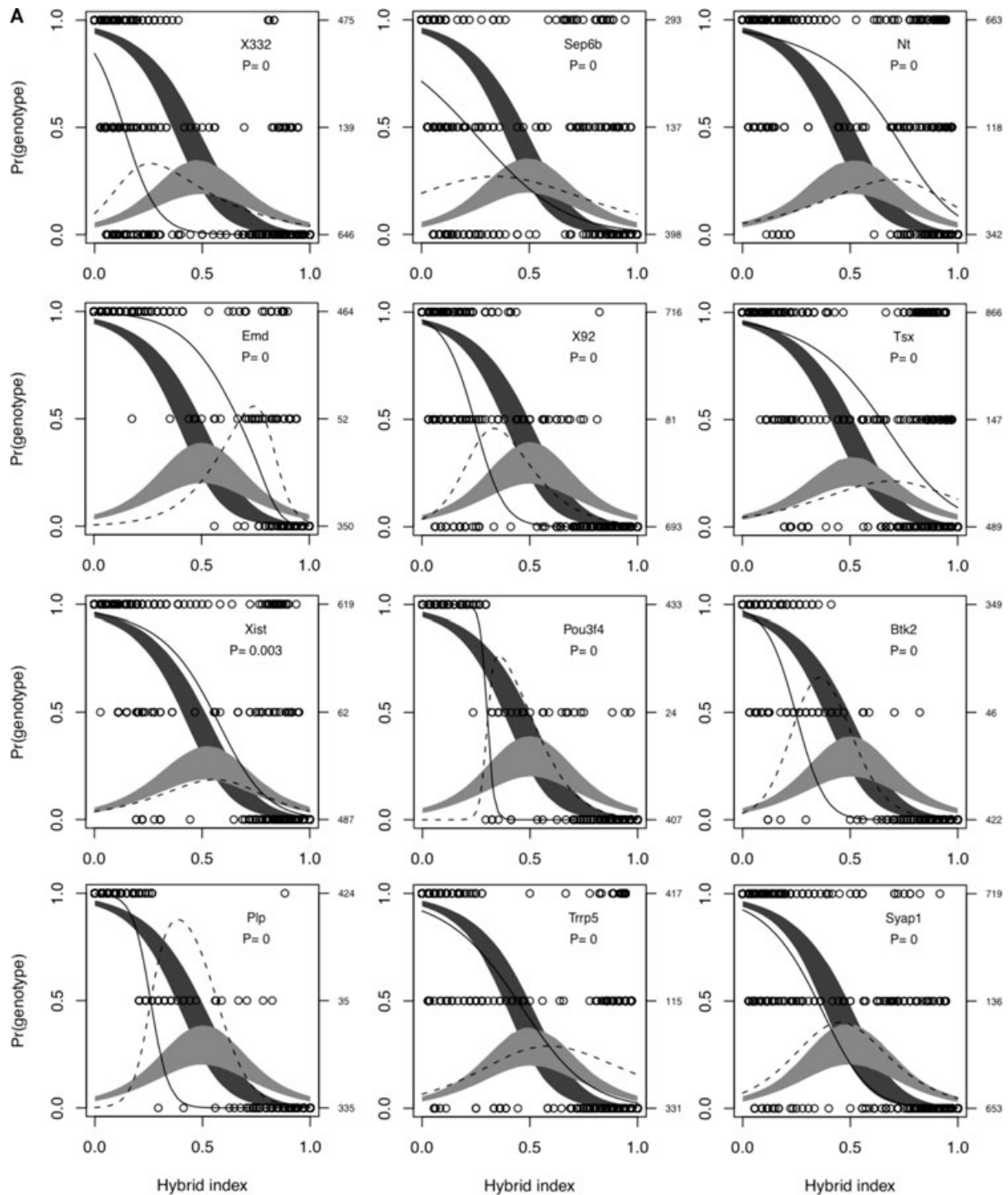
Figure 6. (A) Genomic clines for the 12 loci. Solid and dashed lines represent estimated clines based on the observed *domesticus* homozygotes and heterozygotes, respectively. The 95% confidence intervals for expected homozygote and heterozygote clines given neutral introgression are shown as dark gray and light gray areas, respectively. The *P*-value for the test of departure from neutral expectations is given on each panel. Circles show the raw data with counts of each genotype on the right vertical axis; only female data are presented here. The hybrid index is expressed as a proportion of *musculus* alleles. To minimize potential bias caused by undue influence of the seven "outlier" loci, the analyses were carried out separately for each of these "outliers"; however, the differences from a single analysis of the whole dataset where all these loci were included are only negligible. (B) Barton's concordance for the same 12 loci. For each individual per locus hybrid index is fitted against hybrid index calculated over all X loci: the equal-introgression-over loci-expectation lies on the diagonal. Per locus fits can deviate from this toward either genetic background, (parameter α, eq. 1), and by having more or less abrupt change (parameter β, eq. 1). Blue and Red support envelopes enclose the maximum likelihood fits for males and females, respectively.
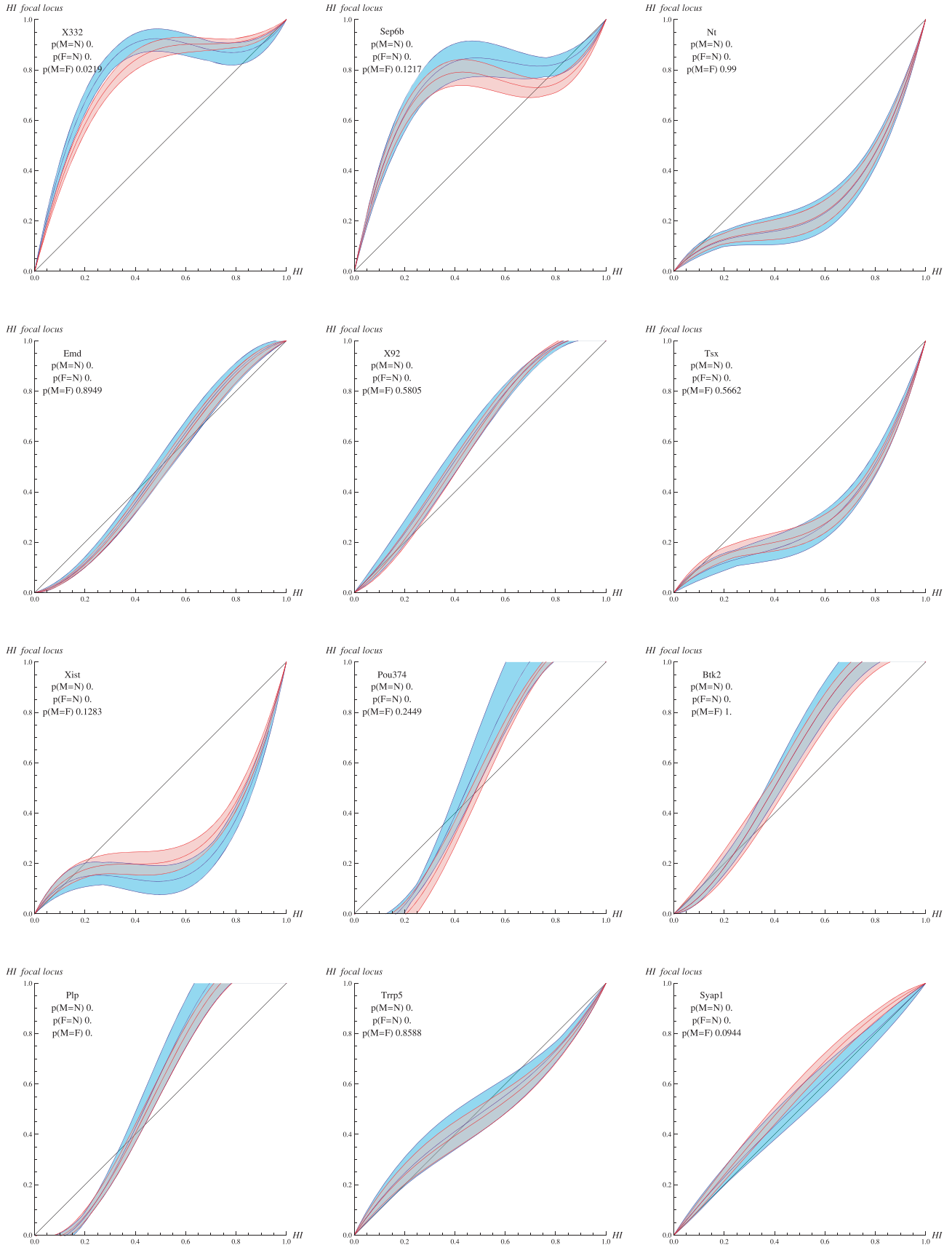
B

Figure 6. Continued

TWO-DIMENSIONAL GEOGRAPHIC PATTERNS OF X CHROMOSOME VARIATION IN THE CZECH HMHZ

The geographic cline approach to identifying genomic regions affected by selection against hybrids is appealing in its simplicity but assumes a simple history of admixture and a smooth transition of a trait from one taxon to the other. Our 2D Geneland analyses of 24 loci (Fig. 3, Table 1) indicate neither is the case for the majority of the X chromosome loci studied. Islands of *domesticus* alleles in *musculus* territory (four loci) are consistent with a history of westward zone movement, while salients into *domesticus* territory (two loci) and a zone centre line bulging toward that territory (four loci) suggest this movement is ongoing.

This does not mean previous models of a straight zone centre line and smooth clinal change in this part of the HZ were inappropriate. The match of seven loci to precisely this description, including even the orientation of that centre line (cf. Macholán et al. 2008), is striking, given the Geneland analysis has no prior information about the zone other than $K = 2$. (These loci include *Btk*, used previously to model the zone centre position, Macholán et al. (2008).) Rather, we feel the complex spatial patterns on the X chromosome reflect a very active role of the chromosome at the subspecies boundary, perhaps in particular in the Czech region where there has been the Y chromosome invasion. The five loci with complicated "boiling" patterns at the zone centre (*Sep6b*, *X347*, *X65*, *Syap1*, *Glra2*) illustrate this view. At the risk of overextending this analogy: if the tension zone is thought of as the boundary between two fluids, independent markers might escape from one to the other without disturbing their boundary—akin to evaporation. In contrast, the escape of markers tied up in epistatic associations will require stronger selective forces ($\sim$ higher temperatures), and such escapes may well disturb the tension zone boundary—akin to boiling.

THE UTILITY OF NONGEOGRAPHIC INTROGRESSION ANALYSES: BARTON'S CONCORDANCE AND GENOMIC CLINES

Barton's concordance analysis not only reflects the 2D geographic patterns just described, but allows us to refine our understanding of those patterns (Fig. 4). The geographic island introgression loci are also genomic island introgression loci (four loci, [Gradual, *musculus*] quadrant). The X locus forming a large geographic salient into *domesticus* territory, in a region where the Y has already invaded, is the most extreme of a four-locus set of (Gradual, *domesticus*) quadrant introgressors, consistent with the wide wave of advance of an advantaged allele, carrying flanking hitch-hikers. In Figure 7, we show the geographic hitch-hiking component of this introgression by plotting the uninterrupted spans of *musculus* X markers around this locus, dispersing into *domesticus* territory and being reduced by recombination.

The two most extreme loci in the (Abrupt, *domesticus*) quadrant are *Pola1* and its proximal neighbor *Emd*. Figure 4 of Payseur et al. (2004) shows these two loci as defining a clear reduced X introgression region in the Bavarian transect, with the *Pola1* geographic cline width estimate slightly more extreme than *Emd*, whereas Barton's concordance analysis (applied to the Czech zone, Fig. 4) shifts this emphasis very much toward *Emd*. Following Payseur et al.'s candidate speciation gene approach, we could seek to characterize genes of interest around *Emd*, rather than *Pola1*. It is perhaps worth noting that in NCBI build 37 of the mouse genome, of a total five gene annotations for X linked paternal imprinting, two are on either side of *Emd*, 0.4 cM apart (*Mecp2*, *Emd*, *G6pdx*, 71.27, 71.51, 71.73 Mb respectively). Dysfunction in the protein product of *Mecp2* had been shown to induce changes in the expression levels of thousands of genes, and the majority of these genes ($\sim$85%) appeared to be activated (Chahrour et al. 2008), making it a fine candidate for speciation through change in (perhaps sex-specific) gene expression levels. In the (Abrupt, *domesticus*) quadrant there are no clear outliers, but three neighboring pairs of markers *Btk2*, *Btk* (previously used to mark the zone centre), *Pou3f4*, *X92*, and distal *Glra2*, *Syap1*, this last being the only gene where concordance MLEs for the sexes lie in different quadrants.

Where asymmetry of introgression is clear in the genomic clines approach, it is always consistent with Barton's concordance. It is much more difficult to rank abruptness of introgression from the genomic clines output, but perhaps a future version of INTROGRESS will make estimates equivalent to Barton's (α, β) available to the user. We suggest the difficulty in interpreting genomic clines output lies not with the approach itself, which after all is in essence the same as Barton's concordance. Rather, it lies in the inference framework chosen. In the current case, it seems clear that estimating parameters that relate to an explicit and understandable model (likelihood-based inference) is more practical than demonstrating deviation from a null, and providing users with a guide to what certain idealized deviations look like. The potential advantage of the genome clines approach above Barton's concordance is identification of asymmetric patterns in the distribution of heterozygotes. INTROGRESS fits unimodal distributions to the heterozygote data for each locus. These will be constrained on either side by the peaks of heterozygote deficit indicated in Figure S1. However, as with abruptness of introgression, relative effects for the heterozygotes at different loci cannot easily be ranked from the INTROGRESS output.

CONSISTENCY OF ARCHITECTURE OF REPRODUCTIVE ISOLATION: IMPLICATIONS FOR PATTERN AND PROCESS

Teeter et al. (2010) suggest the genomic architecture of reproductive isolation might dramatically differ between parts of the
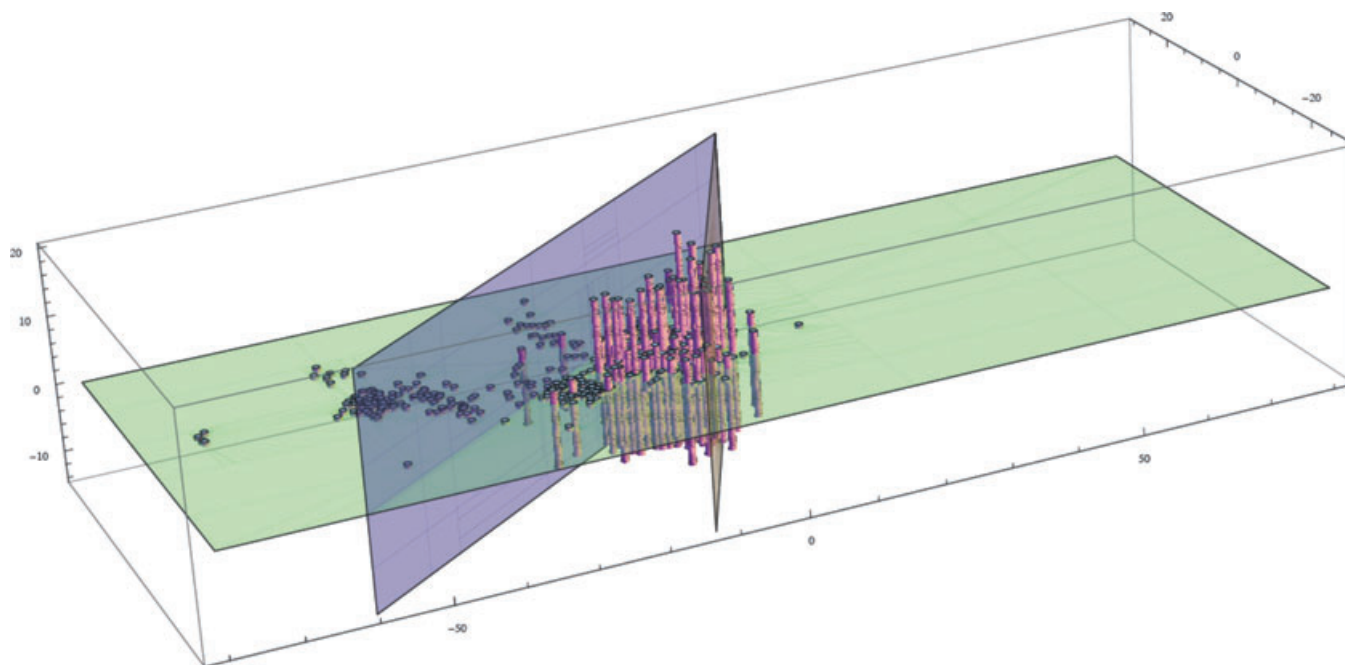
Figure 7. The distribution of *musculus* flanking regions around a presumed introgressing factor. An invading factor in the proximal X chromosome is hypothesized at map position 14 cM between *Sep6b* and *X332*. On the vertical axis is the map distance on the chromosome while the horizontal axes indicate geographic position. The green plane shows the map position of the putative introgressing factor; the gray and purple planes show the consensus centre and the centre of the cline for the Y chromosome, respectively (see Macholán et al. 2008 for details). The red columns represent the size of contiguous *musculus* blocks that occur on predominantly *domesticus* background; the shortest columns show where such X chromosomes are sampled without any introgression. It is apparent that the average block size decreases with distance from the consensus centre: large blocks are carried with the introgressing factor, but their flanks are reduced as it recombines onto the *domesticus* background. Note: for clarity, geographic positions of the blocks are scattered within 1 km of their sampling location avoiding superposition of X chromosomes sampled from the same locality.

HMHZ. Although this may be true, our analysis of the Czech zone lying between their two focal transects finds the same reduced central-X reduced introgression region as Payseur et al. (2004) found for one of their transects, and the loci involved, *Emd* and *Pola1*, are among the loci they listed as not differing significantly between their transects. Thus, over the three transects, with respect to these loci, there is some evidence of common architecture of reproductive isolation.

The X locus Teeter et al. (2010) suggest differs across transects is *Pou3f4*, and indeed our analysis shows very abrupt introgression (*Pou3f4* is the second most extreme after *Emd*) whereas Payseur et al. (2004) show relatively gradual change for this locus. This might be taken as evidence for diversity in architecture of reproductive isolation. We feel it is more likely, however, that Payseur et al.'s result for *Pou3f4* suffers from the same uncertainty demonstrated for *Btk2*. Our results in 2D and Barton's concordance clearly suggest *Btk2* introgression is abrupt whereas Payseur et al.'s (2004) Fig. 4 shows a gradual change (wide cline) for this locus, but in their erratum stating "a few mice had been assigned to the wrong localities" (http://eebweb. arizona.edu/nachman/attachments/erratum_figure4.pdf) change

at *Btk2* becomes almost as abrupt as at *Pola1*. This extreme sensitivity of estimates to small details of geographic sampling from 1D transects (an issue explored by Dufková et al., 2011) is not taken into account in the error bars shown in either version of Payseur et al.'s (2004) Figure 4 and will also affect genomic clines, or for that matter concordance estimates, made using the same data. Teeter et al. (2010) did not include the new speciation gene candidate *Btk2* in their study. Thus, over the three transects, and considering *Emd, Pola1, Btk2,* and *Pou3f4* we conclude there is some evidence of common architecture of reproductive isolation, and no reliable evidence to the contrary.

Payseur et al. (2004) identified the X-inactive transcript gene (*Xist*) as an outlier of introgression into *musculus* territory and genetic background. Their use of a stepped (as opposed to sigmoid) 1D geographic cline model was necessary to draw out the *Xist* pattern, but the linear transect sampling was not sufficient for further clarification. Here, *Xist* is seen to be one of four loci that form geographic and genomic islands of *domesticus* introgression within *musculus* territory (the *Xist* genomic island includes *Tsx*, 52 kb distant). Drawing loci and linear transect samples at random, Teeter et al. (2010) risk sometimes hitting such an island,

and sometimes not. This may well explain a large component of their differential patterns when comparing the Bavarian and Saxony transects. Payseur et al. (2004) suggest *Xist* is an example of adaptive introgression, but the island pattern of introgression observed here leads us to modify that conclusion: *domesticus Xist* is not striking out into *musculus* territory, rather, it may be sufficiently advantaged to hold on for a time in local islands as the westward advance of the HZ incorporated them into *musculus* territory. The clearest example of classic introgression of an advantaged gene in the current study is *X332* on the proximal X chromosome. If *X332* introgression is associated with the westward Y chromosome invasion in the Czech region (Macholán et al. 2008), then the proximal X introgression pattern we observe may be unique among the Bavarian, Czech and Saxony transects, and provide further clues to help explain how a Y chromosome can invade across a HZ in apparent disregard of Haldane's rule. Given the distortion of the sex ratio in the area where the Y, and to a lesser extent *X332*, have introgressed, Macholán et al. (2008) suggested a role for genetic conflict. It is tantalizing that male introgression of *X332* is significantly greater than female, and perhaps fitting that in all quadrants of Barton's concordance analysis, the heterogametic sex in a HZ has the more extreme X chromosome introgression patterns.

We pointed out that if universally advantaged genes are intermingled in tight linkage with speciation genes, they might still be in the process of escaping this background even many generations after secondary contact, a prediction being effective recombination rates bearing little relation to map distance. The sharply contrasting patterns of introgression for two of four pairs of physically very close loci in our study are consistent with this prediction, and this is worth considering because of the implications for genome sampling strategy in HZ studies (even coverage of candidate genes?). In the current study, effects at clear outlier loci (*Xist, Emd*) would not have been detected using markers even less than 0.5 Mb distant (*X65, DXMit18.2*), suggesting even coverage must be dense, and thus widening the circumstances where a candidate gene approach is preferable.

ACKNOWLEDGMENTS

LITERATURE CITED

Abe, K., H. Noguchi, K. Tagawa, M. Yuzuriha, A. Toyoda, T. Kojima, K. Ezawa, N. Saitou, M. Hattori, Y. Sakaki, et al. 2004. Contribution of Asian mouse subspecies *Mus musculus molossinus* to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SKIP analysis. Genome Res. 14:2439–2447.

Arntzen, J. W., and G. P. Wallis. 1991. Restricted gene flow in a moving hybrid zone of the newts *Triturus cristatus* and *T. marmoratus* in western France. Evolution 45:805–826.

Barton, N. H., and G. M. Hewitt. 1985. Analysis of hybrid zones. Annu. Rev. Ecol. Syst. 16:113–148.

Barton, N. H., and B. O. Bengtsson. 1986. The barrier to genetic exchange between hybridising populations. Heredity 56:357–376.

Barton, N. H., and K. S. Gale. 1993. Genetic analysis of hybrid zones. Pp. 13–45 *in* R. G. Harrison, ed. Hybrid zones and the evolutionary process. Oxford Univ. Press, Oxford, U.K.

Bateson, W. 1909. Heredity and variation in modern lights. Pp. 85–101 *in* A. C. Seward, ed. Darwin and modern science. Cambridge Univ. Press, Cambridge, UK.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Lond. B 57:289–300.

Boursot, P., J.-C. Auffray, J. Britton-Davidian, and F. Bonhomme. 1993. The evolution of house mice. Annu. Rev. Ecol. Syst. 24:119–152.

Božíková, E., P. Munclinger, K. C. Teeter, P. K. Tucker, M. Macholán, and J. Piálek. 2005. Mitochondrial DNA in the hybrid zone between *Mus musculus musculus* and *Mus musculus domesticus*: a comparison of two transects. Biol. J. Linn. Soc. 84:363–378.

Bridle, J. R., S. J. E. Baird, and R. K. Butlin. 2001. Spatial structure and habitat variation in a grasshopper hybrid zone. Evolution 55:1832–1843.

Buerkle, C. A., and C. Lexer. 2008. Admixture as the basis for genetic mapping. Trends Ecol. Evol. 23:686–694.

Buggs, R. J. A. 2007. Empirical study of hybrid zone movement. Heredity 99:301–312.

Chahrour, M., S. Y. Jung, C. Shaw, X. Zhou, S. T. C. Wong, J. Qin, and H. Y. Zoghbi. 2008. MeCP2, a key contributor to neurological disease, activates and represses transcription. Science 320:1224–1229.

Counterman, B. A., D. Ortíz-Barrientos, and M. A. F. Noor. 2004. Using comparative genomic data to test for fast-X evolution. Evolution 58:656–660.

Coyne, J. A., and H. A. Orr. 1989. Two rules of speciation. Pp. 180–207 *in* D. Otte and J. Endler, eds. Speciation and its consequences. Sinauer Associates, Inc., Sunderland, MA.

———. 2004. Speciation. Sinauer Associates, Inc. Sunderland, MA.

Cucchi, T., J.-D. Vigne, and J.-C. Auffray. 2005. First occurrence of the house mouse (*Mus musculus domesticus* Schwarz & Schwarz, 1943) in the western Mediterranean: a zooarchaeological revision of sub-fossil house mouse occurrences. Biol. J. Linn. Soc. 84:429–445.

Dietrich, W. F., J. Miller, R. Steen, M. A. Merchant, D. Damron-Boles, Z. Husain, R. Dredge, M. J. Daly, K. A. Ingalls, T. J. O'Connor, et al. 1996. A comprehensive genetic map of the mouse genome. Nature 380:149–152.

Dobzhansky, T. 1936. Studies on hybrid sterility. 11. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. Genetics 21:113–135.

Dod, B., L. S. Jermiin, P. Boursot, V. H. Chapman, J. T. Nielsen, and F. Bonhomme. 1993. Counterselection on sex chromosomes in the *Mus musculus* European hybrid zone. J. Evol. Biol. 6:529–546.

Dod, B., C. Smadja, R. C. Karn, and P. Boursot. 2005. Testing for selection on the androgen-binding protein in the Danish mouse hybrid zone. Biol. J. Linn. Soc. 84:447–459.

Dufková, P., Macholán, M., Piálek, J. 2011. Inference of selection and stochastic effects in the house mouse hybrid zone. Evolution, doi: 10.1111/j.1558-5646.2011.01222.x.

Endler, J. A. 1977. Geographic variation, speciation and clines. Princeton Univ. Press, Princeton, NJ.

Espregueira Themudo, G. E., and J. W. Arntzen. 2007. Newts under siege: range expansion of *Triturus pygmeus* isolates populations of its sister species. Diversity Distrib. 13:580–586.

Forsdyke, D. R. 2011. The B in 'BDM.' William Bateson did not advocate a genic speciation theory. Heredity 106:202.

Galtier, N., B. Nabholz, S. Glémin, and G. D. D. Hurst. 2009. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. Mol. Ecol. 18:4541–4550.

Gompert, Z., and C. A. Buerkle. 2009. A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. Mol. Ecol. 18:1207–1224.

———. 2010. INTROGRESS: a software package for mapping components of isolation in hybrids. Mol. Ecol. Res. 10:378–384.

Good, J. M., M. D. Dean, and M. W. Nachman. 2008. A complex genetic basis to X-linked hybrid male sterility between two species of house mice. Genetics 179:2213–2228.

Grula, J. W., and O. R. Taylor. 1980. Some characteristics of hybrids derived from the sulphur butterflies, *Colias eurytheme* and *Colias philodice*: phenotypic effects of the X chromosome. Evolution 34:673–687.

Guillot, G., F. Mortier, and A. Estoup. 2005. Geneland: a program for landscape genetics. Mol. Ecol. Notes 5:712–715.

Haldane, J. B. S. 1922. Sex ratio and unisexual sterility in animal hybrids. J. Genet. 12:101–109.

Harr, B. 2006. Genomic islands of differentiation between house mouse subspecies. Genome Res. 16:730–737.

Ibrahim, K. M., R. A. Nichols, and G. M. Hewitt. 1996. Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. Heredity 77:282–291.

Jiggins, C. D., M. Linares, R. E. Naisbit, C. Salazar, Z. H. Yang, and J. Mallet. 2001. Sex-linked hybrid sterility in a butterfly. Evolution 55:1631–1638.

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

Lindblad-Toh, K., E. Winchester, M. J. Daly, D. G. Wang, J. N. Hirschhorn, J. P. Laviolette, K. Ardlie, D. E. Reich, E. Robinson, P. Sklar, et al. 2000. Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. Nat. Genet. 24:381–386.

Macholán, M., B. Kryštufek, and V. Vohralík. 2003. The location of the *Mus musculus/M. domesticus* hybrid zone in the Balkans: clues from morphology. Acta Theriol. 48:177–188.

Macholán, M., P. Munclinger, M. Šugerková, P. Dufková, B. Bímová, E. Božíková, J. Zima, and J. Piálek. 2007. Genetic analysis of autosomal and X-linked markers across a mouse hybrid zone. Evolution 61:746–771.

Macholán, M., S. J. E. Baird, P. Munclinger, P. Dufková, B. Bímová, and J. Piálek. 2008. Genetic conflict outweighs heterogametic incompatibility in the mouse hybrid zone? BMC Evol. Biol. 8:271–284.

Mouse Genome Sequencing Consortium. 2002. Initial sequencing and comparative analysis of the mouse genome. Nature 420:520–562.

Muller, H. J. 1940. Bearing of the *Drosophila* work on systematics. Pp. 185–268 in J. S. Huxley, ed. The new systematics. Clarendon Press, Oxford.

———. 1942. Isolating mechanisms, evolution, and temperature. Biol. Symp. 6:71–125.

Munclinger, P., E. Božíková, M. Šugerková, J. Piálek, and M. Macholán. 2002. Genetic variation in house mice (*Mus*, Muridae, Rodentia) from the Czech and Slovak Republics. Folia Zool. 51:81–92.

Munclinger, P., P. Boursot, B. Dod. 2003. B1 insertions as easy markers for mouse population studies. Mamm. Genome 14:359–366.

Nagylaki, T. 1975. Conditions for the existence of clines. Genetics 80:595–615.

———. 1976. Clines with variable migration. Genetics 83:867–886.

Nolte, A. W., Z. Gompert, and C. A. Buerkle. 2009. Variable patterns of introgression in two sculpin hybrid zones suggest that genomic isolation differs among populations. Mol. Ecol. 18:2615–2627.

Oka, A., A. Mita, N. Sakurai-Yamatani, H. Yamamoto, N. Takagi, T. Takano-Shimizu, K. Toshimori, K. Moriwaki, K., and T. Shiroishi. 2004. Hybrid breakdown caused by substitution of the X chromosome between two mouse subspecies. Genetics 166:913–924.

Orr, H. A. 1996. Dobzhansky, Bateson, and the genetics of speciation. Genetics 144:1331–1335.

———. 1997. Haldane's rule. Annu. Rev. Ecol. Syst. 28:195–218.

Orr, H. A., and S. Irving. 2005. Segregation distortion in hybrids between the Bogota and USA subspecies of *Drosophila pseudoobscura*. Genetics 169:671–682.

Orr, H. A., J. P. Masly, and N. Phadnis. 2007. Speciation in *Drosophila*: From phenotypes to molecules. J. Hered. 98:103–110.

Payseur, B. A. 2010. Using differential introgression in hybrid zones to identify genomic regions involved in speciation. Mol. Ecol. Res. 10:806–820.

Payseur, B. A., J. G. Krenz, and M. W. Nachman. 2004. Differential patterns of introgression across the X chromosome in a hybrid zone between two species of house mice. Evolution 58:2064–2078.

Phadnis, N., and H. A. Orr. 2009. A single gene causes both male sterility and segregation distortion in *Drosophila* hybrids. Science 323:376–379.

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

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

Piálek, J., M. Vyskočilová, B. Bímová, D. Havelková, J. Piálková, P. Dufková, V. Bencová, L'. Ďureje, T. Albrecht, H. C. Hauffe, et al. 2008. Development of unique house mouse resources suitable for evolutionary studies of speciation. J. Hered. 99:34–44.

Pletcher, M. T., P. McClurg, S. Batalov, A. I. Su, S. W. Barnes, E. Lagler, R. Korstanje, X. S.Wang, D. Nusskern, M. A. Bogue, et al. 2004. Use of a dense single nucleotide polymorphism map for in silico mapping in the mouse. PLoS Biol. 2:2159–2169.

Porter, A. H., R. Wenger, H. Geiger, A. Scholl, and A. M. Shapiro. 1997. The *Pontia daplidice-edusa* hybrid zone in northwestern Italy. Evolution 51:1561–1573.

Prager, E. M., P. Boursot, and R. D. Sage. 1997. New assays for Y chromosome and p53 pseudogene clines among East Holstein house mice. Mamm. Genome 8:279–281.

Prowell, D. P. 1998. Sex linkage and speciation in Lepidoptera. Pp. 309–319 in D. J. Howard and S. H. Berlocher, eds. Endless forms: species and speciation. Oxford Univ. Press, Oxford, U.K.

R Development Core Team. 2007. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 3-900051-07-0, http://www.R-project.org.

Raufaste, N., A. Orth, K. Belkhir, D. Senet, C. Smadja, S. J. E. Baird, F. Bonhomme, B. Dod, and P. Boursot. 2005. Inference of selection and migration in the Danish house mouse hybrid zone. Biol. J. Linn. Soc. 84:593–616.

Rieseberg, L. H., J. Whitton, and K. Gardner. 1999. Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. Genetics 152:713–727.

Sage, R. D., W. R. Atchley, and E. Capanna. 1993. House mice as models in systematic biology. Syst. Biol. 42:523–561.

Shifman, S., J. T. Bell, R. R. Copley, M. S. Taylor, R. Mott, and J. Flint. 2006. A high-resolution single nucleotide polymorphism genetic map of the mouse genome. PLoS Biol. 4:e395.

Slatkin, M. 1973. Gene flow and selection in a cline. Genetics 75:733–756.

———. 1975. Gene flow and selection on a two-locus system. Genetics 81:787–802.

Storchová, R., S. Gregorová, D. Buckiová, D. Kyselová, P. Divina, and J. Forejt. 2004. Genetic analysis of X-linked hybrid sterility in the house mouse. Mamm. Genome 15:515–524.

Storchová R., J. Reif, and M. W. Nachman. 2010. Female heterogamy and speciation: reduced introgression of the Z chromosome between two species of nightingales. Evolution 64:456–471.

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

———. 1991. The genetic structure of the hybrid zone between the firebellied toads *Bombina bombina* and *B. variegata*: comparisons between transects and between loci. Evolution 45:237–261.

Tao, Y., D. L. Hartl, and C. C. Laurie. 2001. Sex-segregation distortion associated with reproductive isolation in *Drosophila*. Proc. Natl. Acad. Sci. USA 98:13183–13188.

Tao, Y., S. Chen, D. L. Hartl, and C. C. Laurie. 2003. Genetic dissection of hybrid incompatibilities between *Drosophila simulans* and *D. mauritiana*. I. Differential accumulation of hybrid male sterility effects on the X and autosomes. Genetics 164:1383–1397.

Teeter, K. C., B. A. Payseur, L. W. Harris, M. A. Bakewell, L. M. Thibodeau, J. E. O'Brien, J. G. Krenz, M. A. Sans-Fuentes, M. W. Nachman, and P. K. Tucker. 2008. Genome-wide patterns of gene flow across a house mouse hybrid zone. Genet. Res. 18:67–76.

Teeter, K. C., L. M. Thibodeau, Z. Gompert, C. A. Buerkle, M. W. Nachman and P. K. Tucker. 2010. The variable genomic architecture of isolation between hybridizing species of house mice. Evolution 64:472–485.

Tucker, P. K., R. D. Sage, J. Warner, A. C. Wilson, and E. M. Eicher. 1992. Abrupt cline for sex chromosomes in a hybrid zone between two species of mice. Evolution 46:1146–1163.

Vanlerberghe, F., P. Boursot, J. Catalan, S. Guerasimov, F. Bonhomme, B. Botev, and L. Thaler. 1988a. Analyse génétique de la zone d'hybridation entre les deux sous-espèces de souris *M. m. domesticus* et *M. m. musculus* en Bulgarie. Genome 30:427–437.

Vanlerberghe, F., P. Boursot, J. T. Nielsen, and F. Bonhomme. 1988b. A steep cline for mitochondrial DNA in Danish mice. Genet. Res. 52:185–193.

Vanlerberghe, F., B. Dod, P. Boursot, M. Bellis, and F. Bonhomme. 1986. Absence of Y-chromosome introgression across the hybrid zone between *Mus musculus domesticus* and *Mus musculus musculus*. Genet. Res. 48:191–197.

Wolfram, S. 1992. Mathematica: a system for doing mathematics by computer. Addison-Wesley Publishing Company, Inc., Redwood City.

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

Zouros, E., K. Lofdahl, and P. A. Martin. 1988. Male hybrid sterility in *Drosophila*: interactions between autosomes and sex chromosomes in crosses of *D. mojavensis* and *D. arizonensis*. Evolution 42:1321–1331.

Associate Editor: N. Barton

Supporting Information

The following supporting information is available for this article:

Table S1. List of sampling sites.
Table S2. List of analyzed markers (in parentheses, abbreviations used in the text).
Table S3. List of new insertions diagnostic for *musculus/domesticus* subspecies.
Table S4. Results of Barton's concordance analysis.
Figure S1. Individual-based estimates of heterozygote deficit (F) against individual hybrid index (HI), a treatment analogous to Barton's concordance approach for allele frequencies.
Figure S2. Genomic clines for all 24 loci.

Supporting Information may be found in the online version of this article.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.