

GENETIC ANALYSIS OF AUTOSOMAL AND X-LINKED MARKERS ACROSS A MOUSE HYBRID ZONE

Miloš Macholán,^{1,2} Pavel Munclinger,^{1,3} Monika Šugerková,¹ Petra Dufková,⁴ Barbora Bímová,^{3,5} Eva Božíková,³ Jan Zima,⁵ and Jaroslav Piálek⁵

¹Laboratory of Mammalian Evolutionary Genetics, Institute of Animal Physiology and Genetics, Academy of Sciences of the Czech Republic, 60200 Brno, Czech Republic

²E-mail: macholan@iach.cz

³Biodiversity Research Group, Department of Zoology, Charles University, Prague, Czech Republic

⁴Department of Genetics, University of South Bohemia, Czech Republic

⁵Department of Population Biology, Institute of Vertebrate Biology, ASCR, Brno and Studenec, Czech Republic

Received November 10, 2005

Accepted December 11, 2006

In this paper, we present results of the first comprehensive study of the introgression of both autosomal and sex-chromosome markers across the central European portion of the hybrid zone between two house mouse subspecies, *Mus musculus musculus* and *M. m. domesticus*. More than 1800 individuals sampled from 105 sites were analyzed with a set of allozyme loci (hopefully representing neutral or nearly neutral markers) and X-linked loci (which are assumed to be under selection). The zone center is best modeled as a single straight line independent of fine-scale local geographic or climatic conditions, being maintained by a balance between dispersal and selection against hybrids. The width (w) of the multilocus autosomal cline was estimated as 9.6 km whereas the estimate for the compound X-chromosome cline was about 4.6 km only. As the former estimate is comparable to that of the Danish portion of the zone (assumed to be much younger than the central European one), zone width does not appear to be related to its age. The strength (B) of the central barrier was estimated as about 20 km; with dispersal (σ) of about 1 km/gen^{1/2}, this means effective selection (s^*) is approximately 0.06–0.09 for autosomal loci and about 0.25 for X-linked loci. The number of loci under selection was estimated as $N = 56$ –99 for autosomes and about 380 for X-linked loci. Finally, we highlight some potential pitfalls in hybrid zone analyses and in comparisons of different transects. We suggest that conclusions about parts of the mouse genome involved in reproductive isolation and speciation should be drawn with caution and that analytical approaches always providing some estimates should not be used without due care regarding the support or confidence of such estimates, especially if conclusions are based on the difference between these estimates. Finally, we recommend that analysis in two-dimensional space, dense sampling, and rigorous treatment of data, including inspection of likelihood profiles, are essential for hybrid zone studies.

KEY WORDS: Gene flow, hybridization, mouse, *Mus musculus domesticus*, *Mus musculus musculus*, selection.

Free gene flow among individuals of the same species and its absence between individuals of different species is the central tenet of the biological species concept (Dobzhansky 1937; Mayr 1942; Coyne and Orr 2004). Separation of biological species requires the

evolution of genetic barriers that either directly cause a substantial reduction in exchange of alleles or facilitate further evolution of genetic differences eventually leading to a reproductive isolation. Therefore, the study of the origin and evolution of genetic

barriers, and the effects these barriers have on gene flow, is the main prerequisite for understanding speciation.

To distinguish between causes and consequences of genetic isolation, one should focus on cases in which the isolation is not complete. Excellent areas for such studies are naturally occurring zones of contact between genetically diversified populations ("hybrid zones;" Barton and Hewitt 1985; Harrison 1990, 1993; Arnold 1997; Lijtmaer et al. 2003). An advantage of this approach is that mixing parental genomes in hybrid zones brings together new gene combinations that could not have been under selective scrutiny before hybridization. Accordingly, fitness of hybrids can vary considerably relative to fitness of their parents.

Hybrid zone theory predicts that selection acting on particular traits can be estimated using an analysis of changes in allele frequencies along a geographic transect or cline. In the simplest models based on analysis of divergence at one or two loci, the relationship between cline shape and selection is straightforward: the stronger the selection, the narrower the cline (Slatkin 1973, 1975; Nagylaki 1975, 1976; Endler 1977). However, hybrid zones typically involve divergence at more than a single locus or a few loci. If we assume that in a hybrid zone many loci are under weak selection and clines in gene frequency of these loci coincide, then the influx of different parental gene combinations into the zone causes linkage disequilibrium between them. Because the strength of these gene associations is proportional to the gradient of gene frequencies, only weak linkage disequilibrium occurs at the tails of a cline in which the gradient is shallow. However, as the gradient gets steeper towards the center of the zone, the disequilibrium increases; at the same time, selection on an individual locus is strengthened by its associations with other loci. This results in a sharp step at the center of each cline, even for neutral loci. This part of the cline is steeper than it would be if selection affected each locus separately and strengthens the barrier to introgression of foreign alleles (Szymura and Barton 1986, 1991; Barton and Gale 1993; Baird 1995; Kruuk et al. 1999). On the other hand, negligible disequilibria at the edges of the cline result in long introgression tails and selection outside the central region is regarded as acting at each locus separately. Importantly, by assessing differences in cline shapes between traits, we can detect areas of the genome that are under strong selection and thus potentially contribute to speciation (Barton and Hewitt 1985; Barton and Gale 1993; Payseur et al. 2004).

Several lines of evidence indicate that sex chromosomes carry more genes causing disruption of fertility and/or viability in hybrids than autosomes and hence can be under stronger selection (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; Harr 2006). One of the possible explanations of this phenomenon comes from the dominance theory of

Haldane's rule (Haldane 1922; Orr 1997; Coyne and Orr 2004) based on the assumption that if alleles causing hybrid problems are recessive, their effect will be much stronger on the X chromosome that is hemizygous in males ("large X-effect"; Orr 1997). Stronger selection is thus expected to be manifested by more limited introgression and hence narrower clines of X-linked markers. However, to detect differences in selection pressures between different parts of the genome, we must know, in the first place, exact positions of analyzed markers.

The house mouse, *Mus musculus*, represents a very useful model for the study of differential introgression of various parts of the genome across zones of contact. First, the mouse genome has recently been sequenced (Mouse Genome Sequencing Consortium 2002) and a dense genetic map is available for this species (Dietrich et al. 1996; Lindblad-Toh et al. 2000; Abe et al. 2004; Pletcher et al. 2004). Second, there are two mouse subspecies, *M. m. musculus* and *M. m. domesticus* (sometimes referred to as species, *M. musculus* and *M. domesticus*), which hybridize in Eurasia. In Europe, the contact front is about 2500 km long and runs across the Jutland Peninsula and from the Baltic coast in East Holstein (northern Germany), through Central Europe and the Balkan Peninsula to the Black Sea coast (Boursot et al. 1993; Sage et al. 1993; Macholán et al. 2003). Two transects have been studied intensively thus far, one in Denmark (Ursin 1952; Selander and Yang 1969; Hunt and Selander 1973; Vanlerberghe et al. 1986, 1988b; Dod et al. 1993, 2005; Lanneluc et al. 2004; Raufaste et al. 2005) and one near Munich in southern Germany (Tucker et al. 1992; Payseur et al. 2004; Payseur and Nachman 2005). Introgression of sex-chromosome markers has been shown to be limited compared to autosomal markers (Vanlerberghe et al. 1986; Tucker et al. 1992; Dod et al. 1993); however, only a few recent papers have used rigorous likelihood-based statistical methods to fit clines to the data and estimate cline and fitness parameters (Dod et al. 2005; Raufaste et al. 2005).

In this paper, we present results of the first comprehensive study of the central European portion of the *musculus/domesticus* hybrid zone based on the study of more than 1800 individuals sampled from 105 sites and analyzed with a set of allozyme loci, representing neutral or nearly neutral markers, and X-chromosome markers, which are assumed to be linked to loci under selection. Rate of dispersal, linkage disequilibrium, fitness of hybrids, selection against hybrids, and the number of loci under selection are estimated from the data. We highlight some potential pitfalls in hybrid zone analyses and suggest that conclusions about parts of the mouse genome involved in reproductive isolation and speciation should be drawn with caution. Finally we recommend that analysis in two-dimensional space, dense sampling, and rigorous treatment of data including inspection of likelihood profiles is essential for hybrid zone studies.

Materials and Methods

SAMPLING

Mice were trapped with metal and/or wooden live traps during several trapping seasons between 1991 and 2003 although the majority of the individuals (95.5%) were collected since 1997. In total, 1841 mice were collected from 105 sites, scattered across an area ca. 110 km long and 40 km wide covering western Bohemia (Czech Republic) and northeastern Bavaria (Germany) (Fig. 1). There were slightly more females (948) than males (893) in the sample, but this difference was not significant at the 95% and even the 85% level (chi-square test: $\chi^2 = 1.645$, $P = 0.200$). The sampling sites are listed in the Appendix and shown in Figure 1. Geographic coordinates of all the sites are available at <http://www.iach.cz/legs>.

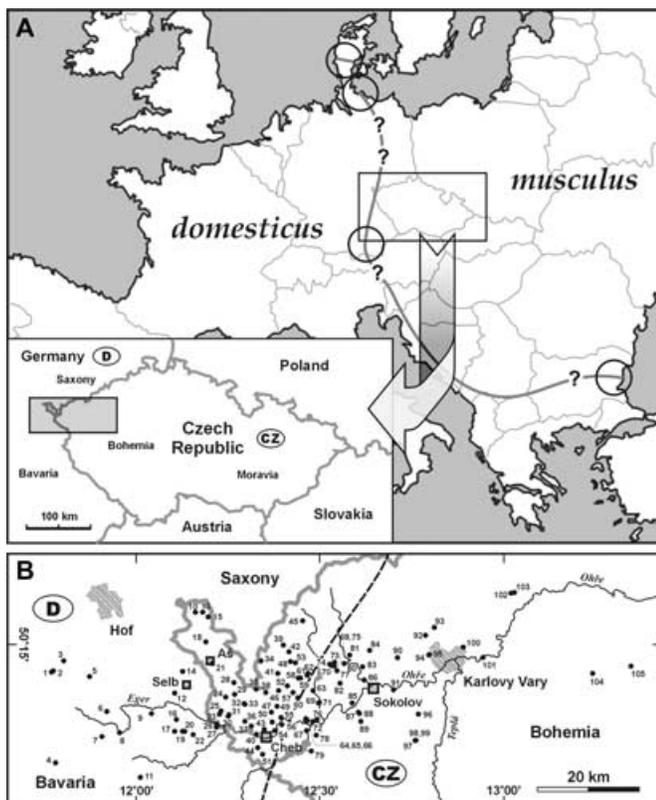


Figure 1. (A) The course of the *musculus/domesticus* hybrid zone in Europe. Circles indicate previously studied transects. In the insert, the position of the study area is indicated. (B) Location of sampling sites. Locality numbers correspond to those in the Appendix. The thick dashed line is an approximate zone center defined as a 0.5-isocline derived from the bicubic spline smoothing of allele frequencies at each site. Intersection of this isocline with the bottom margin of the figure was used as a starting point for subsequent maximum likelihood-based positioning of the zone center (see text for details).

AUTOSOMAL MARKERS

Mice were sacrificed and dissected either in the field or in the laboratory. Samples of kidney and muscle were frozen in liquid nitrogen and kept at -80°C until processed. Seven enzymatic loci that appeared to discriminate the two taxa in previous studies (Bonhomme et al. 1984; Munclinger et al. 2002) were used as autosomal markers: isocitrate dehydrogenase-1 (E.C. 1.1.1.42; *Idh1*, chromosome 1; $N = 1485$), glucose dehydrogenase-1 (E.C. 1.1.1.47; *Gpd1*, Chr. 4; $N = 1436$), superoxide dismutase-1 (E.C. 1.15.1.1; *Sod1*, Chr. 16; $N = 1478$), nucleoside phosphorylase (E.C. 2.4.2.1; *Np*, Chr. 14; $N = 1503$), esterase-1 and 2 (E.C. 3.1.1.1; *Es1*, *Es2*, Chr. 8; $N = 1416$ and 1439 , respectively), and mannose phosphate isomerase (E.C. 5.3.1.8; *Mpi*, Chr. 9; $N = 1485$). All the allozymes were scored after standard horizontal starch gel electrophoresis (Harris and Hopkinson 1976; Pasteur et al. 1988) using samples of the C57BL/6J inbred strain as standards (see Munclinger et al. 2002 for details).

X-CHROMOSOME MARKERS

DNA was isolated from frozen or ethanol-preserved tissues using proteinase K digestion and subsequent extraction with phenol-chloroform and ethanol precipitation (Hoelzel and Green 1992). Five X-chromosome markers were scored, three of them being subspecies-specific short interspersed nuclear elements or SINEs. Genetic positions of these markers (in cM) refer to closest known markers in the Whitehead-MIT F2 intercross (Dietrich et al. 1996; <http://www.broad.mit.edu/>) to allow an easy comparison with results in Payseur et al. (2004).

The first SINE is a B1 insertion in the *Btk* gene (43.7 cM; $N = 1655$ individuals), which is fixed in *M. m. domesticus* and absent in *M. m. musculus* (Munclinger et al., 2002, 2003). The second marker, a B1 insertion in the *Tsx* gene (29.5 cM; $N = 1627$), is not fixed in all *domesticus* populations, but it is a diagnostic marker in Central Europe (Munclinger et al. 2003). The primers for *Btk* and *Tsx* markers are described in Munclinger et al. (2003). The third marker is a B2 insertion 3' downstream the *Syap1* gene (55.7 cM; $N = 1375$). The primers used were 5'-TGGCTGAGTCACCACTTGTT-3' and 5'-TGGGGAATGACATTGAGGT-3'. This insertion is fixed in *M. m. domesticus* and absent in *M. m. musculus* (P. Munclinger, unpubl.).

Two SNPs, *Nt* (Payseur et al. 2004; $N = 1237$) and *DXMit18.2* (Lindblad-Toh et al. 2000; $N = 1354$), were also used. These markers are mapped to intergenic regions of the X chromosome at 20.8 cM and 29.5 cM, respectively (note that the map position is the same for *DXMit18.2* and *Tsx* so that independence of those markers may be questioned).

The PCR conditions for fragments potentially containing SINEs followed Munclinger et al. (2003). The temperature profile was modified only for the insertion in the *Tsx* gene, where

after 3 min incubation at 95°C, 35 cycles were performed at 95°C (3 min), 55°C (30 sec), and 72°C (30 sec). The conditions for *Nt* and *DXMit18.2* were those described in Payseur et al. (2004). All PCR products were run on 2% agarose gels and visualized by ethidium bromide staining.

MICROSATELLITES

To get an estimate of the rate of dispersal σ independent from that estimated from a cline shape, genotypes at four microsatellite loci were used. Dispersal was estimated under the isolation-by-distance model (Wright 1943) using two methods. First, the regression between the $\theta/(1-\theta)$ statistic (Weir and Cockerham 1984), where θ denotes the fixation index, and the logarithms of geographic distances between populations was calculated (Rousset 1997). The θ rather than R_{ST} (Slatkin 1995) was used as recommended by Gaggiotti et al. (1999) for situations in which sample sizes and number of loci are small and because of uncertainties concerning the mutation model of microsatellites. The second method used the \hat{a} parameter (Rousset 2000) estimated by comparing individuals rather than population samples. In both cases, the regression slopes b equal $1/(4\pi D\sigma^2)$, where D is population density. Under the isolation-by-distance model (Wright 1943) we could estimate the slope, b , of the regression of two variables, $\theta/(1-\theta)$ (Weir and Cockerham 1984), and \hat{a} (Rousset 2000), on geographic distance. The dispersal was then calculated as $\sigma = \sqrt{1/(4\pi D\hat{b})}$. The detailed information about the loci and methods is given in Supplementary Material available online.

FITTING THE CLINE IN TWO DIMENSIONS

To pool samples from successive years we must assume that the zone has been sufficiently stable and that allele frequencies at individual loci have not significantly changed from year to year. Although we have no information about the stability of the zone in Central Europe, there is no evidence of its movement in Denmark since the 1960s (Raufaste et al. 2005). We further tested the assumption of temporal homogeneity of allele frequencies in polymorphic samples of sufficient size: Neuenreuth (1999: $N = 17$; 2001: $N = 13$), Lužná (1999: $N = 16$; 2000: $N = 13$; 2001: $N = 27$; 2002: $N = 11$), Hůrka (2002: $N = 42$; 2003: $N = 7$), Plesná (1997: $N = 114$; 1999: $N = 19$), Milhostov (2000: $N = 14$; 2003: $N = 7$), Kaceřov 2 (1999: $N = 11$; 2001: $N = 26$; 2003: $N = 5$). Only frequencies of *Sod1* from Plesná and Kaceřov 2 were proven to be significantly different between successive years after a Bonferroni correction (6 sites \times 6 loci; one locus, *Es2*, was not considered for reasons given in Results); therefore, samples from different years were pooled.

Aside from the sampling error, allele frequencies between sites can be expected to fluctuate randomly due to a drift. Therefore, to avoid giving undue weight to very large samples, allele frequencies at each site were weighed in proportion to the effective sample size, N_e . According to Szymura and Barton (1986, 1991),

N_e is defined as an inverse of the variance in average frequency of one taxon's alleles (here *musculus* is chosen) at the i th locus around the average cline, $V = 1/n + F_{ST}/k$, where n is the number of all alleles summed over k scored loci, and $F_{ST} = \text{var}(p_i)/p_iq_i$. The latter term is called the standardized variance of an allele frequency (Szymura and Barton 1986, 1991) and represents the residual variation around the regression of allele frequencies at individual loci against the average of all loci in each sample.

To minimize potential bias resulting from incorrect orientation of a linear transect across the hybrid zone, its orientation in two-dimensional space was estimated as follows. First, the bicubic spline smoothing procedure using the 3D Contour Plot routine of the Statistica software package (StatSoft, Inc., Prague, Czech Republic 2000) was applied to the data in the two-dimensional plane in which each site was defined by the x and y coordinates and the third coordinate was given by allele frequencies averaged across six allozyme loci (excluding *Es2*, see Results). The position of the zone center was then defined as the 0.5-isocline (Fig. 1B).

In the second step, a maximum likelihood (ML) method using a Metropolis algorithm was used (for details, see below). In this method, the position of the cline center can be approximated either by a single straight line or by a more-or-less angled line made up by various numbers of segments (Bridle et al. 2001). Obviously, the more segments a given cline center has, the better overall fit we should expect because a greater number of parameters allow the zone position to be defined at a finer geographical scale. However, adding further segments into the model increases the support interval around parameter estimates as they must be calculated from the same amount of data. Therefore, new segments should only be added until there is no further significant improvement of fit. This can be tested with the likelihood ratio test (LRT) in which the test statistic, $LR = 2(\ln L_0 - \ln L_1)$, can be approximated by the chi-square distribution with one degree of freedom per each segment added. The chi-square approximation is valid only when the two hypotheses compared are nested. In this case, multisegment clines may only be considered nested if the simpler (null) hypothesis is a special case of the more parameter-rich (alternative) hypothesis. This means that more complex clines must be derived from simpler ones by subdividing existing segments. Such a series of segments may be, for example, $\{1, 2, 4, 8, \dots\}$ (i.e., an eight-segment cline is created by halving each part of a four-segment cline, which in turn was created by halving a two-segment cline etc.) or $\{1, 3, 6, 12, \dots\}$ or $\{1, 3, 9, 27, \dots\}$ or $\{1, 5, 10, 20, \dots\}$ etc., where individual segments are of equal length. Conversely, series such as $\{1, 2, 3, 4, \dots\}$ do not represent sets of nested hypotheses and LRT cannot be applied. Although one can use another criterion that is not constrained by this requirement, for instance AIC (Akaike 1973), comparing nonnested families of hypotheses can lead to likelihood oscillations rather than to

smooth increase of likelihood values as more parameters are added (Bridle et al. 2001). For each series, the total length of the cline remains constant to ensure the hypotheses to be comparable and to keep the total number of parameters at a computationally reasonable level.

We analyzed three nested series: {1, 2, 4}, {1, 3, 6}, and {1, 3, 9}, and for each number of segments we estimated ML using three cline models of increasing complexity: sigmoid, symmetrical stepped, and asymmetrical stepped models with two, four, and six parameters, respectively (see the next section for details). Thus, in total $6 \times 3 = 18$ hypotheses were tested with LRT. During computations, the cline parameters were free to vary; however, the same cline width was maintained across all segments, that is, it was not allowed to differ in different parts of the cline. Similarly, although angles between segments could vary freely, the position of the starting point was fixed. We used the intersection of the 0.5-isocline with the southern edge of the study area as the start (cf. Fig. 1B and 2A). In this way, together with fixing the total cline length and keeping the length of each segment constant, we minimized the number of parameters and made convergence to a likelihood peak computationally feasible.

When the course of the zone in two-dimensional space was estimated, distances from each site to the cline center were measured. Subsequently, the distance of the most distant site on the *domesticus* side (i.e., Straas 1) was set to zero and all other distances were recalculated accordingly. By this procedure the three-parameter space was collapsed to a two-parameter situation in which each site was defined by its distance from Straas 1 along an imaginary linear transect perpendicular to the course of the hybrid zone. Both a three-dimensional contour plot and ML approach gave almost identical results (Pearson product-moment correlation $R = 0.997$) with average difference $\Delta d = 0.44$ km (range 0–6.64 km) so only maximum-likelihood estimates (MLE) of distances were used in all subsequent analyses. The Analyse 1.3 program (Barton and Baird 1995; available at <http://helios.bto.ed.ac.uk/evolgen/Mac/Analyse/Version1.3.html>) was used for the ML estimations.

COMPARISON OF CLINE SHAPES

The simplest cline model fitted to the data (referred here to as “Sig” model) is defined by a hyperbolic tangent function

$$p(x_i) = \frac{1}{2} \left[1 + \tanh \left(\frac{2(x_i - c)}{w} \right) \right], \quad (1)$$

where $p(x_i)$ is the allele frequency at i th site, $(x_i - c)$ is the distance of this site from the cline center (c), and w is the cline width, given as an inverse of the maximum slope of the curve. This is a general single-locus model giving a smooth sigmoid curve (or a straight line if plotted on a logit scale) that can be used for modeling clines caused by either heterozygote disadvantage or

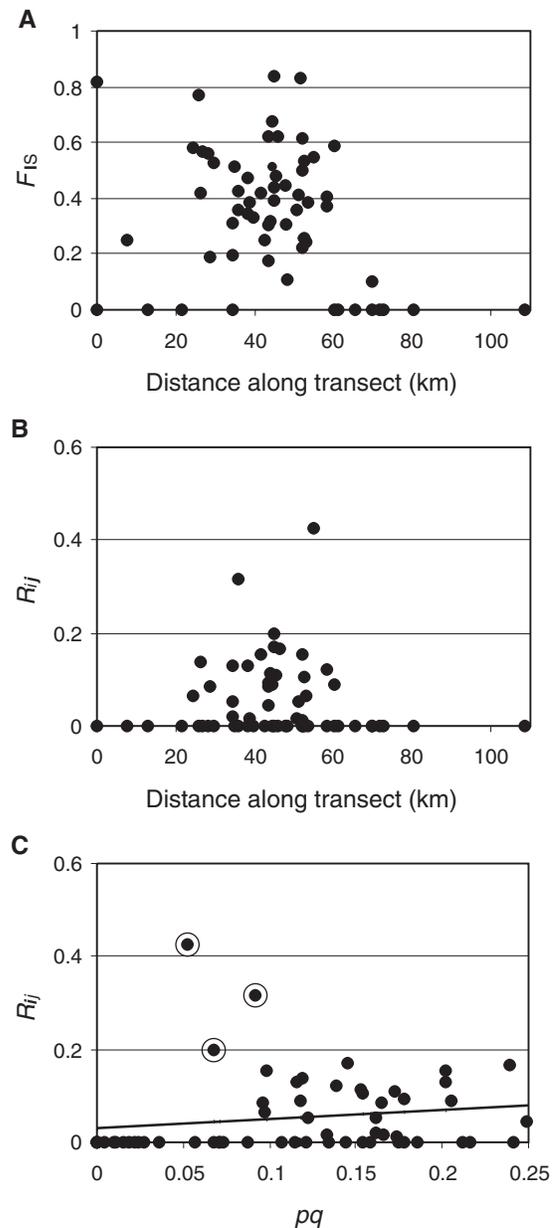


Figure 2. Values of F_{1s} (A) and standardized linkage disequilibrium averaged over all pairs of loc. (B) R_{ij} plotted against the distance along the transect. (C) R_{ij} plotted against the product of allele frequencies, $\bar{p}\bar{q}$, averaged over six autosomal loci. The regression line is given by the equation $y = 0.0301 + 0.1965x$. The slope of the regression line is not significantly different from zero; after removing the three outliers (marked with circles) the slope of the new regression ($y = 0.0023 + 0.3196x$) becomes significant.

extrinsic selection favoring different alleles in different places (Haldane 1948; Bazykin 1969; Kruuk et al. 1999; Barton and Shpak 2000). This model is based on the diffusion approximation (Haldane 1948; Fisher 1950; Bazykin 1969), which assumes weak selection acting on each locus separately with no epistasis (Barton and Shpak 2000).

However, when dispersal is balanced by selection on several to many loci, we should expect a pattern different from the simple model outlined above. As pointed out above, if clines at these loci coincide, linkage disequilibria cause the selection on individual loci to be strengthened, resulting in a characteristic sharp step at the center of each cline. When plotted on a logit scale, the whole cline can be described by three straight lines. According to this “stepped” model, most change occurs around the center, yet foreign alleles introgress far beyond the zone. The central segment of a cline is described by the sigmoid function (eq. 1) whereas the regions outside the center are characterized by exponential decay towards the tails described by two parameters: the rate of decay, θ , and the size of the central step, $B = \Delta p/p'$, where Δp is the difference in allele frequency across the central step and $p' = \delta p/\delta x$ is the gradient of allele frequency p along distance x at the edges outside this step (Nagylyki 1976; Barton 1986; Szymura and Barton 1986). The parameter B describes the strength of the central selective barrier to diffusion of alleles from one genetic background to another and is usually expressed as the physical distance (in kilometers) to which a foreign allele would introgress without the barrier. The parameter θ then measures the strength of selection acting separately on the characters outside the central region. The left tail can be expressed by the equation (after Raufaste et al. 2005)

$$p(x_i) = \frac{w}{2\sqrt{\theta_0}} \cdot \frac{\Delta u}{B_0} \exp \left[\frac{2(x_i - c)}{w} \sqrt{\theta_0} \right], \quad (2)$$

and similarly for the right tail

$$p(x_i) = 1 - \frac{w}{2\sqrt{\theta_1}} \cdot \frac{\Delta u}{B_1} \exp \left[\frac{-2(x_i - c)}{w} \sqrt{\theta_1} \right], \quad (3)$$

where Δu is given as

$$\Delta u = \frac{\frac{2}{w} B_0 \sqrt{\theta_0} B_1 \sqrt{\theta_1}}{B_0 \sqrt{\theta_0} + B_1 \sqrt{\theta_1} + \frac{2}{w} B_0 \sqrt{\theta_0} B_1 \sqrt{\theta_1}}, \quad (4)$$

which simplifies to

$$\Delta u = \frac{B\sqrt{\theta}}{w + B\sqrt{\theta}} \quad (5)$$

when the model is symmetrical. Note that even though Δp (Barton 1986; Porter et al. 1997) and Δu (defined Raufaste 2001 see also Raufaste et al. 2005) both describe the central cline step, the two parameters have slightly different meaning, which is not obvious at first glance. Because Δu is defined as the difference of allele frequencies between the intersections of introgression tails with an imaginary vertical line passing through the cline center ($x = c$) whereas Δp is the distance between the intersections of the tails with the central cline segment itself, it should hold $\Delta u \leq \Delta p$. The rate of decay, θ , can be expressed in terms of the ratio between the selection acting on an individual locus itself (s_e) and the effective selection on the locus at the center due to association with other

loci (s^*): $\theta = s_e/s^*$ (hence, $\theta = 1$ when selection affects the loci separately; Szymura and Barton 1986).

In this paper, we used all three models, that is, sigmoid (Sig, described by two parameters: c and w), symmetrical stepped (Sstep, with four parameters: c , w , B , and θ), and asymmetrical stepped model (Astep, with six parameters: c , w , B_0 , B_1 , θ_0 , and θ_1 , where subscripts 0 and 1 denote the *domesticus* and *musculus* side, respectively). The best-fitting model was then chosen using LRT with two degrees of freedom.

As in the case of two-dimensional clines we used the ML method (Edwards 1992) for fitting one of the three hybrid zone models to the data. However, because searching over the whole four- or even six-parameter space cannot be done analytically, we applied a Markov chain Monte Carlo (MCMC) method based on a modified Metropolis–Hastings algorithm (Metropolis et al. 1953; Hastings 1970). Several runs of chains of 10^3 – 10^6 steps (depending on the complexity of the likelihood surface) were carried out for all parameters and MLEs were then scored. Following Edwards (1992), two-unit support limits were used as approximately 95% confidence intervals. For this purpose, we followed an approach suggested by Phillips et al. (2004): for each model and each locus, the likelihood surface was explored stepwise along an axis for one of the parameters with the other parameters free to vary at each step and the resulting likelihood profile was then constructed (see also Takami and Suzuki 2005). The same procedure was repeated for all the cline parameters. The support limits were found at $L_{\max} - 2$ (2LL). In a few allozyme loci, we had to make some arbitrary decisions (cf. *Idh1*, *Mpi*, and *Np* in Fig. 5): in these cases, log-likelihoods increased monotonically from a width of several kilometers until nearly zero and hence we would arrive at an extreme value of few meters only, substantially lower than individual dispersal. Because a central step of infinite steepness can be placed between any two discrete sampling points, analyses using the stepped models can often only place an upper bound on the width (S. J. E. Bird, pers. comm.). In this study, we arbitrarily accepted only those values of w that were higher than 5 km and for which we could identify an apparent peak on the likelihood profile.

Coincidence of clines was tested as follows. First, a composite likelihood profile for the cline center position was constructed by summing likelihood profiles for all individual loci. Then the ML value of this composite profile (hereafter $LL\Sigma$) was found and compared with the sum of all MLEs for individual loci (hereafter ΣLL). If the clines are staggered the value of $LL\Sigma$ should be significantly lower than the value of ΣLL according to LRT with $n - 1$ degrees of freedom where n is the number of loci. The same procedure was used for testing the concordance of clines using likelihood profiles for the cline width. As described in Phillips et al. (2004), summing log-likelihoods over loci assumes that these loci are independent, otherwise the tightness of support limits on

consensus parameters may be overestimated. This assumption is fulfilled for the six autosomal markers (excluding *Es2*), which are on a different chromosome each, and for comparisons between the consensus allozyme and X-chromosome clines. Even though individual X loci are all linked on the same chromosome, the rate of recombination between them seems to be sufficiently high for rendering reliable estimates (with a possible exception of two loci, *Tsx* and *DXMit18.2*, which map to the same position on the X chromosome).

ESTIMATION OF DISPERSAL AND SELECTION PARAMETERS

From the estimates of cline shape parameters we can infer several important parameters of process in a contact zone such as selection on marker loci, selection on selected loci, effective selection, number of genes under selection, or mean fitness of hybrids. To do so, we need to make several assumptions, the most restrictive being that fitness is not frequency-dependent and linkage disequilibria are weak ($D_{ij} \ll 1$). Here we use ML estimates of the standardized linkage disequilibrium, $R_{ij} = D_{ij} / \sqrt{p_i q_i p_j q_j}$, averaged across all allozyme loci. We simplified the estimation of linkage disequilibria by neglecting higher-order disequilibria. Values of F_{IS} were estimated from the data rather than assuming them to be zero during the computation of R_{ij} . Only samples of six or more individuals ($N = 1409$ mice from 62 sites) were used for estimating linkage disequilibria (Hatfield et al. 1992).

If we assume that selected loci are randomly scattered over all chromosomes and that the rate of recombination between two loci x map units apart is $[1 - \exp(-2x)]/2$, we can compute the harmonic mean recombination rate among genes (Szymura and Barton 1991; Barton and Gale 1993) according to the equation given in Porter et al. (1997)

$$\frac{1}{\bar{r}} = \frac{2(C-1)}{C} + \frac{C}{R^2} \sum_{y=0}^{R/C} r_0 \cdot \ln \left(\frac{e^{2y} - 1}{2r_0} \right), \quad (6)$$

where C is the effective number of chromosomes (here we take all 20 pairs as the effective number of chromosomes), R/C is the average chromosome length, and r_0 is the minimum distance between a marker locus and a locus under selection (note r_0 is missing in the numerator in eq. 7 of Porter et al. 1997). According to the Ob \times Cast F2 intercross and the Copeland/Jenkins map of a *M. m. domesticus* \times *M. spretus* backcross (Broad Institute, <http://www.broad.mit.edu/cgi-bin/mouse/index>) the mean chromosome length of the mouse is 0.7 Morgans. This value is consistent with the mean number of chiasmata per chromosome estimated by Polani (1972) to 1.35. This gives the average chromosome length of 0.675 M. The summation in equation (6) was done over steps of size $r_0 = 10^{-4}$. This value is arbitrary, but as the result depends only on $\ln r_0$, it does not make much difference if

we use 10^{-3} or 10^{-5} . The resulting harmonic mean recombination rate is then $\bar{r} = 0.402$.

The rate of dispersal can be estimated from the average linkage disequilibrium, recombination rate, and the cline width (Szymura and Barton 1986; Barton and Gale 1993)

$$\sigma = w \sqrt{\frac{R_{ij} \bar{r}}{1 + \bar{r}}}, \quad (7)$$

where σ is expressed as the standard deviation of distances between birthplaces of parents and their offspring measured after migration. The linkage disequilibrium was averaged across samples from the central step segment of the composite cline of six allozyme loci (about center ± 5 km).

Assuming intrinsic selection against hybrid genotypes, the effective selection pressure on a locus at the center of the zone is (Barton and Gale 1993; Raufaste et al. 2005)

$$s^* = 8 \left(\frac{\sigma}{w} \right)^2. \quad (8)$$

If we further assume that selection of strength s acts multiplicatively against heterozygotes at n loci, that the number of selected loci is large and that linkage disequilibrium is generated predominantly by dispersal, then the barrier B can be approximated as $B \approx (n\sigma\sqrt{2s})/\bar{r}$ (Barton and Shpak 2000) and $ns/\bar{r} = 2 \ln(B/w)$ (Barton 1986; Barton and Bengtsson 1986; Szymura and Barton 1986, 1991). From these equations we can derive formula for s and n (Raufaste et al. 2005)

$$s = \frac{8\sigma^2}{B^2} \left[\ln \left(\frac{B}{w\Delta u} \right) \right]^2, \quad (9)$$

where Δu is the height of the central step estimated according to equation (4), and

$$n = \frac{\bar{r} B^2}{4\sigma^2 \ln \left(\frac{B}{w} \right)}. \quad (10)$$

The total selection is then given as $S = ns$.

The strength of the barrier can be expressed in terms of the cline width, mean fitness of hybrids relative to mean fitness of populations outside the zone, and the rate of recombination as $B = w\Delta p (\bar{W}_H/\bar{W}_P)^{-1/\bar{r}}$ (Barton and Bengtsson 1986). If we fix the mean fitness of both parental populations arbitrarily at 1, we get

$$\bar{W}_H = \left(\frac{w\Delta p}{B} \right)^{\bar{r}}. \quad (11)$$

Where $P = 0.5$, a half of individuals will be heterozygous at any one of the n loci so the mean fitness of hybrids can also be approximated as $\bar{W}_H \approx \exp(-S/2)$ (Szymura and Barton 1986, 1991).

Except for the microsatellite data, the Analyse 1.3 program was used to perform all the computations given above. Because

the current version of Analyse does not allow direct processing of haplodiploid data, we used pooled allele frequencies as input in the case of the X-linked markers. However, this step is only possible if male and female frequencies are not significantly different. This assumption was tested with Fisher exact test using Statistica (StatSoft, Inc. 2000). As long as neither of the pairwise tests proved significant differences between male and female frequencies, the two sexes were pooled together for subsequent statistical treatment.

For comparison, the parameters for the X-linked loci were also estimated from individual genotypes using the ClineFit program (Porter et al. 1997; available at <http://www-unix.oit.umass.edu/~aporter/software/>). Several independent runs with changing parameters (burn-in: parameters tries per step ≥ 500 ; sampling for support limits: replicates saved ≥ 3000 , and 30–50 replicates between saves) were performed for each locus. Because this program does not directly estimate the strength of the barrier, B_0 , this parameter was derived from the equation

$$B_0 = \frac{\Delta p}{\left(\frac{\partial A}{\partial x}\right)}, \quad (12)$$

evaluated at x_0 , where Δp is the height of the central step as specified above and x_0 is the position of the intersection between the central step and the left introgression tail, that is, the intersection between equations (1) and (2). The derivative ($\partial A/\partial x$) is given as

$$\frac{\partial A}{\partial x} = \frac{4\sqrt{\theta_0}}{w} \exp\left[\frac{4(x_0 - [c + z_0])\sqrt{\theta_0}}{w}\right], \quad (13)$$

where $c + z_0$ is the distance of a vertical asymptote for the exponential decay on the left side of the zone from the cline center (Porter et al. 1997; but note incorrect typing of parentheses in their eqs. [10a] and [10b]). Analogically, by substituting z_1 for z_0 and θ_1 for θ_0 we can compute B_1 for the right side.

Results

ALLELE AND GENOTYPE FREQUENCIES AT AUTOSOMAL LOCI

Individual genotypes at all autosomal and X-linked loci are available at <http://www.iapg.cz/legs>. The *Es2* locus appeared not to behave as a good diagnostic marker because too many “foreign” alleles occurred far within territories of both taxa (see also Munclinger et al. 2002) and was excluded from subsequent analyses. Allele frequencies of *musculus* alleles averaged across the remaining six autosomal loci (hybrid index, HI_6) as well as five X-linked loci (HI_X) for each population are given in the Appendix.

Populations within the zone harbored mixtures of individuals with complex hybrid ancestry. No individuals heterozygous at all autosomal loci scored (which would indicate presence of F_1 hybrids) were found, nor were any females heterozygous for the five

X-chromosome markers. Introgressed diagnostic alleles appeared at low frequencies even at the edges of the transect, about 60 km from the hybrid zone center.

Two rare alleles occur in the hybrid zone: *Idh*¹⁵⁵ on the *musculus* side (Anenská Ves: $P = 0.036$; Krajková 1: $P = 0.125$; Krajková 2: $P = 0.111$; Nová Role: $P = 0.500$; Rudolec 2: $P = 0.089$; Staré Sedlo: $P = 0.088$); and *Np*¹¹⁰ on the *domesticus* side (Benk: $P = 0.025$; Křižovatka: $P = 0.141$; Neuenreuth: $P = 0.077$; Plesná: $P = 0.211$; Straas 2: $P = 0.015$; Thierstein: $P = 0.139$). This can be an example of the so-called rare allele phenomenon (Sage and Selander 1979); that is, alleles absent in adjacent “pure” populations (called hybridzymes; Woodruff 1989) appear in hybrids (see also Hunt and Selander 1973). Barton and Hewitt (1985) noted that rare alleles have been found in 19 of 23 electrophoretic studies of hybrid zones. According to these authors, such genetic variants are the result of increased mutation rate, intragenic recombination (see also Godinho et al. 2006), or relaxed selection within the zone. However, although frequencies of both *Idh*¹⁵⁵ and *Np*¹¹⁰ tended to increase towards the center of the zone, the slopes of regression lines were not significantly different from zero and occurrence of these alleles was limited to the edges of the zone (i.e., they were missing in populations with hybrid indices between 0.25 and 0.75).

No “pure” mice, those homozygous at all autosomal and X-linked loci, were found on the “wrong” side of the hybrid zone. Presence of such individuals would indicate the occasional long-distance migration. However, a single individual from a site 10.5 km from the center on the *musculus* side was homozygous for *domesticus* alleles at all X-linked and five autosomal loci whereas heterozygous at the sixth autosomal locus. Another individual from the same population was found to possess the “pure” *domesticus* genotype at all X-marker loci. On the *domesticus* side, one animal from a site 9.5 km from the center was found to be homozygous for *musculus* alleles at all five X loci.

At several sites, allele frequencies at one of the loci differed markedly from others, suggesting a possible founder event or establishment of a migrant into an existing population. Although in many populations we could not establish whether this phenomenon was simply an artifact of small sample size, there were few notable exceptions. The first one was the population in Plesná, from which the largest sample was analyzed ($N = 134$). This population ($HI_6 = 0.211$; Appendix) was characterized by a high frequency of the *musculus* allele at *Gpd1* ($p[Gpd1] = 0.961$) on a predominantly *domesticus* background (the value of the hybrid index based on remaining five loci was $HI_5 = 0.079$). A similar situation was found in Dolní Luby, about 6 km east of Plesná ($N = 10$; $HI_6 = 0.202$; $p[Gpd1] = 0.850$, $HI_5 = 0.080$). Even more dramatic examples of suspected founder events were found in several populations within the *musculus* territory, between the 60th and 75th km of the transect, with high frequencies of *domesticus*

alleles at either the *Nt* or *Tsx* (or both) loci and fixed *musculus* alleles at the remaining X-linked loci scored.

ALLELIC AND GENOTYPIC DISEQUILIBRIA AT AUTOSOMAL LOCI

Significant heterozygote deficit was found at all six autosomal loci with a mean value of $\bar{F}_{IS} = 0.409$ (2LL support limit: 0.376–0.442), yet there was significant heterogeneity both among loci ($\Delta L_5 = 17.94$, $P \ll 0.001$) and among samples ($\Delta L_{61} = 106.64$, $P = 0.0003$). The lowest value of F_{IS} was found for *Mpi* (0.285; 2LL = 0.205–0.365) and the highest at *Es1* (0.604; 2LL = 0.492–0.691). We hypothesized that these high F_{IS} values may have been caused by three outliers—Lehsten on the *domesticus* side and Buškovice and Kostelní Bříza on the *musculus* side. In the two former cases, a single individual among genetically pure animals was found to be homozygous for a foreign allele at one locus (Np^{90} in Lehsten, *Es1*⁹⁴ in Buškovice) whereas three animals exhibited two *domesticus* alleles, *Mpi*¹⁰⁰ in the pure *musculus* background in Kostelní Bříza. However, when the five spurious individuals were excluded from the analysis, the overall decrease in F_{IS} was negligible: $\bar{F}_{IS} = 0.405$ (0.372–0.438; heterogeneity among loci: $\Delta L_5 = 18.76$, $P \ll 0.001$; and among samples: $\Delta L_{61} = 46.92$, $P = 0.0044$). Importantly, there is an apparent peak of F_{IS} values in the zone center (Fig. 2A).

Values of the mean standardized linkage disequilibria between pairs of autosomal loci are listed in Table 1. Overall linkage disequilibrium averaged across the samples was $\bar{R}_{ij} = 0.0391$ (2LL = 0.0244–0.0532; heterogeneity among sites: $\Delta L_{61} = 42.77$, $P = 0.0208$; heterogeneity among loci: $\Delta L_5 = 14.06$, $P < 0.001$). As for F_{IS} , there was a peak of high values of R_{ij} in the central part of the zone (Fig. 2B). When only the central populations were considered the average disequilibrium increased to $\bar{R}_{ij} = 0.0584$ (2LL = 0.0385–0.0748; heterogeneity among sites: $\Delta L_{22} = 23.55$, $P = 0.0014$; heterogeneity among loci: $\Delta L_5 = 9.19$, $P = 0.0025$). As “central populations” we considered sites from an area characterized by a steep gradient of allele frequencies, that is, a belt approximately of ± 5 km from the center (see below). This sub-

sample consisted of 23 sites and 478 individuals. Figure 2C shows the disequilibrium plotted against the product of averaged allele frequencies, pq , for each site of $N \geq 6$. The regression line is given by the equation $y = 0.0301 + 0.1965x$ ($R^2 = 0.0282$; slope not significantly different from zero at the 95% and even 85% levels, $P = 0.1690$). This result seems to be due to the three outliers characterized by strong heterozygote deficit at three sites. Indeed, when these outliers are removed, the slope becomes significantly different from zero (regression equation: $y = 0.0023 + 0.3196x$; $R^2 = 0.172$; $P = 0.0011$). The linkage disequilibrium estimated from the regression equation at $pq = 0.25$ is $R_{ij} = 0.0822$.

CLINE SHAPES

Allozyme cline in two dimensions

The allozyme data were best fitted to a cline with the straight line oriented 22° clockwise from the northsouth direction (Fig. 3A). The line runs between site numbers 55, 56, 60, 61, and 62 on the *domesticus* side, and 63, 64, 65, 67, 68, 69, 70, and 71 on the *musculus* side (Fig. 3B). The cross-section of this two-dimensional cline was symmetrical and stepped. The two alternative models as well as more segmented clines did not provide significantly better fits.

Single- and multilocus autosomal clines

All clines were centered within approximately 5 km of one another, between the 45th and 50th km of the transect (Table 2). Sigmoid clines were considerably wider (ca. 24 km on average) than stepped clines (ca. 10 km on average). Clines at each of the autosomal loci are shown in Figure 4 and the parameter estimates are given in Table 2. Three loci (*Es1*, *Mpi*, *Np*) revealed symmetrical (Sstep) clines whereas two loci (*Gpd1*, *Idh1*) gave a significantly better fit with asymmetrical (Astep) clines. The sixth locus (*Sod1*) was best described with the simple sigmoid (Sig) model.

Likelihood profiles for cline widths are shown in Figure 5. This figure illustrates that likelihoods for the sigmoid curves tend to peak at higher estimates of w than those for stepped clines (cf. also Table 2). This is caused by the fact that the Sig model

Table 1. Values of standardized variance of fluctuations about the fitted cline (F_{ST}) and the average standardized linkage disequilibrium between each pair of loci (\bar{R}_{ij}).

Locus	F_{ST}	\bar{R}_{ij}						α	β
		<i>Es1</i>	<i>Gpd1</i>	<i>Idh1</i>	<i>Mpi</i>	<i>Np</i>	<i>Sod1</i>		
<i>Es1</i>	0.0937	–						0.0538	–0.0610
<i>Gpd1</i>	0.3184	0.040	–					0.2435	–0.2387
<i>Idh1</i>	0.1440	0.052	0.013	–				0.1091	0.0601
<i>Mpi</i>	0.1050	0.043	0.043	0.098	–			–0.1230	0.0698
<i>Np</i>	0.1572	0.087	0.000	0.013	0.042	–		–0.0076	–0.1526
<i>Sod1</i>	0.1572	0.050	0.025	0.010	0.077	0.017	–	–0.3005	0.2825
All	0.1693	0.0391							

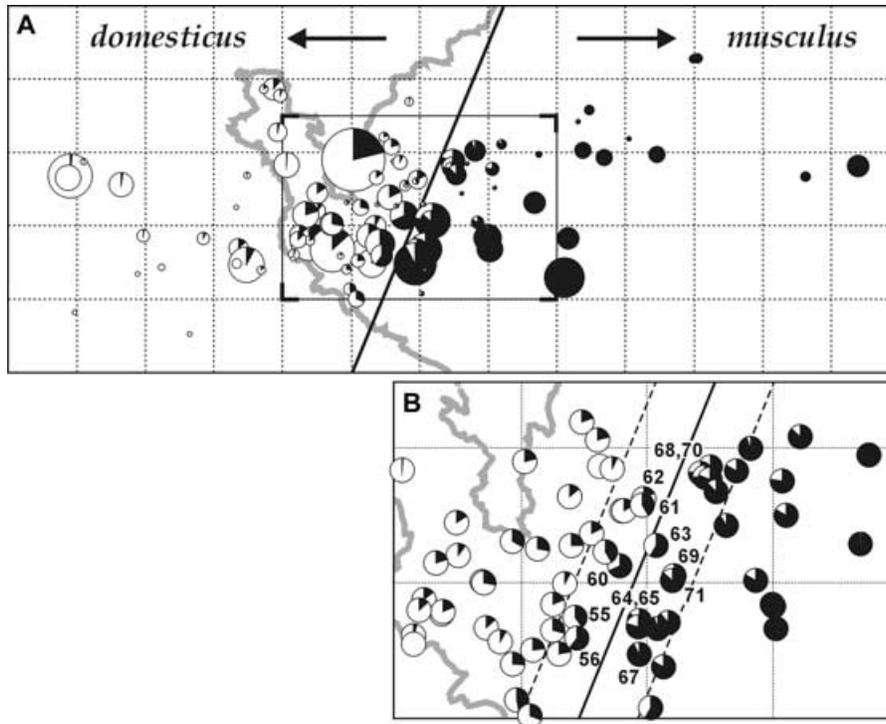


Figure 3. (A) Values of the hybrid index (HI_6), defined as the frequency of *musculus* alleles averaged over six allozyme loci, at each locality. The pies are proportional to sample sizes. The black line indicates the center of the average six-allozyme cline. The data were best fitted with the Sstep model as a straight line (solid line) oriented 22° clockwise from the north–south direction. In the lower right corner of the figure, the width of the cline, together with two-unit support limits, and the log-likelihood value are shown. The grid is $10 \text{ km} \times 10 \text{ km}$. Below, the cline course is shown in more detail with numbers of adjacent sites indicated (B). Note that for clarity, the pies are not proportional to sample sizes. The two dashed lines delimit the central step of the cline (about $\pm 5 \text{ km}$ from the center). Samples falling into this belt were regarded as “central.”

consists of two parameters only and thus is less flexible when fitting the data. Profiles for both stepped models are very similar in their height and peak position in most cases, as expected from the predominantly symmetric pattern of the clines.

When superimposed, the clines appear relatively coincident and concordant (Fig. 6; note that only a part of the transect, between the 20th and 80th km, is shown to render the pictures more informative). However, when the allele frequencies at each of the six loci (p_i) were plotted against the average frequency (\bar{p}) the points were largely scattered around the diagonal suggesting considerable deviations of individual frequencies from the mean (figure not shown). We can fit polynomials to these points according to the formula $p_i = \bar{p} + 2\bar{p}\bar{q}[\alpha + (\bar{p} - \bar{q})\beta]$ (Szymura and Barton 1986, 1991), where α describes an increase of *musculus* alleles above the average (i.e., a shift of the cline to the *domesticus* territory) and β describes narrowing the cline below the average (Table 1). These polynomials were less coincident on the western than on the eastern side of the hybrid zone with *Idh1* ($\alpha = 0.1091$) and *Gpd1* ($\alpha = 0.2435$) most markedly shifted to the *domesticus* territory and *Sod1* ($\alpha = -0.3005$) shifted to the *musculus* territory (cf. Table 1).

Coincidence and concordance of clines was tested using the LRT where the null (constrained) hypothesis is represented by MLE of summed (composite) likelihood profiles ($LL\Sigma$) and the alternative (unconstrained) hypothesis is represented by the sum of MLEs for individual loci (ΣLL) for the cline center position and width, respectively. LRTs were performed for all three cline models yet for simplicity we present here only tests for the most parameter-rich Astep model. The autosomal clines appeared very similar in the position and width: all the clines have coincident estimates of c with the exception of *Sod1* ($\Delta L_5 = 14.71$, $P \ll 0.001$; without *Sod1*: $\Delta L_4 = 0.54$, $P = 0.8962$).

The six-locus autosomal cline was best described by a symmetrical step model with $w = 9.62 \text{ km}$ ($2LL = 0.0\text{--}16.9$). This symmetry was also suggested by t -tests of B_0 versus B_1 values and θ_0 versus θ_1 values, respectively (B : $t_{10,\alpha} = 1.2263$, $P = 0.2482$; θ : $t_{10,\alpha} = 0.7039$, $P = 0.4976$). According to equation (7) and given the average standardized linkage equilibrium from the zone center $\bar{R}_{ij} = 0.0584$ and harmonic recombination $\bar{r} = 0.402$, the multilocus cline width yields an estimate of the $\sigma = 1.0472 \text{ km/gen}^{1/2}$. This estimate is somewhat higher than the average dispersal estimated from microsatellite genotypes

Table 2. Cline shape parameter estimates for each autosomal locus and all six loci. The parameters were estimated for three models: sigmoid (Sig), symmetrical stepped (Sstep), and asymmetrical stepped (Astep); the best-fit model is indicated with an asterisk. For the Sstep model parameters are B_0/w and θ_0 the same for both sides of a line, for the Astep model B_0/w and θ_0 denote parameters for the western, *domesticus*, side whereas B_1/w and θ_1 describe the eastern, *musculus*, side. Two-unit support limits are given in parentheses for each estimate.

Locus	Model	LnL	c (km)	w (km)	B_0/w	θ_0	B_1/w	θ_1
<i>EsI</i>	Sig	-68.118	46.05 (44.7-47.4)	22.51 (19.0-26.9)	-	-	-	-
	Sstep	-52.477*	47.02 (46.0-47.8)	10.67 (6.4-14.5)	5.93 (3.1-39.9)	.0804 (.009-.341)	-	-
	Astep	-50.106	47.00 (46.1-50.6)	9.92 (0-13.6)	3.64 (.8-32.2)	.0881 (.006-.323)	5.68 (1.0-inf.)	.1112 (.000-.663)
<i>GpdI</i>	Sig	-39.772	46.12 (44.1-50.6)	20.97 (16.0-28.0)	-	-	-	-
	Sstep	-31.966	46.90 (45.3-48.6)	6.07 (0-14.6)	6.89 (.7-inf.)	.0403 (.000-.408)	-	-
	Astep	-28.810*	47.27 (46.0-50.8)	7.95 (0-12.4)	3.63 (0-32.2)	.0919 (.016-.341)	1605.28 (9.4-inf.)	3.126×10^{-5} (6.3×10^{-7} -239)
<i>IdhI</i>	Sig	-75.701	44.90 (43.2-46.6)	25.80 (19.7-34.7)	-	-	-	-
	Sstep	-65.015	47.41 (45.9-48.8)	10.72 (0-16.2)	6.65 (1.0-46.1)	.0521 (.007-.326)	-	-
	Astep	-49.788*	48.31 (46.9-50.0)	6.43 (0-10.8)	5.31 (.1-25.0)	.0171 (.000-.079)	1.29 (.1-inf.)	.6392 (.000-1.000)
<i>Mpi</i>	Sig	-77.375	48.07 (46.7-49.5)	22.78 (19.0-27.5)	-	-	-	-
	Sstep	-63.650*	48.52 (47.4-49.7)	9.29 (3.1-14.3)	3.20 (6-12.5)	.1008 (.013-.377)	-	-
	Astep	-62.741	48.72 (47.2-51.5)	9.64 (0-13.2)	3.50 (.7-19.4)	.0822 (.009-.339)	2.41 (.3-25.6)	.1937 (7.0×10^{-4} -1.000)
<i>Np</i>	Sig	-85.406	48.08 (46.0-50.2)	33.65 (27.4-42.0)	-	-	-	-
	Sstep	-78.396*	48.52 (47.2-50.8)	10.22 (0-19.5)	1.93 (.2-16.5)	.0758 (.001-.453)	-	-
	Astep	-77.654	49.95 (47.8-51.1)	10.81 (0-22.0)	2.28 (.1-10.0)	.0743 (.016-.450)	7.09 (.1-inf.)	.0471 (5.0×10^{-4} -1.000)
<i>SodI</i>	Sig	-55.478*	50.12 (48.8-51.6)	18.07 (14.6-22.6)	-	-	-	-
	Sstep	-55.478	50.12 (48.8-51.6)	18.07 (0-22.6)	339.17 (0-inf.)	.9980 (.000-1.000)	-	-
	Astep	-55.478	50.12 (48.8-51.6)	18.07 (14.6-22.6)	289.24 (0-inf.)	.7795 (.000-1.000)	692.87 (0-inf.)	.9895 (.000-1.000)
All	Sig	-17.719	47.18 (45.3-49.1)	23.96 (19.0-30.8)	-	-	-	-
	Sstep	-10.909*	47.85 (46.3-49.4)	9.62 (0-16.9)	2.62 (.2-15.5)	.1083 (.010-.682)	-	-
	Astep	-9.918	47.97 (46.6-51.5)	9.89 (0-16.4)	2.54 (.2-25.1)	.0906 (.001-.580)	1.73 (4.5×10^{-2} -inf.)	.2289 (.003-.500)

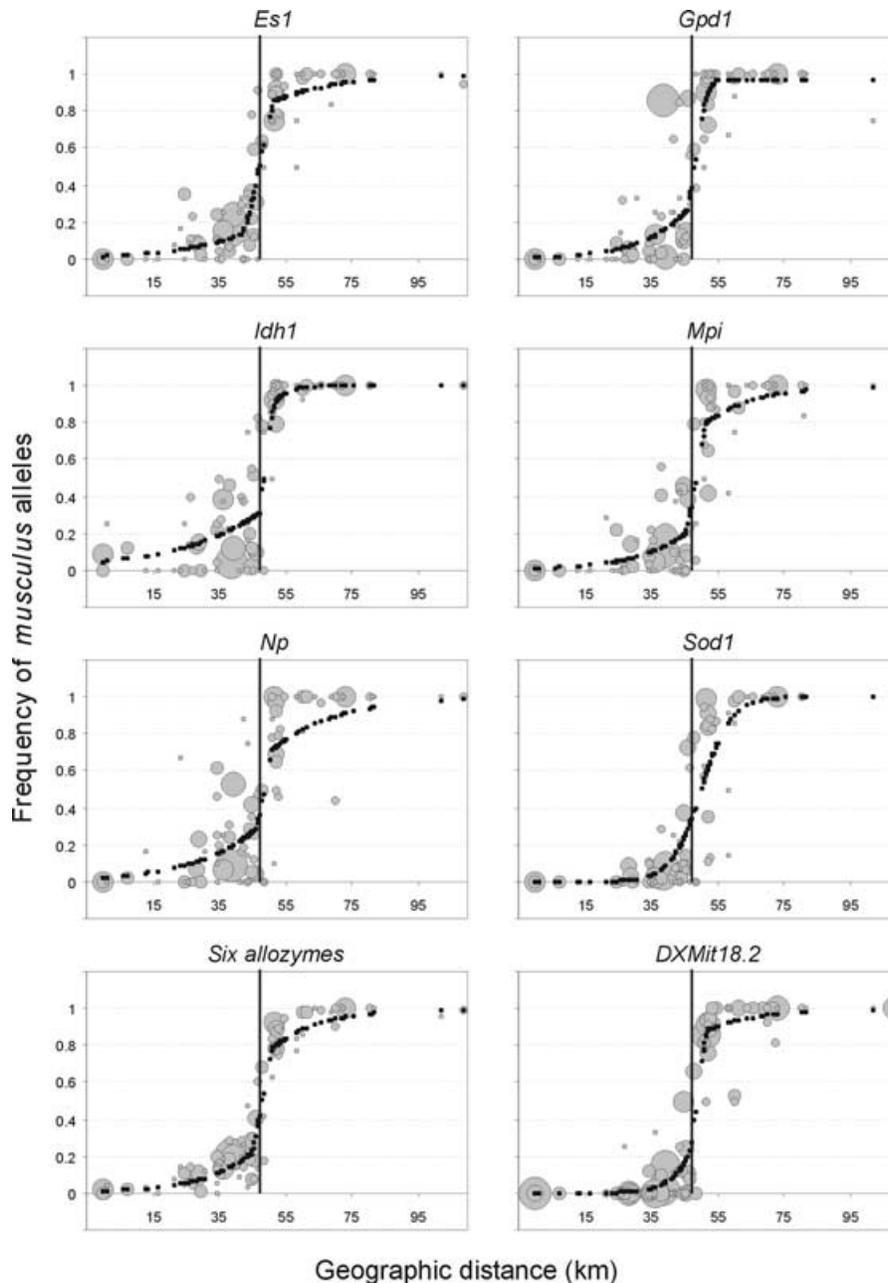


Figure 4. Frequencies of *musculus* alleles at six autosomal loci, their mean, and those at five X-linked loci, plotted against geographic distance along the one-dimensional transect across the hybrid zone. Gray bubbles represent real data (the area of each bubble is proportional to sample size) whereas the black dots show the best fit of the same data based on one of the three cline models used in this paper. Vertical lines depict the position of the average six-allozyme cline. *Nt*(-) and *Tsx*(-) denote the data with outliers excluded.

($\sigma = 0.8172 \text{ km/gen}^{1/2}$; see Supplementary Material available online for details).

X-chromosome clines

The transitions of individual X-chromosome markers and fitted clines are shown in Figure 4. All the clines were significantly asymmetrical. This asymmetry is most conspicuous in the *Nt* and *Tsx* loci, due to low frequencies of *musculus* alleles at several eastern sites, between 60th and 75th km of the transect, within

the *musculus* territory. Importantly, high incidence of *domesticus* alleles either at the *Nt* or *Tsx* (or both) loci occurred in populations with fixed *musculus* alleles at the remaining loci, suggesting recent founder events after human-mediated long-distance migration (see above). Therefore, samples with high frequencies of *domesticus* alleles on fixed *musculus* background were excluded as outliers for the two loci (*Nt*: nos. 71, 72, 88, 93, 97–99; *Tsx*: 72, 88, 89, 97, 99). These reduced datasets are hereafter denoted as *Nt*(-) and *Tsx*(-), respectively. When outliers are removed, both

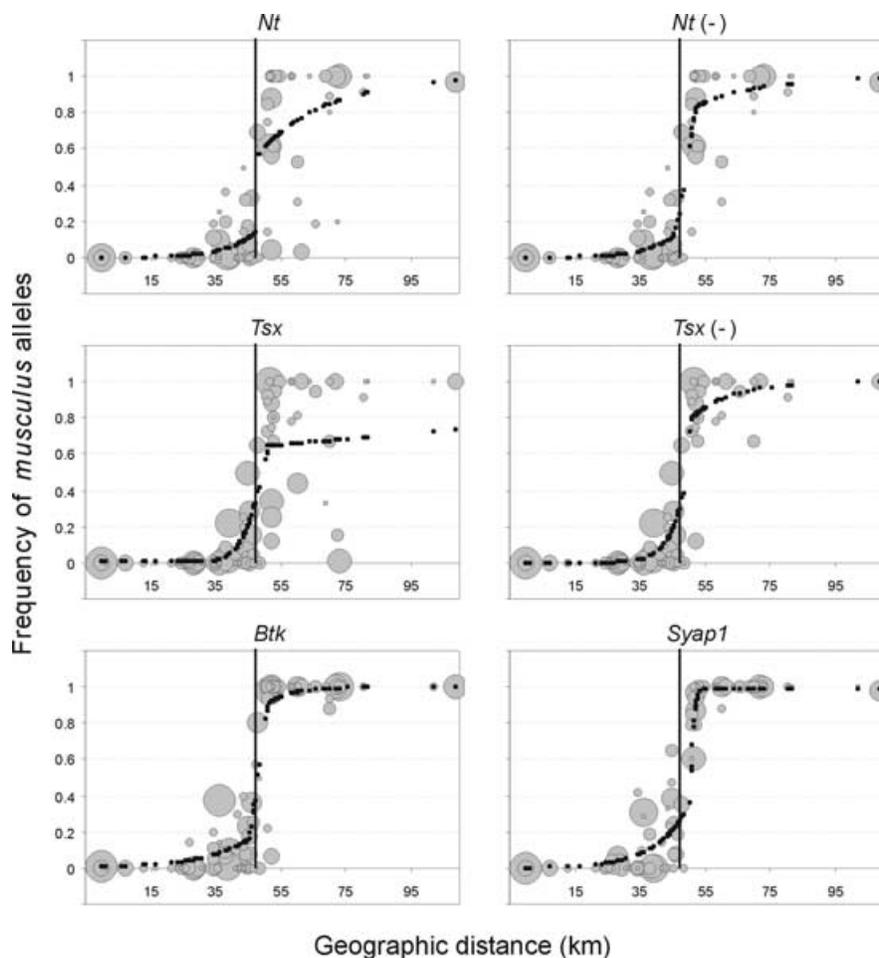


Figure 4. Continued.

the *Nt(-)* and *Tsx(-)* clines are much more symmetrical, but the Astep model still fit the data significantly better than the Sstep model (cf. Table 3).

After superposition, all the X-marker clines are apparently very similar to each other both in their position and their shape regardless of the model used (Fig. 6, bottom row). The Astep model revealed high coincidence, with only the cline center for *Btk* significantly different from that of the others ($\Delta L_3 = 12.27$, $P \ll 0.001$; excluding *Btk*: $\Delta L_3 = 3.90$, $P = 0.0504$). Discordance of the clines was caused by the significantly narrower *Syap1* cline (Astep: $\Delta L_4 = 5.22$, $P = 0.0337$; without *Syap1*: $\Delta L_3 = 2.34$, $P = 0.1963$). Indeed, the width of the *Syap1* cline was almost six times lower than that of *Nt(-)* or almost 10 times lower than the *Tsx* cline uncorrected for outliers. The compound five-locus X-chromosome cline was best described by an asymmetrical step model with $w = 4.61$ km (2LL = 3.7–5.0).

Comparison of multilocus autosomal and X-chromosome clines

When w is estimated from a multilocus cline and cline centers for particular loci are noncoincident, the resulting value of w and,

hence, σ may be overestimated (see Porter et al. 1997). In our study, the compound autosomal cline width may be overestimated due to higher w and different position of the center of the *Sod1* cline relative to remaining loci and thus it could be argued that we should exclude this locus from fitting a multilocus cline (see also Dod et al. 2005; Raufaste et al. 2005). However, excluding *Sod1* revealed virtually the same w estimate for the best-fit Sstep model ($w = 9.52$ km, 2LL = 1.2–16.2); thus there is no reason to exclude this marker from the analysis. Both step models yielded coincident autosomal and X-chromosome clines, that is, although the stepped multilocus clines for the five analyzed X-chromosome markers were apparently much narrower than the compound autosomal clines fitted with the same models (about 35% and about 47% of the autosomal cline width for the Sstep and Astep model, respectively; cf. Tables 2 and 3), the differences were insignificant.

COMPARISON OF ANALYSE AND CLINEFIT RESULTS

To allow the comparison of the results of the two programs, we used equations (12) and (13) to obtain estimates of parameters B_0 and B_1 from the ClineFit output. The ML estimates of the six

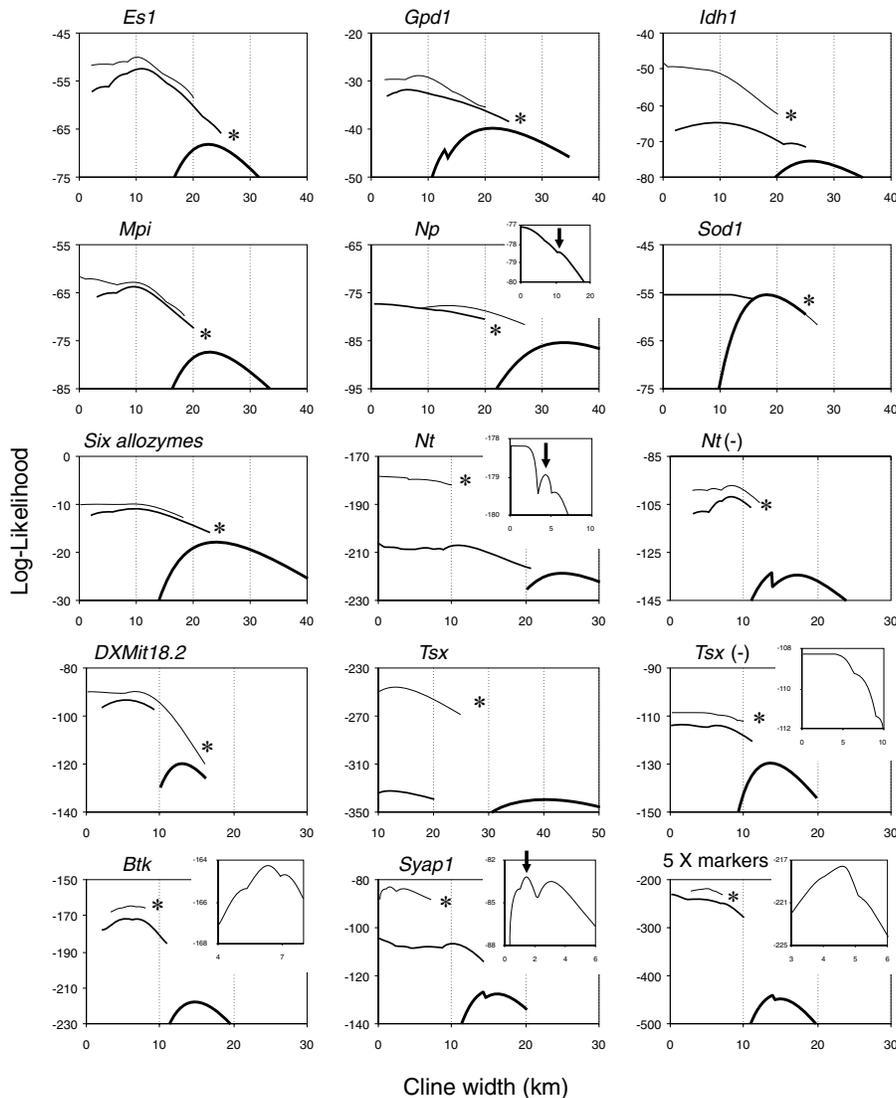


Figure 5. Marginal likelihood surfaces along the cline width (w) axis for all autosomal and X-linked markers as well as for average allozyme and X-chromosome data. The profiles are made of maximum-likelihood values with w fixed in 0.1-km increments, whereas all other parameters are free to vary at each step. All three models are compared for each dataset and the best-fitting model, chosen using LRT, is indicated with an asterisk (cf. also Tables 3 and 4). Inserts show details of the profiles and arrows point to local likelihood peaks that can be considered “real.”

parameters and their two-unit support limits are given in Table S1 (see Supplementary Material available online). As expected, the highest congruence was found in the estimates of the parameter c : the difference was 18.6 m for the Sig model and 301.4 m for the Astep model, representing 0.097% and 1.574% of the average cline width, respectively. The congruence in the estimates of cline widths was slightly lower: the difference between Analyse and ClineFit estimates were 1.85 km for Sig and 1.51 km for Astep, representing 9.67% and 26.21% of the average cline width, respectively. Estimates of B and theta parameters were considerably less congruent, ranging from 132.04% (θ_0) to 194.67% (B_1) of average values of the respective parameters.

ESTIMATING SELECTION PARAMETERS FROM CLINE SHAPES

The strength of the central barrier to gene flow into the *domesticus* territory estimated both from the multilocus autosomal and X-chromosome cline was approximately 20 km (Table 4). In Table 4, we used both dispersal rate estimates (σ) for the allozyme data, that is, σ based on the cline parameters (eq. 7) as well as σ estimated independently from the microsatellite data (see Supplementary Material available online); for X-chromosome data only the latter estimate was used. Both the clines were fitted with the Astep model to allow their comparison. On the *musculus* side, the barrier estimated from the autosomal data was also about 20 km whereas

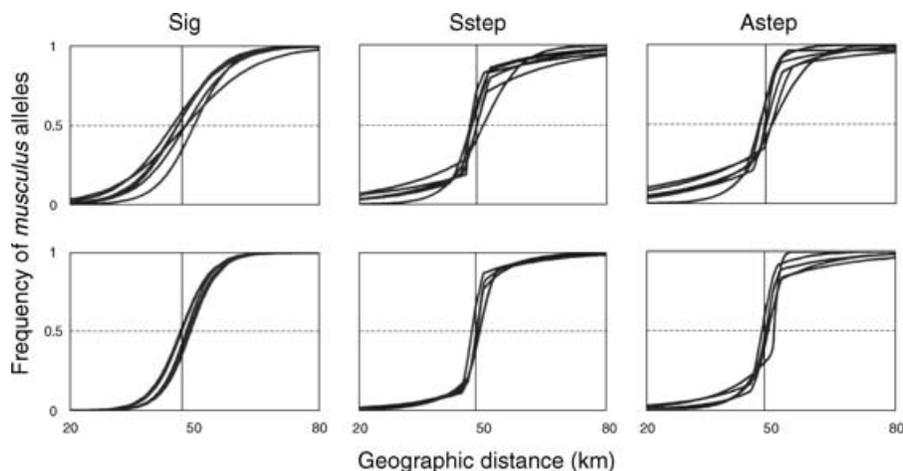


Figure 6. Superimposed Sig, Sstep, and Astep clines for six autosomal (top) and five X-linked (bottom) loci. Note that only central parts of clines (between the 20th and 80th kilometer of the transect) are shown.

the multilocus cline based on the five X-linked markers gave an estimate six times higher (124 km). These parameters can be interpreted as the distance a neutral allele would introgress without the barrier. This barrier can also be interpreted as the time (in generations) that a neutral allele is delayed when crossing the zone: $T_0 = (B_0/\sigma)^2$ on the *domesticus* side and $T_1 = (B_1/\sigma)^2$ on the *musculus* side, respectively (Barton and Gale 1993). This delay was of the order of few to many hundreds of generations except for X-linked loci, where the time required to reach the *musculus* side exceeded 20,000 generations (Table 4). This means that given two generations per year, the average autosomal gene that has crossed the barrier in either direction will have taken about 160 years when the higher dispersal estimate is used or about 400 years when the lower estimate is considered. On the other hand, diffusion of an average X-linked gene would take roughly 2500–4000 years.

Raufaste et al. (2005) used a more conservative estimate of a single generation per year assuming that the autumn generation does not contribute to migration. However, it is known that in temperate zones of Central Europe some mice continue to breed throughout the year (Pelikán 1981) so there can be up to three generations per year. Even though it is doubtful that animals born in late autumn and during winter move outdoors, the second generation of late summer/early autumn is likely to contribute to migration.

The effective selection pressure on an average enzyme locus in the center of the zone was $s^* \approx 5.6\text{--}9.0\%$, whereas the effective selection on the X chromosome was 3.5-times higher. Outside the central part of the zone, the selection acting directly on each allele was $s_e \approx 0.8\text{--}1.4\%$ for autosomes and $s_e \approx 1.1\%$ for X-linked markers. Finally, the selection against hybrids acting at loci responsible for reproductive isolation (s) was $s \approx 2.8\text{--}4.9\%$ for autosomal and about 2.7% for X-chromosome markers. The estimates of the number of loci under selection were between 56

and 99 for autosomes and almost 380 for the X chromosome. The mean fitness of hybrids in the center of the zone relative to the mean fitnesses of pure populations outside the zone (eq. 11), when the fitnesses of *domesticus* and *musculus* were considered $\bar{W}_D = \bar{W}_M = 1$, was $\bar{W}_H \approx 60\%$ for autosomal and $\bar{W}_H \approx 35\%$ for the X-linked loci.

Discussion

CLINE SHAPE AND SELECTION AGAINST HYBRIDS IN THE MOUSE HYBRID ZONE

The central European portion of the *musculus/domesticus* hybrid zone, studied across a two-dimensional field area stretching from eastern Bavaria (Germany) to western Bohemia (Czech Republic), is best modeled as a single straight line oriented 22° clockwise from the north–south direction. This simple pattern, independent of fine-scale local geographic conditions, is consistent with the synanthropic niche of the house mouse and absence of any climatic or other environmental gradient throughout the study area. It is also consistent with the notion of the zone as being maintained by dispersal and selection against hybrids (so-called tension zone; Key 1968) rather than by extrinsic habitat selection. The position and orientation of the hybrid zone in our study area matches fairly well with those in other sections of the zone near Munich in southern Germany (Sage et al. 1986; Tucker et al. 1992) and that east of Leipzig in eastern Germany (K. C. Teeter, pers. comm.). Thus, the zone appears to run in a straight line from Munich through the western tip of the Czech Republic to the Elbe River where it turns north–northwest to Kiel Bay in East Holstein (Prager et al. 1993).

Changes in allele frequencies along a one-dimensional transect in five of the six allozyme loci and all five X-linked loci studied were best fitted with one of the stepped cline models.

Table 3. Cline shape parameter estimates for each X-chromosome locus and the average X-linked locus.

Locus	Model	LnL	c (km)	w (km)	B_0/w	θ_0	B_1/w	θ_1
<i>Nr</i>	Sig	-218.861	52.52 (51.5-53.6)	25.21 (22.4-28.5)	-	-	-	-
	Sstep	-207.232	50.24 (49.6-51.5)	11.44 (1-14.4)	3.45 (7-10.3)	.1365 (.040-.380)	-	-
	Astep*	-178.921*	47.64 (46.7-48.9)	4.35 (1-7.6)	8.41 (1.6-inf.)	.0462 (.015-.300)	4.69 (1.8-inf.)	.0110 (1.35×10^{-4} -.045)
<i>Nr(-)</i> ^a	Sig	-133.364	49.39 (48.750.4)	13.65 (12.9-13.8)	-	-	-	-
	Sstep	-101.650	49.40 (48.8-49.9)	8.13 (6.9-9.8)	8.05 (2.6-27.9)	.0841 (.020-.238)	-	-
	Astep*	-97.080*	49.30 (47.7-48.6)	8.13 (6.8-10.0)	3.59 (.8-20.0)	.2177 (.050-.730)	7.95 (2.39-89.8)	.0417 (.004-.152)
<i>DXMit</i>	Sig	-119.695	48.32 (47.7-49.0)	12.90 (11.4-14.6)	-	-	-	-
	Sstep	-93.471	48.53 (47.8-49.3)	5.16 (2.3-8.3)	3.51 (1.1-11.1)	.1124 (.020-.380)	-	-
	Astep*	-89.539*	48.68 (47.8-52.0)	6.77 (1-8.6)	1.30 (.3-6.0)	.3853 (.010-.970)	8.09 (2.0-55.0)	.0505 (.004-.182)
<i>Tsx</i>	Sig	-339.964	56.02 (54.4-57.8)	40.01 (35.3-45.7)	-	-	-	-
	Sstep	-332.555	50.65 (49.8-51.3)	11.92 (9.2-15.5)	6.41 (4-18.3)	.0376 (.009-.112)	-	-
	Astep*	-245.758*	49.31 (48.4-50.2)	13.23 (11.0-15.8)	494.75 (49.9-inf.)	.0054 (.004-1.000)	26.15 (7.28-inf.)	.0011 (2.10×10^{-6} -.011)
<i>Tsx(-)</i> ^a	Sig	-129.620	48.84 (48.1-49.6)	13.62 (11.6-15.7)	-	-	-	-
	Sstep	-113.958	48.84 (48.1-50.1)	6.13 (0-8.5)	1.67 (4-inf.)	.1984 (.001-.562)	-	-
	Astep*	-108.276*	48.96 (48.0-50.3)	3.48 (0-8.8)	1.32 (.1-inf.)	.1304 (.010-1.000)	8.05 (1.2-inf.)	.0139 (.000-.150)
<i>Btk</i>	Sig	-217.672	46.95 (46.3-47.5)	14.73 (13.3-16.4)	-	-	-	-
	Sstep	-171.780	47.49 (47.0-48.3)	5.34 (3.7-8.1)	8.75 (3.5-21.6)	.0524 (.017-.161)	-	-
	Astep*	-164.230*	47.81 (47.2-48.4)	6.36 (4.4-8.2)	7.43 (2.7-20.2)	.0533 (.016-.160)	5.45 (1.2-22.0)	.1847 (.040-.685)
<i>Syp1</i>	Sig	-126.199	47.24 (46.5-47.9)	14.14 (13.3-14.3)	-	-	-	-
	Sstep	-106.440	48.04 (47.4-48.7)	9.90 (6.8-11.9)	5.32 (1.3-25.5)	.1700 (.040-.484)	-	-
	Astep*	-83.155*	51.40 (49.7-52.0)	1.40 (5-5.1)	3.59 (1.7-15.0)	.0085 (.003-.063)	2.80×10^5 (210-inf.)	3.283×10^{-8} (.000-.010)
All ^b	Sig	-441.114	48.02 (47.7-48.4)	13.96 (13.5-14.1)	-	-	-	-
	Sstep	-240.437	48.80 (48.5-49.4)	3.33 (1-5.6)	9.84 (5.2-inf.)	.0253 (.016-.074)	-	-
	Astep*	-217.598*	49.56 (49.4-49.3)	4.61 (3.7-5.0)	4.35 (2.9-6.4)	.0669 (.039-.098)	26.93 (12.7-65.4)	.0181 (.007-.039)

^a(-)^c denotes data with outliers excluded (see text for explanation).

^bIn this case, the *Nr(-)* and *Tsx(-)* loci were used together with *DXMit*, *Btk*, and *Syp1* for fitting the average cline. Two-unit support limits are given in parentheses.

Table 4. Parameters estimated from the multilocus autosomal and X-chromosome cline, respectively, fitted with the Astep model in both cases. For autosomal data, the dispersal was estimated either from equation (7) or based on the distribution of microsatellite alleles, whereas the latter estimate was only used for the X-chromosome data. All selection parameters and the mean fitness of hybrids were averaged for the left and right sides of the zone.

Parameters	Autosomes (six loci)		Chromosome X (five loci)
	Cline shape	Microsatellites	
Standardized linkage disequilibrium, R_{ij}	.0584 (.039–.075)	–	–
Harmonic mean recombination rate, r		.402	
Cline width, w (km)	9.89 (5.69–15.04)		4.61 (3.71–5.00)
Barrier to flow into <i>domesticus</i> , B_0 (km)	25.14 (5.18–109.54)		20.05 (14.35–27.88)
Barrier to flow into <i>musculus</i> , B_1 (km)	17.14 (7.53–50.29)		124.01 (18.06–329.87)
Introgression tail (<i>domesticus</i> side), θ_0	.0906 (.0229–.2662)		.0669 (.0393–.0980)
Introgression tail (<i>musculus</i> side), θ_1	.2289 (.020–.4693)		.0181 (.0066–.0395)
Dispersal rate, σ (km.gen ^{-1/2})	1.047	.817	.817
Delay on the <i>domesticus</i> side, T_0 (gen.)	577.6 (23.3–12,677.5)	1025.9 (56.5–30,748.8)	602.0 (308.2–1163.5)
Delay on the <i>musculus</i> side, T_1 (gen.)	268.5 (15.1–2.0 × 10 ⁵)	556.2 (36.6–4.9 × 10 ⁵)	23,029.5 (5272.3–1.6 × 10 ⁵)
Effective selection, s^*	.0895 (.0391–.2842)	.0564 (.0236–.1649)	.2517 (.2139–.3913)
Selection on marker loci, s_e	.0143 (.0027–.0353)	.0081 (.0013–.0216)	.0107 (.0080–.0144)
Selection on selected loci, s	.0488 (.0049–.5732)	.0278 (.0019–.2417)	.0272 (.0191–.0440)
Number of loci under selection, n	56 (0–545)	99 (0–1435)	378 (128–1952)
Total selection on hybrids, S	2.716 (0–164.562)	2.752 (0–164.562)	10.25 (4.349–45.967)
Mean fitness of hybrids, \bar{W}_H (%)	59.26 (26.37–90.01)		35.23 (29.82–41.97)

The stepped cline shape may be explained by three nonexclusive circumstances. First, the central step could be caused by a geographic barrier inhibiting or impeding migration. Such a barrier was, for example, found to affect the cline shape in the hybrid zone between two chromosome races of the grasshopper *Podisma pedestris* in the Alpes Maritimes, France (Barton and Gale 1993). In the *M. musculus* contact zone, the presence of a geographic barrier has been reported from Bulgaria (Vanlerberghé et al. 1986, 1988a), southern Germany (Sage et al. 1986), and Denmark (Raufaste et al. 2005).

Coincidence of a hybrid zone position with a geographic barrier is in agreement with the theoretical assumptions of the tension-zone model, which predicts that the zone be trapped either at a physical barrier or in an area of low population density. However, sometimes it is questionable whether a physical obstacle per se can explain cline shape and the strength of barrier to gene flow. For instance, in the case of the *Podisma* hybrid zone, the barrier strength of two small streams at Lac Autier and near Col de la Lombarde was estimated 150 and 95 m, respectively, yet the barrier estimated from cline shapes was significantly stronger: 1700 and 1240 m, respectively (Barton and Gale 1993). The geographic barriers were thus augmented by selection against hybrids.

In the Czech-Bavarian transect across the mouse hybrid zone studied here, the widest body of water is the Ohře (Eger) River. However, this river flows with the direction of gene flow between the taxa, rather than cutting across it, so it can hardly act as a

barrier to gene flow (see Fig. 1). On the other hand, there are two left-bank tributaries of the Ohře River near the zone center, which run in a north-south direction, that is, perpendicular to the main axis of the transect. On the western stream (Fig. 1), there is a water reservoir surrounded by steep slopes covered by mixed forests; the valley of the eastern stream is characterized by steep, up to 100-m high, slopes covered by woods and substantially limited connections between both sides both for mice and humans. Whether these geographic barriers affect the cline shape will be tested elsewhere.

A second possible explanation for the steep central segments of the clines revealed in this study is epistatic interactions among loci. According to Barton and Gale (1993), fitness under epistasis can be described as $W(\bar{p}) = 1 - s[4\bar{p}(1 - \bar{p})]^\beta$, where \bar{p} is the mean frequency of an allele (*domesticus* in this case) at a given site. When β is large ($\gg 1$) and a moderate-to-large number of loci interact, only individuals with nearly F₁-hybrid genotypes from the central area suffer reduced fitness whereas backcrosses are almost as fit as “pure” individuals from parental populations, causing the cline to be shallow at the edges, where animals with 50% mean frequency are rare, and steep in the center. The shape of the multilocus cline based on six allozyme loci corresponds to $\beta > 4$ (fig. 2-2c in Barton and Gale 1993). However, although one can imagine physiological pathway(s) that would embrace several to many loci interacting in their effect on fitness it is difficult to model the precise interactions that may be relevant to two given genomes meeting in a particular contact zone.

A third explanation seems to be the most likely: a permanent influx of parental genotypes into a hybrid zone causes non-random associations between allele combinations, or linkage disequilibria. When a moderate-to-large number of loci are involved, dispersal generates stronger disequilibria than epistasis (Barton 1983); this is expected also from the fact that dispersal is a first-order effect whereas epistatic selection is (at most) second order. Because linkage disequilibrium is proportional to the gradient in allele frequency it increases towards the zone center. Strong associations among loci cause selection pressure on them to be higher in the center than at the edges where the loci are affected separately. This synergistic effect is manifested by a steep central part of the cline. Increased linkage disequilibria have been found in hybrid zones of various species, including *Uroderma* (Barton 1982), *Caledia* (Shaw et al. 1985), *Rana* (Kocher and Sage 1986), *Bombina* (Szymura and Barton 1986, 1991), *Gryllus* (Rand and Harrison 1989), *Heliconius* (Mallet et al. 1990), *Pontia* (Porter et al. 1997), *Carlia* (Phillips et al. 2004), and *Mus* (Payseur et al. 2004; Raufaste et al. 2005). Nonrandom associations are also suggested by the correlations between unrelated traits that cannot be explained by pleiotropic effects. For example, such correlations have been found between genotype at enzyme loci and mating calls or belly color in *Bombina* (Szymura and Barton 1991) or between genotype and diagnostic morphological traits in *Mus* (M. Macholán, unpubl. results).

Porter et al. (1997) give an alternative explanation for the occurrence of stepped clines: strong selection acting on a small number of loci. In this case, the loci under selection would have steep, narrow clines without introgression tails whereas neutral or nearly neutral marker loci would recombine away from the selected loci resulting in shallow and long introgression tails. In the case of the mouse hybrid zone, this could be true for the Y chromosome that is characterized by an abrupt transition between the subspecies with no or negligible introgression (Vanlerberghe et al. 1986; Dod et al. 1993, 2005). However, the clines for most of the X-linked markers analyzed, although apparently narrower than the clines of autosomal loci, were shown to resemble the latter in having rather long and shallow introgression tails (Fig. 4). Although it is possible that the recombination rates between the marker and selected X-linked loci are sufficient to escape strong selection pressure, the apparently high number of selected X-linked loci (although probably overestimated, see below) does not seem to be consistent with this hypothesis (Table 4).

The estimates of effective selection and the number of selected X-linked loci ($s^* \approx 0.25$; $N \approx 380$) were much higher than those of the autosomes ($s^* \approx 0.06$ – 0.09 ; $N = 56$ – 99) in agreement with the notion of the X chromosome as being involved in reproductive isolation (Gruha 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). Indeed, although the

difference between the width of the compound allozyme and X-chromosome cline was insignificant, the former was two to three times wider than the latter (cf. Tables 2 and 3), consistent with predictions of the “large X-effect” hypothesis (Orr 1997; Coyne and Orr 2004) and corroborating findings of limited introgression of X-linked markers in the German and Danish portions of the mouse hybrid zone (Tucker et al. 1992; Dod et al. 1993). On the other hand, the great differences in cline width among individual X-linked loci (even after correcting for outliers; Table 3) suggest differential selection acting on different parts of the X chromosome (see also Payseur et al. 2004; Storchová et al. 2004; Payseur and Nachman 2005; Harr 2006).

The number of X-linked loci under selection may seem too high. For example, Payseur et al. (2004) argue that the number of selected loci in this hybrid zone is probably moderate given the incompleteness of reproductive isolation and recent divergence of the two hybridizing taxa. We should keep in mind that all our estimations were based on an assumption that epistatic interactions between loci can be neglected. Even though our information on epistasis is very limited, several studies indicate that such interactions occur between X-linked genes (Oka et al. 2004; Storchová et al. 2004) and between these genes and autosomal loci (Forejt 1981, 1996; Montagutelli et al. 1996; Britton-Davidian et al. 2005; Payseur and Hoekstra 2005). Therefore, the number of X-linked loci under selection in the Czech-Bavarian hybrid zone is probably overestimated.

On the other hand, from equations (9) and (10) it follows that estimates of both parameters, s and n , are strongly influenced by the square of the barrier (B^2). The average barrier for the five X-linked loci studied here was about 3.5 times stronger than the barrier for autosomes (Table 4). However, the reliability of B and θ estimates is compromised by the fact that they are negatively correlated (Fig. 7; see also Szymura and Barton 1986; Dod et al. 2005) so we can arrive at the same or very similar likelihood value with large B and small θ as with small B and large θ . Therefore, estimates of these parameters should be taken with caution.

Our estimates of selection, number of selected loci, and fitness of hybrids are comparable to those from Denmark based on eight autosomal loci ($s^* \approx 0.03$ – 0.07 ; $s \approx 0.01$ – 0.04 ; $N = 46$ – 120 ; $\overline{W}_H \approx 45\%$; Raufaste et al. 2005). The lower estimate of hybrid fitness from Denmark can arise from the slightly different estimation methods used by Raufaste et al. (2005); indeed, if we approximate \overline{W}_H as $\approx \exp(-S/2)$ we get a much lower value ($\overline{W}_H \approx 25\%$). However, as pointed out by Raufaste et al. (2005), the estimates of selection, the number of selected loci, and fitness of hybrids are highly derived and quite heavily dependent on model assumptions so that they should be considered only approximate. Overall, low values of hybrid fitness estimated in this study are consistent with more direct measures. For example, hybrid males have been shown to suffer high rates of infertility (reduced testis size and

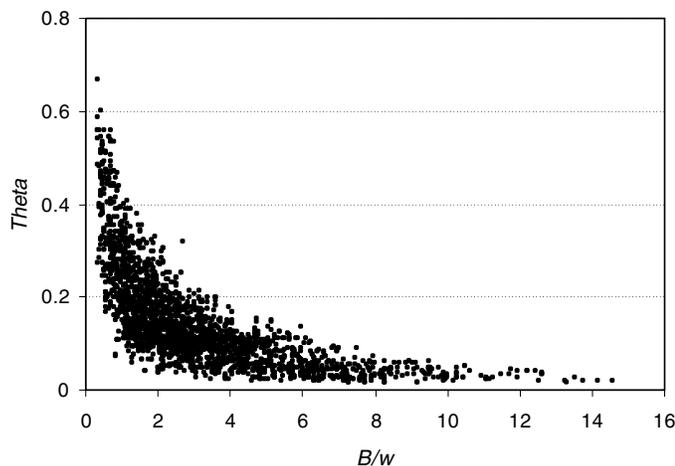


Figure 7. Values of B/w plotted against the rate of decay, θ . Each point represents the joint ML estimate falling within the 2LL support limit after one run of biased random walk using *Analyse*; the run was based on the mean frequencies of six allozymes and the Sstep model.

sperm count) and hybrid females have also been shown to have reduced fertility (Britton-Davidian et al. 2005; J. Piálek, unpubl. results).

Comparison of our results with those from southern Germany (Payseur et al. 2004; Payseur and Nachman 2005) is more difficult. The barriers estimated from their data approach infinity, precluding any meaningful estimation of selection and fitness parameters. Moreover, it is not clear to what extent the results from the German transect are affected by the presence of a strong geographic barrier (the presence of the Isar River passing right through the zone center [Sage et al. 1986, 1993; cf. also appendix and fig. 1 in Tucker et al. 1992]), by a one-dimensional sampling design, as well as by markedly lower numbers of sampled localities and animals analyzed.

Regardless of the transect analyzed, one characteristic has been reported from all parts of the zone studied so far: asymmetric transition of diagnostic alleles between the two taxa. The introgression of *domesticus* alleles into the *musculus* territory has been consistently more gradual than vice versa (Vanlerberghe et al. 1986; Tucker et al. 1992; Dod et al. 1993, 2005). Interestingly, clines of four of six autosomal loci scored and the six-locus compound cline were found to be symmetrical in this study (it should also be pointed out that asymmetry of the *Gpd1* cline was most probably caused by a high frequency of the *musculus* allele at Plesná and surrounding sites, located on the *domesticus* side of the zone; cf. Fig. 4 and Appendix; genotype data available at <http://www.iach.cz/legs>). In addition, although two autosomal and all X-linked loci were fitted with asymmetric clines, this asymmetry was skewed towards the *domesticus* side, that is, in the opposite direction than in other transects.

This apparent asymmetry may be caused by recent movement of the zone. For example, if the zone moves from the *domesticus* to *musculus* territory, cline shapes of neutral loci will have a higher variance on the *domesticus* side because some loci will lag behind. This is also suggested by the larger scatter of points on the *domesticus* side when the allele frequencies at the six loci (p_i) are plotted against the average frequency (\bar{p}) (figure not shown). Alternatively, the same pattern can be explained by the movement of the zone in the opposite direction: in this case, clines of positively selected loci would be faster than neutral or counterselected loci, causing a “wave of advance” (Piálek and Barton 1997; but see Dasmahapatra et al. 2002). To some extent, asymmetry in cline tails can be a product of interspecific differences in behavior such as assortative mating in one subspecies and aggressive dominance in the other. It has been demonstrated that wild *M. m. musculus* prefer homosubspecific urine signals (Smadja and Ganem 2002, 2005) and, to some extent, salivary androgen-binding proteins (Bímová et al. 2005; Smadja and Ganem 2005) whereas no such preferences were found in *M. m. domesticus*. Furthermore, in the interaction *M. m. domesticus* males usually dominate less aggressive *M. m. musculus* males (Thuesen 1977; Van Zegeren and Van Oortmerssen 1981; B. Bímová, unpubl. data). The asymmetrical divergence of the behavioral isolation can result from differential selection acting on both subspecies or, alternatively, can itself cause the differential selection leading to asymmetric gene flow. However, if these behavioral differences caused asymmetry of the mouse hybrid zone we would expect the clines to be skewed to the *musculus* side.

DISPERSAL

From equations (8) to (10) it is obvious that estimates of the fitness parameters listed in Table 4 depend on an accurate estimate of the dispersal rate. Here we used two independent estimations, one derived from linkage disequilibrium and cline width (eq. 7) and the other one based on microsatellite genotypes. It may be argued that strong heterozygote deficit in the zone center, demonstrated by high F_{IS} values (Fig. 2A), inflates the estimate of R_{ij} and hence σ . To deal with this problem properly, one should estimate F_{IS} and R_{ij} together with frequencies for each locus in a single analysis; however, the genotype space is very large so even huge samples would be insufficient for likelihood estimates to be informative. Therefore, we used at least the nonzero option for F_{IS} available in *Analyse* to get an improved estimate of R_{ij} .

Another potential source of differences in estimates of the dispersal rate can be the choice of sites from which linkage disequilibrium is computed. For example, if we adopted an approach similar to that of Szymura and Barton (1986) and estimated disequilibria from all samples polymorphic for studied loci, we would get $\sigma = 0.922$, the figure close to $\sigma = 0.817$ based on microsatellite data. Also other parameters would get closer

to microsatellite-derived estimates (cf. Table 4, middle column): $s^* = 0.0693$ (0.0303–0.2201); $s_e = 0.0111$ (0.0021–0.0273); $s = 0.0378$ (0.0038–0.4439); $N = 72$ (0–39,284). However, according to our opinion, what one should concern oneself about is selection acting *within* the hybrid zone, that is, within an area delimited approximately by the central step of the multilocus allozyme cline. This approach also enabled us to compare the results with those published from the Danish contact zone (Raufaste et al. 2005).

Dispersal exceeding 1 km/gen^{1/2} might seem to be too high for the house mouse. Movement of mice is mainly limited to several tens of meters (see Berry 1981; Sage 1981; Hauffe et al. 2000; Pocock et al. 2004, 2005; and references therein). However, studies of population structure and migration in the mouse have been generally limited to a single farm neglecting long-range migrations (but see Pocock et al. 2004, 2005). As pointed out by Barton and Hewitt (1985), rare long-distance movements and frequent extinction-recolonization events may greatly inflate dispersal, and hence most of published measurements are usually underestimates. Indeed, several authors (e.g., Berry 1968; Pearson 1963) have reported migrations over 1000 m, or even up to 2400 m (Tomich 1970, cited in Sage 1981).

There is an interesting point that has not been discussed before. Higher dispersal in the zone center can be consistent with lower population densities in the zone center expected from the tension-zone predictions. However, there is a question whether the mouse densities in central populations can per se explain the increased dispersal rate. Although we have no exact data available, our rough estimates (see below) do not indicate any substantial decrease of the mean population density within the zone. Then taking into account equation (7) one can hypothesize that the harmonic recombination rate may be lower in the hybrid zone than outside it so there is no need to expect higher dispersal of mice caused by decreased population density. One of the possible causes of lower recombination could be an assortative mating between individuals of the same subspecies.

To get an estimate of dispersal rate independent of cline parameters, the parameter was inferred by analyzing variation at four microsatellite loci under the isolation-by-distance model. The resulting value of $\sigma \approx 800$ m/gen^{1/2} was closer to values reported from Denmark ($\sigma \approx 500$ –800 m/gen^{1/2}; Raufaste et al. 2005) than to dispersal estimated here from the multilocus cline ($\sigma \approx 1000$ m/gen^{1/2}; Table 4). However, the former value should be treated with caution. First, it is questionable whether four loci can give unbiased results. Second, and more importantly, our estimates of population density are only approximate and may be afflicted with large error.

It might be argued that using neutral models for estimating σ in the center of a hybrid zone in which selection is involved is inappropriate (Wright 1978). This problem may not be trivial because

selection against hybrids will influence F_{ST} in the zone center and hence estimates of σ (Wright 1978). Therefore, we first compared the density estimated from the central populations with that estimated from a cluster of 11 sites outside the zone center ($HI_6 < 0.2$) on the *domesticus* side, sampled in 2001 (nine sites sampled, $N = 106$) and 2002 (nine sites, $N = 107$). The distance between the sites does not exceed 15 km. The data yielded ≈ 2 individuals/km² and dispersal ≈ 750 m/gen^{1/2}, close to the estimates from the center (≈ 1.7 individuals/km²; see Supplementary Material available online).

According to simulations by Slatkin and Barton (1989), when migration rate is $m = 0.01$, selection of $s = 0.02$ acting against heterozygotes in populations of $N = 50$ decreases from $F_{ST} = 0.33$, estimated under the neutral model, to $F_{ST} = 0.26$ in intermediate allele frequencies ($P = 0.5$). The authors conclude that selection must be of the same order as the migration rate before F_{ST} to be substantially distorted (Slatkin and Barton 1989). Although we know neither N nor m , the selection coefficient ($s \approx 0.028$; Table 4) and F_{ST} estimated from the zone center based on the neutral model ($F_{ST} = 0.257$; M. Šugerková, unpubl. results) are very close to parameter values used in Barton and Slatkin's simulation.

Cline width and the age of the house mouse hybrid zone

The age of the zone could potentially affect its width and estimation of key parameters of selection and fitness of hybrids. After a secondary contact, we may expect a rapid diffusion of neutral alleles across the zone and homogenization of diverged gene pools. However, such a process may take quite a long time because it is difficult for an immigrant allele to enter an established deme (Endler 1977; Barton and Bengtsson 1986). As a result, even clines for perfectly neutral markers, unlinked to any selected loci, will look like those under selection immediately after the contact and their width will decay proportionally to the time because secondary contact due to dispersal at a rate $w = \sqrt{2\pi\sigma^2t}$, where t is time in generations (Endler 1977; Barton and Hewitt 1985).

It is widely accepted that the *musculus/domesticus* hybrid zone has resulted from the secondary contact of the two taxa in the Holocene. According to Kratochvíl (1986), Auffray et al. (1990), and others, *M. m. domesticus* colonized Europe from Levant through the eastern and western Mediterranean whereas *M. m. musculus* has followed a route north of the Black Sea. From this scenario it follows that different parts of the zone may differ in age. Even though exact dating of the origin of the mouse hybrid zone in Europe is doubtful and individual estimates vary considerably (Sokal et al. 1991: 4000–5000 BP; Cucchi et al. 2005: 1000 BC–300 AD; Hunt and Selander 1973: 1850s), there is little doubt that the northern part of the zone is much younger than the

central and southern parts; therefore, we can expect zone width to be correlated with its age. For example, Sage et al. (1993) noticed data from various transects to suggest widening of the zone from the east to the west and pointed out that “The width difference is inversely related to the postulated age of the zone . . .” (Sage et al. 1993: p. 534). However, the results on which Sage and his collaborators based their conclusions often suffered from insufficient sampling and lack of any statistical analysis.

Although the compound autosomal cline in Denmark was found to be slightly narrower (eight loci: $w = 8.9$ km, 2LL = 7.7–12.4; Raufaste et al. 2005) than in Central Europe (six loci: $w = 9.6$ km, 2LL = 0.0–16.9; this study), the support limits are widely overlapping. In addition, the estimate by Raufaste et al. (2005) may have been influenced by excluding *Idh1* and *Sod1*, that is, two loci best fitted with rather wide sigmoid clines, from multilocus cline analyses. More importantly, cline widths for the Danish data are likely to be reduced by the presence of a geographic barrier. Interestingly, an area of the highest gradient of change in allele frequency in the Danish mouse hybrid zone was noncoincident with its center. There, the central step of particular clines of most allozyme loci was shifted entirely or almost entirely below frequency 0.5. This phenomenon was caused by the presence of a steep-edged river valley just south of the zone center (Dod et al. 2005; Raufaste et al. 2005).

Although it could be argued that the Danish zone may be too young for the cline shape of neutral markers to be unbiased, according to Raufaste et al. (2005) molecular data suggest that the time elapsed from the initial contact of the two subspecies in Denmark is sufficient for neutral alleles to cross the zone. Thus, it appears that the width of (at least) the Danish and central European sections of the mouse hybrid zone is independent of age.

Comparison of Analyse and ClineFit and inspection of likelihood profiles

With Analyse, individual genotypes at the X-linked loci need to be transformed to frequency data and male and female allele frequencies are pooled before processing. Instead, the ClineFit program (Porter et al. 1997) allows haplodiploid data to be processed directly. Moreover, the sets of estimated parameters one gets as output are not identical in the two programs, while the exponential decays are described in terms of their slopes (θ values) and the points at which they intersect the upper and lower boundaries of the central step (B values), ClineFit describes the left and right sides of the cline as exponential functions with asymptotes on the x-axis (Z values) and slopes (θ values) describing the exponential decay away from the asymptotes.

Regardless of different philosophies in processing data, the two programs should give very similar results (A. Porter, pers. comm.) provided allele frequencies in males and females do not

significantly differ and the same weighting is applied. As expected, the highest congruence of resulting estimates was found for the position of cline centers whereas the least congruent were estimates of B and θ (cf. Tables 3 and 4). More surprising, however, is the fact that in independent ClineFit analyses performed by three of the present authors (M. Macholán, P. Dufková, J. Piálek) two arrived at the same or very similar estimates of cline width (~ 1 km for all five loci) even after running several chains when proceeding from default values yet these runs apparently resulted in being “trapped” on lower peaks. Close inspection of likelihood profiles appears to be necessary in such cases. However, for a higher number of parameters, some likelihood surfaces can be rather complex. More importantly, some peaks may be of nearly the same height. This can be illustrated, for example, with cline width for *Syap1* (Fig. 5). In addition, a peak on the maximum-likelihood surface can be very broad and flat. In *Tsx(-)*, for instance, the same likelihood values were obtained for widths from 4 km to a few meters (Fig. 5). According to our opinion, constructing likelihood profiles used in this study (see Materials and Methods; Phillips et al. 2004) renders a more precise image of the likelihood surface than simple plotting of log-likelihoods against parameter values.

CONCLUDING REMARKS

The abovementioned conclusions are quite sobering. Even though we analyzed a rather large dataset consisting of 1237–1655 specimens from 105 sites scattered across a 110 km long and 40 km wide transect, support limits of some parameters were large potentially due to one or more of the following factors: varying geographic setting, method of generating a 1D cline, random drift, founder events after extinctions and recolonizations, and/or human-mediated long-distance movements (this held especially for parameters B and θ ; cf. Tables 2 and 3). Often, the same data could be fitted with identical or nearly identical likelihood with central cline steps oriented in various directions leading to a wide range of width estimates. The latter problem was typical for X-chromosome data: for example, in the data for all five X-linked markers, most points were located near extreme values (i.e., either 0% or 100% of *domesticus* alleles); intermediate points, necessary for correct fitting the central segment of the stepped cline, were scarce. In such a case, several more-or-less equivalent estimates of w are possible. When the likelihood peak is flat, the problem is exacerbated because the range of possible results is continuous. Unfortunately, this means that in analyses based on models with a central step we can often only place an upper bound on the width so identification of regions on the X chromosome (and in the whole genome in general), which are under strong selection and likely to contribute to reproductive isolation may be considerably complicated.

We contend that two-dimensional analysis is essential for hybrid zone studies. It is not difficult to demonstrate that orientation of a transect across a hybrid zone can substantially alter the results even in simple cases like that surveyed here. Additionally, when the course of the zone is complicated (e.g., Bridle et al. 2001) sampling along a single straight line, regardless whether correctly oriented or not, can potentially lead to highly biased results. Fitting a cline to data from sufficiently sampled two-dimensional area should therefore be a prerequisite for subsequent analyses. This may differ depending on the scale and clarity of change in traits over the field area. However, the presence of apparent outliers in the *Nt* and *Tsx* loci brought about most probably by human-mediated founder events (see above) is cautionary because they can affect the shape of resulting clines substantially even in such a densely sampled transect as that reported in this study. Similar outliers have been found also in some other segments of the mouse hybrid zone (Payseur et al. 2004; Božíková et al. 2005; Dod et al. 2005; Raufaste et al. 2005).

To conclude, we strongly suggest, first, to analyze hybrid zones across a two-dimensional transect and to use a rigorous method of fitting a cline in two-dimensional space. This procedure should include tests of the best-fitting number of cline segments because the zone can have a rather complex spatial pattern, and there is no reason to exclude this possibility a priori even in commensal species such as the house mouse. Second, we should be cautious about our results, especially if we want to localize strongly selected regions with the aim of identifying loci responsible for reproductive isolation (speciation genes), because very different parameter combinations can yield similar or even identical ML values. Inspection of likelihood profiles thus appears to be an essential step. Finally, differences between various sections of the hybrid zone highlight the importance of analyzing and comparing a number of transects before any generalizations can be made.

ACKNOWLEDGMENTS

We thank A. Mack, G. Lang, R. Fritsch, Ms. Brázdová, and K. Brož who greatly facilitated our field work as well as to many other people for their help in collecting mice. MM wishes to express his gratitude to S. J. E. Baird and A. H. Porter who proved to be very patient in explaining intricacies of their computer programs, and to P. Boursot for providing N. Raufaste's Ph.D. thesis. B. A. Payseur and M. W. Nachman provided primers for the SNP markers before publication and their help is gladly acknowledged. We also gratefully acknowledge N. H. Barton, P. Boursot, and H. C. Hauffe for reviewing earlier versions of the manuscript and for fruitful discussions. Comments by the Associate Editor and two reviewers greatly improved the quality of the paper. The work was supported by Institutional Research Plan of the Institute of Animal Physiology and Genetics No. A VOZ50450515, and by the grants of the Grant Agency of AS CR No. A6045902 (to MM and JP), the Czech Science Foundation Nos. 206/03/0205 and 206/06/0707 (to MM) and partly No. 206/03/D148 (to PM).

LITERATURE CITED

- Abe, K., H. Noguchi, K. Tagawa, M. Yuzuriha, A. Toyoda, T. Kojima, K. Ezawa, N. Saitou, M. Hattori, Y. Sakaki, K. Moriwaki, and T. Shiroishi. 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.
- Akaike, H. 1973. Information theory as an extension of the maximum likelihood principle. Pp. 267–281 in B. N. Petrov and F. Csaki, eds. Second international symposium on information theory. Akademiai Kiado, Budapest.
- Arnold, M. L. 1997. Natural hybridisation and evolution. Oxford Univ. Press, Oxford, U.K.
- Auffray, J.-C., F. Vanlerberghe, and J. Britton-Davidian. 1990. The house mouse progression in Eurasia: a palaeontological and archaeozoological approach. *Biol. J. Linn. Soc.* 41:13–25.
- Baird, S. J. E. 1995. A simulation study of multilocus clines. *Evolution* 49:1038–1045.
- Barton, N. H. 1982. The structure of the hybrid zone in *Uroderma bilopatum*. *Evolution* 36:454–471.
- . 1983. Multilocus clines. *Evolution* 37:454–471.
- . 1986. The effects of linkage and density-dependent regulation on gene flow. *Heredity*. 57:415–426.
- Barton, N. H., and S. J. E. Baird. 1995. Analyse—An application for analyzing hybrid zones. Freeware, Edinburgh, U.K. Available at <http://helios.bto.ed.ac.uk/evolgen/Mac/Analyse>.
- 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.
- Barton, N. H., and G. M. Hewitt. 1985. Analysis of hybrid zones. *Annu. Rev. Ecol. Syst.* 16:113–148.
- Barton, N. H., and M. Shpak. 2000. The effect of epistasis on the structure of hybrid zones. *Genet. Res.* 75:179–198.
- Bazykin, A. D. 1969. Hypothetical mechanism of speciation. *Evolution* 23:685–687.
- Berry, R. J. 1968. Ecology of an island population of the house mouse. *J. Anim. Ecol.* 37:445–470.
- . 1981. Town mouse, country mouse: adaptation and adaptability in *Mus domesticus* (*M. musculus domesticus*). *Mamm. Rev.* 11:91–136.
- Bímová, B., R. C. Karn, and J. Piálek. 2005. The role of salivary androgen-binding protein in reproductive isolation between two subspecies of house mouse: *Mus musculus musculus* and *Mus musculus domesticus*. *Biol. J. Linn. Soc.* 84:349–361.
- Bonhomme, F., J. Catalan, J. Britton-Davidian, V. M. Chapman, K. Moriwaki, E. Nevo, and L. Thaler. 1984. Biochemical diversity and evolution in the genus *Mus*. *Biochem. Genet.* 22:275–303.
- 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.
- Britton-Davidian, J., F. Fel-Clair, J. Lopez, P. Alibert, and P. Boursot. 2005. Postzygotic isolation between the two European subspecies of the house mouse: estimates from fertility patterns in wild and laboratory-bred hybrids. *Biol. J. Linn. Soc.* 84:379–393.

- 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* Schwartz & Schwartz, 1943) in the western Mediterranean: a zooarchaeological revision of sub-fossil house mouse occurrences. *Biol. J. Linn. Soc.* 84:429–445.
- Dasmahapatra, K. K., M. J. Blum, A. Aiello, S. Hackwell, N. Davies, E. P. Bermingham, and T. Mallett. 2002. Inferences from a rapidly moving hybrid zone. *Evolution* 56:741–753.
- 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, C. A. Evans, M. M. DeAngelis, D. M. Levinson, L. Kruglyak, N. Goodman, N. G. Copeland, N. A. Jenkins, T. L. Hawkins, L. Stein, D. C. Page, and E. S. Lander. 1996. A comprehensive genetic map of the mouse genome. *Nature* 380:149–152.
- Dobzhansky, T. 1937. *Genetics and the origin of species*. Columbia Univ. Press, New York, NY.
- Dod, B., L. S. Jermiin, P. Boursot, V. H. Chapman, J. Tonnes-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.
- Edwards, A. W. F. 1992. *Likelihood*. Cambridge Univ. Press, Cambridge, MA.
- Endler, J. A. 1977. *Geographic variation, speciation and clines*. Princeton Univ. Press, Princeton, NJ.
- Fisher, R. A. 1950. Gene frequencies in a cline determined by selection and diffusion. *Biometrics* 6:353–361.
- Forejt, J. 1981. Hybrid sterility gene located in the T/t-H-2 supergene on chromosome 17. Pp. 103–131 in R. A. Reisfeld and S. Ferrone, eds. *Current trends in histocompatibility. Immunogenetic and molecular profiles*. Plenum Press, London, U.K.
- . 1996. Hybrid sterility in the mouse. *Trends Genet.* 12:412–417.
- Gaggiotti, O. E., O. Lange, K. Rassmann, and C. Gliddon. 1999. A comparison of two indirect methods for estimating average levels of gene flow using microsatellite data. *Mol. Ecol.* 8:1513–1520.
- Godinho, R., B. Mendonca, E. G. Crespo, and N. Ferrand. 2006. Genealogy of the nuclear B-fibrinogen locus in a highly structured lizard species: comparison with mtDNA and evidence for intragenic recombination in the hybrid zone. *Heredity* 96:454–463.
- Grola, J. W., and O. R. Taylor. 1980. Some characteristics of hybrids derived from the sulfur butterflies, *Colias eurytheme* and *Colias philodice*: phenotypic effects of the X chromosome. *Evolution* 34:673–687.
- Haldane, J. B. S. 1922. Sex ratio and unisexual sterility in animal hybrids. *J. Genet.* 12:101–109.
- . 1948. The theory of a cline. *J. Genet.* 48:277–284.
- Harr, B. 2006. Genomic islands of differentiation between house mouse subspecies. *Genome Res.* 16:730–737.
- Harris, H., and D. A. Hopkinson. 1976. *Handbook of enzyme electrophoresis in human genetics*. North-Holland Publ. Comp., Oxford, U.K.
- Harrison, R. G. 1990. Hybrid zones—windows on evolutionary process. *Oxf. Surv. Evol. Biol.* 7:69–128.
- . 1993. *Hybrid zones and the evolutionary process*. Oxford Univ. Press, Oxford, U.K.
- Hastings, W. K. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57:97–109.
- Hatfield, T., N. H. Barton, and J. B. Searle. 1992. A model of a hybrid zone between two chromosomal races of the common shrew (*Sorex araneus*). *Evolution* 46:1129–1145.
- Hauffe, H. C., J. Piálek, and J. B. Searle. 2000. The house mouse chromosomal hybrid zone in Valtellina (SO): a summary of past and present research. *Hystrix* 11:17–25.
- Hunt, W. G., and R. K. Selander. 1973. Biochemical genetics of hybridisation in European house mice. *Heredity* 31:11–33.
- 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.
- Key, K. H. 1968. The concept of stasipatric speciation. *Syst. Zool.* 17:14–22.
- Kocher, T. D., and R. D. Sage. 1986. Further genetic analyses of a hybrid zone between leopard frogs (*Rana pipiens* complex) in central Texas. *Evolution* 40:21–33.
- Kratochvíl, J. 1986. Die intraspezifische Evolution der Art *Mus domesticus*. *Acta Sc. Nat. Brno* 20(1):1–49.
- 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.
- Lanneluc, I., E. Desmarais, P. Boursot, B. Dod, and F. Bonhomme. 2004. Characterization of a centromeric marker on mouse chromosome 11 and its introgression in a *domesticus/musculus* hybrid zone. *Mamm. Genome* 15:924–934.
- Lijtmaer, D. A., B. Mahler, and P. L. Tubaro. 2003. Hybridization and postzygotic isolation patterns in pigeons and doves. *Evolution* 57:1411–1418.
- Lindblad-Toh, K., E. Winchester, M. J. Daly, D. G. Wang, J. N. Hirschhorn, J. P. Lavolette, K. Ardlie, D. E. Reich, E. Robinson, P. Sklar, N. Shah, D. Thomas, J. B. Fan, T. Gingeras, J. Warrington, N. Patil, T. J. Hudson, and E. S. Lander. 2000. Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nature 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.
- Mallet, J., N. Barton, G. M. Lamas, J. C. Santisteban, M. M. Muedas, and H. Eeley. 1990. Estimates of selection and gene flow from measures of cline width and linkage disequilibrium in *Heliconius* hybrid zones. *Genetics* 124:921–936.
- Mayr, E. 1942. *Systematics and the origin of species*. Columbia Univ. Press, New York, NY.
- Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. 1953. Equations of state calculations by fast computing machines. *J. Chem. Phys.* 21:1087–1091.
- Montagutelli, X., R. Turner, and J. H. Nadeau. 1996. Epistatic control of non-Mendelian inheritance in mouse specific crosses. *Genetics* 143:1739–1752.
- Mouse Genome Sequencing Consortium. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* 420:520–562.
- 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, and 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.
- Oka, A., A. Mita, N. Sakurai-Yamatani, H. Yamamoto, N. Takagi, T. Takano-Shimizu, K. Toshimori, K. Moriwaki, and T. Shiroishi. 2004. Hybrid

- breakdown caused by substitution of the X chromosome between two mouse subspecies. *Genetics* 166:913–924.
- Orr, H. A. 1997. Haldane's rule. *Annu. Rev. Ecol. Syst.* 28:195–218.
- Pasteur, N., G. Pasteur, F. Bonhomme, J. Catalan, and J. Britton-Davidian. 1988. Practical isozyme genetics. Ellis Horwood, Ltd., Chichester, U.K.
- Payseur, B. A., and H. E. Hoekstra. 2005. Signatures of reproductive isolation in patterns of single nucleotide diversity across inbred strains of mice. *Genetics* 171:1905–1916.
- 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.
- Payseur, B. A., and M. W. Nachman. 2005. The genomics of speciation: investigating the molecular correlates of X chromosome introgression across the hybrid zone between *Mus domesticus* and *Mus musculus*. *Biol. J. Linn. Soc.* 84:523–534.
- Pearson, O. P. 1963. History of two local outbreaks of feral house mice. *Ecology* 44:540–549.
- Pelikán, J. 1981. Patterns of reproduction in the house mouse. Pp. 205–230 in R. J. Berry, ed. *Biology of the house mouse*. Academic Press, London.
- 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.
- Piálék, 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.
- 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, R. J. Mural, B. Paigen, and T. Wiltshire. 2004. Use of a dense single nucleotide polymorphism map for in silico mapping in the mouse. *PLoS Biol.* 2:2159–2169.
- Pocock, M. J. O., H. C. Hauffe, and J. B. Searle. 2005. Dispersal in house mice. *Biol. J. Linn. Soc.* 84:565–583.
- Pocock, M. J. O., J. B. Searle, and P. C. L. White. 2004. Adaptations of animals to commensal habitats: population dynamics of house mice *Mus musculus domesticus* on farms. *J. Anim. Ecol.* 73:878–888.
- Polani, P. E. 1972. Centromere localization at meiosis and the position of chiasmata in the male and female mouse. *Chromosoma* 36:343–374.
- 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., R. D. Sage, U. Gyllensten, W. K. Thomas, R. Hübner, C. S. Jones, L. Noble, J. B. Searle, and A. C. Wilson. 1993. Mitochondrial-DNA sequence diversity and the colonization of Scandinavia by house mice from East Holstein. *Biol. J. Linn. Soc.* 50:85–122.
- 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.
- Rand, R. G., and D. M. Harrison. 1989. Ecological genetics of a mosaic hybrid zone—mitochondrial, nuclear, and reproductive differentiation of crickets by soil type. *Evolution* 43:432–449.
- Raufaste, N. 2001. Barrières au flux génique et sélection dans une zone hybride: Etude théorique et expérimentale chez la souris domestique. Université Montpellier II, Montpellier, France.
- Raufaste, N., A. Orth, K. Belkhir, D. Senet, C. Smadja, S. J. E. Bird, F. Bonhomme, B. Dod, and P. Boursot. 2005. Inferences of selection and migration in the Danish house mouse hybrid zone. *Biol. J. Linn. Soc.* 84:593–616.
- Rousset, F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* 145:1219–1228.
- . 2000. Genetic differentiation between individuals. *J. Evol. Biol.* 58:58–62.
- Sage, R. D. 1981. Wild mice. Pp. 39–90 in H. L. Foster, J. D. Small, and J. G. Fox, eds. *The mouse in biomedical research*. Volume I. Academic Press, New York, NY.
- Sage, R. D., W. R. Atchley, and E. Capanna. 1993. House mice as models in systematic biology. *Syst. Biol.* 42:523–561.
- Sage, R. D., and R. K. Selander. 1979. Hybridization between species of the *Rana pipiens* complex in central Texas. *Evolution* 33:1069–1088.
- Sage, R. D., J. B. Whitney, III, and A. C. Wilson. 1986. Genetic analysis of a hybrid zone between *domesticus* and *musculus* mice (*Mus musculus* complex): hemoglobin polymorphisms. *Curr. Top. Microbiol. Immunol.* 127:75–85.
- Selander, R. K., and S. Y. Yang. 1969. Protein polymorphism and genic heterozygosity in a wild population of the house mouse (*Mus musculus*). *Genetics* 63:653–657.
- Shaw, D. D., D. J. Coates, M. L. Arnold, and P. Wilkinson. 1985. Temporal variation in the chromosomal structure of a hybrid zone and its relationship to karyotypic repatterning. *Heredity* 55:293–306.
- 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.
- . 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139:457–462.
- Slatkin, M., and N. H. Barton. 1989. A comparison of three indirect methods for estimating average levels of gene flow. *Evolution* 43:1349–1368.
- Smadja, C., and G. Ganem. 2002. Subspecies recognition in the house mouse: a study of two populations from the border of a hybrid zone. *Behav. Ecol.* 13:312–320.
- . 2005. Asymmetrical reproductive character displacement in the house mouse. *J. Evol. Biol.* 18:1485–1493.
- Sokal, R. R., N. L. Oden, and A. C. Wilson. 1991. Genetic evidence for the spread of agriculture in Europe by demic diffusion. *Nature* 351:143–145.
- StatSoft, Inc. 2000. *Statistica for Windows* (computer program manual). Tulsa, OK.
- Storčhová, 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.
- 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 fire-bellied toads *Bombina bombina* and *B. variegata*: comparisons between transects and between loci. *Evolution* 45:237–261.
- Takami, Y., and H. Suzuki. 2005. Morphological, genetic and behavioural analyses of a hybrid zone between the ground beetles *Carabus lewisianus* and *C. albrecthi* (Coleoptera, Carabidae): asymmetrical introgression caused by movement of the zone? *Biol. J. Linn. Soc.* 86:79–94.
- 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.
- Thuesen, P. 1977. A comparison of the agonistic behaviour of *Mus musculus musculus* L. and *Mus musculus domesticus* Ruddy (Mammalia, Rodentia). *Vidensk. Meddr. dansk. naturh. Foren.* 140:117–128.
- 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.
- Ursin, E. 1952. Occurrence of voles, mice and rats (Muridae) in Denmark, with a special note on a zone of intergradation between two subspecies

- of the house mouse (*Mus musculus* L.). Vid. Medd. Dansk. Naturhist. Foren. 114:217–244.
- 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.
- 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.
- van Zegeren, K., and G. A. van Oortmerssen. 1981. Frontier disputes between the West- and East-European house mouse in Schleswig-Holstein, West Germany. Z. Säugetierkd. 46:363–369.
- Woodruff, D. S. 1989. Genetic anomalies associated with *Cerion* hybrid zones: the origin and maintenance of new electromorphic variants called hybridzymes. Biol. J. Linn. Soc. 36:281–294.
- Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370.
- Wright, S. 1943. Isolation by distance. Genetics 28:114–138.
- . 1978. Evolution and the genetics of populations, vol 4: variability within and among natural populations. Univ. of Chicago Press, Chicago.
- 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: B. Nürnberger

Supplementary Material

The following supplementary material is available for this article:

Appendix S1. Dispersal Estimated from Microsatellite Data.

This material is available as part of the online article from:

<http://www.blackwell-synergy.com/doi/abs/10.1111/j.1558-5646.2007.00065.x>

(This link will take you to the article abstract.)

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

APPENDIX. List of localities studied; D_{fit} = distance along transect; HI_6 = frequency of *musculus* alleles at six allozyme loci; HI_X = frequency of *musculus* alleles at five X-linked loci.

No.	Locality	D_{fit}	HI_6	HI_X	No.	Locality	D_{fit}	HI_6	HI_X
1	Straas 1	.00	0	0	54	Dolnice	45.55	.222	.184
2	Straas 2	.11	.015	.004	55	Doubí	45.93	.404	.27
3	Münchberg	1.41	.042	0	56	Jindřichov	46.56	.603	.16
4	Grassemann	6.30	0	0	57	Děvín	46.58	.417	0
5	Benk	7.40	.025	0	58	Mlýnek 2	46.83	.133	0
6	Lehsten	12.91	.028	0	59	Mlýnek 1	47.02	.167	.556
7	Meierhof	13.60	0	0	60	Milhostov	47.91	.684	.744
8	Roeslau	16.63	0	0	61	Kopanina 2	48.23	.417	.25
9	Hebanz	21.35	.073	0	62	Kopanina 1	48.25	.174	0
10	Trojmezí	23.19	.147	0	63	Hluboká	50.24	.583	1
11	Kleinwendern	23.27	0	0	64	Nebanice 1	50.75	.625	.833
12	Unterweissenbach	24.36	0	0	65	Nebanice 3	50.83	.842	.933
13	Hranice	24.50	.109	0	66	Nebanice 2	51.22	.798	.864
14	Plössberg	24.52	.042	0	67	Obilná	51.58	.923	.989
15	Krásňany	25.82	.061	0	68	Krajková 1	51.75	.776	1
16	Thierstein	26.44	.139	0	69	Kaceřov 1	52.13	.865	.945
17	Höchstädt-Zelch	26.80	0	.071	70	Krajková 2	52.14	.769	.875
18	Smrčína	26.97	.038	0	71	Kaceřov 2	52.21	.776	1
19	Neuenreuth	28.32	.077	0	72	Mostov	52.30	.889	1
20	Neuenreuth8	28.65	.108	0	73	Dolina	52.48	.810	.909
21	Aš	29.57	.011	0	74	Krajková 3	52.64	.753	.9
22	Kothigenbibersbach	30.53	.118	0	75	Krajková 4	52.69	.833	.9
23	Libá 1	34.23	.133	.067	76	Chotíkov	52.92	.889	.833
24	Polná	34.28	.202	0	77	Anenská Ves	53.47	.878	.974
25	Libá 2	34.29	.142	.037	78	Lipoltov	53.72	.833	.5
26	Hammermühle 1	34.40	.052	.1	79	Okrouhlá	53.83	.583	1
27	Hohenberg	34.50	0	0	80	Hřebeny	54.47	.833	1
28	Skalka (Hazlov)	34.91	.159	0	81	Boučí	54.95	.944	1
29	Hazlov	35.80	.102	0	82	Habartov	54.97	.917	1
30	Hůrka 2	35.99	.188	.077	83	Lomnice	58.13	.774	1
31	Hůrka 1	36.00	.133	.217	84	Dolní Nivy	58.37	.870	1
32	Poustka 2	38.21	.189	.033	85	Hlavno	58.41	.833	.889
33	Poustka 1	38.34	.280	.091	86	Svatava	59.19	.833	1
34	Plesná	38.56	.211	.015	87	Rudolec 1	60.24	.864	.92
35	Vojtanov	39.49	.333	0	88	Rudolec 2	60.45	.981	1
36	Lužná	39.51	.198	.131	89	Kostení Bříza	61.20	.978	1
37	Klest	40.95	.083	0	90	Vintřívov	64.03	1	1
38	Starý Rybník	41.50	.280	.15	91	Staré Sedlo	65.42	.995	.977
39	Luby	41.77	.196	0	92	Nová Role	67.88	.750	1
40	Skalka (Cheb)	42.40	.261	0	93	Děpoltovice	69.06	.978	1
41	Křižovatka	42.58	.141	.05	94	Počerny 1	69.83	.904	.793
42	Dolní Luby	43.30	.208	0	95	Počerny 2	69.83	1	.967
43	Střížov	43.56	.232	.3	96	Horní Slavkov	71.64	1	1
44	Dolní Pelhřimov	43.58	.470	.188	97	Nová Ves 1 (So.)	72.68	1	1
45	Čirá	43.78	.028	0	98	Nová Ves 2 (So.)	72.80	1	1
46	Suchá	43.85	.250	0	99	Nová Ves 3 (So.)	72.83	1	.937
47	Horní Ves	43.91	.189	.077	100	Dalovice	75.71	1	1
48	Spálená	44.18	0	0	101	Sedlečko	80.22	1	.958
49	Nový Drahov	44.56	.079	.018	102	Osvinov	81.28	.972	1
50	Dlouhé Mosty	44.56	.289	.369	103	Stráž nad Ohří	81.80	1	1
51	Svatý Kříž	44.93	.305	0	104	Podb. Rohozec	101.70	.958	1
52	Nová Ves (Cheb)	45.17	.176	.088	105	Buškovice	108.50	.990	1
53	Nový Kostel	45.17	.073	0					