



HAL
open science

Inferences of selection and migration in the Danish house mouse hybrid zone

Nathalie Raufaste, Annie Orth, Khalid Belkhir, David Senet, Carole Smadja, Stuart Baird, François Bonhomme, Barbara Dod, Pierre Boursot

► **To cite this version:**

Nathalie Raufaste, Annie Orth, Khalid Belkhir, David Senet, Carole Smadja, et al.. Inferences of selection and migration in the Danish house mouse hybrid zone. *Biological Journal of the Linnean Society*, 2005, 84 (3), pp.593-616. 10.1111/j.1095-8312.2005.00457.x . hal-01946507

HAL Id: hal-01946507

<https://hal.science/hal-01946507>

Submitted on 27 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The genus *Mus* as a model for evolutionary studies

Edited by J. Britton-Davidian and J. B. Searle

Inferences of selection and migration in the Danish house mouse hybrid zone

NATHALIE RAUFASTE¹, ANNIE ORTH¹, KHALID BELKHIR¹, DAVID SENET¹,
CAROLE SMADJA², STUART J. E. BAIRD^{1†}, FRANÇOIS BONHOMME¹, BARBARA DOD¹
and PIERRE BOURSOT^{1*}

¹Laboratoire Génome Populations Interactions Adaptation (UMR 5171 IFREMER-CNRS-UMII), and
²Institut des Sciences de l'Évolution (UMR 5554 CNRS-UMII), Université Montpellier II, France

Received 30 October 2003; accepted for publication 7 October 2004

We analysed the patterns of allele frequency change for ten diagnostic autosomal allozyme loci in the hybrid zone between the house mouse subspecies *Mus musculus domesticus* and *M. m. musculus* in central Jutland. After determining the general orientation of the clines of allele frequencies, we analysed the cline shapes along the direction of maximum gradient. Eight of the ten clines are best described by steep central steps with coincident positions and an average width of 8.9 km (support limits 7.6–12.4) flanked by tails of introgression, indicating the existence of a barrier to gene flow and only weak selection on the loci studied. We derived estimates of migration from linkage disequilibrium in the centre of the zone, and by applying isolation by distance methods to microsatellite data from some of these populations. These give concordant estimates of $\sigma = 0.5\text{--}0.8$ km generation^{-1/2}. The barrier to gene flow is of the order of 20 km (support limits 14–28), and could be explained by selection of a few per cent at 43–120 underdominant loci that reduces the mean fitness in the central populations to 0.45. Some of the clines appear symmetrical, whereas others are strongly asymmetrical, and two loci appear to have escaped the central barrier to gene flow, reflecting the differential action of selection on different parts of the genome. Asymmetry is always in the direction of more introgression into *musculus*, indicating either a general progression of *domesticus* into the *musculus* territory, possibly mediated by differential behaviour, or past movement of the hybrid zone in the opposite direction, impeded by potential geographical barriers to migration in *domesticus* territory. © 2005 The Linnean Society of London, *Biological Journal of the Linnean Society*, 2005, 84, 593–616.

ADDITIONAL KEYWORDS: gene flow – genetic barrier – hybridization – *Mus musculus domesticus* – *Mus musculus musculus* – speciation.

INTRODUCTION

Hybrid zones have been referred to as ‘windows on the evolutionary process’ (Harrison, 1990) because they allow us to study the interplay between migration and selection on the evolution of genetic differentiation and adaptation. When parapatric taxa meet

and hybridize, selection against unfit hybrids can counteract the homogenizing effect of migration and lead to the establishment of frequency clines of diagnostic characters at the boundary between their distribution areas. A detailed population genetics theory has been developed to model the expected patterns of allele frequency changes in such situations (e.g. for reviews see Barton & Hewitt, 1985; Barton & Gale, 1993), showing that they can constitute barriers to gene flow between the taxa, of increasing intensity with the number of loci involved in hybrid unfitness, and with the spread of these loci in the genome.

*Corresponding author. E-mail: boursot@univ-montp2.fr

†Present address: INRA CBGP, Campus de Baillarguet, 34988 Montpellier sur Lez, France

Briefly, this is because migration brings parental genotypes to the centre of the hybrid zone where neutral loci are thus in linkage disequilibrium with loci under selection in hybrids, which impedes free introgression of these neutral loci and causes an abrupt change of their allele frequency in the centre. However, these loci can eventually extricate themselves from this negative genetic background by recombination, and form long tails of introgression into the foreign territory. The resulting clines (described by the shapes of the central step and of the tails of introgression) are relatively independent of the type of selection against hybrids (which is usually unknown), and can be used to quantify the intensity of the selection maintaining the zone and of the resulting barrier to gene flow. Their estimation can also provide some information about the number of loci involved in the selection of hybrids, an important characteristic of the mechanisms leading to incompatibilities between differentiating genomes.

Since the pioneering work of Hunt & Selander (1973), several authors have studied the genetics of the hybrid zone between the two European subspecies of the house mouse, *Mus musculus domesticus* and *M. m. musculus*, that are thought to have come into secondary contact in Europe after a period of independent geographical expansion from the Middle East, with *M. m. domesticus* colonizing the Mediterranean basin and Western Europe while *M. m. musculus* was expanding across central Europe (e.g. see Boursot *et al.*, 1993, for a review). In previous studies, cline widths were roughly quantified by visually inspecting the variations of synthetic morphological or genetic hybrid indexes: 90% of the genetic transition occurs over 20 km in Denmark (Hunt & Selander, 1973), 75% over 20 km in southern Germany (Sage, Whitney & Wilson, 1986b), 80% over 36 km in Bulgaria (Vanlerberghe *et al.*, 1988), and 60% of genetic and morphological variation in 20–40 km in East Holstein (Prager *et al.*, 1993). The first attempt to estimate cline widths involved the application of a simple sigmoid model on the south German transect (Tucker *et al.*, 1992). They found narrower cline widths for the sex chromosome markers (Y chromosome, 4 km; X chromosome markers from 4 to 10 km) compared with autosomal allozymes (from 6.4 to 21.2 km). A similar contrast between the sex chromosomes and autosomal loci was also found in Denmark and Bulgaria (Vanlerberghe *et al.*, 1986; Dod *et al.*, 1993). However, none of these studies had enough samples both in the centre and in the tails of introgression for a detailed analysis of the cline shape to be realistic. In addition, sampling was often carried out in a linear fashion in an arbitrary direction across the transect, allowing comparisons between markers, but not the calculation of cline parameters along the line of maximum slope. Further-

more, none of these studies included estimations of migration, nor of linkage disequilibrium that could be combined with cline widths to estimate the intensity of selection against hybrid mice. Here we analyse a large dataset on the Danish hybrid zone characterized for ten diagnostic allozyme loci, and derive independent estimates of migration using microsatellite loci.

MATERIAL AND METHODS

MICE

Mice were live trapped inside buildings using multi-capture wire traps, during several field trips from 1984 to 2000. The location of the sampling sites in the Jutland peninsula is indicated on Figure 1 and the list of localities with their Universal Transverse Mercator (UTM) coordinates are given in Appendix 1.

PROTEIN ELECTROPHORESIS

Mice were killed and dissected in the field, and liver, kidney, heart, plasma and blood cells were kept in liquid nitrogen for further preparations. Protein extractions, separation by starch gel electrophoresis (or acrylamide gels in the case of Amylase) and detection of enzyme activity in the gels followed standard protocols, such as described in Pasteur *et al.* (1987). The loci were chosen for their ability to distinguish between the two subspecies in previous studies on house mice in the Jutland peninsula and on a broader geographical scale (Hunt & Selander, 1973; Bonhomme *et al.*, 1984; Britton-Davidian, 1990; Din *et al.*, 1996). The alleles were identified by comparison with standards obtained from mice of known genotypes, and each locality was characterized by the frequency of *M. m. musculus* alleles.

ORIENTATION OF THE CLINES

The general orientation of the maximum gradient of allele frequency across the hybrid zone was determined by fitting the allele frequency data to a simple sigmoid model, where the logit transform of the allele frequencies is a linear function of the two-dimensional (2D) geographical coordinates. The model was fitted by maximum likelihood, assuming a binomial error on the estimations of allele frequencies, using the computer package GLIM4 (the Numerical Algorithm Group). This orientation procedure determined the direction of maximum gradient of allele frequency, assuming the centre of the hybrid zone is a straight line, and that the frequency change is sigmoid. The coordinate of each locality was then calculated by projection on this direction of maximum gradient.

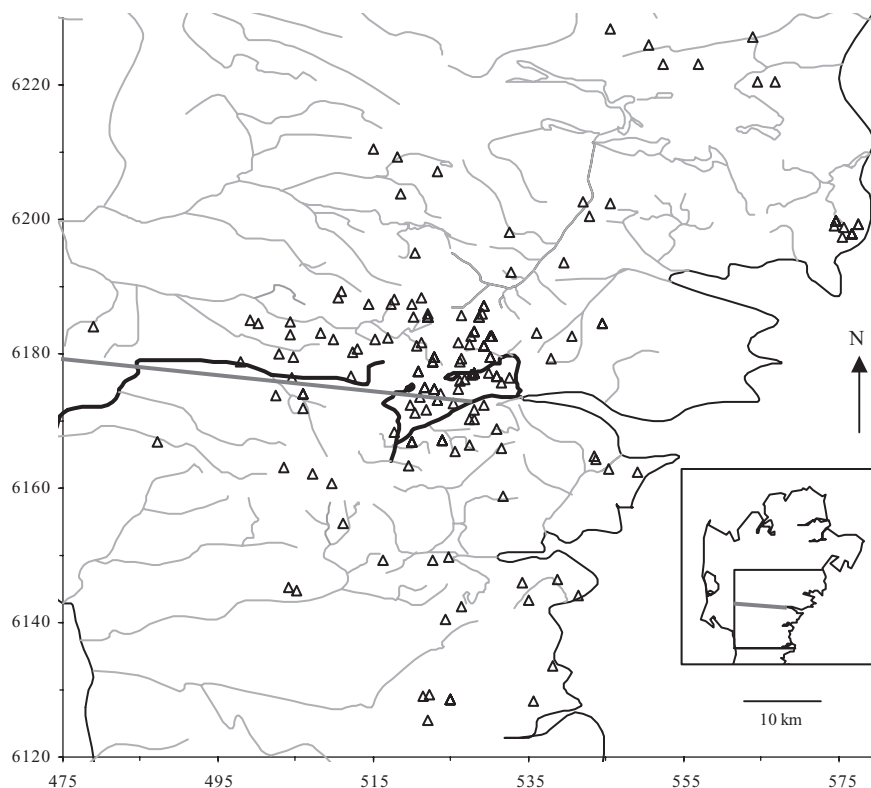


Figure 1. Location of the sampling sites are shown by open triangles on the map, and the axes give their UTM coordinates. Inset: the location of the study area in the Jutland Peninsula. The grey line running across the map is the position of the centre of the hybrid zone that ends in the east at the head of the Vejle Fjord. The rivers are drawn on this map, and some are highlighted by thicker lines (see text).

FITTING CLINE SHAPE

The *musculus* allele frequencies for each locus in the different localities along the 1D transect were fitted to various models of cline shapes by maximum likelihood estimation, using the computer package Analyse (by N. Barton & S. Baird, <http://helios.bto.ed.ac.uk/evolgen/Mac/Analyse/>).

Sample sizes were corrected according to:

$$N_e = \frac{2N}{2N * F_{st} + F_{is} + 1} \quad (1)$$

(adapted from Szymura & Barton, 1986, 1991), where N is the number of individuals sampled in the locality. F_{IS} is the deficit of heterozygotes (set to zero if not positive) and is used to correct for the non-independence between sampling of alleles when there is inbreeding. F_{ST} represents the fluctuations of allele frequencies between loci that are not accounted for by differences in their cline shapes. It represents the residual variation around the regression of allele frequencies at individual loci in each locality against the average of all loci, and was estimated using the 'concordance' procedure in the Analyse package. The above correction is designed for a single locus. When data for sev-

eral loci were pooled, the effective sample size was taken as the sum of effective sample sizes for the different loci.

The Analyse package is then used to compare the likelihood of the allele frequency data under three different models of cline shape: a sigmoid cline, and clines in three parts, with a central sigmoid part and two exponential tails of introgression, either identical on both sides or different. The first model has two parameters, w , the width of the cline (inverse of the maximum slope), and c , the geographical position of the centre. The second model has two additional parameters (four in total) describing the shape of the exponential tails, and the last model has two such parameters for each tail (six parameters in total). As the models are nested, a likelihood ratio test can be applied to choose the model that best explains the data with a minimum of parameters, by assuming that twice the difference of log-likelihood between two models follows a chi-squared distribution with the number of degrees of freedom equal to the difference in the number of parameters between the two models.

The Analyse program uses a Metropolis random exploration algorithm to find the maximum likelihood

estimate of the parameters, so it was run many times on each dataset, with different starting conditions and different settings of the parameters controlling the exploration algorithm, in order to explore the parameter space as thoroughly as possible. Two log-likelihood support limits of parameter estimates were determined by inspecting the results of 10 000–50 000 explorations of the parameter space with the algorithm. The likelihood profiles (Hilborn & Mangel, 1997) for position and width were explored using the ‘crosssection’ option of Analyse, as described in Phillips, Baird & Moritz (2004): the parameter of interest (c or w) is set to a fixed value, and the other parameters are searched until the maximum likelihood of the data is found for this value of c (w). The procedure is repeated for a range of values of c (w) until the range of relevant values is covered, and the likelihood profile can thus be generated for the parameter of interest.

ESTIMATING SELECTION PARAMETERS

The parameters of the fitted cline shapes can be used to infer some population parameters of interest, using the existing theory of tension zones, as detailed in a number of papers (Barton & Hewitt, 1985; Szymura & Barton, 1986, 1991; Barton & Gale, 1993; Kruuk *et al.*, 1999; Barton & Shpak, 2000). We will summarize here the part of this complex theory that was used.

As we will see, a cline shape in three parts, with a central sigmoid part and two exponential tails, best describes most loci studied. This is the shape expected for a locus under weak selection (creating the exponential tails of introgression) submitted to the influence of several loci under stronger selection, creating the central barrier to gene flow, and the central step in the clines. Although some of the theory summarized below is derived for neutral loci, the hypothesis of weak selection on the studied loci needs to be introduced to derive equilibrium cline shapes (the only equilibrium for neutral loci would be even frequency, e.g. Szymura & Barton, 1986, 1991).

The central sigmoid step can be described by two parameters, c , the position of the centre, and w , the width of the cline (inverse of the maximum slope). The following relation relates the total selection acting on the locus to the width of the cline:

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

where σ^2 is the migration parameter (variance of distance from parent to offspring). This is a nuisance parameter, but it is possible to estimate it by comparing the clines for different loci and their linkage disequilibria. It is expected to be proportional to the linkage disequilibrium and the rate of recombination between the loci, and to the gradients of allele fre-

quencies for these loci. Standardized linkage disequilibria (linkage disequilibrium standardized by the maximum possible value given allele frequencies) between pairs of loci were estimated by maximum likelihood with the program Analyse, in localities from the centre of the hybrid zone, where linkage disequilibrium is expected to be maximum, and where the gradient of allele frequencies is estimated as the inverse of cline width. The migration parameter was estimated for each locality and each pair of loci. The estimates obtained are then averaged over pairs of loci and localities to obtain a single estimate of the migration parameter. Because the statistical properties of such an average are not known, we also used the variance of the individual hybrid index to estimate the average linkage disequilibrium between loci (Barton & Gale, 1993), \bar{D} .

The equations of the exponential tails of introgression allow the inference of other important parameters. The equation of the left tail of the cline is:

$$p(x) = \alpha_0 \exp\left(\frac{2(x-c)}{w} \sqrt{\theta_0}\right), \quad (3)$$

and that of the right tail

$$p(x) = 1 - \alpha_1 \exp\left(\frac{-2(x-c)}{w} \sqrt{\theta_1}\right). \quad (4)$$

It can be seen that parameters θ represent the square of the ratio between the expected rate of decay in the tails without a barrier and the actual rate of decay. They can thus be used to estimate the ratio between the selection on the locus under study itself and the total selection experienced by this locus, including the influence of other loci:

$$\theta = \frac{s_{\text{locus}}}{s^*}. \quad (5)$$

The program Analyse does not give estimates of the α parameters but of parameters B/w , the ratio of the barrier to gene flow to the width of the clines. The relationship between these parameters is:

$$B_0 = \frac{w}{2\sqrt{\theta_0}} \cdot \frac{1 - \alpha_0 - \alpha_1}{\alpha_0}, \quad (6)$$

and similarly for B_1 .

By making the hypothesis that selection on the loci studied is weak, that the number of loci under selection is not too small ($1/n \ll 1$) and that selection acts against heterozygotes, one can derive approximate estimates of the number of loci under selection creating the barrier and the intensity of the selection on each locus using the two following relationships (e.g. Barton & Shpak, 2000):

$$s = \frac{8\sigma^2}{B^2} \ln^2\left(\frac{B}{w\Delta u}\right) \quad \text{and} \quad (7)$$

$$\sqrt{\frac{n}{\bar{r}}} = \frac{B}{2\sigma\sqrt{\ln\left(\frac{B}{w\Delta u}\right)}} \quad (8)$$

where Δu is the height of the central step of allele frequency, and is estimated as the difference in the frequencies of the two exponential tails taken at the centre ($x = c$). We thus have $\Delta u = 1 - \alpha_0 - \alpha_1$, which becomes:

$$\Delta u = \frac{\frac{2}{w}\sqrt{\theta_0}B_0\sqrt{\theta_1}B_1}{\sqrt{\theta_0}B_0 + \sqrt{\theta_1}B_1 + \frac{2}{w}\sqrt{\theta_0}B_0\sqrt{\theta_1}B_1}. \quad (9)$$

Note that in the literature, Δu is sometimes omitted, supposing that selection is strong and that it is close to 1 (Szymura & Barton, 1986, 1991). There remains one nuisance parameter in the equations above, which is \bar{r} , the average recombination rate between the neutral (or quasi-neutral in that case) locus and the loci under selection. Several methods have been suggested to estimate \bar{r} (Barton & Hewitt, 1985; Barton & Bengtsson, 1986). Here we take the conservative value of 0.5 that gives the upper limit of n , the number of loci under selection. In the framework of this underdominance model, the average fitness of the central populations can be estimated (taking the fitness of the parental populations arbitrarily at 1):

$$\overline{W}_H = \exp\left(-\frac{ns}{2}\right). \quad (10)$$

MICROSATELLITE TYPING

Genomic DNA was extracted from the spleen using standard proteinase K/phenol-chloroform methods. Polymorphism was studied at six microsatellite loci that are described in Table 1. Five of them are common with the previous study of Dallas *et al.* (1995). The loci were amplified by PCR. One of the amplification primers was fluorescently labelled (Cy5) and allele sizes were measured after migration in a denaturing acrylamide gel on an automated sequencer

(Pharmacia). Allele sizes for all loci were measured relative to the corresponding allele of the Balb/c mouse laboratory strain, which was arbitrarily attributed size 50. We intentionally selected loci with 3–5-bp repeats (except *D17Mit41*, 2 bp, Table 1), because allele size determination is more reliable than for 2-bp repeats.

ESTIMATION OF MIGRATION FROM MICROSATELLITE ALLELE FREQUENCIES

We estimated the migration parameter σ under isolation by distance models, using three different methods applied to the microsatellite data. The first method calculates the regression between $F_{ST}/(1 - F_{ST})$ and the logarithm of geographical distance between populations, r (Rousset, 1997). It applies to pairs of populations in a 2D habitat, and the slope of the regression gives an estimate of $1/4\pi D\sigma^2$, where D is the local population density, and σ^2 is the dispersion parameter. We chose F_{ST} rather than R_{ST} (Slatkin, 1995) because it is thought to be more conservative when sample sizes are small, and only a few loci are used (Gaggiotti *et al.*, 1999), but also because of its small variance and of the uncertainties about the mutation model underlying the justification of R_{ST} . The second method calculates the regression of statistic \hat{a}_r over the logarithm of geographical distance, r (Rousset, 2000b), the slope of which also provides an estimate of $1/4\pi D\sigma^2$. Here, however, pairs of individuals rather than of populations are compared, and statistic \hat{a}_r is given by $(Q_w - Q_r)/(1 - Q_w)$, where Q_r is the probability of identity by state of two alleles separated by a distance r , and Q_w is the probability of identity by state of the two alleles from the same individual. The third method used here calculates the regression of a spatial autocorrelation statistic, Moran's I , over the logarithm of geographical distance r (Hardy & Vekemans, 1999). It also applies to pairs of individuals, and the slope of the regression can be used as an estimate of $(1 - F_{IT})/2\pi D\sigma^2(1 + F_{IT})$. Computer software GENETIX (<http://www.univ-montp2.fr/~genetix/genetix/genetix.htm>) was used to

Table 1. Description of the microsatellite loci studied

Locus	Name of locus in Dallas <i>et al.</i> (1995)	Balb/c allele in Dallas <i>et al.</i> (1995)	Motif size (bp)	No. of alleles	Chromosome	Position from centromere (cM)
<i>Cypla2</i>	34	4	4	11	9	3.1
<i>Ckmm</i>	105	1	5	7	7	4.5
<i>Kir2</i>	126	4	4	5	15	58.7
<i>Gfap</i>	150	4	3	16	11	62
<i>D15Mit16</i>	<i>D15Mit16</i>	3	4	11	15	61.7
<i>D17Mit41</i>	–	–	2	10	17	53

calculate F_{ST} and convert data files between different formats, while GENEPOP (ver. 3.3, Raymond & Rousset, 1995) allowed the calculation of \hat{a}_r . A program was written to calculate Moran's I , according to the formula in Hardy & Vekemans (1999).

RESULTS

ALLOZYME ANALYSES

A total of 170 localities, listed in Appendix 1 with their geographical coordinates, were studied here and their location is indicated in Figure 1. Most yielded five mice or fewer (see Appendix 1 and the distribution of sample sizes in Table 2). The results of the genetic analysis at the ten enzymatic autosomal loci (*Amy*, *Es1*, *Es10*, *Es2*, *Gpd*, *Idh*, *Mpi*, *Np*, *Pgm* and *Sod*) are given in Appendix 2. A total of 1233 mice were studied, but only 996–1187 were successfully typed, depending on the locus.

Samples from the same locality but different collection years were treated separately. An independent study (to be published elsewhere) failed to find any evidence that this hybrid zone has moved since it was sampled in the late 1960s by Hunt & Selander (1973) and our sampling period which extended from 1984 to 2000. We therefore considered here that the zone was stable enough for us to analyse all sampling years together, in order to increase the quality of the sampling along the transect.

ORIENTATION OF THE TRANSECT

When fitting a sigmoid cline to the data in two dimensions, the centre of the hybrid zone was found to run west-north-west to east-south-east, at a 7° angle clockwise from the west–east direction (as drawn on the map of Fig. 1). We will see below that most loci do not fit the simple sigmoid cline model used here, but rather more complex models with a central step of allele frequency.

Table 2. The distribution of sample sizes

No. of mice	No. of localities
1–5	103
6–10	26
11–15	19
16–20	10
21–25	4
26–30	5
31–35	1
36–40	1
41–45	0
46–50	0
51–55	1

We were thus concerned that the use of this simple model could lead to an erroneous determination of the direction of the clines. Using the more complex models, we tested several orientations with angles from 0 to 20° and found in all cases that the 7° orientation gave the best likelihood (data not shown). The coordinates of the localities along the transect were thus calculated by projecting them onto an axis perpendicular to this central line. However, as can be seen on the map, the centre of the hybrid zone reaches at its eastern end the head of a deep fjord (the Vejle Fjord). For localities further east of this point our standard procedure would certainly underestimate their distance to the centre of the hybrid zone. Thus, the straight line distance between the locality considered and the head of the fjord was added to (north of the hybrid zone centre) or subtracted from (south of the hybrid zone centre) the transect coordinate of the head of the fjord. The amended transect coordinates are listed in Appendix 1, and are used in the following analyses.

CLINE SHAPES

The 1D coordinates of the localities calculated as described above were used to analyse the shape of the cline of allele frequency along the transect. The three models implemented by the computer program Analyse were fitted successively: sigmoid cline (two parameters), symmetric stepped cline (four parameters) and asymmetric stepped cline (six parameters). The likelihood ratio test was used to compare the models and the choice was made at the 5% significance level. Table 3 shows the type of model retained for each locus using this criterion. The sigmoid model (two parameters) could not be rejected for *Idh* and *Sod*, while a stepped symmetric model (four parameters) was retained for *Es1* and *Mpi*, but the stepped asymmetric model (six parameters) provided the best fit for the six remaining loci. The best positions of the centres for each locus and their 2 Log-Likelihood (LL) support limits are given in Table 3. We then tested whether an acceptable common centre for all loci could be found. To do this, we calculated the best LL of each locus dataset for a range of centres spanning the confidence intervals of all loci, leaving all the other parameters free to vary. We then summed the best LLs over loci for each of the centres tested, and determined the position that provides the best summed LL (method described in Phillips *et al.*, 2004). This model, with a common centre at position 6191.8 (Table 3), is significantly worse than that where the centre is free to vary at each locus ($P = 1.4 \times 10^{-12}$). Inspection of the individual tests for each locus between the free and constrained centre models identifies two clear outliers (Table 3): *Idh* and *Sod*, the two loci for which sigmoid clines were retained. We reiterated the above proce-

Table 3. Maximum likelihood position of the centre for each locus, and the likelihood search of a common centre

Locus	Best centre for each locus				Best common centre all loci				Best common centre 8 loci					
	Model	Centre	2LL limits	LL	Centre	LL	2DLL	d.f.	P	Centre	LL	2DLL	d.f.	P
<i>Amy</i>	6	6189.2	(6188.0–6191.3)	-83.55	6191.8	-86.75	6.39	1	0.011	6191.2	-85.13	3.16	1	0.076
<i>Es1</i>	4	6192.0	(6190.8–6193.3)	-62.07	6191.8	-62.15	0.15	1	0.700	6191.2	-62.91	1.68	1	0.195
<i>Es10</i>	6	6191.6	(6189.3–6193.2)	-70.62	6191.8	-70.75	0.26	1	0.607	6191.2	-70.89	0.54	1	0.462
<i>Es2</i>	6	6191.1	(6190.1–6196.7)	-99.70	6191.8	-99.87	0.33	1	0.564	6191.2	-99.60	0.20	1	0.656
<i>Gpd</i>	6	6191.2	(6185.5–6194.8)	-85.43	6191.8	-86.05	1.24	1	0.266	6191.2	-85.43	0.00	1	1.000
<i>Idh</i>	2	6198.9	(6196.1–6201.7)	-82.75	6191.8	-94.27	23.05	1	1.6×10^{-6}					
<i>Mpi</i>	4	6190.4	(6189.3–6191.5)	-76.04	6191.8	-79.07	6.07	1	0.014	6191.2	-77.11	2.14	1	0.143
<i>Np</i>	6	6191.7	(6190.7–6194.6)	-75.53	6191.8	-75.55	0.03	1	0.862	6191.2	-76.03	0.99	1	0.320
<i>Pgm</i>	6	6193.8	(6191.6–6196.3)	-67.30	6191.8	-69.24	3.88	1	0.049	6191.2	-69.54	4.48	1	0.034
<i>Sod</i>	2	6199.0	(6196.6–6201.4)	-81.30	6191.8	-98.21	33.84	1	6.0×10^{-9}					
All loci	best for each locus			-784.29	6191.8	-821.90	75.23	9	1.4×10^{-12}					
All loci except <i>Idh</i> and <i>Sod</i>	best for each locus			-620.24						6191.2	-626.64	12.79	7	0.077

ture for all loci except these two, and found an acceptable common centre for these loci (at position 6191.2, $P = 0.077$, Table 3).

Using this common centre we determined the best cline shapes for these eight loci. The data points and the fitted cline shapes are plotted on Figure 2, and the cline parameters are given in Table 4. A variety of shapes are observed among the clines with a central step. The height of this step in allele frequency varies, the highest being for *Amy*. Although two loci, *Es1* and *Mpi*, have relatively symmetric introgression patterns, all the other loci show an asymmetry that is always in the same direction. They are characterized by a steep central change in allele frequency that occurs mostly on the *domesticus* side, where introgression past this barrier is less extensive than on the *musculus* side (this is particularly pronounced for *Es10*, *Gpd* and *Pgm*). The most pronounced asymmetry of introgression is seen at the *Idh* and *Sod* loci, whose centre is displaced almost 8 km into the *musculus* territory as compared with the common centre of the other loci. For these two loci, no model with a central step (around frequency 0.5) fitted the data better than the simpler sigmoid model (see Fig. 3).

It can also be seen in Figure 2 that there is considerable dispersion of the data points around the model. This is reflected in the poor precision of the parameter estimates, given in Table 4 with their two LL support limits. The variations between loci and the support limits of the parameters are particularly large for the left side, with the left barrier parameter absurdly large for several loci. The parameters defining the right side of the cline are generally better estimated and more consistent across loci, but the support limits still remain rather wide. When the average of the eight loci with a common centre is used (average hybrid indices reported in Appendix 1), more reliable parameter estimates with reasonably narrow support intervals are obtained (Table 4). The parameters defining the right and left sides are similar to each other, and the average cline for these eight loci is shown graphically in Figure 4. The Y chromosome data for the same samples (data from Dod *et al.*, 2005, this issue) are also reported in this figure, showing that the centre of the allozyme clines corresponds to the major and abrupt change for this chromosome.

ESTIMATING SELECTION PARAMETERS FROM CLINE SHAPES

Given the estimates of cline widths derived above for the different loci, it is theoretically possible to estimate the migration parameter σ using linkage disequilibria between loci in the centre of the hybrid zone. We estimated the standardized linkage disequilibria between pairs of loci in each locality for the eight loci

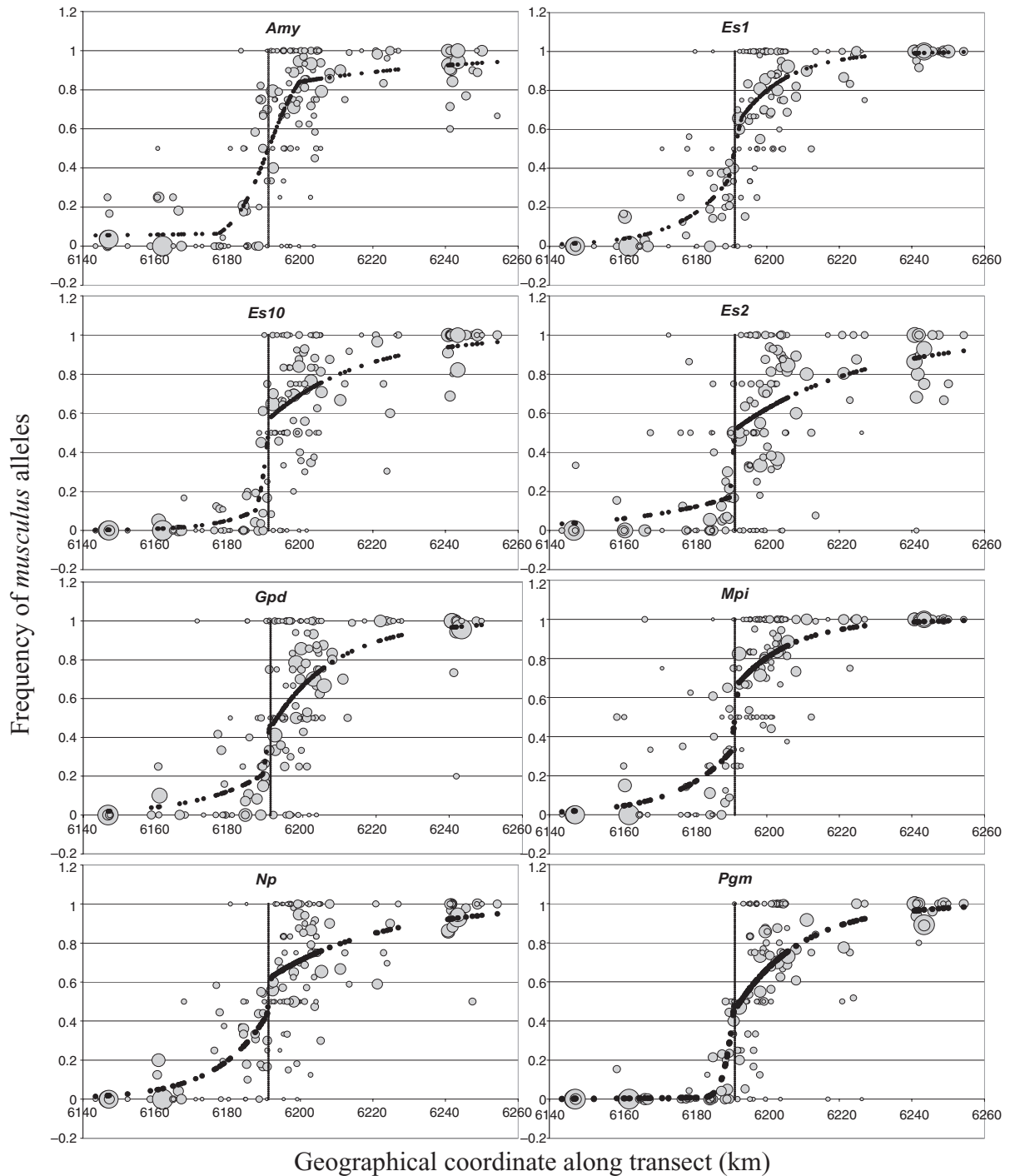


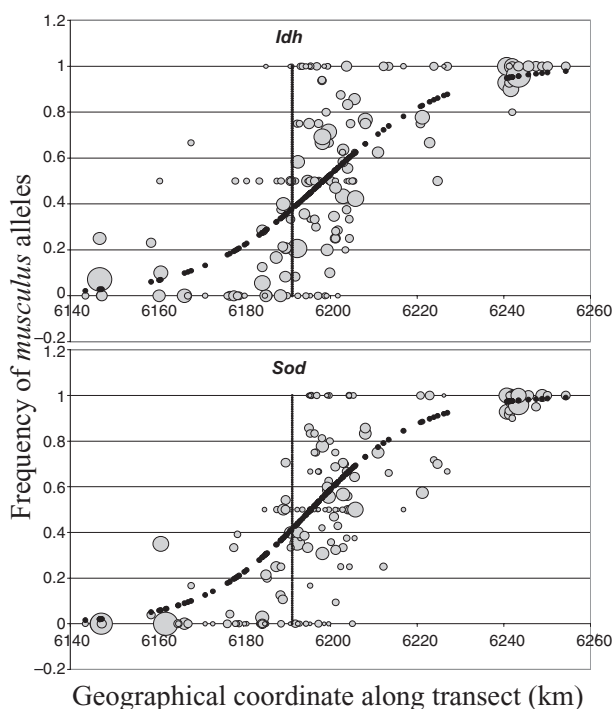
Figure 2. Variations of *Mus musculus musculus* allele frequencies along the transect across the hybrid zone, for eight autosomal loci. Shaded circles represent the real data, and the area of the circles is proportional to the effective number of alleles sampled (after correction for F_{IS} and F_{ST} , see text). The filled circles represent the best fit with a common centre for the eight loci, the position of which is shown by the thin vertical line.

that have coincident centres (*Amy*, *Es1*, *Es2*, *Es10*, *Gpd*, *Mpi*, *Np* and *Pgm*) using the maximum likelihood procedure in the Analyse package. With the estimated cline width for each locus given in Table 4, we

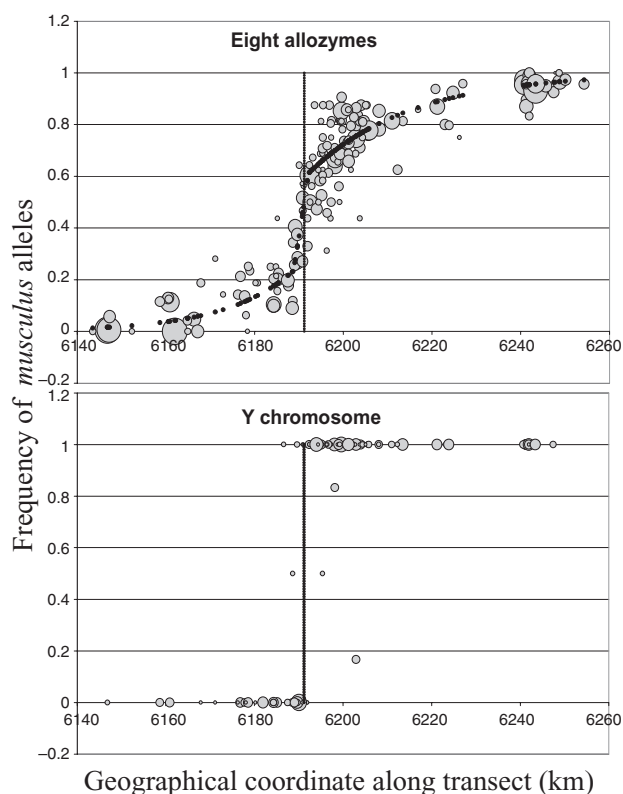
then obtained estimates of the migration parameter. All pairs of loci are unlinked on the genetic map, except *Es1* and *Es2*, at 9 cM on chromosome 8, and *Es10* and *Np*, at 22 cM on chromosome 14, and these

Table 4. Maximum likelihood estimates of cline parameters, and their two log-likelihood support limits, for eight loci with a common centre at position 6191.2

Locus	w	$B0/w$	θ_0	$B1/w$	θ_1
<i>Amy</i>	20.6 (15.2–26.4)	123.51 (2.27–inf)	0.002 (0.000–0.284)	9.67 (2.74–108.45)	0.037 (0.001–0.157)
<i>Es1</i>	11.0 (2.2–25.2)	0.67 (0.03–5.57)	0.172 (0.010–0.999)		
<i>Es10</i>	6.4 (4.9–11.7)	5.15 (1.16–332.67)	0.085 (0.004–0.457)	3.95 (1.81–7.54)	0.016 (0.008–0.055)
<i>Es2</i>	3.8 (1.3–9.2)	14.35 (1.91–123.26)	0.004 (0.000–0.058)	6.50 (1.66–23.25)	0.003 (0.000–0.019)
<i>Gpd</i>	6.3 (3.4–13.0)	2.59 (0.92–27.71)	0.029 (0.002–0.138)	1.09 (0.68–3.05)	0.032 (0.009–0.135)
<i>Mpi</i>	6.2 (2.2–18.9)	2.16 (0.35–9.01)	0.041 (0.004–0.453)		
<i>Np</i>	5.9 (3.8–44.6)	0.91 (0.15–inf)	0.045 (0.000–1.000)	2.39 (0.39–7.05)	0.009 (0.002–0.715)
<i>Pgm</i>	6.8 (4.0–9.1)	146.49 (9.98–inf)	0.016 (0.001–0.098)	2.19 (1.91–5.04)	0.035 (0.010–0.061)
Eight loci	8.9 (7.7–12.4)	2.25 (1.09–3.76)	0.072 (0.043–0.163)	2.22 (1.31–3.05)	0.033 (0.023–0.063)

**Figure 3.** Allele frequency variations along the transect for two autosomal loci that do not fit the general cline position and shape of the eight loci in Fig. 2. Symbols are as in Fig. 2.

values of recombination were used to estimate σ . The results were then averaged over pairs of loci and localities. We chose localities from the centre of the hybrid zone with more than ten mice in the sample (localities numbers 102, 106, 120 and 142: Appendix 1). This led to an average estimate of the migration parameter σ of 0.75 km generation^{-1/2}. We also applied the other method based on the variance of the hybrid index (Barton & Gale, 1993) to 145 mice from the 24 most central populations. The variance of hybrid index was 0.021, and the part due to heterozygosity 0.014. The excess variance is attributed to linkage disequilib-

**Figure 4.** Allele frequency variations along the transect, for the average of eight autosomal loci (*Amy*, *Es1*, *Es2*, *Es10*, *Gpd*, *Mpi*, *Np* and *Pgm*), and the Y chromosome (data from Dod *et al.*, 2005, this issue). Symbols are as in Fig. 2.

rium, and with an average cline width of 8.9 km, we get $D = 0.015$ and $\sigma = 0.77$ and 0.63 km generation^{-1/2} (whether sampling occurred before or after migration, Barton & Gale, 1993), in good agreement with the previous estimate.

In order to calculate the intensity of selection needed to balance dispersal against selection and recombination, cline shape has to be estimated. To do this we used the parameters of the average cline for

the eight coincident loci. We have seen above that, individually, most of the clines are asymmetrical, but that this asymmetry is tempered when the average introgression over the eight loci is considered. Using either the right or the left parameters of this cline fit gives consistent estimates of selection parameters. The barrier to gene flow on both sides is of the order of $B \sim 20$ km (2LL support limits 12–33 km on the left side and 14–28 on the right side), and the height of the central step in allele frequency equals 0.45. The total selection acting on the loci studied is estimated as $s^* = 0.040\text{--}0.059$ (eqn 2, with $\sigma = 0.63\text{--}0.77$, see above). The selection acting on each of the loci causing the genetic barrier is estimated as $s = 0.021\text{--}0.030$ (eqn 7). The number of loci under selection creating the central barrier is estimated to be $n \sim 52\text{--}78$ by assuming that the average recombination between the loci studied and the selected loci is $r = 0.5$ (eqn 8). This is thus an overestimate of the number of loci. The average fitness of hybrid central populations is estimated to be $W_H = 0.45$ (eqn 10). Finally, the selection acting on the loci under study is estimated using eqn (5) to be on average of the order of $s_{locus} = 0.003\text{--}0.004$, or $0.001\text{--}0.002$ depending on whether the left or right cline parameters, respectively, are considered. These low values are in agreement with the hypothesis that was made of weak selection on these loci in order to apply the approximations needed for the above parameter estimations.

ESTIMATING MIGRATION FROM MICROSATELLITE DATA

In order to estimate migration independently of the cline analyses, we typed some of the populations at six microsatellite loci and applied various methods of estimation of migration under isolation by distance. We selected the two sampling years with the largest samples (1992 and 1998) and analysed the results for each year separately. Here we are interested in the part of the genetic differentiation between populations and individuals that results from migration and drift alone, so we want to remove the effect of selection in the hybrid zone. Ideally, this could be done by analysing the residual variation of allele frequencies around the cline fits for polymorphic, yet diagnostic, markers. However, most microsatellites tested did not meet these criteria. We thus chose to study allele frequency variation among groups of populations more or less aligned in a direction perpendicular to that of the clinal gradient of the hybrid zone, by restricting the comparisons to pairs of populations whose average index of hybridization for the allozymes differed by less than 10% (Appendix 1).

The 1992 sample consisted of 185 individuals from 14 different localities. For 1998 we used 436 mice from 64 localities. Only 113 mice could be typed for locus

D17Mit14 in the 1992 sample (381 in the 1998 sample), but from 171 to 184 were typed at the five other loci (from 421 to 436 in 1998). The number of different alleles found at each locus is indicated in Table 1. The genotypes of these 621 mice at the six loci are given in Appendix 3. They include some 1992 data already reported in Dallas *et al.* (1995). We retyped a fraction of the mice from the earlier study to establish the correspondence of allele sizes between the two studies.

For the isolation by distance methods, we not only restricted the comparisons to pairs of populations with similar hybridization indexes, as argued above, but also to pairs not too far from each other, because it is recommended that pairs be generally separated by no more than 10σ (Rousset, 2000b), and in no case by more than 20σ (Rousset, 2000a). We chose a cut-off at 15 km. By using the F_{ST} method on pairs of populations selected in this way, we derived an estimate of $D\sigma^2$ of 3.7 individuals on the 1992 dataset, and of 2.7 individuals for 1998. This method relies on estimates of population frequencies, and in the regression, data points are not weighted according to sample sizes, so these values must be considered with caution. The \hat{a}_r and Moran I methods do not have this drawback, because they are based on comparisons between individuals, but they appear to be less adapted to the habitat structure of mice, which is fragmented rather than continuous. They gave consistent estimates of $D\sigma^2$, but showed differences between 1992 ($D\sigma^2 = 1.4$ and 1.7 for the two methods, respectively) and 1998 ($D\sigma^2 = 3.7$ and 3.9). The 1998 estimate is presumably more reliable because many more localities were sampled, so that the range of pairwise distances between individuals is covered much better.

In order to derive estimates of σ , we must now attempt to estimate the density of mice, which we can do by using our trapping data. In 1998, the year with most intense trapping, we underwent a random prospecting of farms, and found that 25% of them had mice. The average number of mice trapped per farm with mice was 7.2 (with a large variance, see Table 2). Because we intensively trapped in all cases, and stopped the effort only after several unsuccessful trapping nights, we believe this is a reasonable estimate of the mouse population present. It necessarily provides a lower limit of the number of mice existing at the time of trapping, but it includes at least two generations (we counted both adults and young), and so probably overestimates population densities per generation. The density of farms in the prospected areas was accurately determined by counting them on a map, leading to an average of 3.1 farms km^{-2} . This gives an estimated density D of 5.8 mice km^{-2} . The average $D\sigma^2$ estimated using the three different isolation by distance methods ranged from 1.4 to 3.9, which gives $\sigma = 0.51\text{--}0.82$ km generation $^{-1/2}$. If our density esti-

mates are really lower limits these estimates of σ are upper limits, but they are compatible with the estimates derived from the cline analyses. With the lowest value, $\sigma = 0.51$ km generation^{-1/2}, we estimate that the barrier would be created by a larger number ($n = 120$) of loci under weaker selection ($s = 0.013$) than what is found when estimating migration from linkage disequilibrium. The highest value of $\sigma = 0.82$ gives $n = 46$ loci with $s = 0.035$. The results of the two approaches are comparable, given the numerous sources of uncertainty associated with these estimates.

DISCUSSION

For eight of the ten allozyme loci studied, the changes in allele frequencies can be best described by steep central steps of allele frequency that are coincident and flanked by smooth tails of introgression on either side. This indicates the presence of a barrier to gene flow in the centre of the zone. We assumed that this barrier is caused by selection against hybrids, creating a tension zone, and applied the existing theory to estimate the selection and migration parameters shaping this hybrid zone.

As quantifying migration is essential to infer selection, we also estimated migration independently from the cline analysis, using microsatellites and comparing populations of similar positions in the overall gradient of the hybrid zone. The results we obtained using the different approaches are in good agreement ($\sigma = 0.5\text{--}0.8$ km generation^{-1/2}). There are few relevant data in the literature with which to compare this estimate. Many capture–recapture experiments give discordant results (Lidicker & Patton, 1987), and often such studies concern small study areas and would miss long-distance migrants (Baker, 1981; see Pocock, Hauffe & Searle, 2005, this issue). Myers (1974) witnessed colonization events between grids separated by 92 m. According to Berry & Jakobson (1974) 25% of the mice on an island mate within 50–100 m from their birthplace. Carlsen (1993) captured mice in farms and surrounding fields and reports movements of up to 130 m. A few studies were able to detect longer distance migrants. Walkowa, Adamczyk & Chelkowska (1989) trapped 6.5% of the mice living in an enclosure in surrounding fields, up to 300 m away. Cassaing & Croset (1985) recaptured most mice within 50–480 m in outdoor populations. Auffray *et al.* (1990a) report on mice that had covered distances of up to 700 m, and detected immigration from populations up to 820 m apart. Berry (1968) describes more than 30 mice moving 500 m and 14 moving 1500 m. Migrations over 1000 m are also reported by other authors (Pearson, 1963; Tomich, 1970). Although most mice probably reproduce close to their birthplace, they can apparently easily reach regions several hundreds

of metres away. In addition, humans can passively transport them over much longer distances. In Figure 5 it can be seen that substantial linkage disequilibrium is sometimes found rather far from the centre of the hybrid zone, and it could be due to such long-distance migrations. However, we found no mouse heterozygous at all loci in our dataset, and thus no F₁ mice that could result from such long-distance migrations. Potential first-generation backcrosses, heterozygous at half of their loci, were rare: only one was found on the *domesticus* side of the zone (among 151 mice from localities with a coordinate lower than 6182 and with data available at eight loci or more). Six such mice were found on the *musculus* side (among 359 mice from localities with a coordinate above 6200), but only from the localities closest to the central step of the hybrid zone (coordinates from 6200–6205).

The average barrier to gene flow appears relatively moderate (20 km) compared with migration (0.5–0.8 km generation^{-1/2}), but varies considerably between loci in our estimates, as well as the level of introgression and degree of asymmetry. Two of the loci we studied (*Idh* and *Sod*) did not show the typical central step, but rather wide clines with extensive introgression into *musculus*, which could be an indication that they have escaped the central barrier. The delay to introgression of neutral alleles across such a barrier is expected to be of the order of 500–1600 generations (B^2/σ^2 ; Barton & Hewitt, 1985, and references therein). There are presumably two reproduction periods per year, one indoors in autumn that we observed during our trapping campaigns and a second one in spring, mostly in the fields. This would, however, represent only one generation a year in terms of migration. According to what is believed about the patterns of expansion of these subspecies (reviewed in Boursot

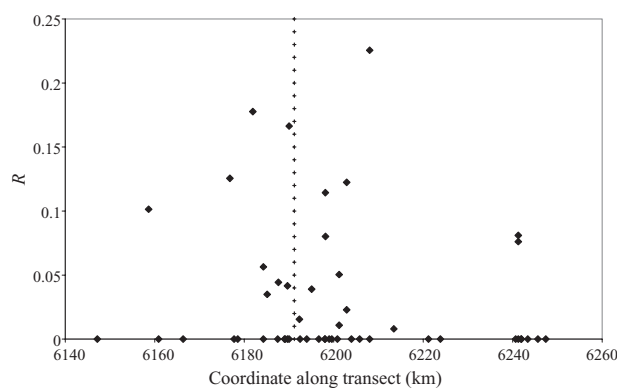


Figure 5. Average standardized linkage disequilibrium (R) between the eight loci of Fig. 2 in localities along the transect. Only localities with eight mice or more were selected. Crosses indicate the position of the centre of the hybrid zone.

et al., 1993), being at the most northern end of the hybrid zone the transect in Jutland should be youngest. The few available archaeological records point to an earliest occurrence of house mice in north-western Europe around 2000 years BP (Auffray, Vanlerberghe & Britton-Davidian, 1990b; Auffray & Britton-Davidian, 1992; Cucchi, Vigne & Auffray, 2005, this issue). Unless the hybrid zone in Jutland is much younger than this, which cannot be formally excluded, there would have been time for neutral genes to escape the barrier. Our results also imply that the barrier created by this hybrid zone would be very weak against selectively favourable alleles [$T = \log(B^2\pi s/2\sigma^2)/2s$ only about 100 years for a selective advantage s of 1%]. It will be interesting to determine whether such selective introgression has occurred. Because the loci we studied were not chosen at random, but for their diagnosticity in this part of the hybrid zone, they are not suited to address this issue.

With a number of approximations, our analyses provide a rather low estimate of fitness for the central populations ($W_H = 0.45$). Indirect evidence of the lower fitness of hybrid populations comes from their observed higher load of intestinal parasites (Sage *et al.*, 1986a; Moulia *et al.*, 1991) and the confirmation of the probable genetic origin of such differences (Moulia *et al.*, 1993). More direct evidence comes from the observation of a slightly reduced testis size in central populations, as well as very high sterility rates of males and substantial reduction of female fertility in F_1 crosses between Danish strains of the two subspecies (Britton-Davidian *et al.*, 2005, this issue). However, it has also been shown that the natural and artificial hybrids display a better general bilateral symmetry of skull morphology, which was taken as evidence of heterosis on this character (Alibert *et al.*, 1994, 1997; Auffray *et al.*, 1996).

This relatively narrow hybrid zone is expected to be created by a moderate number of loci, and if their number was low enough, they could be searched by scanning the genome. The range of values of n that we derived (46–120) might appear to be rather high for this to be feasible. However, it is based on the assumption of selection against heterozygotes, neglecting potentially strong epistasis between loci. Furthermore, in deriving our estimates of the number of loci under selection, we used the parameters of the average cline for eight loci. Averaging over loci whose clines differ in width and symmetry (although asymmetry is always in the same direction) probably tends to overestimate the width, and thus the number of loci under selection (as does the approximation of $r = 0.5$ between the studied and selected loci). The very limited introgression of Y and X chromosome markers in several transects (Tucker *et al.*, 1992; Dod *et al.*, 1993, 2005 (this issue); Prager, Boursot & Sage, 1997) clearly

points to the possibility that at least some loci are under strong enough selection to be detected using this approach. With progress in the knowledge of the mouse genome (Waterston *et al.*, 2002), and the discovery of numerous nucleotide variations between the genomes of the house mouse subspecies (Wade *et al.*, 2002), it should be possible to identify the regions of the genome most involved in selection in this hybrid zone (see Paysseur & Nachman, 2005, this issue). Given the difficulties in fitting cline models to these data however, it may prove difficult to identify the loci under selection.

An intriguing aspect of our cline shape analyses is that the central step hardly overlaps frequency 0.5, but rather occurs almost entirely below this value, on the *domesticus* side (particularly for *Es10*, *Es2*, *Gpd* and *Pgm*, Fig. 2). We wondered whether this could result from the presence of geographical barriers to migration. The steep edged river valley just south of the inferred centre that is highlighted on the map (Fig. 1) could act as such. So could the large area of heathland that occupies the area between the east- and westward-flowing rivers. There were only a few scattered farms in this area and in spite of intensive trapping efforts we were unable to find mice there, which is reflected by the absence of samples in this area (Fig. 1). These geographical accidents could combine to restrict gene flow in most of the region slightly south of the inferred centre, creating an abrupt change in allele frequencies. Depending on the frequencies reached south and north of this geographical barrier, these steps could either be confounded into a single step by the inference method, producing a symmetrical cline (*Amy*, *Es1*, *Mpi* and perhaps *Np*), or the geographical rather than the genetic step could be captured by the method, leading to asymmetric clines with a step on the *domesticus* side (*Es10*, *Gpd*, *Pgm*). This would imply that the hybrid zone might be moving south towards the geographical barrier. Loci not too tightly linked to selected loci would be lagging behind the selection front, leaving long traces on the *musculus* side. Loci more tightly linked to selected loci would tend to follow the front, and leave less trace. As the latter loci will reach higher *musculus* frequencies north of the geographical barrier, they should tend to show more introgression into *domesticus*, a counterintuitive conclusion because they are submitted to stronger selection. As the *Idh* and *Sod* loci appear to have escaped the barrier to gene flow, the centre of their cline could represent the position of the original contact between the subspecies, about 8 km north of the present centre. If these loci are not subject to selection, the width of their clines depends only on the time t elapsed since the initial contact between the subspecies ($w = 2.51\sigma\sqrt{t}$). With an average width of 54 km for these two loci, migration between 0.5 and 0.8 km generation^{-1/2}, and considering one migration

episode a year (see above), we find that the initial contact would have occurred between 700 and 1800 years ago, which is compatible with the archaeological data. This would represent an imperceptible movement during the period that has elapsed since the Hunt & Selander (1973) study. However, the Y chromosome data appear difficult to reconcile with this scenario. Because it appears to be the locus under strongest selection in this transect, its cline should have been the first to be trapped in the presumed density trough, which does not seem to be the case. Barriers to migration such as the one we suspect here could occur elsewhere in Jutland. The complex interaction between such barriers, genetic drift between the barriers and selection could uncouple the dynamics of epistatically interacting loci, producing staggered clines, explaining the differences we see between loci. In such a situation it is not necessary that the major transition of allele frequency of neutral loci occur at 50%. Three Robertsonian (Rb) chromosomal fusions segregate in the *domesticus* populations just south of the centre in the transect we studied, and Fel-Clair *et al.* (1996) found that their frequencies decreased in a series of staggered clines before they could reach the centre of the hybrid zone. Experimental crosses have suggested a complex negative interaction between the centromeres of the two subspecies (Lenormand *et al.*, 1997), and a centromeric marker on a chromosome not involved in a Rb fusion seemed, on the basis of the limited sample available, to show no introgression into *domesticus* (Fel-Clair *et al.*, 1998). It could thus be that asymmetrical centromeric interactions contribute to impede introgression into *domesticus* in this transect, in addition to potential geographical barriers.

However, greater introgression into *musculus* for several diagnostic loci was noted in the broader study of the Jutland hybrid zone by Hunt & Selander (1973), but also in other geographical regions, despite less intense sampling (Bulgaria, Vanlerberghe *et al.*, 1988; Bavaria, Tucker *et al.*, 1992; East Holstein, Prager *et al.*, 1993). This apparently general pattern calls for a general explanation. Fisher (1937) predicted that in the wave of advance of advantageous alleles, the leading edge should be longer than the trailing edge. The asymmetry observed in the mouse hybrid zone could thus witness a general movement of the zone in a direction opposite to that predicted above, and reflect an overall advantage of *domesticus* over *musculus*. Asymmetric behavioural selection would best account for such a situation. There is laboratory evidence that mice have the capacity to discriminate urine odours of their own subspecies from those of the other (Smadja & Ganem, 2002; Ganem *et al.*, 2005, this issue). Odour preference tests performed on wild-derived mice from the tails of the Danish hybrid zone have revealed a

clear assortative choice in *M. m. musculus*, but no preference in *domesticus* (Smadja, Catalan & Ganem, 2004). The consequences of this behavioural difference on gene flow are not straightforward, but one could think that it should prevent introgression into *musculus*, rather than promote it. However, other behavioural asymmetries, such as a greater aggressiveness of *domesticus* males, could act in the opposite direction. It has also been suggested that the choosiness of the *musculus* females could result from reinforcement in the hybrid zone (Smadja *et al.*, 2004), in which case this behaviour could be a consequence of the asymmetry of introgression, rather than its cause.

A comparison with patterns observed in other parts of this hybrid zone would be very informative, as it would allow attribution of the confirmed patterns to deterministic forces, and the others to local historical, geographical or stochastic accidents.

ACKNOWLEDGEMENTS

Tonnes Nielsen provided invaluable support to the fieldwork. Patrick Makoundou performed most microsatellite analyses. Many thanks to Guila Ganem for sharing some samples, and to the Danish farmers for their kind co-operation.

REFERENCES

- Alibert P, Fel-Clair F, Manolakou K, Britton-Davidian J, Auffray J-C. 1997.** Developmental stability, fitness, and trait size in laboratory hybrids between European subspecies of the house mouse. *Evolution* **51**: 1284–1295.
- Alibert P, Renaud S, Dod B, Bonhomme F, Auffray J-C. 1994.** Fluctuating asymmetry in the *Mus musculus* hybrid zone: a heterotic effect in disrupted co-adapted genomes. *Proceedings of the Royal Society of London B* **258**: 53–59.
- Auffray J-C, Alibert P, Latieule C, Dod B. 1996.** Relative warp analysis of skull shape across the hybrid zone of the house mouse (*Mus musculus*) in Denmark. *Journal of Zoology* **240**: 441–455.
- Auffray J-C, Belkhir K, Cassaing J, Britton-Davidian J, Croset H. 1990a.** Outdoor occurrence in Robertsonian and standard populations of the house mouse. *Vie et Milieu* **40**: 111–118.
- Auffray J-C, Britton-Davidian J. 1992.** When did the house mouse colonize Europe? *Biological Journal of the Linnean Society* **45**: 187–190.
- Auffray J-C, Vanlerberghe F, Britton-Davidian J. 1990b.** House mouse progression in Eurasia: a palaeontological and archaeozoological approach. *Biological Journal of the Linnean Society* **41**: 13–25.
- Baker AEM. 1981.** Gene flow in house mice: behavior in a population cage. *Behavioural Ecology and Sociobiology* **8**: 83–90.
- Barton N, Bengtsson BO. 1986.** The barrier to genetic exchange between hybridising populations. *Heredity* **56**: 357–376.

- Barton NH, Gale KS. 1993.** Genetic analysis of hybrid zones. In: Harrison RG, ed. *Hybrid zones and the evolutionary process*. Oxford: Oxford University Press, 13–45.
- Barton NH, Hewitt GM. 1985.** Analysis of hybrid zones. *Annual Review of Ecology and Systematics* **16**: 113–148.
- Barton NH, Shpak M. 2000.** The effect of epistasis on the structure of hybrid zones. *Genetical Research* **75**: 179–198.
- Berry RJ. 1968.** Ecology of an island population of house mouse. *Journal of Animal Ecology* **37**: 445–470.
- Berry RJ, Jakobson ME. 1974.** Vagility in an island population of house mouse. *Journal of Zoology* **173**: 341–354.
- Bonhomme F, Catalan J, Britton-Davidian J, Chapman VM, Moriwaki K, Nevo E, Thaler L. 1984.** Biochemical diversity and evolution in the genus *Mus*. *Biochemical Genetics* **22**: 275–303.
- Boursot P, Auffray J-C, Britton-Davidian J, Bonhomme F. 1993.** The evolution of house mice. *Annual Review of Ecology and Systematics* **24**: 119–152.
- Britton-Davidian J. 1990.** Genic differentiation in *M. m. domesticus* populations from Europe, the Middle East and North Africa: geographic patterns and colonization events. *Biological Journal of the Linnean Society* **41**: 27–45.
- Britton-Davidian J, Fel-Clair F, Lopez J, Alibert P, Boursot P. 2005.** Postzygotic isolation between the two European subspecies of the house mouse: estimates from fertility patterns in wild and laboratory-bred hybrids. *Biological Journal of the Linnean Society* **84**: 379–393.
- Carlsen M. 1993.** Migrations of *Mus musculus musculus* in Danish Farmland. *Zeitschrift für Säugetierkunde* **58**: 172–180.
- Cassaing J, Croset H. 1985.** Organisation spatiale, compétition et dynamique des populations sauvages de souris (*Mus spretus* Lataste et *Mus musculus domesticus* Ruddy) du Midi de la France. *Zeitschrift für Säugetierkunde* **50**: 271–284.
- Cucchi T, Vigne J-D, Auffray J-C. 2005.** First occurrence of the house mouse (*Mus musculus domesticus* Schwarz & Schwarz, 1943) in the Western Mediterranean: a zooarchaeological revision of sub-fossil house mouse occurrences. *Biological Journal of the Linnean Society* **84**: 429–445.
- Dallas JF, Dod B, Boursot P, Prager E, Bonhomme F. 1995.** Population genetic structure and gene flow in Danish house mice. *Molecular Ecology* **4**: 311–320.
- Din W, Anand R, Boursot P, Darviche D, Jouvin-Marche E, Dod B, Orth A, Talwar GP, Cazenave P-A, Bonhomme F. 1996.** Origin and radiation of the house mouse: clues from nuclear genes. *Journal of Evolutionary Biology* **9**: 519–539.
- Dod B, Jermiin LS, Boursot P, Chapman VM, Nielsen JT, Bonhomme F. 1993.** Counterselection on sex chromosomes in the *Mus musculus* European hybrid zone. *Journal of Evolutionary Biology* **6**: 529–546.
- Dod B, Smadja C, Karn RC, Boursot P. 2005.** Testing for selection on the androgen-binding protein in the Danish mouse hybrid zone. *Biological Journal of the Linnean Society* **84**: 447–459.
- Fel-Clair F, Catalan J, Lenormand T, Britton-Davidian J. 1998.** Centromeric incompatibilities in the hybrid zone between house mouse subspecies from Denmark: evidence from patterns of NOR activity. *Evolution* **52**: 592–603.
- Fel-Clair F, Lenormand T, Catalan J, Grobert J, Orth A, Boursot P, Viroux M-C, Britton-Davidian J. 1996.** Genomic incompatibilities in the hybrid zone between house mice in Denmark: evidence from steep and non-coincident chromosomal clines for Robertsonian fusions. *Genetical Research* **67**: 123–134.
- Fisher RA. 1937.** The wave of advance of advantageous genes. *Annals of Eugenics* **7**: 355–369.
- Gaggiotti OE, Lange O, Rassmann K, Gliddon C. 1999.** A comparison of two indirect methods for estimating average levels of gene flow using microsatellite data. *Molecular Ecology* **8**: 1513–1520.
- Ganem G, Ginane C, Ostrowski M-F, Orth A. 2005.** Assessment of mate preference in the house mouse with reference to investigations on assortative mating. *Biological Journal of the Linnean Society* **84**: 461–471.
- Hardy OJ, Vekemans X. 1999.** Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. *Heredity* **83**: 145–154.
- Harrison RG. 1990.** Hybrid zones: windows on evolutionary processes. In: Futuyma D, Antonovics J, eds. *Oxford surveys in evolutionary biology* vol. 7. Oxford: Oxford University Press, 70–128.
- Hilborn R, Mangel M. 1997.** *The ecological detective: confronting models with data*. Princeton: Princeton University Press.
- Hunt WG, Selander RK. 1973.** Biochemical genetics of hybridisation in European house mice. *Heredity* **31**: 11–33.
- Kruuk LE, Baird SJ, Gale KS, Barton NH. 1999.** A comparison of multilocus clines maintained by environmental adaptation or by selection against hybrids. *Genetics* **153**: 1959–1971.
- Lenormand T, Fel-Clair F, Manolakou K, Alibert P, Britton-Davidian J. 1997.** Chromosomal transmission bias in laboratory hybrids between wild strains of the two European subspecies of mice. *Genetics* **147**: 1279–1287.
- Lidicker WZ Jr, Patton JL. 1987.** Patterns of dispersal and genetic structure in populations of small rodents. In: Chepko-Sade BD, Haplin ZT, eds. *Mammalian dispersal patterns*. Chicago: University of Chicago Press, 144–161.
- Mouliac C, Aussel J-P, Bonhomme F, Boursot P, Nielsen JT, Renaud F. 1991.** Wormy mice in a hybrid zone: a genetic control of susceptibility to parasite infection. *Journal of Evolutionary Biology* **4**: 679–687.
- Mouliac C, Le Brun N, Dallas J, Orth A, Renaud F. 1993.** Experimental evidence of genetic determinism in high susceptibility to intestinal pinworms infection in mice: a hybrid zone model. *Parasitology* **106**: 387–393.
- Myers JH. 1974.** Genetic and social-structure of feral house mouse populations on Grizzly-Island, California. *Ecology* **55**: 747–759.
- Pasteur N, Pasteur G, Bonhomme F, Catalan J, Britton-Davidian J. 1987.** *Manuel de génétique par électrophorèse des protéines*. Paris: Lavoisier.
- Payseur BA, Nachman MW. 2005.** The genomics of speciation: investigating the molecular correlates of X chromosome

- introgression across the hybrid zone between *Mus domesticus* and *Mus musculus*. *Biological Journal of the Linnean Society* **84**: 523–534.
- Pearson OP. 1963.** History of 2 local outbreaks of feral house mice. *Ecology* **44**: 540–549.
- Phillips BL, Baird SJE, Moritz C. 2004.** When vicars meet: a narrow contact zone between morphologically cryptic phylogeographic lineages of the rainforest skink, *Carlia rubrigularis*. *Evolution* **58**: 1536–1548.
- Pocock MJO, Hauffe HC, Searle JB. 2005.** Dispersal in house mice. *Biological Journal of the Linnean Society* **84**: 565–583.
- Prager EM, Boursot P, Sage RD. 1997.** New assays for Y chromosome and p53 pseudogene clines among East Holstein house mice. *Mammalian Genome* **8**: 279–281.
- Prager EM, Sage RD, Gyllensten U, Thomas WK, Hübner R, Jones CS, Noble L, Searle JB, Wilson AC. 1993.** Mitochondrial DNA sequence diversity and the colonization of Scandinavia by house mice from East Holstein. *Biological Journal of the Linnean Society* **50**: 85–122.
- Raymond M, Rousset F. 1995.** Genepop (Version 1.2) – Population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**: 248–249.
- Rousset F. 1997.** Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* **145**: 1219–1228.
- Rousset F. 2000a.** Genetic approaches to the estimation of dispersal rates. In: Clobert J, Danchin E, Dhondt AA, Nichols JD, eds. *Dispersal*. Oxford: Oxford University Press, 18–28.
- Rousset F. 2000b.** Genetic differentiation between individuals. *Journal of Evolutionary Biology* **13**: 58–62.
- Sage RD, Heyneman D, Lim K-C, Wilson AC. 1986a.** Wormy mice in a hybrid zone. *Nature* **324**: 60–63.
- Sage RD, Whitney JB III, Wilson AC. 1986b.** Genetic analysis of a hybrid zone between *domesticus* and *musculus* mice (*Mus musculus* complex): hemoglobin polymorphisms. In: Potter M, Nadeau JH, Cancro MP, eds. *The wild mouse in immunology*. Berlin: Springer-Verlag, 75–85.
- Slatkin M. 1995.** A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**: 457–462.
- Smadja C, Catalan J, Ganem G. 2004.** Strong premating divergence in a unimodal hybrid zone between two subspecies of the house mouse. *Journal of Evolutionary Biology* **17**: 165–176.
- Smadja C, Ganem G. 2002.** Subspecies recognition in the house mouse: a study of two populations from the border of a hybrid zone. *Behavioral Ecology* **13**: 312–320.
- Szymura JM, Barton NH. 1986.** Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina orientalis* and *B. variegata*, near Cracow in Southern Poland. *Evolution* **40**: 1141–1159.
- Szymura JM, Barton NH. 1991.** The genetic structure of the hybrid zone between the fire-bellied toads *Bombina orientalis* and *B. variegata*: comparisons between transects and between loci. *Evolution* **45**: 237–261.
- Tomich PQ. 1970.** Movement patterns of field rodents in Hawaii. *Pacific Science* **24**: 195–234.
- Tucker PK, Sage RD, Warner J, Wilson AC, Eicher EM. 1992.** Abrupt cline for sex chromosomes in a hybrid zone between two species of mice. *Evolution* **46**: 1146–1163.
- Vanlerberghe F, Boursot P, Catalan J, Gerasimov S, Bonhomme F, Botev BA, Thaler L. 1988.** Analyse génétique de la zone d'hybridation entre les deux sous-espèces de souris *Mus musculus domesticus* et *Mus musculus musculus* en Bulgarie. *Genome* **30**: 427–437.
- Vanlerberghe F, Dod B, Boursot P, Bellis M, Bonhomme F. 1986.** Absence of Y-chromosome introgression across the hybrid zone between *Mus musculus domesticus* and *Mus musculus musculus*. *Genetical Research* **48**: 191–197.
- Wade CM, Kulbokas EJ 3rd, Kirby AW, Zody MC, Mullikin JC, Lander ES, Lindblad-Toh K, Daly MJ. 2002.** The mosaic structure of variation in the laboratory mouse genome. *Nature* **420**: 574–578.
- Walkowa W, Adamczyk K, Chelkowska H. 1989.** Characteristics of migrants in a free-living population of the house mouse. *Acta Theriologica* **34**: 305–313.
- Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, Antonarakis SE, Attwood J, Baertsch R, Bailey J, Barlow K, Beck S, Berry E, Birren B, Bloom T, Bork P, Botcherby M, Bray N, Brent MR, Brown DG, Brown SD, Bult C, Burton J, Butler J, Campbell RD, Carninci P, Cawley S, Chiaromonte F, Chinwalla AT, Church DM, Clamp M, Clee C, Collins FS, Cook LL, Copley RR, Coulson A, Couronne O, Cuff J, Curwen V, Cutts T, Daly M, David R, Davies J, Delehaunty KD, Deri J, Dermitzakis ET, Dewey C, Dickens NJ, Diekhans M, Dodge S, Dubchak I, Dunn DM, Eddy SR, Elnitski L, Emes RD, Eswara P, Eyraes E, Felsenfeld A, Fewell GA, Flicek P, Foley K, Frankel WN, Fulton LA, Fulton RS, Furey TS, Gage D, Gibbs RA, Glusman G, Gnerre S, Goldman N, Goodstadt L, Grafham D, Graves TA, Green ED, Gregory S, Guigo R, Guyer M, Hardison RC, Haussler D, Hayashizaki Y, Hillier LW, Hinrichs A, Hlavina W, Holzer T, Hsu F, Hua A, Hubbard T, Hunt A, Jackson I, Jaffe DB, Johnson LS, Jones M, Jones TA, Joy A, Kamal M, Karlsson EK *et al.* 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**: 520–562.**

APPENDIX 1

List of the 170 localities studied, with their number, name, year of collection, longitude, latitude (in UTM coordinates), coordinate on the transect (see text), number of mice studied and average hybrid index (% *musculus* alleles) for the eight loci with coinciding cline centres (*Amy*, *Es1*, *Es10*, *Es2*, *Gpd*, *Mpi*, *Np* and *Pgm*).

Locality no.	Locality name	Year	Longitude	Latitude	Transect coordinate	<i>N</i>	HI eight loci
186	Ringtved	2000	522.1	6125.5	6143.5	1	0.000
113	Simmersted	1985	525.0	6128.5	6146.8	2	0.031
114	Simmersted	1990	525.0	6128.5	6146.8	14	0.005
115	Simmersted A	1990	525.0	6128.5	6146.8	4	0.000
181	Kastvra	2000	521.4	6129.1	6146.9	2	0.000
182	Kastvra A	2000	522.2	6129.2	6147.2	14	0.006
187	Sillerup	2000	535.7	6128.4	6147.3	7	0.057
189	Slusen	2000	538.2	6133.7	6152.3	1	0.000
92	Oedis	1984	524.4	6140.5	6158.6	13	0.115
88	Maltbaek Mark	1992	505.1	6144.8	6160.5	4	0.129
93	Oedis F	1985	526.3	6142.3	6160.7	2	0.125
87	Maltbaek	1992	504.1	6145.3	6160.9	10	0.113
184	Lauritzminde	2000	541.4	6144.1	6162.0	17	0.000
175	Binderup	2000	535.1	6143.3	6162.2	3	0.000
188	Skartved	2000	534.2	6145.9	6164.8	2	0.042
190	Soender Stenderup	2000	538.7	6146.5	6164.9	1	0.000
174	Baeklund	2000	516.2	6149.3	6166.4	33	0.047
179	Gelballe	2000	522.6	6149.3	6167.1	5	0.000
154	Vranderup	1998	524.8	6149.8	6167.9	3	0.188
33	Gesten	1998	511.2	6154.7	6171.1	2	0.281
133	Trelde	1998	549.1	6162.3	6172.9	1	0.143
140	Vejlby	1998	545.4	6163.0	6176.2	2	0.143
10	Baekke	1998	509.7	6160.6	6176.8	12	0.213
151	Viuf	1998	531.7	6158.9	6177.7	9	0.136
153	Vorbasse sk	1992	507.1	6162.2	6178.1	1	0.063
101	Rands A	1998	543.7	6164.2	6178.4	1	0.000
152	Vorbasse	1992	503.5	6163.2	6178.6	23	0.251
100	Rands	1998	543.4	6164.8	6178.9	4	0.234
124	Starup	1998	487.1	6166.8	6180.2	1	0.188
24	Egtved 1	1990	519.5	6163.3	6180.7	1	0.188
7	Ammistbol	1990	525.6	6165.5	6183.6	4	0.250
128	Toerskind	1984	519.9	6166.9	6184.3	18	0.104
129	Toerskind	1985	519.9	6166.9	6184.3	21	0.099
130	Toerskind	1991	519.9	6166.9	6184.3	4	0.203
8	Ammistbol A	1992	527.3	6166.5	6184.8	1	0.250
137	Tufkaer Mark	1998	531.6	6166.1	6184.9	1	0.214
94	Oedsted	1990	523.9	6167.3	6185.1	2	0.156
95	Oedsted	1991	523.9	6167.3	6185.1	1	0.438
96	Oedsted	1992	523.9	6167.3	6185.1	14	0.201
14	Bindeballe	1992	517.6	6168.3	6185.4	5	0.225
117	Skoldbjerg	1992	505.9	6171.9	6187.5	12	0.198
52	Hojen Kirke	1998	531.0	6168.9	6187.6	13	0.178
72	Jerlev V	1992	527.5	6170.2	6188.5	8	0.090
71	Jerlev A	1998	528.0	6170.3	6188.6	5	0.118
118	Smakaer	1998	520.4	6171.2	6188.6	2	0.344
134	Troellund	1992	502.5	6173.8	6189.0	14	0.257
102	Ravning	1992	521.9	6171.6	6189.1	10	0.405
125	Store Lihme	1987	519.7	6172.3	6189.6	9	0.375
99	Plougslund A	1992	505.9	6174.0	6189.7	12	0.286
98	Plougslund	1992	506.0	6174.1	6189.7	17	0.263
70	Jerlev	1998	528.0	6171.7	6190.0	1	0.643

APPENDIX 1 *Continued*

Locality no.	Locality name	Year	Longitude	Latitude	Transect coordinate	<i>N</i>	HI eight loci
73	Kaerbolling	1987	525.4	6172.7	6190.7	1	0.571
145	Vilstrup	1992	529.4	6172.3	6190.8	3	0.271
20	Bredsten	1985	523.2	6173.1	6190.8	5	0.516
21	Bredsten	1987	523.2	6173.1	6190.8	2	0.467
26	Englesholm Mark	1998	521.1	6173.6	6191.0	1	0.563
13	Balle A	1987	523.7	6174.1	6191.9	1	0.438
44	Grene Kirke	1998	504.4	6176.5	6191.9	6	0.330
11	Balle	1992	522.8	6174.7	6192.3	17	0.604
12	Balle	1998	522.8	6174.7	6192.3	2	0.643
120	Soedover	1985	521.6	6175.0	6192.5	10	0.500
121	Soedover	1992	521.6	6175.0	6192.5	1	0.500
80	Lildfrost	1998	525.9	6174.8	6192.9	2	0.563
15	Braendgaarde	1998	512.2	6176.7	6193.1	4	0.672
45	Grinsted	1998	497.9	6178.9	6193.5	1	0.875
142	Vester Hornstrup	1998	526.2	6175.9	6194.0	39	0.473
138	Uhre	1998	531.6	6175.6	6194.4	2	0.625
143	Vester Hornstrup A	1998	526.8	6176.2	6194.4	1	0.500
31	Gammelby	1992	520.8	6177.4	6194.8	3	0.583
32	Gammelby	1998	520.8	6177.4	6194.8	1	0.688
25	Elkaer	1998	504.8	6179.5	6195.0	1	0.813
106	Rugballe	1992	527.8	6176.8	6195.1	14	0.527
107	Rugballe A	1992	527.7	6177.0	6195.3	1	0.688
22	Brunbjerg	1998	502.8	6180.1	6195.3	1	0.688
61	Hovertoft	1998	532.5	6176.5	6195.4	3	0.750
108	Rugballe B	1991	528.0	6177.1	6195.4	3	0.708
109	Rugballe B	1992	528.0	6177.1	6195.4	3	0.604
110	Rugballe B	1998	528.0	6177.1	6195.4	3	0.583
60	Hover Kirke A	1998	531.0	6176.7	6195.4	1	0.625
81	Lille Hoegsholt	1998	531.0	6176.8	6195.4	1	0.875
59	Hover Kirke	1998	529.9	6177.1	6195.6	2	0.688
97	Oelgod	1998	478.9	6184.0	6196.3	1	0.313
135	Trollerup	1992	522.7	6178.7	6196.3	3	0.458
136	Trollerup	1998	522.7	6178.7	6196.3	2	0.719
82	Lindeballe	1992	512.3	6180.3	6196.7	10	0.613
67	Jelling bis	1987	526.1	6178.8	6196.9	3	0.875
83	Lindeballe A	1992	513.0	6180.7	6197.1	1	0.438
74	Kiddelund	1998	522.9	6179.5	6197.2	3	0.804
89	Moelvang	1990	522.9	6179.6	6197.3	2	0.500
90	Moelvang	1992	522.9	6179.6	6197.3	1	0.750
66	Jelling	1990	526.5	6179.3	6197.3	3	0.816
50	Hoerup	1990	530.2	6179.5	6198.0	8	0.672
51	Hoerup	1991	530.2	6179.5	6198.0	25	0.711
48	Hoegelund	1998	509.8	6182.1	6198.1	14	0.668
47	Hjortlund	1998	504.2	6182.8	6198.2	10	0.639
30	Gadbjerg	1987	520.5	6181.2	6198.6	1	0.667
91	Norskov	1992	515.3	6182.1	6198.8	2	0.656
78	Langelund S	1998	508.3	6183.2	6199.0	13	0.561
131	Tofthoj	1987	521.2	6181.7	6199.1	1	0.500
119	Smidstrup	1992	516.9	6182.4	6199.3	5	0.688
29	Filskov A	1992	500.1	6184.5	6199.3	1	0.813
16	Brandbjerg	1985	529.3	6181.2	6199.6	4	0.817
17	Brandbjerg	1991	529.3	6181.2	6199.6	28	0.853
18	Brandbjerg	1992	529.3	6181.2	6199.6	1	0.688
63	Hygum	1990	527.4	6181.4	6199.6	1	0.813

APPENDIX 1 *Continued*

Locality no.	Locality name	Year	Longitude	Latitude	Transect coordinate	<i>N</i>	HI eight loci
64	Hygum A	1998	526.0	6181.6	6199.6	2	0.906
28	Filskov	1992	499.2	6185.1	6199.8	2	0.781
19	Bredal	1998	538.0	6179.3	6199.9	5	0.688
77	Langelund F	1998	504.3	6184.8	6200.2	7	0.736
65	Hygum B	1998	527.7	6182.6	6200.8	17	0.857
146	Vindelev M	1985	530.3	6182.6	6201.1	8	0.729
147	Vindelev M	1987	530.3	6182.6	6201.1	1	0.857
148	Vindelev M	1991	530.3	6182.6	6201.1	2	0.688
149	Vindelev S	1991	530.1	6182.7	6201.2	16	0.740
150	Vindelev S	1992	530.1	6182.7	6201.2	17	0.658
111	Sandvad	1990	528.0	6183.5	6201.7	7	0.795
112	Sandvad	1991	528.0	6183.5	6201.7	1	0.625
84	Lindved	1990	536.1	6183.0	6202.4	4	0.828
103	Riis A	1998	520.1	6185.5	6202.8	4	0.688
36	Givskud Z	1987	522.0	6185.4	6202.9	6	0.865
37	Givskud Z	1991	522.0	6185.4	6202.9	15	0.746
38	Givskud Z	1998	522.0	6185.4	6202.9	26	0.801
35	Givskud E	1991	522.0	6186.0	6203.5	5	0.838
2	Aadel A	1991	528.7	6185.4	6203.7	1	0.438
3	Aadel A	1998	528.7	6185.4	6203.7	4	0.712
62	Hvejsel	1998	526.5	6185.7	6203.8	4	0.813
9	Baastrup	1987	540.5	6182.6	6204.0	20	0.769
34	Give Mark	1992	514.4	6187.5	6204.0	6	0.875
116	Skaerhoved	1998	517.2	6187.4	6204.3	1	0.750
1	Aadel	1990	529.0	6186.0	6204.4	3	0.813
122	Soendersthoved	1998	510.5	6188.3	6204.4	4	0.810
104	Riis	1998	519.9	6187.4	6204.6	4	0.813
144	Vibjerg	1998	517.7	6188.1	6205.1	1	0.875
46	Hedegaard	1998	510.8	6189.2	6205.3	1	0.875
4	Aalsted Moelle	1991	529.2	6187.2	6205.6	4	0.797
6	Aalsted Moelle	1998	529.2	6187.2	6205.6	9	0.762
105	Riis Mark	1998	521.3	6188.4	6205.8	13	0.777
85	Loesning	1984	544.5	6184.5	6208.1	15	0.782
86	Loesning	1985	544.5	6184.5	6208.1	14	0.853
127	Toerring Mark	1998	532.8	6192.2	6211.0	6	0.814
126	Thyregod	1998	520.4	6195.0	6212.2	2	0.625
53	Honum	1998	539.7	6193.5	6213.4	54	0.813
76	Klovborg	1998	532.6	6198.2	6216.9	1	0.857
27	Estrupholm	1998	518.4	6203.9	6220.8	2	0.938
139	Underup	1998	543.0	6200.6	6221.2	30	0.869
132	Traeden	1998	542.0	6202.7	6222.9	3	0.800
141	Vestbirk	1998	545.5	6202.5	6223.9	28	0.797
178	Gammel Hampen	2000	523.3	6207.2	6224.7	7	0.923
196	Vester Gludsted	2000	518.0	6209.3	6226.2	1	0.750
192	Store Noerlund	2000	514.9	6210.5	6227.0	5	0.958
49	Hoejby	1992	575.4	6197.4	6240.7	11	0.955
39	Gosmer F	1991	574.5	6199.1	6240.8	7	0.973
41	Gosmer N	1991	574.7	6199.8	6241.3	21	0.941
42	Gosmer N	1992	574.7	6199.8	6241.3	15	0.896
43	Gosmer N	1998	574.7	6199.8	6241.3	16	0.871
180	Gosmer N	2000	574.7	6199.8	6241.3	7	0.957
40	Gosmer J	1991	575.7	6198.8	6241.6	5	0.963
55	Hov	1985	576.6	6197.9	6241.9	6	0.990
56	Hov	1987	576.6	6197.9	6241.9	18	0.959

APPENDIX 1 *Continued*

Locality no.	Locality name	Year	Longitude	Latitude	Transect coordinate	<i>N</i>	HI eight loci
57	Hov	1991	576.6	6197.9	6241.9	2	1.000
58	Hov	1992	576.6	6197.9	6241.9	5	0.833
123	Spoettrup	1998	577.5	6199.2	6243.4	14	0.931
191	Spoettrup	2000	577.5	6199.2	6243.4	9	0.958
177	Elleskovhuse	2000	552.3	6223.2	6245.7	28	0.949
75	Klank	1985	556.8	6223.2	6247.4	10	0.925
194	Tovstrup	2000	550.5	6226.0	6247.7	1	1.000
197	Voel Oestermark	2000	545.5	6228.3	6248.7	2	1.000
173	Abo	2000	564.5	6220.5	6248.8	10	0.965
185	Ormslev	2000	566.7	6220.5	6250.1	7	0.974
176	Borum Oestergaard	2000	563.9	6227.3	6254.3	4	0.957

