Measures of linkage disequilibrium among neighbouring SNPs indicate asymmetries across the house mouse hybrid zone

LIUYANG WANG,* KEN LUZYNSKI,* JOHN E. POOL,† VÁCLAV JANOUŠEK,‡ PETRA DUFKOVÁ,§ MARTINA M. VYSKOČILOVÁ,¶ KATHERINE C. TEETER,*‡‡ MICHAEL W. NACHMAN,** PAVEL MUNCLINGER,‡ MILOŠ MACHOLÁN,¶ JAROSLAV PIÁLEK§ and PRISCILLA K. TUCKER*

*Department of Ecology and Evolutionary Biology and Museum of Zoology, University of Michigan, Ann Arbor MI 48109 USA, †Department of Evolution and Ecology, University of California Davis, Davis, CA 95616, USA, ‡Department of Zoology, Faculty of Science, Charles University in Prague, Prague, Czech Republic, §Department of Population Biology, Institute of Vertebrate Biology, Academy of Sciences CR, Studenec, Czech Republic, ¶Laboratory of Mammalian Evolutionary Genetics, Institute of Animal Physiology and Genetics, ASCR, Brno, Czech Republic, **Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, 85721 AZ, USA

Abstract

Theory predicts that naturally occurring hybrid zones between genetically distinct taxa can move over space and time as a result of selection and/or demographic processes, with certain types of hybrid zones being more or less likely to move. Determining whether a hybrid zone is stationary or moving has important implications for understanding evolutionary processes affecting interactions in hybrid populations. However, direct observations of hybrid zone movement are difficult to make unless the zone is moving rapidly. Here, evidence for movement in the house mouse *Mus musculus domesticus* × *Mus musculus musculus* hybrid zone is provided using measures of LD and haplotype structure among neighbouring SNP markers from across the genome. Local populations of mice across two transects in Germany and the Czech Republic were sampled, and a total of 1301 mice were genotyped at 1401 markers from the nuclear genome. Empirical measures of LD provide evidence for extinction and (re)colonization in single populations and, together with simulations, suggest hybrid zone movement because of either geography-dependent asymmetrical dispersal or selection favouring one subspecies over the other.

Introduction

Hybrid zones, narrow regions where genetically distinct populations overlap and reproduce (Barton & Hewitt 1985, 1989), provide an opportunity to explore the evolutionary processes that isolate taxa. Barton & Hewitt (1985, 1989) argued that most hybrid zones are 'tension zones' (Key 1968), i.e. clines maintained by a balance between dispersal and selection against hybrids. Tension zones can move according to differences in population density, rate of dispersal and/or direction of dispersal and are more likely to move as a result of changes in population structure than by asymmetric selection (Barton 1979; Barton & Hewitt 1985). Tension zones can also be stationary owing to a reduction in dispersal or a decrease in population density in the zone resulting from environmental factors or lower fitness of hybrids (Hewitt 1975; Barton 1979; Barton & Hewitt 1985). Barton (1979) predicted that tension zones are more likely to remain where they first formed than to move.

Correspondence: Priscilla K. Tucker, Fax: 734-763-4080;
E-mail: ptuck@umich.edu
‡‡Present address: Department of Biology, Northern Michigan University, Marquette, MI, 49855 USA.

Determining whether a hybrid zone is stationary or moving has important implications for understanding evolutionary processes affecting interactions in hybrid populations. However, the extent to which tension zones move is not well established, given the difficulties detecting hybrid zone movement (Buggs 2007). The most reliable method of detecting hybrid zone movement is direct observation over time, but this requires long-term studies and/or fast-moving hybrid zones. Hybrid zone movement can also be inferred based on historical records and/or patterns of introgression, i.e. the spread of an allele from one taxon into the gene pool of the other, using molecular markers. For example, a moving zone will leave in its wake a tail of clines at unlinked neutral loci (Arntzen & Wallis 1991; Buggs 2007). However, this pattern could also be interpreted as evidence for adaptive introgression of alleles at various loci away from a stationary hybrid zone (reviewed in Buggs 2007). Thus, inferring hybrid zone movement based on patterns of gene flow, alone, can lead to uncertain conclusions about the nature and direction of introgression from one taxon to another.

Availability of genomic data from naturally occurring organisms provides an opportunity to use the measures of linkage disequilibrium (LD) among 'neighbouring' alleles on a genome-wide scale to assess population history such as changes in population size and patterns of gene exchange (Slatkin 2008). For example, LD is created when two genetically distinct populations first hybridize and it will decay over time as a result of recombination (Reich *et al.* 2001). Assuming low mutation rates, a comparison of genome-wide measures of LD between adjacent diallelic SNPs across hybrid populations could provide evidence for the relative age of the hybrid populations sampled. If genome-wide LD increases asymmetrically across a hybrid zone, one can infer hybrid zone movement in the direction of the higher estimates. This information can then provide a context for understanding the evolutionary significance of introgression, e.g. whether it is adaptive.

House mice, *Mus musculus musculus* and *Mus musculus domesticus* (also referred to, in the literature, as *Mus musculus* and *Mus domesticus*), hybridize in a narrow zone of secondary contact that extends from Denmark to Bulgaria. Hybrid mice have also been recently discovered in Norway (Jones *et al.* 2010). The history of initial contact is unknown. However, because house mice are commensal in nature, their population dynamics, including range expansions, are likely to be influenced by human migration patterns. For example, the establishment of *M. m. domesticus* in western Europe is attributed to the 1st millennium BC coincident with an increased maritime exchange between the eastern and western Mediterranean regions (Cucchi *et al.* 2005). The

first reliably dated occurrence of *M. m. musculus* in southeastern Europe is attributed to the 4th millennium BC (Cucchi *et al.* in press). The hybrid zone is likely a tension zone with selection against hybrids because of genetic incompatibilities including hybrid male sterility (Forejt 1996; Storchová *et al.* 2004; Britton-Davidian *et al.* 2005; Vyskočilová *et al.* 2005, 2009; Good *et al.* 2008a,b, 2010; Mihola *et al.* 2009), hybrid female sterility (Britton-Davidian *et al.* 2005) and increased parasite loads of hybrids (Sage *et al.* 1986a; Moulia *et al.* 1993), although a recent study has shown the opposite pattern (Baird et al., submitted). The best studied incompatibility is hybrid male sterility, where complex epistatic interactions are involved in spermatogenic failure (Forejt 1996; Good *et al.* 2008a). These include X-autosome interactions that result in asymmetric sterility in some reciprocal crosses (Britton-Davidian *et al.* 2005; Good *et al.* 2008a), typically when the maternal line is
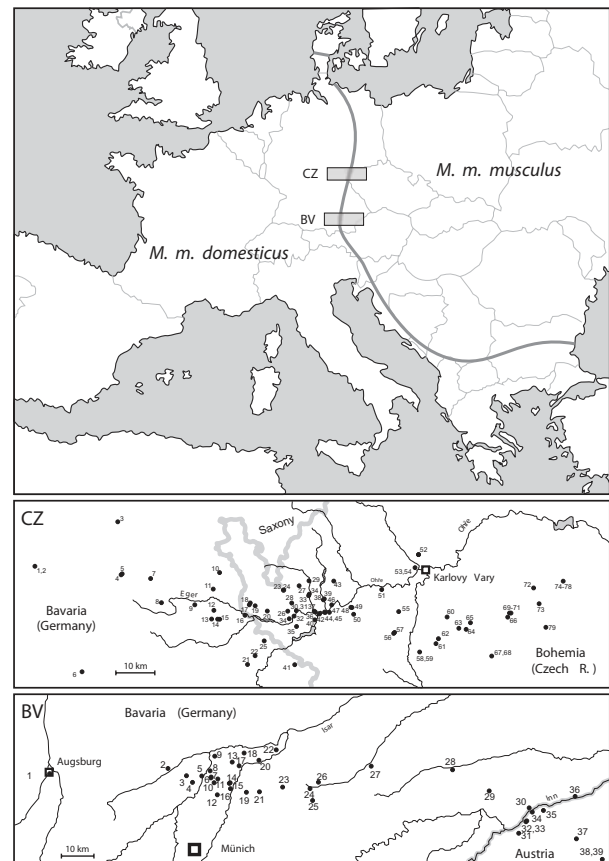

Fig. 1 The location of the house mouse hybrid zone in Europe and the collecting localities for the two transects studied. The black line demarcates the hybrid zone, and the shaded rectangles indicate the location of the two transects (top map). Collecting localities for the Czech (CZ) transect and the southern Bavaria transect (BV) are denoted by numbers that are cross-referenced in Table S1 (Supporting information).

*M. m. musculu*s or when the *M. m. musculus* X is introgressed on to the background of *M. m. domesticus* (Storchová *et al.* 2004; Good *et al.* 2008b). Several studies have argued that there has been reinforcement for mate choice preferences (Smadja & Ganem 2005; Vošlajerová Bímová *et al.*, 2011).

Multiple studies of several transects across the hybrid zone have been conducted (Hunt & Selander 1973; Sage *et al.* 1986b; Vanlerberghe *et al.* 1986, 1988a,b; Tucker *et al.* 1992; Dod *et al.* 1993, 2005; Fel-Clair *et al.* 1996; Orth *et al.* 1996; Boissinot & Boursot 1997; Prager *et al.* 1997; Payseur *et al.* 2004; Božíková *et al.* 2005; Raufaste *et al.* 2005; Macholán *et al.* 2007, 2008, 2011; Teeter *et al.* 2008, 2010), and comparisons have been made between transects when identical markers or markers from the same chromosomes have been used (Božíková *et al.* 2005; Teeter *et al.* 2010; Dufková *et al.* 2011). One study (Teeter *et al.* 2010) compared two transects, one spanning a portion of the state of Bavaria in Germany and Austria (the BV transect, Fig. 1) and a second spanning portions of the states of Thuringia, Saxony-Anhalt and Saxony in Germany, and found different patterns of introgression for multiple autosomal markers. The authors concluded that the genetic basis of isolation may be dependent on local genetic or environmental effects. Another study (Božíková *et al.* 2005) compared the BV transect with a transect located 200 km to the north spanning a portion of Bavaria, Germany and the Czech Republic (the CZ transect, Fig. 1). Asymmetric introgression of mitochondrial DNA from *M. m. musculus* to *M. m. domesticus* was noted in the CZ transect, but introgression of mitochondrial DNA from *M. m. domesticus* to *M. m. musculus* was found in the BV transect. A comparison of the nonrecombining portion of the Y chromosome between these same two transects (Tucker *et al.* 1992; Macholán *et al.* 2008) suggests asymmetric introgression from *M. m. musculus* to *M. m. domesticus* in the CZ transect and no introgression in the BV transect. These differences may also reflect local effects.

Here, we utilize genetic (SNP) resources available for house mice to investigate the genome-wide patterns of LD in naturally occurring hybrid populations from the BV and CZ transects to determine whether there is evidence for hybrid zone movement and, if so, whether observed differences in the patterns of introgression between transects can be explained by dispersal history. We greatly expand on all previous studies of this hybrid zone (Tucker *et al.* 1992; Payseur *et al.* 2004; Božíková *et al.* 2005; Macholán *et al.* 2007, 2008, 2011; Teeter *et al.* 2008, 2010; Dufková *et al.* 2011) by interrogating 1401 X-linked and autosomal SNPs, spaced approximately 2 Mb apart across the genome. This high-resolution genomic perspective allows us to study the detailed patterns of LD for the first time.

Material and methods

Samples

House mice were collected from the CZ and BV transects (Fig. 1, Table S1, Supporting information). The BV transect, located 200 km south of the CZ transect, is 176 km long and extends from south-central Bavaria, Germany, to western Austria (Tucker *et al.* 1992; Payseur *et al.* 2004; Božíková *et al.* 2005). A total of 437 mice were collected by Richard Sage in a roughly linear east–west manner from 41 localities in 1984, 1985 and 1992. The CZ transect is 130 km long × 50 km wide and extends from northeastern Bavaria, Germany, to western Bohemia, Czech Republic (Božíková *et al.* 2005; Macholán *et al.* 2007, 2008). A total of 895 mice were collected from 79 localities between 1991 and 2003, with >95.5% collected since 1997. To calculate the distances in CZ, Cartesian coordinates were projected onto a one-dimensional axis parallel to the most likely direction of change in allele frequencies and ordered with increasing distance from the western-most locality (Dufková *et al.* 2011). Sampling in both transects involved setting traps in suitable areas or where mouse activity was evident, typically within dwellings. While sampling methods may have differed slightly between transects, they were consistent within a transect and are thus not likely to impact the observed trends within transects.

SNP development

A total of 1536 SNPs (Fig. 2) from autosomal and X-linked chromosomes with fixed differences between *Mus musculus musculus* and *Mus musculus domesticus* were selected from a beta version of the Mouse Diversity Array (Yang *et al.* 2009). The diagnostic SNPs were determined by genotyping seven *M. m. musculus* and nine *M. m. domesticus* mice collected from naturally occurring European populations outside the hybrid zone (Table S2, Supporting information; Yang *et al.* 2011). We aimed for a density of approximately 1 SNP every 2 Mb across the genome. However, gaps in coverage exist as a result of the absence of SNPs in certain regions of the genome (e.g. 10.488-Mb gap in the proximal region of the X chromosome) on the Mouse Diversity Array and an inadvertent loss of data during the SNP selection process (e.g. a 23.261-Mb gap in the distal portion of chromosome 12). The mean and median gap sizes between two SNPs were 1.6 and 1.8 Mb, respectively, on the final 1536 SNP array.
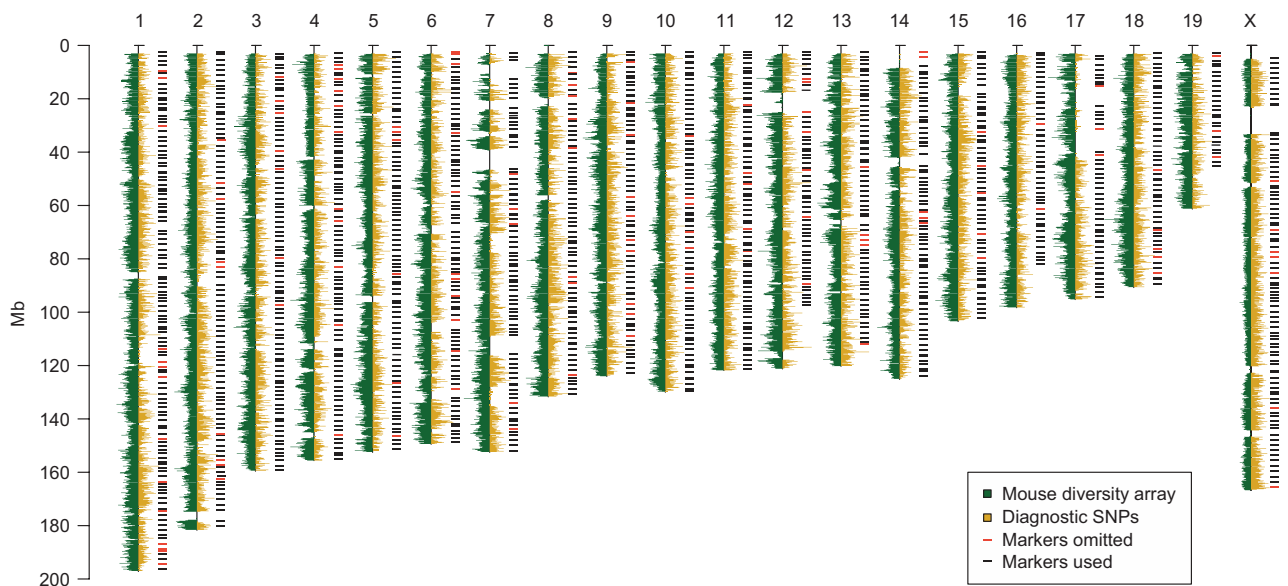
Fig. 2 Genomic location of genetic markers used in this study. Markers diagnostic for *Mus musculus musculus* and *Mus musculus domesticus* were identified from the Mouse Diversity Array (Yang *et al.* 2009), and 1536 SNPs (black and red lines) from autosomes and the X chromosome were selected. A total of 135 SNPs (red lines) failed resulting in data recovery from 1316 autosomal and 85 X-linked SNP markers.

Genotyping

DNA was isolated from mouse tissues using proteinase K digestion and subsequent extraction using either a phenol/chloroform protocol or the Qiagen® DNeasy® kit. Double-stranded DNA was quantitated with the PicoGreen® Assay, and samples were diluted to concentrations of 50–100 ng/μL. Mice were genotyped at autosomal and X-linked SNPs using the Illumina® Goldengate® Assay on an Illumina® Beadstation 500 at the University of Michigan Genotyping Core. Genotyping results were analysed using Illumina's Genome Studio Genotyping Module software. Individual samples with less than a 96% call rate success for SNP markers were eliminated, and individual SNPs with Gencall scores <0.15 were also removed. The standard cluster files for all SNPs were evaluated by eye, and SNP cluster positions were manually edited in some cases.

Data analysis

Population delimitation. Samples collected from populations in close geographic proximity were subjected to a permutation-based statistical test to determine whether they were genetically differentiated following Hudson *et al.* (1992) and using the program DnaSP v. 5.10 (Librado & Rozas 2009). Populations were pooled for subsequent analyses if they were not genetically differentiated ($P = 0.01$). These included three pairs of subpopulations in BV and four pairs of subpopulations

in CZ (Table S3, Supporting information). Only two of the pooled populations, Augs and Simb from the BV transect, had sample sizes, from a single collecting year, large enough ($n \geq 10$) to be used in LD analyses.

Hybrid index. The ancestry of individual mice from each transect was summarized using a hybrid index (0–1), which is the fraction of autosomal alleles at all successfully genotyped SNP loci inherited from *M. m. musculus* using the R package INTROGRESS (Buerkle 2005; Gompert & Buerkle 2010).

Geographic cline analyses. The software ClineFit (Porter *et al.* 1997) was used to estimate geographic clines for each SNP marker in each transect. A two-parameter model that uses cline centre and width to describe cline shape along the length of each transect was used instead of the more complex six-parameter model (Barton & Hewitt 1985; Barton & Bengtsson 1986; Barton & Gale 1993) because the likelihood surface for the more complex models can be uninformative and optima can be difficult to find (Raufaste *et al.* 2005; Macholán *et al.* 2007). The western-most localities in both transects were set at zero distance.

Linkage disequilibrium (LD) analyses. Haplotype reconstruction and missing data imputation for each individual were performed using the program PHASE v. 2.1 (Stephens *et al.* 2001; Stephens & Scheet 2005). Haplotype estimation using this program is not sensitive to

departures from HWE (Stephens *et al.* 2001; Stephens and Donnelly, 2003; Marchini *et al.*, 2006). The default setting, which incorporates the general model with varying recombination rate (Li & Stephens 2003), and the –X10 option, which increases the final runs 10 times longer than the main iteration, were used. Predicted phases with best pair probabilities >0.80 were retained for each individual.

The squared correlation of alleles at two sites, r^2, was used to quantify linkage disequilibrium between adjacent markers separated by <2 Mb. It is less sensitive to allele frequency differences between loci and shows less inflation than D' (Weir 1996) in small samples (Ardlie *et al.* 2002). Gametic r^2, the squared correlation of allelic indicators at two loci in a haploid gamete (Weir *et al.* 2004), was calculated from autosomal (males and females) and X-linked (females only) loci. Because LD can be affected by sample size and allele frequency, r^2 was calculated from each locality using the same number ($n = 10$) of randomly chosen individuals collected in the same year and by removing alleles from each locality with a minor allele frequency ≤0.05. SNPs with a proportion of missing data >0.5 within each locality were also removed. In addition, localities with $n \geq 10$ but with <40 possible pairwise SNP comparisons were removed from the analysis. These included six localities, all from the edges of the hybrid zone in CZ. The resulting data set included eight localities in BV and 14 in CZ. For some localities, a sufficient number of individuals ($n \geq 10$) were collected in multiple years, making it possible to generate r^2 estimates for different years from the same locality. In addition, for a subset of localities, a sufficient number of individuals were collected in a single year to estimate r^2 from a sample of $n = 15, 19$ and/or 20 mice.

Estimates of r^2 were calculated using the equation $r^2 = \frac{D_{AB}^2}{P_A(1-P_A)P_B(1-P_B)}$, where D is the coefficient of linkage disequilibrium for a pair of alleles (A and B) at two loci and P_A and P_B are the frequencies, respectively, of allele A at the first locus and allele B at the second locus (Hill & Robertson 1968). Mean r^2 was calculated from 100 replicates of a randomly selected sample of 10 individuals per locality. Standard errors of mean r^2 were calculated using $\sqrt{\mathrm{var}(r^2)/n}$. The physical distance (Mb) at which r^2 drops below 0.5, the 'half-length of LD' (Reich *et al.* 2001), was estimated to compare haplotype block size between populations along each transect.

The age, in generations, of each population was calculated based on the decline of r^2 using equation $\frac{r_t^2}{r_0^2} = (1-c)^t$, where the linkage disequilibrium at time t (r_t^2 or current LD) is compared with initial time 0 (r_0^2) given a certain recombination rate c (Hedrick 2005). Assuming the target pairwise SNPs are in complete linkage disequilibrium at original time, i.e. $r_0^2 = 1$, and

the rate of recombination is known, the age for two SNPs at r_t^2 can be estimated. A genome-wide average recombination rate of 0.63 cM/Mb for mice (Shifman *et al.* 2006) was used. Analyses were conducted in the R GENETICS package (Warnes *et al.* 2011).

Simulation analyses. We conducted computer simulations to compare our data with various models that incorporate hybrid zone movement. Forward simulations were conducted using a modification of the program developed by Pool & Nielsen (2009). Each generation, each individual (or chromosome, in this haploid framework) randomly chooses two parents. The number of recombination events along a 100 cM chromosome is determined by a simple exponential process, and recombination breakpoints are distributed randomly. One hundred evenly spaced ancestry-informative markers were simulated on each chromosome, and genome-scale data sets were assembled by combining data from 10 independent runs. Ten stepping-stone populations were simulated, each containing 1000 individuals. [Simulations conducted using a population of 100 individuals produced qualitatively similar results.] The five 'left-most' populations were initially composed entirely of subspecies *A*, and the five 'right-most' populations were initially composed entirely of subspecies *B*.

Under the null model of symmetric migration with equal dispersal rates, the migration rate between any adjacent pair of populations is simply defined by *m*. Fitness of the two subspecies are equal to 1, but individuals with a hybrid index of 0.5 (e.g. F1 hybrids) have a fitness $h = 0.5$. In general, the hybrid component of fitness is given by $h + (\mathrm{abs}(i-0.5)/0.5) \times (1-h)$, where i indicates an individual's hybrid index (proportion of subspecies *B* ancestry), such that hybrid fitness is a linear function of hybrid index between 0.5 and either 0 or 1. The value of h used here is close to the fitness of hybrid mice estimated by Raufaste *et al.* (2005) from the centre of a Danish hybrid zone transect using cline theory. This null model was varied in three ways to reflect potential causes of a moving hybrid zone (in each case, advancing from right to left):

1 A geography-dependent asymmetric dispersal model where 'rightward' migration rates are a constant proportion (u) lower than 'leftward' migration rates, regardless of the subspecies composition of each population. Thus, rightward migration rates were determined by um, where $0 \leq u \leq 1$.

2 An ancestry-dependent asymmetric dispersal model in which difference in dispersal rates between species led subspecies *A* to have a migration rate a

proportion d of the base migration rate m defined by subspecies B (where $0 \leq d \leq 1$). The migration probability of a hybrid individual was governed by $m(d + i(1-d))$, where i represents the hybrid index.

3 A selection model in which subspecies A has a lower fitness than subspecies B. Defining the fitness of subspecies B as 1, the fitness of subspecies A is denoted by f. The subspecies component of fitness for a hybrid individual is a linear function of hybrid index, given by $f + i(1-f)$. Subspecies and hybrid components of fitness are combined multiplicatively, such that an individual's fitness equals $(f + i(1-f)) \times (h + (\mathrm{abs}(i-0.5)/0.5) \times (1-h))$.

Simulations were run for 500 generations, with summary data recorded at intervals of 10 generations. Hybrid index and linkage disequilibrium (mean r^2) were analysed at each interval, with the goal of testing whether patterns observed in the empirical data could reflect a shifting hybrid zone under these conditions. Summaries were calculated based on the samples of 20 chromosomes from each population. LD was calculated between adjacent SNPs that each had a minor allele frequency ≥ 0.05, and LD was reported only for populations with at least 25 SNP pairs contributing to mean r^2.

Two base migration probabilities were examined ($m = 0.002$ or 0.01) in conjunction with alterations to at most one of the following parameters. For u or d, the default value of 1 was modified to 0.95, 0.75, 0.5 or 0.25. For f, the default value of 1 was modified to 0.99, 0.95, 0.9 or 0.8.

Results

Data matrices

Of the original 1536 SNP markers, 135 did not yield reliable genotyping calls resulting in a genotype matrix consisting of 1316 autosomal SNPs and 85 X-linked SNPs. The mean and median gap sizes between two SNPs for these markers are approximately 1.86 Mb (Fig. 2). Of the 895 mice genotyped from the Czech (CZ) transect, SNP data were recovered for 869 samples, and of the 437 mice genotyped from the BV transect, SNP data were recovered for 432 samples. Individual genotyping calls for some samples were eliminated resulting in a data matrix consisting of 1 817 678 genotyped markers or 0.41% missing data (Fig. S1A, B, Supporting information). The final phased data matrix consisted of 3 533 871 SNP alleles from autosomal and X-linked loci. The former matrix was used in the geographic cline analyses, and the latter matrix, in r^2 analyses.

Hybrid index

Hybrid indices plotted against inter-subspecific heterozygosity (Fig. 3a, b, Table S1, Supporting information)
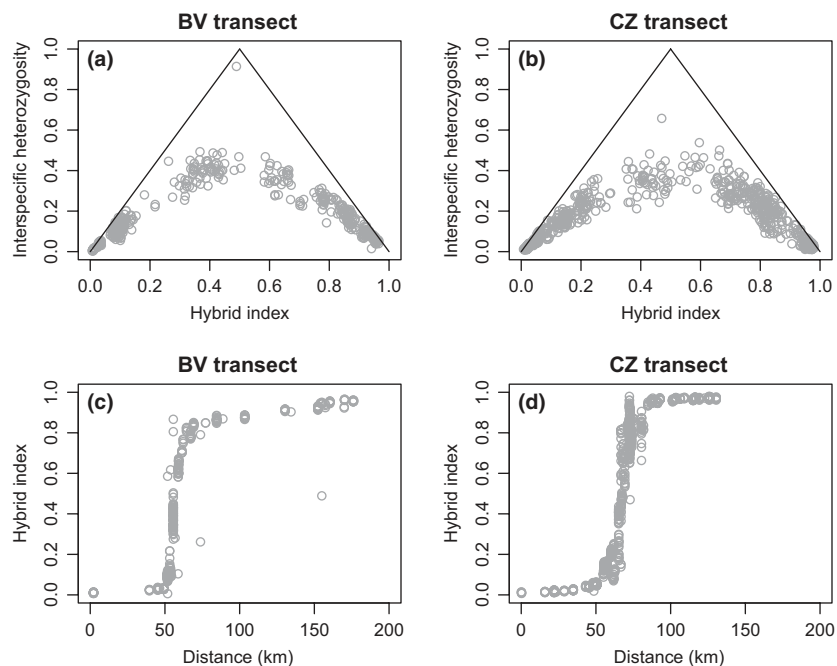

Fig. 3 Hybrid indices plotted against inter-subspecific heterozygosity (a and b) and distance from the western-most locality (c and d) in kilometres for individual mice.

indicate that sampling from the BV and CZ transects yielded similar distributions of admixture. Both transects show a nearly continuous distribution of hybrid genotypes and provide evidence that most hybrids are the result of multiple generations of intercrossing and/or backcrossing. One exception is a mouse collected from the eastern end of the Bavaria transect. It has a hybrid index score near 0.5 and is almost completely heterozygous. This hybrid individual is likely to be only a few generations removed from a long-distance dispersal event of a *Mus musculus domesticus* mouse to the eastern edge of the hybrid zone. Hybrid index plotted against distance from the western-most locality (Fig. 3c, d) demonstrates that the proportion of *Mus musculus musculus* alleles increases with geographic distance from the western-most locality with a characteristic abrupt change in both transects. However, the proportion of *M. m. musculus* alleles increases more gradually east of this abrupt transition in BV vs. CZ.

Geographic cline analysis

Estimates of cline width and centre were calculated for all SNP markers. A small number of SNPs (37 for BV and 11 for CZ) with estimated cline widths greater than the actual length of the transects are likely to be unreliable estimates and were excluded from the calculations of mean/median cline centre and width (Table 1, Table S5, Supporting information). In BV, the mean/median cline width and centre for X-linked markers are 10.23/5.87 km and 57.26/56.22 km, respectively, and the mean/median cline width and centre for autosomal markers are 45.99/24.03 km and 66.21/61.05 km, respectively. In CZ, the mean/median cline width and centre for X-linked markers are 14.7/11.79 km and 68.03/68.36, respectively, and the mean/median cline width and centre for autosomal markers are 26.45/21.03 and 67.29/66.95, respectively. These estimates provide evidence for steep clines in accord with a tension zone

model with selection against hybrids owing to genetic incompatibilities (Fig. S2, Supporting information). The simple two-parameter model returns larger estimates of cline width than the six-parameter model (Macholán *et al.* 2007; Teeter *et al.* 2010) and does not, in and of itself, capture asymmetry in cline shape. However, when cline width is plotted against cline centre for each SNP marker (Fig. 4), asymmetry is revealed. In BV, cline centre shifts to the east as cline width increases for many markers. In CZ, a different pattern is found. Cline centre shifts in both directions as cline width increases. The two-parameter mean cline centres for the X-linked SNPs from BV and CZ are very similar to the mean cline centre of 55.10 km and 68.26, respectively, as calculated for 13 X-linked markers for the same transects analysed with the six-parameter model (Payseur *et al.* 2004; Dufková *et al.* 2011).

When cline centres of individual markers are compared with median cline centre for all SNPs in each transect (Fig. 5), several interesting patterns emerge. First, a large number of markers (approximately 50%) show minimal deviations in cline centre. Because cline centre and cline width are correlated, these SNPs are in genome regions exhibiting narrow clines, on average, and therefore, they may be in, or closely linked to, regions associated with reproductive isolation. Second, many markers (>300) show a marked deviation in cline centre to the east in BV but little or no deviation from median cline centre in CZ. If the hybrid zone is moving from east to west in BV, we might predict these markers to be neutral. Third, a few markers show deviations in cline centre in the same direction, either from east to west or from west to east, in both transects. These markers may be in genome regions with genes that confer strong fitness advantages and, as such, escape barriers to gene flow, regardless of the direction of hybrid zone movement. Finally, excluding the large number of markers showing a marked deviation in cline centre to the east in BV, there are a few markers with deviations

Table 1 Mean and median cline centre and width for SNP markers for the Bavaria (BV) and Czech (CZ) transects

	BV			CZ		
	No. of markers†	Cline centre mean/median km	Cline width mean/median in km	No. of markers†	Cline centre mean/median km	Cline width mean/median km
Autosomal SNPs	1281	66.21/61.05	45.99/24.03	1306	67.29/66.95	26.45/21.03
X-linked SNPs	84	57.26/56.22	10.23/5.87	85	68.03/68.36	14.70/11.79
X-linked and autosomal SNPs, combined	1365	65.66/60.65	43.79/22.60	1391	67.34/67.01	25.73/20.38

†SNPs with estimated cline widths greater than the actual length of the transects are unreliable estimates and were excluded from calculations.
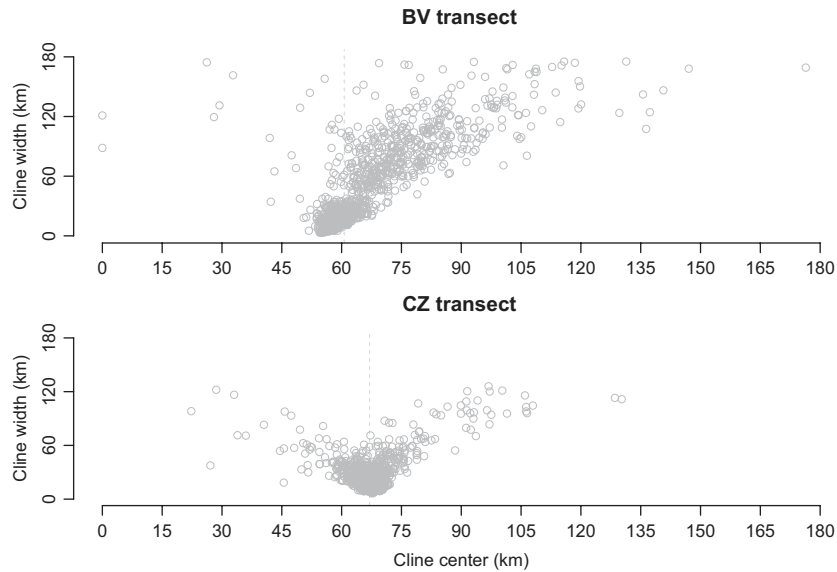
Fig. 4 Estimates of cline width plotted against the estimates of cline centre for all SNP markers for the BV and CZ transects. Dashed line indicates median cline centre calculated from all markers.
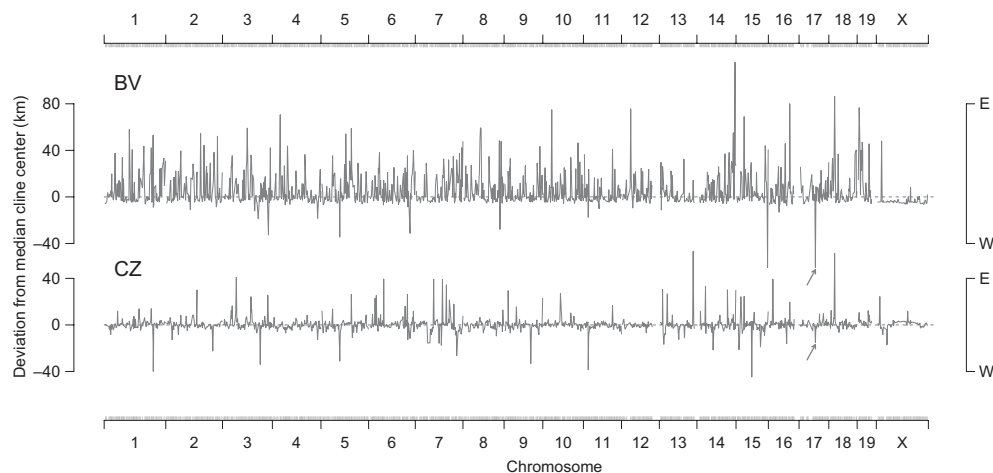

Fig. 5 Deviations from median cline centre (dashed line) for all SNP markers in BV and CZ. Some markers show similar patterns in both transects, e.g. a marker on Chr 17 (arrows) indicates the introgression of *Mus musculus musculus* alleles onto the background of *Mus musculus domesticus*.

in cline centre that are transect specific. These markers may indicate regions with genes conferring fitness advantages on a local scale.

Linkage disequilibrium analyses

The decay of LD with physical distance for SNP markers was measured in a single population (Neufahrn) located in the central portion of the BV transect and is summarized in Fig. S3 (Supporting information). LD, measured separately for samples collected in 1984 and 1985, showed similar patterns of decay for similar sam-

ple sizes (n = 10, 15 or 25) between years and is slightly lower when sample size is increased by five or 10 individuals. Because mean and median gap size among SNPs for all chromosomes are approximately 1.8 Mb, we compared the r^2 values of all pairwise SNPs with a genomic distance ≤2 Mb, as this measure includes most of the adjacent SNPs. For populations sampled across the BV transect, mean r^2 ranged from 0.2679 to 0.7093, with values generally increasing from east to west (Table 2). For populations sampled across the CZ transect, mean r^2 ranged from 0.3471 to 0.8217, with values generally increasing from west to east. Similar results

Table 2 Values of r^2, D', half-length of LD (HLD) or block size in Mb, estimated age in generations, mean hybrid index and mean heterozygosity (He) based on samples ($n = 10$) collected at specific localities in specific years. Locality code refers to map location in Fig. 1

Pop.	Locality code	Collection date	Distance (km)†	SNP pairs‡	r^2	SE of r^2	D'	SE of D'	HLD (Mb)	Age§	Hybrid index (mean)	He (mean)
BV transect												
Appe	5	1992	49.20	3697	0.5979	0.0067	0.9239	0.0039	7.2704	41	0.0312	0.3199
GessZ	7	1992	51.80	20227	0.6401	0.0028	0.9647	0.0011	9.2058	35	0.0990	0.1310
GessW	6	1984	51.80	19762	0.6754	0.0027	0.9601	0.0011	13.7778	31	0.1001	0.1276
GessW	6	1985	51.80	20018	0.6136	0.0029	0.9432	0.0014	7.4431	39	0.0862	0.1174
GessW	6	1992	51.80	27064	0.7093	0.0022	0.9656	0.0009	24.7567	27	0.1057	0.1048
Neuf	12	1984	55.60	95382	0.5296	0.0011	0.8910	0.0008	3.4506	50	0.3881	0.3958
Neuf	12	1985	55.60	96334	0.5345	0.0011	0.8876	0.0008	3.6275	49	0.4050	0.3716
Rudl	20	1984	67.20	44161	0.2680	0.0014	0.8376	0.0014	0.5021	104	0.7833	0.2591
Sonn	24	1984	84.60	24320	0.2679	0.0019	0.8796	0.0017	0.6093	104	0.8655	0.1775
Brun	27	1992	103.40	18627	0.2908	0.0023	0.9070	0.0017	0.7532	97	0.8777	0.1502
Simb	30	1985	152.40	10511	0.3384	0.0036	0.8890	0.0025	1.0802	85	0.9225	0.0966
CZ transect												
NEU8	15	2006	50.27	5037	0.4749	0.0056	0.9205	0.0031	1.9613	59	0.0537	0.0641
NEU8	15	2007	50.27	4800	0.4776	0.0059	0.9453	0.0026	1.9778	58	0.0516	0.0589
HOHE1	16	2007	54.85	17520	0.4766	0.0029	0.9523	0.0013	1.9610	58	0.1033	0.1357
HUR1	19	2002	57.07	29218	0.4876	0.0022	0.9277	0.0013	1.9751	57	0.1452	0.1618
HUR1	19	2006	57.07	24848	0.5516	0.0026	0.9596	0.0011	3.7503	47	0.1453	0.1517
LUZN	20	1999	60.13	45676	0.5066	0.0018	0.9073	0.0011	2.0946	54	0.2077	0.2500
LUZN	20	2006	60.13	39333	0.5020	0.0020	0.9073	0.0012	2.0053	54	0.2062	0.2223
WOHL	22	2007	61.56	14222	0.4081	0.0034	0.9235	0.0018	1.3238	71	0.0818	0.1070
DLMO	26	2001	65.02	83631	0.5053	0.0012	0.9003	0.0008	2.0317	54	0.3960	0.3516
JIND	32	2000	67.12	74912	0.5508	0.0014	0.9227	0.0008	3.6702	47	0.4818	0.3363
MIL1	33	2000	67.94	95337	0.5933	0.0012	0.9240	0.0007	6.5661	41	0.6474	0.3643
MIL1	33	2001	67.94	89200	0.5770	0.0013	0.9237	0.0007	5.5899	43	0.6823	0.3663
NEB3	36	2007	71.30	44937	0.5729	0.0018	0.9374	0.0009	5.0091	44	0.8246	0.1564
KAC2	39	2005	72.19	53100	0.7565	0.0015	0.9776	0.0005	86.4630	22	0.8997	0.1547
OBIL1	40	2000	72.23	59820	0.5857	0.0016	0.9295	0.0009	5.1163	42	0.6946	0.2780
OBIL1	40	2007	72.23	65873	0.5196	0.0015	0.9079	0.0009	2.3240	52	0.7452	0.2940
MOST2	42	2002	72.64	47816	0.5864	0.0018	0.9469	0.0008	5.3077	42	0.7794	0.2651
MOST2	42	2006	72.64	36195	0.6784	0.0020	0.9740	0.0006	10.2532	31	0.8260	0.1848
CHL1	46	2007	73.68	46462	0.7683	0.0016	0.9738	0.0006	25.8102	21	0.7935	0.1976
RUD2	48	2003	80.29	42926	0.6221	0.0018	0.9536	0.0008	8.0806	37	0.8277	0.2213

†Distance is calculated from the western-most locality. ‡Total number of SNP comparisons at ≤2-Mb distance is based on 100 permutations of samples at each locality. SNPs with minor allele frequency <0.05 and missing values >50% were excluded. §The age, in generations, of each population was calculated based on the decline of r^2 using equation $\frac{r_t^2}{r_0^2} = (1 - c)^t$ as described in the Material and methods and is an estimate of age at the point in time when the population was sampled. BV, Bavaria; CZ, Czech.

are found for a subset of the same populations in BV and CZ when sample size is increased by increments of 5, i.e. $n = 10$, $n = 15$ and $n = 20$ (Table S6, Supporting information). When r^2 values are plotted against distance from mean cline centre, estimated using ClineFit, they are greatest in populations to the west of cline centre in the BV transect and to the east of cline centre in the CZ transect (Fig. 6). Similarly, when r^2 values from these same populations are plotted against mean hybrid index per locality, they increase as the proportion of *M. m. musculus* alleles decrease across the BV transect and as the proportion of *M. m. musculus* increase across the CZ transect (Fig. 7a, b). The distribution of populations from which we could measure values of r^2, i.e. a sufficient number of SNPs with a minor allele frequency ≥0.05 allowing for >40 pairwise comparisons, differed between transects. In CZ, these populations included a region 17 km to the west and 13 km to the east of median cline centre, whereas in BV, these populations included a region 12 km to the west but over 90 km to the east of median cline centre. This difference reflects the pronounced asymmetry in cline shape in BV in
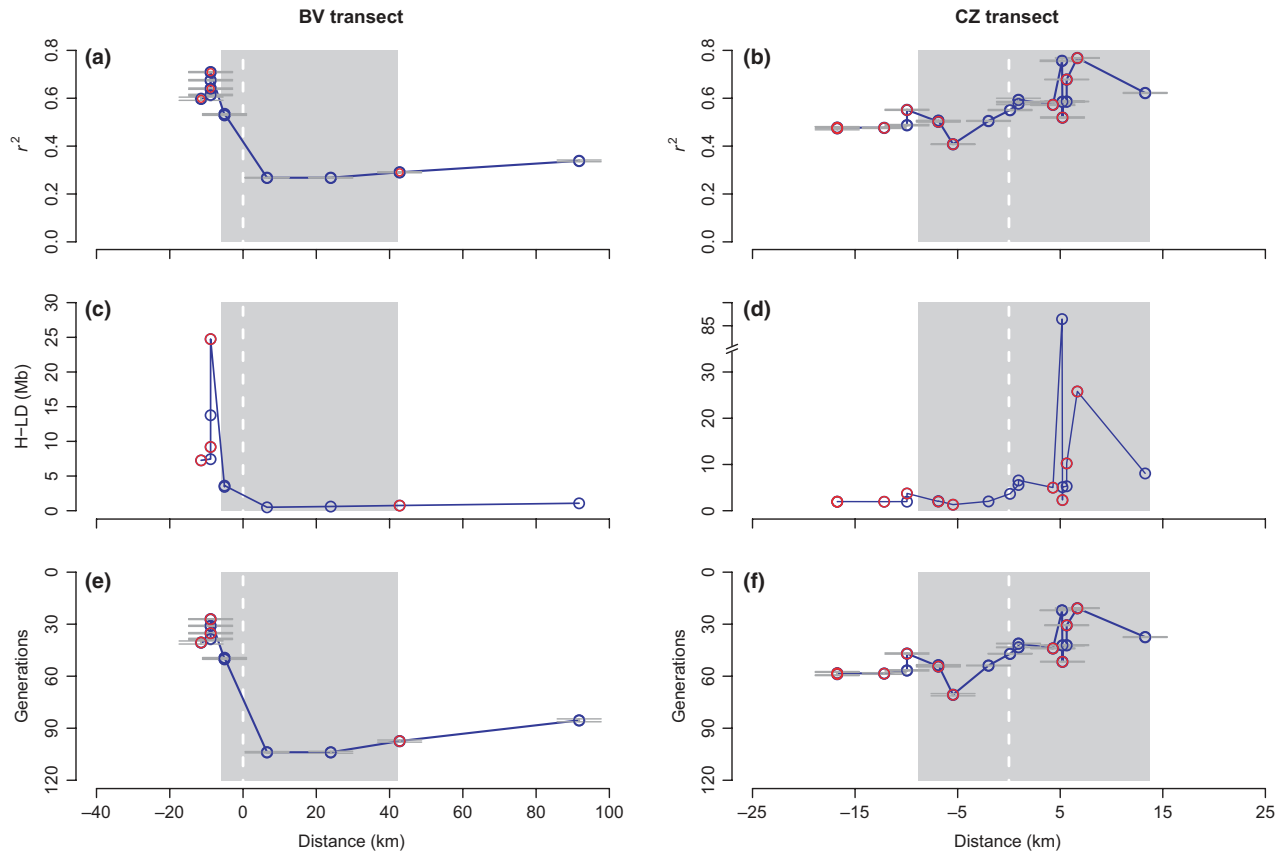
Fig. 6 Mean r^2 values of all pairwise SNPs with a genomic distance ≤2 Mb and half-length of LD (H-LD) in Mb is plotted for eight populations sampled in the Bavaria (BV) transect and 20 populations sampled in the CZ transect (a–b). Two populations in BV and six populations in CZ were sampled over multiple years. Red circles = populations sampled in 1992 in BV and 2007 in CZ. Blue circles = populations sampled earlier in time (see Table 2 for details). Standard errors for each population (bars) are also shown for each r^2 and population age estimate. Half-length of LD (H-LD) is plotted for the same populations (c–d). Population ages in generations for the same populations (e–f) are based on r^2 estimates. The dash line indicates median cline centre. The 95th percentile of cline centres for all SNP markers, except those markers with cline width larger than the transect, are shaded to show the marked asymmetry in the distribution of cline centres in BV.

comparison with CZ as demonstrated by the relationship between cline centre and width (Fig. 4) and the asymmetrical distribution of clines centres relative to median cline centre (Fig. 6).

Estimates of the physical distance (Mb) at which r^2 drops below 0.5 (H-LD) across the BV transect range from 0.5021 to 24.7567 Mb, with values increasing from east to west (Table 2; Fig. 6). Estimates across the CZ transect range from 1.2787 to 86.4630 Mb with the highest values to the east (Table 2; Fig. 6). The extreme value for KAC2, a population in the CZ transect, is likely due to the presence of several individuals heterozygous at many SNP loci in an otherwise predominantly *M. m. musculus* population. This locality is a storage site for agricultural products. The mixture of mice with different genotypes may have resulted from passive long-distance travel.

The population age (in generations) for BV, sampled in 1984, 1985 and 1992, varied from as young as 27 gen-

erations (GessW sampled in 1992) to as old as 104 generations (Rudl and Sonn, both sampled in 1984) (Table 2, Fig. 6). The population age for the CZ transect, sampled between 1999 and 2007, varied from as young as 21 generations (CHL1, sampled in 2007) to as old as 71 generations (WOHL sampled in 2007). Interestingly, population age estimates for mice collected from GessW in 1984 and 1985 are older than an estimate for mice collected from the same locality in 1992, suggesting this population experienced an episode(s) of extinction and (re)colonization. Similar patterns, more or less extreme, are observed for Neuf in the BV transect and NEU8, HUR1 and MOST2 in the CZ transect. In addition, estimates of population age are identical for one population in the CZ transect (LUZN) sampled in 1999 and 2006, again suggesting that these populations experienced extinction and (re)colonization events. The remaining populations (MIL1 and OBIL1) in the CZ transect for which multi-year data were collected
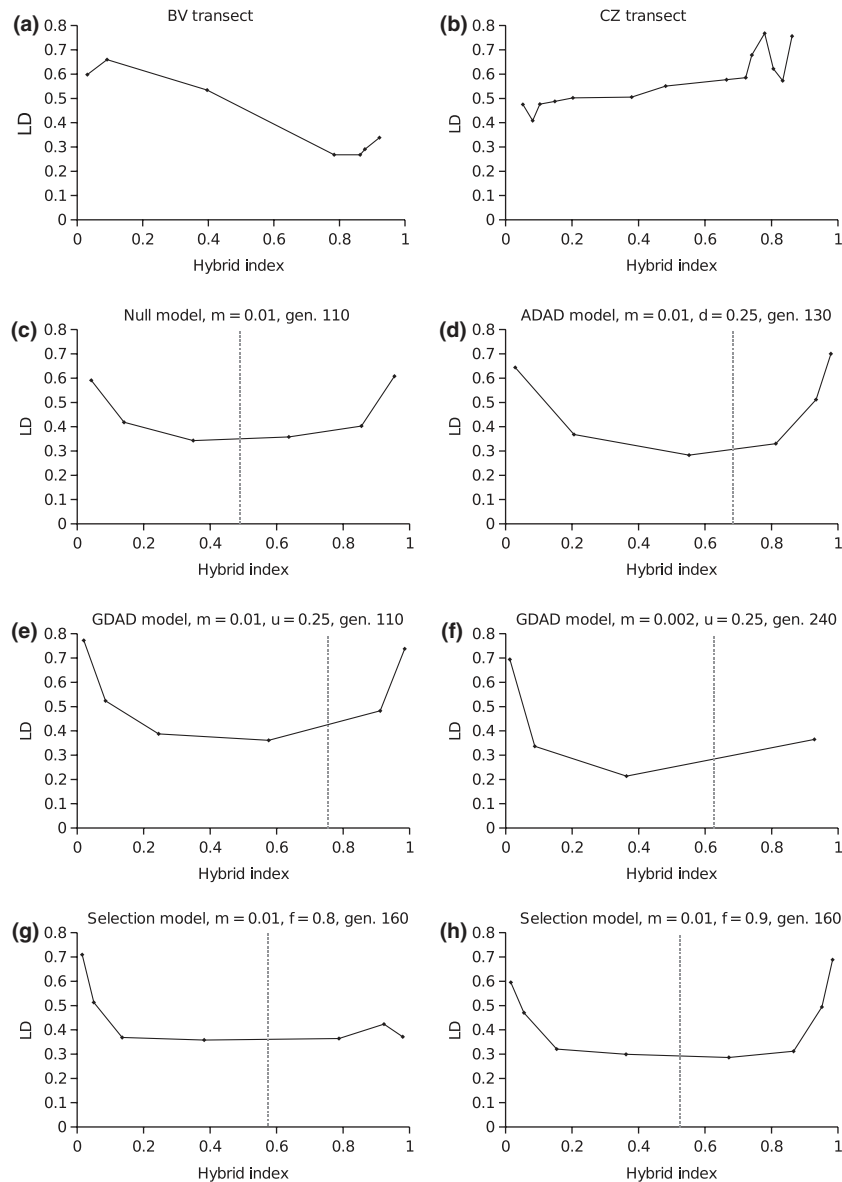
Fig. 7 The relationship between hybrid index (HI) and linkage disequilibrium (measured by mean r^2) in the empirical data (top panels, a–b, give the mean HI's and r^2 values for localities with identical distances including those sampled over multiple years, see Table 2) and in simulated data (panels c–h, see Material and methods). Models of symmetric gene flow (Null model, panel c) and biased dispersal based on subspecies ancestry (ancestry-dependent asymmetric dispersal model, ADAD, refer to panel d) failed to re-create the asymmetric pattern of HI vs. LD observed in the BV and CZ transects. Some models involving either biased dispersal based on geography alone (geography-dependent asymmetric dispersal model, GDAD, panels e–f) or fitness differences between subspecies (Selection model, panels g–h) were able to generate a qualitatively similar pattern. Among 10 simulated populations, only those with LD based on ≥25 SNP pairs were plotted here. The dotted lines indicate the populations where initial contact occurred.

show an increase in age with time as expected if the population remains relatively stable.

Simulation analyses

Forward simulations were conducted to identify processes that might generate the asymmetric pattern of linkage disequilibrium (LD) observed across the BV and CZ transects. The results in Fig. 7 illustrate the patterns observed among 26 parameter combinations (see Material and methods section). In some cases, models that involved a geographic effect or a subspecies fitness difference were capable of generating asymmetric patterns of LD across the hybrid zone. However, these results were sensitive to the particular parameter values and generations examined, and in some cases, hybrid zones

shifted without creating asymmetric LD. In addition, the simulated data did not recapitulate the empirical pattern of populations with intermediate hybrid index having higher LD than populations at one end of the transect. Nonetheless, these simulations clearly show that some models of hybrid zone movement can capture a key aspect of our data, namely that patterns of LD are distributed asymmetrically across transects of this hybrid zone.

Discussion

LD and hybrid zone movement—direction and timing

Comparing measures of LD among 'neighbouring' SNPs across hybrid populations of house mice to assess evidence for hybrid zone movement is based on the idea that LD is highest when two genetically distinct populations first hybridize and then decays over time as a result of recombination (Reich *et al.* 2001). Our measures of LD between adjacent diagnostic SNPs across two transects in the house mouse hybrid zone sampled at different points in time provide a snapshot documenting the history of the sampled populations. They suggest hybrid zone movement as manifested in several trends in the data. Values of r^2 and haplotype block size (half-length of LD) measured from populations sampled in 1984, 1985 and again in 1992 across the BV transect change in a unidirectional mode across geographic distance, from east to west and across hybrid index from *Mus musculus musculus* to *Mus musculus domesticus*, a pattern consistent with hybrid zone movement from east to west over time. The highest values of r^2 do not coincide with mean cline centre calculated for autosomes or for X-linked markers, as might be expected in a stationary zone (Barton & Hewitt 1985). Values of r^2 and haplotype block size measured from populations sampled more recently, from 1999 to 2007, across the CZ transect, change across geographic distance from west to east and across hybrid index from *M. m. domesticus* to *M. m. musculus*, although these trends are not as pronounced as in BV. The range of values of r^2 is lower in CZ than in BV and the lowest values of r^2 in CZ are higher than the lowest values in BV, suggesting less stability in CZ populations and possibly past movement in both directions. However, the highest values of r^2 are east of the cline centre suggesting more recent movement from west to east. Simulations indicate that under certain conditions, particularly when geography or selection halted the spread of hybrid chromosomes in one direction but not the other, a moving hybrid zone may produce aspects of the LD patterns observed in the BV and CZ transects. Unique aspects of the empirical data, however, suggest that fur-

ther study using different models is warranted to more fully understand the evolutionary forces behind potential hybrid zone movement.

Age estimates provide some information on the relative stability of mouse populations. Some populations sampled later in time are, in fact, estimated to be younger than populations sampled earlier in time. Consistent with findings from other studies of population size (e.g. Dallas *et al.* 1998; Hauffe *et al.* 2000), these data suggest that house mouse populations are likely to consist of small demes that are subject to extinction–(re)colonization events. The older age estimates for populations on the eastern side of the BV transect suggest that these populations are more stable than populations on the western side of the BV transect. Raufaste *et al.* (2005) combined cline theory and migration rates to calculate initial contact between *M. m. musculus* and *M. m. domesticus* in central Jutland between 700 and 1800 years ago, an estimate consistent with archaeological data (Cucchi *et al.* 2005). Estimates of age across the BV and CZ transects are much younger, suggesting the possibility that age of initial contact is not easily captured from the estimates of LD when populations are unstable. Nevertheless, the estimates of LD using SNPs diagnostic for *M. m. musculus* and *M. m. domesticus* provide information on relative recency of mixing of the two genomes. The clear pattern of increase in the values of r^2 across geographic distance, from east to west in BV, suggests that mixing has occurred more recently in the west in this transect.

Hybrid zone movement and clinal variation

Asymmetries in cline shape have been provided as evidence for hybrid zone movement in a number of hybrid zones in nature (reviewed in Buggs 2007). However, asymmetry is viewed as equivocal evidence for movement because it could also explain adaptive introgression away from a stationary front (Dasmahapatra *et al.* 2002). The shift in cline centre, either to the east or to the west, as cline width increases (Fig. 4), is suggestive of asymmetrical introgression in the house mouse hybrid zone. Across the BV transect, asymmetry is especially pronounced and previous studies have interpreted this pattern as a possible example of adaptive introgression of *M. m. domesticus* alleles away from a stationary front (e.g. Teeter *et al.* 2008). However, coupled with evidence for increased LD from east to west, we now suggest that *M. m. domesticus* alleles are lagging behind a westward moving front and, as such, they may represent 'neutral' footprints of this movement. It is interesting to note that several markers in BV show a shift in cline centre to the west. Given the direction of hybrid zone movement, these markers may

be in genomic regions where *M. m. musculus* alleles are conferring fitness advantages in hybrid mice. For CZ, asymmetrical introgression is not as pronounced as in BV and occurs in both directions depending on the SNP marker. This may be a signature of movement in both directions as suggested by the pattern of LD discussed earlier or evidence for adaptive introgressions. In a previous study, Macholán *et al.* (2011) found evidence for X-linked introgression in both directions. However, these authors found qualitative differences between the east- and west-introgressing markers. The west-introgressing markers were geographically localized and displayed significant linkage disequilibria. The east-introgressing loci formed isolated and mutually independent islands of *M. m. domesticus* alleles on a *M. m. musculus* background. This pattern was interpreted as being a result of westward zone movement leaving some alleles behind as footprints of this movement (Macholán *et al.* 2011).

The role of selection and geography in hybrid zone movement

Of the 23 cases of possible hybrid zone movement reviewed by Buggs (2007), selection was proposed as the mechanism responsible for the pattern in all instances. These, as well as more recent studies (Chatfield *et al.* 2010), suggest, in particular, that mating asymmetries and responses to environmental changes (García-París *et al.* 2003; Leaché & Cole 2007) may be selective explanations for hybrid zone movement. In house mice, behaviour of both males and females is thought to play a role in reproductive isolation (Vošlaje-rová Bímová *et al.* 2011) and asymmetric behavioural selection has been proposed as a possible explanation for hybrid zone movement (Raufaste *et al.* 2005; Teeter *et al.* 2008 and 2010), although it is difficult to imagine how these could result in differences in hybrid zone movement between transects. Asymmetric sterility because of genetic incompatibilities between the X chromosome and autosomes in some reciprocal crosses (Storchová *et al.* 2004; Britton-Davidian *et al.* 2005; Good *et al.* 2008a,b) might also play a role. In this case, the differences in the direction of hybrid zone movement may result from known instances of polymorphism of hybrid sterility alleles (Forejt and Ivanyi, 1975; Britton-Davidian *et al.* 2005; Good *et al.* 2008b; Vy-skočilová *et al.* 2009). Alternatively, differences in hybrid zone movement between transects may result from selective effects associated with responses to the local (extrinsic) environment. However, because our simulations are equivocal regarding the role of selection vs. geography-dependent dispersal in hybrid zone movement, it is also possible that dispersal asymmetry

is the result of human-mediated transportation of mice and land-use change. This possibility was suggested as an explanation for clinal patterns in the Danish hybrid zone transect (Raufaste *et al.* 2005). Alongside selection, it could also account for the remarkable differences between the CZ and BV transects. Continued sampling over time of the same populations in CZ and BV may help determine the role of selection vs. geography in hybrid zone movement. If selection is responsible for the observed patterns, one might expect to see continued evidence of movement over time. If, however, movement is geography dependent, one might expect that changes in the landscape may alter or even halt movement.

References

Ardlie KG, Kruglyak L, Seielstad M (2002) Patterns of linkage disequilibrium in the human genome. *Nature Reviews Genetics*, **3**, 299–309.

Arntzen JW, Wallis GP (1991) Restricted gene flow in a moving hybrid zone of the newts *Triturus cristatus* and *T. marmoratus* in western France. *Evolution*, **45**, 805–826.

Barton NH (1979) The dynamics of hybrid zones. *Heredity*, **43**, 341–359.

Barton N, Bengtsson BO (1986) The barrier to genetic exchange between hybridizing populations. *Heredity*, **57**, 357–376.

Barton NH, Gale KS (1993) Genetic analysis of hybrid zones. In: *Hybrid Zones and the Evolutionary Process* (ed. Harrison RG), pp. 13–45. Oxford University Press, New York.

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, **16**, 113–148.

Barton NH, Hewitt GM (1989) Adaptation, speciation and hybrid zones. *Nature*, **341**, 497–503.

Boissinot S, Boursot P (1997) Discordant phylogeographic patterns between the Y chromosome and mitochondrial DNA in the house mouse: selection on the Y chromosome? *Genetics*, **146**, 1019–1034.

Božíková E, Munclinger P, Teeter KC, Tucker PK, Macholán M, Piálek J (2005) Mitochondrial DNA in the hybrid zone between *Mus musculus musculus* and *Mus musculus domesticus*: a comparison of two transects. *Biological Journal of the Linnean Society*, **84**, 363–378.

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.

Buerkle CA (2005) Maximum-likelihood estimation of a hybrid index based on molecular markers. *Molecular Ecology Notes*, **5**, 684–687.

Buggs RJA (2007) Empirical study of hybrid zone movement. *Heredity*, **99**, 301–312.

Chatfield MWH, Kozak KH, Fitzpatrick BM, Tucker PK (2010) Patterns of differential introgression in a salamander hybrid zone: inferences from genetic data and ecological niche modeling. *Molecular Ecology*, **19**, 4265–4282.

Cucchi T, Vigne JD, Auffray JC (2005) First occurrence of the house mouse (*Mus musculus domesticus* Schwarz & Schwarz, 1943) in the Western Mediterranean: a zooarchaeological revision of subfossil occurrences. *Biological Journal of the Linnean Society*, **84**, 429–445.

Cucchi T, Balasescu A, Bem C, Radu V, Vigne JD, Tresset A (in press) New insights into the invasive process of the eastern house mouse (*Mus musculus musculus*) in Europe: evidences from the burnt houses of Chalcolithic Romania. *The Holocene*, in press.

Dallas JF, Bonhomme F, Boursot P, Britton-Davidian J, Bauchau V (1998) Population genetic structure in a Robertsonian race of house mice: evidence from microsatellite polymorphism. *Heredity*, **80**, 70–77.

Dasmahapatra KK, Blum MJ, Aiello A *et al.* (2002) Inferences from a rapidly moving hybrid zone. *Evolution*, **56**, 741–753.

Dod B, Jermiin LS, Boursot P, Chapman VH, Tonnes-Nielsen J, 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.

Dufková P, Macholán M, Piálek J (2011) Inference of selection and stochastic effects in the house mouse hybrid zone. *Evolution*, **65**, 993–1010.

Fel-Clair F, Lenormand T, Catalan J *et al.* (1996) Genomic incompatabilities 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.

Forejt J (1996) Hybrid sterility in the mouse. *Trends in Genetics*, **12**, 412–417.

Forejt J, Ivanyi P (1975) Genetic studies on male sterility of hybrids between laboratory and wild mice (*Mus musculus* L.). *Genetical Research*, **24**, 189–206.

García-París M, Alcobendas M, Buckley D, Wake DB (2003) Dispersal of viviparity across contact zones in Iberian populations of fire salamanders (*Salamandra*) inferred from discordance of genetic and morphological traits. *Evolution*, **57**, 129–143.

Gompert Z, Buerkle CA (2010) INTROGRESS: a software package for mapping components of isolation in hybrids. *Molecular Ecology Resources*, **10**, 378–384.

Good JM, Dean MD, Nachman MW (2008a) A complex genetic basis to X-linked hybrid male sterility between two species of house mice. *Genetics*, **179**, 2213–2228.

Good JM, Handel MA, Nachman MW, Feder J (2008b) Asymmetry and polymorphism of hybrid male sterility during the early stages of speciation in house mice. *Evolution*, **62**, 50–65.

Good JM, Giger T, Dean MD, Nachman MW (2010) Widespread over-expression of the X chromosome in sterile F1 hybrid mice. *PLoS Genetics*, **6**, e1001148.

Hauffe HC, Piálek J, Searle JB (2000) The house mouse chromosomal hybrid zone in Valtellina (SO): a summary of past and present research. *Hystirx*, **11**, 17–25.

Hedrick PW (2005) *Genetics of Populations*. Jones & Bartlett, Sudbury, MA.

Hewitt GM (1975) A sex-chromosome hybrid zone in the grasshopper *Podisma pedestris* (Orthoptera: Acrididae). *Heredity*, **35**, 375–387.

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics*, **38**, 226–231.

Hudson RR, Boos DD, Kaplan NL (1992) A statistical test for detecting geographic subdivision. *Molecular Biology and Evolution*, **9**, 138–151.

Hunt WG, Selander RK (1973) Biochemical genetics of hybridisation in European house mice. *Heredity*, **31**, 11–33.

Jones EP, van der Kooij J, Solheim R, Searle JB (2010) Norwegian house mice (*Mus musculus musculus/domesticus*): distributions, routes of colonization and patterns of hybridization. *Molecular Ecology*, **19**, 5252–5264.

Key KHL (1968) The concept of stasipatric speciation. *Systematic Zoology*, **17**, 14–22.

Leaché AD, Cole CJ (2007) Hybridization between multiple fence lizard lineages in an ecotone: locally discordant variation in mitochondrial DNA, chromosomes, and morphology. *Molecular Ecology*, **16**, 1035–1054.

Li N, Stephens M (2003) Modelling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics*, **165**, 2213–2233.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Macholán M, Munclinger P, Šugerková M *et al.* (2007) Genetic analysis of autosomal and X-linked markers across a mouse hybrid zone. *Evolution*, **61**, 746–771.

Macholán M, Baird SJE, Munclinger P, Dufková P, Bímová B, Piálek J (2008) Genetic conflict outweighs heterogametic incompatibility in the mouse hybrid zone? *BMC Evolutionary Biology*, **8**, 271.

Macholán M, Baird SJE, Dufková P, Munclinger P, Vošlajerová Bímová B, Piálek J (2011) Assessing multilocus introgression patterns: a case study on the mouse X chromosome in central Europe. *Evolution*, **65**, 1428–46.

Marchini J, Cutler D, Patterson N *et al.* (2006) A comparison of phasing algorithms for trios and unrelated individuals. *American Journal Human Genetics*, **78**, 437–450.

Mihola O, Trachtulec Z, Vlček C, Schimenti JC, Forejt J (2009) A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science*, **323**, 373–375.

Moulia C, LeBrun N, Dallas J, Orth A, Renaud F (1993) Experimental evidence of genetic determinism in high susceptibility to intestinal pinworm infection in mice: a hybrid zone model. *Parasitology*, **106**, 387–393.

Orth A, Lyapunova E, Kandaurov A *et al.* (1996) Polytypic species *Mus musculus* in Transcaucasia. *Comptes Rendus De L*

Academie Des Sciences Serie Iii-Sciences de la Vie-Life Sciences, **319**, 435–441.

Payseur BA, Krenz JG, Nachman MW (2004) Differential patterns of introgression across the X chromosome in a hybrid zone between two species of house mice. *Evolution*, **58**, 2064–2078.

Pool JE, Nielsen R (2009) Inference of historical changes in migration rate from the lengths of migrant tracts. *Genetics*, **181**, 711–719.

Porter AH, Wenger R, Geiger H, Scholl A, Shapiro AM (1997) The *Pontia daplidice–edusa* hybrid zone in northwestern Italy. *Evolution*, **51**, 1561–1573.

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.

Raufaste N, Orth A, Belkhir K *et al.* (2005) Inferences of selection and migration in the Danish house mouse hybrid zone. *Biological Journal of the Linnean Society*, **84**, 593–616.

Reich DE, Cargill M, Bolk S *et al.* (2001) Linkage disequilibrium in the human genome. *Nature*, **411**, 199–204.

Sage RD, Heyneman D, Lim KC, Wilson AC (1986a) Wormy mice in a hybrid zone. *Nature*, **324**, 60–63.

Sage RD, Whitney JB 3rd, Wilson AC (1986b) Genetic analysis of a hybrid zone between *domesticus* and *musculus* mice (*Mus musculus* complex): hemoglobin polymorphisms. *Current Topics in Microbiology and Immunology*, **127**, 75–85.

Shifman S, Bell JT, Copley RR *et al.* (2006) High-resolution single nucleotide polymorphism genetic map of the mouse genome. *PLoS Biology*, **4**, 2227–2237.

Slatkin M (2008) Linkage disequilibrium understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics*, **9**, 477–485.

Smadja C, Ganem G (2005) Asymmetrical reproductive character displacement in the house mouse. *Journal of Evolutionary Biology*, **18**, 1485–1493.

Stephens M, Donnelly P (2003) A comparison of bayesian methods for haplotype reconstruction from population genotype data. *American Journal of Human Genetics*, **73**, 1162–1169.

Stephens M, Scheet P (2005) Accounting for decay of linkage disequilibrium in haplotype inference and missing data imputation. *American Journal of Human Genetics*, **76**, 449–462.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, **68**, 978–989.

Storchová R, Gregorová S, Buckiová D, Kyselová V, Divina P, Forejt J (2004) Genetic analysis of X-linked hybrid sterility in the house mouse. *Mammalian Genome*, **15**, 515–524.

Teeter KC, Payseur BA, Harris LW *et al.* (2008) Genome-wide patterns of gene flow across a house mouse hybrid zone. *Genome Research*, **18**, 67–76.

Teeter KC, Thibodeau LM, Gompert Z, Buerkle AC, Nachman MW, Tucker PK (2010) The variable genomic architecture of isolation between hybridizing species of house mice. *Evolution*, **64**, 472–485.

Tucker PK, Sage RD, Warner JH, 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, 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.

Vanlerberghe F, Boursot P, Catalan J (1988a) Analyse génétique de la zone d'hybridation entre les deux sous-especes de souris *Mus musculus domesticus* et *Mus musculus musculus* en Bulgarie. *Genome*, **30**, 427–437.

Vanlerberghe F, Boursot P, Nielsen JT, Bonhomme F (1988b) A steep cline for mitochondrial DNA in Danish mice. *Genetical Research*, **52**, 185–193.

Vošlajerová Bímová B, Macholán M, Baird SJE *et al.* (2011) Reinforcement selection acting on the European house mouse hybrid zone. *Molecular Ecology*, **20**, 2403–2424.

Vyskočilová M, Trachtulec Z, Forejt J, Piálek J (2005) Does geography matter in hybrid sterility in house mice? *Biological Journal of the Linnean Society*, **84**, 663–674.

Vyskočilová M, Pražanová G, Piálek J (2009) Polymorphism in hybrid male sterility in wild-derived *Mus musculus musculus* strains on proximal chromosome 17. *Mammalian Genome*, **20**, 83–91.

Warnes G, Gorjanc G, Leisch F, Man M (2011) Genetics: population genetics. R package version 1.3.5. http://CRAN.R-project.org/package=genetics.

Weir BS (1996) *Genetic Data Analysis II*. Sinauer Publications, Sunderland; pp. 120 (pp 1–445).

Weir BS, Hill WG, Cardon LR (2004) Allelic association patterns for a dense SNP map. *Genetic Epidemiology*, **27**, 442–450.

Yang H, Ding YM, Hutchins LN *et al.* (2009) A customized and versatile high-density genotyping array for the mouse. *Nature Methods*, **6**, 663–666.

Yang H, Wang J, Didion JP *et al.* (2011) Subspecific origin and haplotype diversity in the laboratory mouse. *Nature Genetics*, DOI 10.1038/847.

The authors are evolutionary biologists with research interests in the genetics of speciation. They utilize naturally occurring hybrid populations and wild-derived laboratory strains of house mice as study organisms. J.E.P. is interested in developing methods to detect natural selection and population history from genomic diversity data.

Data accessibility

The genotype data matrix is available from http://hdl.handle.net/2027.42/83674.

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Genotype matrices for the BV transect (A) and the CZ transect (B). Each column is an individual mouse and each row is an individual SNP. Mice are arranged geographically from west (left) to east (right) and SNPs are arranged by chromosome. Genotypes are color-coded in the following way. Green is homozygous *M. m. domesticus*, blue is homozygous *M. m. musculus*, red is a heterozygous genotype, and white is

missing data [Correction added after online publication 14 June 2011: the labelling in Fig. S1 was corrected].

Fig. S2 A comparison of geographic clines estimated using a 2-parameter model between the Bavaria and Czech transects. The autosomal clines were calculated based on median values of cline center and width for all SNPs (see Table 1). The X-linked clines were calculated for 84 or 85 SNP markers in the BV and CZ transects, respectively. For comparative purposes the median cline center, estimated from autosomal markers, was set to 0 km in both transects.

Fig. S3 Decay of LD for samples sizes of 10 (black), 15 (red) and 25 (blue) from a single population (Neufahrn) sampled in two separate years. Dashed lines = mice from 1984, solid lines = mice from 1985.

Table S1 ID's (Field #), sex, locality information, collection year, heterozygosity and hybrid index for all samples.

Table S2 Sample ID's and localities of mice genotyped with the Mouse Diversity Array and used to sort for fixed differences between *M. m. musculus* and *M. m. domesticus* mice.

Table S3 Result of F_{ST} analyses using program DnaSP v. 5.10.

Table S4 List of SNP markers used in the study and their chromosomal locations (mouse genome-Build 37).

Table S5 Estimates of cline center and width for all autosomal and X-linked markers.

Table S6 Values of r^2, D', half-length of LD (HLD), estimated age in generations, for samples ($n = 10$, 15, 19 or 20) collected at specific localities in specific years. *Distance is calculated from the western-most locality. **Total number SNP comparisons at ≤2Mb distance are based on 100 permutations of samples at each locality. SNPs with minor allele frequency ≤0.05 and missing values >50% are excluded. ***The age, in generations, of each population was calculated based on the decline of r^2 as described in the Material and methods.

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