Evolution, 58(9), 2004, pp. 2064–2078

DIFFERENTIAL PATTERNS OF INTROGRESSION ACROSS THE X CHROMOSOME IN A HYBRID ZONE BETWEEN TWO SPECIES OF HOUSE MICE

BRET A. PAYSEUR,[1,2] JAMES G. KRENZ,[1,3] AND MICHAEL W. NACHMAN[1]
[1]*Department of Ecology and Evolutionary Biology, BioSciences West Building, University of Arizona, Tucson, Arizona 85721*

Abstract.—A complete understanding of the speciation process requires the identification of genomic regions and genes that confer reproductive barriers between species. Empirical and theoretical research has revealed two important patterns in the evolution of reproductive isolation in animals: isolation typically arises as a result of disrupted epistatic interactions between multiple loci and these disruptions map disproportionately to the X chromosome. These patterns suggest that a targeted examination of natural gene flow between closely related species at X-linked markers with known positions would provide insight into the genetic basis of speciation. We take advantage of the existence of genomic data and a well-documented European zone of hybridization between two species of house mice, *Mus domesticus* and *M. musculus*, to conduct such a survey. We evaluate patterns of introgression across the hybrid zone for 13 diagnostic X-linked loci with known chromosomal positions using a maximum likelihood model. Interlocus comparisons clearly identify one locus with reduced introgression across the center of the hybrid zone, pinpointing a candidate region for reproductive isolation. Results also reveal one locus with high frequencies of *M. domesticus* alleles in populations on the *M. musculus* side of the zone, suggesting the possibility that positive selection may act to drive the spread of alleles from one species on to the genomic background of the other species. Finally, cline width and cline center are strongly positively correlated across the X chromosome, indicating that gene flow of the X chromosome may be asymmetrical. This study highlights the utility of natural populations of hybrids for mapping speciation genes and suggests that the middle of the X chromosome may be important for reproductive isolation between species of house mice.

Key words.—Cline, gene flow, mammals, reproductive isolation, speciation.

Received December 17, 2003. Accepted June 21, 2004.

A complete understanding of the process of speciation requires elucidation of the underlying genetic details. Ultimately, evolutionary biologists would like to know the identities, phenotypic effects, and genomic locations of genetic changes that confer reproductive isolation between newly formed species. This is a formidable challenge, but recent work has identified a handful of genes that cause fitness reductions in interspecific hybrids and, hence, may contribute to the evolution of reproductive isolation.

Fine-scale mapping experiments suggest that *Odysseus*, a gene containing a homeobox domain, is associated with hybrid male sterility in crosses between *Drosophila sechellia* and *D. mauritiana* (Ting et al. 1998). Additionally, *Hmr* causes lethality and female sterility in hybrids between *D. melanogaster* and its sibling species (Barbash et al. 2003). Hybrids between *Xiphophorus maculatus* platyfish and *X. helleri* swordtails display melanomas (which can be fatal) caused by overexpression of the *Xmrk-2* gene when a particular allele at a repressor gene is not present (Malitschek et al. 1995), earmarking this locus as a candidate for the evolution of reproductive isolation via hybrid inviability. Finally, Presgraves and colleagues (Presgraves 2003; Presgraves et al. 2003) used a hybrid rescue mutation and crosses between *D. simulans* and strains of *D. melanogaster* containing small, individual deletions to identify a disrupted interaction between the autosomal, nucleoporin *Nup96* gene in *D. simulans* and a locus on the *D. melanogaster* X chromosome as a cause of hybrid lethality. Patterns of nucleotide variation within and between the two species suggest that positive selection has driven species divergence at this gene. Extensive divergence time between *D. melanogaster* and *D. simulans* prohibits the conclusion that *Nup96* contributed to the original development of reproductive isolation between the two species, but this locus seems to be a strong candidate for a gene associated with speciation (Noor 2003).

Although the identities of genes contributing to reproductive isolation have been established in only a few cases, considerable empirical and theoretical effort has revealed generalities that point toward productive directions in research on the genetic basis of speciation. First, reproductive isolation typically results from substitutions at different, interacting loci (as proposed by Bateson 1909; Dobzhansky 1936; Muller 1940, 1942; Orr 1996). Numerous examples of such complementary genes have been described in plants and animals (Hollingshead 1930; Dobzhansky 1936; Wu and Beckenbach 1983; Christie and Macnair 1984; Orr 1987; Pantazidis and Zouros 1988; Orr and Coyne 1989; Wittbrodt et al. 1989; Perez and Wu 1995; True et al. 1996). Theoretical work indicates that these incompatible substitutions can accumulate rapidly, particularly as the number of participating loci increases (Orr 1995).

Second, the loci contributing to reproductive isolation in animals are found disproportionately on the X chromosome (Coyne and Orr 1989). Crosses between species pairs, primarily in *Drosophila*, have consistently mapped loci with major effects on hybrid sterility and inviability, in males and females, to the X chromosome (Dobzhansky 1936; Crow 1942; Grula and Taylor 1980; Zouros et al. 1988; Orr 1989; Orr and Coyne 1989). A recent introgression experiment between *D. simulans* and *D. mauritiana* suggested that the X

[2] Present address: Department of Molecular Biology and Genetics, Biotechnology Building Room 227, Cornell University, Ithaca, New York 14853; E-mail: bp69@cornell.edu.

[3] Present address: USDA-Agricultural Research Service, Hatfield Marine Science Center, 2030 SE Marine Science Dr, Newport, Oregon 97365.

chromosome has a density of factors contributing to hybrid male sterility that is 2.5 times that of the autosomes (Tao et al. 2003). Furthermore, the disproportionate effect of the X chromosome is observed whether males or females are the heterogametic sex. For example, genes underlying reproductive isolation between butterfly species also appear to be overrepresented on the Z chromosome (the butterfly analog of the X chromosome; Sperling 1994; Prowell 1998; Jiggins et al. 2001). Less direct but more taxonomically widespread evidence for the involvement of the X chromosome in reproductive isolation comes from the observation of Haldane's (1922) rule: when in the offspring of two different animal races one sex is absent, rare, or sterile, that sex is the heterogametic sex. This generalization has been upheld across a variety of animal species, including *Drosophila*, butterflies, moths, mammals, amphibians, reptiles, and birds (Haldane 1922; Laurie 1997; Orr 1997). Although some cases of Haldane's rule may be due to incompatibility of the Y chromosome with a heterospecific genetic background, incompatibilities involving X-linked recessive mutations are expected to be especially important (Turelli and Orr 1995, 2000). Finally, X-linked markers show reduced introgression across a number of hybrid zones (Hagen 1990; Sperling and Spence 1991; Tucker et al. 1992; Dod et al. 1993), suggesting that interactions including loci on the X chromosome are disrupted in hybrids.

These two generalizations—that reproductive isolation is caused by disrupted epistatic interactions and that such interactions are likely to involve the X chromosome—suggest a productive avenue to the genetic dissection of reproductive isolation: a targeted examination across the X chromosome for substitutions that are incompatible with the genomic background of the alternative species. A particularly exciting approach is to study natural variation in levels of gene flow between newly formed species across the X chromosome. Hybrid zones provide excellent arenas for such an investigation and offer three advantages relative to standard laboratory crossing designs. First, genotypic distributions in hybrid zones typically reflect more generations of recombination than in F_2 or backcross laboratory designs, allowing more precise mapping of individual regions causing reproductive isolation. Second, incompatibilities with small fitness effects may be easier to detect in natural populations than in the laboratory. Third, the identification of genomic regions conferring reproductive isolation in hybrid zones does not require any knowledge or assumptions about the relevant phenotype. Indeed, the study of hybrid zones has contributed substantially to ideas about the genetic architecture of speciation (Endler 1977; Barton 1983; Barton and Hewitt 1985; Szymura and Barton 1986, 1991; Harrison 1990). For example, the number of loci underlying reproductive isolation can be estimated from clines that show stepped allele frequency patterns across hybrid zones (Barton and Hewitt 1985; Szymura and Barton 1986; Barton and Gale 1993). In one approach, the dispersal rate is estimated from linkage disequilibrium and the width and size of the step in allele frequency are calculated. These numbers can then be compared to yield estimates of the strength of selection against hybrids and the number of loci involved. Application of this approach to hybrid zone data from a small number of molecular markers has suggested a polygenic basis for reproductive isolation in some cases (e.g., toads: Szymura and Barton 1986, 1991) and a smaller number of factors in others (e.g., butterflies: Porter et al. 1997).

Hybrid zones can also be used to identify the individual genomic regions involved in speciation by investigating patterns of differential introgression at multiple loci with known map positions. Under this approach, markers linked to genes causing reproductive isolation are expected to introgress at lower rates than markers unlinked to such genes. Rieseberg et al. (1999) used this rationale to locate 26 genomic regions contributing to reproductive isolation between two sunflower species, *Helianthus petiolaris* and *H. annuus*. Sixteen of these regions were associated with pollen sterility, providing a biological explanation for potential reductions in hybrid fitness caused by these genetic changes. The power of this differential introgression approach would be even greater in species for which genome sequences are available (providing knowledge about genes linked to surveyed markers), but the combination of genomic information with the propensity to form natural hybrid zones is rare.

One prominent exception is a pair of closely related house mouse species, *M. domesticus* and *M. musculus* (also referred to as *M. musculus domesticus* and *M. musculus musculus*, respectively). The genome sequence of the C57BL/6J inbred strain, which is largely a genetic hybrid between *M. domesticus* and *M. musculus* (Yonekawa et al. 1980; Ferris et al. 1982; Bishop et al. 1985; Wade et al. 2002), was recently described (Mouse Genome Sequencing Consortium 2002) and a dense genetic map is available (Dietrich et al. 1996). These two species diverged approximately 350,000 years (700,000 generations) ago (She et al. 1990) and are distinguished morphologically by differences in relative tail length (longer in *M. domesticus*) and craniofacial shape (longer and narrower in *M. domesticus*; Macholan 1996). *Mus domesticus* ranges across western Europe, northern Africa, and the middle East, whereas the range of *M. musculus* extends throughout eastern Europe and northern Asia. These two species form a hybrid zone that stretches across Europe, from the Jutland Peninsula to the Bulgarian coast of the Black Sea (Boursot et al. 1993; Sage et al. 1993). This zone represents a region of secondary contact between the two species, having formed as a consequence of the spread of human agriculture and shipping into Europe during the Neolithic transition (Ammerman and Cavalli-Sforza 1984; Sage et al. 1993). The zone is estimated to be 6000 years old at its southeastern edge (Sage et al. 1993) and 250 years old at its northwestern edge (Hunt and Selander 1973).

Several lines of evidence suggest partial reproductive isolation between these two species. First, crosses between wild *M. musculus* and some laboratory inbred strains (which are derived primarily from *M. domesticus*) yield sterile hybrid males (but females are fertile, consistent with Haldane's rule; Forejt and Ivanyi 1975; Forejt 1996). Additionally, hybrid males produced by crossing wild individuals from both species are sometimes sterile (J. Pialek, pers. comm.). Second, hybrids from multiple transects of the European hybrid zone harbor more parasites than do pure *M. domesticus* or *M. musculus* individuals (Sage et al. 1986; Moulia et al. 1993; Moulia et al. 1995). Finally, changes in allele frequencies of di-

agnostic molecular markers across the European hybrid zone occur very rapidly relative to the geographic extent of the species ranges (Hunt and Selander 1973; Vanlerberghe et al. 1986; Tucker et al. 1992). These observations indicate that the hybrid zone is primarily maintained by a balance between selection against hybrids and dispersal, making it an excellent natural arena for studying the genetics of speciation.

The European hybrid zone between *M. domesticus* and *M. musculus* has been intensively studied for more than half a century, beginning with phenotypic characterization of populations with individuals showing differences in tail length (Ursin 1952) and investigation of patterns of allozyme variation across the zone (Selander and Yang 1969; Hunt and Selander 1973; Schnell and Selander 1981). Examination of multiple transects of the European hybrid zone has yielded intriguing insights into the genetic basis of reproductive isolation between *M. domesticus* and *M. musculus*. First, while most loci show reduced gene flow relative to the geographic extent of the species ranges, there is clear heterogeneity in levels of introgression between genomic regions (Vanlerberghe et al. 1986; Tucker et al. 1992; Boursot et al. 1993; Dod et al. 1993; Sage et al. 1993). Hence, the genome appears to be semipermeable, with some regions tolerant of gene flow between species and others not. Second, Y chromosome introgression is usually inhibited (Vanlerberghe et al. 1986; Tucker et al. 1992; Dod et al. 1993; but see Munclinger et al. 2002), suggesting a role for Y-linked incompatibilities in mouse speciation. Because of the lack of recombination, however, genes on the Y chromosome underlying reproductive isolation will be difficult to find using comparative introgression (or mapping) approaches. Finally, the X chromosome shows reduced gene flow across three transects (Tucker et al. 1992; Dod et al. 1993; Munclinger et al. 2002). Using mice from southern Germany and western Austria, Tucker et al. (1992) studied gene flow at two loci on the X chromosome. Both markers displayed reduced introgression relative to most autosomal loci, with the locus mapping to the proximal (centromeric) part of the chromosome exhibiting the largest reduction. Dod et al. (1993) analyzed clines across a Danish transect at three X-linked loci and identified one locus in the central part of the chromosome showing the least introgression. An X-linked marker also shows reduced gene flow across a transect in the Czech and Slovak Republics (Munclinger et al. 2002). These studies suggest that exchange of the X chromosome between *M. domesticus* and *M. musculus* is retarded in nature and there is variation in the degree of introgression within the X chromosome. Therefore, a study of gene flow between these two species at loci situated across the X chromosome is a productive strategy for identifying genomic regions involved in reproductive isolation.

In this paper, we report patterns of introgression for 13 diagnostic X-chromosomal markers across a southern German transect of the hybrid zone between *M. domesticus* and *M. musculus*. Our results highlight substantial variation in gene flow across the X chromosome. We identify one locus exhibiting clearly reduced introgression across the center of the hybrid zone, a candidate region for reproductive isolation, and one region displaying signs of increased gene flow from *M. domesticus* to *M. musculus*, potentially associated with positive selection on a heterospecific genomic background.

Materials and Methods

Samples and Marker Identification

Mice were live-trapped in 1984, 1985, and 1992 by R. Sage. A map of the collecting localities for the mice used in this study is provided in Figure 1, and the localities are listed in Table 1. Genomic DNAs and tissues were generously provided by R. Sage.

Informative single nucleotide polymorphisms (SNPs) were identified by polymerase chain reaction (PCR) amplification and direct sequencing of two individuals each of *M. domesticus* and *M. musculus* from outside the hybrid zone, searching for restriction fragment length polymorphisms (RFLPs), and surveying candidate RFLPs (via digestion of PCR products) for fixed differences in a larger panel including up to 10 individuals from each species. Eleven markers were identified in introns, with names corresponding to the genes in which they were located. Two additional markers were found in intergenic regions, including *Nt* (named by us) and one marker (*DXMit18.2*) from Lindblad-Toh et al. (2000).

Loci were amplified by PCR under the following conditions: 95°C for 15 min, 94°C for 30 sec, annealing temperature (see Supplementary Table 1 available online only at http://dx.doi.org/10.1554/03-738.1.s1) for 30 sec, and 72°C for 30 sec, with the last three steps being repeated for 40 cycles. The panel included *M. domesticus* from Italy and Spain and *M. musculus* from the Czech Republic and Serbia. RFLPs that showed fixed differences in this panel were selected for genotyping in the hybrid zone. Loci were found in the genome sequence using the UCSC mouse genome browser (February 2003 version; www.genome.ucsc.edu). For each locus, the genetic position (in cM) was taken to be that of the physically closest marker in the sequence that was mapped in the Whitehead-MIT F_2 intercross (Dietrich et al. 1996; Fig. 2). Locations on genetic and sequence-based maps, PCR annealing temperatures, and restriction enzymes for each locus are listed in Supplementary Table 1 available online. Individual genotypes are provided in Supplementary Table 2 available online only at: http://dx.doi.org/10.1554/03-738.1.s2.

Data Analysis

Frequencies of *M. domesticus* alleles were calculated separately for each locality and each locus (Supplementary Table 3 available online only at http://dx.doi.org/10.1554/03-738.1.s3). Locality distances were measured along a straight line running (west to east) through the transect starting at a point at its western edge. Measures of introgression were estimated using an approach developed by Szymura and Barton (1986) and modified by Porter et al. (1997) to account for sex linkage. The model relates allele frequency (p) and geographic distance (x) using three equations. Equation (1b) describes the sinusoidal shape in the center of the cline and equations (1a, c) describe the exponential change in allele frequency on either side of this center:

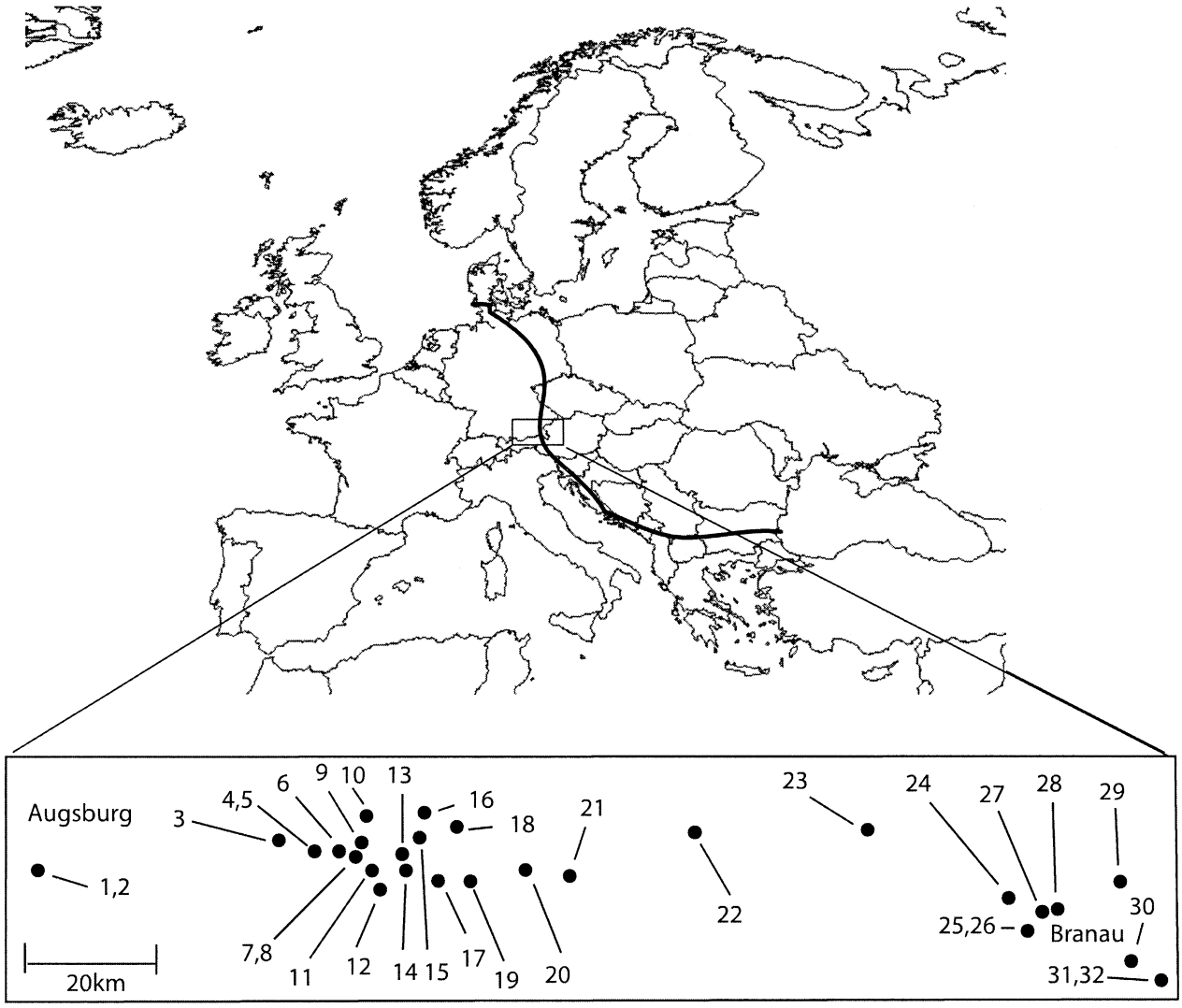

FIG. 1. Map of collecting localities along a transect through the hybrid zone between *Mus domesticus* and *M. musculus* in southern Germany and western Austria. Locality numbers correspond to those in Table 1. The dark line shows the approximate position of the hybrid zone throughout Europe (modified from Sage et al. 1993).

$$p = \exp\left\{\frac{4[x - (c + z_L)]\sqrt{\theta_L}}{w}\right\}, \quad (1a)$$

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

$$p = 1 - \exp\left\{\frac{-4[x - (c + z_R)]\sqrt{\theta_R}}{w}\right\}. \quad (1c)$$

In this model, c is the location of the center of the zone, w is the width of the zone (1/slope at c), z_L and z_R represent distances from c to a vertical asymptote for the exponential decay on the left and right sides of the zone (respectively), and θ_L and θ_R are the rates of exponential decay on the left and right sides of the zone relative to the shape of the central cline from equation (1b). In total, these equations allow estimation of six parameters that provide information about introgression patterns. Cline width (w) describes the rate of change in allele frequency in the center of the zone, where the frequency changes most rapidly (Endler 1977). Models of selection against hybrids produce w proportional to the ratio between dispersal (σ, the standard deviation of the distance between parent and offspring) and the square root of the strength of selection (s) (Bazykin 1969; Slatkin 1973; Barton and Gale 1993). The center of a cline (c) is the point at which allele frequency changes most rapidly and provides information about the overall geographic location of allele frequency gradients (Endler 1977). θ_L and θ_R describe the exponential rate of change in allele frequency in the western and eastern tails of the cline, respectively, while z_L and z_R estimate the distance over which this change occurs (as measured from c). Strong nonrandom associations among loci (linkage disequilibrium) are generated in the center of the hybrid zone by dispersal and mating between individuals from parental populations from different sides (which differ in allele frequency) and selection against hybrid genotypes. Moving out from the center of the hybrid zone, individuals exhibit genotypes resulting from continued backcrossing and

TABLE 1. Hybrid zone localities, distances from western edge of transect, and numbers of chromosomes sampled (averaged across loci).

Locality	Distance (km)	Number of chromosomes sampled
1. Augsburg (A), Germany	2.2	17
2. Augsburg (B), Germany	2.2	11
3. Ebersbach, Germany	39.6	9
4. Kammerberg-Zandt, Germany	45.2	7
5. Kammerberg-Hartl, Germany	45.2	6
6. Appercha, Germany	49.2	27
7. Gesselthausen-Ziigletrumm, Germany	51.8	109
8. Gesselthausen-Warta, Germany	51.8	63
9. Giesenbach, Germany	52.4	4
10. Ebserspoint, Germany	53.2	16
11. Massenhausen, Germany	53.8	15
12. Neufahrn bei Freising, Germany	55.6	103
13. Pulling-Petryszak, Germany	58.8	12
14. Achering, Germany	59.2	30
15. Freising, Germany	61.4	6
16. Tuntenhausen, Germany	62.2	9
17. Gut Wildschwaig, Germany	64.4	6
18. Rudlfing, Germany	67.2	30
19. Schwaig, Germany	69.2	23
20. Tittenkofen, Germany	77.2	3
21. Sonnendorf, Germany	84.6	32
22. Brundl, Germany	103.4	23
23. Attenham, Germany	130.4	13
24. Simbach, Germany	152.4	27
25. Ranshofen-Holfinger, Austria	155.0	6
26. Ranshofen-Penias, Austria	155.6	4
27. Branau, Austria	157.4	5
28. Nofing, Austria	160.4	24
29. Aufhausen, Austria	170.0	3
30. Rodham, Austria	170.2	10
31. Leitham-Fuchs, Austria	176.0	8
32. Leitham-Hubinger, Austria	176.4	3

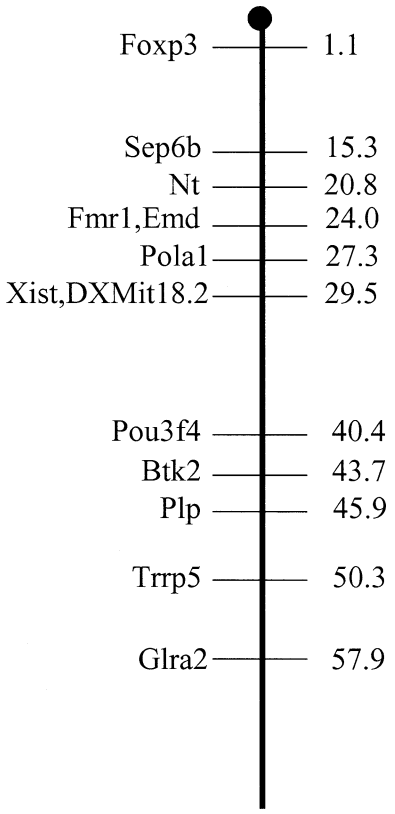

FIG. 2. Loci surveyed in this study and their genetic positions (cM) on the X chromosome.

recombination; consequently, the patterns of change in allele frequency in the tails of the cline (θ_L, θ_R, z_L, and z_R) are more indicative of locus-specific forces acting on purer genomic backgrounds.

Parameters were estimated for each locus separately using likelihood and ClineFit software kindly provided by A. Porter (available at http://www-unix.oit.umass.edu/~aporter/software/), which allows the analysis of loci with haplodiploid inheritance patterns (such as those on the X chromosome). The likelihood of observed allele frequencies given the parameter values was assumed to follow a binomial distribution, with expected allele frequencies derived from equations (1a–c) (Porter et al. 1997).

Numerical searching of the parameter space used a simulated annealing (Metropolis) algorithm (for details, see Szymura and Barton 1986; Porter et al. 1997). For each locus, the fits of two models to the data were compared: one with two parameters (*c* and *w*) and one with six parameters (*c*, *w*, z_L, z_R, θ_L, and θ_R). Likelihood ratio tests indicated that the six-parameter model provided a statistically better fit to the data for 12 of the 13 loci. Hence, all subsequent analyses were performed using the six-parameter model.

This likelihood framework allowed comparison of cline shape between pairs of loci as follows. For a given comparison, the second locus was constrained to the parameter estimates for the first locus and the resulting likelihood was estimated. This likelihood was compared to the likelihood estimated using an unconstrained model (in which the maximum likelihood estimate of the parameter was obtained). To assess statistical significance, twice the difference in log-likelihood values between the constrained and unconstrained models was compared to a chi-square distribution with six degrees of freedom (since the constrained and unconstrained models differed by six free parameters). Using a Bonferroni correction for the performance of multiple (156) tests, significantly different cline shapes were those with *P*-values less than 0.0003. Additionally, individual parameter estimates (such as cline width) were compared between locus pairs using two-unit support limits (which correspond roughly to 95% confidence limits; Edwards 1992) derived from the likelihood searches.

We also compared levels of introgression to local recombination rate and gene density. We estimated recombination rates by comparing the genetic (using the Whitehead-MIT map; Dietrich et al. 1996) and physical (using the February 2003 version of the genome sequence; Mouse Genome Sequencing Consortium 2002) positions of microsatellites from across the X chromosome. Recombination rate (in cM/Mb) at a given locus was estimated as the slope of a linear regression of genetic versus physical position, including five markers on each side. We estimated gene density by counting the number of predicted genes in a 2-Mb window centered on each locus using data from the UCSC mouse genome browser (February 2003 version; www.genome.ucsc.edu).

TABLE 2. Cline parameter estimates (two-unit support limits) for loci surveyed in this study.

Locus	c	w	θ_L	θ_R	z_L	z_R
Foxp3	56.25 (55.61–56.99)	6.14 (4.63–8.02)	0.49103 (0.00028–0.99932)	0.00029 (0.00004–0.00163)	13.9 (1.9–998.9)	260.4 (80.4–667.6)
Sep6b	55.86 (55.33–56.58)	7.53 (6.10–9.69)	0.02066 (0.00016–0.85684)	0.00006 (0.00003–0.00237)	40.0 (2.1–995.4)	982.0 (143.7–999.9)
Nt	54.89 (54.13–55.39)	4.02 (2.78–5.40)	0.00002 (0.00001–0.99780)	0.01656 (0.00494–0.05092)	999.2 (10.0–999.9)	7.4 (4.1–20.7)
Fmr1	54.73 (54.08–55.07)	3.60 (2.30–4.56)	0.00002 (0.00001–0.98508)	0.00233 (0.00084–0.00712)	980.2 (2.8–999.9)	23.1 (12.8–48.6)
Emd	54.12 (53.84–55.24)	1.16 (0.64–3.89)	0.00846 (0.00184–0.51683)	0.00256 (0.00073–0.03859)	8.6 (1.4–27.1)	5.7 (2.7–12.9)
Pola1	53.62 (53.34–53.88)	0.37 (0.21–0.91)	0.00000 (0.00000–0.00001)	0.00034 (0.00008–0.00221)	999.9 (317.0–999.9)	4.6 (2.4–11.2)
Xist	54.54 (53.93–55.21)	3.25 (2.02–4.65)	0.00001 (0.00001–0.61781)	<0.00001 (0.00000–0.00001)	999.6 (7.5–999.6)	999.9 (554.2–999.9)
DXMit18.2	54.49 (54.08–54.78)	2.95 (2.57–3.99)	0.45499 (0.00068–0.99983)	0.00009 (0.00021–0.00053)	0.9 (0.9–998.7)	230.0 (86.3–600.5)
Pou3f4	55.73 (55.14–56.66)	5.29 (3.89–7.90)	0.00003 (0.00001–0.75318)	0.15332 (0.03662–0.58601)	999.9 (150.9–999.9)	2.3 (1.7–17.9)
Btk2	55.78 (54.88–56.59)	5.42 (4.19–7.55)	0.00004 (0.00002–0.00469)	0.06026 (0.01751–0.16597)	999.9 (55.1–999.9)	3.8 (3.0–9.8)
Plp	56.26 (55.26–57.15)	5.72 (4.24–7.82)	0.67528 (0.00029–0.99881)	0.00205 (0.00073–0.00416)	558.7 (2.1–998.8)	21.9 (17.6–37.8)
Trrp5	54.75 (54.06–55.13)	3.59 (2.12–4.55)	0.00002 (0.00001–0.12893)	0.02494 (0.00391–0.06957)	999.9 (11.9–999.9)	6.4 (3.2–31.8)
Glra2	55.32 (54.99–55.62)	2.49 (1.21–4.32)	0.03634 (0.00638–0.22112)	0.00011 (0.00002–0.00042)	7.5 (1.3–25.0)	135.3 (51.4–334.4)

We asked whether alleles at different loci were nonrandomly associated by estimating linkage disequilibrium. Because levels of linkage disequilibrium are strongly influenced by allele frequencies and sample sizes, we restricted the analyses to one population in the center of the hybrid zone (Neufahrn bei Freising), which was polymorphic for all loci and contained a large number of sampled individuals. Linkage disequilibrium could not be directly calculated because phase was unknown in females. Instead, linkage disequilibrium (D, which can range from -0.25 to 0.25) was estimated for each two-locus combination using Hill's (1974) likelihood method, with modifications for sex-linked loci suggested by Porter et al. (1997). This approach assumes Hardy-Weinberg equilibrium to infer gametic frequencies; only two of 13 loci appeared to violate this assumption (*Fmr1* and *Xist* each exhibited a deficit of heterozygotes). We compared the likelihood of the data assuming that D equals the maximum likelihood estimate to the likelihood of the data assuming that D equals zero. Under the null hypothesis of no disequilibrium, twice this difference in log-likelihood is approximately distributed as a chi-square with one degree of freedom.

To account for the performance of multiple tests, we used a critical P-value of 0.0006 (0.05/78). Because D strongly depends on allele frequencies, we also calculated D' (Lewontin 1964), the observed value of D divided by its maximum possible value given observed allele frequencies. D' can range from -1 to 1.

RESULTS

Differential Introgression among X-Linked Loci

Cline shape parameter estimates for each locus are provided in Table 2 and the relationship between allele frequency and geographic position in the hybrid zone transect is depicted in Figure 3. The positions of the clines, as measured by c, are similar: consideration of all loci suggests a range of less than 3 km (average c = 55.10 km; range = 53.62–56.25 km). In general, cline widths are low (average w = 3.96 km), with clear variation among loci (range = 0.37–7.53 km). Cline widths are more variable than cline positions (w, CV = 51.08; c, CV = 1.51). The four tail parameters also suggest variation in the degree of introgression.

Further insight into patterns of gene flow is obtained by comparing cline shapes to chromosomal positions (Figs. 3, 4). Both w and c exhibit positional effects along the X chromosome, with adjacent markers often showing similar values. Cline width displays a clear reduction near the center of the chromosome, primarily associated with the *Pola1* marker positioned at 27.3 cM on the genetic map and 78.4 Mb in the sequence. Two-unit support limits for w at this locus overlap slightly with its proximal neighbor, which also exhibits a low w, but not with any other locus on the chromosome (Table 2; Fig. 4). Cline center shows a very similar pattern (Fig. 5), with the location of the cline at *Pola1* being significantly shifted to the west (further into *M. domesticus* territory). Hence, the cline is statistically narrower and westwardly shifted at *Pola1* relative to the remainder of the X chromosome, identifying this marker as being situated near to a candidate locus for reproductive isolation between *M. domesticus* and *M. musculus*. A list of known (i.e., confirmed) genes in this region is provided in Table 3.

Statistical comparisons of overall (six-parameter) cline shapes reveal clear variation in patterns of introgression among loci, with a large number of pairwise comparisons yielding significant differences (Table 4). Two loci, *Xist* and *Plp*, consistently show patterns of introgression that are un-

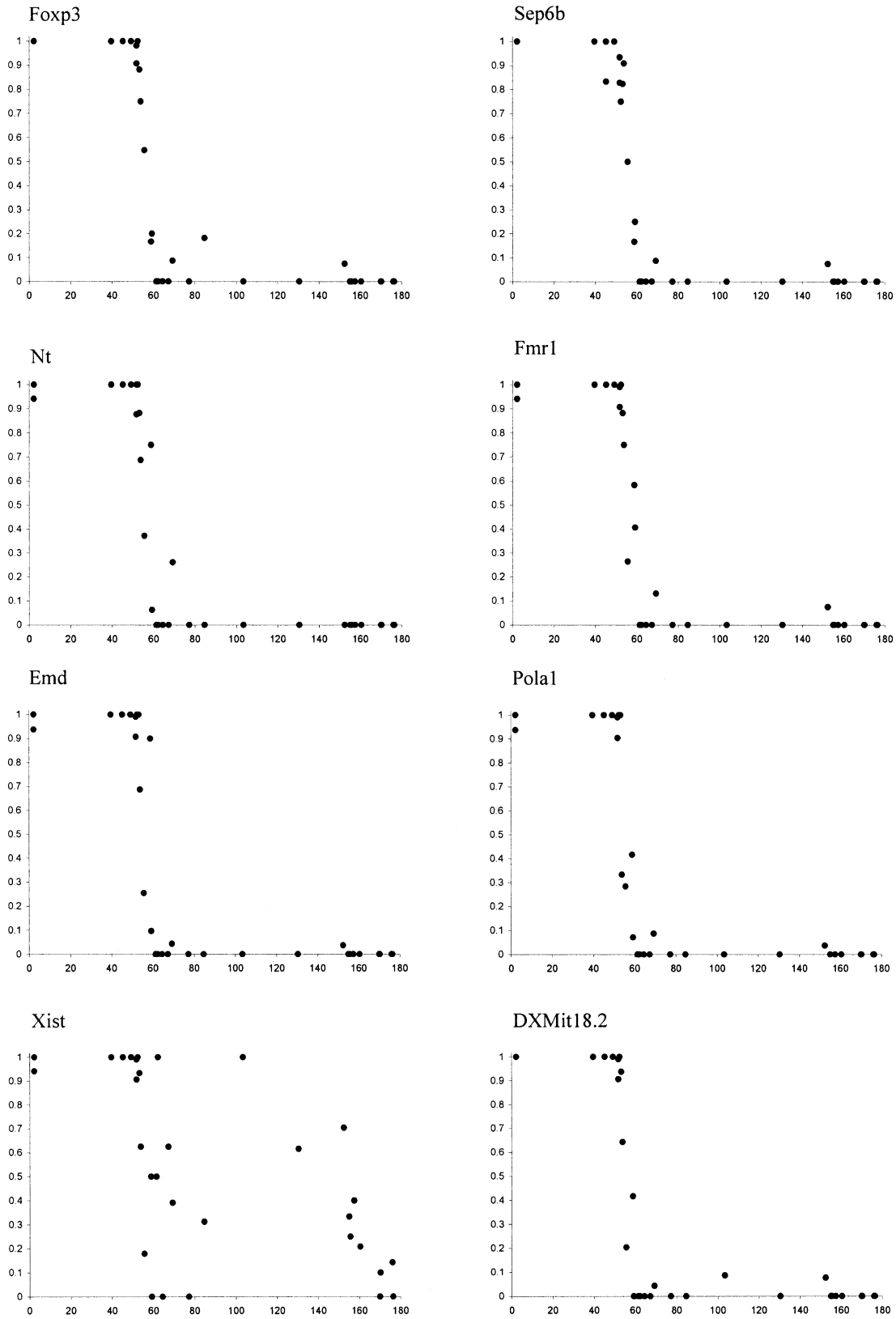

FIG. 3. Scatterplots of *Mus domesticus* allele frequency versus geographic position (km) for 13 X-linked loci. The loci are arranged by increasing physical position in the sequence.

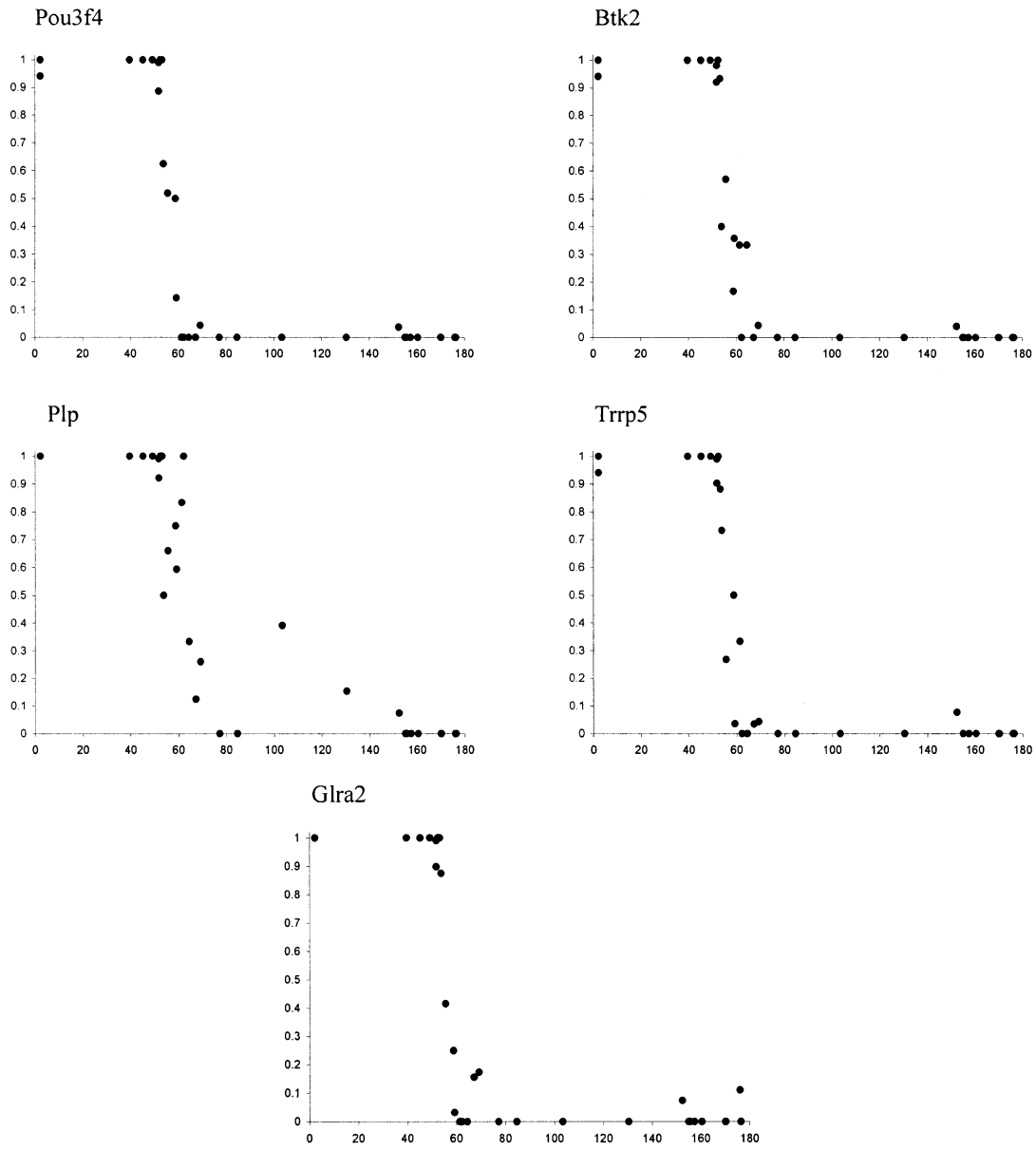

FIG. 3. Continued.

usual for the X chromosome. Both loci show *M. domesticus* alleles farther to the east than do the other loci. *Xist* shows an especially gradual change in allele frequency in the *M. musculus* tail of the cline (two-unit support limits for θ_R do not overlap with other loci across the chromosome). *Xist* does not appear unusual in terms of *w*, suggesting a decoupling of patterns of introgression in the center of the zone from those outside the center. The *Plp* cline is centered more to the east than that of the other loci, and the change in allele frequency in the *M. domesticus* tail is unusually abrupt, although consideration of two-unit support limits indicates that no individual parameters show statistical evidence of departure from the remainder of the chromosome. The *Pola1* locus also displays signs of differentiation in this overall cline shape comparison: data from *Pola1* fit the maximum likelihood parameter estimates from this locus significantly better than they fit estimates from 10 of 12 other loci (Table 4).

Asymmetrical Introgression

We also evaluated the relationship between the degree of introgression and the geographic location of the clines. Cline width and cline center are strongly positively correlated across the X chromosome (Spearman's $\rho = 0.89$; $P < 0.0001$; Fig. 6).

Introgression, Recombination Rate, and Gene Density

If a large number of loci underlying reproductive isolation are scattered randomly throughout the genome, markers linked to more genes might be expected to display less in-

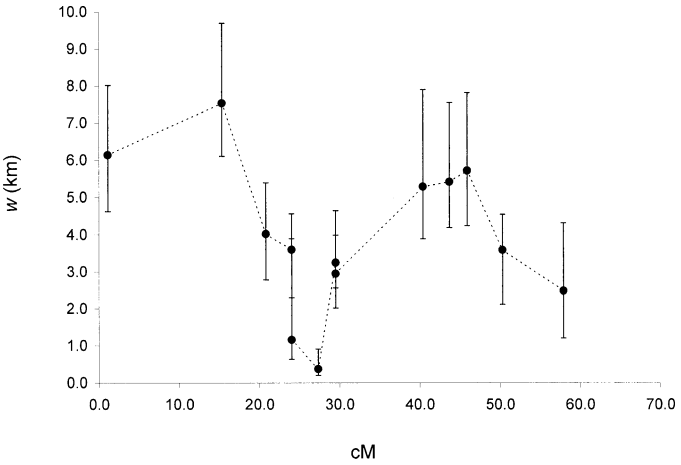

FIG. 4. Scatterplot of cline width (km) versus genetic position (cM) for 13 X-linked loci.

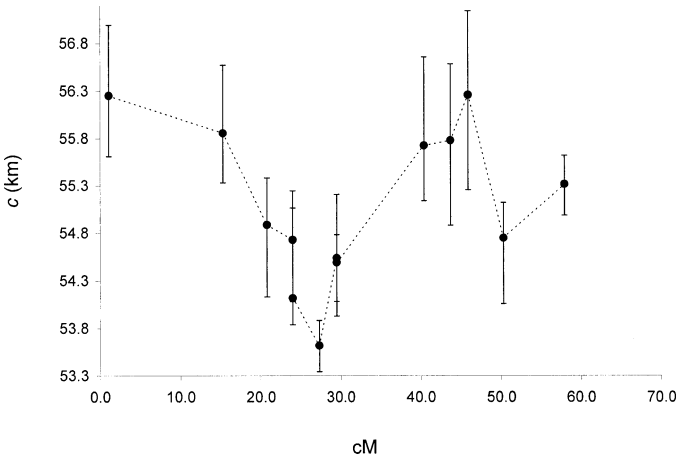

FIG. 5. Scatterplot of cline center (km) versus genetic position (cM). Distances run from west to east along the hybrid zone transect.

trogression. However, there is no relationship between cline shape and local recombination rate or gene density for markers on the X chromosome ($P > 0.05$ in all comparisons). This pattern is illustrated by the locus showing the least introgression, which is situated in an environment with a recombination rate and gene density typical for the X chromosome.

Linkage Disequilibrium

Analyses of linkage disequilibrium yield clear signs of nonrandom associations between pairs of loci in the center of the hybrid zone (Table 5). Fifty-three of 78 locus pairs exhibit linkage disequilibrium significant at the $P = 0.05$ level, and 34 of these comparisons remain significant after correcting for the performance of multiple tests. Most linkage disequilibrium estimates are positive, indicating that alleles from the same species tend to cosegregate. Because linkage disequilibrium decays as a function of recombinational distance, we might expect that pairs of loci more closely situated on the genetic map should show increased evidence of nonrandom association. This prediction is supported by the data: levels of linkage disequilibrium are negatively correlated with distances (cM) between loci (D, Spearman's $\rho = -0.50$; $P < 0.0001$; D', $\rho = -0.54$; $P < 0.0001$; Fig. 7). Additionally, locus pairs contained in the Bonferroni-corrected significant set have reduced genetic distances relative to other locus pairs (significant set average = 12.2 cM; average, nonsignificant set average = 22.9 cM; $P < 0.0001$; Mann-Whitney U). However, some pairs of loci separated by large recombinational distances also show linkage disequilibrium (e.g., *Emd* and *Glra2*; $D = 0.095$; $D' = 0.64$; distance = 33.9 cM).

DISCUSSION

We conducted a detailed survey of gene flow across the X chromosome in the European hybrid zone between *M. domesticus* and *M. musculus* and found clear variation in patterns of introgression across the chromosome. We identified multiple loci with unusual characteristics of gene flow, including a candidate region for reproductive isolation and a candidate region for adaptive introgression. Comparison of cline shapes across loci also suggested that the exchange of X-linked genes between *M. domesticus* and *M. musculus* is not symmetrical, indicating that interactions between the *M. domesticus* X chromosome and the *M. musculus* autosomes may be disrupted in hybrids.

Maintenance of the Hybrid Zone

Patterns of gene flow at 13 X-linked loci point to the conclusion that this hybrid zone between *M. domesticus* and *M. musculus* is primarily maintained by a balance between migration into the center of the zone and selection against hybrids. As reported in previous studies of this hybrid zone

TABLE 3. Known genes in a window encompassing 1 Mb on either side of the *Pola1* locus.

Gene name	Gene description	Function	Position (bp)
Arx	aristaless related homeobox gene (*Drosophila*)	transcription factor, development of forebrain and testes	78100519
Pola1	polymerase (DNA directed), alpha	replication	78118764
Pdk3	pyruvate dehydrogenase kinase, isoenzyme 3	pyruvate metabolism	78578614
Zkx	zinc finger protein	transcription factor, expressed in gonads	78888629
Eif2s3x	eukaryotic translation initiation factor 2, subunit 3, structural gene	translation	79004497
Maged1	melanoma antigen, family D	transcription factor, expressed in testes and placenta or cancerous cells	79349471

TABLE 4. *P*-values for likelihood ratio tests for differences in six-parameter cline shape. Columns indicate source of data used in likelihood ratio tests. Rows indicate source of parameter estimates used in likelihood ratio tests. Asterisks indicate that cline shapes are different at the Bonferroni-adjusted significance level ($P = 0.0003$).

	Foxp3	*Sep6b*	*Nt*	*Fmr1*	*Emd*	*Pola1*	*Xist*	*DXMit18.2*	*Pou3f4*	*Btk2*	*Plp*	*Trrp5*	*Glra2*
Foxp3		0.058	<0.001*	<0.001*	0.021	<0.001*	<0.001*	<0.001*	0.293	0.019	<0.001*	<0.001*	0.017
Sep6b	0.106		<0.001	<0.001*	0.370	<0.001*	<0.001*	<0.001*	0.503	0.015	<0.001*	0.002	<0.001*
Nt	<0.001*	<0.001		0.983	0.901	0.001	<0.001*	<0.001*	0.057	0.148	<0.001*	0.618	0.002
Fmr1	<0.001*	<0.001*	0.433		0.999	<0.001*	<0.001*	0.003	<0.001*	<0.001	<0.001*	0.650	0.004
Emd	<0.001*	<0.001*	0.083	0.114		<0.001*	<0.001*	0.002	0.004	<0.001	<0.001*	0.502	0.001
Pola1	<0.001*	<0.001*	0.008	0.006	0.644		<0.001*	<0.001*	0.002	0.003	<0.001*	0.084	<0.001*
Xist	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*		<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*
DXMit18.2	<0.001*	<0.001*	<0.001*	<0.001*	0.999	<0.001*	<0.001*		<0.001*	<0.001*	<0.001*	0.465	0.001
Pou3f4	<0.001*	0.024	0.017	0.002	0.021	<0.001*	<0.001*	<0.001*		0.558	<0.001*	0.010	<0.001*
Btk2	<0.001*	0.009	0.142	0.021	0.007	<0.001*	<0.001*	<0.001*	0.701		<0.001*	0.003	<0.001*
Plp	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*		<0.001*	<0.001*
Trrp5	<0.001*	<0.001*	0.522	0.640	0.999	0.011	<0.001*	0.002	0.006	<0.001	<0.001*		<0.001*
Glra2	0.005	<0.001*	0.002	<0.001	0.999	<0.001*	<0.001*	0.002	0.002	<0.001*	<0.001*	0.169	

TABLE 5. Pairwise estimates of linkage disequilibrium. *D* and *D′* values are above and below the diagonal, respectively. Statistical significance is presented only for *D*.

	Foxp3	*Sep6b*	*Nt*	*Fmr1*	*Emd*	*Pola1*	*Xist*	*DXMit18.2*	*Pou3f4*	*Btk2*	*Plp*	*Trrp5*	*Glra2*
Foxp3		0.12[2]	−0.08[1]	0.00	0.00	0.01	−0.04[1]	−0.05[1]	0.02	−0.02	−0.04	−0.03	−0.01
Sep6b	0.53		0.02	0.10[1]	0.09[1]	0.09[1]	0.04	0.04	−0.01	−0.05	−0.03	0.04	0.10[1]
Nt	−0.42	0.08		0.12[2]	0.11[2]	0.10[2]	0.08[2]	0.10[2]	0.01	0.08[1]	0.08[1]	0.08[1]	0.08[1]
Fmr1	−0.03	0.76	0.69		0.18[2]	0.17[2]	0.08[2]	0.09[2]	0.09[1]	0.03	0.08[1]	0.12[2]	0.10[2]
Emd	−0.04	0.67	0.69	0.96		0.18[2]	0.09[2]	0.09[2]	0.10[2]	0.04	0.07[1]	0.13[2]	0.10[2]
Pola1	0.04	0.60	0.53	0.90	0.99		0.11[2]	0.09[2]	0.08[1]	0.03	0.06[1]	0.12[2]	0.12[2]
Xist	−0.46	0.45	0.67	0.61	0.64	0.86		0.10[2]	0.04	0.03	0.03	0.10[2]	0.08[2]
DXMit18.2	−0.45	0.39	0.78	0.57	0.59	0.58	0.70		0.05[1]	0.08[2]	0.06[1]	0.11[2]	0.07[2]
Pou3f4	0.09	−0.04	0.06	0.67	0.78	0.55	0.41	0.46		0.18[2]	0.18[2]	0.12[2]	0.01
Btk2	−0.13	−0.23	0.50	0.26	0.32	0.20	0.32	0.86	0.78		0.19[2]	0.09[1]	−0.03
Plp	−0.29	−0.21	0.63	0.84	0.81	0.57	0.49	0.87	0.99	0.98		0.09[2]	0.01
Trrp5	−0.20	0.26	0.45	0.62	0.67	0.60	0.72	0.70	0.89	0.74	0.99		0.10[2]
Glra2	−0.07	0.48	0.37	0.65	0.64	0.69	0.72	0.59	0.05	−0.15	0.07	0.64	

[1] Likelihood ratio test is significant at $P = 0.05$ level.
[2] Likelihood ratio test is significant at $P = 0.0006$ (Bonferroni-adjusted) level.

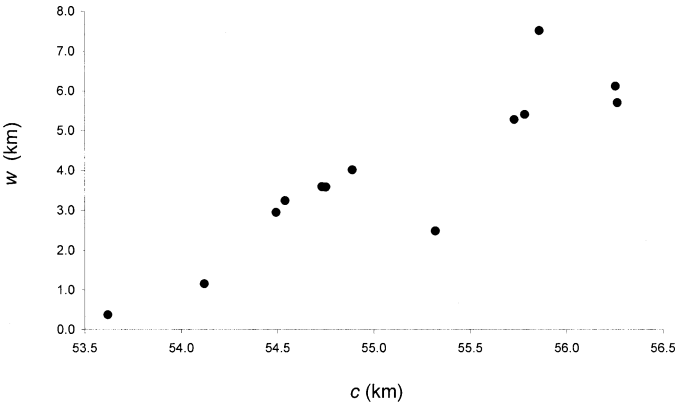

FIG. 6. Scatterplot of cline width (km) versus cline center (km).

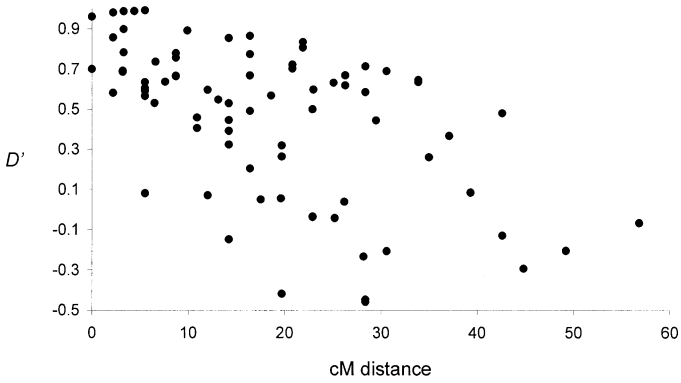

FIG. 7. Scatterplot of D' versus distance in cM.

(Hunt and Selander 1973; Vanlerberghe et al. 1986; Tucker et al. 1992), allele frequency clines at most loci are very steep, implying restricted exchange of the X chromosome between these species. Additionally, we detected significant linkage disequilibrium at 44% (34 of 78) of all two-locus combinations on the X chromosome. Some of the loci showing nonrandom associations are separated by large recombinational distances, an observation that seems unlikely if the hybrid zone is maintained by a simple environmental gradient. Moreover, both *M. domesticus* and *M. musculus* are commensal with humans and live in seemingly identical habitat; the hybrid zone is not associated with any known ecotone.

Evidence for the Effects of Linkage in Patterns of Introgression

Because genetic and physical maps are unavailable for most species that form hybrid zones in nature, the effects of genomic location on marker introgression have rarely been evaluated. A notable exception is the study of Rieseberg et al. (1999). In addition to identifying specific genomic regions potentially involved in reproductive isolation (as well as some regions introgressing more readily than expected), this research uncovered stronger associations (linkage disequilibria) among adjacent loci than between unlinked markers. These associations decayed with genetic map distance, a predicted signature of recombination in the hybrid zone.

The chromosomal arrangement of patterns of introgression in our study provides a similar illustration of the effects of genetic linkage on gene flow in the hybrid zone between *M. domesticus* and *M. musculus*. Adjacent markers show significant autocorrelation in cline widths at a lag of one locus (Spearman's $\rho = 0.60$; $P = 0.04$). Furthermore, tightly linked loci exhibit similar patterns: the two pairs of loci that map to 24.0 cM and 29.5 cM show very similar cline widths and cline centers (Table 2; Figs. 2–4). Finally, more closely linked loci exhibit higher levels of linkage disequilibrium in the center of the hybrid zone. This general correspondence between inferred gene flow and chromosomal position suggests that recombination has played a role in structuring patterns of introgression, providing support for this interlocus approach.

Candidate Region for Reproductive Isolation

Because cline width measures the extent of gene flow in the geographic portion of the zone in which hybrids are most likely to occur (the center) and because of the theoretical relation between cline width and selection against hybrids (balanced by dispersal; Barton and Gale 1993), this parameter provides the most direct information about reproductive isolation. By using a six-parameter model, we have attempted to control for the effects of variation in other gene flow parameters on the estimation of cline width (Szymura and Barton 1986; Porter et al. 1997). The *Pola1* locus, mapping to 27.3 cM, shows an unusually low cline width for the X chromosome. Additionally, the interval including the four loci surrounding *Pola1*, spanning 5.5 cM (24.0–29.5 cM) and 35.7 Mb (52.9–88.6Mb), displays a chromosomal valley in cline width (Fig. 4). In the center of the hybrid zone, *Pola1* also displays strong linkage disequilibrium with the four other loci in this valley, *Fmr1, Emd, Xist,* and *DXMit18.2*. These results suggest that *Pola1* lies in a candidate region for reproductive isolation.

Two additional lines of evidence support the conclusion that this region of the X chromosome has played an important role in the development of reproductive barriers between *M. domesticus* and *M. musculus*. Dod et al. (1993) surveyed three X-linked loci, *Hprt* (a diagnostic allozyme marker), *DXPas1*, and *DXPas2* (two anonymous, diagnostic RFLP markers) across a Danish transect of the hybrid zone. Although sequences and precise positions for the DNA markers are not available, Dod et al. (1993) note that the locus with the most restricted gene flow, *DXPas2*, is located a few cM distal to *Xist*, a gene that has been mapped in mice. This places *DXPas2* about 4 cM away from the locus showing the most reduced introgression in our study. Although these results may reflect the effects of different genes on reproductive isolation, this broad concordance in patterns of gene flow across independent transects of the hybrid zone points toward the center of the X chromosome as a clear candidate region for speciation in house mice.

One of the benefits of an approach that uses differential patterns of introgression to identify genomic regions involved in reproductive isolation is that it does not require knowledge of the phenotypes contributing to speciation. Nevertheless, information that links genomic regions experiencing reduced introgression to phenotypes related to reproductive isolation

can provide corroborating evidence that the restricted gene flow is connected to the speciation process (Rieseberg and Buerkle 2002). Recently, Forejt and colleagues (Storchova et al. 2004) introgressed individual genomic segments between the PWD/Ph laboratory strain (which is of *M. musculus* origin) and C57BL/6J (which is primarily of *M. domesticus* origin). Two results are noteworthy when viewed in light of our data. First, placement of the PWD/Ph X chromosome on to the genomic background of C57BL/6J causes male sterility, further implicating the X chromosome as significant in the genetics of reproductive isolation between *M. domesticus* and *M. musculus*. Second, quantitative trait loci (QTL) associated with measures of hybrid male sterility, including the number of offspring and sperm head morphology, map to the same chromosomal region we have identified as potentially important in the genetics of speciation.

The availability of the mouse genome sequence allows us to determine the gene content of the chromosomal region showing reduced introgression. A list of confirmed genes with functional information available in a window of 1 Mb on either side of the *Pola1* locus is provided in Table 3. In addition to these genes, 14 predicted genes of unknown function are located in this region. Based on their location, these loci represent candidate genes for reproductive isolation between *M. domesticus* and *M. musculus*.

In a recent study, we used two criteria to identify additional candidate genes for reproductive isolation (Payseur and Nachman 2004). First, because laboratory crosses suggest that *M. domesticus* and *M. musculus* may be isolated by hybrid male sterility, we surveyed the region of reduced introgression for genes expressed in the male germ line. Second, we looked for genes with high rates of protein evolution (in comparison with rat) in the region of reduced introgression, reasoning that targets of positive selection may often be involved in reproductive isolation (Ting et al. 1998; Barbash et al. 2003; Presgraves et al. 2003). We found three genes expressed solely in the male germ line (*Tktl1, Halapx,* and *Tex11*) and four genes with relatively high rates of protein evolution (*Pbsn, Pet2, Mm.21705* [Unigene identifier], *ENSMUSG00000050332.1* [Ensembl identifier]) that mapped to the general chromosomal region showing reduced introgression (Payseur and Nachman 2004). The identification of these candidate genes is accompanied by two caveats. First, all of these genes are more than 1 Mb from *Pola1* and some are more than 5 Mb from *Pola1*. Second, *M. domesticus* and *M. musculus* may be isolated through mechanisms other than hybrid sterility in nature, suggesting that different criteria may be more appropriate for the selection of candidate genes.

Candidate Locus for Adaptive Introgression

In addition to locating candidate regions for reproductive isolation, patterns of differential introgression can identify adaptive gene flow by uncovering genomic regions that mix between species at an unusually high rate. Comparison of allele frequency clines in Figure 3 identifies *Xist* as one such locus. In particular, *Xist* is polymorphic in an unusual number of populations on the *M. musculus* side of the hybrid zone (this result is quantitatively described by the very slow decay in *M. domesticus* allele frequency (θ_R) in *M. musculus* territory; Table 2). This result raises the intriguing possibility that *M. domesticus* alleles at a gene mapping to the *Xist* region may outcompete *M. musculus* alleles on a primarily *M. musculus* genomic background. Adaptive introgression could, in principle, be symmetrical in this case, but less intensive sampling on the *M. domesticus* side of the zone makes this proposition difficult to test. The causative locus does appear to be tightly linked: the introgression patterns displayed by *Xist* are not seen at *Pola1* (2.2 cM proximal) or at *DXMit18.2* (identical genetic map position; only 167 kb distal).

Fitting our data to the six-parameter model allows us to separate evolutionary forces acting in the center of the hybrid zone from those acting outside the center. Patterns of introgression at *Xist* track those of closely linked loci in the center of the zone: *Xist, Pola1,* and *DXMit18.2* have similar cline centers and cline widths. Theory predicts that associations between unlinked loci will often be strong in the center of a hybrid zone, due to the continual input of genotypes from populations with different allele frequencies and selection against hybrids (Barton 1983). Clearly, this effect will be exacerbated for linked loci. Outside the center of the hybrid zone, the degree of allele frequency differentiation and the frequency of hybridization are smaller, resulting in a relaxation of interlocus associations (Barton 1983). Hence, the parameters measured in the center of the hybrid zone reflect genomewide forces and measure introgression on a more hybrid genetic background (a more relevant arena for mapping genomic regions associated with reproductive isolation), while the parameters in the tails of hybrid zone clines indicate more locus-specific forces measured on a backdrop of purer species genomes. *Xist* provides a nice example of this logic because this locus harbors *M. domesticus* alleles in a number of individuals who carry *M. musculus* alleles at the remaining 12 X-linked loci. *Mus domesticus Xist* alleles that reach *M. musculus* populations must first pass through the center of the zone. As a result, *M. domesticus* alleles found on the *M. musculus* side of the zone are likely to be older and have experienced more recombination than those found in the center of the zone. This temporal pattern of gene flow provides another mechanism for decoupling this locus from other loci on the X chromosome.

An alternative explanation for the unusual pattern of introgression at *Xist* is that a problem in the RFLP assay causes the genotypes to be incorrectly assessed. To rule out this possibility, we sequenced the RFLP site from *Xist* PCR products for a number of individuals scored as *M. domesticus* on the *M. musculus* side of the zone. In all cases, RFLP patterns matched the sequence at the restriction site. Additionally, the sequences revealed that these individuals carry *M. domesticus* alleles at other sites in the PCR fragment. Hence, inferences about introgression at *Xist* from RFLP patterns appear to be robust.

Another potential explanation for the *Xist* results is that the scored difference at this locus is not diagnostic of *M. domesticus* and *M. musculus* and merely reflects patterns of neutral gene flow at a site retaining ancestral polymorphism in one or both species. Limited availability of samples from outside the hybrid zone makes this possibility difficult to eliminate. However, three lines of evidence argue against this interpretation. First, *Xist* was not polymorphic in our initial

sample of *M. domesticus* and *M. musculus* sampled from outside the hybrid zone. Second, if the observed difference at this locus is polymorphic in one or both species just outside the hybrid zone and patterns of introgression result mainly from neutral diffusion of this polymorphism, we might expect this variation to be visible in localities on the *M. domesticus* side of the transect as well. However, the long tail of introgression appears to be restricted to the *M. musculus* side of the hybrid zone (although the *M. domesticus* side has not been sampled as thoroughly). Third, if the unusual tail of the *Xist* cline primarily reflects neutral introgression, similar patterns should be found in markers from across the genome. Although the introgression of *M. domesticus* alleles can be seen at autosomal markers genotyped in the same transect (Tucker et al. 1992), the extent of gene flow appears to be much greater at *Xist.* Finally, it is possible that the assayed SNP does not map to the inferred genomic location due to errors in assembly of the draft genome sequence. Although we cannot definitively rule out this possibility, it seems clear that the SNP is X-linked: heterozygous genotypes are only observed in females.

Overall, the data seem most consistent with the notion that alleles at this locus are experiencing positive selection on a heterospecific genomic background. Although the pattern is less dramatic, the *Plp* locus also displays suggestive evidence of adaptive introgression. Empirical studies, particularly in plants, indicate that hybridization can provide a source of genetic variation for adaptation (Arnold 1997; Rieseberg et al. 2003). Theory indicates that advantageous alleles responding to environmental selection can escape barriers to gene flow caused by heterozygote disadvantage at the same locus in the center of a hybrid zone via drift (Pialek and Barton 1997). A situation more directly comparable to our study, with one locus showing a substantial barrier to gene flow in the center of the hybrid zone and a linked locus showing signs of adaptive introgression just outside the center, has yet to be modeled. Given the apparent permeability of the genome (Rieseberg et al. 1999; Martinsen et al. 2001), theoretical results in this area would be useful.

Asymmetrical Introgression

There is a strong, positive correlation between cline width and cline center across the X chromosome, indicating that loci with reduced introgression have clines shifted further into the *M. domesticus* side of the hybrid zone. Using data from this transect, Tucker et al. (1992) noted greater interlocus variation in gene flow on the *M. musculus* side of the hybrid zone. Specifically, *M. domesticus* alleles were sometimes found in *M. musculus* localities at autosomal loci but not at markers on the X or Y chromosome. A similar discrepancy between sex chromosomes and autosomes was also observed in the Danish transect (Vanlerberghe et al. 1988; Dod et al. 1993). Taken together, these results suggest that the extent of gene flow of particular X-linked regions may be determined primarily by the fitness of *M. domesticus* alleles on a hybrid or *M. musculus* genomic background. However, the small number of sampled localities in the western part of the transect challenges the measurement of asymmetrical introgression, suggesting caution in this interpretation. Additionally, differences in the stringency of mate choice may affect observed patterns of gene flow between *M. domesticus* and *M. musculus*. For example, behavioral studies of mice from the edges of the Danish transect suggest that *M. musculus* females are more particular than *M. domesticus* females (Smadja and Ganem 2002).

Introgression and Recombination Rate

The degree of introgression is not correlated with local recombination rate or gene density across the X chromosome. The restriction of our test to the X chromosome limits the generality of our conclusions. Nevertheless, this pattern argues against the idea that reproductive isolation between *M. domesticus* and *M. musculus* is due to many loci situated throughout the genome, with the probability of detection just a function of the number of genes linked to a surveyed locus. Viewed in combination with the incompleteness of reproductive isolation between these species and their recent divergence time (350,000 years; She et al. 1990), these results suggest that the isolation is relatively young and that we may expect to find a small to moderate number of genes associated with speciation in this case. The collection and analysis of comparable data from the autosomes will allow a more rigorous assessment of this idea.

Predictions and Future Research

Our results provide a number of testable predictions regarding the genetic basis of reproductive isolation between *M. domesticus* and *M. musculus*. First, we might expect a gene or genes in the region with strongly reduced introgression to show unusual patterns. For example, such genes might display aberrant expression patterns in hybrids between *M. domesticus* and *M. musculus*. Additionally, we predict that studies designed to map the genetic basis of postzygotic or prezygotic isolation between these two species will detect loci of moderate to large effect in this region of the X chromosome. Our results also suggest that surveys of nucleotide variaton at the *Xist* locus in mouse populations from just outside the hybrid zone may reveal evidence of positive, directional selection.

Acknowledgments

We thank R. Sage for generously donating mouse DNAs. P. Tucker and K. Teeter provided useful input about the hybrid zone. We appreciate A. Porter's dedicated assistance in using his ClineFit program. We thank J. Forejt and J. Pialek for permission to cite unpublished research. R. Harrison, N. Barton, J. Pialek, P. Tucker, K. Teeter, J. Good, and one anonymous reviewer provided useful comments on the manuscript. This work was supported by National Science Foundation Integrative Graduate Education and Research Training (IGERT) fellowships in Genomics and in Biology, Mathematics, and Physics to BAP and by grants from the National Science Foundation to MWN.

Literature Cited

Ammerman, A. J., and L. L. Cavalli-Sforza. 1984. The Neolithic transition and the genetics of populations in Europe. Princeton Univ. Press, Princeton, NJ.

Arnold, M. L. 1997. Natural hybridization and evolution. Oxford Univ. Press, New York.
Barbash, D. A., D. F. Siino, A. M. Tarone, and J. Roote. 2003. A rapidly evolving MYB-related protein causes species isolation in *Drosophila*. Proc. Natl. Acad. Sci. USA 100:5302–5307.
Barton, N. H. 1983. Multilocus clines. Evolution 37:454–471.
Barton, N. H., and K. S. Gale. 1993. Genetic analysis of hybrid zones. Pp. 13–45 *in* R. G. Harrison, ed. Hybrid zones and the evolutionary process. Oxford Univ. Press, Oxford, U.K.
Barton, N. H., and G. M. Hewitt. 1985. Analysis of hybrid zones. Annu. Rev. Ecol. Syst. 16:113–148.
Bateson, W. 1909. Heredity and variation in modern lights. Pp. 85–101 *in* A. C. Seward, ed. Darwin and modern science. Cambridge Univ. Press, Cambridge, U.K.
Bazykin, A. D. 1969. Hypothetical mechanism of speciation. Evolution 23:685–687.
Bishop, C. E., P. Boursot, B. Baron, F. Bonhomme, and D. Hatat. 1985. Most classical *Mus musculus domesticus* laboratory mouse strains carry a *Mus musculus musculus* Y chromosome. Nature 325:70–72.
Boursot, P., J.-C. Auffray, J. Britton-Davidian, and F. Bonhomme. 1993. The evolution of house mice. Annu. Rev. Ecol. Syst. 24: 119–152.
Christie, P., and M. R. Macnair. 1984. Complementary lethal factors in two North American populations of the yellow monkeyflower. J. Hered. 75:510–511.
Coyne, J. A., and H. A. Orr. 1989. Two rules of speciation. Pp. 180–207 *in* D. Otte and J. Endler, eds. Speciation and its consequences. Sinauer Associates, Sunderland, MA.
Crow, J. F. 1942. Cross fertility and isolating mechanisms in the *Drosophila mulleri* group. Univ. Texas Pub. 4228:53–67.
Dietrich, W. F., J. Miller, R. Steen, M. A. Merchant, D. Damron-Boles, Z. Husain, R. Dredge, M. J. Daly, K. A. Ingalls, T. J. O'Connor, C. A. Evans, M. M. DeAngelis, D. M. Levinson, L. Kruglyak, N. Goodman, N. G. Copeland, N. A. Jenkins, T. L. Hawkins, L. Stein, D. C. Page, and E. S. Lander. 1996. A comprehensive genetic map of the mouse genome. Nature 380: 149–152.
Dobzhansky, T. 1936. Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. Genetics 21:113–135.
Dod, B., L. S. Jermiin, P. Boursot, V. H. Chapman, J. Tonnes-Nielsen, and F. Bonhomme. 1993. Counterselection on sex chromosomes in the *Mus musculus* European hybrid zone. J. Evol. Biol. 6:529–546.
Edwards, A. W. F. 1992. Likelihood. Johns Hopkins Univ. Press, Baltimore, MD.
Endler, J. A. 1977. Geographic variation, speciation, and clines. Princeton Univ. Press, Princeton, NJ.
Ferris, S. D., R. D. Sage, and A. C. Wilson. 1982. Evidence from mtDNA sequences that common laboratory strains of inbred mice are descended from a single female. Nature 295:163–165.
Forejt, J. 1996. Hybrid sterility in the mouse. Trends Genet. 12: 412–417.
Forejt, J., and P. Ivanyi. 1975. Genetic studies on male sterility of hybrids between laboratory and wild mice (*Mus musculus* L.). Genet. Res. 24:189–206.
Grula, J. W., and O. R. Taylor. 1982. Some characteristics of hybrids derived from the sulfur butterflies, *Colias eurytheme* and *Colias philodice*: phenotypic effects of the X chromosome. Evolution 34:673–687.
Hagen, R. H. 1990. Population structure and host use in hybridizing subspecies of *Papilio glaucus* (Lepidoptera: Papilionidae). Evolution 44:1914–1930.
Haldane, J. B. S. 1922. Sex-ratio and unisexual sterility in hybrid animals. J. Genet. 12:101–109.
Harrison, R. G. 1990. Hybrid zones: windows on evolutionary process. Oxf. Surv. Evol. Biol. 7:69–128.
Hill, W. G. 1974. Estimation of linkage disequilibrium in randomly mating populations. Heredity 33:229–239.
Hollingshead, L. 1930. A lethal factor in *Crepis* effective only in interspecific hybrids. Genetics 15:114–140.
Hunt, W. G., and R. K. Selander. 1973. Biochemical genetics of hybridisation in European house mice. Heredity 31:11–33.
Jiggins, C. D., M. Linares, R. E. Naisbit, C. Salazar, Z. H. Yang, and J. Mallet. 2001. Sex-linked hybrid sterility in a butterfly. Evolution 55:1631–1638.
Laurie, C. C. 1997. The weaker sex is heterogametic: 75 years of Haldane's rule. Genetics 147:937–951.
Lewontin, R. C. 1964. The interaction of selection and linkage. I. General considerations; heterotic models. Genetics 49:49–67.
Lindblad-Toh, K., E. Winchester, M. J. Daly, D. G. Wang, J. N. Hirschhorn, J. P. Laviolette, K. Ardlie, D. E. Reich, E. Robinson, P. Sklar, N. Shah, D. Thomas, J. B. Fan, T. Gingeras, J. Warrington, N. Patil, T. J. Hudson, and E. S. Lander. 2000. Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. Nat. Genet. 24:381–386.
Macholan, M. 1996. Morphometric analysis of European house mice. Acta Theriol. 41:255–275.
Malitschek, B., D. Fornzler, and M. Schartl. 1995. Melanoma formation in *Xiphophorus*: a model system for the role of receptor tyrosine kinases in tumorigenesis. Bioessays 17:1017–1023.
Martinsen, G. D., T. G. Whitham, R. J. Turek, and P. Keim. 2001. Hybrid populations selectively filter gene introgression between species. Evolution 55:1325–1335.
Moulia, C., N. LeBrun, J. Dallas, A. Orth, and F. Renaud. 1993. Experimental evidence of genetic determinism in high susceptibility to intestinal pinworm infection in mice: a hybrid zone model. Parasitology 106:387–393.
Moulia, C., N. LeBrun, C. Loubes, R. Marin, and F. Renaud. 1995. Hybrid vigor in parasites of interspecific crosses between two mice species. Heredity 74:48–52.
Mouse Genome Sequencing Consortium. 2002. Initial sequencing and comparative analysis of the mouse genome. Nature 420: 520–562.
Muller, H. J. 1940. Bearing of the *Drosophila* work on systematics. Pp. 185–268 *in* J. S. Huxley, ed. The new systematics. Clarendon Press, Oxford, U.K.
———. 1942. Isolating mechanisms, evolution, and temperature. Biol. Symp. 6:71–125.
Munclinger, P., E. Bozikova, M. Sugerkova, J. Pialek, and M. Macholan. 2002. Genetic variation in house mice (*Mus*, Muridae, Rodentia) from the Czech and Slovak Republics. Folia Zool. 51: 81–92.
Noor, M. A. F. 2003. Genes to make new species. Nature 423: 699–700.
Orr, H. A. 1989. Genetics of sterility in hybrids between two subspecies of *Drosophila*. Evolution 43:180–189.
———. 1995. The population genetics of speciation: the evolution of hybrid incompatibilities. Genetics 139:1805–1813.
———. 1996. Dobzhansky, Bateson, and the genetics of speciation. Genetics 144:1331–1335.
———. 1997. Haldane's rule. Ann. Rev. Ecol. Syst. 28:195–218.
Orr, H. A., and J. A. Coyne. 1989. The genetics of postzygotic isolation in the *Drosophila virilis* group. Genetics 121:527–537.
Pantazidis, A. C., and E. Zouros. 1988. Location of an autosomal factor causing sterility in *Drosophila mojavensis* males carrying the *Drosophila arizonensis* Y chromosome. Heredity 60: 299–304.
Payseur, B. A., and M. W. Nachman. 2004. The genomics of speciation: investigating the molecular correlates of X chromosome introgression across a hybrid zone between *Mus domesticus* and *Mus musculus*. Biol. J. Linn. Soc., *In press.*
Perez, D. E., and C.-I. Wu. 1995. Further characterization of the *Odysseus* locus of hybrid sterility in *Drosophila*: one gene is not enough. Genetics 140:201–206.
Pialek, J., and N. H. Barton. 1997. The spread of an advantageous allele across a barrier: the effects of random drift and selection against heterozygotes. Genetics 145:493–504.
Porter, A. H., R. Wenger, H. Geiger, A. Scholl, and A. M. Shapiro. 1997. The *Pontia daplidice-edusa* hybrid zone in northwestern Italy. Evolution 51:1561–1573.
Presgraves, D. C. 2003. A fine-scale genetic analysis of hybrid incompatibilities in *Drosophila*. Genetics 163:955–972.
Presgraves, D. C., L. Balagopalan, S. M. Abymayr, and H. A. Orr.

2003. Adaptive evolution drives divergence of a hybrid inviability gene between two species of *Drosophila*. Nature 423: 715–719.
Prowell, D. P. 1998. Sex linkage and speciation in Lepidoptera. Pp. 309–319 *in* D. J. Howard and S. H. Berlocher, eds. Endless forms: species and speciation. Oxford Univ. Press, Oxford, U.K.
Rieseberg, L. H., and A. C. Buerkle. 2002. Genetic mapping in hybrid zones. Am. Nat. 159:S37–S50.
Rieseberg, L. H., J. Whitton, and K. Gardner. 1999. Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. Genetics 152:713–727.
Rieseberg, L. H., O. Raymond, D. M. Rosenthal, Z. Lai, K. Livingstone, T. Nakazato, J. L. Durphy, A. E. Schwarzbach, L. A. Donovan, and C. Lexer. 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. Science 301: 1211–1216.
Sage, R. D., D. Heyneman, K.-C. Lim, and A. C. Wilson. 1986. Wormy mice in a hybrid zone. Nature 324:60–63.
Sage, R. D., W. R. Atchley, and E. Capanna. 1993. House mice as models in systematic biology. Syst. Biol. 42:523–561.
Schnell, G. D., and R. K. Selander. 1981. Environmental and morphological correlates of genetic variation in mammals. Pp. 60–69 *in* J. Joule and M. H. Smith, eds. Mammalian population genetics. Univ. of Georgia Press, Athens, GA.
Selander, R. K., and S. Y. Yang. 1969. Protein polymorphism and genic heterozgyosity in a wild population of the house mouse (*Mus musculus*). Genetics 63:653–657.
She, J. X., F. Bonhomme, P. Boursot, L. Thaler, and F. Catzeflis. 1990. Molecular phylogenies in the genus *Mus*: comparative analysis of electrophoretic, scnDNA hybridization, and mtDNA RFLP data. Biol. J. Linn. Soc. 41:83–103.
Slatkin, M. 1973. Gene flow and selection in a cline. Genetics 75: 733–756.
Smadja, C., and G. Ganem. 2002. Subspecies recognition in the house mouse: a study of two populations from the border of a hybrid zone. Behav. Ecol. 13:312–320.
Sperling, F. A. H. 1994. Sex-linked genes and species differences in Lepidoptera. Can. Entomol. 126:807–818.
Sperling, F. A. H., and J. R. Spence. 1991. Structure of an asymmetric hybrid zone between two water strider species (Hemiptera: Gerridae: Limnoporus). Evolution 45:1370–1383.
Storchova, R., S. Gregorova, D. Buckiova, V. Kyselova, P. Divina, and J. Forejt. 2004. Genetic analysis of X-linked hybrid sterility in the house mouse. Mamm. Genome 15:515–524.
Szymura, J. M., and N. H. Barton. 1986. Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina bombina* and *Bombina variegata*, near Cracow in southern Poland. Evolution 40: 1141–1159.
———. 1991. The genetic structure of the hybrid zone between fire-bellied toads *Bombina bombina* and *B. variegata*: comparisons between transects and loci. Evolution 45:237–261.
Tao, Y., S. Chen, D. L. Hartl, and C. C. Laurie. 2003. Genetic dissection of hybrid incompatibilities between *Drosophila simulans* and *D. mauritiana*. I. Differential accumulation of hybrid male sterility effects on the X and autosomes. Genetics 164: 1383–1397.
Ting, C. T., S. C. Tsaur, M. L. Wu, and C.-I. Wu. 1998. A rapidly evolving homeobox at the site of a hybrid sterility gene. Science 282:1501–1504.
True, J. R., B. S. Weir, and C. C. Laurie. 1996. A genome-wide survey of hybrid incompatibility factors by the introgression of marked segments of *Drosophila mauritiana* chromosomes into *Drosophila simulans*. Genetics 142:819–837.
Tucker, P. K., R. D. Sage, J. Warner, A. C. Wilson, and E. M. Eicher. 1992. Abrupt cline for sex chromosomes in a hybrid zone between two species of mice. Evolution 46:1146–1163.
Turelli, M., and H. A. Orr. 1995. The dominance theory of Haldane's rule. Genetics 140:389–402.
———. 2000. Dominance, epistasis and the genetics of postzygotic isolation. Genetics 154:1663–1679.
Ursin, E. 1952. Occurrence of voles, mice, and rats (Muridae) in Denmark, with a special note on a zone of intergradation between two subspecies of the house mouse (*Mus musculus* L.). Vid. Medd. Dansk. Naturhist. Foren. 114:217–244.
Vanlerberghe, F., B. Dod, P. Boursot, M. Bellis, and F. Bonhomme. 1986. Absence of Y-chromosome introgression across the hybrid zone between *Mus musculus domesticus* and *Mus musculus musculus*. Genet. Res. 48:191–197.
Vanlerberghe, F., P. Boursot, J. Catalan, S. Guerasimov, F. Bonhomme, B. Botev, and L. Thaler. 1988. Analyse genetique de la zone d'hybridation entre les deux sous-especes de souris *M. m. domesticus* et *M. m. musculus* en Bulgarie. Genome 30: 427–437.
Wade, C. M., E. J. Kulbokas, A. W. Kirby, M. C. Zody, J. C. Mullikin, E. S. Lander, K. Lindblad-Toh, and M. J. Daly. 2002. The mosaic structure of variation in the laboratory mouse genome. Nature 420:574–578.
Wittbrodt, J., D. Adam, B. Malitschek, W. Mqaueler, F. Raulf, A. Telling, S. M. Robertson, and M. Schartl. 1989. Novel putative receptor tyrosine kinase encoded by the melanoma-inducing Tu locus in *Xiphophorus*. Nature 341:415–421.
Wu, C.-I., and A. T. Beckenbach. 1983. Evidence for extensive genetic differentiation between the sex-ratio and the standard arrangement of *Drosophila pseudoobscura* and *D. persimilis* and identification of hybrid sterility factors. Genetics 105:71–86.
Yonekawa, H., K. Moriwaki, O. Gotoh, J. Watanabe, J.-I. Hayashi, N. Miyashita, M. L. Petras, and Y. Tagashira. 1980. Relationship between laboratory mice and subspecies *Mus musculus domesticus* based on restriction endonuclease cleavage patterns of mitochondrial DNA. Jpn. J. Genet. 55:289–296.
Zouros, E., K. Lofdahl, and P. A. Martin. 1988. Male hybrid sterility in *Drosophila*: interactions between autosomes and sex chromosomes in crosses of *D. mojavensis* and *D. arizonensis*. Evolution 42:1321–1331.

Corresponding Editor: R. Harrison