Patterns of DNA Variability at X-Linked Loci in *Mus domesticus*

Michael W. Nachman

Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721

Manuscript received March 7, 1997
Accepted for publication August 14, 1997

ABSTRACT

Introns of four X-linked genes (*Hprt, Plp, Glra2,* and *Amg*) were sequenced to provide an estimate of nucleotide diversity at nuclear genes within the house mouse and to test the neutral prediction that the ratio of intraspecific polymorphism to interspecific divergence is the same for different loci. *Hprt* and *Plp* lie in a region of the *X* chromosome that experiences relatively low recombination rates, while *Glra2* and *Amg* lie near the telomere of the *X* chromosome, a region that experiences higher recombination rates. A total of 6022 bases were sequenced in each of 10 *Mus domesticus* and one *M. caroli.* Average nucleotide diversity (π) for introns within *M. domesticus* was quite low (π = 0.078%). However, there was substantial variation in the level of heterozygosity among loci. The two telomeric loci, *Glra2* and *Amg,* had higher ratios of polymorphism to divergence than the two loci experiencing lower recombination rates. These results are consistent with the hypothesis that heterozygosity is reduced in regions with lower rates of recombination, although sampling of additional genes is needed to establish whether there is a general correlation between heterozygosity and recombination rate as in *Drosophila melanogaster.*

A central goal of population genetics is to understand the evolutionary forces that shape genetic variation in natural populations. If much of phenotypic evolution is guided by natural selection, then in principle it should be possible to detect evidence of selection in the DNA as well. Statistical analyses of DNA sequences may be used to detect selective forces that are presumably too weak to observe directly. Recent studies with Drosophila have used samples of DNA sequences within and between closely related species to look for evidence of natural selection at the molecular level (reviewed in AQUADRO 1993; BROOKFIELD and SHARP 1994; KREITMAN and AKASHI 1995; HUDSON 1996). The neutral theory (KIMURA 1968, 1983) serves as a useful null hypothesis in these studies because it makes a number of simple, testable predictions. For example, under neutrality, the ratio of polymorphism within a species to divergence between species is expected to be the same for different genes or gene regions (HUDSON *et al.* 1987). Differences from this neutral expectation may be due to balancing selection (HUDSON *et al.* 1987), directional selection (BEGUN and AQUADRO 1991; BERRY *et al.* 1991), or purifying selection (CHARLESWORTH *et al.* 1993). Although most population-level, DNA-based studies of selection have been conducted with Drosophila, recent work has also included studies on *Zea mays* (GAUT and CLEGG 1993a,b), *Escherichia coli* (GUTTMAN and DYKHUIZEN 1994), humans (HAMMER 1995; NACHMAN *et al.* 1996; HEY 1997), and house mice (NACHMAN and AQUADRO 1994; NACHMAN *et al.* 1994b).

One important result from *Drosophila melanogaster* is the finding of an overall positive correlation between levels of nucleotide diversity and rates of recombination for different genes (BEGUN and AQUADRO 1992). This is unexpected under a simple, neutral model, but may be explained by either selective sweeps of favorable mutations and associated genetic hitchhiking (MAYNARD-SMITH and HAIGH 1974; BEGUN and AQUADRO 1992) or background selection against deleterious mutations (CHARLESWORTH *et al.* 1993, 1995; CHARLESWORTH 1994; HUDSON 1994; HUDSON and KAPLAN 1994, 1995).

D. melanogaster provides an unusually good model system for studying patterns of polymorphism and divergence. Many genes have been characterized and sequenced, intron-exon boundaries are often known, and introns in Drosophila are typically smaller than mammalian introns. The physical and genetic maps are quite dense and well integrated; comparison of these maps for different regions of the genome provides a reasonably detailed picture of regional variation in recombination rate (LINDSLEY and ZIMM 1992). Because population sizes are large, average heterozygosity is moderately high, making it relatively easy to detect variation in heterozygosity among genomic regions. Finally, by using balancer chromosomes and isofemale lines, single alleles can be easily sampled. Because of the technical advantages associated with Drosophila, there have been few studies of polymorphism and divergence at nuclear genes in other species. This is unfortunate since the extension of population-level studies of DNA sequence variation to other species has the potential to reveal general patterns in population genetics. Despite its importance, remarkably few data exist on the average level of heterozygosity at nuclear genes in species outside of Drosophila.

Corresponding author: Michael W. Nachman, Department of Ecology and Evolutionary Biology, Biosciences West Bldg., University of Arizona, Tucson, AZ 85721. E-mail: nachman@ccit.arizona.edu

The present study was undertaken with the aim of extending population-level studies of nuclear gene variation to mice. The house mouse, *Mus domesticus*, is a logical mammalian model. As in Drosophila, many genes have been cloned, sequenced and mapped. The genetic map for the mouse is quite dense, primarily because of the inclusion of over 6000 $(CA)_n$ microsatellite markers (DIETRICH *et al.* 1992, 1994). By choosing *X*-linked genes in males, it is possible to circumvent the problem of sampling heterozygotes. The chief disadvantages of working with mice are that estimates of recombination rates are still imprecise and that average heterozygosity is expected to be low, given the presumably smaller effective population size of mice relative to Drosophila.

This article documents patterns of nucleotide polymorphism at four *X*-linked loci within *M. domesticus* and patterns of divergence in comparisons between *M. domesticus* and *M. caroli*. These taxa are believed to have diverged from each other ~2.5 million years ago (SHE *et al.* 1990). These data represent the first step in an on-going study to look for evidence of selection in patterns of nuclear DNA sequence variation within and between species of Mus. Here I show that the average level of nucleotide heterozygosity in *M. domesticus* is quite low, that there is significant variation in the level of heterozygosity among loci, and that ratios of polymorphism to divergence differ among some genes, contrary to neutral expectations.

MATERIALS AND METHODS

Samples and chromosome analysis: Ten male *M. domesticus* were wild caught from within their native range in Italy. Males only were chosen so that single alleles could be PCR-amplified from *X*-linked loci. Because *M. domesticus* consists of numerous cryptic chromosomal races (NACHMAN and SEARLE 1995), all mice were karyotyped as described previously (NACHMAN *et al.* 1994a) and only mice of the standard $2n = 40$ chromosomal race were included in this study. Collecting localities and specimen numbers are as follows: (1) Northern Italy ($N = 2$), Lombardia, Staffora Valley, Canova (specimens MWN 1036, 1037); (2) Central Italy ($N = 6$), Lazio, Cassino (specimens MWN 1103, 1114, 1120, 1126), 3.5 km SW of Sette Vene (specimen MWN 1083), Campania, Policastro (specimen MWN 1204); (3) Southern Italy ($N = 2$), Sicily, 6.0 km S of Milazzo (specimen MWN 1213), Sicily, Salina (specimen MWN 1248). Specimens were deposited in the collections of the Museum of Zoology at The University of Michigan. Genomic DNA from a single male *M. caroli* (strain 926) was purchased from the Jackson Laboratory DNA Resource.

DNA preparation: Genomic DNA from wild caught mice was prepared from frozen spleen or liver tissue following SAMBROOK *et al.* (1989) with modifications. Tissues were ground in liquid nitrogen and the resulting powder was suspended in 10 ml of extraction buffer (10 mM Tris pH 8.0, 100 mM EDTA, 0.5% SDS, 20 μg/ml RNAse) and incubated at 55° for 1 hr. Proteinase K (100 μg/ml) was added, and the solution was incubated for up to 12 hr at 55° and extracted three times with phenol/chloroform. DNA was ethanol precipitated, dried, and resuspended in T.E. (pH 8.0) to a final concentration of 500 μg/ml.

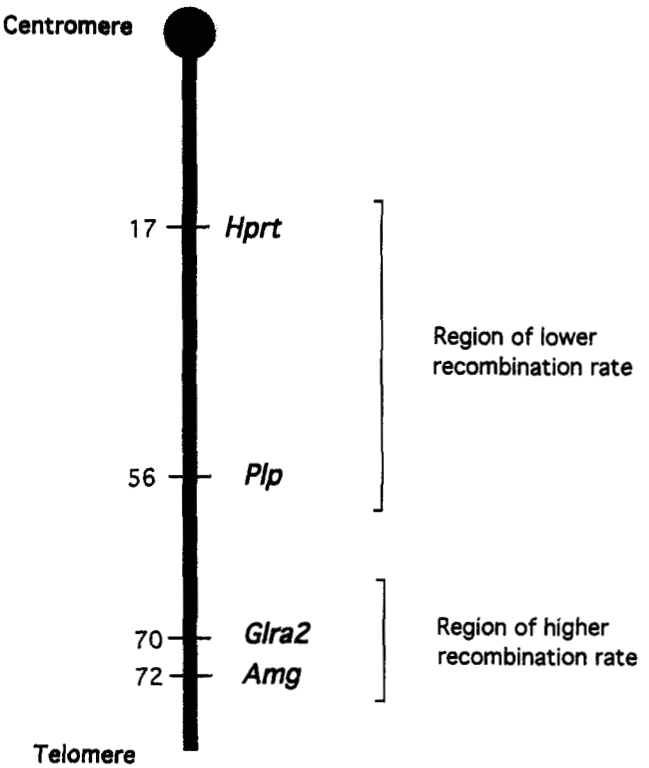

FIGURE 1.—Map of the *M. domesticus X* chromosome showing the genetic locations (cM) of the four genes in this study.

PCR amplification of X-linked loci: Introns and small portions of exons of four *X*-linked genes were amplified directly from genomic DNA. Loci were chosen to lie in regions experiencing different recombination rates on the *X* chromosome (NACHMAN and CHURCHILL 1996). Hypoxanthine phosphoribosyltransferase (*Hprt*) and myelin proteolipid protein (*Plp*) lie in the central portion of the mouse *X* and experience low recombination rates, and glycine receptor α_2 subunit (*Glra2*) and amelogenin (*Amg*) lie near the telomere and experience higher recombination rates (Figure 1). The amplified regions of each of these genes are shown in Figure 2 and amplification primers are listed in Table 1. Primers were designed to lie in conserved regions based on interspecific sequence comparisons. Primers in exons (or short stretches of intron sequence immediately adjacent to exons) were designed from published sequences for *Hprt* (MELTON *et al.* 1984), *Plp* (IKENAKA *et al.* 1988), *Glra2* (MATZENBACH *et al.* 1994), and *Amg* (SNEAD *et al.* 1985). Exon-intron boundaries and the approximate sizes of the introns were known for each of these genes, but intron sequence had not previously been generated except for short stretches of *Glra2* (MATZENBACH *et al.* 1994). The identity of amplified products was confirmed by comparison to published exon sequences.

DNA was amplified using PCR (SAIKI *et al.* 1986, 1988) in 100 μl reaction volumes with 40 cycles of 94° for 1 min, 55° for 1 min, 72° for 2 min. Taq polymerase (Perkin Elmer-Cetus) was used with conditions as specified by the supplier, and the reaction mixture was overlaid with mineral oil. Following the reaction, oil was removed using a chloroform extraction, and the double-stranded PCR product was precipitated with 33 μl ammonium acetate (10 M) and 133 μl 100% cold ethanol, washed once in 80% ethanol, and resuspended in 30 μl ddH_2O for direct double-stranded sequencing.

DNA sequencing: Direct sequencing of double-stranded PCR products was done using the dideoxy chain termination

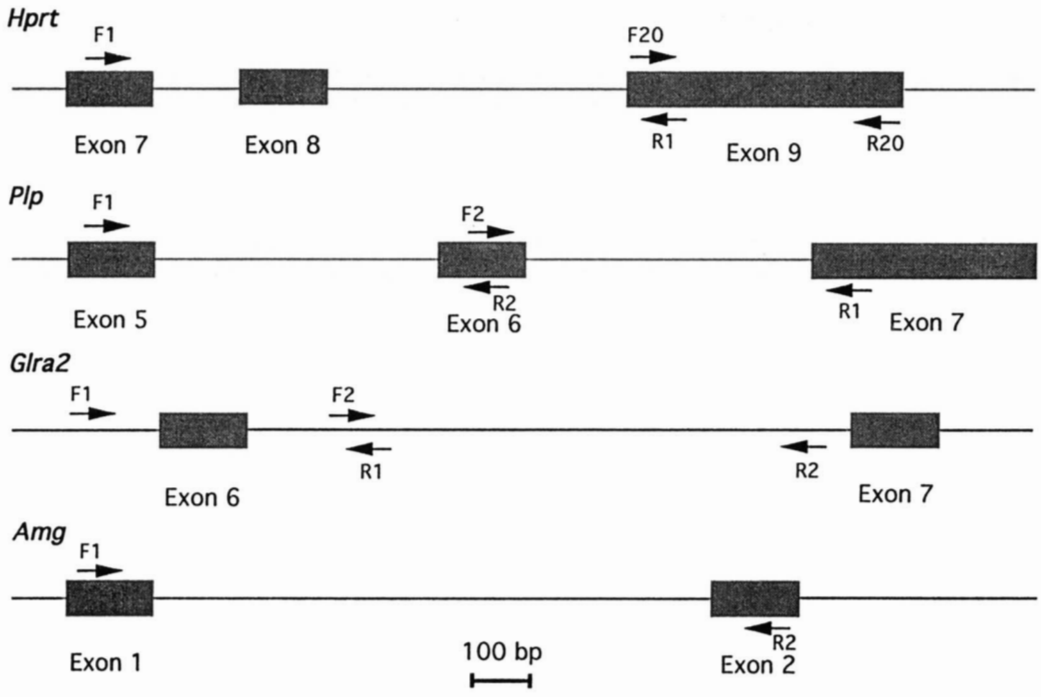

FIGURE 2.—Physical maps of portions of *Hprt*, *Plp*, *Glra2*, and *Amg*. The locations of PCR primers used in this study are indicated by arrows.

method (SANGER *et al.* 1977) with Sequenase 2 enzyme and kit (BRL) according to the protocol supplied by the manufacturer with slight modifications. The sequencing primer was annealed to template DNA (annealing step of Sequenase protocol) by heating at 100° for 3 min, freezing directly in dry ice/ethanol bath, and thawing in presence of enzyme and elongation reagents. Sequencing primers were 17-mers and were spaced approximately every 200 bp. The following numbers of bases were sequenced for each locus: *Hprt* (1290 bp), *Plp* (1595 bp), *Glra2* (1996 bp), *Amg* (1141 bp).

Data analysis: Sequences were aligned by eye, and the numbers and frequencies of all polymorphic sites were counted. Two measures of nucleotide variability, π (NEI and LI 1979) and θ (WATERSON 1975), were calculated for each locus. Nucleotide diversity, π, is based on the average number of nucleotide differences between any two sequences drawn from a sample, and θ is based on the total number of segregating sites in a sample. Under equilibrium conditions with respect to mutation and drift, both π and θ estimate the neutral parameter $4N_e\mu$ for autosomal loci, or $3N_e\mu$ for *X*-linked loci, where N_e is the effective population size and μ is the neutral mutation rate. Tajima's *D* statistic (TAJIMA 1989) was calculated to test this neutral expectation for each locus. Linkage disequilibrium (D') was calculated for a set of independent pairwise comparisons between polymorphic sites among the 10 individuals (LEWONTIN 1964), and the significance of D' was assessed using Fisher's exact tests. Ratios of polymorphism to divergence were compared with the expectations under a neutral model using the HKA test (HUDSON *et al.* 1987). Polymorphism data were from the 10 *M. domesticus* alleles and divergence was based on the single *M. caroli* sequence and a single randomly chosen *M. domesticus*.

TABLE 1

PCR amplification primers used in this study

Locus	Primer	Sequence
Hprt	F1	TCC TCA TGG ACT AAT TAT GGA CAG
	R1	GGT ACT CTC ATA TAA CAT CAG GAG
	F20	TTT TAT AGC ACG TTT GTG TCA T
	R20	CCT CTT AGA TGC TGT TAC TGA T
Plp	F1	CCT TCT GTC CAT CTG CAA AAC AGC
	R1	TAA GGA CGG CGA AGT TGT AAG TGG
	F2	GAC CTT CCA CCT GTT TAT TGC TGC
	R2	ACG CAG CAA TAA ACA GGT GGA AGG
Glra2	F1	CAA ATT GGA TTG TCT TAT TGA ACT GCC
	R1	CTC ATT GAG ACT GGT ATT GGT TTC C
	F2	TAA TGA CAC AAG TCG GTG AAA CCC
	R2	ATC TCA CCC TAC ACA ATC AAC TCC
Amg	F1	GCC CCT ACC ACC TCA TCC TGG G
	R2	CGG CTG CCT TAT CAT GCT CTG

RESULTS

A total of 6022 bases were sequenced in each of 10 *M. domesticus* and one *M. caroli*. The aligned *M. caroli* and *M. domesticus* sequences for each of the four genes are shown in Figure 3, and the polymorphic sites within the sample of 10 *M. domesticus* are shown in Table 2. A total of 13 polymorphic nucleotide sites and six polymorphic insertion-deletion variants were observed among the 10 *M. domesticus*. Measures of polymorphism for each locus are summarized in Table 3. Average nucleotide diversity (over all sites) was $\pi = 0.078\%$ and ranged from a low of $\pi = 0$ at *Plp* to a high of $\pi = 0.16\%$ at *Amg*. Average diversity based on numbers of segregating sites was $\theta = 0.076\%$ and ranged from a low of $\theta = 0$ at *Plp* to a high of $\theta = 0.142\%$ at *Glra2*. The similarity in estimates of π and θ, both overall and for each gene separately, indicates that there is no substantial skew in the frequency distribution of segregating sites. Tajima's *D*, which is based on differences between π and θ (TAJIMA 1989), is shown in Table 3 for each locus. None of these comparisons revealed sig-

Hprt

```
1
AATCATTTAT NNATCTTATA ACTGAAAAAG TTAACTATGA AATGGCTTAG GGTTACTTTA GAGAACATTT GTGAGGAAGG TACATTTGCA AGAATTAGAA
.......... .......... .......... .......... .......... A......... .......... .......... .......... .........T
101
TTGAAAGTAT TATTTTCTCT AAAGAATAAA CTATGATTAT TTATTTACAG TTGTTGGATT TGAAATTCCA GACAAGTTTG TTGTTGGATA TGCCCTTGAC
.......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
201
TATAATGAGT ACTTCAGGGA TTTGAATGTA AGTAATGCTT TTTTTTTTCT CGCTCTCATT TTTCAAAAAC CCACATAAAA ATTGAGGAAA GGGAAGAATT
.......... .......... .......... .......... .......... .......... .......... .......... ...T...... ..........
301
GTTTTCTCCT TCCAGCACCT CGTAATTTGA CCCGACTGAT GGTTCCCATT AGTCACATAA AGCTGTAGTC AAGTACAGAC GTCCTTAGAA CTGGAACTTG
.......... ..T....... ......G... ..A....... .......... .......... .......... .......C.T .......... ..........
401
GCCAGGCTAC GGTGACACTT CTTGCTGGCT GAAGTAGTTG AACAGCTTTA ATATACAACA ACTGTCTTAT TGTTATTTCA CATGATATAG TCATAAGTAA
.........G .......... .......... ......A... ......C... ........G. .......... .......... .......... C.........
501
GAAATAAACG ATTCAGTTTC AATTTTGTAA TTGACTCTGC TTCAGATCCC TGCCTCCCAT TCTGGAGGCA GGAAGGAGTC TTCTGGGTGA GTTTATAAGC
........T. .......... .......... .......... .......T.. .........C .......... .......... .......... ..........
601
ATGGTCCAGG AGCAGTTAGA AAGTCTCCAT GGGCAATATT ACCTAAAGCG TATTTTTTAT TTCACTTACC TGTTTTGATT TCAGCTTTAT GTTTCTCAAG
.......... .......... ...A..A... .......... .......... .G........ .......†.. .......... .......... ..........
701
TTACGTACTA AGACAAAGAA AAAGCAAACT AATGGACTAT AGTTCTGCTT TGTTTTCAAA TTATGTCCTC TTCAAGTTGC TGGTCCAAAA TACCTTGTTT
....A..... .......... .......... .......... .A........ .......... A......... .......... .......... ..........
801
GGTAGAAACC CAGACAACGT AGGAGGACCC TTTAATGCTG TTATTACCTC TTTAGAATCC GCTTTTCTAA CGAGAATGTC TTTTATTTTA TAGCACGTTT
.....G.... .......... ........T. .......... .....G.... .........T .......... T.T....... .......... .....T....
901
GTGTCATTAG Tgaaactgga aaagCCAAAT ACAAAGCCTA AGATGAGCGC AAGTTGAATC TGCAAATACG AGGAGTCCTG TTGATGTTGC CAGTAAAATT
.......... .                ..... .......... .......... .......... ......C... .......... .......... ..........
1001
AGCAGGTGTT CTAGTCCTGT GGCCATCTGC CTAGTAAAGC TTTTTGCATG AACCTTCTAT GAATGTTACT GTTTTATTTT TAGAAATGTC AGTTGCTGCG
.......... .......... .......... .......... .......... .......... ........T. .......... .....C.... ..........
1101
TCCCCAGACT TTTGATTTGC ACTATGAGCC TATAGGCCAG CCTACCCTCT GGTAGATTGT CGCTTATCTT GTAAGAAAAA CAAATCTCTT AAATTACCAC
.......... .......... .......... ......G... ...†...... .......... .......... ..G....... .......... ..........
1201
TTTTAAATAA TAATACTGAG ATTGTATCTG TAAGAAGGAT TTAAAGAGAA GCTATATTAG TTTTTTAATT GGTATTTTAA TTTTTATATA TTCAGGAGAG
.G........ .......... .......... .......... .......... .......... .......... .......... .......... ..........
1301
AAA
...
```

***Plp* Intron 5**

```
1
TTTCTACCTA TAATAGAAGT TTCTCCTTAT GAACTTAAGT TTTAAGGATT ATATTATTTC CTTATTGGAC TTTGTATTAG CAGATTTCCC TTCTACAAAA
.......... .......... .......C.. .......... ....G..... .......... ...G...... .......... .......... ..........
101
TAATGGTGAA ATGTACTTGA ATCAATGAGC ATACTTATTA AAAGTCAAAG TACTTATTAA CCCCAGAGCT GCACTAGAAG TCTTCCTTTC CCAAACTCCT
.......... .......... ........A. .......... .......... .......... .......... .......... .......... .........C
201
CTCATTAAAA GAAACATGTG TAGCTTTAAG TAAAGATAAT ACTTAGCAGT AAAATTTCCC AAATGAAAAT TTTCATTTTG GTTAAAAACG TACCAAGAAA
.......... .......... .......... .......... .......... .......... .......... .......... .......-.. ..........
301
ATGGGACATG CACAGAGACA CACATGGGCT CTATAAAATG CTTTTAATTT GGAACTCCTT AATTTCCAAT GAGAGGTGAA ATAAGTATAT TTGAAAAGAA
.......... ....T..... .......... .......... .......... A......... .......... .......... .........C ...GG-----
401          *
AAGAAAAGAA AAGAAAAGAG AAGAAAGAGC TTCATTGAGG ACAGACGCAT GCCACCGCAG CCAGGGAATA AGAAGCAAAT CTTGGGGAGG AAGTCTGAAA
-------A.. ..A....A.C .†....-T.. .......... .....T.... ....T..T.. .......... .A.......G .......... .........†
501
CCGTCTTTCA AGGGAAAGTA TTGGCTTTGA CCTTTGGGCT CCAGGTTCCA ACTGTAGGAA ATGAGTTCAT TATGTGGGAT CATTAATCAC TTTTCTGCAT
.......... ..A....... .......... .......... .......A.. .......... .......... G......... .......... ....G.....
601
CTTCGAGATG CTGTGGGGGT GAGGTACTAT GATAGTAGCA AATGGGAGGT TGACACTTAG GGAAGTGAAC ATTTAGTTAG AGACACAGAA TCAACATACA
....T..... .C.....-.. .......... .C........ .......... .......... .......... .......... .....T.... ..........
701
GGAATGGGTG TCAGCCCATC CAGACAGCCG TGTTCACTGT TTACAGTGGA TTGTGCTTGC TTTTCTGTTC TAAGAAATAA
......-... .......... .......... .......GC. .......... .......... .......... ..........
```

FIGURE 3.—Aligned sequences of *M. domesticus* and *M. caroli* for *Hprt*, *Plp*, *Glra2*, and *Amg*. The consensus sequence (from the 10 individuals surveyed) of *M. domesticus* is shown. Polymorphic nucleotide sites in *M. domesticus* are underlined and polymorphic insertion-deletion variation in *M. domesticus* is indicated by an asterisk. Small regions that were not sequenced in this study are indicated by bases in lowercase (positions 912–924 in *Hprt* and positions 813–881 in *Glra2*), the lowercase sequences shown are from published sequences (*Hprt*, MELTON *et al.* 1984; *Glra2*, MATZENBACH *et al.* 1994). The total lengths reported in Table 3 represent only those bases sequenced in this study. The *M. caroli* sequence is indicated by dots where it is identical to *M. domesticus;* only differences are shown. Deletions are indicated by a dash. Insertions are indicated by a dagger. The following insertions in *M. caroli* come after the indicated nucleotide position: *Hprt*, 668 CCCGGGCC, 1144 TCAGTT; *Plp* intron 5, 422 GAAA, 500 A; *Plp* intron 6, 244 TTTAATTCACCCAGGG, 477 A, 623 TGTTATGTTATTCCTAAATAACATATC, 705 A, 734 T; *Glra2*, 153 TGA, 1633 TG, 1744 GAACA; *Amg*, 304 AC, 521 AAAA.

***Plp* Intron 6**

```
1
GTTATTTTAA GATAATATTA GAAAAGAAGT GGTCCAGGGA TAGCATTAGG CCGAAAGACT AGCAGAGAGA CTCCTTACAG CCTTGAAAGG CTGTAGGCAA
..........  ..........  ..........  ..........  ..........  AT........  ..A.......  ..........  ..........  ..........
101
GCTGTGAGAA GGCACAGCAA GGTGTGGGGT TACACATCTC TGCCGTGAGT TGTCAAGCAT TGACCAGGGC ATGGCACTGT AAACATGATC TACTGTATCC
..C.......  ..........  .........C  ..........  .A..A.....  ..........  ....A.....  ....A.....  ..........  .G........
201
TAGGGTTCTA GATAGTATAA ATGCTCTCGG TAGTCTCTGG TGGGAAACAG GAGGGACTCG TCATATTAGA ATTGTTCTGA CTATGATGCA GTCCTGTGGC
....A.....  ..........  ........AA  ...C......  ...†......  .........-  .TG.......  ..........  ..G.......  ..........
301
CCAGGTCCTT CAGGTAGTGT TAATGTAGTG CCAATGCAGT GTGCTCCTTG TGCAGAGGTG AAGCCTTAGG TGCCAGTGTG CAAGGTGATG TATCTACCTC
..........  ..........  .........A  ....C.....  ..........  .........A  ..........  ........G.  ..........  ..C.......
401
AGATCTCTGA GGTAGCACTT AACTACAAGG AGGAGCTCAC AGAATCCTCC TCAGTACGGC AGAGGGGTTC ACAAAAAGCC AGCACACTTG CACTTTGTGA
.A........  ..........  ..........  ..........  ...C......  ..........  .....T....  ......†...  ..........  T.........
501
AATCAGTTGA TTATAAGAGA GTGGAGGGCT ATGGTGCATG AAGTAGCTGT GTGTGTGCAG CACATGCATA GCCAATGTCC TTCAGGAAAG GTCAGTGGTT
....C.....  ..........  ..........  ..........  ..........  .......T..  ..........  ..T.......  .........T  ..........
601
AGCAGAACTT AGCTGTGGAA TACATTGTAT GGATATGTTA GTTAGGAATG GGCTATTGTG GAAATGTCTA ACAGAAGATT TCTGTTCACA AATATAGACA
..........  .........G  .T†--.....  ..........  T.........  ..........  .G........  ..........  ...A......  ......T--.
701
AAAAATAGGG ACGTGCCTTT AATTTTATTT ATTTATTTTT TTGCAGATAG ACTATTTCCT TTTATTTAAT TGTGATTTTG CTTCAGCTCT CCCTCATTCC
....†C....  .....TT...  ........--  ...†......  ..........  ..........  ..........  .........A  ..........  ..........
801
CAAAAGGTTT TGAGT
..........  .....
```

Glra2

```
1
GGTTTCGGTG CTTGTTTACT TGTTTGCTCT TCAGAGAGGG TTAGTGAATA AAGAGGCCAG AGATAAAAGA TTCTATAGAT TCCTTTCTCT CGTTATCCAC
..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ......A...
101
ACGAAAAAAA TTAATCAGAA ATACCTTGCT GTGTTCAGTT TCTCTGTGTC ATCAGACACG TAGTTATGAT GACTACAGCA GCTCTTATAA AAACTGTACC
.........C  ..........  ..........  ..........  ......---.  ..†T......  ..........  ...C......  ..........  .G.......T
201
TGAAATGATT CCTTCCTCTG CTTGCCCACC TGTCATTACA TTTTGCCTTT TCTCAGAAAA GATACATTGT TTGTTACTAC TTCTTTCCTC CAGTTGGGTA
.........--  --....C...  ..........  ..........  ..C...AC..  ..........  ..........  ..........  ..........  ..........
301
CACCATGAAT GACCTGATAT TTGAGTGGTT AAGTGATGGT CCAGTACAAG TTGCTGAAGG ACTCACCTTG CCCCAGTTTA TTTTGAAAGA AGAGAAGGAG
..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
401
CTTGGTTATT GCACAAAGCA TTACAACACT GGTAAGGTTT TCTTTCTTTC ATTTGCTAAG CCCAAGAGCT CTCCATTTGT ATATGAGATT AGCTGCCCTA
..C.......  ..........  ..........  ..........  ..----....  .........-  ..........  ..........  ..........  ..G.A.....
501
TTATCTTAAC AGTTGCACAT TTATGAAGCA AATTGAATTT AAGAAAACTC TCAACATGTA GGGTTACTGT ACCCTATGGA ATAGGCTTGT GAGTGCAGCA
.........T  ..........  ..........  G.........  ........A.  ..........  ..........  ..........  ..........  ..........
601
GATTGTCTCC CTCAAACACA GAAAAATTAG GAATCTTTTA CTCTTCTCTT AATGAGAGTC TGCTGTCGAA TGAGTCTAAG GAAAAGTGTA GACCTCATTA
..........  ..........  ...G......  ..........  ..........  ..........  .....GT...  ..........  ....G.....  ..........
701
ATTGTACAGA AAACATAAAT CCTTGGCTAC TCTTTGCATC TACTCTTCAC ATTGCATCCT AAGGTCAATT TGCATTCCTG GCCTTAATGG AAGAAAGTGA
..........  ..........  T.........  ..........  ..........  ..........  CG........  ..........  ......G...  ..........
801
TAATGACACA AGtcggtgaa acccagaaat agagacatgc tctgcagttc tggaaaccaa taccagtctc aatgagcatc aTCTGGAGAG AATTCAGCTG
..........  G.                                                                                  .........  ..........
901         *
GCCCTAAAGT CAGGTGCTAT CTTTAGCACT GAAAGACATA TGGAGTAGGC AAGGTTGAGG TAGTGGCAAC AAGTGGGCAT CTGTTGTTGC CTGGAAGATT
..T.......  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
1001
ATACCTGACC AGAAAAACTG ATGTCTTTTA TTCTAAATGT TCTGGTCTTT TGATAACATG AATTGTGATG GTAGAGATAC AGTCATCACA AACTTCCTTG
..........  ..G.......  ..........  ..........  ..........  ..........  ..G.......  ..........  ..........  ..........
1101
GGTCTCTTGT GAGTAGAACA TTCTACAGCA GCCAACAGTG ACTTCCCATG AAGTAACTAA AACCTTATTC CAAGGTACTC TTCTAAGACT TCAACCCAGT
..........  ..........  ..........  ..........  ......A...  ..........  .....G....  ..........  ..........  ..........
1201
CTTCTATCTT TGTAGTGAAT AATACAAATG ATTTTCCAGG GTTTAGCTCA AAAGTCAATA AAAAGAGAGA AAAATGTAGA AAGTGGCAAG AGCATCTACT
..........  .........A  ..........  ..........  .....A....  ..........  ...G......  ..........  ..........  C.........
```

FIGURE 3.—*Continued*

nificant deviations from a neutral, equilibrium model ($P > 0.05$ for each locus and for all sites together), although the power of this test is low with small sample sizes (BRAVERMAN *et al.* 1995; SIMONSEN *et al.* 1995).

Measuring the extent of linkage disequilibrium in these data is complicated by the small number of segregating sites and the small sample size. With small samples and mostly rare polymorphisms, it is often impossible for purely numerical reasons to obtain significant results using the conventional contingency-table test (LEWONTIN 1995). Excluding polymorphisms present in only one individual (*i.e.*, singletons), there were 13

***Glra2* (continued)**

```
1301
TAATAATAAT GTTAAAAGTT GTCGTACCTT CCTATACACT CTTTTCTTTA TAGATGGAGC AGCACTACAG TTTCTGAGAA GGAATACTCC AAGGTTCTTC
......C... .......... .......... .......... .......... .......... .......... .......... .......... ..........
1401
AAGCAATTCT CAAAAGACTA GAAAGTTTCC ACAATGAACA GAGTGTACAT ATACTCCCTA ATTGTACCCT TTGGTCTTTA CCATTGAGAT TATAAGACAT
...A...... .......... .......... .......... .......... ......T... .A......T. .....T.... .......... ..C.......
1501
ACAACTATTC AAACACTCAG GATATTCTTC TTGCTTATTA GAAGGGGGTA GGGGAATAGT TTACTCAGTT AAGTGCTTGT CATGCAAGCA TGTAGACCTG
.......... .......... .......... .......... .......A.. .T.......C .......... .......... .........G ..........
1601
AGTTCAGTCC CCAGGACCCA GGGAATGATG ACATGTGCAT ATAGCTTAAG CACTGTGGAG GTAGACACAG GTAGCCAGCC AAGAGCCTGC CTCAAAACTC
.......... .......... ......A... ..†....--. .......... .......... .......... .......... .......... .......T..
1701
AAGATGGAGT CAACTTGAGG TTGACCTCTG GCTTTCACAT GCATGAACAC ACACATGAAC CCCAACACTA AGCATGCTCC CTGTCTCTCT GTGTCTCTCT
...G...... .......... .......... A.......... ...†...... .......... .......... .........T .......... ..........
1801
CTGCCTCTCT GTGTGTCTCT GTCTCTCTTC ATTGGTTACA ATTGTTTCAT ATTACATGAT ATGTTATATG TTACATAGTA TATAGTATAT GTTAATGATT
.......... .......... .......... .C........ .......... .......... .......... .......... .......... ..........
1901
GTACATGATG TGTATAATAC ATATAACACA AACATTCTTG AAATTTTTAA TCAGTCTTAA AAATGCTATA GTTCTATATA GAAAAATCTG TGATGGATCT
...A.....A ......T... .......... .......... .C........ .......... ..T....... .....G.... .......... .........AC
2001
AATCATATGA TTCCAGTATG AAGGCTCTTT CATGCTTACA CAATCATTTC CATTGTTAGA AGAAT
.......... C......... .......... .......... .......... .......... .....
```

Amg

```
1
TGTTTCTCAA ATGGTCACAC TTTTTCAGTT CCCACCAGCA GCTTTAAGCC CTGACTGATG TCCTCAAACC TGCATTTCTC CGCACCTTCC TACCTGGTCA
..C....... .......... .......... .......... .......... .......... .......... .......... .......... ..........
101
CTCTGACTCA ACCTCTACTT TTAAACCTAG TTATAAATGT ATCCACCATG AACTACCACT CAGGAAGGCT CCATGATAGG GCAGAGAGTG AACTCAGCTT
.......... G......... .......... C......... .......... .......... .......... .......... .......... ..........
201                                                                                    •
CAACAGTCTT GTTCTATGCC AGCCTATAAA ATTCAAGGAG TAAGTAAGAT GTTAAAATCT TTTCTAGCTG GAAAAAAATT GCTGATTCAA AGACATCCAC
.......... .......... .......... .......... .......... .......... .......... ........A. .......... ..........
301
ACACTGTGGG TTTTCTCATT GGCTATGCTA AGTGATCAAT GAAAGTGGTT ACCAAAGTCT CTACAGAGTG ATGGAACACA GAAATGTTCT AGAAACAGGA
...†.....- .......... .......... .......... .......... .......... ....G..... ....G.AG.. .......... ..........
401
TTTGGGGTTC TCCTAAACTG CCTGCAAACT ATATGACCCT GGGAAAATCA GTTCCCCACT CTGTAACTTT GGCTTCCTCA CCTAAACAAG GGAAACAATG
..C..C.... .......... .......... .......... .......... .......... .......... .......... .......... .........-
501
ATAAGAACAT TTCCAAAAAA ATATCATCCA GAAGAGCTTG AAAAGTGCTG TGTCACTCAT AACTTGAGTA AGGTTAATAG CTTTTTGATG TATTGGTTAA
--...T.... .......... †.......... .......... .........A .......... .......... .......... .......T.. ..........
601                                                                                                              •
ATGATGTACT ATATGAAAGA GCTATCTGAG ACTAATTAAT AAGACTATAC TGTCCAGAAG GAATGAAACT CTTTACTTAT CTCATGGGAT CAGATAAAAA
......C.T. .......... .......... .......... .......... .......... .......... .......... .......... .....---..
701
AAAACCTAAG TCTGTTAACT AGTATATACA GAAGAAAAAT ATCATAAGGA TCTGAGGCAA ATTTTAGTGT TTTATCTGAA AAGGTTGATT GATATATTTG
.......... .......... .A.G...G.. .......... G......... .A...A.... .......... .......... .......... ..........
801
ACCAAGAAAT CTATGTCTGA AGGATGAAAG ACAATGTTTC TACCTCATTT TTCCATTGTG ATATCAAATA AACAAGTAAT GATTATTATA TCACATTATT
.......... .......... .......... .......... .......C.. .........--. .......... .......... .......... ..........
901                                                                                    •
ATAAGTTTTC AATTGTTTTG CTCTTGCAAT GAAGAGGATT AAACTGAATT ACTGCTTTTG TTACTTCAGC TGAGTGAATG AACCTACCTG AGTTTTCATA
.......... .......... .......... .......... ...T...... .......... .......... .......... .......... ..........
1001                  •
ACATTACCTA TAAAAGTGAA GGTTGATACC TGCCTCCAAT CATGCGAGAA AGCATGAAAA GTGTATGGGA TGGATGTAAC ACAGTGCCTG ACACATAGAA
.......... .......... ..-....... .......... .......... .......... .......... ...------- .......... ..........
1101
AGCATCCTAC AAATGTTTAC CTTCTTCTTT CTTTTGTAGA A
.......... .......... .......... .......... .
```

FIGURE 3.—*Continued*

polymorphic sites (Table 2) and these were all found in *Glra2* and *Amg*, which are separated by ~2 cM. It is conventional to compute D' (LEWONTIN 1964) for all $n(n-1)/2$ pairwise comparisons among n segregating sites in a sample to look for groups of sites in linkage disequilibrium (*e.g.*, SCHAEFFER and MILLER 1993; KIRBY and STEPHAN 1996). This does not take into account that multiple comparisons are nonindependent, although the dependence may be weak if recombination is high. Here, the significance of D' was calculated using Fisher's exact test for a set of 12 independent pairwise comparisons among the 13 nonunique sites. Independent tests were constructed by taking the sites in order and comparing the pairs 1-2, 2-3, 3-4, . . . , 12-13 (LEWONTIN 1995). In no case were all four gametic types represented, and 10 out of 12 comparisons were significant ($P < 0.05$). Five out of six comparisons were significant within *Glra2*, all five comparisons were sig-

TABLE 2

Polymorphic nucleotide sites and polymorphic indels within *Mus domesticus*

	Hprt	*Plp*	*Glra2*	*Amg*
	1	0	000111122	00000001
	0	4	499157900	25677790
	0	1	901290305	74902272
	9	2	512804403	28663924
Individual		*a*	*b*	*c c dc*
Consensus	T	A	GGAGAAATT	-C-CGAG-
MWN 1036	.	.	.....	.....
MWN 1037	.	.	.....	.....
MWN 1083	.	.	.....	A..AA.-T
MWN 1103	.	.	C	.TA.AG-T
MWN 1114	.	.	.....	.TA.AG-T
MWN 1120	.	.	.....	.....
MWN 1126	.	.	C	.TA.AG-T
MWN 1204	.	.	AA-AGGGC.	.....
MWN 1213	.	.	AA..GGGC.	.....
MWN 1248	G	-	.....	.TA.AG-T

[a] Animal 1248 has the following 5-bp deletion after *Plp* intron 5 nucleotide 412, represented by a single dash: GAAAA.

[b] Animal 1204 has the following 6-bp deletion after *Glra2* nucleotide 912, represented by a single dash: AGGTGC.

[c] Single base-pair insertions follow *Amg* nucleotides 272, 696, and 1024.

[d] Animals 1083, 1103, 1114, 1126, and 1248 have the following 18-bp deletion after *Amg* nucleotide 972, represented by a single dash: AGTGAATGAACCTACCTG.

nificant within *Amg*, and the single comparison between *Glra2* and *Amg* was not significant. It is important to realize that multiple comparisons within genes do not contain separate information in the case of complete linkage. These data indicate that there is linkage disequilibrium among sites within *Glra2* and among sites within *Amg*, but no evidence for linkage disequilibrium between *Glra2* and *Amg*, although there is little power to detect it.

It is possible to construct a single most parsimonious network linking all individuals (Figure 4). This sample contains mice from Northern Italy (1036, 1037), Central Italy (1083, 1103, 1114, 1120, 1126, and 1204) and Southern Italy (1213, 1248). However, there is no association between haplotypes and geography. For example, the two mice from Southern Italy are among the most distantly related haplotypes in the network. Similar discordance between geography and phylogeny is also seen in mitochondrial DNA (NACHMAN *et al.* 1994) and *Y*-chromosome sequences (TUCKER *et al.* 1989).

Divergence between *M. domesticus* and *M. caroli* averaged 3.06% across all 6022 bases, and varied from a low of 2.02% at *Amg* to a high of 4.39% at *Plp* (Table 3). These values are lower than the divergence estimate of 7% between these taxa based on DNA-DNA hybridization experiments (RICE and STRAUSS 1973; SHE *et al.* 1990). This difference may be due in part to the different regions of the genome that were sampled in this study (*X*-linked introns and small portions of exons, Figure 2) *vs.* DNA hybridization studies that used total single copy DNA. It is possible to estimate the average neutral mutation rate (μ) from these data, assuming that intron sequences are free of constraint. Given the divergence time between *M. domesticus* and *M. caroli* of 2.5 million years (SHE *et al.* 1990) and assuming that mice have two generations per year, $\mu = 3.06 \times 10^{-9}$. Under the neutral expectation for *X*-linked genes that $= 3N_e\mu$, the average effective population size estimated from these data is $N_e = 8.5 \times 10^4$.

The ratio of polymorphism to divergence is very similar for *Hprt* (1:35) and *Plp* (0:70), and also for *Glra2* (8:56) compared to *Amg* (4:23) (Table 3). To test specifically whether ratios of polymorphism to divergence differed between these chromosomal regions (*i.e.*, center of *X vs.* telomeric region), HKA tests were used (Table 4). Both comparisons involving *Plp* were significant while comparisons involving *Hprt* were not significant. Given the small number of segregating sites involved, the power of the test to reject the null hypothesis is low. Likewise, the sensitivity of the test result to single observations is high. For example, had there been zero polymorphic sites at *Hprt* instead of one, the *Hprt-Amg* test would have been significant at the 0.05 level. A single, four-locus HKA test incorporating all the data was not significant ($\chi^2_{[6]} = 7.7$, $P > 0.10$).

Recombination rates for each of the four loci are

TABLE 3

Levels of polymorphism and divergence at four *X*-linked loci

	MGD map position			Polymorphism					Divergence		
Locus	(cM)	Recombination rate (cM/Mb)	Length (bp)	Base substitutions	Insertion/ deletions	π (%)	θ (%)	Tajima's *D*	Base substitution	Insertion/ deletion	*D* (%)
Hprt	17	0.25	1290	1	0	0.016	0.027	−1.07	35	2	2.71
Plp	56	0.30	1595	0	1	0.000	0.000	—	70	16	4.39
Glra2	70	0.51	1996	8	1	0.135	0.142	−0.20	56	8	2.81
Amg	72	0.44	1141	4	4	0.160	0.124	1.18	23	8	2.02
Total	—	—	6022	13	6	0.078	0.076	0.10	184	34	3.06

Polymorphism was based on a random sample of 10 *M. domesticus*. Divergence was between a randomly chosen allele from *M. domesticus* and a randomly chosen allele from *M. caroli*.

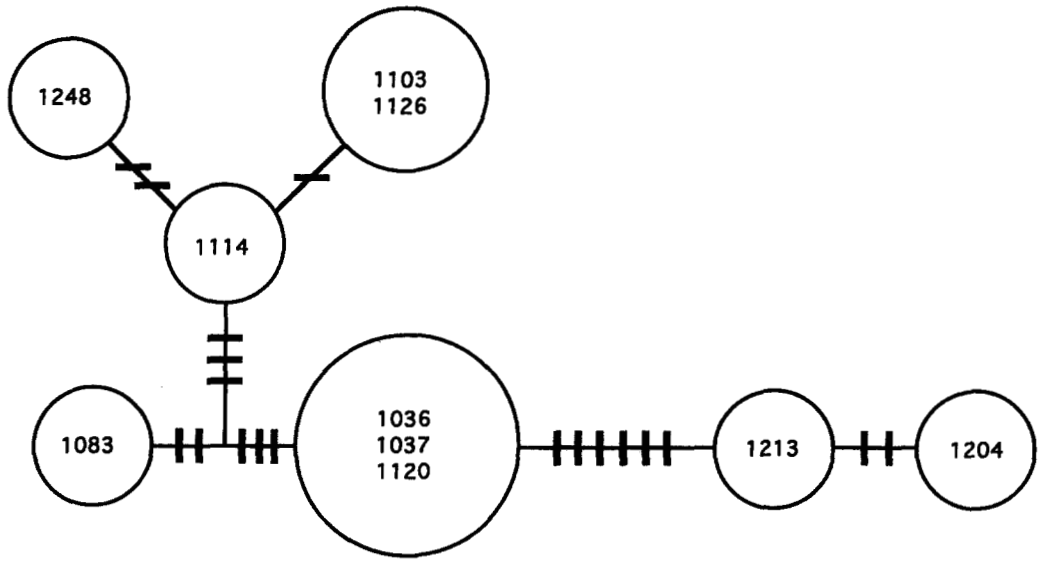

FIGURE 4.—Single most parsimonious network linking all haplotypes. The size of each circle represents the frequency of that haplotype in the population. Numbers correspond to individuals; the collecting localities for these individuals are given in MATERIALS AND METHODS. Hash marks indicate mutations (both nucleotide changes and insertion-deletion changes).

shown in Table 3. These rates are derived from the density of microsatellite markers on the linkage map of the *X* chromosome and rely on the assumption that (CA)*n* dinucleotide repeats are randomly distributed physically on the *X* (NACHMAN and CHURCHILL 1996). These estimates are in general agreement with cytological studies of chiasmata that indicate that recombination rates are elevated near the telomere of the mouse *X* (POLANI 1972; JAGIELLO and FANG 1987).

Figure 5 depicts scatterplots of π *vs.* recombination rate, θ *vs.* recombination rate, and divergence *vs.* recombination rate. These show a trend with higher heterozygosity for the two genes experiencing higher recombination rates, yet no comparable trend when comparing divergence to recombination rate. As such, the data are consistent with the correlation between heterozygosity and recombination rate observed in *D. melanogaster* (BEGUN and AQUADRO 1992; AQUADRO *et al.* 1994). It is premature, however, to conclude based on four loci that there is a general correlation between heterozygosity and recombination rate in mice; additional loci will need to be surveyed to determine whether this is so.

DISCUSSION

These results indicate that average nucleotide heterozygosity in mice is quite low, yet there is substantial variation in π among loci. Moreover, not all of the observations are consistent with a simple, neutral model of molecular evolution. Ratios of polymorphism to divergence differ between chromosomal regions, and genes experiencing lower recombination rates tend to be less polymorphic.

Nucleotide variability in *M. domesticus*: Estimates of nucleotide variability from nuclear genes in different organisms are surprisingly rare. By far the best studied species in this regard is *D. melanogaster*, where $\pi = 0.4\%$ on average for all sites and $\pi = 1.3\%$ on average for silent sites (MORIYAMA and POWELL 1996). In *D. melanogaster* nucleotide diversity ranges from a low of zero in low recombination regions of the genome (*e.g., cubitus interruptus* $\pi = 0$, BERRY *et al.* 1991; *yellow-achaete* $\pi = 0.08\%$, BEGUN and AQUADRO 1991) to a high of over 1% for some genes in some populations (*e.g., vermillion* introns from a population in Kenya $\pi = 1.78\%$, BEGUN and AQUADRO 1995). Average nucleotide diversity is higher in *D. simulans* ($\pi = 3.0\%$ for silent sites) and *D. pseudoobscura* ($\pi = 2.8\%$ for silent sites; MORIYAMA and POWELL 1996). Nucleotide diversity has been estimated in humans from comparison of published cDNA sequences and is much lower than in Drosophila (LI and SADLER 1991). The highest diversity observed in humans was at fourfold degenerate sites ($\pi = 0.11\%$).

TABLE 4

HKA tests comparing genes in different chromosomal regions

Loci	θ_1 (%)	θ_2 (%)	S_1/ES_1	S_2/ES_2	D_1/ED_1	D_2/ED_2	T	χ^2	P
Hprt-Glra2	0.089	0.102	1/3.2	8/5.8	35/32.8	56/58.2	27.6	1.59	>0.10
Hprt-Amg	0.078	0.066	1/2.9	4/2.1	35/33.1	23/24.9	31.8	2.16	>0.10
Plp-Glra2	0.093	0.068	0/4.2	8/3.8	70/65.8	56/60.2	43.6	5.49	<0.05*
Plp-Amg	0.064	0.034	0/2.9	4/1.1	70/67.1	23/25.9	64.8	8.46	<0.01**

Subscripts denote locus 1 and locus 2, respectively. *S*, observed segregating sites; *ES*, expected segregating sites; *D*, observed divergence, *ED*, expected divergence; *T*, divergence time as in HUDSON *et al.* (1987).

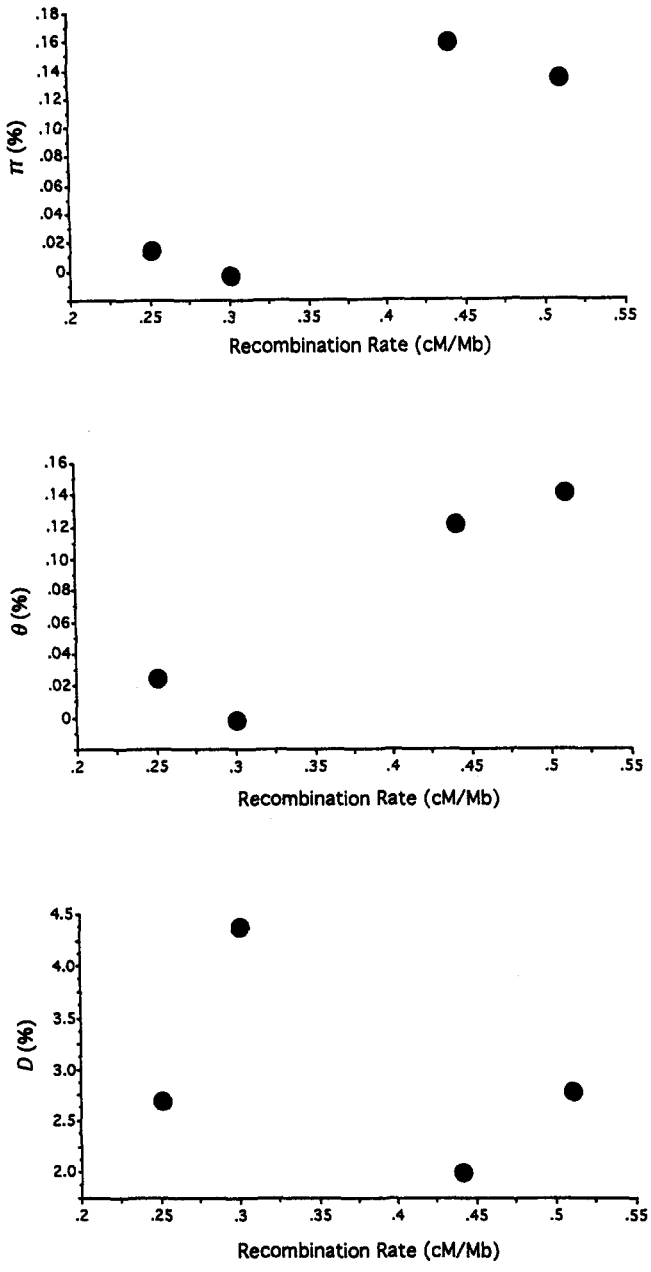

FIGURE 5.—Scatterplots of π (NEI and LI 1979) *vs.* recombination rate, θ (WATERSON 1975) *vs.* recombination rate, and nucleotide divergence (D) between *M. domesticus* and *M. caroli* *vs.* recombination rate. Data are from Table 3.

Heterozygosity from *X*-linked loci must be multiplied by 4/3 to compare directly to heterozygosity from autosomal loci to account for the different effective populations sizes of the *X* and autosomes. The value obtained for mice (0.078% × 4/3 = 0.104%) is nearly identical to the value observed at fourfold degenerate sites in humans (0.11%; LI and SADLER 1991). In mammals, fourfold degenerate sites, introns, and pseudogenes all appear to have roughly similar levels of selective constraint (LI and GRAUR 1991). If mutation rates are similar for humans and mice, and if fourfold degenerate sites and introns experience similar selective constraints, then the similarity in nucleotide diversity suggests that mice and humans have similar long term effective population sizes. This appears to be approximately correct; most estimates of N_e for humans are on the order of 10^4 (*e.g.*, TAKAHATA 1994; HAMMER 1995), while the estimate of N_e for mice from these data is 8.5×10^4. This estimate is slightly below estimates from mouse mitochondrial DNA ($N_e = 3.2 \times 10^5$, assuming a sex ratio of one), which has higher nucleotide diversity ($\pi = 0.69\%$), but also has a higher mutation rate ($\mu = 2.1 \times 10^8$; calculated from data in NACHMAN *et al.* 1994 for mice from the same regions in Italy).

The average value of nucleotide diversity hides substantial variation among loci. Excluding *Plp*, at which no polymorphisms were observed, π still varies by one order of magnitude among the remaining three loci (Table 3). This variation does not simply reflect differences in functional constraint; the average level of divergence between *M. domesticus* and *M. caroli* varied only twofold among these loci. Substantial variation in π among loci is also seen in the *D. melanogaster* genome (AQUADRO 1992), and it is thus reasonable to expect that significant variation among loci will be present in other organisms as well. This variation in π suggests that there is substantial variation in N_e among loci.

The levels of nucleotide diversity reported here for mice raise an issue previously noted in the analysis of nucleotide diversity in humans (LI and SADLER 1991). Although π for mice and humans appears to be at least one order of magnitude lower than for several species of Drosophila, levels of allozyme heterozygosity are similar among these species (Table 5). Under neutral, equilibrium conditions both nucleotide diversity (π) and allozyme heterozygosity (H) are expected to increase as a function of population size ($\pi = 4N_e\mu$ under an infinite sites model; $H = 4N_e\mu/(4N_e\mu + 1)$ under an infinite alleles model). Nucleotide diversity for nuclear genes is larger for Drosophila than for either Mus or Homo, consistent with the idea that flies have larger effective population sizes than mice or humans. This difference is also apparent in mtDNA nucleotide diversity, given that Drosophila mitochondrial mutation rates are considerably lower than in mammals (SHARP and LI 1989), but this difference is not seen in the allozyme data. One explanation for these observations is that most DNA mutations are strictly neutral but that many allozyme differences involve slightly deleterious mutations that are eliminated more effectively in larger populations. This idea has been invoked to explain differences at *Xdh* between *D. melanogaster* and *D. simulans* (AQUADRO *et al.* 1988) and has also been invoked to explain differences between *D. melanogaster* and *H. sapiens* (LI and SADLER 1991). One difficulty with this hypothesis is that it requires selection coefficients, s, to fall within a very narrow range of values (close to $N_e s = 1$; OHTA 1992; GILLESPIE 1994). If the spectrum of selection coefficients is such that $N_e s = 1$ for many new

TABLE 5

Nuclear gene nucleotide diversity, mitochondrial nucleotide diversity, and allozyme heterozygosity from Mus, Homo, and Drosophila

Species	π_N (%)[a]	π_M (%)[b]	H
Mus domesticus	0.1	0.77	0.09
Homo sapiens	0.1	0.32	0.12
Drosophila melanogaster	1.3	0.64	0.10
Drosophila simulans	3.0	0.49	0.09
Drosophila pseudoobscura	2.8	2.13	0.12

Data from the following: *M. domesticus* (this study; NACHMAN *et al.* 1994a; SELANDER *et al.* 1969); *H. sapiens* (LI and SADLER 1991; CANN *et al.* 1987; NEI and ROYCHOUDHURY 1982); *D. melanogaster* (MORIYAMA and POWELL 1996; RAND and KANN 1996; SINGH and RHOMBERG 1987); *D. simulans* (MORIYAMA and POWELL 1996; RAND and KANN 1996; CHOUDHARY *et al.* 1992); *D. pseudoobscura* (MORIYAMA and POWELL 1996; JENKINS *et al.* 1996; LEWONTIN and HUBBY 1966). π_N, nuclear gene nucleotide diversity; π_M, mitochondrial nucleotide diversity; H, allozyme heterozygosity.

[a] Nucleotide diversity for *X*-linked genes was multiplied by 4/3 to be comparable to autosomal loci. π_N is from introns for mice, fourfold degenerate sites for humans, and silent sites for the three species of Drosophila.

[b] π_M is from the control region for mice, RFLP data from across the genome for humans, silent sites for *D. melanogaster* and *D. simulans*, and the small ribosomal RNA gene for *D. pseudoobscura*.

mutations in one of the species in Table 5, then $N_e s$ should be very different for some of the other species, given the presumed differences in N_e. While a slightly deleterious model (OHTA 1992) may account for some of the data, it is difficult to reconcile with the range of variation shown in Table 5. It is possible that the distribution of selection coefficients may vary among species, but there is no obvious biological reason why this distribution should vary in such a way that a substantial proportion of new mutations always fall in the neighborhood of $N_e s = 1$. An alternative explanation for the data in Table 5 is that allozyme frequencies are maintained at similar levels in a diverse array of species by selection. This is a difficult proposition to test, but may be supported by studies of individual loci (*e.g.*, *Ldh* in fish, DIMICHELE and POWERS 1982a,b; *Pgi* in butterflies, WATT 1983; WATT *et al.* 1986; *Adh* in Drosophila, BERRY and KREITMAN 1993).

Linkage disequilibrium: There was more linkage disequilibrium among sites within *Glra2* and *Amg* than is generally observed among polymorphic sites in high recombination regions of the *D. melanogaster* genome (*e.g.*, BEGUN and AQUADRO 1995). Several factors may account for the higher linkage disequilibrium in mice, including lower recombination rates, smaller N_e, or recent changes in population structure. *M. domesticus* recolonized Europe from the Middle East only within the last 10,000 years (AUFFRAY *et al.* 1990), and populations are likely to have undergone expansions and mixing since their invasion. There is considerable discordance between geography and mtDNA genealogies among *M. domesticus* populations, suggesting that ancestral polymorphism is not fully sorted (NACHMAN *et al.* 1994a). Recent admixtures or population expansions may have contributed to increase linkage disequilibrium.

Neutral and nonneutral evolution: Many of the results are in good agreement with the predictions of a strictly neutral model. Within genomic regions, ratios of polymorphism to divergence are similar for different genes. For example, the observed ratios at *Glra2* (1:7) and *Amg* (1:6) are nearly identical. One possible effect of selection is a change in the frequency distribution of polymorphisms: directional selection can produce an excess (over neutral expectations) of rare polymorphisms and balancing selection can produce an excess of intermediate frequency polymorphisms. However, there is no evidence in these data for a skew in the frequency distribution of polymorphic sites at any of the genes. Tajima's D is not significantly negative or positive in any of the tests (Table 3), although the power of the test is low with small samples and few segregating sites (BRAVERMAN *et al.* 1995; SIMONSEN *et al.* 1995) so the inability to reject should not be construed as strong support for a neutral, equilibrium model.

Indeed not all of the data are compatible with neutrality. Comparisons between genes in the middle of the *X* and genes near the telomere of the *X* show that ratios of polymorphism to divergence are quite different, and two of these four comparisons are significant using the HKA test (Table 4). Comparisons between *Hprt* and either *Glra2* and *Amg* are not significant, but the trend is in the same direction as observed in comparisons between *Plp* and either *Glra2* and *Amg* (*i.e.*, telomeric genes have higher ratios of polymorphism to divergence). Because the average level of nucleotide diversity is quite low in *M. domesticus*, detecting a reduction in π is expected to be difficult. The highest value of π for *M. domesticus* ($\pi = 0.16\%$ at *Amg*) is one order of magnitude lower than the highest values of π in *D. melanogaster* ($\pi = 1.85\%$ at *vermillion* in Africa; BEGUN and AQUADRO 1995). In this context, the observation of any significant reduction in π is notable and implies that selection has acted to alter ratios of polymorphism to divergence. This does not imply that selection is acting on any of the genes studied. Instead, differences in ratios of polymorphism to divergence are probably due to the effects of selection on linked sites (MAYNARD SMITH and HAIGH 1974; CHARLESWORTH *et al.* 1983), as described below.

Recombination rates and variability: Ideally, recombination rates would be estimated by comparing the genetic and physical distance between loci sampled at regular intervals along a chromosome. This requires a dense linkage map and a dense physical map, and these data are currently available only for a few species such as *D. melanogaster*. No comparable data exist for any

species of mammal. The estimates of recombination rates in Table 3 are from NACHMAN and CHURCHILL (1996) and are based on the density of microsatellite $(CA)_n$ markers along a genetic map (DIETRICH *et al.* 1992, 1994), the total genetic length of the *X* chromosome and the total physical length of the *X* chromosome. These estimates suffer from several layers of uncertainty. First, the method assumes that $(CA)_n$ repeats are uniformly distributed physically along the *X* chromosome. Second, the estimates depend on the inherent accuracy of the genetic map, the correct integration of the four loci studied (*Hprt, Plp, Glra2,* and *Amg*) with the microsatellite markers, and the correct assessment of the overall genetic and physical length of the *X* chromosome. Finally, the estimates are based on a single (OB × CAST) F_2 intercross and may or may not reflect recombinational distances between these same markers in wild mice. It is not uncommon to observe differences in recombinational distances between the same set of markers in different strains of laboratory mice (DAVISSON *et al.* 1989). Despite these uncertainties, two observations suggest that the estimates in Table 3 may be reasonable. First, cytological observations of chiasmata show that, in general, telomeres tend to have increased recombination rates in mice (POLANI 1972). Second, comparison of the genetic and physical location of markers (that have been physically localized by *in situ* hybridization to G-banded metaphase chromosomes) suggests that the telomere of the mouse *X* has elevated recombination rates compared to the central region of the *X* chromosome (Table 4 in NACHMAN and CHURCHILL 1996).

Comparison of recombination rates with the levels of polymorphism and divergence for the four genes shows that the ratios of polymorphism to divergence are lowest at the low recombination genes, *Hprt* and *Plp* (Table 3 and Figure 5). These results are consistent with previous findings from *D. melanogaster,* where a survey of 20 loci from throughout the genome showed a positive correlation between π and recombination rate but no correlation between divergence and recombination rate (BEGUN and AQUADRO 1992). Those results were based on data from the literature and included a diverse sample of loci and populations. More recently, *D. melanogaster* from a single population in Maryland were surveyed and revealed a positive correlation between nucleotide variability and recombination rate among 15 loci on the third chromosome (AQUADRO *et al.* 1994). Caution is warranted in interpreting the scatterplots in Figure 5. While there is a strong trend in the data, with only four data points there is no basis at present for concluding that there is a general correlation between π and recombination rate.

The correlation between nucleotide variability and recombination rate observed in *D. melanogaster* may be explained by either of two very different processes. One hypothesis involves selective sweeps and the associated genetic hitchhiking of linked neutral sites (MAYNARD SMITH and HAIGH 1974; KAPLAN *et al.* 1989; BEGUN and AQUADRO 1992). The other hypothesis is background selection, or selection that removes deleterious mutations and linked neutral variation from the population (CHARLESWORTH *et al.* 1993, 1995; CHARLESWORTH 1994; HUDSON and KAPLAN 1994, 1995). It is somewhat unfortunate that both processes may explain the observed pattern since they have fundamentally different biological implications. If hitchhiking is the correct (or predominant) explanation, this suggests that the substitution rate for adaptive mutations may be quite high. Background selection, on the other hand, is a logical extension of neutral theory and invokes no special role for adaptive mutations. Of course these two processes are not mutually exclusive and both undoubtedly occur to some extent.

The data presented here are in some ways quite different from the observations in Drosophila. In *D. melanogaster,* a 10-fold reduction in recombination rate corresponds approximately to a fivefold reduction in heterozygosity (BEGUN and AQUADRO 1992; AQUADRO *et al.* 1994). The plots in Figure 5 suggest that a twofold reduction in recombination rate corresponds approximately to a 10-fold reduction in heterozygosity. Such an extreme difference in heterozygosity is not expected under current models of either background selection or genetic hitchhiking unless the number of selected mutations and the strength of selection are quite high. For example, under a simple model of background selection, the reduction in heterozygosity corresponds to f_o, the equilibrium frequency of mutation free gametes in the population. For *X*-linked genes, this is given by

$$f_o = \exp\{-3U/[2s(2h + 1)]\},$$

where U is the deleterious mutation rate for the region of the *X* in question, s is the average selection coefficient and h is the dominance factor (CHARLESWORTH *et al.* 1993). The *X* chromosome constitutes 6.2% of the total *M. domesticus* genome (EVANS 1989). As an extremely conservative upper estimate, we assume that complete linkage extends over one third of the *X* chromosome. This is conservative because recombination rates are not zero (Table 3), and because the region of low recombination on the *X* is probably smaller than one third of the chromosome (NACHMAN and CHURCHILL 1996). This suggests that, at most, the portion of the *X* under consideration for background selection constitutes ~2% of the mouse genome. Estimates of U, s, and h are not available for mice, but we can use estimates of these parameters from *D. melanogaster* to estimate the expected reduction in heterozygosity, f_o, for the mouse *X* chromosome under background selection. If U for the total genome is approximately one, then for this portion of the *X*, $U = 0.02$. Assuming $s = 0.1$ and $h = 0.2$ (CHARLESWORTH *et al.* 1993), the expected reduction in heterozygosity is $f_o = 0.66$. In fact, we observe a much

greater reduction in heterozygosity ($f_o = 0.1$). We can also use the observed reduction in heterozygosity and $s = 0.1$ and $h = 0.2$ to calculate the deleterious mutation rate that is consistent with the data. In this case, $U = 0.22$ for the portion of the X chromosome under consideration. This is not intended as an accurate estimate, but serves to illustrate that the data are compatible with background selection only if the deleterious mutation rate or the strength of selection is very high. Similarly, these data are consistent with models of selective sweeps (WIEHE and STEPHAN 1993; STEPHAN 1995) only with high mutation rates or strong selection.

There are several alternative explanations for the plots in Figure 5. One is simply that the estimates of recombination rate are poor. Another is that the four data points do not provide an accurate picture of the underlying relationship between heterozygosity and recombination rate, assuming such a relationship exists. For example, it is possible to randomly choose four loci from among the 20 in the data of BEGUN and AQUADRO (1992) and come up with a very different picture of the relationship between heterozygosity and recombination rate. Finally, it is possible that the genes surveyed here are closely linked to loci recently under selection. Variation at *Hprt* and *Plp* may have been reduced as a consequence of one or more selective sweeps at nearby loci, or alternatively, variation at *Amg* and *Glra2* may be maintained by association with one or more balanced polymorphisms.

Studies aimed at better documenting the relationship between recombination rate and nucleotide variability in mice are underway. Future studies are needed in Drosophila and other organisms to help distinguish the relative contributions of genetic hitchhiking and background selection to observed patterns of variation. Such studies may include detailed surveys in regions of low recombination (HAMBLIN and AQUADRO 1996), comparison of *X*-linked and autosomal loci (AQUADRO *et al.* 1994; CHARLESWORTH 1994), studies on the frequency spectrum of mutations in low recombination regions (CHARLESWORTH *et al.* 1993), or comparison of loci with different mutation rates (SLATKIN 1995).

Finally, it is worth noting that the possibility of a correlation between recombination rate and nucleotide variability has important implications for molecular evolution, regardless of the cause of the relationship. For example, the relationship implies that the effective population size for different genomic regions varies in a predictable fashion. Since the efficacy of selection depends on N_e, mutations arising in regions with smaller N_e (low recombination regions) will be more influenced by drift. Conversely, the rate of adaptive evolution is expected to be faster in regions of high N_e. In principle, these predictions could be tested by comparing interspecific substitutions rates and patterns among genes that have experienced different recombination rates.

I thank S. L. CROWELL for sequencing a portion of *Glra2*, C. F. AQUADRO, D. J. BEGUN, M. J. FORD, and M. F. HAMMER for discussions, and R. C. LEWONTIN and D. M. RAND for comments on the manuscript. This work was supported by the National Science Foundation.

LITERATURE CITED

AQUADRO, C. F., 1993 Molecular population genetics of Drosophila, pp. 222–266 in *Molecular Approaches to Pure and Applied Entomology*, edited by J. OAKESHOTT and M. WHITTEN. Springer-Verlag, New York.

AQUADRO, C. F., 1992 Why is the genome variable? Insights from Drosophila. Trends Genet. **8:** 355–362.

AQUADRO, C. F., D. J. BEGUN and E. C. KINDAHL, 1994 Selection, recombination, and DNA polymorphism in Drosophila, pp. 46–56 in *Non-Neutral Evolution. Theories and Molecular Data,* edited by B. GOLDING. Chapman and Hall, New York.

AQUADRO, C. F., K. M. LADO and W. A. NOON, 1988 The *rosy* region of *Drosophila melanogaster* and *Drosophila simulans*. I. Contrasting levels of naturally occurring DNA restriction map variation and divergence. Genetics **119:** 875–888.

AUFFRAY, J-C., F. VANLERBERGHE and J. BRITTON-DAVIDIAN, 1990 The house mouse progression in Eurasia: a palaeontological and archaeozoological approach. Biol. J. Linn. Soc. **41:** 13–25.

BEGUN, D. J., and C. F. AQUADRO, 1991 Molecular population genetics of the distal portion of the X chromosome in Drosophila: evidence for genetic hitchhiking of the yellow-achaete region. Genetics **129:** 1147–1158.

BEGUN, D. J., and C. F. AQUADRO, 1992 Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. Nature **356:** 519–520.

BEGUN, D. J., and C. F. AQUADRO, 1995 Molecular variation at the *vermillion* locus in geographically diverse populations of *Drosophila melanogaster* and *D. simulans*. Genetics **140:** 1019–1032.

BERRY, A., and M. KREITMAN, 1993 Molecular analysis of an allozyme cline—alcohol dehydrogenase in *Drosophila melanogaster* on the east coast of North America. Genetics **134:** 869–893.

BERRY, A. J., J. W. AJIOKA and M. KREITMAN, 1991 Lack of polymorphism on the Drosophila fourth chromosome resulting from selection. Genetics **129:** 1111–1117.

BRAVERMAN, J. M., R. R. HUDSON, N. L. KAPLAN, C. H. LANGLEY and W. STEPHAN, 1995 The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. Genetics **140:** 783–796.

BROOKFIELD, J. F. Y., and P. M. SHARP, 1994 Neutralism and selectionism face up to DNA data. Trends Genet. **10:** 109–111.

CANN, R. L., M. STONEKING and A. C. WILSON, 1987 Mitochondrial DNA and human evolution. Nature **325:** 31–36.

CHARLESWORTH, B., 1994 The effect of background selection against deleterious mutations on weakly selected, linked variants. Genet. Res. Camb. **63:** 213–227.

CHARLESWORTH, B., M. T. MORGAN and D. CHARLESWORTH, 1993 The effect of deleterious mutations on neutral molecular variation. Genetics **134:** 1289–1303.

CHARLESWORTH, D., B. CHARLESWORTH and M. T. MORGAN, 1995 The pattern of neutral molecular variation under the background selection model. Genetics **141:** 1619–1632.

CHOUDHARY, M., M. B. COULTHART and R. S. SINGH, 1992 A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*. VI. Patterns and processes of genic divergence between *D. melanogaster* and its sibling species, *Drosophila simulans*. Genetics **130:** 843–853.

DAVISSON, M. T., T. H. RODERICK and D. P. DOOLITTLE, 1989 Recombination percentages and chromosomal assignments, pp. 432–505 in *Genetic Variants and Strains of the Laboratory Mouse,* Ed. 2, edited by M. F. LYON and A. G. SEARLE. Oxford University Press, Oxford.

DIETRICH, W. F., H. KATZ, S. E. LINCOLN, H-S. SHIN, J. FRIEDMAN *et al.*, 1992 A genetic map of the mouse suitable for typing intraspecific crosses. Genetics **131:** 423–447.

DIETRICH, W. F., J. C. MILLER, R. G. STEEN, M. MERCHANT, D. DAMRON *et al.*, 1994 A genetic map of the mouse with 4,006 simple sequence length polymorphisms. Nature Genet. **7:** 220–225.

DIMICHELE, L., and D. A. POWERS 1982a LDH-B genotype-specific

hatching times of *Fundulus heteroclitus* embryos. Nature **296:** 563–564.

DIMICHELE, L., and D. A. POWERS 1982b Physiological basis for swimming endurance differences between LDH-B genotypes of *Fundulus heteroclitus*. Science **216:** 1014–1016.

EVANS, E. P., 1989 Standard idiogram, pp. 576–578 in *Genetic Variants and Strains of the Laboratory Mouse*, Ed. 2, edited by M. F. LYON and A. G. SEARLE. Oxford University Press, Oxford.

GAUT, B. S., and M. T. CLEGG, 1993a Nucleotide polymorphism in the *Adh1* locus of pearl millet (*Pennisetum glaucum*) (Poaceae). Genetics **135:** 1091–1097.

GAUT, B. S., and M. T. CLEGG, 1993b Molecular evolution of the *Adh1* locus in the genus *Zea*. Proc. Natl. Acad. Sci. USA **90:** 5095–5099.

GILLESPIE, J. H., 1994 Substitution processes in molecular evolution. III. Deleterious alleles. Genetics **138:** 943–952.

GUTTMAN, D. S., and D. E. DYKHUIZEN, 1994 Detecting selective sweeps in naturally occurring *Escherichia coli*. Genetics **138:** 993–1003.

HAMBLIN, M. T., and C. F. AQUADRO, 1996 High nucleotide sequence variation in a region of low recombination in Drosophila simulans is consistent with the background selection model. Mol. Biol. Evol. **13:** 1133–1140.

HAMMER, M., 1995 A recent common ancestry for human Y chromosomes. Nature **378:** 376–378.

HEY, J., 1997 Mitochondrial and nuclear genes present conflicting portraits of human origins. Mol. Biol. Evol. **14:** 166–172.

HUDSON, R. R., 1994 How can the low levels of DNA sequence variation in regions of the Drosophila genome with low recombination rates be explained? Proc. Natl. Acad. Sci. USA **91:** 6815–6818.

HUDSON, R. R., 1996 Molecular population genetics of adaptation. pp. 291–309 in *Adaptation*, edited by M. R. ROSE and G. V. LAUDER. Academic Press, San Diego.

HUDSON, R. R., and N. L. KAPLAN, 1994 Gene trees with background selection. pp. 140–153 in *Non-Neutral Evolution. Theories and Molecular Data*, edited by B. GOLDING. Chapman and Hall, New York.

HUDSON, R. R., and N. L. KAPLAN, 1995 Deleterious background selection with recombination. Genetics **141:** 1605–1617.

HUDSON, R. R., M. KREITMAN and M. AGUADE, 1987 A test of neutral molecular evolution based on nucleotide data. Genetics **116:** 153–159.

IKENAKA, K., T. FURUICHI, Y. IWASAKI, A. MORIGUCHI, H. OKANO *et al.*, 1988 Myelin proteolipid protein gene structure and its regulation of expression in normal and jimpy mutant mice. J. Mol. Biol. **199:** 587–596.

JAGIELLO, G., and J. S. FANG, 1987 Observations on chiasmata in mouse diplotene oocytes and spermatocytes. Cytologia (Tokyo) **52:** 283–292.

JENKINS, T. M., C. J. BASTEN and W. W. ANDERSON, 1996 Mitochondrial gene divergence of Columbian *Drosophila pseudoobscura*. Mol. Biol. Evol. **13:** 1266–1275.

KAPLAN, N. L., R. R. HUDSON and C. H. LANGLEY, 1989 The "hitchhiking effect" revisited. Genetics **123:** 887–899.

KIRBY, D. A., and W. STEPHAN, 1996 Multi-locus selection and the structure of variation at the *white* gene of *Drosophila melanogaster*. Genetics **144:** 635–645.

KIMURA, M., 1968 Evolutionary rate at the molecular level. Nature **217:** 624–626.

KIMURA, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.

KREITMAN, M., and H. AKASHI, 1995 Molecular evidence for natural selection. Annu. Rev. Ecol. Syst. **26:** 403–422.

LEWONTIN, R. C., 1964 The interaction of selection and linkage. I. General considerations; heterotic models. Genetics **43:** 419–434.

LEWONTIN, R. C., 1995 The detection of linkage disequilibrium in molecular sequence data. Genetics **140:** 377–388.

LEWONTIN, R. C., and J. L. HUBBY, 1966 A molecular approach to the study of genic heterozygosity in natural populations II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. Genetics **54:** 595–609.

LI, W. H., and D. GRAUR, 1991 *Fundamentals of Molecular Evolution*. Sinauer Associates, Sunderland, MA.

LI, W. H., and L.A. SADLER, 1991 Low nucleotide diversity in man. Genetics **129:** 513–523.

LINDSLEY, D. L., and G. G. ZIMM, 1992 *The Genome of Drosophila melanogaster*. Academic Press, San Diego.

MATZENBACH, B., Y. MAULET, L. SEFTON, B. COURTIER, P. AVNER *et al.*, 1994 Structural analysis of mouse glycine receptor α subunit genes. J. Biol. Chem. **269:** 2607–2612.

MAYNARD SMITH, J., and J. HAIGH, 1974 The hitch-hiking effect of a favourable gene. Genet. Res. **23:** 23–35.

MELTON, D. W., D. S. KONECKI, J. BRENNAND and C. T. CASKEY, 1984 Structure, expression, and mutation of the hypoxanthine phosphoribosyltransferase gene. Proc. Natl. Acad. Sci. USA **81:** 2147–2151.

MORIYAMA, E. N., and J. R. POWELL, 1996 Intraspecific nuclear DNA variation in Drosophila. Mol. Biol. Evol. **13:** 261–277.

NACHMAN, M. W., and C. F. AQUADRO, 1994 Polymorphism and divergence at the 5′ flanking region of the sex determining locus, *Sry*, in mice. Mol. Biol. Evol. **11:** 539–547.

NACHMAN, M. W., and J. B. SEARLE, 1995 Why is the house mouse karyotype so variable? Trends Ecol. Evol. **10:** 397–402.

NACHMAN, M. W., and G.A. CHURCHILL. 1996 Heterogeneity in rates of recombination across the mouse genome. Genetics **142:** 537–548.

NACHMAN, M. W., S. N. BOYER, J. B. SEARLE and C. F. AQUADRO, 1994a Mitochondrial DNA variation and the evolution of Robertsonian chromosomal races of house mice, *Mus domesticus*. Genetics **136:** 1105–1120.

NACHMAN, M. W., S. N. BOYER and C. F. AQUADRO, 1994b Non-neutral evolution of the NADH dehydrogenase subunit 3 gene in mice. Proc. Natl. Acad. Sci. USA **91:** 6364–6368.

NACHMAN, M. W., W. M. BROWN, M. STONEKING and C. F. AQUADRO, 1996 Nonneutral mitochondrial DNA variation in humans and chimpanzees. Genetics **142:** 953–963.

NEI, M., and W.-H. LI, 1979 Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. USA **76:** 5269–5273.

NEI, M., and A. K. ROYCHOUDHURY, 1982 Genetic relationship and evolution of human races. Evol. Biol. **14:** 1–59.

OHTA, T., 1992 The nearly neutral theory of molecular evolution. Annu. Rev. Ecol. Syst. **23:** 263–286.

POLANI, P. E., 1972 Centromere localization at meiosis and the position of chiasmata in the male and female mouse. Chromosoma **36:** 343–374.

RAND, D. M., and L. M. KANN, 1996 Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from Drosophila, mice, and humans. Mol. Biol. Evol. **13:** 735–748.

RICE, N. R., and N. A. STRAUSS, 1973 Relatedness of mouse satellite deoxyribonucleic acid to deoxyribonucleic acid of various Mus species. Proc. Natl. Acad. Sci. USA **70:** 3546–3550.

SAIKI, R. K., T. L. BUGAWAN, G. T. HORN, K. B. MULLIS and H. A. ERLICH, 1986 Analysis of enzymatically amplified Beta-globin and HLA-DQα DNA with allele-specific oligonucleotide probes. Nature **324:** 163–166.

SAIKI, R. K., D. H. GELFAND, S. STOFFEL, S. J. SCHARF, R. HIGUCHI *et al.*, 1988 Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239:** 487–491.

SAMBROOK, J., E. F. FRITSCH and T. MANIATIS, 1989 *Molecular Cloning: A Laboratory Manual*, Ed. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

SANGER, F., S. NICKLON and A. R. COULSON, 1977 DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:** 5463–5467.

SCHAEFFER, S. W., and E. L. MILLER, 1993 Estimates of linkage disequilibrium and the recombination parameter determined from segregating nucleotide sites in the alcohol dehydrogenase region of Drosophila pseudoobscura. Genetics **135:** 541–552.

SELANDER, R. K., W. G. HUNT and S. Y. YANG, 1969 Protein polymorphism and genic heterozygosity in two European subspecies of the house mouse. Evolution **23:** 379–390.

SHARP, P. M., and W. H. LI, 1989 On the rate of DNA sequence evolution in Drosophila. J. Mol. Evol. **28:** 398–402.

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.

SIMONSEN, K. L., G. A. CHURCHILL and C. F. AQUADRO, 1995 Properties of statistical tests of neutrality for DNA polymorphism data. Genetics **141:** 413–439.

SINGH, R. S., and L. R. RHOMBERG, 1987 A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*. II. Estimates of heterozygosity and patterns of geographic differentiation. Genetics **117:** 255–271.

SLATKIN, M., 1995 Hitchhiking and associative overdominance at a microsatellite locus. Mol. Biol. Evol. **12:** 473–480.

SNEAD, M. L., E. C. LAU, M. ZEICHNER-DAVID, A. C. FINCHAM, S. L. C. WOO *et al.*, 1985 DNA sequence for cloned cDNA for murine amelogenin reveal the amino acid sequence for enamel-specific protein. Biochem. Biophys. Res. Commun. **129:** 812–818.

STEPHAN, W., 1995 An improved method for estimating the rate of fixation of favorable mutations based on DNA polymorphism data. Mol. Biol. Evol. **12:** 959–962.

TAJIMA, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585–595.

TAKAHATA, N., 1994 Repeated failures that led to the eventual success in human evolution. Mol. Biol. Evol. **11:** 803–805.

TUCKER, P. K., B. K. LEE and E. M. EICHER, 1989 *Y* chromosome evolution in the subgenus Mus (genus Mus). Genetics **122:** 169–179.

WATERSON, G. A., 1975 On the number of segregating sites in genetical models without recombination. Theoret. Popul. Biol. **7:** 256–276.

WATT, W. B., 1983 Adaptation at specific loci. II. Demographic and biochemical elements in the maintenance of the *Colias* PGI polymorphism. Genetics **103:** 691–724.

WATT, W. B., K. DONOHUE and P. A. CARTER, 1996 Adaptation at specific loci. VI. Divergence *vs.* parallelism of polymorphic allozymes in molecular function and fitness-component effects among *Colias* species (Lepidoptera, Pieridae). Mol. Biol. Evol. **13:** 699–709.

WIEHE, T. H. E., and W. STEPHAN, 1993 Analysis of a genetic hitchhiking model, and its application to DNA polymorphism data from *Drosophila melanogaster*. Mol. Biol. Evol. **10:** 842–855.

Communicating editor: A. G. CLARK