Y Chromosome Variation of Mice and Men

Michael W. Nachman

Department of Ecology and Evolutionary Biology, University of Arizona

DNA sequences from the nonrecombining portion of the Y chromosome were compared with autosomal and X-linked sequences from mice and humans to test the neutral prediction that ratios of polymorphism to divergence are the same for different genes. Intraspecific variation within *Mus domesticus* was compared with divergence between *M. domesticus* and *Mus caroli* for *Sry*, a region 5′ to *Sry*, and four X-linked genes, *Hprt, Plp, Amg,* and *Glra2*. None of these comparisons revealed significantly reduced variation on the Y chromosome. Intraspecific variation within humans was compared with divergence between humans and chimpanzees for three Y-linked loci (*Zfy,* the YAP region, and the *Sry* region), seven X-linked loci (*Il2rg, Plp, Hprt, Gk, Ids, Pdha1,* and *Dmd*), and the β-globin locus on chromosome 11. In these comparisons, the observed level of variation on the human Y chromosome was slightly lower than expected, but was significantly lower in only one case (*Sry* region vs. *Dmd*). These results suggest that the levels of variability on the Y chromosome in mice and humans are close to expected values given the effective population size and mutation rates for these loci. There is at most only a modest reduction in variability that may be attributed to natural selection (either genetic hitchhiking or background selection).

Introduction

The nonrecombining portion of the mammalian Y chromosome (NRY) may be affected by a number of evolutionary processes that are typically uncommon in genomic regions that undergo recombination. Selection acting at any site within the NRY will affect the entire NRY. Thus, genetic hitchhiking (Maynard Smith and Haigh 1974; Kaplan, Hudson, and Langley 1989) and background selection (Charlesworth, Morgan, and Charlesworth 1993) may be important in shaping patterns of variation in this region of the genome. Genetic hitchhiking refers to the fixation of an advantageous mutation and the associated fixation of linked, neutral mutations. The strength of this effect will depend in part on the average selection coefficient for adaptive mutations, the rate of adaptive evolution, and the rate of recombination (Wiehe and Stephan 1993; Stephan 1995). Background selection (Charlesworth, Morgan, and Charlesworth 1993; Charlesworth 1994; Hudson and Kaplan 1995) is the selective removal of deleterious mutations from a population and the associated removal of linked neutral mutations. The strength of this effect depends on the deleterious mutation rate for the region in question, the average selection coefficient and dominance factor, and the frequency of recombination. Both processes may contribute to reduced levels of nucleotide polymorphism in regions of the genome with little recombination (e.g., Berry, Ajioka, and Kreitman 1991; Aguade and Langley 1994; Aquadro, Begun, and Kindahl 1994; Wayne and Kreitman 1996; Stephan et al. 1998).

There are several reasons for questioning the importance of these theoretical expectations for reduced variability due to selection for the NRY. First is the observation that there are relatively few functional genes and, therefore, presumably few targets for selection on the NRY. A systematic search for functional genes on the human NRY revealed 12 novel genes, bringing the total to 20 (Lahn and Page 1997). The human NRY contains approximately 60 Mb of DNA (Hammer and Zegura 1996) and thus has a gene density of 0.33 genes/Mb. The average gene density for the entire human genome (3×10^9 bp) can be calculated from the estimated total of 70,000 genes (Bird 1995) as 23 genes/Mb. Thus, the NRY gene density may be approximately two orders of magnitude lower than average. Second, the effective population size for the NRY is one-fourth the effective population size of the autosomes. The efficacy of selection is weaker in smaller populations (Kimura 1983). Even if average selection coefficients for both adaptive and deleterious mutations are the same on the NRY and on autosomes, selection may play less of a role on the NRY because of differences in effective population size.

Population surveys of Y chromosome sequence variation have been conducted for humans (Dorit, Akashi, and Gilbert 1995; Hammer 1995; Whitfield, Sulston, and Goodfellow 1995) and house mice (Lundrigan and Tucker 1994; Nachman and Aquadro 1994) and have typically revealed little variability at the nucleotide level. However, this lack of variability has generally not been evaluated in the context of polymorphism and divergence at other loci, and thus it has been impossible to separate population-level effects from locus-specific effects (Hudson, Kreitman, and Aguade 1987).

Here, I investigate whether there is evidence for selection on the NRY in humans and house mice by comparing levels of polymorphism and divergence at Y-linked loci with polymorphism and divergence at autosomal and X-linked loci undergoing different amounts of recombination. A neutral model predicts that the ratio of polymorphism to divergence should be the same for different genes, since both are a function of the neutral mutation rate (Hudson, Kreitman, and Aguade 1987). Selection, in contrast, can lead to an uncoupling of levels of polymorphism and divergence (Kaplan, Hudson, and Langley 1989). Previous studies using this approach

Abbreviation: NRY, nonrecombining portion of the Y chromosome.

Key words: evolution, *Mus domesticus,* genetic hitchhiking, background selection, the neutral theory, Y chromosome.

Address for correspondence and reprints: Michael W. Nachman, Department of Ecology and Evolutionary Biology, Biosciences West Building, University of Arizona, Tucson, Arizona 85721. E-mail: Nachman@u.arizona.edu.

Mol. Biol. Evol. 15(12):1744–1750. 1998
 ISSN: 0737-4038

Table 1
Population Samples of Y Chromosome Nucleotide Variation in Mice

Locus[a]	*n*	Length (bp)	*S*	π (%)	θ (%)	Tajima's *D*	Fu and Li's *D*	Fu's F_S	Divergence[b] (%)	Reference
Sry 5′ ..	20	1,063	2	0.041	0.053	−0.483	−0.634	−0.464	4.18	Nachman and Aquadro (1994)
Sry	6	515	0	0.000	0.000	—	—	—	2.91	Lundrigan and Tucker (1994)

[a] *Sry* 5′ includes only noncoding sequence, and *Sry* includes only coding sequence.
[b] Divergence is between *Mus domesticus* and *Mus caroli.*

to detect selection on the NRY (Nachman and Aquadro 1994; Hammer 1995) involved comparisons between mitochondrial loci and Y-linked loci and were complicated by the different inheritance patterns of these loci and the fact that neither the NRY nor mtDNA undergoes recombination. In this paper, I demonstrate that there is only a modest reduction in variability on the NRY compared with loci that undergo recombination in mice and humans.

Materials and Methods

NRY Sequences

Mus domesticus Y chromosome sequences (table 1) were taken from Lundrigan and Tucker (1994) and Nachman and Aquadro (1994). Lundrigan and Tucker (1994) sequenced 515 bp of the *Sry* coding regions of six *M. domesticus,* and Nachman and Aquadro (1994) sequenced 1,063 bp of noncoding DNA 5′ to *Sry* in 20 *M. domesticus.* Both samples contain wild-caught mice from natural populations in Western Europe. Divergence data are from comparisons between *M. domesticus* and *M. caroli.* These species are not sister taxa but are closely related, with an average sequence divergence of approximately 3% for X-linked introns (Nachman 1997). The *M. caroli Sry* sequence is from Lundrigan and Tucker (1994). For the region 5′ to *Sry,* a fragment was PCR-amplified for *M. caroli* that corresponds to positions 6643–7920 in the sequence of Gubbay et al. (1992). *Mus caroli* DNA was obtained commercially from the Jackson Laboratory, and PCR reaction conditions were as described in Nachman and Aquadro (1994). The double-stranded PCR product was manually sequenced for 901 bp in both directions using the dideoxy chain termination method (Sanger, Nicklen, and Coulson 1977) with Sequenase 2 enzyme and kit (USB) according to the protocol supplied by the manufacturer. This sequence has been deposited in GenBank (accession numbers AF087040, AF087041).

Human Y chromosome sequences (table 2) were taken from Dorit, Akashi, and Gilbert (1995), Hammer (1995), and Whitfield, Sulston, and Goodfellow (1995). Dorit, Akashi, and Gilbert (1995) sequenced 729 bp of the *Zfy* intron for 38 individuals from around the world, Hammer (1995) sequenced 2,638 bp of noncoding DNA in the Y Alu-insertion polymorphism (YAP) region in a worldwide sample of 16 individuals, and Whitfield, Sulston, and Goodfellow (1995) sequenced 18,300 bp of noncoding DNA in the *Sry* region for a sample of five individuals (three Africans, one European, and one Melanesian). Divergence was calculated for each of these regions by comparing human and chimpanzee sequences (Dorit, Akashi, and Gilbert 1995; Hammer 1995; Whitfield, Sulston, and Goodfellow 1995).

Autosomal and X-Linked Sequences

Sequences for *M. domesticus* X-linked loci (table 3) are from Nachman (1997). Introns of four X-linked genes (*Hprt, Plp, Glra2,* and *Amg*) were sequenced for 10 male *M. domesticus* from wild populations in Italy. An average of 1,506 bp was sequenced for each locus. Rates of recombination are approximately twofold higher at *Amg* and *Glra2* than at *Hprt* and *Plp.* Divergence for these same loci was calculated from comparisons between *M. domesticus* and *M. caroli* sequences (Nachman 1997).

Polymorphism data for human autosomal and X-linked sequences (table 4) come from Harding et al. (1997) and Nachman et al. (1998). Harding et al. (1997) sequenced 2,320 bp of the β-globin locus on chromosome 11 in a worldwide sample of 349 alleles. Nachman et al. (1998) sequenced 11,365 bp from introns of seven X-linked genes (*Il2rg, Plp, Hprt, Gk, Ids, Pdha1,* and *Dmd*) in a worldwide sample of 10 males. These seven genes experience different rates of recombination, and there is a positive correlation between nucleotide heterozygosity and recombination rate (Nachman et al. 1998). The lowest recombination rate is observed for *Il2rg* (0.27 cM/Mb), and the highest rate is observed for *Dmd* (7.74 cM/Mb). Divergence for both data sets comes from comparison of human and chimpanzee sequences.

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 two sequences randomly drawn from a sample, and θ is based on the proportion 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, where N_e is the effective population size and μ is the neutral mutation rate. Tajima's (1989) *D,* Fu and Li's (1993) *D,* and Fu's (1997) F_S statistics were calculated to test for deviations from a neutral frequency distribution for the NRY loci. Ratios of polymorphism to divergence were compared with the expectations under a neutral model using the HKA test (Hudson, Kreitman, and Aguade 1987). This test was modi-

Table 2
Population Samples of Y Chromosome Nucleotide Variation in Humans

Locus[a]	*n*	Length (bp)	*S*	π (%)	θ (%)	Tajima's *D*	Fu and Li's *D*	Fu's F_S	Divergence[b] (%)	Reference
Zfy intron	38	729	0	0.000	0.000	—	—	—	0.686	Dorit, Akashi, and Gilbert (1995)
YAP region	16	2,638	3	0.037	0.034	0.231	−0.106	−0.297	1.895	Hammer (1995)
Sry region	5	18,300	3	0.008	0.008	−0.145	−0.505	−1.648	1.320	Whitfield, Sulston, and Goodfellow (1995)

[a] The YAP region and the *Sry* region contain noncoding sequence only.
[b] Divergence is between *Homo* and *Pan*.

Table 3
Population Samples of X-Linked Nucleotide Variation in Mice

Locus[a]	*n*	Length (bp)	*S*	π (%)	θ (%)	Divergence[b] (%)
Hprt	10	1,290	1	0.016	0.027	2.71
Plp	10	1,595	0	0.000	0.000	4.39
Amg	10	1,141	4	0.160	0.124	2.02
Glra2	10	1,996	8	0.135	0.142	2.81

[a] All sequences are from introns; data are from Nachman (1997). Genes are listed in order of increasing recombination rate.
[b] Divergence is between *Mus domesticus* and *Mus caroli*.

fied to account for a fourfold difference in effective population size for Y–autosome comparisons and for a threefold difference in effective population size for Y–X comparisons. These calculations assume a sex ratio of one.

Results

House Mice

Two studies have measured Y chromosome sequence variation in natural populations of *Mus* (table 1). Both show low levels of variability ($\pi = 0$, Lundrigan and Tucker 1994; $\pi = 0.041\%$, Nachman and Aquadro 1994). No polymorphisms were detected in *Sry*, and only two segregating sites were detected in the region 5′ to *Sry*. Of these two, one was a singleton (5%) and the other was present in 4 out of 20 individuals (20%), resulting in values of Tajima's *D*, Fu and Li's *D*, and Fu's F_S close to the neutral expectation of zero (table 1).

HKA comparisons between NRY sequences and X-linked sequences for mice are given in Table 5. Each NRY locus was compared with each X-linked locus (with the exception of *Sry* vs. *Plp*, for which neither locus showed any polymorphisms). Only one of the seven tests showed significantly different ratios of polymorphism and divergence: *Sry* 5′ vs. *Plp* ($P = 0.049$ after correction for multiple tests). In this comparison, significantly more polymorphism than expected was ob-

Table 4
Population Samples of Autosomal and X-Linked Nucleotide Variation in Humans

Locus[a]	*n*	Length (bp)	*S*	π (%)	θ (%)	Divergence[b] (%)
β-globin	349	2,320	19	0.180	0.130	1.60
Il2rg	10	1,147	0	0.000	0.000	0.78
Plp	10	769	2	0.095	0.092	0.65
Hprt	10	2,485	4	0.038	0.057	0.97
Gk	10	1,861	1	0.019	0.019	0.64
Ids	10	1,909	0	0.000	0.000	0.26
Pdha1	10	1,657	5	0.137	0.107	0.84
Dmd	10	1,537	8	0.187	0.184	0.85

[a] β-globin data are from Harding et al. (1997); other genes are from Nachman et al. (1998) and include intron sequences only, listed in order of increasing recombination rate.
[b] Divergence is between *Homo* and *Pan*.

Table 5
HKA Tests Between Y-Linked and X-Linked Loci in Mice

Comparison	θ_1(%)	θ_2(%)	S_1/ES$_1$	S_2/ES$_2$	D_1/ED$_1$	D_2/ED$_2$	T	χ^2	P
Sry 5′–Hprt	0.13	0.07	2/1.0	1/2.0	40/41.0	35/34.0	35.6	1.26	NS
Sry 5′–Plp	0.05	0.05	2/0.4	0/1.6	40/41.6	70/68.4	90.4	7.24	0.007
Sry 5′–Glra2	0.31	0.18	2/2.3	8/7.7	40/39.7	56/56.3	14.8	0.05	NS
Sry 5′–Amg	0.33	0.14	2/2.5	4/3.5	40/39.5	23/23.5	13.6	0.15	NS
Sry–Hprt.....	0.04	0.03	0/0.1	1/0.9	15/14.9	35/35.1	82.3	0.11	NS
Sry–Glra2.....	0.18	0.18	0/0.5	8/7.5	15/14.5	56/56.5	15.3	0.50	NS
Sry–Amg.....	0.20	0.14	0/0.6	4/3.4	15/14.4	23/23.6	13.9	0.60	NS

NOTE.—Subscripts denote locus 1 (NRY) and locus 2 (X-linked), respectively. S = observed segregating sites; ES = expected segregating sites; D = observed divergence; ED = expected divergence; T = divergence time.

served at *Sry* 5′, and significantly less polymorphism than expected was observed at *Plp*. *Plp* lies in a region of the X chromosome with a low recombination rate (approximately 0.25 cM/Mb) and also shows significantly reduced variability when compared with *Glra2* or *Amg* (Nachman 1997). None of the comparisons in table 5 provide evidence for a significant reduction in variability on the NRY relative to other genomic regions.

Humans

Three studies have measured DNA sequence variation at the NRY in population samples of humans (table 2). These three studies consist of different, and, in some respects, complementary, sampling strategies. Dorit, Akashi, and Gilbert (1995) sequenced a relatively small number of bases (729) at *Zfy* for a relatively large number of individuals (38), while Whitfield, Sulston, and Goodfellow (1995) sequenced a large number of bases (18,300) in the *Sry* region for a small number of individuals (5), and Hammer's (1995) survey of the YAP region is intermediate in both respects. All three studies found low levels of nucleotide diversity (*Zfy*, π = 0; YAP, π = 0.037%; *Sry* region, π = 0.008%), and all were lower than the average value of 0.11% reported by Li and Sadler (1991) for autosomal fourfold-degenerate sites. For the YAP region, Tajima's D was slightly positive, while Fu and Li's D and Fu's F_S were slightly negative; all three were close to the neutral expectation of zero (table 2). For the *Sry* region, all three tests were slightly but not significantly negative. The power of these tests is low with small samples or few segregating sites (Braverman et al. 1995; Simonsen, Churchill, and Aquadro 1995; Fu 1997).

HKA comparisons are presented in table 6. Each NRY locus was compared with each X-linked or autosomal locus except when neither locus under comparison was polymorphic. Of the 22 HKA comparisons, three were significant and showed reduced variation at *Sry* relative to X-linked genes with moderate to high rates of recombination (*Hprt, Pdha1,* and *Dmd*). After correcting for multiple tests, only the *Sry–Dmd* comparison remained significant (P = 0.044). *Dmd* has the highest rate of recombination and is also the most variable gene surveyed. In five of the 22 comparisons, the

Table 6
HKA Tests Between Y-Linked and Other Loci in Humans

Comparison	θ_1(%)	θ_2(%)	S_1/ES$_1$	S_2/ES$_2$	D_1/ED$_1$	D_2/ED$_2$	T	χ^2	P
Zfy–β globin	0.05	0.12	0/0.4	19/18.6	5/4.6	37/37.4	11.9	0.42	NS
Zfy–Plp	0.08	0.09	0/0.6	2/1.4	5/4.4	5/5.6	7.6	0.89	NS
Zfy–Hprt	0.05	0.07	0/0.4	4/3.6	5/4.6	24/24.4	13.4	0.40	NS
Zfy–Gk.....	0.02	0.02	0/0.2	1/0.8	5/4.8	12/12.2	30.2	0.20	NS
Zfy–Pdha1.....	0.09	0.12	0/0.7	5/4.3	5/4.3	14/14.7	6.5	0.83	NS
Zfy–Dmd	0.14	0.21	0/1.1	8/6.9	5/3.9	13/14.1	3.6	1.40	NS
YAP–β globin....	0.15	0.12	3/3.4	19/18.6	50/49.6	37/37.4	11.9	0.04	NS
YAP–*Il2rg*.....	0.10	0.04	3/2.1	0/0.9	50/50.9	9/8.1	19.4	1.10	NS
YAP–*Plp*.....	0.18	0.07	3/3.9	2/1.1	50/49.1	5/5.9	10.3	0.82	NS
YAP–*Hprt*.....	0.15	0.07	3/3.2	4/3.8	50/49.8	24/24.2	12.7	0.01	NS
YAP–*Gk*	0.12	0.04	3/2.6	1/1.4	50/50.4	12/11.6	16.2	0.19	NS
YAP–*Ids*	0.11	0.01	3/2.5	0/0.5	50/50.5	5/4.5	16.8	0.65	NS
YAP–*Pdha1*	0.21	0.10	3/4.5	5/3.5	50/48.5	14/15.5	8.7	0.87	NS
YAP–*Dmd*.....	0.28	0.15	3/6.1	8/4.9	50/46.9	13/16.1	6.2	2.49	NS
Sry–β globin.....	0.08	0.10	3/7.2	19/14.8	207/202.8	37/41.2	16.9	1.94	NS
Sry–Il2rg.....	0.03	0.02	3/2.6	0/0.4	207/207.4	9/8.6	47.8	0.40	NS
Sry–Plp	0.05	0.03	3/4.5	2/0.5	207/205.5	5/6.5	27.4	4.82	0.027
Sry–Hprt	0.05	0.04	3/4.9	4/2.1	207/205.1	24/25.9	25.3	1.63	NS
Sry–Gk.....	0.03	0.02	3/3.3	1/0.7	207/206.7	12/12.3	37.6	0.16	NS
Sry–Ids	0.03	0.01	3/2.8	0/0.2	207/207.2	5/4.8	45.1	0.23	NS
Sry–Pdha1.....	0.07	0.05	3/6.2	5/1.8	207/203.8	14/17.2	19.7	5.48	0.020
Sry–Dmd.....	0.09	0.08	3/8.4	8/2.6	207/201.6	13/18.4	14.3	10.12	0.002

NOTE.— Subscripts denote locus 1 (NRY) and locus 2 (X-linked or autosomal), respectively. S = observed segregating sites; ES = expected segregating sites; D = observed divergence; ED = expected divergence; T = divergence time.

NRY locus showed more variation than expected, and in the remaining 17 comparisons, the NRY locus showed less variation than expected. In all cases (except that of *Sry–Dmd*), these differences were slight. The number of comparisons in which the NRY locus showed more or less than expected is not different from random (Fisher's exact test, $P = 0.11$).

Discussion

For both mice and humans, there is at most a modest reduction in nucleotide heterozygosity on the NRY relative to nuclear loci which undergo recombination. Previous studies which have argued that the Y has unusually low levels of variability (Lundrigan and Tucker 1994; Dorit, Akashi, and Gilbert 1995; Whitfield, Sulston, and Goodfellow 1995; Burrows and Ryder 1997) were not based on ratios of polymorphism to divergence at multiple loci and thus did not fully account for differences in effective population size or neutral mutation rate.

The small number of segregating sites at nuclear genes in general makes it difficult to conduct statistical tests. In particular, the power of the single-locus tests presented here is low, given the level of variability observed and the small sample sizes (Braverman et al. 1995; Simonsen, Churchill, and Aquadro 1995; Fu 1997). It is also likely that the HKA test lacks power in such situations, although the power of the HKA test has not been investigated. Inability to reject the null model in general should not be construed as evidence that a locus is evolving neutrally; it may simply be that the tests employed are not sensitive enough to detect the effects of weak selection (Wayne and Simonsen 1998). One way to assess the power of the HKA test in these comparisons is to ask whether significant results would have been obtained if no variation had been observed at the *Sry* 5′ region in mice or the YAP region in humans. Under this scenario, the human YAP region would have rejected neutrality, but the mouse *Sry* 5′ region would not have rejected neutrality in comparisons against high-recombination X-linked loci (*Dmd* and *Pdha1* in humans, *Amg* and *Glra2* in mice). Thus, the power to detect a significant reduction in variability at the NRY in mice is limited. Nonetheless, the general agreement between observed and expected values for the NRY loci in tables 5 and 6 and the significant excess of variation observed for the mouse *Sry* 5′ region when compared with *Plp* demonstrate that there is no consistent reduction in variability at NRY loci such as that seen at loci in other regions of low recombination (e.g., Berry, Ajioka, and Kreitman 1991; Begun and Aquadro 1992; Aguade and Langley 1994; Wayne and Kreitman 1996; Nachman 1997; Stephan et al. 1998).

The frequency distribution tests generally yielded negative values, though none were significant (Tables 1 and 2). Negative test values are expected with either directional selection or a recent population expansion. Similar negative values are also generally observed for mitochondrial genes in humans (e.g., Whittam et al. 1986; Excoffier 1990), but not for most autosomal loci (Hey 1997; Nachman et al. 1998).

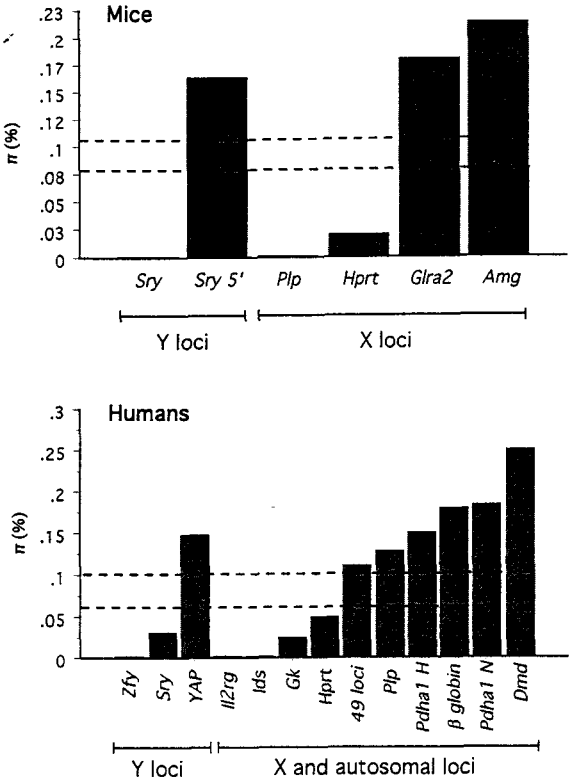

FIG. 1.—Nucleotide diversity at NRY, X-linked, and autosomal loci in mice (top) and humans (bottom). Nucleotide diversity for Y-linked loci was multiplied by 4 and nucleotide diversity at X-linked loci was multiplied by 4/3 to account for differences in effective population size. The lower dotted line in each plot gives an average value of π for NRY loci and the upper dotted line gives an average value of π for other loci. Loci are as in tables 1–4; "Pdha1 H" is from Hey (1997) and "49 loci" data are from Li and Sadler (1991). Loci are listed in order of increasing nucleotide diversity. The 49 loci of Li and Sadler (1991) are mostly autosomal, and the β-globin locus is on chromosome 11.

One of the HKA comparisons involving mice and one of the HKA comparisons involving humans were significant. The one significant comparison for mice (*Sry* 5′ vs. *Plp*) revealed an excess of variability on the NRY, while the one significant comparison for humans (*Sry* vs. *Dmd*) revealed a deficiency of variability on the NRY. For both mice and humans, there is a range of variation observed at different NRY loci, and this range includes values that are both lower and higher than the average level of variability at nuclear loci which experience meiotic recombination (fig. 1). The average level of variability on the NRY for both mice and humans is slightly but not significantly lower than the average level of variability at X-linked and autosomal loci (once differences in effective population size are factored out; fig. 1). Thus, while levels of variability on the Y are low, they are well within the range of variation observed at other nuclear loci and are not uniformly low, as might be expected of genes in regions without recombination.

Events other than selection may also lead to a rejection of the null model in an HKA test involving the NRY. If there is greater variance in male reproductive

success than in female reproductive success, male effective population size will be smaller than female effective population size. This would result in reduced variation on the Y- relative to the X-linked genes or autosomal loci.

The comparison between the *Sry* region and *Dmd* for humans revealed less variation at *Sry* than expected. Because the NRY evolves as a single linkage group, all NRY loci are expected to share the same evolutionary history. It is thus somewhat surprising that the *Sry* region rejects the null model, while neither *Zfy* nor the YAP region show a significant reduction in variability compared to *Dmd*. What might account for this discrepancy? *Zfy*, the YAP region, and the *Sry* region were all sampled from different individuals and from different ethnic groups. While all three samples and the *Dmd* sample contained African and non-African individuals, the *Sry* sample only contained five individuals and may represent a nonrepresentative subset of Y chromosome variation in humans.

The results presented here may be contrasted with results for *Drosophila melanogaster*, for which every gene in a region of low recombination appears to have reduced variability. In particular, Berry, Ajioka, and Kreitman (1991) documented reduced variation at the cubitus interruptus locus on the small fourth chromosome. This chromosome represents about 1% of the *Drosophila* genome and is thought to contain approximately 75 genes (Hochman 1976). A recent search for functional genes on the NRY in humans revealed a total of only 20 loci (Lahn and Page 1997).

Is it possible that the absence of reduced variability on the NRY is simply a result of few targets for selection, either for genetic hitchhiking or for background selection? This is a difficult question to answer given our limited knowledge of the adaptive substitution rate and the deleterious mutation rate. Aquadro and Begun (1993), using data from 20 loci throughout the *D. melanogaster* genome, estimate one adaptive fixation every 3×10^4 generations for the entire *D. melanogaster* genome. Expressed per locus (assuming roughly 10,000 loci in the *Drosophila* genome), we obtain one adaptive fixation per locus every 3×10^8 generations. We can estimate the expected coalescent time under neutrality for a sample of n alleles from humans. For a haploid, hemizygous locus such as the NRY, it is $N(1 - 1/n)$ generations, where N is the effective species population size (Tajima 1983). Given an estimate of $N = 10^4$ for humans (Hammer 1995), we see that under neutrality we expect a sample of 16 alleles to coalesce to a common ancestral allele in 9.4×10^3 generations. If we assume that there are 20 loci on the human NRY (Lahn and Page 1997) and if we use the adaptive substitution rate estimates from *Drosophila*, we expect an adaptive fixation for the human NRY approximately every 1.5×10^7 generations. This number is three orders of magnitude greater than the average coalescent time under neutrality. This calculation, while extremely rough, suggests that the absence of evidence for hitchhiking may be simply due to the small number of genes on the NRY.

A similar argument can be made for deleterious mutations. In the simplest model, background selection is expected to reduce heterozygosity by a fraction, $f_0 = \exp(-U/2hs)$, where U is the deleterious mutation rate for the region in question, h is the average dominance factor, and s is the average selection coefficient for deleterious mutations (Charlesworth, Morgan, and Charlesworth 1993). If the deleterious mutation rate is proportional to the number of loci present, and if the genomic deleterious mutation rate for humans is approximately 1 (Kondrashov and Crow 1993; Drake et al. 1998), then the deleterious mutation rate for the NRY is approximately $20/70{,}000 = 2.86 \times 10^{-4}$. Assuming $hs = 0.02$ (Charlesworth, Morgan, and Charlesworth 1993), $f_0 = 0.993$, which in practical terms will be indistinguishable from heterozygosity in the absence of background selection.

Acknowledgments

I thank M. F. Hammer and the members of my lab for discussion. This work was supported by NSF.

LITERATURE CITED

AGUADE, M., and C. H. LANGLEY. 1994. Polymorphism and divergence in regions of low recombination in Drosophila. Pp. 67–76 *in* B. GOLDING, ed. Non-neutral evolution. Theories and molecular data. Chapman and Hall, New York.

AQUADRO, C. F., and D. J. BEGUN. 1993. Evidence for and implications of genetic hitchhiking in the *Drosophila* genome. Pp. 159–178 *in* N. TAKAHATA and A. G. CLARK, eds. Mechanisms of molecular evolution. Sinauer, Sunderland, Mass.

AQUADRO, C. F., D. J. BEGUN, and E. C. KINDAHL. 1994. Selection, recombination, and DNA polymorphism in Drosophila. Pp. 46–56 *in* B. GOLDING, ed. Non-neutral evolution. Theories and molecular data. Chapman and Hall, New York.

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.

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.

BIRD, A. P. 1995. Gene number, noise reduction and biological complexity. Trends Genet. **11**:94–100.

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 polymorphims. Genetics **140**:783–796.

BURROWS, W., and O. A. RYDER. 1997. Y-chromosome variation in great apes. Nature **385**:125–126.

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.

DORIT, R. L., H. AKASHI, and W. GILBERT. 1995. Absence of polymorphism at the *Zfy* locus on the human Y chromosome. Science **268**:1183–1185.

DRAKE, J. W., B. CHARLESWORTH, D. CHARLESWORTH, and J. F. CROW. 1998. Rates of spontaneous mutation. Genetics **148**:1667–1686.

EXCOFFIER, L. 1990. Evolution of human mitochondrial DNA: evidence for departure from a pure neutral model of populations at equilibrium. J. Mol. Evol. **30**:125–139.

FU, Y. X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking, and background selection. Genetics **147**:915–925.

FU, Y. X., and W.-H. LI. 1993. Statistical tests of neutrality of mutations. Genetics **133**:693–709.

GUBBAY, J., N. VIVIA, A. ECONOMOU, D. JACKSON, P. GOODFELLOW, and R. LOVELL-BADGE. 1992. Inverted repeat structure of the Sry locus in mice. Proc. Natl. Acad. Sci. USA **89**:7953–7957.

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

HAMMER, M. F., and S. L. ZEGURA. 1996. The role of the Y chromosome in human evolutionary studies. Evol. Anthropol. **5**:116–134.

HARDING, R. M., S. M. FULLERTON, R. C. GRIFFITHS, J. BOND, M. J. COX, J. A. SCHNEIDER, D. S. MOULIN, and J. B. CLEGG. 1997. Archaic African *and* Asian Lineages in the genetic ancestry of modern humans. Am. J. Hum. Genet. **60**:772–789.

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

HOCHMAN, B. 1976. The fourth chromosome of *Drosophila melanogaster*. Pp. 903–928 *in* M. ASHBURNER and E. NOVITSKI, eds. The genetics and biology of Drosophila. Vol. 1b. Academic Press, 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.

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

KIMURA, M. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge, England.

KONDRASHOV, A. S., and J. F. CROW. 1993. A molecular approach to estimating the human deleterious mutation rate. Hum. Mutat. **2**:229–234.

LAHN, B. T., and D. C. PAGE. 1997. Functional coherence of the human Y chromosome. Science **278**:675–680.

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

LUNDRIGAN, B. L., and P. K. TUCKER. 1994. Tracing paternal ancestry in mice, using the Y-linked sex-determining locus, *Sry*. Mol. Biol. Evol. **11**:483–492.

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

NACHMAN, M. W. 1997. Patterns of DNA variability at *X*-linked loci in *Mus domesticus*. Genetics **147**:1303–1316.

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., V. L. BAUER, S. L. CROWELL, and C. F. AQUADRO. 1998. DNA variability and recombination rates at *X*-linked loci in humans. Genetics (in press).

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.

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

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.

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.

STEPHAN, W., L. XING, D. A. KIRBY, and J. M. BRAVERMAN. 1998. A test of the background selection hypothesis based on nucleotide data from Drosophila ananassae. Proc. Natl. Acad. Sci. USA **95**:5649–5654.

TAJIMA, F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics **105**:437–460.

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

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

WAYNE, M. L., and M. KREITMAN. 1996. Reduced variation at concertina, a heterochromatic locus in Drosophila. Genet. Res. Camb. **68**:101–108.

WAYNE, M. L., and K. L. SIMONSEN. 1998. Statistical tests of neutrality in the age of weak selection. Trends Ecol. Evol. **13**:236–240.

WHITFIELD, L. S., J. E. SULSTON, and P. E. GOODFELLOW. 1995. Sequence variation of the human Y chromosome. Nature **378**:379–380.

WHITTAM, T. S., A. G. CLARK, M. STONEKING, R. L. CANN, and A. C. WILSON. 1986. Allelic variation in human mitochondrial genes based on patterns of restriction site polymorphism. Proc. Natl. Acad. Sci. USA **83**:9611–9615.

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.

DAVID M. RAND, reviewing editor

Accepted September 15, 1998