ELSEVIER

Gene 300 (2002) 31–42

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

Natural selection at linked sites in humans

Bret A. Payseur*, Michael W. Nachman

Department of Ecology and Evolutionary Biology, Biosciences West Building, University of Arizona, Tucson, AZ 85721, USA

Received 21 December 2001; received in revised form 24 May 2002; accepted 17 July 2002

Abstract

Theoretical and empirical work indicates that patterns of neutral polymorphism can be affected by linked, selected mutations. Under background selection, deleterious mutations removed from a population by purifying selection cause a reduction in linked neutral diversity. Under genetic hitchhiking, the rise in frequency and fixation of beneficial mutations also reduces the level of linked neutral polymorphism. Here we review the evidence that levels of neutral polymorphism in humans are affected by selection at linked sites. We then discuss four approaches for distinguishing between background selection and genetic hitchhiking based on (i) the relationship between polymorphism level and recombination rate for neutral loci with high mutation rates, (ii) relative levels of variation on the X chromosome and the autosomes, (iii) the frequency distribution of neutral polymorphisms, and (iv) population-specific patterns of genetic variation. Although the evidence for selection at linked sites in humans is clear, current methods and data do not allow us to clearly assess the relative importance of background selection and genetic hitchhiking in humans. These results contrast with those obtained for *Drosophila*, where the signals of positive selection are stronger.

Keywords: Nucleotide polymorphism; Background selection; Genetic hitchhiking; Selective sweeps; Recombination rate

1. Introduction

The observation that only a small fraction of nucleotide sites in the human genome code for genes suggests that the evolutionary dynamics of most human DNA polymorphisms may be governed primarily by an interaction between the forces of mutation and genetic drift. This 'neutral theory' (Kimura, 1983) has been largely successful in explaining many patterns of molecular evolution in humans and other species. However, because of linkage, sites may not evolve independently of one another. Consequently, in many regions of the genome, polymorphisms that do not affect fitness may occasionally be linked to those that do. Theoretical and empirical work indicates that this linkage can substantially affect levels of neutral polymorphism, whether selection acts primarily against deleterious mutations (as envisioned by Kimura) or often favors advantageous mutations. This result means that we can use patterns of polymorphism at neutral sites to detect selection acting at the molecular level.

Neutral variants linked to deleterious mutations will indirectly experience selective pressure to be removed from populations. This idea, termed background selection (Charlesworth et al., 1993), predicts that genomic regions of reduced recombination will exhibit decreased polymorphism levels. Additionally, neutral variants linked to selectively favored polymorphisms can be driven to high frequency and fixed. Such genetic hitchhiking (Maynard Smith and Haigh, 1974) also predicts a reduction of polymorphism in regions of low recombination. If selection is sufficiently common, both theories indicate that levels of polymorphism should be positively correlated with recombination rate overall (Hudson and Kaplan, 1995; Wiehe and Stephan, 1993).

Nucleotide polymorphism data from *Drosophila melanogaster* strongly support these predictions. The tiny fourth chromosome and the tip of the X chromosome, two genomic regions that rarely recombine, have reduced nucleotide variation (Aguade et al., 1989; Begun and Aquadro, 1991; Berry et al., 1991; Jensen et al., 2002; but see Wang et al., 2002). Furthermore, there is an overall positive correlation between nucleotide polymorphism and recombination rate throughout the *D. melanogaster* genome (Begun and Aquadro, 1992; Moriyama and Powell, 1996). Similar (but weaker) patterns are observed in other *Drosophila* species (Begun and Aquadro, 1991; Berry et al., 1991; Hilton et al., 1994; Stephan and Langley, 1989), mice (Nachman, 1997), sea beets (Kraft et al., 1998), tomatoes (Stephan and Langley, 1998), goatgrasses (Dvorak et al., 1998), and

Abbreviations: SNP, single nucleotide polymorphism.

* Corresponding author. Tel.: +1-520-626-4747; fax: +1-520-621-9190.
E-mail address: payseur@email.arizona.edu (B.A. Payseur).

0141-933/02/$ - see front matter
PII: S0378-1119(02)00849-1

maize (Tenaillon et al., 2001). Recent work has demonstrated that nucleotide variation and recombination rate are also correlated in humans (Nachman et al., 1998; Przeworski et al., 2000; Nachman, 2001).

At present, it remains unclear whether background selection, genetic hitchhiking, or some combination of these processes best explains the reduced level of genetic variation observed in genomic regions with little recombination. Unfortunately, the detailed information on levels and patterns of polymorphism and recombination needed to address this problem are available in only a few species. At present, the two best candidates are *D. melanogaster* and humans. Here, we review the evidence that selection acting at linked sites shapes patterns of variation across the human genome. We discuss the results of work aimed at distinguishing between background selection and genetic hitchhiking as causes of observed patterns in humans, and we suggest avenues for further research.

2. Selection at linked sites in humans

2.1. The effect of selection on neutral nucleotide polymorphism in humans

Two estimates of the neutral parameter, $4N_e\mu$ (where N_e is the effective population size and μ is the per-generation mutation rate), are $\hat{\theta}_w$ (Watterson, 1975) and $\hat{\pi}$ (Nei and Li, 1979). The proportion of segregating nucleotides in a population sample, corrected for sample size ($\hat{\theta}_w$), is independent of the frequencies of segregating sites, while the average number of nucleotide differences between two randomly chosen sequences in a population ($\hat{\pi}$) includes information about polymorphism frequencies. Background selection and genetic hitchhiking each predict reduced values of $\hat{\pi}$ and $\hat{\theta}_w$ in genomic regions experiencing low recombination rates. This prediction can be tested in humans by comparing polymorphism levels from population surveys of non-coding sequences to recombination rate. Human recombination rates can be estimated by comparing the positions of markers on genetic and physical maps (e.g. Payseur and Nachman, 2000). Using this approach, Nachman (2001) demonstrated a strong positive correlation between nucleotide variation and recombination rate ($\hat{\theta}_w$, $R^2 = 0.63$, $P < 0.001$; $\hat{\pi}$, $R^2 = 0.54$, $P < 0.001$) for 17 loci scattered throughout the genome. Recombination rate and divergence (measured by comparing human and chimpanzee sequences) were not correlated, arguing against a neutralist interpretation that recombination is mutagenic.

The effects of background selection and genetic hitchhiking on neutral polymorphism depend (inversely) on the local recombination rate. The magnitude of these effects may also depend on the number of selective targets linked to neutral variants. Therefore, if selection acts disproportionately on coding regions (relative to non-coding regions), the number of genes in a genomic region may be related to observed levels of polymorphism. Specifically, background selection and genetic hitchhiking both predict reduced polymorphism in gene-rich regions. Using the same set of loci as Nachman (2001) and gene density estimates from the human genome sequence, Payseur and Nachman (2002) tested this prediction. There is a significant negative correlation between the residuals of the regression reported above (comparing $\hat{\theta}_w$ and recombination rate) and gene density ($R^2 = 0.25$, $P = 0.04$). Moreover, there is weak evidence that $\hat{\theta}_w$ alone is negatively correlated with gene density ($R^2 = 0.17$, $P = 0.10$). These results provide further support for the importance of selection acting at linked sites and suggest, perhaps not surprisingly, that genes rather than non-coding regions may be most frequently targeted by selection. The combined ability of recombination rate and gene density to explain levels of polymorphism (adjusted $R^2 = 0.68$, $P = 0.0001$) motivates attempts to distinguish between models such as background selection and genetic hitchhiking. Below we discuss four approaches for distinguishing between background selection and genetic hitchhiking; the predictions under each model are summarized in Table 1.

2.2. The relationship between polymorphism level and recombination rate for neutral loci with high mutation rates

The level of nucleotide polymorphism under background selection (Charlesworth et al., 1993) is given by

$$\pi = 4f_0\pi_0$$

π_0 is the neutral level of variation and f_0, the proportion of

Table 1
Predictions of background selection and genetic hitchhiking models

Comparison	Background selection	Genetic hitchhiking
1. Polymorphism levels in low-recombination regions for loci with high mutation rates	Reduced	Depends on mutation rate and strength of selection
2. Relative polymorphism levels on the X chromosome and the autosomes	Higher on the X chromosome	May be lower on the X chromosome
3. Frequency spectra in low-recombination regions	Not skewed in large populations	Skewed toward an excess of rare variants
4. Patterns of polymorphism in different populations	Similar inter-locus patterns of polymorphism in different populations	Different inter-locus patterns of polymorphism in different populations

gametes free of deleterious mutations, is given by

$$f_0 = \exp\left(-\sum_i \frac{q_i}{(1+\rho_i)^2}\right)$$

where $q_i = u_i/h_i s_i$ is the equilibrium mutant allele frequency at the ith selected locus, $\rho_i = r_i(1 - h_i s_i)/h_i s_i$, u_i is the mutation rate at the ith locus, $h_i s_i$ is the heterozygous fitness effect of a mutant allele at the ith locus, and r_i is the recombination frequency between the neutral locus and the ith selected locus (Hudson and Kaplan, 1995; Nordborg et al., 1996). The strength of background selection is indicated by the reduction in f_0, which measures the effective population size of the affected region. f_0 does not depend on the mutation rate at the neutral locus; consequently, background selection should cause a reduction of polymorphism in low-recombination regions without regard to mutation rate. This observation suggests that if background selection is responsible for the correlation between nucleotide variation and recombination rate in humans, a similar correlation between polymorphism and recombination rate should also be observed for markers with higher mutation rates, such as microsatellites (Slatkin, 1995).

In contrast, the reduction of polymorphism in a low-recombination region affected by genetic hitchhiking is related to the mutation rate at the neutral locus and the frequency of selective sweeps (Wiehe and Stephan, 1993). For example, if mutation rates are high, variation will be quickly restored after a hitchhiking event. In this case, selective sweeps must occur at a relatively high frequency for their effects to be visible. Wiehe (1998) and Schlötterer and Wiehe (1999) modeled the conditions under which positive selection at one site will reduce levels of linked microsatellite variation. With the high mutation rates characteristic of human microsatellites (e.g. 10^{-4}; Banchs et al., 1994) genetic hitchhiking is generally unlikely to be detectable through a reduction in levels of polymorphism unless selective sweeps are very frequent. For example, if $s = 0.01$, $N_e = 10^4$, and $\mu = 10^{-4}$, microsatellite polymorphism will be reduced to 10% of its neutral value only in the case of complete linkage (i.e. no recombination). With even a little recombination, smaller s, or higher μ, little or no effect on levels of microsatellite polymorphism is expected. However, since we know little about the rate of selective sweeps in humans, the effect of hitchhiking on human microsatellite variation is difficult to predict. Therefore, the comparison of human microsatellite variation to recombination rate in humans is best construed as a test of background selection.

Payseur and Nachman (2000) used published data to estimate levels of microsatellite variation (based on a sample of 28 unrelated Europeans) and recombination rates throughout the human genome. They reported no strong overall correlation between microsatellite variation and recombination rate. Additionally, they showed that there is little difference in polymorphism levels between loci in regions of very high and very low recombination rates. A similar result was obtained by Yu et al. (2001a), who measured variation in recombination rate by comparing the genetic and sequence-based physical map positions of markers from the first draft of the human genome. These authors also found no correlation between recombination rate and microsatellite variation. Although inter-locus variation in mutation rate may have partly obscured a stronger effect in both studies, it seems unlikely that background selection is a primary determinant of microsatellite polymorphism levels in humans.

2.3. Relative polymorphism levels on the X chromosome and the autosomes

Models of background selection and genetic hitchhiking make different qualitative predictions about the relative levels of neutral variation on the X chromosome and on the autosomes (Aquadro et al., 1994). Under background selection, the X chromosome is expected to be more variable than the autosomes (once differences in effective population size are taken into account: heterozygosity on the X chromosome is multiplied by 4/3 to account for the fact that there are three X chromosomes in the population for every four autosomes when the breeding sex ratio is one). Under genetic hitchhiking, the X chromosome may be less variable than the autosomes.

Deleterious recessive mutations will be maintained at lower frequencies and removed from the population more quickly on the X chromosome than on the autosomes (e.g. Crow and Kimura, 1970). Due to the hemizygous nature of the X chromosome, selection acts every generation in males to remove these deleterious mutations. At mutation-selection equilibrium, there will be a larger fraction of chromosomes that are free of deleterious mutations for the X chromosome than for the autosomes, causing f_0 to be larger for the X chromosome relative to the autosomes. Thus, background selection predicts higher levels of polymorphism on the X chromosome than on the autosomes (Charlesworth et al., 1993; Charlesworth, 1996). The magnitude of the difference will depend on the average recessivity of deleterious mutations.

The predictions concerning relative levels of variation on the X chromosome and the autosomes under genetic hitchhiking are more complicated, but generally indicate that the X chromosome should be less variable than the autosomes (Aquadro et al., 1994; Begun and Whitley, 2000). The actual effect depends on a number of factors, including the average dominance of beneficial mutations, sojourn times as beneficial mutations move to fixation, whether adaptive evolution results from new mutations or from standing variation, and possible differences in gene density on the X chromosome and the autosomes. If, on average, beneficial mutations are recessive, then fixation rates will be higher on the X chromosome than on the

autosomes (Charlesworth et al., 1987) and genetic hitchhiking will be more frequent, causing reduced variation on the X chromosome relative to the autosomes. However, even if beneficial mutations are not recessive on average, sojourn times for beneficial mutations are expected to be shorter on the X chromosome than on the autosomes due to the partial haploid nature of the X chromosome (Avery, 1984). If beneficial mutations spread through a population more quickly, there will be less opportunity for recombination and consequently, there will be greater hitchhiking effects (Begun and Whitley, 2000).

Using published nucleotide data (Nachman, 2001), we compared X-linked and autosomal polymorphism at mostly non-coding sites in humans. As in previous sections, we concentrated only on studies that sampled more than ten individuals ($n = 17$ loci). In these data, there is no evidence for a difference in the level of nucleotide heterozygosity between X-linked and autosomal loci (multiplying X-linked values by 4/3; two-tailed t-test; $\hat{\pi}$, $P = 0.99$; $\hat{\theta}_w$, $P = 0.23$). The mean, corrected value of $\hat{\pi}$ for the X chromosome is 0.101% and the mean value of $\hat{\pi}$ for the autosomes is also 0.101%; the mean, corrected value of $\hat{\theta}_w$ for the X chromosome is 0.128% and the mean value of $\hat{\theta}_w$ for the autosomes is 0.098%.

These comparisons assume that the X chromosome and the autosomes have equal mutation rates and equal recombination rates. However, male-driven molecular evolution (Makova and Li, 2002; Miyata et al., 1987; Nachman and Crowell, 2000a) or selection for modifiers of mutation rate (McVean and Hurst, 1997) may reduce the mutation rate on the X chromosome, and the X chromosome and the autosomes may experience different recombination rates. To assess the effects of differences in mutation rate and recombination rate on our comparisons, we performed two additional sets of analyses. First, we adjusted nucleotide polymorphism for variation in mutation rate by dividing each value by divergence (measured by comparing human and chimpanzee sequences). Comparison of these ratios for the X chromosome and the autosomes ($n = 15$) yields no significant differences, although values of $\hat{\theta}_w$ (corrected for divergence) suggest a marginally significant increase in polymorphism on the X chromosome ($\hat{\pi}_{corrected}$, mean value for the X chromosome = 0.100, mean value for the autosomes = 0.080, two-tailed t-test, $P = 0.54$; $\hat{\theta}_{w\ corrected}$, mean value for the X chromosome = 0.125, mean value for the autosomes = 0.082, two-tailed t-test, $P = 0.09$). It is important to bear in mind that if neutral mutation rates are lower on the X chromosome, deleterious mutation rates will also be reduced and background selection may be a weaker force. However, comparison of divergence values between X-linked and autosomal loci suggests no strong difference in neutral mutation rates for these data ($P = 0.25$).

Second, we compared the residuals of a regression of nucleotide polymorphism on the recombination rate for the X chromosome and the autosomes (multiplying X-linked recombination rates by 2/3 to account for the fact that the X chromosome spends 2/3 of its time in a recombining sex; this correction was used in all subsequent analyses) and again found no significant differences in polymorphism ($\hat{\pi}$, mean value of residuals for the X chromosome = 0.003, mean value of residuals for the autosomes = −0.004, two-tailed t-test, $P = 0.75$; $\hat{\theta}_w$, mean value of residuals for the X chromosome = 0.014, mean value of residuals for the autosomes = −0.010, two-tailed t-test, $P = 0.11$). Therefore, our results do not appear to be overly sensitive to the assumption that the X chromosome and the autosomes have equal mutation rates and recombination rates.

An additional assumption of our polymorphism comparisons is that the neutral ratio of effective population sizes for the X chromosome and the autosomes is 3/4. However, differences between the sexes in variance in reproductive success can cause deviations from this ratio (Charlesworth, 2001). In humans, where variance in reproductive success may be higher in males than in females, the ratio of X chromosome to autosome effective population sizes may be larger than 3/4, and multiplying X-linked polymorphism by 4/3 may represent an over-correction. Therefore, this procedure may bias our results against detecting a reduction in polymorphism on the X chromosome. Choosing the appropriate correction factor for these comparisons is challenging, given our lack of detailed knowledge of relative male and female effective population sizes. Differences in patterns of migration between the sexes may also complicate attempts to compare variation on the X chromosome and the autosomes. Finally, differences in effective population size could cause the X chromosome and the autosomes to respond in different ways to demographic changes. For example, if human populations have recently undergone a bottleneck, the lower effective population size of the X chromosome could allow X-linked variation to recover from this event more rapidly than variation on the autosomes. Fay and Wu (1999) described this effect in the context of disparities between human frequency spectra of mitochondrial and autosomal loci. The effect is expected to be less severe for X-autosome comparisons since the X chromosome has 3/4 the effective population size of autosomes, while the comparable value for mitochondrial DNA is 1/4.

Interestingly, if one considers only European populations (using data for 14 of the same loci and two additional loci; see Section 2.4), the X chromosome is about half as variable as the autosomes ($\hat{\pi}$, mean for the X chromosome = 0.048%, mean for the autosomes = 0.096%, two-tailed t-test, $P = 0.12$; $\hat{\theta}_w$, mean for the X chromosome = 0.049%, mean for the autosomes = 0.083%, two-tailed t-test, $P = 0.16$), although one recent study found similar levels of nucleotide variation on the X chromosome and the autosomes of Europeans (Yu et al., 2002). Geographic patterns of variation are discussed in more detail below.

A genome-wide assessment of single nucleotide polymorphisms (SNPs) in humans was recently reported (International SNP Map Working Group, 2001). In this

study, SNPs were not identified in a true population sample, but primarily through the comparison of two to three individuals. Using these data (Table 2 from International SNP Map Working Group, 2001), and correcting for differences in effective population size between the X chromosome and the autosomes, average $\hat{\pi}$ for the X chromosome (0.063%) is approximately 20% lower than average $\hat{\pi}$ for the autosomes (0.077%). The comparison of estimates of $4N_e\mu$ using variances in allele size for 209 X-linked and 5048 autosomal microsatellites (data from Dib et al., 1996), again multiplying X-linked values by 4/3, also reveals a weak, marginally significant reduction in X-linked polymorphism (mean for the X chromosome = 25.2, mean for the autosomes = 28.1, Mann–Whitney U-test, $P = 0.09$).

Overall, the comparisons between the X chromosome and the autosomes do not clearly distinguish background selection and genetic hitchhiking models. Considering only European populations (from data analyzed in this paper, the genomic SNP data, and the microsatellite data), there is at best weak evidence for a reduction of variation on the X chromosome relative to the autosomes. However, this is not the case in analyses of data for African populations (below), where the X chromosome may be more polymorphic than the autosomes.

2.4. The frequency spectrum of polymorphisms

Background selection and genetic hitchhiking are expected to affect the frequency spectrum of polymorphisms in different ways. Under most conditions, background selection is not expected to cause a substantial skew in the frequency spectrum; the effect is analogous to a reduction in effective population size (Charlesworth et al., 1995). In contrast, simple models of genetic hitchhiking in which a new adaptive mutation quickly spreads through the population to fixation are expected to cause a strong skew in the frequency distribution of polymorphisms with an excess of low-frequency variants (Tajima, 1989a; Braverman et al., 1995; Simonsen et al., 1995). Following a simple, complete selective sweep, all variation will be eliminated, and new variation will arise solely through the input of new mutations that begin at low frequencies. Over time, the distribution of polymorphisms will return to equilibrium. Tajima's D (Tajima, 1989a) is a statistic that summarizes aspects of the frequency distribution of polymorphisms and is based on the standardized difference between $\hat{\pi}$ and $\hat{\theta}_w$. At equilibrium, D is expected to be approximately 0. When there is an excess of rare variants D takes on negative values, and when there is an excess of intermediate-frequency variants D takes on positive values. If simple genetic hitchhiking is largely responsible for the correlation between nucleotide heterozygosity and recombination rate, we also expect a positive correlation between Tajima's D and recombination rate. If background selection is largely responsible (and effective population sizes are large), then no correlation between Tajima's D and recombination rate is predicted. Nachman (2001) compared Tajima's D for human nucleotide polymorphism data with recombination rate and detected a weak, positive association (Fig. 1; $R^2 = 0.17$, $P = 0.10$). Hence, the trend is in the direction predicted by genetic hitchhiking. However, this observation does not provide clear support for genetic hitchhiking for two reasons. First, the trend is very weak and not statistically significant. Second, if selection against deleterious mutations is weak and effective population sizes are small (as may be the case in humans), background selection can also cause a skew in the frequency spectrum in regions of low recombination.

Summaries of the frequency spectrum other than Tajima's D may be more useful for distinguishing between background selection and genetic hitchhiking. For example, Fay and Wu (2000) argued that genetic hitchhiking models uniquely predict an excess of high-frequency, derived alleles, and devised a test statistic (H) to measure departures from a neutral frequency spectrum in this direction. Therefore, a correlation between H and recombination rate would not be expected under background selection. However, although the signal of genetic hitchhiking may be easier to differentiate from other forces by using the H test, this signal is expected to persist for a shorter amount of time than that seen in Tajima's D (Przeworski, 2002), suggesting that the power to detect such an association may be reduced.

2.5. Human nucleotide polymorphism in Africa and Europe

Another approach to evaluating the relative significance of background selection and genetic hitchhiking is to compare patterns of polymorphism among different populations (e.g. Schlötterer and Wiehe, 1999). Because background selection is an equilibrium process that involves recurrent deleterious mutations, all populations are expected to respond in a roughly similar fashion. Alternatively, genetic hitchhiking may involve the fixation of beneficial mutations in one population only, or the fixation of different beneficial mutations in different populations. Thus, population-specific deviations from neutrality at particular loci may identify candidate regions for genetic hitchhiking.

Published sequence-based surveys of human nucleotide polymorphism differ in the populations sampled, but most contain information about variation in Africa and Europe. We assessed levels and patterns of nucleotide diversity for Africans and Europeans at 16 loci (Table 2).

Consistent with previous studies, average nucleotide diversity is greater in Africa ($\hat{\pi} = 0.112\%$, $\hat{\theta}_w = 0.134\%$) than in Europe ($\hat{\pi} = 0.069\%$, $\hat{\theta}_w = 0.067\%$), and these differences are statistically significant (one-tailed, paired t-test: $\hat{\pi}$, $P = 0.026$; $\hat{\theta}_w$, $P = 0.004$). The observation that $\hat{\pi}$ values are more similar than $\hat{\theta}_w$ values suggests that polymorphic sites are segregating at higher average frequencies in Europe than in Africa. Indeed, the average

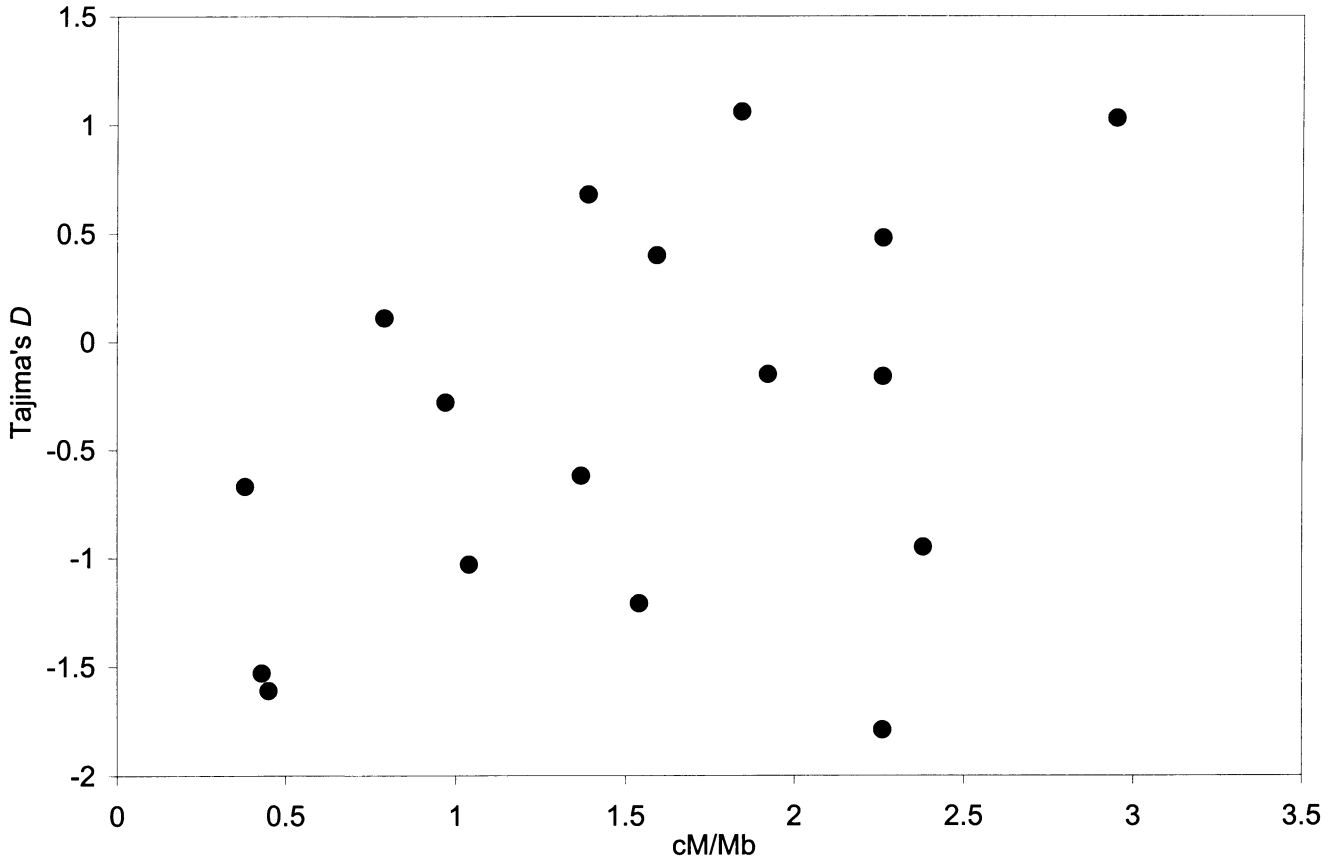

Fig. 1. Scatterplot of Tajima's D vs. recombination rate (cM/Mb).

Tajima's D is higher in Europe ($D = -0.063$) than in Africa ($D = -0.420$), although these values are close to zero in both groups, and the difference is not statistically significant (two-tailed, paired t-test, $P = 0.47$).

We looked for population-specific differences in several ways. First, we compared polymorphism to recombination rate for African and European samples separately.[1] Recombination rate estimates were taken from Nachman (2001). Nucleotide diversity is strongly correlated with recombination rate in Africans ($\hat{\pi}$, $R^2 = 0.44$, $P = 0.007$; $\hat{\theta}_w$, $R^2 = 0.51$, $P = 0.003$). These relationships are similar in magnitude to the correlations observed when polymorphism data are pooled across groups (Nachman, 2001), reflecting the fact that much of the nucleotide diversity in humans is located within Africa. In contrast, there is no significant association between European nucleotide diversity and recombination rate ($\hat{\pi}$, $P = 0.36$; $\hat{\theta}_w$, $P = 0.36$). One explanation for this result is that the power to detect an association may be too low in Europeans because polymorphism levels are small. Two arguments suggest that this is not the case. First, although average polymorphism levels are higher in Africa, the magnitude of the difference is relatively small. Second, the range of variation in nucleotide polymorphism is similar between the African and European groups. The coefficient of variation, the standard deviation scaled by the average, is similar in Europe ($\hat{\pi}$, CV = 83.3; $\hat{\theta}_w$, CV = 67.9) and Africa ($\hat{\pi}$, CV = 70.4; $\hat{\theta}_w$, CV = 73.7). The absence of a correlation between nucleotide diversity and recombination rate in Europe seems to be caused by three high-recombination genes with low nucleotide diversity (*Pdha1*, *FIX*, and *Dmdl7*). All of these regions have been identified as likely candidates for recent selective sweeps in non-African populations (Harris and Hey, 1999, 2001; Nachman and Crowell, 2000b). The absence of a correlation in Europe is inconsistent with background selection. However, background selection has only been modeled for populations at equilibrium. Since European populations are unlikely to be at equilibrium, theoretical studies of background selection in non-equilibrium situations (e.g. population bottlenecks and expansions) would be useful.

Second, we compared levels of variation in Africa and Europe for each locus; scatterplots of $\hat{\pi}$ and $\hat{\theta}_w$ for Europe vs. Africa are shown in Fig. 2. For both measures of polymorphism there is a general correspondence between populations: loci that are more variable in Africa tend to also be more variable in Europe. Nevertheless, correlation analyses indicate no significant association between poly-

[1] This comparison did not include the 16p13.3 locus (Alonso and Armour, 2001). This locus lies at the tip of chromosome 16, making it difficult to reliably estimate recombination rate.

Table 2
Data used for nucleotide polymorphism comparisons between Africa and Europe

Locus	N[a]		π (%)[b]		θ (%)[c]		D[d]		Reference
	Africa	Europe	Africa	Europe	Africa	Europe	Africa	Europe	
β-Globin	103	46	0.115	0.130	0.132	0.108	−0.358	0.690	Harding et al., 1997
Lpl	48[e]	48	0.200	0.160	0.180	0.130	0.812	1.007	Clark et al., 1998
Ace	10[e]	12	0.108	0.073	0.103	0.061	0.227	1.522	Rieder et al., 1999
Apoe	48[e]	48	0.044	0.056	0.058	0.053	−0.736	0.160	Fullerton et al., 2000
22q11.2	40	44	0.085	0.077	0.128	0.074	−1.070	−0.084[f]	Zhao et al., 2000
1q24	40	21	0.076	0.045	0.076	0.044	0.000	0.145	Yu et al., 2001b
Duffy	24	17	0.040	0.131	0.062	0.108	−0.970	0.680	Hamblin and Di Rienzo, 2000
16p13.3	40	10	0.238	0.044	0.432	0.097	−1.549[g]	−1.739[g]	Alonso and Armour, 2001
Pdha1	18	6	0.273	0.011	0.225	0.014	0.862	−0.933	Harris and Hey, 1999
Xq13.3	22	7	0.047	0.043	0.086	0.046	−1.718	−0.312	Kaessmann et al., 1999
Zfx	113	93	0.103	0.081	0.161	0.096	−0.830	−0.510	Jaruzelska et al., 1999
Dmd I44	10	10	0.231	0.192	0.220	0.157	0.223	0.985	Nachman and Crowell, 2000b
Dmd I7	10	10	0.107	0.012	0.117	0.020	−0.409	−1.176	Nachman and Crowell, 2000b
FIX	18	5	0.031	0.021	0.062	0.017	−1.660	1.225	Harris and Hey, 2001
Msn	10	10	0.047	0.017	0.041	0.031	0.566	−1.562	[h]
Alas	10	10	0.054	0.005	0.055	0.009	−0.106	−1.112	[h]

[a] Number of chromosomes sampled.
[b] Average pairwise difference between two randomly chosen sequences (Nei and Li, 1979), expressed as a percentage. X-linked values are multiplied by 4/3.
[c] Proportion of segregating sites, corrected for sample size (Watterson, 1975). X-linked values are multiplied by 4/3.
[d] Tajima's D (Tajima, 1989a).
[e] The sample is African-Americans.
[f] Calculated from total non-African sample ($N = 44$).
[g] $P < 0.05$.
[h] M.W.N., S. D'Agostino, C. Tillquist, and M. Hammer, unpublished results.

morphism levels in Europe and Africa for either $\hat{\pi}$ or $\hat{\theta}_w$ ($\hat{\pi}$, $P = 0.26$; $\hat{\theta}_w$, $P = 0.10$). Again, the absence of a correlation is caused by a few loci that have low variation in Europe but not in Africa. There is also no correlation between European and African values of Tajima's D ($P = 0.88$).

Third, we compared X-linked and autosomal variation in Africa and Europe (Fig. 3). Levels of X-linked polymorphism are not significantly different from levels of autosomal polymorphism in Africans or Europeans ($P > 0.05$ in all tests), consistent with results using data pooled across populations. However, relative levels of polymorphism are different in Africans and Europeans. In Africans, average X-linked variation ($\hat{\pi} = 0.112\%$, $\hat{\theta}_w = 0.121\%$) is slightly greater than autosomal variation ($\hat{\pi} = 0.095\%$, $\hat{\theta}_w = 0.106\%$). Conversely, in Europeans, average X-linked variation ($\hat{\pi} = 0.048\%$, $\hat{\theta}_w = 0.049\%$) is considerably less than autosomal variation ($\hat{\pi} = 0.096\%$, $\hat{\theta}_w = 0.083\%$).

All three observations above suggest 'locus-by-population' interactions that are not predicted under background selection. However, these interactions appear to be primarily attributable to only a few loci; patterns at the majority of loci may still be governed by similar forces in different populations.

3. Discussion

Empirical work has demonstrated that natural selection acting at linked sites shapes patterns of neutral polymorphism across the human genome. Selection affects neutral polymorphism most dramatically in regions of reduced recombination and may be more frequent in gene-dense regions. The nature of the selection, however, is not well understood. Is the reduction of neutral polymorphism in low-recombination regions primarily caused by consistent purging of deleterious mutations from the genome or by frequent events of positive selection?

We considered four criteria for distinguishing between background selection and genetic hitchhiking in humans: the relationship between polymorphism level and recombination rate for loci with high mutation rates, relative levels of variation on the X chromosome and the autosomes, the relationship between the frequency spectrum of polymorphisms and recombination rate, and patterns of polymorphism in African and European populations.

Unfortunately, with the available data, these analyses do not clearly distinguish between background selection and genetic hitchhiking. Microsatellite polymorphism is not reduced in regions of low recombination. This observation is inconsistent with background selection, but does not speak to the importance of genetic hitchhiking. Polymorphism data from European populations point toward a potential reduction of variation on the X chromosome relative to the autosomes, while the autosomes appear to be less variable than the X chromosome in African populations and world-wide samples. None of the comparisons between the X chromosome and the autosomes yield statistically significant differences. There is weak evidence that

Tajima's *D* and recombination rate are positively correlated. However, this trend is not statistically significant, and may be predicted under background selection if selection is weak. Finally, population-specific patterns of polymorphism suggest three loci with unusually low diversity in Europeans. Although this result suggests that genetic hitchhiking may be driving patterns at these loci, it does not rule out the possibility that background selection is primarily responsible for polymorphism levels at the remaining loci.

The analyses described here have also been applied to *Drosophila*, and the results provide a useful contrast with observed patterns in humans. First, in some studies of *D. melanogaster*, microsatellite diversity and recombination rate appear to be correlated (Schug et al., 1998; but see also Michalakis and Veuille, 1996). The discrepancy between this pattern and the results for humans may be due to lower mutation rates at microsatellites in *Drosophila* (Schug et al., 1997). Reduced mutation rates would allow the signals of both background selection and genetic hitchhiking events to persist over longer time periods. Thus, the correlation between microsatellite polymorphism and recombination rate in *Drosophila* may be consistent with both models. Second, comparisons of X-linked and autosomal nucleotide variation in *Drosophila* have yielded different results in different species. *Drosophila simulans* exhibits a clear reduction in X-linked variation (Begun and Whitley, 2000), consistent with genetic hitchhiking. In *D. melanogaster*, results depend on which populations are surveyed (Andolfatto, 2001a). In African populations, there is a trend toward higher diversity on the X chromosome relative to the autosomes, while in non-African populations, X-linked loci appear to be less variable than autosomal loci (this pattern is also observed at microsatellite loci; Kauer et al., 2002). In this regard, patterns in humans are more similar to those in *D. melanogaster* than to those in *D. simulans*. Third, Andolfatto and Przeworski (2001) reported a significant positive correlation between nucleotide frequency spectra and recombination rates in *D. melanogaster*, providing support for hitchhiking. The trend in humans is weaker but in the same direction. Finally, relative levels of polymorphism in African and non-African populations of *Drosophila* often differ at individual loci (Andolfatto, 2001a). This result is congruent with observed patterns in humans.

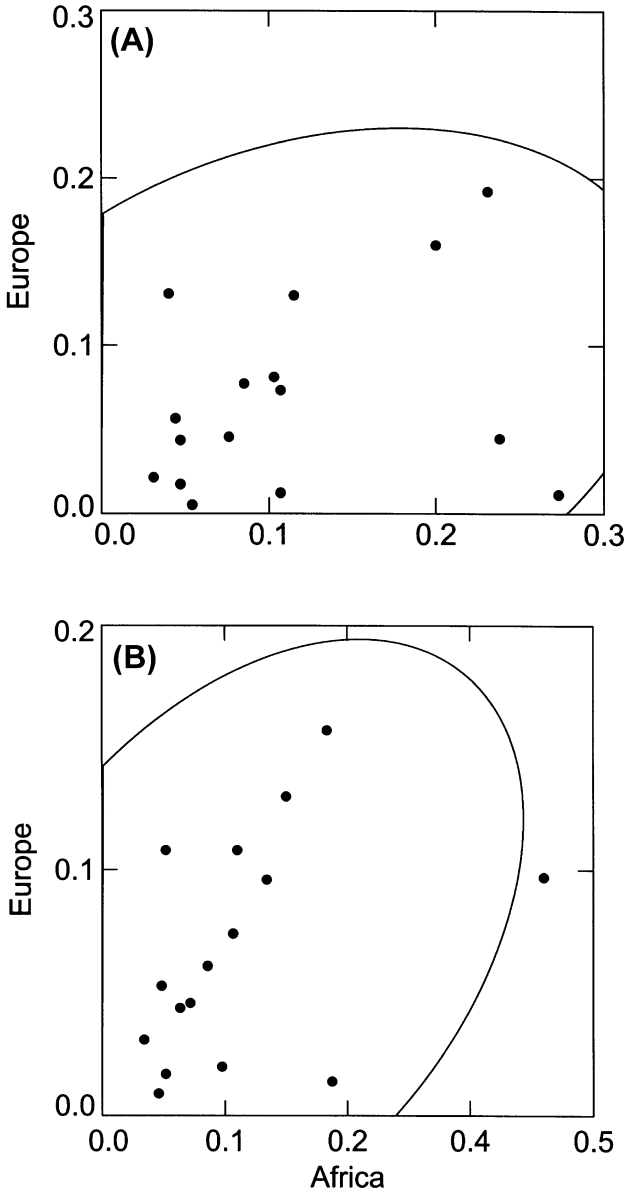

Fig. 2. Scatterplots of European and African nucleotide polymorphism levels with 95% density ellipses. (A) π (%). (B) θ (%). The sole outlier in (B) is *Pdha1*.

Although nucleotide polymorphism is positively correlated with recombination rate in both *Drosophila* and in humans, in general, the signal of positive selection appears to be stronger in *Drosophila* (Andolfatto, 2001b). This is seen in the skews in the frequency spectrum, patterns of geographic variation, and differences between the X chromosome and the autosomes. Are there biological differences between flies and humans that might lead to a stronger signature of positive selection in patterns of DNA sequence variation in flies? One obvious and important difference is in the effective population size, which has been estimated at 10^4 for humans (Nachman et al., 1998) and 10^6 for *D. melanogaster* (Kreitman, 1983). Since rates of adaptive evolution are proportional to population size, we expect, a priori, higher adaptive fixation rates in flies than in humans. A smaller effective population size for humans also means that humans have lower average levels of neutral nucleotide polymorphism. This makes it more difficult to detect differences in the frequency spectrum, differences between populations, or differences between the X chromosome and the autosomes.

Attempts at distinguishing between background selection and genetic hitchhiking would also benefit from theoretical studies incorporating more complex models of selection. For example, many current models deal with populations at equilibrium, although humans and flies are known to have

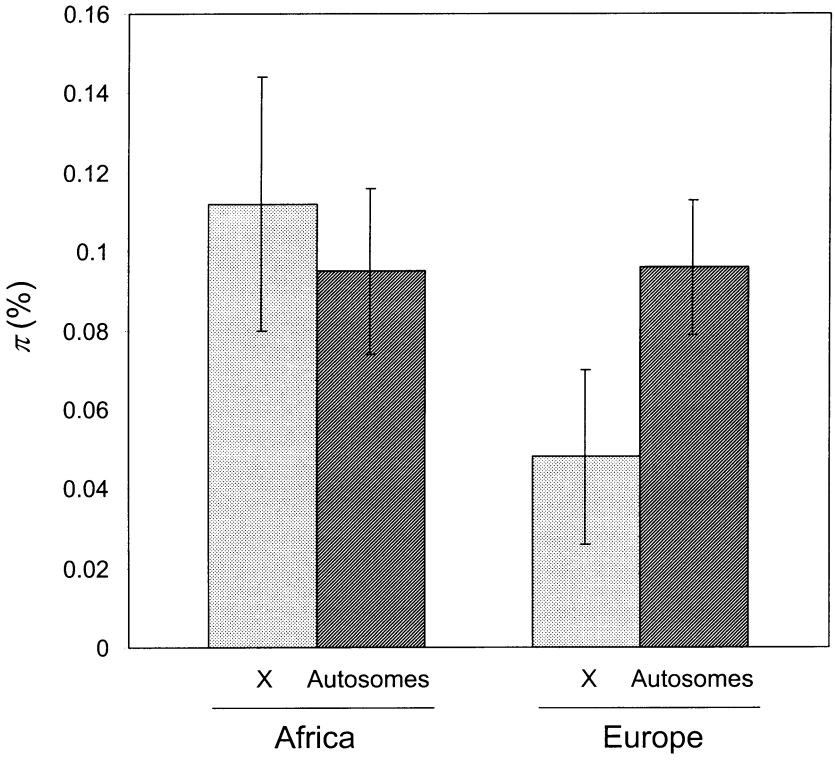

Fig. 3. Average values of π (%) for X-linked loci and autosomal loci in Africa and Europe. Error bars indicate ± one standard error.

undergone major range expansions and changes in population size in their recent evolutionary history. Although the effects of changes in the effective population size or population subdivision on variation at a single locus have been studied (e.g. Tajima, 1989b), expected patterns across loci are less clear. The development of coalescent-based, multi-locus methods for demographic inference (e.g. Beaumont, 1999) and empirical attempts to estimate the degree of inter-locus variance by sampling loci in regions with similar recombination rates in the same individuals (e.g. Frisse et al., 2001) are promising, but we need more theoretical work that makes explicit predictions for different demographic scenarios. Second, the approaches considered in this paper are based on the predictions of a genetic hitchhiking model assuming strong, constant, positive selection. However, positive selection acting on individual loci may often be quite weak (Ohta, 1973; Przeworski et al., 2000) or temporally variable (Gillespie, 1994), suggesting that theoretical investigation of other models of positive selection may be useful. Third, it seems unavoidable that both background selection and genetic hitchhiking play some role in generating observed patterns. Recent theoretical work treating these forces simultaneously suggests that genetic hitchhiking may be the dominant force in low-recombination regions, while background selection may be more important in high-recombination regions (Kim and Stephan, 2000). Additional theoretical studies incorporating both positive and negative selection will be useful.

Finally, independent estimates of the deleterious mutation rate and the adaptive substitution rate may allow us to gauge the likelihood that background selection and/or genetic hitchhiking are important in humans. For example, the genomic deleterious mutation rate in humans has been estimated at approximately two mutations per genome per generation (Eyre-Walker and Keightley, 1999; Nachman and Crowell, 2000a). We can calculate the expected reduction in heterozygosity due to background selection alone in a region of low recombination using the formula:

$$f_0 = \exp\left(-\frac{U}{2hs}\right)$$

(a special case of the more general formula from Section 2.2, where $r_i = 0$ and hs represents the sum of terms across i loci; Charlesworth et al., 1993). For the centromeric region of the human X chromosome, the approximate size of the region of low recombination (less than 0.5 cM/Mb; Payseur and Nachman, 2000) is 4.9 Mb, comprising about 0.16% of the genome. Thus, the deleterious mutation rate for this region is 3.2×10^{-3} ($0.16\% \times 2$). Assuming $hs = 0.02$ (Crow and Simmons, 1983), the expected reduction in neutral variation ($1 - f_0$) is only 8%. In contrast, the observed level of nucleotide variation in this region of the genome is about 64% lower than average (Kaessmann et al., 1999). This discrepancy suggests that background selection alone may not be sufficient to account for the observed low levels of nucleotide variation. However, if selection coefficients are smaller, the fit of the background selection model to the data will be improved. In order for background selection to predict a 64% reduction of variation in this region of the genome, hs would need to be approximately

1.6×10^{-3}. Although we know little about the average strength of selection in humans, this value does not seem unreasonable.

There are few estimates of the rate of adaptive evolution in humans, but Fay et al. (2001) recently calculated that approximately 35% of non-synonymous substitutions between humans and Old World monkeys were fixed by positive selection. Humans have 31,778 protein-coding genes (International Human Genome Sequencing Consortium, 2001), and the average number of non-synonymous sites per gene is 733 (International Human Genome Sequencing Consortium, 2001). The average non-synonymous divergence between humans and chimpanzees is approximately 0.32% (Eyre-Walker and Keightley, 1999; Ohta, 1995), so the total number of non-synonymous substitutions between chimpanzees and humans is roughly 74,500, and the number of non-synonymous differences fixed by positive selection may be approximately 26,000. Is this number of adaptive substitutions sufficient to account for the observed differences in levels of polymorphism seen in humans? Nachman (2001) used the observed correlation between nucleotide variation and recombination rate and the model of Wiehe and Stephan (1993) to estimate that 30,000 adaptive substitutions may have occurred since the divergence of chimpanzees and humans. Although these estimates are very rough, the general correspondence between this number and the independent estimate based on the results of Fay et al. (2001) suggests that rates of adaptive evolution and associated genetic hitchhiking in humans may be sufficiently high to account for much of the observed variation in levels of nucleotide polymorphism.

Acknowledgements

We thank Giorgio Bernardi for hosting the symposium that helped bring this paper to fruition. We also thank Santos Alonso, Andy Clark, Anna Di Rienzo, Rosalind Harding, Jody Hey, and Zhongming Zhao for assistance with data compilation, and two anonymous reviewers for helpful comments on the manuscript. This work was supported by a National Science Foundation Integrative Graduate Education and Research Training Grant Fellowship in Biology, Mathematics, and Physics to B.A.P. and by National Science Foundation grants to M.W.N.

References

Aguade, M., Miyashita, N., Langley, C.H., 1989. Reduced variation in the *yellow-achaete-scute* region in natural populations of *Drosophila melanogaster*. Genetics 122, 607–615.

Alonso, S., Armour, J.A.L., 2001. A highly variable segment of human subterminal 16p reveals a history of population growth for modern humans outside Africa. Proc. Natl. Acad. Sci. USA 98, 864–869.

Andolfatto, P., 2001a. Contrasting patterns of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *Drosophila simulans*. Mol. Biol. Evol. 18, 279–290.

Andolfatto, P., 2001b. Adaptive hitchhiking effects on genome variability. Curr. Opin. Genet. Dev. 11, 635–641.

Andolfatto, P., Przeworski, M., 2001. Regions of lower crossing over harbor more rare variants in African populations of *Drosophila melanogaster*. Genetics 158, 657–665.

Aquadro, C.F., Begun, D.J., Kindahl, E.C., 1994. Selection, recombination, and DNA polymorphism in *Drosophila*. In: Golding, B., (Ed.), Non-Neutral Evolution: Theories and Molecular Data, Chapman and Hall, London, pp. 46–56.

Avery, P.J., 1984. The population-genetics of haplo-diploids and X-linked genes. Genet. Res. 44, 321–341.

Banchs, I., Bosch, A., Guimera, J., et al., 1994. New alleles at microsatellite loci in CEPH families mainly arise from somatic mutations in the lymphoblastoid cell lines. Hum. Mutat. 3, 365–372.

Beaumont, M.A., 1999. Detecting population expansion and decline using microsatellites. Genetics 153, 2013–2029.

Begun, D.J., Aquadro, C.F., 1991. Molecular population genetics of the distal portion of the X chromosome in Drosophila. Genetics 129, 1147–1158.

Begun, D.J., Aquadro, C.F., 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rate in *D. melanogaster*. Nature 356, 519–520.

Begun, D.J., Whitley, P., 2000. Reduced X-linked nucleotide polymorphism in *Drosophila simulans*. Proc. Natl. Acad. Sci. USA 97, 5960–5965.

Berry, A.J., Ajioka, J.W., Kreitman, M., 1991. Lack of polymorphism on the Drosophila fourth chromosome resulting from selection. Genetics 129, 1111–1117.

Braverman, J.M., Hudson, R.R., Kaplan, N.L., Langley, C.H., Stephan, W., 1995. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. Genetics 140, 783–796.

Charlesworth, B., 1996. Background selection and patterns of genetic diversity in *Drosophila melanogaster*. Genet. Res. 68, 131–149.

Charlesworth, B., 2001. The effect of life-history and mode of inheritance on neutral genetic variability. Genet. Res. 77, 153–166.

Charlesworth, B., Coyne, J.A., Barton, N.H., 1987. The relative rates of evolution of sex chromosomes and autosomes. Am. Nat. 130, 113–146.

Charlesworth, B., Morgan, M.T., Charlesworth, D., 1993. The effect of deleterious mutations on neutral molecular variation. Genetics 134, 1289–1303.

Charlesworth, D., Charlesworth, B., Morgan, M.T., 1995. The pattern of neutral molecular variation under the background selection model. Genetics 141, 1619–1632.

Clark, A.G., Weiss, K.M., Nickerson, D.A., et al., 1998. Haplotype structure and population genetic inferences from nucleotide sequence variation in human lipoprotein lipase. Am. J. Hum. Genet. 63, 595–612.

Crow, J.F., Kimura, M., 1970. An Introduction to Population Genetics Theory, Harper and Row, New York.

Crow, J.F., Simmons, M.J., 1983. The mutation load in *Drosophila*. In: Ashburner, M., Carson, H.L., Thompson, J.J.N. (Eds.), The Genetics and Biology of *Drosophila*, Academic Press, New York, pp. 1–35.

Dib, C., Faure, S., Fizames, C., et al., 1996. A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature 380, 152–154.

Dvorak, J., Luo, M.-C., Yang, Z.-L., 1998. Restriction fragment length polymorphism and divergence in the genomic regions of high and low recombination in self-fertilizing and cross-fertilizing Aegilops species. Genetics 148, 423–434.

Eyre-Walker, A., Keightley, P.D., 1999. High genomic deleterious mutation rates in hominids. Nature 397, 344–347.

Fay, J.C., Wu, C.-I., 1999. A human population bottleneck can account for the discordance between patterns of mitochondrial versus nuclear DNA variation. Mol. Biol. Evol. 16, 1003–1005.

Fay, J.C., Wu, C.-I., 2000. Hitchhiking under positive Darwinian selection. Genetics 155, 1405–1413.

Fay, J.C., Wyckoff, G.J., Wu, C.-I., 2001. Positive and negative selection on the human genome. Genetics 158, 1227–1234.

Frisse, L., Hudson, R.R., Bartoszewicz, A., Wall, J.D., Donfack, J., Rienzo, A.D., 2001. Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium levels. Am. J. Hum. Genet. 69, 831–843.

Fullerton, S.M., Weiss, K.M., Clark, A.G., et al., 2000. Apolipoprotein E variation at the sequence haplotype level: implications for the origin and maintenance of a major human polymorphism. Am. J. Hum. Genet. 67, 881–900.

Gillespie, J.H., 1994. Alternatives to the neutral theory. In: Golding, B., (Ed.), Non-Neutral Evolution: Theories and Molecular Data, Chapman and Hall, New York, pp. 1–17.

Hamblin, M.T., Di Rienzo, A., 2000. Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. Am. J. Hum. Genet. 66, 1669–1679.

Harding, R.M., Fullerton, S.M., Griffiths, R.C., Bond, J., Cox, M.J., Schneider, J.A., Moulin, D.S., Clegg, J.B., 1997. Archaic African and Asian lineages in the genetic ancestry of modern humans. Am. J. Hum. Genet. 60, 772–789.

Harris, E.E., Hey, J., 1999. X chromosome evidence for ancient human histories. Proc. Natl. Acad. Sci. USA 96, 3320–3324.

Harris, E.E., Hey, J., 2001. Human populations show reduced DNA sequence variation at the factor IX locus. Curr. Biol. 11, 774–778.

Hilton, H., Kliman, R.M., Hey, J., 1994. Using hitchhiking genes to study adaptation and divergence during speciation with the *Drosophila melanogaster* species complex. Evolution 48, 1900–1913.

Hudson, R.R., Kaplan, N.L., 1995. Deleterious background selection with recombination. Genetics 141, 1605–1617.

International Human Genome Sequencing Consortium, 2001. Initial sequencing and analysis of the human genome. Nature 409, 860–921.

International SNP Map Working Group, 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409, 928–933.

Jaruzelska, J., Zietkiewicz, E., Batzer, M., Cole, D.E.C., Moisan, J.P., Scozzari, R., Tavare, S., Labuda, D., 1999. Spatial and temporal distribution of the neutral polymorphisms in the last Zfx intron: analysis of haplotype structure and genealogy. Genetics 152, 1091–1101.

Jensen, M.A., Charlesworth, B., Kreitman, M., 2002. Patterns of genetic variation at a chromosome 4 locus of *Drosophila melanogaster* and *D. simulans*. Genetics 160, 493–507.

Kaessmann, H., Heibig, F., von Haeseler, A., Paabo, S., 1999. DNA sequence variation in a non-coding region of low recombination on the human X chromosome. Nat. Genet. 22, 78–81.

Kauer, M., Zangerl, B., Dieringer, D., Schlötterer, C., 2002. Chromosomal patterns of microsatellite variability contrast sharply in African and non-African populations of *Drosophila melanogaster*. Genetics 160, 247–256.

Kim, Y., Stephan, W., 2000. Joint effects of genetic hitchhiking and background selection on neutral variation. Genetics 155, 1415–1427.

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

Kraft, T., Sall, T., Magnusson-Rading, I., Nilsson, N.-O., Hallden, C., 1998. Positive correlation between recombination rates and levels of genetic variation in natural populations of sea beet (*Beta vulgaris* subsp. *maritima*). Genetics 150, 1239–1244.

Kreitman, M., 1983. Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. Nature 304, 412–417.

Makova, K.D., Li, W.-H., 2002. Strong male-driven evolution of DNA sequences in humans and apes. Nature 416, 624–626.

Maynard Smith, J., Haigh, J., 1974. The hitch-hiking effect of a favorable gene. Genet. Res. 23, 23–35.

McVean, G.T., Hurst, L.D., 1997. Evidence for a selectively favourable reduction in the mutation rate of the X chromosome. Nature 386, 388–392.

Michalakis, Y., Veuille, M., 1996. Length variation of CAG/CAA trinucleotide repeats in natural populations of *Drosophila melanogaster* and its relation to recombination rate. Genetics 143, 1713–1725.

Miyata, T., Hayashida, H., Kuma, K., Mitsuyasu, K., Yasunaga, T., 1987. Male-driven molecular evolution: a model and nucleotide sequence analysis. Cold Spring Harbor Symp. Quant. Biol. 52, 863–867.

Moriyama, E.N., Powell, J.R., 1996. Intraspecific nuclear DNA variation in Drosophila. Mol. Biol. Evol. 13, 261–277.

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

Nachman, M.W., 2001. Single nucleotide polymorphisms and recombination rate in humans. Trends Genet. 17, 481–485.

Nachman, M.W., Crowell, S.L., 2000a. Estimate of the mutation rate per nucleotide in humans. Genetics 156, 297–304.

Nachman, M.W., Crowell, S.L., 2000b. Contrasting evolutionary histories of two introns of the Duchenne muscular dystrophy gene, *Dmd*, in humans. Genetics 155, 1855–1864.

Nachman, M.W., Bauer, V.L., Crowell, S.L., Aquadro, C.F., 1998. DNA variability and recombination rates at *X*-linked loci in humans. Genetics 150, 1133–1141.

Nei, M., Li, W.-H., 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. USA 76, 5269–5273.

Nordborg, M., Charlesworth, B., Charlesworth, D., 1996. The effect of recombination on background selection. Genet. Res. 67, 159–174.

Ohta, T., 1973. Slightly deleterious substitutions in evolution. Nature 246, 96–98.

Ohta, T., 1995. Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. J. Mol. Evol. 40, 56–63.

Payseur, B.A., Nachman, M.W., 2000. Microsatellite variation and recombination rate in the human genome. Genetics 156, 1285–1298.

Payseur, B.A., Nachman, M.W., 2002. Gene density and human nucleotide polymorphism. Mol. Biol. Evol. 19, 336–340.

Przeworski, M., 2002. The signature of positive selection at randomly chosen loci. Genetics 160, 1179–1189.

Przeworski, M., Hudson, R.R., Rienzo, A.D., 2000. Adjusting the focus on human variation. Trends Genet. 16, 296–302.

Rieder, M.J., Taylor, S.L., Clark, A.G., Nickerson, D.A., 1999. Sequence variation in the human angiotensin converting enzyme. Nat. Genet. 22, 59–62.

Schlötterer, C., Wiehe, T., 1999. Microsatellites, a neutral marker to infer selective sweeps. In: Goldstein, D.B., Schlötterer, C. (Eds.), Microsatellites: Evolution and Applications, Oxford University Press, Oxford.

Schug, M.D., Mackay, T.F.C., Aquadro, C.F., 1997. Low mutation rates of microsatellite loci in *Drosophila melanogaster*. Nat. Genet. 15, 99–102.

Schug, M.D., Mackay, T.F.C., Aquadro, C.F., 1998. Mutation and evolution of microsatellites in *Drosophila melanogaster*. Genetica 102/103, 359–367.

Simonsen, K.L., Churchill, G.A., Aquadro, C.F., 1995. Properties of statistical tests of neutrality for DNA polymorphism data. Genetics 141, 413–429.

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

Stephan, W., Langley, C.H., 1989. Molecular genetic variation in the centromeric region of the X chromosome in three *Drosophila ananassae* populations. I. Contrasts between the *vermillion* and *forked* loci. Genetics 121, 89–99.

Stephan, W., Langley, C.H., 1998. DNA polymorphism in Lycopersicon and crossing-over per physical length. Genetics 150, 1585–1593.

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

Tajima, F., 1989b. The effect of change in population size on DNA polymorphism. Genetics 123, 597–601.

Tenaillon, M.I., Sawkins, M.C., Long, A.D., Gaut, R.L., Doebley, J.F., Gaut, B.S., 2001. Patterns of DNA sequence polymorphism along

chromosome 1 of maize (*Zea mays* ssp. *mays* L.). Proc. Natl. Acad. Sci. USA 98, 9161–9166.

Wang, W., Thornton, K., Berry, A., Long, M.Y., 2002. Nucleotide variation along the *Drosophila melanogaster* fourth chromosome. Science 295, 134–137.

Watterson, G.A., 1975. On the number of segregating sites in genetical models without recombination. Theor. Pop. Biol. 7, 256–276.

Wiehe, T.H.E., 1998. The effect of selective sweeps on the variance of the allele distribution of a linked multi-allele locus – hitchhiking of microsatellites. Theor. Pop. Biol. 53, 272–283.

Wiehe, T.H.E., Stephan, W., 1993. Analysis of a genetic hitchhiking model, and its application to DNA polymorphism data from *Drosophila melanogaster*. Mol. Biol. Evol. 10, 842–854.

Yu, A., Zhao, C., Fan, Y., Jang, W., Mungall, A.J., Deloukas, P., Olsen, A., Doggett, N.A., Ghebranious, N., Broman, K.W., Weber, J.L., 2001a. Comparison of human genetic and sequence-based physical maps. Nature 409, 951–953.

Yu, N., Fu, Y.-X., Sambuughin, N., Ramsay, M., Jenkins, T., Leskinen, E., Patthy, L., Jorde, L.B., Kuromori, T., Li, W.-H., 2001b. Global patterns of human DNA sequence variation in a 10-kb region on chromosome 1. Mol. Biol. Evol. 18, 214–222.

Yu, N., Chen, F.-C., Ota, S., Jorde, L.B., Pamilo, P., Patthy, L., Ramsay, M., Jenkins, T., Shyue, S.-K., Li, W.-H., 2002. Larger genetic differences within Africans than between Africans and Eurasians. Genetics 161, 269–274.

Zhao, Z., Jin, L., Fu, Y.-X., et al., 2000. Worldwide DNA sequence variation in a 10-kb noncoding region on human chromosome 22. Proc. Natl. Acad. Sci. USA 97, 11354–11358.