Microsatellite Variation and Recombination Rate in the Human Genome

Bret A. Payseur and Michael W. Nachman

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

Manuscript received June 2, 1999
Accepted for publication July 19, 2000

ABSTRACT

Background (purifying) selection on deleterious mutations is expected to remove linked neutral mutations from a population, resulting in a positive correlation between recombination rate and levels of neutral genetic variation, even for markers with high mutation rates. We tested this prediction of the background selection model by comparing recombination rate and levels of microsatellite polymorphism in humans. Published data for 28 unrelated Europeans were used to estimate microsatellite polymorphism (number of alleles, heterozygosity, and variance in allele size) for loci throughout the genome. Recombination rates were estimated from comparisons of genetic and physical maps. First, we analyzed 61 loci from chromosome 22, using the complete sequence of this chromosome to provide exact physical locations. These 61 microsatellites showed no correlation between levels of variation and recombination rate. We then used radiation-hybrid and cytogenetic maps to calculate recombination rates throughout the genome. Recombination rates varied by more than one order of magnitude, and most chromosomes showed significant suppression of recombination near the centromere. Genome-wide analyses provided no evidence for a strong positive correlation between recombination rate and polymorphism, although analyses of loci with at least 20 repeats suggested a weak positive correlation. Comparisons of microsatellites in lowest-recombination and highest-recombination regions also revealed no difference in levels of polymorphism. Together, these results indicate that background selection is not a major determinant of microsatellite variation in humans.

Theoretical studies suggest that the joint effects of selection and linkage may lead to broadscale patterns of genetic variation in different genomic regions. Background (purifying) selection on deleterious mutations may reduce levels of linked neutral variation, particularly in genomic regions of reduced recombination (Charlesworth et al. 1993; Charlesworth 1994; Hudson and Kaplan 1995). This process reflects a mutation-selection equilibrium in which heterozygosity is reduced as a function of the deleterious mutation rate, the average selection coefficient and dominance factor for harmful mutations, and the rate of recombination for a given region. Fixation of beneficial mutations may also reduce levels of linked neutral variation as a result of genetic hitchhiking (Maynard Smith and Haigh 1974; Kaplan et al. 1989; Stephan 1995).

A key distinction between the effects of harmful and beneficial mutations on linked neutral variation is that background selection is an equilibrium process while genetic hitchhiking is not (Slatkin 1995; Wiehe 1998).

Consequently, background selection predicts a positive correlation between recombination rate and levels of neutral polymorphism regardless of the neutral mutation rate. Thus, this association is predicted both for polymorphisms at single nucleotide sites (where mutation rates may be on the order of $10^{-8}$; Drake et al. 1998) and for polymorphisms at microsatellite loci (where mutation rates may be on the order of $10^{-4}$; Banchs et al. 1994). In contrast, genetic hitchhiking will reduce variability at linked neutral sites sporadically, and the recovery to steady-state heterozygosity will depend in part on the neutral mutation rate. If selective sweeps are common and the mutation rate is low (as for single nucleotide polymorphisms), a positive correlation is expected between recombination rate and polymorphism. However, if sweeps are rare or if mutation rates are very high (as for microsatellites), a positive correlation is not expected (Wiehe 1998).

Empirical data from several sources show that nucleotide polymorphism is reduced in genomic regions experiencing low rates of recombination. The best evidence for this comes from Drosophila melanogaster. Nucleotide polymorphism is significantly reduced at the tip of the X chromosome (Aguade et al. 1989; Begun and Aquadro 1991) and on the small fourth chromosome (Berry et al. 1991), both of which experience low rates of recombination. More generally, there is an overall positive correlation between recombination rate and nucleotide diversity throughout the D. melanogaster genome (Begun and Aquadro 1992; Aquadro et al. 1994; Moriyama and Powell 1996). Reduced nucleotide variability in low-recombination regions has also been observed in D. simulans and D. pseudoobscura.
D. ananassae (Stephan and Langley 1989), D. simulans (Begun and Aquadro 1991; Berry et al. 1991), D. mauristiana (Hilton et al. 1994), and D. sechellia (Hilton et al. 1994). There is weaker evidence for a positive association between recombination rate and nucleotide diversity in mice (Nachman 1997), humans (Nachman et al. 1998; Przeworski et al. 2000), sea beets (Kraft et al. 1998), tomatoes (Stephan and Langley 1998), and goatgrasses (Dvorak et al. 1998).

One empirical study in which the effect of recombination rate on microsatellite variation has been assessed is in D. melanogaster, where Schug et al. (1998) documented a strong, positive association. However, microsatellite mutation rates in Drosophila are quite low (6 × 10⁻⁶; Schug et al. 1997) and thus the observed correlation may be consistent with either genetic hitchhiking or background selection (Schug et al. 1998).

Here, we assess the relationship between microsatellite variability and recombination rate in humans. Microsatellite mutation rates in humans are known to be high (e.g., 10⁻⁴; Banchs et al. 1994) and the wealth of genetic and physical mapping data makes it possible to estimate recombination rates in different genomic regions. Dut et al. (1996) have constructed a genetic map of the human genome based on 5264 (CA), microsatellite markers genotyped in eight CEPH families comprising 134 individuals. The sex-averaged length of the genetic map is 3699 cM and the average interval size is 1.6 cM. Two different physical maps of the human genome have been assembled from radiation-hybrid (RH) cell panels (Gyapay et al. 1996; Stewart et al. 1997). RH panels consist of somatic cell hybrids, with each cell line containing a random set of fragments of irradiated human genomic DNA in a hamster background. Markers that are physically close to one another are expected to have relatively few chromosomal breaks between them and thus co-occur in hybrid cell lines more often than markers that are far apart. Cell lines are genotyped for numerous markers and the relative positions in centirays (cR) of markers are inferred from the degree of co-occurrence. RH maps provide probabilistic statements about the physical locations of markers and rely on the assumption that radiation-induced break points are randomly distributed. However, there is some evidence that radiosensitivity is influenced by local chromatin composition and that the distribution of radiation-induced breakage events may not be Poisson (Teague et al. 1996). Collins et al. (1996) have also assembled physical maps of the human genome, with marker positions determined largely from in situ hybridization of probes to high-resolution, G-banded metaphase chromosomes. Finally, complete sequences of portions of the human genome (e.g., Dunham et al. 1999) now provide the exact physical location of many markers.

In this article, we compare recombination rates and levels of microsatellite polymorphism in different regions of the human genome. We find no evidence for a strong association between recombination rate and microsatellite polymorphism, suggesting that background selection is not a primary determinant of microsatellite variability in humans.

MATERIALS AND METHODS

Recombination rates: Recombination rates can be estimated from comparison of genetic and physical maps. Error in estimates of recombination rate may derive from one or both of these sources. The ultimate physical map of the human genome will eventually come from the complete sequence of the whole genome; at the time of this writing the sequences of two human chromosomes are available (Dunham et al. 1999; Hattori et al. 2000). Chromosome 21 shows little regional variation in recombination rate (Hattori et al. 2000) while chromosome 22 has several regions with considerably elevated recombination rates (Dunham et al. 1999).

Because chromosome 22 has been completely sequenced and because it displays significant variation in recombination rate, we first analyzed patterns of microsatellite variation on this chromosome in light of its recombinational landscape (as described in detail below). The estimates of recombination rate for chromosome 22 are among the best for the human genome since they derive from physical distances measured directly in base pairs. Even in this situation, however, recombination rates may be inaccurate because of imprecision in the genetic map (which is based on pedigrees rather than crosses and thus is constructed from fewer meioses). Moreover, the relatively low density of markers that have been integrated on both the genetic map and complete sequence implies that small-scale variation in recombination rate may go undetected.

In the second part of this article, we extend our analysis to include the whole genome. Since the complete sequence is not yet available we rely on physical maps based on cytogenetic data (Collins et al. 1996; http://cedar.genetics.soton.ac.uk/public_html/ldb.html, subsequently referred to as “Morton’s map”) or radiation hybrid panels (Gyapay et al. 1996; Stewart et al. 1997; Deloukas et al. 1998; http://www.ncbi.nlm.nih.gov/genemap). There are currently two large-scale RH maps of the human genome, the Genebridge4 (GB4) map (Gyapay et al. 1996) and the Stanford G3 map (Stewart et al. 1997). Although the G3 panel contains a slightly higher level of resolution, many more markers have been placed on the GB4 map. To assess the degree of concordance between different physical maps we have compared the positions of co-occurring microsatellites on the GB4 map, the G3 map, Morton’s map, and the complete sequence of chromosome 22 (Table 1). In all cases, these different physical maps were highly significantly correlated (Kendall’s correlation analyses, \( P < 10^{-4} \) for each comparison). This suggests that these physical maps, which were constructed independently, are relatively accurate.

Recombination rates were calculated separately using four different physical maps: the complete sequence of chromosome 22, the GB4 radiation hybrid map, the G3 radiation hybrid map, and Morton’s map. Recombination rates were calculated two different ways. In the first approach, we used a sliding window encompassing five markers on either side of the locus of interest. When the locus of interest was situated at or near the edge of a chromosome, windows were constructed to include five markers proximal to the locus of interest and zero to five markers distal to the locus of interest. For example, for a microsatellite situated at the end of a chromosome, only six markers were included in the window. Thus, recombination rates estimated using the sliding-window method at the edges of chromosomes are based on fewer data.
and may be somewhat biased in these regions (NACHMAN and CHURCHILL 1996). For each sliding window, a linear function was fit to the points representing genetic (DIB et al. 1996) and physical map positions, and the slope of this line was taken as the estimate of recombination rate. We also used the sliding-window approach with third- and fifth-order polynomial curve fitting and estimated recombination rate as the derivative of the function at the locus of interest. Using polynomial rather than linear functions had little effect on recombination rate estimates and so we report only the estimates from linear functions. In the second approach (subsequently referred to as the “whole-chromosome” method), we fit a third-order polynomial to the genetic and physical positions of all the markers on a chromosome. The derivative of this function, evaluated for each marker, was taken as the estimate of recombination rate. Estimating recombination rate by simultaneously using all markers on a given chromosome has a smoothing effect (KLIMAN and HEY 1993), and probably obscures much regional variation, which is better captured by the sliding-window method. Conversely, estimates under the sliding-window approach are more strongly impacted by errors in the genetic and physical mapping locations of individual markers.

All analyses were completed using both the sliding-window and the whole-chromosome estimates of recombination rates. Although sliding-window and whole-chromosome recombination rates differed in some cases, none of our conclusions were affected by these differences; in most cases we report results from the sliding-window analyses only. All analyses were also completed using estimates of recombination rate derived from each of the four physical maps. While estimates differed in some cases, none of our conclusions were affected by these differences and we report results only from the complete sequence of chromosome 22, the GB4 map, and Morton’s map. Results from the G3 map were similar and are available upon request from the authors. Recombination rates calculated on the basis of the GB4 map were converted from centimorgans per centimorgan to centimorgans per megabase to facilitate comparison to other estimates of human recombination rates using the conversion factors for individual chromosomes given by HUDSON et al. (1995).

**Microsatellite variation:** Data on microsatellite variation were obtained from Dib et al. (1996; http://www.genethon.fr). All microsatellites consist of (CA)n dinucleotide repeats, genotyped in unrelated European individuals (56 autosomes and 108 X chromosomes). Dib et al. (1996) used an initial screen for polymorphism to isolate markers: only loci that contained at least three distinct alleles among 8 chromosomes were included. This screen may have created an ascertainment bias for our analyses by excluding loci that are monomorphic or nearly monomorphic (see RESULTS).

The number of alleles and observed heterozygosity for each locus were taken from Dib et al. (1996). The variance in allele size for each marker was calculated as

\[
\frac{n}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2
\]

(SOKAL and ROHLF 1995), where \( f_i \) is the frequency of the \( i \)th allele, \( x_i \) is the number of repeats at the \( i \)th allele, \( \bar{x} \) is the average number of repeats weighted by frequency, and \( n \) is the number of alleles. For a subset of markers placed on the physical maps, microsatellite length was determined by counting the longest string of CA dinucleotide repeats from the published sequence (Dib et al. 1996). Each locus was scored as a perfect (i.e., uninterrupted) repeat or an imperfect repeat.

Because background selection is an equilibrium model, its predictions are only strictly valid for randomly mating populations. One of the CEPH families from which data have been drawn for this study descends from an admixed population (BEGOVICH et al. 1992), potentially violating the assumption of random mating. However, inclusion of this family is unlikely to have biased our results for two reasons. First, the contribution of this family to the sample is only four chromosomes. Second, linkage disequilibrium induced by nonrandom mating would cause an overestimation of degrees of freedom for statistical analyses, increasing the likelihood of rejecting the null hypothesis of no correlation between recombination rate and microsatellite variation. Because we find no strong statistical evidence for such an association, this bias is conservative relative to our conclusions.

**Statistical analyses:** Recombination rate, heterozygosity, number of alleles, and variance in allele size for microsatellites were nonnormally distributed (Shapiro-Wilk’s test, \( P < 0.01 \) for each distribution). Nonparametric Kendall’s correlation analyses were used to characterize the relationships between measures of microsatellite polymorphism and recombination rate. Analyses were conducted for chromosome 22 first (on the basis of recombination rates estimated from physical positions in the complete sequence) and then at the genomic level for the entire dataset and separately for each chromosome (on the basis of recombination rates estimated from physical positions on the GB4 map and on the Morton map). Because we performed multiple tests, we adjusted the statistical significance level for all analyses. There were 12 tests performed per chromosome, and we used a Bonferroni correction (SOKAL and ROHLF 1995) by adjusting the significance level to \( P = 0.05/12 = 0.004 \).

Heterogeneity in mutation rates might obscure a relationship between microsatellite variation and recombination rate (SCHUG et al. 1998). In humans, evidence has accumulated for a positive association between microsatellite mutation rate

---

### TABLE 1

**Comparison of the physical position of co-occurring microsatellite markers on the GB4, G3, and Morton physical maps and the complete sequence of chromosome 22**

<table>
<thead>
<tr>
<th>Marker</th>
<th>GB4</th>
<th>G3</th>
<th>Morton</th>
<th>Chromosome 22</th>
</tr>
</thead>
<tbody>
<tr>
<td>GB4</td>
<td>—</td>
<td>0.81 (10^-4)</td>
<td>0.78 (10^-4)</td>
<td>1.00 (10^-4)</td>
</tr>
<tr>
<td>G3</td>
<td>0.94 (10^-4)</td>
<td>—</td>
<td>0.82 (10^-4)</td>
<td>0.92 (10^-4)</td>
</tr>
<tr>
<td>Morton</td>
<td>0.93 (10^-4)</td>
<td>0.95 (10^-4)</td>
<td>—</td>
<td>0.99 (10^-4)</td>
</tr>
<tr>
<td>Chromosome 22</td>
<td>0.99 (10^-4)</td>
<td>0.99 (10^-4)</td>
<td>0.93 (10^-4)</td>
<td>—</td>
</tr>
</tbody>
</table>

Nonparametric Kendall’s \( \tau \) values are given above the diagonal, and parametric correlation coefficients are given below the diagonal; \( P \) values are in parentheses.
and the number of repeats (Weber 1990; Brinkmann et al. 1998; Di Rienzo et al. 1998). In an attempt to control for the effects of mutation rate variation, we completed additional analyses with restrictions on allele size (20 or more repeats) and kind (perfect repeats).

Finally, we used Mann-Whitney U-tests to compare levels of microsatellite variation in high vs. low recombination regions, including markers in the upper and lower 10% of the recombination rate distribution.

RESULTS

Chromosome 22: Mean recombination rates, variance in allele size, heterozygosity, number of alleles, and number of repeats for the 61 microsatellite loci on chromosome 22 are given in Table 2. Sliding-window estimates of recombination rate varied by approximately one order of magnitude, from 0.47 to 4.57 cM/Mb. Higher recombination rates were found primarily in several distinct regions in the middle of the chromosome, as previously reported (Dunham et al. 1999). Variance in allele size ranged from 0.90 to 63.36. There was no correlation between recombination rate and variance in allele size for the 61 loci on chromosome 22 (Kendall’s correlation analyses; $\tau = -0.118; P = 0.181$); a scatterplot of these data is shown in Figure 1. Similar results were obtained using heterozygosity ($\tau = -0.108; P = 0.218$) and number of alleles ($\tau = -0.134; P = 0.126$) as measures of microsatellite polymorphism. Furthermore, when correlation analyses of microsatellite variation and recombination rate were restricted to loci with at least 20 repeats, no association emerged ($P > 0.05$ using each measure of polymorphism). These results stand in contrast to the strong association between microsatellite variation and recombination rate observed in Drosophila ($R^2 = 0.55$; Schug et al. 1998), where only 18 loci were surveyed.

Among the 61 microsatellites considered above, 25 have been placed on Morton’s map (and only a few have been placed on the radiation hybrid maps). A scatterplot comparing recombination rates estimated from the complete sequence of chromosome 22 and those estimated from Morton’s map is shown in Figure 2. There is a positive correlation between these estimates (correlation coefficient $R = 0.578, P = 0.003$; nonparametric Kendall’s $\tau = 0.180, P = 0.207$), suggesting that recombination rates estimated from these different underlying physical maps are consistent with each other. It should be pointed out that this is a relatively weak test since it only includes 25 loci and each sliding window is based on 11 markers (5 on either side of the locus of interest).

Genome-wide recombination rates: Complete tables listing the genetic positions, physical positions, and estimated recombination rates for microsatellite markers placed on the GB4 map and Morton’s map are given at http://eebweb.arizona.edu/nachman/publications/data/microsats.html. Substantial variation in recombination rate was observed both within and among chromosomes. For the GB4 map, the mean recombination rate was 1.55 cM/Mb (sliding-window method) and 1.46 cM/Mb (whole-chromosome method), with low values <0.5 cM/Mb and high values >6 cM/Mb (Table 3). For Morton’s map, the mean recombination rate was

![Figure 1.—Scatterplot of variance in allele size vs. recombination rate for 61 microsatellite loci on chromosome 22. Recombination rates were calculated from the genetic (Dib et al. 1996) and physical positions of loci in the complete sequence of chromosome 22 (Dunham et al. 1999). There is no correlation between variance in allele size and recombination rate (Kendall’s $\tau = -0.118; P = 0.181$).](image)

### TABLE 2

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Sliding-window recombin. rate (cM/Mb)</th>
<th>Whole-chromosome recombin. rate (cM/Mb)</th>
<th>Variance in allele size</th>
<th>Heterozygosity (%)</th>
<th>No. of alleles</th>
<th>No. of repeats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1.82</td>
<td>1.34</td>
<td>11.07</td>
<td>69.71</td>
<td>7.57</td>
<td>19.10</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.47</td>
<td>1.15</td>
<td>0.90</td>
<td>42.00</td>
<td>2.00</td>
<td>12.00</td>
</tr>
<tr>
<td>Maximum</td>
<td>4.57</td>
<td>2.31</td>
<td>63.36</td>
<td>87.00</td>
<td>16.00</td>
<td>34.00</td>
</tr>
<tr>
<td>Standard error of mean</td>
<td>0.13</td>
<td>0.05</td>
<td>1.53</td>
<td>1.45</td>
<td>0.32</td>
<td>0.63</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>1.01</td>
<td>0.26</td>
<td>11.91</td>
<td>11.31</td>
<td>2.53</td>
<td>4.90</td>
</tr>
</tbody>
</table>
Microsatellite Variation in Humans

Figure 2—Scatterplot of recombination rates estimated using Morton’s map vs. recombination rates estimated using the sequence of chromosome 22 (correlation coefficient $R = 0.578$, $P = 0.003$; nonparametric Kendall’s $\tau = 0.180$, $P = 0.207$).

1.72 cM/Mb (sliding-window method) and 1.37 cM/Mb (whole-chromosome method), with low values $<$0.5 cM/Mb and high values $>20$ cM/Mb (Table 4). The highest recombination rates on Morton’s map (i.e., $>10$ cM/Mb) derive from two regions, one on the p arm of chromosome 5 and one on the p arm of chromosome 7. Neither region is well represented on the GB4 map, and this may account for the absence of recombination rates $>10$ cM/Mb on this map. Alternatively, these exceptionally high rates may reflect errors in the physical position of markers on Morton’s map. Several windows on Morton’s map resulted in negative values for recombination rates; these values derive from inconsistencies in marker order on the Genethon (Dib et al. 1996) and Morton maps (Collins et al. 1996).

Estimates of recombination rate from the GB4 and Morton maps were highly correlated (Kendall’s $\tau = 0.215$; $P < 0.0001$), and several general patterns emerged from both maps. First, scatterplots of genetic vs. physical map position for the markers on each chromosome typically reveal sigmoidal curves, GB4 scatterplots are shown in Figure 3, Morton’s map scatterplots are not shown but are similar. This pattern is seen for most metacentric chromosomes and is consistent with a reduction in recombination rate near centromeres, as previously documented (e.g., Nagaraja et al. 1997). Second, several metacentric chromosomes displayed high levels of recombination near one or both telomeres. For example, the highest recombination rates calculated from the GB4 map are for loci at the q telomere of chromosome 2. Third, in general, the acrocentric chromosomes (13, 14, 15, 21, and 22) revealed less variation in rates of recombination than did the metacentric chromosomes.

The mean recombination rates estimated using the sliding-window method were close to the mean values obtained from the whole-chromosome method (Tables 3 and 4). However, the variance in recombination rates was larger for the sliding-window estimates than for the whole-chromosome estimates. Recombination rate estimates from the two approaches were highly correlated (Morton’s map: $\tau = 0.26$, $P < 0.0001$; GB4 map: $\tau = 0.59$, $P < 0.0001$). In subsequent analyses, we report results using the sliding-window approach, but similar results are obtained using whole-chromosome estimates of recombination rate.

**Genome-wide microsatellite variation:** Variation in number of alleles, heterozygosity, and variance in allele size for the microsatellite loci is given in Tables 3 and 4. Substantial variation in levels of microsatellite polymorphism was observed for loci on both maps. For microsatellite loci on the GB4 map the mean variance in allele size was 13.95 and ranged from a minimum of 0.37 to a maximum of 303.21. For loci on the Morton map, the mean variance in allele size was 13.67 and ranged from a minimum of 0.19 to a maximum of 474.29. Under a stepwise mutation model, variance in allele size provides an estimate of the neutral mutation parameter, $2(N_e - 1)\mu$, where $N_e$ is the effective population size and $\mu$ is the neutral mutation rate (Moran 1975). Assuming an average mutation rate of $10^{-4}$ to $10^{-5}$ (Banchs et al. 1994), the mean variance in allele size for loci on the GB4 map (13.95) suggests an effective

### TABLE 3

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Sliding-window recombination rate (cM/Mb)</th>
<th>Whole-chromosome recombination rate (cM/Mb)</th>
<th>Variance in allele size</th>
<th>Heterozygosity (%)</th>
<th>No. of alleles</th>
<th>No. of repeats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1.55</td>
<td>1.46</td>
<td>13.95</td>
<td>70.72</td>
<td>7.68</td>
<td>19.48</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.11</td>
<td>0.20</td>
<td>0.37</td>
<td>21.00</td>
<td>2.00</td>
<td>8.00</td>
</tr>
<tr>
<td>Maximum</td>
<td>7.91</td>
<td>6.57</td>
<td>303.21</td>
<td>92.00</td>
<td>27.00</td>
<td>50.00</td>
</tr>
<tr>
<td>Standard error of mean</td>
<td>0.03</td>
<td>0.02</td>
<td>0.49</td>
<td>0.29</td>
<td>0.07</td>
<td>0.12</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>1.03</td>
<td>0.81</td>
<td>20.19</td>
<td>11.85</td>
<td>2.80</td>
<td>4.83</td>
</tr>
</tbody>
</table>
population size of $0.7 \times 10^5$, an estimate in rough agreement with others based on nucleotide polymorphisms from a variety of independent loci ($\sim 10^6$; e.g., Hamm 1995). The mean heterozygosity on the GB4 map was 70% and ranged from 21 to 92% (Table 3). Effective population size can also be calculated from heterozygosity ($H$) under a stepwise mutation model, where $H = 1 - \frac{1}{\sqrt{1 + 8N_e \mu}}$ (Ohta and Kimura 1973). With mutation rates of $10^{-4}$ to $10^{-5}$, the observed heterozygosity of 0.70 suggests $N_e = 0.13 \times 10^5$. These calculations of $N_e$ assume equilibrium conditions and constant mutation rates among loci, and, as such, are intended to provide only rough estimates. On the GB4 map, the mean number of alleles per locus was 7.68 and the average number of repeats was 19.48, illustrating the fact that humans tend to have longer microsatellites (with potentially higher mutation rates) than D. melanogaster (Schug et al. 1997). Despite some ascertainment bias present in the original choice of markers used to construct the genetic map, the results in Tables 3 and 4 indicate that there is still substantial variation in all measures of polymorphism. Moreover, estimates of $N_e$ from these data are not much higher than estimates of $N_e$ from other genetic data.

Comparison of genome-wide recombination rate and microsatellite variation: Using two genome-wide datasets (loci on Morton’s map and loci on the GB4 map), we compared three measures of microsatellite variability (variance in allele size, heterozygosity, and number of alleles) to recombination rate. We also restricted these analyses to loci with perfect repeats, loci with at least 20 repeats, and loci with at least 20 perfect repeats. Results of these analyses are shown in Tables 5 and 6. Although some correlations exhibited low probabilities (especially for loci on Morton’s map), none were significant when corrected for multiple tests. In all cases, the magnitudes of the correlations were small. Scatterplots of microsatellite polymorphism vs. recombina-

---

**Figure 3.** Scatterplots of genetic vs. physical map position for microsatellites on the GB4 map. Sigmoidal curves for many of the chromosomes are indicative of lower levels of recombination near centromeres of metacentric chromosomes.
Figure 3.—Continued.
TABLE 4

Recombination rate and microsatellite variability for 3180 loci on Morton’s map (COLLINS et al. 1996)

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Sliding-window recombination rate (cM/Mb)</th>
<th>Whole-chromosome recombination rate (cM/Mb)</th>
<th>Variance in allele size</th>
<th>Heterozygosity (%)</th>
<th>No. of alleles</th>
<th>No. of repeats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1.72</td>
<td>1.37</td>
<td>13.67</td>
<td>69.78</td>
<td>7.61</td>
<td>19.33</td>
</tr>
<tr>
<td>Minimum</td>
<td>-1.54*</td>
<td>0.44</td>
<td>0.19</td>
<td>17.00</td>
<td>2.00</td>
<td>8.00</td>
</tr>
<tr>
<td>Maximum</td>
<td>25.45</td>
<td>19.22</td>
<td>474.29</td>
<td>94.00</td>
<td>27.00</td>
<td>50.00</td>
</tr>
<tr>
<td>Standard error of mean</td>
<td>0.04</td>
<td>0.03</td>
<td>0.37</td>
<td>0.22</td>
<td>0.05</td>
<td>0.14</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>2.08</td>
<td>1.48</td>
<td>21.06</td>
<td>12.13</td>
<td>2.81</td>
<td>4.66</td>
</tr>
</tbody>
</table>

\* Using the sliding-window method, 24 loci have negative estimates of recombination rate. These values derive from inconsistencies in marker order on the genetic and physical maps.

Table 5

Correlation analyses of microsatellite variation and recombination rate for 1635 loci on the GB4 radiation hybrid map

<table>
<thead>
<tr>
<th>Measure of variability</th>
<th>Restrictions</th>
<th>Sample size</th>
<th>Kendall’s ( \tau )</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance in allele size</td>
<td>None</td>
<td>1635</td>
<td>0.014</td>
<td>0.380</td>
</tr>
<tr>
<td>Variance in allele size</td>
<td>Perfect</td>
<td>1072</td>
<td>-0.005</td>
<td>0.813</td>
</tr>
<tr>
<td>Variance in allele size</td>
<td>( \geq 20 ) repeats</td>
<td>742</td>
<td>0.050</td>
<td>0.042</td>
</tr>
<tr>
<td>Variance in allele size</td>
<td>Perfect, ( \geq 20 ) repeats</td>
<td>404</td>
<td>0.021</td>
<td>0.529</td>
</tr>
<tr>
<td>Heterozygosity</td>
<td>None</td>
<td>1635</td>
<td>0.008</td>
<td>0.618</td>
</tr>
<tr>
<td>Heterozygosity</td>
<td>Perfect</td>
<td>1072</td>
<td>0.008</td>
<td>0.703</td>
</tr>
<tr>
<td>Heterozygosity</td>
<td>( \geq 20 ) repeats</td>
<td>742</td>
<td>0.008</td>
<td>0.759</td>
</tr>
<tr>
<td>Heterozygosity</td>
<td>Perfect, ( \geq 20 ) repeats</td>
<td>404</td>
<td>0.020</td>
<td>0.548</td>
</tr>
<tr>
<td>No. of alleles</td>
<td>None</td>
<td>1635</td>
<td>0.016</td>
<td>0.330</td>
</tr>
<tr>
<td>No. of alleles</td>
<td>Perfect</td>
<td>1072</td>
<td>0.006</td>
<td>0.998</td>
</tr>
<tr>
<td>No. of alleles</td>
<td>( \geq 20 ) repeats</td>
<td>742</td>
<td>0.037</td>
<td>0.131</td>
</tr>
<tr>
<td>No. of alleles</td>
<td>Perfect, ( \geq 20 ) repeats</td>
<td>404</td>
<td>-0.004</td>
<td>0.966</td>
</tr>
</tbody>
</table>
TABLE 6
Correlation analyses of microsatellite variation and recombination rate for 3180 loci on Morton's map

<table>
<thead>
<tr>
<th>Measure of variability</th>
<th>Restrictions</th>
<th>Sample size</th>
<th>Kendall’s $\tau$</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance in allele size</td>
<td>None</td>
<td>3180</td>
<td>0.019</td>
<td>0.102</td>
</tr>
<tr>
<td>Variance in allele size</td>
<td>Perfect</td>
<td>789</td>
<td>0.015</td>
<td>0.526</td>
</tr>
<tr>
<td>Variance in allele size</td>
<td>$\geq 20$ repeats</td>
<td>529</td>
<td>0.058</td>
<td>0.044</td>
</tr>
<tr>
<td>Variance in allele size</td>
<td>Perfect, $\geq 20$ repeats</td>
<td>296</td>
<td>0.055</td>
<td>0.158</td>
</tr>
<tr>
<td>Heterozygosity</td>
<td>None</td>
<td>3180</td>
<td>0.018</td>
<td>0.138</td>
</tr>
<tr>
<td>Heterozygosity</td>
<td>Perfect</td>
<td>789</td>
<td>0.058</td>
<td>0.015</td>
</tr>
<tr>
<td>Heterozygosity</td>
<td>$\geq 20$ repeats</td>
<td>529</td>
<td>0.059</td>
<td>0.041</td>
</tr>
<tr>
<td>Heterozygosity</td>
<td>Perfect, $\geq 20$ repeats</td>
<td>296</td>
<td>0.085</td>
<td>0.030</td>
</tr>
<tr>
<td>No. of alleles</td>
<td>None</td>
<td>3180</td>
<td>0.022</td>
<td>0.068</td>
</tr>
<tr>
<td>No. of alleles</td>
<td>Perfect</td>
<td>789</td>
<td>0.025</td>
<td>0.293</td>
</tr>
<tr>
<td>No. of alleles</td>
<td>$\geq 20$ repeats</td>
<td>529</td>
<td>0.065</td>
<td>0.025</td>
</tr>
<tr>
<td>No. of alleles</td>
<td>Perfect, $\geq 20$ repeats</td>
<td>296</td>
<td>0.026</td>
<td>0.499</td>
</tr>
</tbody>
</table>

ance in allele size for the high-recombination loci and for the low-recombination loci are shown in Figure 6. Although a slight difference in variation can be seen in this figure (there are more low-polymorphism loci in regions of low recombination, for example), both highly polymorphic and nearly monomorphic loci can be found in each group.

DISCUSSION

Chromosome 22: The complete sequence of chromosome 22 provides the unambiguous physical position of 61 microsatellite markers that have been genetically mapped (Dib et al. 1996). Consequently, recombination rates can be estimated with more certainty for chromosome 22 than for much of the genome. Recombination rates vary by one order of magnitude and levels of microsatellite polymorphism (variance in allele size) vary by a factor of 70 on chromosome 22. Chromosome 22 thus provides a good test of the simple prediction that recombination rates and levels of microsatellite polymorphism should be correlated under background selection. As shown in Figure 1, we find no evidence of such an association. These results stand in contrast to the findings of Schug et al. (1998), who reported a strong positive correlation between variance in microsatellite allele size and recombination rate for 18 markers throughout the D. melanogaster genome. Possible explanations for these differing results are discussed below. It should be noted that with the complete sequence of the human genome, comparable analyses will soon be possible for other human chromosomes.

Recombination rates throughout the human genome: The average rate of recombination across the human genome from comparison of genetic and physical maps is $\sim 1.5$ cM/Mb. This average value, however, masks the substantial variation in recombination rate that exists in different genomic regions. There is broad concordance between the sliding-window and whole-chromosome methods of estimating recombination rates and both methods show several consistent patterns, including the suppression of recombination near centromeres of metacentric chromosomes. Relatively little is known about the recombinational landscape in humans or the scale at which recombination rates vary. Several studies have revealed recombinational hotspots (e.g., Oudet et al. 1992; Harding et al. 1997), suggesting that recombination rates may vary substantially over a scale of several kilobases. If so, then the whole-chromosome approach to quantifying variation in recombination rate is likely to have a smoothing effect that will obscure important differences in recombination rate. However, even the sliding-window approach implemented here will not detect fine-scale variation in rate since the average spacing of markers integrated between the genetic and physical maps is on the order of 1–2 Mb. Moreover, the effective resolution of the RH panels ($\sim 1$ Mb for the GB4 map) is insufficient to detect fine-scale variation. Furthermore, the distances in centirays given on radiation hybrid maps are probabilistic statements about relative physical positions and do not correspond precisely with distances in base pairs. While the complete sequence of the human genome will soon provide the ultimate physical map, estimates of recombination rates in humans will still depend on the precision of genetic maps that are limited by reliance on pedigrees rather than crosses.

Recombination rate and microsatellite variation: Genome-wide analyses are completely consistent with the results from chromosome 22 and do not support the hypothesis of a strong positive correlation between microsatellite polymorphism and recombination rate (Tables 5 and 6, Figure 4). Similarly, comparisons between loci experiencing the highest and lowest levels of recombination are inconsistent with the notion that recombination rate strongly affects levels of microsatellite
polymorphism (Figure 6). For example, both groups contain nearly monomorphic loci and both groups also contain highly polymorphic loci. Recombination rate (as estimated here) does not appear to be a major determinant of microsatellite variability in humans. These patterns can be contrasted with one recent study in D. melanogaster, where variation in recombination rate explained 55% of the variation in variance in allele size for a set of 18 microsatellite loci throughout the genome (Schug et al. 1998). In that study, none of the high-recombination loci were monomorphic and none of the low-recombination loci were highly polymorphic.

At least four factors may contribute to the substantial scatter in Figure 4 and we consider each of these in turn: (i) imprecision of estimates of recombination rate, (ii) ascertainment bias, (iii) variation in mutation rate, and (iv) locus-specific effects (such as selection).

Four observations suggest that imprecise estimates of genome-wide recombination rates are not obscuring an otherwise strong correlation. First, the estimates of very low levels of recombination near centromeres are likely to be reasonably accurate since they agree well with other studies based on different genetic and physical maps (e.g., Wang et al. 1994; Nagaraja et al. 1997). In these lowest-recombination regions, a substantial fraction of the loci exhibit moderate to high levels of polymorphism (Figures 4 and 6). Second, the physical positions of markers on the GB4, G3, and Morton maps are in good agreement (see MATERIALS AND METHODS), and recombination rates estimated from these different maps are strongly correlated. Microsatellite polymorphism and recombination rate are not strongly correlated when recombination rates are calculated from any of these maps. Third, the results from the genome-wide analysis are entirely consistent with the results from chromosome 22, where we have better estimates of recombination rate. Fourth, a positive correlation is observed between nucleotide variability and recombination rates estimated from the GB4 map (Przeworski et al. 2000), suggesting that these recombination rates are sufficiently precise to detect a correlation when one exists.

Ascertainment bias in the original choice of loci may also be hiding a stronger association between recombination rate and microsatellite polymorphism. If all of the monomorphic or nearly monomorphic loci ex-
slipped strand mispairing, for example, mutation rates national distance is about five times higher in D. melano-
crosatellite mutation rates are active topics of debate given region will be a function of the number of genes substantial heterogeneity within a given length class. genomic region in question. As a first approximation, we size for that locus. Second, repeat length is not a perfect important in humans? The strength of background se-
talleles of different size, a randomly chosen allele may Are there inherent differences between Drosophila repeats and by performing multiple regressions including satellites. Our results demonstrate that variation in re-
morphism. If we assume that all loci experience an effec-
tion in allele size (Figure 5) but not between recombi-
cent variables. These corrections may be insuf®cient tion in microsatellite polymorphism in humans. Hence, have tried to correct for variation in underlying muta-
tion rate by restricting analyses to long and perfect repeats and by performing multiple regressions including both recombination rate and repeat length as indepen-
dent variables. These corrections may be insufficient for several reasons. First, repeat length was measured from one sequenced allele and may not represent the mean repeat length for a locus. For loci that have many alleles of different size, a randomly chosen allele may be substantially shorter or longer than the mean allele size for that locus. Second, repeat length is not a perfect predictor of mutation rate, and it is likely that there is substantial heterogeneity within a given length class. The mechanisms responsible for heterogeneity in microsatellite mutation rates are active topics of debate (Di Rienzo et al. 1998). If these loci mutate through slipped strand mispairing, for example, mutation rates may be influenced by genomic variation in chromatin structure. Thus, despite our efforts to control for variation in mutation rate, substantial differences in mutation rate may exist and be responsible for much of the heterogeneity in levels of polymorphism in Figure 4.

A final possibility is that many of the microsatellites are experiencing the effects of selection on closely linked loci. Balancing selection is expected to elevate levels of polymorphism, while directional selection will reduce polymorphism (Hudson et al. 1987). Selection acting haphazardly on different loci at different times may contribute to heterogeneity in levels of polymorphism. This is a formal hypothesis that cannot be excluded, although we know of no evidence to support this hypothesis at present.

Of the analyses of individual chromosomes, only chromosome 4 showed any hint of a positive correlation between recombination rate and microsatellite polymorphism (Figure 5), although this result is not signi®cant when corrections for multiple tests are employed. Chromosome 4 contains one of the largest regions of very low recombination rate in the human genome (~200 cR, Figure 3). Thus, it is possible that the effects of selection at linked sites may be more pronounced for chromosome 4 than for other chromosomes. A positive correlation is seen between recombination rate and vari-
ce in allele size (Figure 5) but not between recombi-
nation rate and heterozygosity (not shown). In D. melano-
gaster, Schug et al. (1998) also found that variance in microsatellite allele size was correlated with recombin-
ation rate, while heterozygosity was not. Heterozygosity ranges from 0 to 1, but most microsatellite loci exhibit a narrow range of high heterozygosities from 0.60 to 0.90. There is no such restriction on variance in allele size, which can take on a wide range of values. Heterozy-
gosity may be a less sensitive statistic than variance in allele size for detecting changes in population size or deviations from equilibrium conditions.

**Background selection and genetic hitchhiking:** Background selection predicts a positive correlation between rate of recombination and levels of genetic variation even for markers with high mutation rates such as micro-
satellites. Our results demonstrate that variation in re-
combination rate is not strongly associated with varia-
tion in microsatellite polymorphism in humans. Hence, background selection does not appear to be a major determinant of levels of microsatellite polymorphism in different genomic regions.

Are there inherent differences between Drosophila and humans that might make background selection less important in humans? The strength of background se-
lection depends on the deleterious mutation rate for the genomic region in question. As a first approximation, we can assume that the deleterious mutation rate for a given region will be a function of the number of genes in that region. Overall, the density of genes per recombi-
national distance is about five times higher in D. melano-
Figure 6.—Distributions of variance in allele size for microsatellites experiencing high (upper 10%) and low (lower 10%) rates of recombination. Data are shown for (A) loci on the GB4 map and (B) loci on Morton’s map. A few loci with very high variances in allele size are not shown for ease of visual presentation.


gaster than in humans (100 genes/cM in D. melanogaster compared to 20 genes/cM in humans; Nachman et al. 1998). For the background selection model, the deleterious mutation rate for a region of reduced recombination is more relevant, however. In D. melanogaster, the largest euchromatic region of low recombination is near the centromere of the third chromosome and corresponds to ~15% of the total amount of euchromatin in the genome (Sorsa 1988). The Drosophila genome contains ~14,000 genes (Adams et al. 2000). Therefore, the centromeric region of the third chromosome may contain ~2100 genes (assuming genes are randomly distributed). The low recombination region of human chromosome four comprises ~25% of the length of this chromosome, or 1.5% of the length of the human genome (Morton 1991). The human genome contains ~70,000 genes (Bird 1995) and thus the centromeric region of chromosome 4 may contain roughly 1050 genes, or about half as many as in the centromeric region of chromosome 3 in D. melanogaster. These calculations are very rough and do not take into account possible differences in the genomic deleterious mutation rate between flies and humans (e.g., Drake et al. 1998). Nevertheless, these calculations suggest that, a priori, we might expect the effects of selection at linked sites to be weaker in humans than in flies.

Can we use these results to evaluate models of genetic hitchhiking in humans? Wiehe (1998) has shown that with mutation rates typical of human microsatellites, the selective effects of hitchhiking will quickly be obscured by mutation. When population sizes are small and mutation rates large, as in human populations, the traces of hitchhiking on levels of microsatellite variability are unlikely to be detected by statistical means. Thus, while microsatellite data are appropriate for testing background selection models, they are unlikely to provide a useful test of genetic hitchhiking. Therefore, the results presented here do not speak to the relative importance of background selection and genetic hitchhiking in humans, but do show that background selection is not strongly influencing genomic patterns of microsatellite diversity. Discerning whether or not genetic hitchhiking is important in humans will require large-scale surveys of markers with lower mutation rates (such as nucleotide polymorphisms) in regions of high and low recombination.
We thank Asher Cutter, Chris Hanus, and Scott Payseur for help with analyses. We thank Andy Clark, Chip Aquadro, and two anonymous reviewers for useful comments. We thank John Collins for providing the chromosome 22 marker positions. Members of the Nachman lab gave constructive suggestions during the course of the project. This work was funded by the National Science Foundation.

LITERATURE CITED


Nagaraja, R., S. Macmillan, J. Kere, S. Jones, S. Griffin et al., 1997 X chromosome map at 75-kb STS resolution, revealing extremes of recombination and GC content. Genome Res. 7: 210–222.


Communicating editor: A. G. Clark