Patterns of Diversity and Recombination Along Chromosome 1
of Maize (*Zea mays* ssp. *mays* L.)

Maud I. Tenaillon,*,[1] Mark C. Sawkins,*,[1] Lorinda K. Anderson,[†] Stephen M. Stack,[†] John Doebley[‡] and Brandon S. Gaut*,[2]

*Department of Ecology and Evolutionary Biology, University of California, Irvine, California 92612, †Department of Biology, Colorado State University, Fort Collins, Colorado 80523 and ‡Department of Genetics, University of Wisconsin, Madison, Wisconsin 53706

ABSTRACT

We investigate the interplay between genetic diversity and recombination in maize (*Zea mays* ssp. *mays*). Genetic diversity was measured in three types of markers: single-nucleotide polymorphisms, indels, and microsatellites. All three were examined in a sample of previously published DNA sequences from 21 loci on maize chromosome 1. Small indels (1–5 bp) were numerous and far more common than large indels. Furthermore, large indels (>100 bp) were infrequent in the population sample, suggesting they are slightly deleterious. The 21 loci also contained 47 microsatellites, of which 33 were polymorphic. Diversity in SNPs, indels, and microsatellites was compared to two measures of recombination: C ($=4Nc$) estimated from DNA sequence data and R based on a quantitative recombination nodule map of maize synaptonemal complex 1. SNP diversity was correlated with C ($r = 0.65$; $P = 0.007$) but not with R ($r = -0.10$; $P = 0.69$). Given the lack of correlation between R and SNP diversity, the correlation between SNP diversity and C may be driven by demography. In contrast to SNP diversity, microsatellite diversity was correlated with R ($r = 0.45$; $P = 0.004$) but not C ($r = -0.025$; $P = 0.55$). The correlation could arise if recombination is mutagenic for microsatellites, or it may be consistent with background selection that is apparent only in this class of rapidly evolving markers.

THE interplay between recombination and selection shapes the degree and distribution of genetic variation in a genome. Two theoretical models have been developed to explain the interaction between these two processes. Under the background selection model, deleterious alleles are continuously eliminated from a population, a process that decreases linked neutral genetic variation (CHARLESWORTH *et al.* 1993; CHARLESWORTH 1994; HUDSON and KAPLAN 1995). In contrast, the hitchhiking model posits that selectively advantageous alleles sweep through a population, thereby reducing genetic variation at sites linked to the advantageous allele (MAYNARD-SMITH and HAIGH 1974; KAPLAN *et al.* 1989).

The common thread for both models is the strong influence of recombination. Both models predict that selection (either positive or negative) reduces polymorphism at linked neutral sites, and both predict that loss of polymorphism is greatest in regions of low recombination. The predicted positive correlation between genetic diversity and recombination has been demonstrated empirically in Drosophila (*Drosophila melanogaster*; KAPLAN *et al.* 1989; BEGUN and AQUADRO 1992), humans (NACHMAN *et al.* 1998; PRZEWORSKI *et al.* 2000), mouse

(*Mus domesticus*; NACHMAN 1997), tomato (*Lycopersicon esculentum*; STEPHAN and LANGLEY 1998), sea beet (*Beta vulgaris*; KRAFT *et al.* 1998), and goatgrass (Aegilops; DVORAK *et al.* 1998).

One difference between the two models is that background selection is an equilibrium process, with continuous removal of deleterious alleles from populations (WIEHE 1998). As a result, genetic variation at sites linked to deleterious sites is expected to remain low. Because of the equilibrium dynamics of this process, a positive correlation should be observed between recombination, c, and levels of genetic variation, irrespective of the mutation rate μ. In contrast, the recovery of linked genetic variation under the hitchhiking model depends on both c and μ. If the rate of selective sweeps is low and μ is high (as in microsatellite markers, for example), theory predicts that neutral genetic variation can be restored between rounds of selection, thus masking the effect of a selective sweep (WIEHE 1998). As a result, the positive correlation between diversity and recombination under the hitchhiking model may be obscured when μ is high (WIEHE 1998; PAYSEUR and NACHMAN 2000). Thus, one way to contrast the background and hitchhiking models is to study molecular markers that have different mutation rates.

Demography can also play a large role in the maintenance and distribution of genetic diversity. Population subdivision and population bottlenecks, as well as other demographic factors, can obscure the relationship be-

---

tween recombination and diversity. For example, BAUDRY *et al.* (2001) compared recombination to diversity in regions of differing recombination rate in five Lycopersicon (tomato) species, two of which inbreed at high to intermediate levels. All five species demonstrated a positive correlation between recombination and diversity, but the type of mating system (and presumably the demographic factors associated with the mating system) had a stronger influence on genetic variation than recombination. Thus, for any one system, the distribution of genetic variation within the genome is a complex function of mutation, recombination, selection, and demography.

In maize (*Zea mays* ssp. *mays* L.), genetic diversity has been studied for 21 loci distributed along the genetic map of chromosome 1. TENAILLON *et al.* (2001) found a positive correlation between nucleotide diversity, as measured by single-nucleotide polymorphisms (SNPs), and estimates of the population-recombination parameter $C = 4Nc$, where N is the effective population size. It was presumed that the positive correlation reflects interplay between recombination and selection, but it is important to note that C, which is inversely related to linkage disequilibrium (LD), is also affected by demographic factors. For example, population subdivision increases LD and thus decreases C; similarly, LD is low (and C high) in expanding populations (reviewed in PRITCHARD and PZEWORSKI 2001). Demography also may not affect all loci equally; for example, historical levels of gene flow can differ among loci (WANG *et al.* 1997). Given the effect of demography on C and the possibility that demographic effects differ among loci, demography could have contributed to the positive correlation observed in maize. Thus, the cause of the positive correlation—*i.e.*, background selection, hitchhiking selection, or demography—is unclear. In the previous study, C was estimated from sequence data, and there were no independent estimates of recombination based on a measure that is related to physical distance along the chromosome. The lack of a physical measure of recombination limits our understanding of the forces that shape maize diversity.

Here we investigate further the relationship between recombination and genetic diversity in maize by studying markers that evolve with different μ than SNPs and also by using a physical measure of recombination along maize chromosome 1. To estimate measures of diversity, we reanalyzed data from the 21 genetic loci examined in the previous study (TENAILLON *et al.* 2001). The DNA sequence of each of these loci was determined for $\sim$25 individuals representing much of the geographic range of cultivated maize. All of the 21 loci contained SNPs, and most of the loci contained both microsatellite and insertion-deletion (indel) variation. Only SNP variation was analyzed previously, but microsatellite and indel variation is extensive, representing 24% of the total aligned length of the 21 loci. Thus, the data of TENAIL-

LON *et al.* (2001) offer a unique opportunity to compare diversity among marker types. Moreover, for the microsatellites there was no ascertainment bias for highly polymorphic microsatellites.

In addition to examining measures of diversity, we report a physical measure of recombination based on a quantitative cytogenetic map of the distribution of recombination nodules (RNs) along synaptonemal complex 1 (SC1). During prophase I of meiosis, an SC forms between each homologous pair of chromosomes, and RNs are found associated with SCs at pachytene (ZICKLER and KLECKNER 1999). RNs mark the physical locations of crossovers along SCs (HERICKHOFF *et al.* 1993; SHERMAN and STACK 1995). SC1 can be identified on the basis of its relative length and arm ratio (ANDERSON and STACK 2001), and the frequency and location of RNs along the length of SC1 have been determined. The frequency distribution of RNs along SC1 provides an estimate of recombination along the physical length of the chromosome. Here we report the first RN map for maize chromosome 1, and we use this map to predict the density of crossovers per physical length for each of the 21 loci from the previous study (TENAILLON *et al.* 2001).

Altogether, this study has three objectives. First, we estimate different measures of genetic variation based on SNP, microsatellite, and indel variation. Second, we report physical estimates of recombination (R), based on a quantitative cytogenetic map, for each of the 21 loci. Third, we investigate the influence of recombination on genetic diversity by comparing estimates of both R and C to estimates of diversity. By taking this approach, we intend to provide a better understanding of the interplay between recombination and diversity in maize and begin to provide insight into the relative importance of hitchhiking selection, background selection, and demographic effects.

MATERIALS AND METHODS

DNA polymorphism

Sequence data and analyses: DNA sequence data for 21 loci were obtained from TENAILLON *et al.* (2001; GenBank nos. AF377345–AF377864). The 21 loci represent seven known genes, six cDNA clones, and eight anonymous restriction fragment length polymorphism (RFLP) clones. All 21 loci were located on the UMC98 genetic map (DAVIS *et al.* 1999), and hence their relative locations have been identified on maize chromosome 1. The length of the loci varied from 248 to 2740 bp, with an average length of 648 bp. All 21 loci were originally targeted for sequencing from a common set of 25 individuals of cultivated maize (*Z. mays* L. ssp *mays*). However, some loci were difficult to amplify in some individuals, and thus the data set was not complete. Nonetheless, $\geq$22 individuals were sequenced from each locus, and all 25 individuals were sequenced for 11 of 21 loci. A full description of the plant material and the sequencing protocols were published in TENAILLON *et al.* (2001). For clarity, we hereafter refer to the 21 loci as *genetic* loci, to differentiate them from *microsatellite* loci.

TENAILLON *et al.* (2001) reported SNP diversity in genetic loci by the sequence statistic θ (WATTERSON 1975). Each estimate of θ ($\hat{\theta}$) was a per site value that was based on all of the aligned sites in the sequence data, but calculation of $\hat{\theta}$ did not include gaps. Because it was based only on aligned nucleotide sites, $\hat{\theta}$ does not incorporate any of the diversity found in either polymorphic microsatellite sites or indels.

Microsatellite analyses: To locate microsatellites in DNA sequence data, we performed searches with RepeatMasker (http://repeatmasker.genome.washington.edu/cgi-bin/Repeat Masker) and Ephemeris version 1.0 (http://www.uga.edu/srel/DNA_Lab/ephemeris_readme.htm), as well as manual searches.

Given the lack of consensus regarding the definition of microsatellites in the literature, we based our definition on the expected frequency of occurrence of a microsatellite. Assuming that all nucleotides are present at equal frequencies, the probability of occurrence of a microsatellite is $p_m = 0.25^{x(m-1)}$, with x the length of the motif (*i.e.*, x = 2 for a dinucleotide repeat) and m the number of repeats. We studied microsatellites for which the expected frequency is less than five microsatellites in 10 kb. This frequency corresponds to microsatellites of length m = 7 for mononucleotide repeats; m = 4 for dinucleotide repeats; m = 3 for tri-, tetra-, and pentanucleotide repeats; and m = 2 for hexanucleotide repeats.

For each microsatellite locus, we calculated the number of alleles in our sample (A), the sample variance in allele size (V), and expected heterozygosity ($H_{microsat}$), on the basis of Nei's unbiased estimate (NEI 1973),

$$H = n(1 - \sum p_i^2)/(n - 1),$$

where n is the number of individuals and p_i is the frequency of the ith allele. Genepop ver. 1.2 (RAYMOND and ROUSSET 1995) was used to determine allele frequencies. [We note that the term "heterozygosity" is a misnomer in this case, because all sequenced individuals were homozygotes. In reality, H measures diversity in the sample and is directly comparable to the polymorphic information content (PIC). We use H instead of PIC because it is more widely used in the literature.] The significance of LD between all pairs of microsatellites was tested using the exact test described in Arlequin ver. 2000 (SCHNEIDER *et al.* 2000).

Indel analyses: Nonmicrosatellite indels were also identified and characterized. SITES (HEY and WAKELEY 1997) was used to determine the number, length, and position of indels in the data set. BLAST searches and RepeatMasker, in conjunction with a maize transposable element database (provided by S. Wessler, University of Georgia), were used to determine if large indels corresponded to known mobile elements.

To determine levels of diversity, all identified indels were scored as present (1) or absent (0). This binary data matrix was then transformed into frequencies, and diversity values were calculated using Nei's measure of heterozygosity (H_{indel}), as previously described. To test the neutral mutation hypothesis on large indels, we calculated Tajima's D separately on all indels <100 bp and indels >100 bp, as suggested by CHARLESWORTH and LANGLEY (1989). We used DnaSp ver. 3 (ROZAS and ROZAS 1999) to calculate Tajima's D.

Recombination rate based on RN map

Maize SC karyotype and distribution of RNs: Maize cultivar Kansas Yellow Saline (KYS) was used for the two-dimensional spreads of SCs. Plants were grown to maturity and anthers containing microsporocytes at pachytene were collected. Spreads of SCs were produced using a modification of the procedure described by PETERSON *et al.* (1999) and examined with an AE 801 electron microscope. The positions of kinetochores and RNs were determined for each SC in a set, and the SCs were measured using the computer program Micro-Measure version 3.2 (REEVES 2001). Based on relative lengths and arm ratios, the 10 maize SCs were assigned to the 10 maize pachytene bivalents (2n = 20). Although the absolute lengths of SCs vary in different sets, the relative lengths and arm ratios remain constant (*e.g.*, SHERMAN and STACK 1995). To compare RN positions on SC1 from different sets of SCs, the position of each RN was measured as a fractional length of either the long or short arm from the kinetochore. Then, using an average length of 45.4 μm and average arm ratio of 1.26 for SC1, the position of each RN on an average SC1 was calculated by multiplying the fractional distance of the RN from the kinetochore by the appropriate arm length (see SHERMAN and STACK 1995 for a similar procedure). In total, the positions of 277 RNs on 110 SC1s were mapped.

Recombination rate along chromosome 1: To construct a frequency map of RNs along the physical length of SC1, the total number of RNs observed in each 0.4-μm segment of SC length was determined. The 0.4-μm segment length was chosen to maximize the total number of segments but minimized the number of segments that had no observed RNs. The Lowess procedure (CLEVELAND 1981) was applied to these data to smooth local variation, as suggested by STEPHAN and LANGLEY (1998). The Lowess procedure smoothes the recombination function by applying weighted least-squares regression to sliding windows that are defined by the number of data points. We applied four different sliding window sizes, ranging from 5 to 11 data points, to examine the influence of window size on recombination rate estimates. As is detailed below, the size of sliding windows made little qualitative difference on results. The Lowess procedure was applied in the R statistical package, version 1.2.0 (http://www.r-project.org/).

Determining recombination rate in the 21 loci: The 21 genetic loci were localized on the RN map using an approach similar to that of STEPHAN and LANGLEY (1998). First, the RN distribution was converted into centimorgan map units, following SHERMAN and STACK (1995). Regions with more observed RNs represented regions with greater centimorgan distances. Second, the RN map and the genetic map of maize chromosome 1 (UMC98; DAVIS *et al.* 1999) were aligned in a linear fashion, such that each arm of the SC corresponded to the appropriate arm of the genetic map. Finally, the ratio between the total length of the RN map in cemtimorgans (125.9) and the total length of the UMC98 map of chromosome 1 (249.2 cM) was used to determine the positions of the 21 loci on the RN map. Once positioned along the RN map, we estimated the recombination rate R for each locus as the predicted frequency of occurrence of RNs per micrometer.

The population-recombination parameter, recombination, and diversity

The population-recombination parameter C was estimated from DNA sequence data by three different methods: (i) HUDSON's (1987) method, with estimates taken from TENAILLON *et al.* (2001); (ii) WALL's (2000) method, in which the estimate maximizes the joint probability of obtaining the observed number of minimum recombination events and haplotypes (program provided by J. Wall); and (iii) the program LDhat (http://www.stats.ox.ac.uk/~mcvean/LDhat/LDhat.html), which employs HUDSON's (2001) method with importance sampling (FEARNHEAD and DONNELLY 2001). Estimates based on the three methods were denoted $\hat{C}_{Hud87}$, $\hat{C}_{Wall00}$, and $\hat{C}_{Hud01}$, respectively. All reported C estimates were per site values. We did not use full-likelihood methods to estimate C because

they are computationally infeasible with high levels of recombination (WALL 2000).

We contrasted estimates of R and C with the diversity measures $\hat{\theta}$, $H_{microsat}$, V, A, and H_{indel}. Correlations were determined among measures with Pearson correlation coefficients (r); the significance of r was determined by 10,000 bootstrap resamplings of observed values.

RESULTS

Microsatellite diversity: A total of 47 microsatellites were identified in 18 of the 21 genetic loci. A description of the microsatellite loci including their genetic diversity (H), number of alleles (A), and variance in allele size (V) is presented in Table 1. The 14 monomorphic and 33 polymorphic microsatellites were further characterized as being located in either coding or noncoding sequence (Table 1). The majority occurred in noncoding regions, with 7 noncoding among 14 monomorphic microsatellites and 30 noncoding among 33 polymorphic microsatellites. All polymorphic markers in coding regions were trinucleotide or hexanucleotide repeats and did not induce frameshifts.

We investigated the relationships between three different measures of microsatellite diversity ($H_{microsat}$, V, and A) calculated for the 33 polymorphic microsatellites. $H_{microsat}$ ranged from 0.17 to 0.88, and the average variance in allele size (V) ranged from 0.08 to 60.69 for the longest microsatellite (Table 1). These ranges are comparable to previous estimates from maize (SMITH *et al.* 1997; SENIOR *et al.* 1998; PROVAN *et al.* 1999; MATSUOKA *et al.* 2002a) and other organisms (INNAN *et al.* 1997; SCHUG *et al.* 1998). The correlation between H and V was significantly positive ($r = 0.37$; $P = 0.019$) but this depends, however, on a single data point corresponding to the longest microsatellite [$(CT)_{11-26}$] in the *tb1* locus (Table 1). The correlation was not significant without this data point ($r = 0.11$; $P = 0.72$), suggesting that H and V are weakly related, at best. Additional results involving V were often dependent on the single *tb1*-based data point, an observation that we reiterate. In contrast to the weak correlation between H and V, there was a strong positive correlation between H and A ($r = 0.80$; $P < 0.001$); for the remainder of this article, we ignore A because results based on H were similar (data not shown).

We examined pairwise LD among microsatellite markers. LD was significant at the 5% level in 33 of 528 pairwise comparisons. However, only five associations remained significant after sequential Bonferroni correction, and only one of these five included a pair of microsatellites located in the same locus, *umc67*. Altogether, these analyses are consistent with previous observations that LD in maize breaks down very rapidly over distance (REMINGTON *et al.* 2001; TENAILLON *et al.* 2001).

We calculated the average number of repeats (ANR) for the 38 perfect microsatellites in Table 1 and compared ANR to measures of microsatellite diversity. There

was no significant difference (*t*-test, $P = 0.15$) between ANR within the 13 monomorphic perfect microsatellites (3.8 repeats) and ANR within the 25 polymorphic perfect ones (4.9 repeats). However, there was a significant positive correlation between ANR and $H_{microsat}$ ($r = 0.55$; $P < 0.001$), and the correlation remained significant when only polymorphic microsatellites ($P < 0.001$) were considered. Finally, ANR and V were positively correlated ($r = 0.51$; $P < 0.001$) for polymorphic microsatellite loci. This correlation relied, however, on the single *tb1* data point, and the correlation was not significantly positive without that data point ($r = -0.48$; $P = 0.99$).

Indel variation: A total of 263 nonmicrosatellite indels were scored in 17 of 21 loci. Indel size ranged from 1 to 640 bp, and the number of indels per genetic locus ranged from 2 to 59 (Table 2). A total of 56% of the indels were 1–2 bp in length, and 92% were <20 bp in length (Figure 1). Of the 21 indels longer than 20 bp, 5 were found to have sequence similarity to previously identified transposable elements, including miniature inverted repeat elements (MITEs). Two families of MITEs were found: a *Tourist* element in 1 individual for each of *asg75*, *umc230*, and *csu381* and a *Stowaway* element in 3 of 23 individuals of *umc67*. In addition, BLAST searches revealed the presence of a *Ds* element in 4 of 23 individuals for *umc128*. H_{indel} values ranged from 0.08 to 0.52, with an average of 0.25 among the polymorphic indels.

A previous study in *D. melanogaster* suggested that the frequency distribution of large indels deviated from the neutral equilibrium model, consistent with selection against large indels (TAJIMA 1989). To determine whether maize indels had a nonneutral frequency spectrum, we first studied the relationship between length and indel diversity. If long indels are deleterious, we expect a negative correlation between H_{indel} and indel length, because large deleterious indels have a low probability of reaching an appreciable population frequency and should therefore have low H_{indel} values. A significant negative correlation was found between H_{indel} and indel length ($r = -0.11$; $P = 0.02$), consistent with this expectation. We note that the longer length variant was present in low frequency (*i.e.*, <15%) for 9 of the 10 large indels.

We also studied the relationship between Tajima's D and indel length. Tajima's D was calculated for three different data sets. The first set included all 253 indels <100 bp. However, because the initial data set had missing data entries, all indels could be identified in a common sample of only 13 individuals. Tajima's D for this data set of 253 indels and 13 individuals was $D_{253-13} = -0.43$, which was not a significant deviation from the neutral expectation of 0.0, assuming no recombination. We also calculated Tajima's D in a data set that included all 25 individuals and a common sample of 55 indels that were <100 bp; D_{55-25} was -0.66 for this data set and again did not deviate from the neutral model, assuming no recombination. Finally, we calculated D for

TABLE 1

Characterization and variability in microsatellites found in 18 of 21 loci

Locus	Microsatellite[a]	C/NC[b]	H	A	V
adh1	$(AC)_4(AT)_{1-2}GT(AT)_3$	NC	0.50	2	1.00
adh1	$(TA)_2TG(TA)_2$	NC	0.00	1	0.00
adh1[c]	$(T)_{6-7}$	NC	0.33	2	0.16
an1[c]	$(GGA)_3$	C	0.00	1	0.00
asg11	$(C)_{5-9}(A)_{3-7}$	NC	0.17	3	6.02
asg26[c]	$(A)_{5-10}$	NC	0.30	3	0.82
asg26[c]	$(A)_{6-8}$	NC	0.37	3	0.49
csu1132[c]	$(CATTTC)_{1-2}$	NC	0.22	2	3.96
csu1138[c]	$(GCTA)_{2-3}$	NC	0.33	2	2.67
csu1150[c]	$(CATC)_{2-3}$	NC	0.51	2	4.11
csu1171[c]	$(T)_{8-10}$	NC	0.53	3	0.34
csu1171[c]	$(CT)_4$	NC	0.00	1	0.00
csu381[c]	$(GT)_{3-4}$	NC	0.42	2	0.84
csu381[c]	$(T)_{8-12}$	NC	0.65	4	0.83
fus6[c]	$(GA)_4$	NC	0.00	1	0.00
fus6[c]	$(T)_{7-11}$	NC	0.61	5	1.33
d8[c]	$(GGC)_3$	C	0.00	1	0.00
d8[c]	$(TGG)_4$	C	0.00	1	0.00
glb1[c,d]	$(AC)_{3-6}$	NC	0.55	3	2.44
glb1[c]	$(ACAGTG)_{1-2}$	NC	0.17	2	2.98
glb1[c]	$(CCA)_{2-4}$	C	0.24	2	4.27
glb1[c]	$(TTA)_{4-7}$	NC	0.68	4	6.12
glb1[c,d]	$(AGG)_{4-5}$	C	0.69	4	3.08
glb1[c,d]	$(GCC)_{2-3}$	NC	0.17	2	0.08
glb1[c]	$(GGA)_3$	C	0.00	1	0.00
glb1[c]	$(CGGTCC)_{1-2}$	NC	0.47	2	8.54
asg62[c]	$(TCTCCT)_{1-2}$	NC	0.28	2	5.04
tb1[c]	$(A)_{10-14}$	NC	0.78	5	1.61
tb1[c]	$(CT)_{11-26}$	NC	0.88	10	60.69
tb1[c]	$(GGA)_4$	C	0.00	1	0.00
tb1[c]	$(AC)_4$	NC	0.00	1	0.00
tb1[c]	$(GCC)_3$	C	0.00	1	0.00
tb1[c]	$(A)_7$	NC	0.00	1	0.00
ts2	$(GAC)_{2-3}C(TG)_{1-5}$	NC	0.47	4	5.07
ts2[c]	$(CG)_4$	C	0.00	1	0.00
ts2	$(GAGACC)_{1-2}(TCGACC)_{1-2}$	C	0.52	3	8.52
umc128[c]	$(AT)_{3-4}$	NC	0.52	2	1.04
umc128[c,d]	$(CA)_{3-4}$	NC	0.51	2	1.03
umc128[c]	$(CA)_{4-5}$	NC	0.24	2	0.47
umc128	$(AACG)_{3-4}C(TAAT)_{1-2}$	NC	0.41	2	3.54
umc128	$(T)_{1-9}(CT)_{5-8}$	NC	0.72	5	3.60
umc230[c]	$(TCCCAA)_{1-2}$	NC	0.17	2	3.12
umc230[c]	$(AAT)_{2-3}$	NC	0.52	2	2.34
umc67	$(GACG)_2C(TACG)_{1-3}$	NC	0.17	2	10.04
umc67[c]	$(CGG)_3$	NC	0.00	1	0.00
umc67[c]	$(AAAC)_3$	NC	0.00	1	0.00
umc67	$(GTT)_{2-3}(T)_{5-7}$	NC	0.38	3	1.08

[a] The repeat motif is in parentheses and the subscript shows the range of variation of the number of repeats observed among individuals.

[b] C, coding region; NC, noncoding region.

[c] Perfect microsatellites.

[d] These loci contained alleles with partial repeats.

10 indels >100 bp that were scored in a common sample of 22 individuals; D_{10-22} was −1.413, which was not significant under the conservative assumption of no recombination, but was substantially lower than D values calculated on short indels. However, the large indels are physically distant from one another, and it is therefore reasonable to apply Tajima's D test assuming free recombination. With free recombination, $D_{10-22} = −1.413$ rep-

TABLE 2

Number of nonmicrosatellite indels found in 17 of the 21 loci

Locus	No. of indels	Indels > 100 bp
asg11	9	1
an1	6	0
tb1	3	0
asg26	14	1
fus6	6	0
csu1150	2	0
umc13	17	1
asg62	4	0
csu1132	7	0
asg75	7	1
adh1	10	1
csu381	13	1
umc128	15	1
ts2	22	0
umc230	34	2
glb1	35	0
umc67	59	1

resents a highly significant departure ($P = 0.008$) from the neutral equilibrium model. Thus, the significant and comparatively low D value based on large indels (>100 bp) is consistent with the hypothesis that the large indels in this sample are selectively deleterious.

Map of the density of RNs per micrometer: Figure 2 plots the frequency of occurrence of RNs per micrometer relative to physical position along SC1. Overlaid on the RN frequency distribution is a smoothed line for the rate of recombination (R = number of RNs per micrometer) derived using the Lowess procedure and a sliding window size of 11 data points. Initially four different sliding window sizes were used. We localized the 21 loci on the RN map and obtained four different estimates (corresponding to the four sliding window sizes) of recombination rate, R, for each locus. Estimates of R based on these four different window sizes were highly and significantly correlated; among the six pairwise comparisons, the lowest correlation was $r = 0.88$ ($P < 0.0001$), which corresponded to the correlation between the most extreme window sizes (*i.e.*, 5 *vs.* 11 data points). Because estimates of R were similar among window sizes, we report results on the basis of a single sliding window size, which we have chosen to be 11 data points.

The values of R, measured in RNs per micrometer, ranged from 0.0099 for *umc67* to 0.1297 for *fus 6* (Table 3). This range is comparable to the range of $\hat{C}_{Hud87}$, which varied from 0.0001 to 0.1337 per base pair (Table 3), and it is also similar to R values reported for tomato (STEPHAN and LANGLEY 1998). The pattern of R along SC1 was characterized by a marked reduction of recombination rate near the centromere and an increase toward telomeres (Figure 2). This pattern, low recombination in the centromere with higher recombination toward telomeres, was confirmed for the 21 genetic loci, because the distance of the loci from the centromere was correlated with their R values ($r = 0.88$; $P < 0.0001$).

Comparing estimates of *R*, *C*, and genetic diversity: Previous studies have demonstrated a positive correlation between recombination rate and genetic diversity, and one purpose of this study was to characterize this correlation in maize. We tested correlations among four
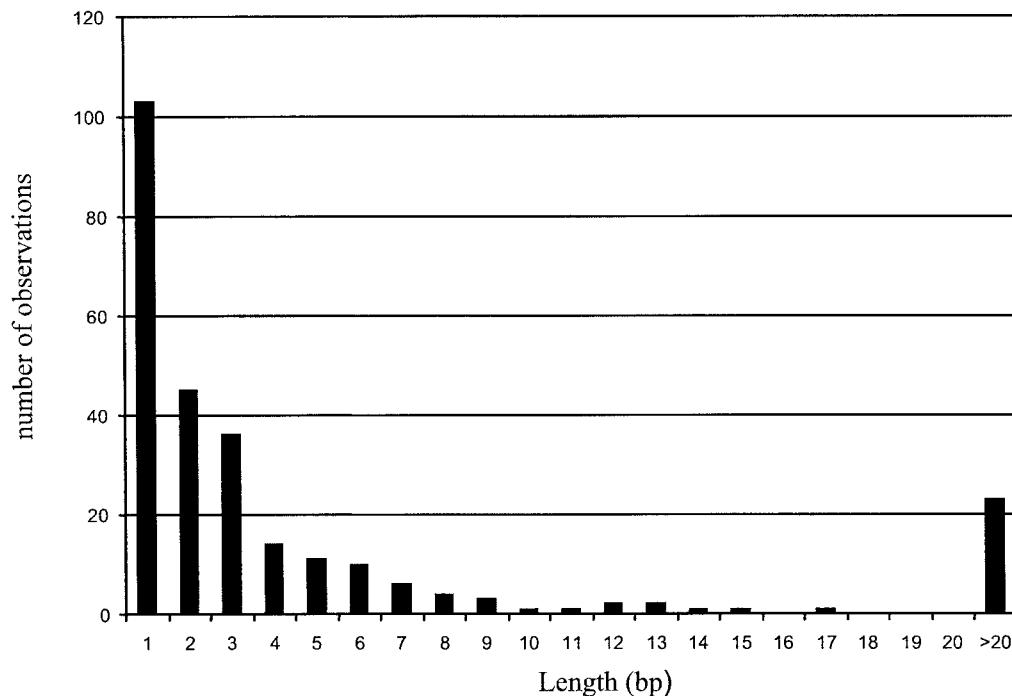

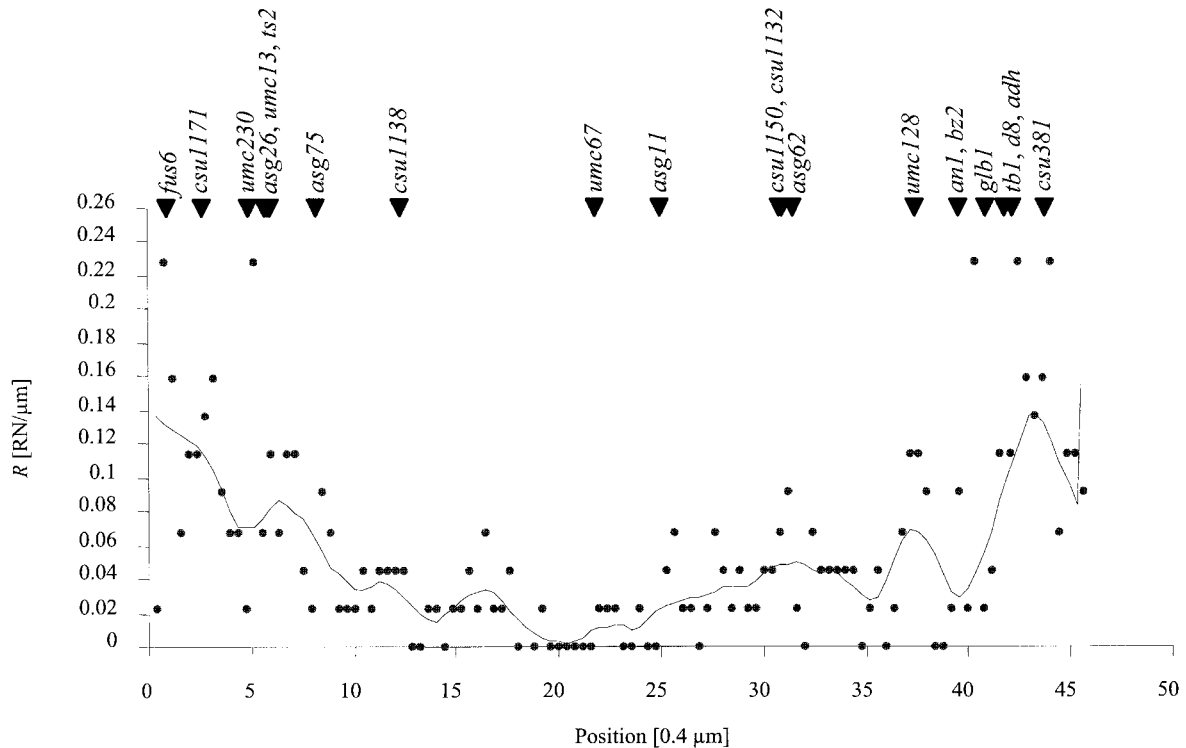FIGURE 1.—Length distribution of indels. For ease of illustration, indels >20 bp in length are grouped.

FIGURE 2.—A map of the distribution of recombination nodules R (RN/μm) along SC1. The short arm of SC1 is to the left, and the centromere is located approximately at position 20. The data points are the frequency of occurrence of RN (no. of RN/ no. of SC observed) per micrometer in each 0.4-μm segment along the SC in the abscissa. The line is the result of the Lowess smoothing procedure with a sliding window containing 11 data points. After alignment to the genetic map (UMC98) of chromosome 1, we determined the positions (indicated by the solid arrows) of the 21 loci along SC1 in the abscissa. The corresponding R values in the ordinate were determined for each of the 21 loci.

different measures of diversity ($\hat{\theta}$, $H_{microsat}$, V, and H_{indel}), three estimates of the population-recombination parameter ($\hat{C}_{Hud87}$, $\hat{C}_{Wall00}$, and $\hat{C}_{Hud01}$), and a physical estimate of recombination (R).

A significant positive correlation between $\hat{C}_{Hud87}$ and $\hat{\theta}$ was described previously (TENAILLON *et al.* 2001). This correlation was based on 18 of the 21 genetic loci because 3 loci (*tb1*, *ts2*, and *d8*) exhibited evidence of deviation from neutral evolution (TENAILLON *et al.* 2001), probably due to artificial selection on these loci during domestication. Independent evidence verifies artificial selection at *tb1* and *d8* (WANG *et al.* 1999; THORNSBERRY *et al.* 2001), and we have additional evidence for artificial selection at *ts2* (M. I. TENAILLON and B. S. GAUT, unpublished data). Because C estimates are valid only for neutral loci (HUDSON 1987; FRISSE *et al.* 2001), correlations with C did not include these 3 loci.

However, HUDSON's (1987) estimator of C can be unreliable, particularly when C values are small per gene (HUDSON 1987; WALL 2000). We therefore utilized two additional estimates of C. We found a significant positive correlation between $\hat{C}_{Wall00}$ and $\hat{\theta}$ ($r = 0.50$; $P < 0.001$) but not between $\hat{C}_{Hud01}$ and $\hat{\theta}$ ($r = 0.007$; $P = 0.46$). The results were similar when $\hat{\theta}$ was based on silent sites, as opposed to all sites (for silent sites: $\hat{\theta}$ *vs.* $\hat{C}_{Hud87}$, $r = 0.67$, $P = 0.007$; $\hat{\theta}$ *vs.* $\hat{C}_{Hud01}$, $r = 0.50$, $P = 0.03$; $\hat{\theta}$ *vs.* $\hat{C}_{Hud01}$,

$r = -0.008$, $P = 0.47$). In general, $\hat{C}_{Hud87}$ and $\hat{C}_{Wall00}$ were highly correlated ($r = 0.68$; $P = 0.005$) but $\hat{C}_{Hud87}$ and $\hat{C}_{Hud01}$ were less correlated ($r = 0.37$; $P = 0.076$). For the remainder of the article we report results with $\hat{C}_{Hud87}$ (Figure 3) but provide results with the other two C estimators when they differ. To sum, at silent sites or all sites, maize nucleotide diversity is correlated with $\hat{C}$ for two of three estimators.

We also compared $\hat{\theta}$ to R, using all 21 genetic loci (Figure 3). There was no significant correlation between $\hat{\theta}$ and R, whether $\hat{\theta}$ was based on all sites (Figure 3) or silent sites ($r = -0.08$; $P = 0.64$). The results were qualitatively similar when the data were limited to the 18 genetic loci for which there was no evidence of artificial selection (data not shown). In addition, R was not strongly correlated with estimates of C (Figure 3), regardless of the C estimator.

Of the eight correlations between recombination and diversity shown in Figure 3, two were both positive and significant at the 5% significance level after multiple-test correction. The first was the correlation between $\hat{\theta}$ and $\hat{C}$, described above. The second positive correlation was between $H_{microsat}$ and R. The correlation remained significant when $H_{microsat}$ was averaged among polymorphic microsatellite loci within a single genetic locus ($r = 0.68$; $P < 0.001$), but the correlation was not as strong

TABLE 3

**Per site estimates of the population-recombination parameter
C and the physical recombination rate R**

Locus	$\hat{\theta}^a$	$\hat{C}_{\text{Hud87}}{}^a$	$\hat{C}_{\text{Wall00}}$	$\hat{C}_{\text{Hud01}}$	R^c
fus6	0.0087	0.0031	0.0000	0.0040	0.1297
csu1171	0.0062	0.0207	0.0002	0.0049	0.1151
umc230	0.0178	0.0713	0.0188	0.0168	0.0707
asg26	0.0092	0.0001	0.0097	0.0041	0.0784
umc13	0.0096	0.0012	0.0000	0.0000	0.0784
ts2	0.0023	0.1337	0.0888	NA^b	0.0784
asg75	0.0071	0.0356	0.0395	0.0565	0.0613
csu1138	0.0131	0.0183	0.0000	0.0021	0.0311
umc67	0.0099	0.0280	0.0339	0.0356	0.0099
asg11	0.0086	0.0114	0.0093	0.0113	0.0228
csu1150	0.0027	0.0053	0.0002	0.0005	0.0477
csu1132	0.0117	0.0001	0.0063	0.0050	0.0481
asg62	0.0259	0.0362	0.0226	0.0090	0.0487
umc128	0.0155	0.0199	0.0513	0.0418	0.0679
an1	0.0048	0.0147	0.0272	0.0525	0.0305
bz2	0.0074	0.0200	0.0000	0.0008	0.0305
glb1	0.0252	0.0897	0.0620	0.0261	0.0615
tb1	0.0051	0.0673	0.0000	0.0148	0.0965
d8	0.0027	0.0210	0.0232	0.0291	0.0965
adh1	0.0117	0.0017	0.0009	0.0039	0.1151
csu381	0.0121	0.0009	0.0041	0.0017	0.1271

^{a} Data are taken from Tenaillon *et al.* 2001; $\hat{\theta}$ is based on all sites.

^{b} NA, not available: We were unable to obtain an estimate for this locus because the approximate-likelihood estimate is apparently beyond the searchable parameter space.

^{c} Estimates are based on the Lowess procedure using a sliding window of 11 data points and measured in RN per micrometer.

when H_{microsat} was averaged over both polymorphic and monomorphic microsatellites within a single genetic locus ($r = 0.28$; $P = 0.14$). Other comparisons of R, $\hat{C}$, and genetic diversity were not significant (Figure 3).

DISCUSSION

In theory, the relative contributions of background and hitchhiking selection can be determined by comparing recombination rates to genetic diversity on the basis of markers that evolve with different rates (Slatkin 1995). However, demography can obscure the relationship between recombination and diversity. Here we have measured genetic diversity for three types of molecular markers in the hope that comparisons among markers would provide insight into the forces shaping maize genetic diversity. Sampling was identical for the three marker types, and hence information among marker types is directly comparable. Recombination was measured both by the population-recombination parameter C and by a quantitative cytogenetic map of chromosome 1.

The quantitative cytogenetic map provides estimates of recombination rate (R) that are related to physical distance along SC1; R is the first quantitative measure of recombination in maize on a chromosomal, rather than a genic, scale. The distribution of R along SC1 indicates that the frequency of exchange per physical unit is reduced in centromeric regions relative to distal chromosomal regions (Figure 2), similar to centromeric suppression observed in other organisms (see Jones 1984 and Resnick 1987 for reviews), including Drosophila (Hudson and Kaplan 1995) and cultivated tomato (Sherman and Stack 1995). The pattern observed in maize chromosome 1 is similar to that reported for large grass genomes such as wheat and barley, in which recombination primarily occurs along the distal half of the chromosomal arm (Gill *et al.* 1996; Kunzel *et al.* 2000). In contrast, the region of centromeric repression is relatively small in rice (Cheng *et al.* 2001).

The distribution of R also suggests substantial heterogeneity in recombination along chromosomal arms (Figure 2). Although the magnitude and scale of recombination needs to be characterized further, heterogeneity in R is consistent with the observation in barley that recombination is mainly confined to a few small areas spaced by large segments in which recombination is severely suppressed (Kunzel *et al.* 2000). Similar recombination hotspots have been previously described in maize (Dooner 1986; Civardi *et al.* 1994; Okagaki and Weil 1997; Fu *et al.* 2002).

The correlation between SNP diversity and recombination estimates: One striking result is that $\hat{C}$ correlates with $\hat{\theta}$ for two of three C estimators. It is unclear why the third estimator, based on Hudson's (2001) method, behaves differently than the first two, but the positive correlation between $\hat{C}$ and $\hat{\theta}$ in two cases indicates that the correlation is not solely an artifact of the estimator ($\hat{C}_{\text{Hud87}}$) used in the previous study (Tenaillon *et al.* 2001). In contrast to $\hat{C}$ and $\hat{\theta}$, we detect no correlation between $\hat{\theta}$ and R or between $\hat{C}$ and R (Figure 3).

Given these results, it is important first to consider differences between C and R. One obvious difference is that the two parameters differ in spatial scale. R is estimated on a chromosomal scale and therefore reflects an "average" recombination rate over large chromosomal regions. In contrast, C is estimated for a particular genetic locus. Maize contains recombination hotspots, particularly in genic regions (Civardi *et al.* 1994; Fu *et al.* 2002), and it is therefore possible that C more accurately incorporates information about recombination on the "local" spatial scale at which θ is measured.

More importantly, R and C measure different quantities. Both R and C describe recombination to some extent, but R measures *only* the recombination rate per physical distance; it is unaffected by population history, selection, and demography. In contrast, C is scaled by population size N, and it is inversely related to LD. Like LD, C is affected by population admixture, population subdivision, fluctuations in population size, and selection, in addition to recombination (reviewed in Pritchard and Przeworski 2001). The lack of correlation
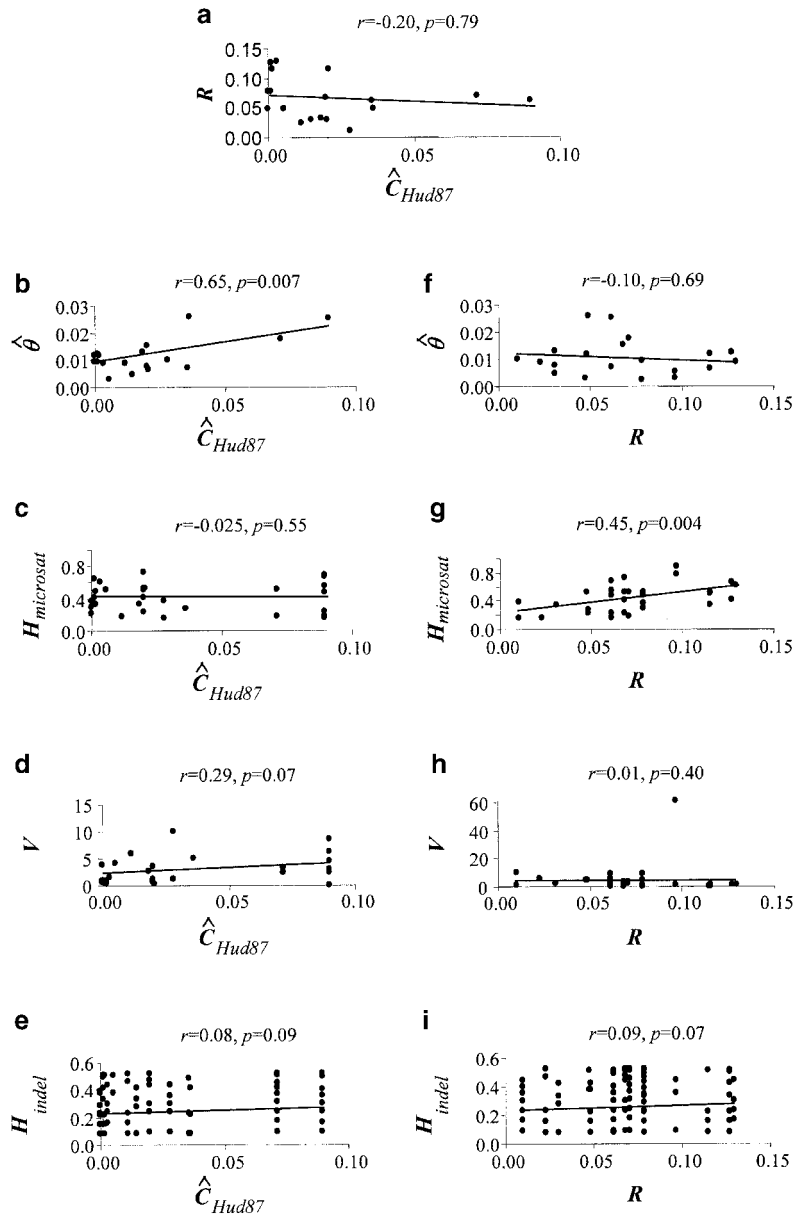
FIGURE 3.—Correlations between two estimates of recombination (R and $\hat{C}_{Hud87}$) and between estimates of recombination and diversity. b–e are based on 18 genetic loci, but f–i contain data from all genetic loci (see RESULTS). Regression lines, Pearson correlation coefficients (r), and P values are given.

between R and $\hat{C}$ may indicate that selection or demographic factors contribute to an uncoupling between LD and recombination.

What could be the evolutionary forces causing a correlation between $\hat{C}$ and $\hat{\theta}$? The correlation could be primarily a function of selection. Under this scenario, hitchhiking or background selection that decreases θ acts in a similar fashion on C. It is clear that C can be affected by selection. For example, balancing selection decreases C (PRITCHARD and PRZEWORSKI 2001); C also decreases briefly after a selective sweep, but some estimators may not detect this effect (PRZEWORSKI 2002). However, if background selection or selective sweeps are prominent enough to cause, in some unknown fashion, a correlation between $\hat{C}$ and $\hat{\theta}$, one expects a correlation between R and θ, as documented in several other systems (BEGUN and AQUADRO 1992; NACHMAN 1997;

NACHMAN et al. 1998). Although we cannot be certain that our estimates of R accurately reflect recombination in the local regions of the 21 genes, we do not observe a correlation between R and $\hat{\theta}$ in maize (Figure 3). Furthermore, under a hitchhiking model Tajima's D should correlate with recombination rate (BRAVERMAN et al. 1995; ANDOLFATTO and PRZEWORSKI 2001), but Tajima's D for the 21 genetic loci is not correlated with either R ($r = -0.06$; $P = 0.60$) or $\hat{C}$ ($r = -0.38$; $P = 0.93$). Altogether, there is no convincing evidence that hitchhiking or background selection contributes to the correlation between $\hat{C}$ and $\hat{\theta}$. It is important to note, however, that these results do not imply that background and hitchhiking selection are *not* acting to shape maize SNP diversity. The signature of background or hitchhiking selection could be overridden by other factors.

Both C and θ $(= 4N\mu)$ contain historical information about population size. Both are also estimated from SNPs that evolve at an estimated rate of $\sim 10^{-9}$ substitutions per site per year (GAUT *et al.* 1996) and therefore encompass relatively long time frames. Some SNPs have been retained in Zea populations for 1 million years or more (GAUT and CLEGG 1993). As a result, the time frame encompassed by both C and θ exceeds maize domestication ~ 7500 (ILTIS 1983) to ~ 9000 years ago (MATSUOKA *et al.* 2002b). Domestication was associated with a bottleneck that decreased SNP diversity in maize $\sim 20\%$ on average relative to its wild ancestor (ZHANG *et al.* 2002). There have also been substantial demographic events since domestication, such as the geographic patterning of extant maize races (*e.g.*, MATSUOKA *et al.* 2002b).

It is not yet clear, however, how demographic events, like a domestication bottleneck or geographic subdivision, affect C and θ jointly. One possibility is that population size N varies among loci because gene flow and other demographic factors vary from locus to locus, both within maize and among its wild relatives (as in the *D. pseudoobscura* complex; WANG *et al.* 1997; MACHADO *et al.* 2002). If demographic effects vary among loci, they could contribute substantially to a correlation between C and θ through N. Variation in N among loci can establish strong positive correlations between C and θ, even in the absence of correlations between θ and c. Simulations with 21 loci suggest that N can vary <10-fold and establish a correlation between C and θ (data not shown). To explore the effect of demography more fully, it will be helpful to have some knowledge of diversity in maize prior to domestication and also of divergence population genetics (KLIMAN *et al.* 2000) in the genus Zea. We are in the process of gathering empirical data from wild relatives and will address the effect of demography on C and θ more thoroughly in future work.

Microsatellite diversity: We identified 47 microsatellite loci in our data, and 33 of these loci were polymorphic. Levels of diversity in these loci, as measured by H_{microsat}, are positively correlated with R. There are at least three possible explanations for this correlation.

The first explanation is based on sampling—*i.e.*, our sample may contain rapidly evolving loci in regions of high recombination by chance alone. This scenario is particularly plausible because the microsatellites in this study likely evolve with different mutation rates. Microsatellite mutation rates (μ) vary considerably by repeat motif (CHAKRABORTY *et al.* 1997; SCHUG *et al.* 1997, 1998), length (SCHLOTTERER *et al.* 1998; SCHUG *et al.* 1998b; UDUPA and BAUM 2001), and base composition (SCHLOTTERER and TAUTZ 1992; GLENN *et al.* 1996; BACHTROG *et al.* 2000). For example, μ is estimated to be $\sim 7.7 \times 10^{-4}$ mutations per generation for dinucleotide repeats in maize but $<5 \times 10^{-5}$ for longer repeat motifs (VIGOUROUX *et al.* 2002).
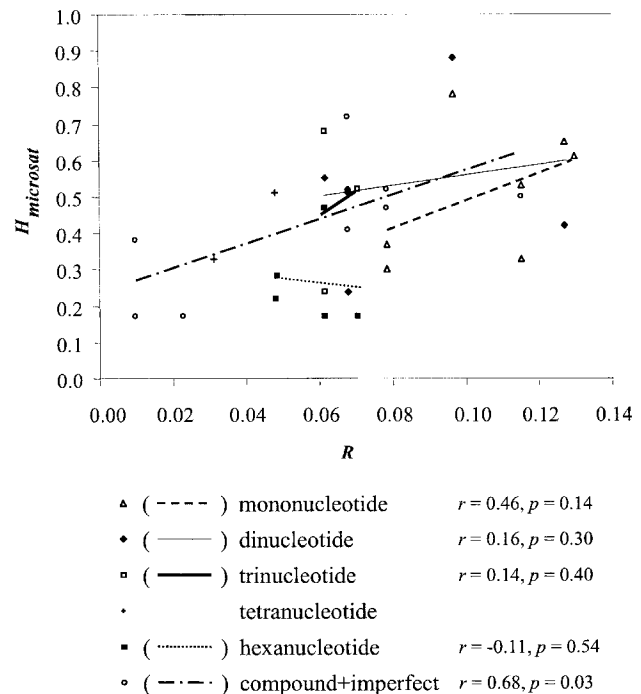
To examine whether any particular class of microsa-

FIGURE 4.—Correlations between microsatellite diversity (H_{microsat}) and R for each microsatellite repeat class. Only two polymorphic loci were available from the tetranucleotide class and none from the pentanucleotide class. Compound and imperfect loci were combined because there is little *a priori* information as to their mutation rates. The regression line, Pearson correlation coefficient (r), and P value are given for each microsatellite class.

tellite is driving the correlation between H_{microsat} and R, we partitioned microsatellite loci into different classes by repeat type, including perfect mono-, di-, tri-, and hexanucleotide repeats, as well as compound + imperfect repeats (Figure 4). The only class exhibiting a positive and significant correlation between H_{microsat} and R was the compound + imperfect class (Figure 4), but this correlation was not significant after multiple test correction. However, four of the five classes exhibited a positive correlation between H_{microsat} and R (Figure 4), suggesting that a positive correlation with R may be a general property of the microsatellite loci in our sample. The mononucleotide repeat class is particularly interesting, both because these loci may evolve rapidly and because they are primarily located in regions with high R (Tables 1 and 3). The mononucleotide class is positively but not significantly correlated with R ($r = 0.46$; $P = 0.14$), but the overall correlation between H_{microsat} and R remains when this class is removed from analysis ($r = 0.42$; $P = 0.02$). Thus, it does not appear that the overall correlation between H_{microsat} and R is driven either by one particular class of microsatellite or by the chance location of rapidly evolving microsatellites (like mononucleotide repeats) in high R regions.

A second possibility for the correlation between H_{microsat} and R is that recombination is itself mutagenic, thereby

causing microsatellite polymorphisms. For example, human data suggest that recombination can lead to the contraction and expansion of trinucleotide repeats (RICHARD and PAQUES 2000). The effect of recombination on microsatellite diversity needs to be investigated further, but mutagenic effects of recombination could underlie the correlation between $H_{microsat}$ and R.

A third possibility is that the correlation is a property of the relationship between recombination and selection. To discuss this possibility, it is first important to note that microsatellite mutation rates have been measured in many organisms, including humans (WEBER and WONG 1993; XU *et al.* 2000), Drosophila (VAZQUEZ *et al.* 2000), chickpea (UDUPA and BAUM 2001), and maize (VIGOUROUX *et al.* 2002). In all of these organisms, microsatellites mutate at a rate $\mu > 10^{-6}$ mutations per generation. Thus, the microsatellites in this study probably mutate at least three orders of magnitude more rapidly than SNPs. The consequence of high mutation rates is profound. Microsatellites are expected to quickly approach an equilibrium between mutation and drift (SLATKIN 1995), and they recover rapidly from demographic and selective events. For example, microsatellites in Drosophila, which mutate relatively slowly at $\mu = 5.1 \times 10^{-6}$ (VAZQUEZ *et al.* 2000) compared to plant microsatellites, are estimated to recover from selective sweeps in <1000 years (NURMINSKY 2001). Although the number of Drosophila generations in 1000 years likely exceeds the number of maize generations since domestication, it is possible that some maize microsatellites may have recovered, at least partially, from the effect of the domestication bottleneck ~7500 (ILTIS 1983) to ~9000 years ago (MATSUOKA *et al.* 2002a). If this is true, it is possible that the signature of ongoing hitchhiking or background selection is no longer dominated by a past demographic event (*i.e.*, a domestication bottleneck) in microsatellite loci as it may be in SNPs.

Finally, we note that comparisons between V and R do not yield a positive correlation (Figure 3). However, when V is based on repeat number, rather than allele size, results with $H_{microsat}$ and V are more comparable. It is desirable to use repeat number, as opposed to allele size, because V based on repeat number is not biased by repeat length. However, V based on repeat number cannot be calculated for several microsatellite loci in our sample because the repeats were imperfect, were compound, or did not evolve in stepwise fashion. For the 21 polymorphic perfect loci that evolve in stepwise fashion (Table 1), V based on repeat number is positively, but not significantly, correlated with R ($r = 0.19$; $P = 0.23$). When *tb1* is dropped from consideration, the correlation is significantly positive ($r = 0.47$; $P = 0.024$), and this result is comparable to that we obtained with $H_{microsat}$. All of our analyses with V—whether based on allele size or repeat number—were heavily influenced by the outlying *tb1* microsatellite locus. Altogether, the reliance on *tb1*, the bias due to repeat length

for V based on allele size and the dependence on stepwise mutations for V based on repeat number, diminish the value of V as a measure of microsatellite diversity for these data.

Indel diversity: Indel diversity in maize is marked by a size distribution that is heavily skewed toward small indels (1–5 bp), with a few large (>100 bp) indels marking the extreme tail of the distribution (Figure 1). Similar distributions have been reported for mammalian and Drosophila nuclear DNA (GU and LI 1995; BERGMAN and KREITMAN 2001), and hence maize is not unique in having a preponderance of small indels. Indel polymorphism was not correlated with $\hat{C}$ or R (Figure 3). Because little is known about indel mutation rates and how μ varies among different indel sizes, it is difficult to interpret the lack of correlation.

It is perhaps more interesting that the population frequency of indels is skewed by size. In our sample, large indels are on average less frequent in the population sample than small indels, suggesting that large indels are slightly deleterious. The 10 large indels also have a lower Tajima's D value than the small indels. Of these 10, only 2 are clearly associated with coding DNA (*adh1* and *csu381*; Table 2); the rest are located in anonymous RFLP marker regions. These results raise an interesting paradox. Greater than 50% of the maize genome consists of retrotransposons (SANMIGUEL *et al.* 1996). Given the preponderance of transposable elements in the maize genome, it seems unlikely that large indels are usually strongly deleterious, yet these population data suggest they are measurably deleterious. The resolution to this problem consists of two components. First, the vast majority of the maize genome consists of retrotransposons that insert into one another (SANMIGUEL *et al.* 1996); presumably the targeting of retrotransposons into nonessential genic regions is evolutionarily favorable for element proliferation. Because of this targeting, retrotransposons may be under different evolutionary dynamics from the indels in our sample, none of which are retrotransposons. Second, MITEs and *Ds* elements, the only identifiable elements in our study, preferentially insert into transcribed regions (BENNETZEN 2000), suggesting some of our genetic loci are near coding regions where insertions are more likely to be deleterious.

The forces affecting genetic diversity in maize: This study offers several insights into the forces contributing to genetic diversity in maize. First, there is no evidence that R and $\hat{\theta}$ are positively correlated, as expected under hitchhiking and background selection models. Assuming R provides reasonable estimates of recombination, it thus appears likely that other effects—perhaps demography—drive the correlation between $\hat{C}$ and $\hat{\theta}$. Second, the correlation between $H_{microsat}$ and $\hat{R}$ suggests either that recombination is mutagenic for microsatellite loci or that a pattern of hitchhiking or background selection is evident in markers that may be partially recovered

from some historical events. However, one must question whether the maize genome has sufficient gene density to permit extensive background selection, because the strength of background selection depends on gene density (Payseur and Nachman 2000). The lower the gene density, the lower the deleterious mutation rate and hence the lower the strength of background selection. In maize, the gene space is restricted to only 20% of the genome (Carels *et al.* 1995; Barakat *et al.* 1997) and consequently the low recombination regions of maize, where the effects of background selection are most pronounced, may contain very few genes.

LITERATURE CITED

Anderson, L. K., and S. M. Stack, 2001 Synaptonemal complex karyotype for maize. Maize Newsl. **75:** 20.

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

Bachtrog, D., M. Agis, M. Imhof and C. Schlotterer, 2000 Microsatellite variability differs between dinucleotide repeat motifs—evidence from *Drosophila melanogaster*. Mol. Biol. Evol. **17:** 1277–1285.

Barakat, A., N. Carels and G. Bernardi, 1997 The distribution of genes in the genomes of Gramineae. Proc. Natl. Acad. Sci. USA **94:** 6857–6861.

Baudry, E., C. Kerdelhue, H. Innan and W. Stephan, 2001 Species and recombination effects on DNA variability in the tomato genus. Genetics **158:** 1725–1735.

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

Bennetzen, J. L., 2000 Transposable element contributions to plant gene and genome evolution. Plant Mol. Biol. **42:** 251–269.

Bergman, C. M., and M. Kreitman, 2001 Analysis of conserved noncoding DNA in Drosophila reveals similar constraints in intergenic and intronic sequences. Genome Res. **11:** 1335–1345.

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

Carels, N., A. Barakat and G. Bernardi, 1995 The gene distribution of the maize genome. Proc. Natl. Acad. Sci. USA **92:** 11057–11060.

Chakraborty, R., M. Kimmel, D. N. Stivers, L. J. Davison and R. Deka, 1997 Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. Proc. Natl. Acad. Sci. USA **94:** 1041–1046.

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

Charlesworth, B., and C. H. Langley, 1989 The population genetics of *Drosophila* transposable elements. Ann. Rev. Genet. **23:** 251–287.

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

Cheng, Z., G. G. Presting, C. R. Buell, R. A. Wing and J. Jiang, 2001 High-resolution pachytene chromosome mapping of bacterial artificial chromosomes anchored by genetic markers reveals the centromere location and distribution of genetic recombination along chromosome 10 of rice. Genetics **157:** 1749–1757.

Civardi, L., Y. Xia, K. J. Edwards, P. S. Schnable and B. J. Nikolau, 1994 The relationship between genetic and physical distances in the clones *a1-sh2* interval of the *Zea mays* L. genome. Proc. Natl. Acad. Sci. USA **91:** 8268–8272.

Cleveland, W. S., 1981 LOWESS: a program for smoothing scatterplots by robust locally weighted regression. Am. Stat. **35:** 54.

Davis, G. L., M. D. McMullen, C. Baysdorfer, T. Musket, D. Grant *et al.*, 1999 A maize map standard with sequenced core markers, grass genome reference points and 932 expressed sequence tagged sites (ESTs) in a 1736-locus map. Genetics **152:** 1137–1172.

Dooner, H. K., 1986 Genetic fine structure of the *bronze* locus in maize. Genetics **113:** 1021–1036.

Dvorak, J., M.-C. Luo and Z.-L. Yang, 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.

Fearnhead, P., and P. Donnelly, 2001 Estimating recombination rates from population genetic data. Genetics **159:** 1299–1318.

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

Fu, H. H., Z. W. Zheng and H. K. Dooner, 2002 Recombination rates between adjacent genic and retrotransposon regions in maize vary by two orders of magnitude. Proc. Natl. Acad. Sci. USA **99:** 1082–1087.

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

Gaut, B. S., B. R. Morton, B. M. McCaig and M. T. Clegg, 1996 Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. Proc. Natl. Acad. Sci. USA **93:** 10274–10279.

Gill, K. S., B. S. Gill and T. R. Endo, 1996 Identification and high-density mapping of gene-rich regions in chromosome group 1 of wheat. Genetics **144:** 1883–1891.

Glenn, T. C., W. Stephan, H. C. Dessauer and M. J. Braun, 1996 Allelic diversity in alligator microsatellite loci is negatively correlated with GC content of flanking sequences and evolutionary conservation of PCR amplifiability. Mol. Biol. Evol. **13:** 1151–1154.

Gu, X., and W. H. Li, 1995 The size distribution of insertions and deletions in human and rodent pseudogenes suggests the logarithmic gap penalty for sequence alignment. J. Mol. Evol. **40:** 464–473.

Herickhoff, L., S. Stack and J. Sherman, 1993 The relationship between synapsis, recombination nodules and chiasmata in tomato translocation heterozygotes. Heredity **71:** 373–385.

Hey, J., and J. Wakeley, 1997 A coalescent estimator of the population recombination rate. Genetics **145:** 833–846.

Hudson, R. R., 1987 Estimating the recombination parameter of a finite population model without selection. Genet. Res. **50:** 245–250.

Hudson, R. R., 2001 Two-locus sampling distributions and their application. Genetics **159:** 1805–1817.

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

Iltis, H. H., 1983 From teosinte to maize: the catastrophic sexual transmutation. Science **222:** 886–894.

Innan, H., R. Terauchi and N. T. Miyashita, 1997 Microsatellite polymorphism in natural populations of the wild plant *Arabidopsis thaliana*. Genetics **146:** 1441–1452.

Jones, G. H., 1984 The control of chiasma distribution. Symp. Soc. Exp. Biol. **38:** 293–320.

Kaplan, N. L., R. R. Hudson and C. H. Langley, 1989 The "hitchhiking effect" revisited. Genetics **123:** 887–899.

Kliman, R. M., P. Andolfatto, J. A. Coyne, F. Depaulis, M. Kreitman *et al.*, 2000 The population genetics of the origin and divergence of the *Drosophila simulans* complex species. Genetics **156:** 1913–1931.

Kraft, T., T. Sall, I. MagnussonRading, N. O. Nilsson and C. Hallden, 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.

Kunzel, G., L. Korzun and A. Meister, 2000 Cytologically integrated physical restriction fragment length polymorphism maps for the barley genome based on translocation breakpoints. Genetics **154:** 397–412.

Machado, C. A., R. M. Kliman, J. A. Markert and J. Hey, 2002 In-

ferring the history of speciate from multilocus DNA sequence data: the case of *Drosophila pseudoobscura* and close relatives. Mol. Biol. Evol. **19:** 472–488.

MATSUOKA, Y., S. E. MITCHELL, S. KRESOVICH, M. GOODMAN and J. DOEBLEY, 2002a Microsatellites in *Zea*—variability, patterns of mutations, and use for evolutionary studies. Theor. Appl. Genet. **104:** 436–450.

MATSUOKA, Y., Y. VIGOUROUX, M. M. GOODMAN, J. SANCHEZ, E. BUCKLER *et al.*, 2002b A single domestication for maize shown by multilocus microsatellite genotyping. Proc. Natl. Acad. Sci. USA **99:** 6080–6084.

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

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

NACHMAN, M. W., V. L. BAUER, S. L. CROWELL and C. F. AQUADRO, 1998 DNA variability and recombination rates at X-linked loci in humans. Genetics **150:** 1133–1141.

NEI, M., 1973 Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. USA **70:** 3321–3323.

NURMINSKY, D. I., 2001 Genes in sweeping competition. Cell. Mol. Life Sci. **58:** 125–134.

OKAGAKI, R. J., and C. F. WEIL, 1997 Analysis of recombination sites within the maize *waxy* locus. Genetics **147:** 815–821.

PAYSEUR, B. A., and M. W. NACHMAN, 2000 Microsatellite variation and recombination rate in the human genome. Genetics **156:** 1285–1298.

PETERSON, D. G., N. L. LAPITAN and S. M. STACK, 1999 Localization of single- and low-copy sequences on tomato synaptonemal complex spreads using fluorescence *in situ* hybridization (FISH). Genetics **152:** 427–439.

PRITCHARD, J. K., and M. PRZEWORSKI, 2001 Linkage disequilibrium in humans: models and data. Am. J. Hum. Genet. **69:** 1–14.

PROVAN, J., P. LAWRENCE, G. YOUNG, F. WRIGHT, R. BIRD *et al.*, 1999 Analysis of the genus *Zea* (Poaceae) using polymorphic chloroplast simple sequence repeats. Plant Syst. Evol. **218:** 245–256.

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

PRZEWORSKI, M., R. R. HUDSON and A. DI RIENZO, 2000 Adjusting the focus on human variation. Trends Genet. **16:** 296–302.

RAYMOND, M., and F. ROUSSET, 1995 Genepop (Version 1.2)—population genetics software for exact tests and ecumenicism. J. Hered. **86:** 248–249.

REEVES, A., 2001 Micromeasure: a new computer program for the collection and analysis of cytogenetic data. Genome **44:** 439–443.

REMINGTON, D. L., J. M. THORNESBERRY, Y. MATSUOKA, L. M. WILSON, S. R. WHITT *et al.*, 2001 Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc. Natl. Acad. Sci. USA **98:** 11479–11484.

RESNICK, M. A., 1987 Investigating the genetic control of biochemical events in meiotic recombination, pp. 157–210 in *Meiosis*, edited by P. B. MOENS. Academic Press, Orlando, FL.

RICHARD, G. F., and F. PAQUES, 2000 Mini- and microsatellite expansions: the recombination connection. EMBO Rep. **1:** 122–126.

ROZAS, J., and R. ROZAS, 1999 DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15:** 174–175.

SANMIGUEL, P., A. TICKHONOV, Y.-K. JIN, A. MELAKE-BERHAN, P. S. SPRINGER *et al.*, 1996 Nested retrotransposons in the intergenic regions of the maize genome. Science **274:** 765–768.

SCHLOTTERER, C., and D. TAUTZ, 1992 Slippage synthesis of simple sequence DNA. Nucleic Acids Res. **20:** 211–215.

SCHLOTTERER, C., R. RITTER, B. HARR and G. BREM, 1998 High mutation rate of a long microsatellite allele in Drosophila melanogaster provides evidence for allele-specific mutation rates. Mol. Biol. Evol. **15:** 1269–1274.

SCHNEIDER, S., D. ROESSLI and L. EXCOFFIER, 2000 *Arlequin: A Software for Population Genetic Data Analysis.* Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Geneva.

SCHUG, M. D., T. F. C. MACKAY and C. F. AQUADRO, 1997 Low mutation rates of microsatellite loci in Drosophila melanogaster. Nat. Genet. **15:** 99–102.

SCHUG, M. D., C. M. HUTTER, K. A. WETTERSTRAND, M. S. GAUDETTE, T. F. C. MACKAY *et al.*, 1998 The mutation rates of di-, tri- and tetranucleotide repeats in Drosophila melanogaster. Mol. Biol. Evol. **15:** 1751–1760.

SENIOR, M. L., J. P. MURPHY, M. M. GOODMAN and C. W. STUBER, 1998 Utility of SSRs for determining genetic similarities and relationships in maize using an agarose gel system. Crop Sci. **38:** 1088–1098.

SHERMAN, J. D., and S. M. STACK, 1995 Two-dimensional spreads of synaptonemal complexes from Solanaceous plants: high-resolution recombination nodule map for tomato (*Lycopersicon esculentum*). Genetics **141:** 683–708.

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

SMITH, J. S. C., E. C. L. CHIN, H. SHU, O. S. SMITH, S. J. WALL *et al.*, 1997 An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L): comparisons with data from RFLPs and pedigree. Theor. Appl. Genet. **95:** 163–173.

STEPHAN, W., and C. H. LANGLEY, 1998 DNA polymorphism in Lycopersicon and crossing-over per physical length. Genetics **150:** 1585–1593.

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

TENAILLON, M. I., M. C. SAWKINS, A. D. LONG, R. L. GAUT, J. F. DOEBLEY *et al.*, 2001 Patterns of DNA sequence polymorphism along chromosome 1 of maize (Zea mays ssp mays L.). Proc. Natl. Acad. Sci. USA **98:** 9161–9166.

THORNSBERRY, J. M., M. M. GOODMAN, J. DOEBLEY, S. KRESOVICH, D. NIELSEN *et al.*, 2001 *Dwarf8* polymorphisms associate with variation in flowering time. Nat. Genet. **28:** 286–289.

UDUPA, S. M., and M. BAUM, 2001 High mutation rate and mutational bias at (TAA)(n) microsatellite loci in chickpea (Cicer arietinum L.). Mol. Genet. Genomics **265:** 1097–1103.

VAZQUEZ, J. F., T. PEREZ, J. ALBORNOZ and A. DOMINGUEZ, 2000 Estimation of microsatellite mutation rates in *Drosophila melanogaster*. Genet. Res. **76:** 323–326.

VIGOUROUX, Y., J. S. JAQUETH, Y. MATSUOKA, O. S. SMITH, W. D. BEAVIS *et al.*, 2002 Rate and pattern of mutation at microsatellite loci in maize. Mol. Biol. Evol. **19:** 1251–1260.

WALL, J. D., 2000 A comparison of estimators of the population recombination rate. Mol. Biol. Evol. **17:** 156–163.

WANG, R. L., J. WAKELEY and J. HEY, 1997 Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relatives. Genetics **147:** 1091–1106.

WANG, R. L., A. STEC, J. HEY, L. LUKENS and J. DOEBLEY, 1999 The limits of selection during maize domestication. Nature **398:** 236–239.

WATTERSON, G. A., 1975 On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **7:** 188–193.

WEBER, J. L., and C. WONG, 1993 Mutation of human short tandem repeats. Hum. Mol. Genet. **2:** 1123–1128.

WIEHE, T., 1998 The effect of selective sweeps on the variance of the allele distribution of a linked multiallele locus: hitchhiking of microsatellites. Theor. Popul. Biol. **53:** 272–283.

XU, X., M. PENG, Z. FANG and X. P. XU, 2000 The direction of microsatellite mutations is dependent upon allele length. Nat. Genet. **24:** 396–399.

ZHANG, L., A. S. PEEK, D. DUNAMS and B. S. GAUT, 2002 Population genetics of duplicated disease-defense genes, *hm1* and *hm2*, in maize (*Zea mays* ssp. *mays* L.) and its wild ancestor (*Zea mays* ssp. *parviglumis*). Genetics **162:** 851–860.

ZICKLER, D., and N. KLECKNER, 1999 Meiotic chromosomes: integrating structure and function. Annu. Rev. Genet. **33:** 603–754.

Communicating editor: D. CHARLESWORTH