

# A Weak Effect of Background Selection on Trinucleotide Microsatellites in Maize

ANNE-CÉLINE THUILLET, MAUD I. TENAILLON, LORINDA K. ANDERSON, SHARON E. MITCHELL, STEPHEN KRESOVICH, STEPHEN M. STACK, BRANDON GAUT, AND JOHN DOEBLEY

From the Department of Genetics, University of Wisconsin, Madison, WI 53706 (Thuillet and Doebley); the Station de Génétique Végétale UMR C8120, Ferme du Moulon, 91190 Gif-Sur-Yvette, France (Tenaillon); the Department of Biology, Colorado State University, Fort Collins, CO 80523–1878 (Anderson and Stack); the Institute for Genomic Diversity, Cornell University, Ithaca, NY 14853 (Mitchell and Kresovich); and the Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697 (Gaut).

Address correspondence to J. Doebley at the address above, or e-mail: jdoebley@wisc.edu.

---

## Abstract

Artificial selection during the domestication of maize is thought to have been predominantly positive and to have had little effect on the surrounding neutral diversity because linkage disequilibrium breaks down rapidly when physical distance increases. However, the degree to which indirect selection has shaped neutral diversity in the maize genome during domestication remains unclear. In this study, we investigate the relationship between local recombination rate and neutral polymorphism in maize and in teosinte using both sequence and microsatellite data. To quantify diversity, we estimate 3 parameters expected to differentially reflect the effects of indirect selection and mutation. We find no general correlation between diversity and recombination, indicating that indirect selection has had no genome-wide impact on maize diversity. However, we detect a weak correlation between heterozygosity and recombination for trinucleotide microsatellites deviating from the stepwise mutation model and located within genes ( $\rho = 0.32$ ,  $P < 0.03$ ). This result can be explained by a background selection hypothesis. The fact that the same correlation is not confirmed for nucleotide diversity suggests that the strength of purifying selection at or near this class of microsatellites is higher than for nucleotide mutations.

---

## Introduction

In many species, DNA polymorphism appears to be reduced in regions of low recombination rate (Begun and Aquadro 1992; Dvorak et al. 1998; Kraft et al. 1998; Nachman et al. 1998; Stephan and Langley 1998; Roselium et al. 2005). Indirect selection may explain this observation, either through genetic hitchhiking if advantageous mutations reach fixation or through background selection if deleterious alleles are swept out of the population (Maynard-Smith and Haigh 1974; Kaplan et al. 1989; Charlesworth et al. 1993; Hudson and Kaplan 1995). Furthermore, a neutral explanation may also hold if recombination is mutagenic or is indirectly related to mutation (Lercher and Hurst 2002; Hellmann et al. 2003; Hellmann et al. 2005; Bussel et al. 2006). Therefore, it can be difficult to distinguish between hitchhiking, background selection, and mutation.

One way to distinguish first between hitchhiking and background selection is to contrast nucleotide diversity with microsatellite diversity (Wiehe 1998). The high mutation rates at microsatellite loci allow them to recover diversity

quickly enough to erase the signature of past hitchhiking events, but not quickly enough to erase that of ongoing background selection. Hence, for such highly mutable loci, a positive correlation between diversity and recombination is expected to be maintained only under background selection. More recently, Innan and Stephan (2003) developed another approach to distinguish between hitchhiking and background selection. It consists in looking at the relationship between diversity and recombination for low recombination rates, as the shape of the curve differs under the 2 models. This approach, to our knowledge, has not been widely applied so far.

In addition, the mutational properties of microsatellites have been extensively studied and can be taken into account when investigating the main factors that shape diversity. The mutation rate of microsatellites is related to their repeated motif (di-, tri-, or tetranucleotides; Chakraborty et al. 1997; Schug et al. 1998), to their sequence structure (perfect or interrupted; Brinkmann et al. 1998), and to allele length in terms of number of repeats (Schug et al. 1998; Vigouroux et al. 2002; Brodehe et al. 2004; Thuillet et al. 2005). Microsatellite markers also vary in their mode of evolution,

which includes the degree to which they conform to the stepwise mutation model (addition or loss of one repeat per mutation) and the frequency of mutational events involving more than one repeat (Di Rienzo et al. 1994).

Maize (*Zea mays* ssp. *mays*) was domesticated from its wild progenitor *Zea mays* ssp. *parviglumis* (teosinte) between 6 250 and 10 000 bp. (Piperno and Flannery 2001; Smith 2001). Domestication often is seen as relying on the positive selection of alleles beneficial to agriculture (Buckler IV et al. 2001). In contrast, indirect selection does not seem to have affected large chromosomal regions in maize, as linkage disequilibrium in maize has been shown to decrease rapidly when physical distance becomes larger (Tenaillon et al. 2001; Flint-Garcia et al. 2003; Clark et al. 2004). Artificial selection during maize domestication is then expected to have been more positive than purifying and to have had little effect on neutral diversity, if any.

Nevertheless, from genome-wide studies conducted to date in maize, it remains difficult to determine to what extent indirect selection and/or mutation during the domestication process affected neutral diversity. A recent study showed that, although microsatellite diversity throughout the genome was predominately shaped by the domestication bottleneck, a weak pattern of indirect selection could be detected at several loci (Vigouroux et al. 2005). On the other hand, a study on the first chromosome reported no correlation between nucleotide diversity and recombination rate but pinpointed a positive correlation between microsatellite diversity and recombination rate that could be explained by a mutational effect of recombination (Tenaillon et al. 2002). However, because only one chromosome was examined, the conclusions of this study cannot be extrapolated to the entire maize genome.

In this paper, we aim at determining whether indirect effects of selection (hitchhiking and background selection) and/or mutation have shaped neutral diversity in maize and in teosinte. We look at the relationship between diversity and recombination in these 2 species for nucleotide polymorphism and microsatellites covering the entire genome. To interpret this relationship, we first check if the loci for which diversity is available represent a random sampling of recombination rates in the maize genome. Second, we use 3 diversity parameters for which the predicted relationship with the local recombination rate differ under hitchhiking, background selection, and/or mutation.

## Materials and Methods

### Plant Material and DNA Markers

We used diversity data from Vigouroux et al. (2005) for microsatellite loci and from Wright et al. (2005) for nucleotide polymorphism. Vigouroux et al. (2005) genotyped 45 landraces of maize representing the pre-Colombian range of maize and 21 teosinte plants corresponding to its presumed ancestor *Z. mays* ssp. *parviglumis*. Diversity in these samples was assayed at 462 microsatellite loci, covering the maize genetic map that comprised 163 dinucleotide, 206

trinucleotide, 59 tetranucleotide, and 34 penta-, hexa-, or heptanucleotide microsatellites or microsatellites with unknown motif type. Wright et al. (2005) sequenced amplicons from 774 genes in 14 maize inbred lines (representing modern maize) and 16 partially inbred teosinte lines. Genetic positions on the IBM2 neighbors genetic map (Intermated B73/Mo17, available on the web site <http://www.maizegdb.org>) are known for 433 of the microsatellite loci and 640 of the sequenced genes. Details about all loci with known recombination rates used in this study are available online as supplementary material (Table S1).

### Recombination Rate Estimates

We estimated local recombination rate along physical chromosomes for all loci from the direct observation of recombination nodules distribution (Anderson et al. 2003). Recombination nodules predict the occurrence of crossing over and were observed by electron microscopy on synaptonemal complexes in extended pachytene chromosomes of the maize inbred line Kansas Yellow Saline. After the method described by Tenaillon et al. (2002), we determined the average number of recombination nodules observed per 0.2  $\mu\text{m}$  chromosomal segment, and we used a Lowess procedure to smooth extreme local variations that could be artifactual (Cleveland 1981; Stephan and Langley 1998; Anderson et al. 2003). This procedure applies weighted least-squares regression in sliding windows along the chromosome. Consistently with the study by Tenaillon et al. (2002), we used a sliding window of 11 segments of 0.2  $\mu\text{m}$  with their corresponding recombination nodule frequencies, as other window sizes do not affect the results. We could thus assign one recombination rate estimate per 0.2  $\mu\text{m}$ .

In order to predict local recombination rates at our loci, we aligned the genetic and physical maps for each chromosome arm in a linear fashion from the centromere as follows. We first converted the recombination nodules frequency in each 0.2- $\mu\text{m}$  segment into centiMorgans (cMs) by multiplying the recombination nodules frequency by 50, each recombination nodule corresponding to one cross over, which equals 50 cM (Sherman and Stack 1995). We then placed our loci on the physical map by aligning the IBM2 neighbors genetic map with the map of recombination nodules converted to cM and for which each location corresponds to a 0.2- $\mu\text{m}$  segment with a specific recombination nodules frequency. We estimated the recombination rate of each locus as the predicted frequency of occurrence of recombination nodules per micrometer within the 0.2- $\mu\text{m}$  window where it is located. Recombination rates are thus obtained in units of recombination nodules per micrometers. We converted values into recombination rate per site by dividing values by  $7.63 \times 10^6$ , which corresponds to an estimation of the number of bases per micrometers in the maize genome (i.e., the total number of bases in the maize genome [2500 Mbp; Arumaganthan and Earle 1991] divided by the total length of the map in micrometers).

## Recombination Rate Representation in Our Sample of Loci

We tested whether our sample of microsatellites correctly represents the entire range of recombination rates by comparing the entire distribution of recombination rate estimates with our sampled distribution of recombination rates using a Kolmogorov–Smirnov test (Sokal and Rohlf 2000).

To test whether the lowest observed recombination rates ( $<6.5 \times 10^{-3}$ ) were equally rare among different types of microsatellite loci, we used a chi-square test among 6 groups of loci: dinucleotide, trinucleotide, and tetranucleotide loci, either evolving under a stepwise mutation model (when mutations involve single repeat changes) or suspected to evolve in a nonstepwise fashion (with their mutational process potentially involving multistep events or with indels in the flanking sequence or interrupting the repeat region). We computed the expected number of loci in a group as the frequency of low recombination rates among all loci times the number of loci in the group, and we compared it with the actual number of loci with a low recombination rate in the same group.

We evaluated the mode of evolution for each locus in each subspecies by calculating its stepwise index. The stepwise index is the maximum proportion of alleles differing by a size equal to a multiple of the repeat size: for a dinucleotide, most alleles will differ from each other by a multiple of 2 bases and some by other multiples. The stepwise index is then the number of alleles that differ from most others by a multiple of 2 bases divided by the total number of alleles. We calculated it with the software Powermarker v3.0 (Liu and Muse 2005). We considered loci as evolving under the stepwise mutation model when the stepwise index was higher than 0.95 in both maize and teosinte and referred to them as “stepwise loci”; we classified loci exhibiting stepwise index values under 0.95 in both groups as evolving in a nonstepwise manner and denoted them “nonstepwise” loci. We did not infer any mutation model for loci that have a stepwise index higher than 0.95 in only one group.

## Diversity Estimates

We calculated the expected heterozygosity for each microsatellite locus in maize and in teosinte as  $H_e = \frac{n}{n-1} [1 - \sum p_i^2]$ , where  $n$  is the sample size and  $p_i$  the allele frequencies (Nei 1987). We estimated the parameter  $\theta = 4N_e\mu$  in teosinte and in maize ( $N_e$  is the effective population size and  $\mu$  the mutation rate) from  $H_e$ , assuming a stepwise mutation model, as  $\hat{\theta}_{H_e} = \frac{1}{2} [(1 - H_e)^{-2} - 1]$  for microsatellites (Ohta and Kimura 1973), and we used Watterson’s  $\theta$  estimator ( $\hat{\theta}_w$ ) for sequence diversity. For both microsatellites and sequences, we calculated the ratio hereafter noted RH of  $\theta$  estimates in both subspecies as  $\hat{\theta}_{H_e}$  in teosinte /  $\hat{\theta}_{H_e}$  in maize (Kauer et al. 2003). We also calculated the variance of allele size at each microsatellite locus for maize and for teosinte.

## Evaluating the Effect of Mutational Variations on RH

We performed a 2-way analysis of variance (ANOVA) on RH with a covariable to test for an effect on RH of 1)

intralocus variation of the mutation rate, 2) interlocus variation in the mode of evolution, and 3) interlocus variation of the mutation rate. We applied the same model (but without the covariable) to  $H_e$  in maize and in teosinte to assess to which extent  $H_e$  is affected by interlocus variation in the mode of evolution and of the mutation rate.

The covariable used in the 2-way ANOVA of RH was the difference in allele size between maize and teosinte (average allele size in teosinte – average allele size in maize in base pair) divided by the size or the repeated motif. This number calculated at each locus is the average difference in number of repeats between both subspecies and should be correlated with the difference in mutation rate in maize and in teosinte (Schug et al. 1998; Vigouroux et al. 2002; Brodehe et al. 2004; Thuillet et al. 2005).

We took interlocus variation of the mode of evolution into account by contrasting stepwise and nonstepwise loci. Loci after a stepwise mutation model in one subspecies and not in the other were not considered because these loci would exhibit RH values that would undoubtedly be affected by the discrepancies in their mode of evolution between maize and teosinte and that thus could not be interpreted in terms of selective or mutational effect when plotted against the recombination rate.

Interlocus variation of the mutation rate was an expected consequence of using loci with the different repeat lengths: dinucleotide loci have been shown to evolve faster than trinucleotide loci, which in turn evolve faster than tetranucleotide loci (Chakraborty et al. 1997; Schug et al. 1998).

In addition, we performed a 1-way ANOVA on the difference in allele size between maize and teosinte to test for intralocus variation of the mutation rate among di-, tri-, and tetranucleotide microsatellites. Finally, we checked whether the allele size difference between maize and teosinte was correlated with the recombination rate.

## Relationship between Diversity and the Local Recombination Rate

We calculated Pearson’s correlation coefficient between the various diversity indexes ( $\hat{\theta}_w$  and  $H_e$  in maize and in teosinte, RH, and variance of allele size) and the local recombination rate. We performed separate analyses for genes, microsatellite stepwise loci, microsatellite nonstepwise loci, and for each chromosome. We further broke down the microsatellites into their various types (di-, tri-, and tetranucleotide loci) evolving under a stepwise model or not (6 correlations tested for each diversity index). To further investigate if recombination increases the proportion of multistep mutational events, we estimated the correlation between the recombination rate and the stepwise index. We performed analyses using the software SYSTAT version 10 (SPSS, Chicago, IL), and we applied Bonferroni correction to account for multiple tests.

## Testing Correlations for Hitchhiking or Background Selection Hypotheses

We applied the test developed by Innan and Stephan (2003) to observed correlations in order to assess whether the data

fit better to a hitchhiking or to a background selection hypothesis. The test determines whether the observed coefficient of correlation between  $\theta$  and  $\theta/rr$  (where  $rr$  is the recombination rate) over a range of rather low recombination rates is more compatible with a hitchhiking model (close to  $-1$ ) or a background selection model (close to  $1$ ). To do this, it considers expected distributions of the coefficient of correlation under hitchhiking and under background selection. Expected distributions are obtained from simulated data by sampling values for  $\theta$  from a normal distribution of mean  $\theta_{HH}$  and standard deviation  $k\theta_{HH}$  under hitchhiking and of mean  $\theta_{BS}$  and standard deviation  $k\theta_{BS}$  under background selection.  $\theta_{HH}$  and  $\theta_{BS}$  are estimated as  $f\theta_{Neu}$ , where  $\theta_{Neu}$  is the neutral expectation of  $\theta$  and  $f$  is defined as  $f=rr/rr+a$  under hitchhiking and  $f=\exp(-u/rr)$  under background selection. The parameters  $k$ ,  $a$ , and  $u$  are estimated from least squares nonlinear regression of the data, plotted as the fraction of maintained diversity under selection ( $f = \theta_{observed}/\theta_{Neu}$ ) versus the recombination rate. In accordance with Innan and Stephan (2003), we chose to examine recombination rates ranging from  $2 \times 10^{-9}$  to  $5 \times 10^{-9}$ ; this included the 17 loci with the lowest recombination rates. We set  $\theta_{Neu}$  to 0.005 in maize and 0.01 in teosinte for nucleotide data and to 6.83 in maize and 8.40 in teosinte for microsatellites. These values are the observed averages for  $\theta_w$  in the entire data set for nucleotide diversity and for  $\hat{\theta}_{H_e}$  for trinucleotide stepwise mutation model loci in our data set, corrected by the expected mutational variance  $\sigma^2$  ( $\hat{\sigma}^2 = \hat{\theta}_{H_e}$  dinucleotide nonstepwise/ $\hat{\theta}_{H_e}$  dinucleotide stepwise). We repeated the tests for values of  $\theta_{Neu}$  of 5 and 9 and of 7 and 11, respectively, in maize and in teosinte. We obtained the expected distributions of the correlations between  $\theta$  and  $\theta/rr$  under hitchhiking and background selection using each time 10 000 replicates.

## Results

### Recombination Rate Estimates

Recombination rates per site and per generation for the 433 microsatellite loci from Vigouroux et al. (2005) and the 640 genes from Wright et al. (2005) ranged from  $6.92 \times 10^{-10}$  to  $3.73 \times 10^{-8}$  with an average of  $1.14 \times 10^{-8}$ .

The recombination nodule map by Anderson et al. (2003) provides recombination rate estimates for an average of 166 0.2- $\mu$ m bins per chromosome. Our marker loci sampled 31 bins per chromosome on average. For all chromosomes, the average recombination rate for our sampled loci (genes or microsatellites) was higher than the average recombination rate for all bins of the chromosome. This result was significant in almost all cases, except on chromosome 6 for genes and microsatellites and on chromosome 10 for genes only (Table 1).

The lack of low recombination rates ( $<6.5 \times 10^{-9}$ ) affected all categories of microsatellites homogeneously (stepwise dinucleotide loci, nonstepwise dinucleotide, step-

**Table 1.** Average recombination rates ( $rr$ ) per chromosome in the bins sampled by our gene sequences and microsatellites versus across the entire chromosome (all bins).  $P$  values are the probabilities that the local recombination rates for our genes or for our microsatellites are drawn from the same distribution as the entire set of recombination rates for the corresponding chromosome (Kolmogorov–Smirnov test). All recombination rates are expressed in units of  $\times 10^{-8}$

Chromosome	rr All bins	rr Genes	$P$ value	rr Microsatellites	$P$ value
1	0.73	1.21	0.0001	1.06	0.0001
2	0.73	0.98	0.0001	1.13	0.0010
3	0.76	1.10	0.0010	1.21	0.0010
4	0.79	1.10	0.0001	1.44	0.0001
5	0.72	0.90	0.0050	1.17	0.0170
6	0.76	0.80	NS	0.94	NS
7	0.79	1.39	0.0001	1.27	0.0001
8	0.84	1.45	0.0001	1.44	0.0001
9	0.89	0.94	NS	1.27	0.0390
10	0.73	1.00	0.0130	1.13	0.0030
Weighted average	0.77	1.10	0.0001	1.21	0.0001

NS, nonsignificant.

wise trinucleotide, nonstepwise trinucleotide, stepwise tetranucleotide, and nonstepwise tetranucleotide loci;  $\chi^2 = 0.59$ , degrees of freedom = 5, nonsignificant at  $\alpha = 0.05$ , Table 2).

### Diversity

The estimates of Watterson's  $\theta$  for nucleotide diversity,  $\theta_w$ , and expected heterozygosity for microsatellites,  $H_e$ , were lower in maize than in teosinte ( $\theta_w$  was 0.0065 in maize and 0.0111 in teosinte;  $H_e$  was 0.64 in maize and 0.73 in teosinte). Values in maize correspond to 59% and 88% of teosinte diversity for  $\theta_w$  and  $H_e$ , respectively. Calculated on stepwise loci only,  $H_e$  values reflect that maize retained 80% of the diversity in teosinte.

From heterozygosity, we estimated  $\theta$  for microsatellites in maize and in teosinte (see Materials and Methods), and for both nucleotide diversity and microsatellites, we estimated the ratio RH for each locus calculated as  $\theta$  in teosinte divided by  $\theta$  in maize. On average, RH equaled 2.09 for genes. For microsatellites, RH equaled 4.03 on all loci but 2.53 on stepwise loci only.

Finally, we calculated the variance in allele size for microsatellites. It was comparable in maize (60.51) and in teosinte (60.14) for all microsatellites but displayed a higher value in maize (32.76) than in teosinte (27.58) when it was calculated for stepwise loci only.

### RH and the Mutational Properties of the Microsatellites

As  $\theta = 4N_e\mu$ , the ratio RH at each locus gives an estimate of the reduction of population size from teosinte in maize, provided that  $\mu$  is constant for a given locus in both species (i.e., there is no intralocus variation of mutation rate) and for microsatellites, under a stepwise mutation model hypothesis

**Table 2.** Repartition of low recombination rates ( $<6.5 \times 10^{-9}$ ) among the different categories of microsatellite loci

Category		Number of loci	Number of loci having low recombination rates
Stepwise loci	Dinucleotides	29	9
	Trinucleotides	65	23
	Tetranucleotides	7	2
Nonstepwise loci	Dinucleotides	101	31
	Trinucleotides	99	29
	Tetranucleotides	41	14
	Total	342	108

(which underlies a proper estimation of  $\theta$ ). Under these 2 conditions, RH in microsatellites has the same expectation for all loci regardless of their mutation rate (i.e., interlocus variation of the mutation rate should then not affect RH).

We tested by ANOVA (see Materials and Methods) the extent to which violation of the 2 assumptions underlying RH interpretation affects its estimation. We compared RH values for 1) loci having different expected intralocus variation of their mutation rate and 2) stepwise loci and nonstepwise loci. Intralocus variation of the mutation rate was reflected by allele size differences between maize and teosinte, as allele length is in maize positively correlated to the mutation rate (Vigouroux et al. 2002). In addition, as dinucleotide, trinucleotide, and tetranucleotide types of loci have in theory different mutation rates, they provide the opportunity to also test whether RH is indeed insensitive to interlocus variation of the mutation rate.

RH was primarily affected by intralocus variation of the mutation rate and, to a lesser extent, was sensitive to departures from the stepwise mutation model hypothesis but was not affected by interlocus variation of the mutation rate represented by the type of locus (Table 3). The observed level of intralocus variation of the mutation rate was mainly due to the 25 loci (7%) displaying the largest differences in allele size between maize and teosinte; RH was affected only by the mode of evolution when they were removed from the analysis ( $F = 4.5$ ;  $P < 0.03$ ,  $R^2 = 0.2$ ).

**Table 3.** Comparison (via ANOVA) of mean RH, mean heterozygosity in maize ( $H_e$  maize), and mean heterozygosity in teosinte ( $H_e$  teosinte) for di-, tri-, and tetranucleotide loci (type) evolving under stepwise mutation model or not (step or nonstep)

	Dinucleotide		Trinucleotide		Tetranucleotide		$R^2$	$F$ ratio	$P_{\text{type}}$
	Step	Nonstep	Step	Nonstep	Step	Nonstep			
RH	3.29	4	2.33	4.34	2.15	5.39	0.25	1.00	0.37
								5.31	0.03
								13.18	0.001
$H_e$ maize	0.68	0.82	0.43	0.61	0.42	0.66	0.65	57.03	0.001
								47.23	0.001
$H_e$ teosinte	0.78	0.88	0.49	0.72	0.53	0.77	0.66	57.68	0.001
								57.69	0.001

$H_e$  on the other hand was greatly and evenly affected by both the structure of the repeated motif (interlocus variation of the mutation rate) and the mode of evolution ( $P < 0.001$  for both factors in maize and in teosinte).

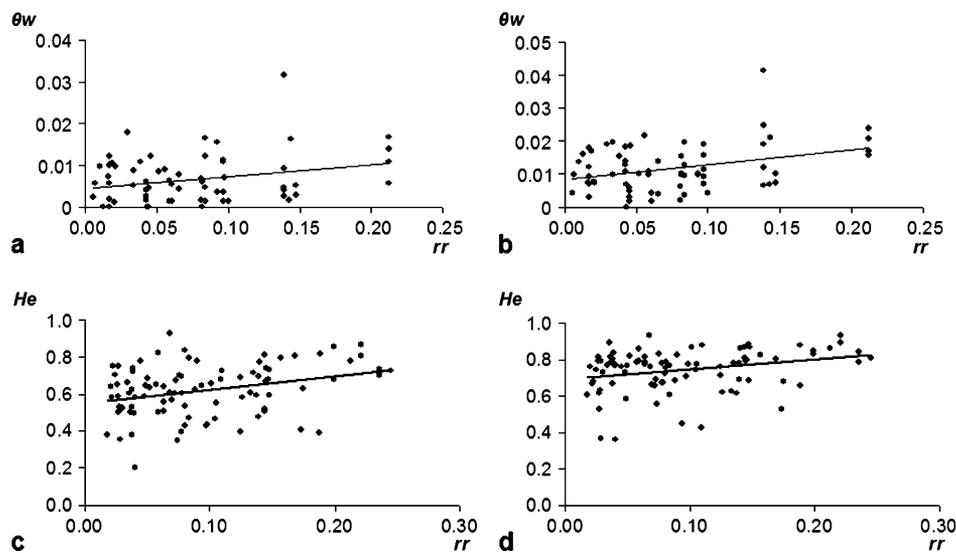
On average, stepwise loci displayed a lower RH than nonstepwise loci, dinucleotide loci had higher heterozygosity than tetranucleotide loci, which had generally higher heterozygosity than trinucleotide loci, and nonstepwise loci had higher heterozygosity than stepwise loci.

Consequently, although  $H_e$  reflected interlocus variation in the mutation rate, RH was likely not affected by such variation because loci known to vary in mutation rate (di- vs. tri- vs. tetranucleotide loci) did not vary accordingly in RH. However, differences in the model of evolution between loci as well as intralocus variation of the mutation rate between maize and teosinte impacted RH for microsatellites. Therefore, plots of RH versus local recombination rate were interpreted for stepwise and nonstepwise loci separately and both with and without the 25 loci having the largest differences in allele size between maize and teosinte.

In addition, the difference in allele size between maize and teosinte was not correlated with local recombination rate. However, among the 25 loci of higher suspected intralocus variation of the mutation rate, 20 are nonstepwise dinucleotide and 5 are stepwise dinucleotide. Dinucleotide loci had on average significantly larger allele size differences between maize and teosinte than tri- and tetranucleotide (1.34, 0.51, and 0.65 repeat difference for di-, tri-, and tetranucleotide loci, respectively;  $P < 0.001$ ).

### Correlation between Diversity and Local Recombination Rate

We observed correlation neither, in maize or in teosinte, between recombination rate and  $\theta_w$  for gene sequences nor between recombination rate and 2 estimates of diversity ( $H_e$  and RH) for stepwise or nonstepwise microsatellites, either including or excluding the 25 loci exhibiting high intralocus variation of the mutation rate. The lack of correlation did not seem to be related to the undersampling of regions of the genome with low recombination rates because we also observed no correlation on chromosomes 6 and 10 alone, the 2 chromosomes for which regions of low recombination



**Figure 1.** Correlations between diversity ( $\theta_w$  or  $H_e$ ) and local recombination rate ( $rr$ ) for genes on chromosome 2 (a, in maize, and b, in teosinte) and for all nonstepwise trinucleotide loci (c, in maize, and d, in teosinte).

rate were well represented by their sample of markers. For microsatellites, the variance in allele size and the stepwise index did not correlate with the recombination rate, as both correlations exhibited a  $\rho$  value close to 0 in maize as well as in teosinte.

Considering individual chromosomes, we detected a weak positive correlation between sequence diversity,  $\theta_w$ , and recombination for chromosome 2 in maize and in teosinte, but after Bonferroni correction for multiple tests, it remained significant only in teosinte ( $\rho = 0.27$ ,  $P < 0.31$  in maize and  $\rho = 0.34$ ,  $P < 0.05$  in teosinte; Figure 1a and b). This correlation was not confirmed for microsatellite loci on the same chromosome. Also, we detected no correlation on this specific chromosome between RH (for genes or for microsatellites) and recombination or between variance in allele size and recombination.

Considering the different groups of microsatellite loci separately (stepwise or nonstepwise di-, tri-, and tetranucleotide loci), significant positive correlations between  $H_e$  and recombination rate were observed only for trinucleotide loci deviating from the stepwise mutation model, both in maize and in teosinte, but they did not remain significant after Bonferroni correction ( $\rho = 0.24$ ,  $P < 0.12$  in maize and  $\rho = 0.22$ ,  $P < 0.18$  in teosinte; Figure 1c and d). No corresponding relationship was observed between RH or the variance in allele size and recombination rate for this category of markers.

Two of the genes sampled from chromosomes 2 and 8 of the trinucleotide loci after a nonstepwise mutation model showed evidence of selection in maize (Vigouroux et al. 2005; Wright et al. 2005). When the 2 genes that showed evidence of selection were excluded, the correlation in maize was slightly lower. When the 8 trinucleotide microsatellites were excluded, the slope was a bit higher in maize and the

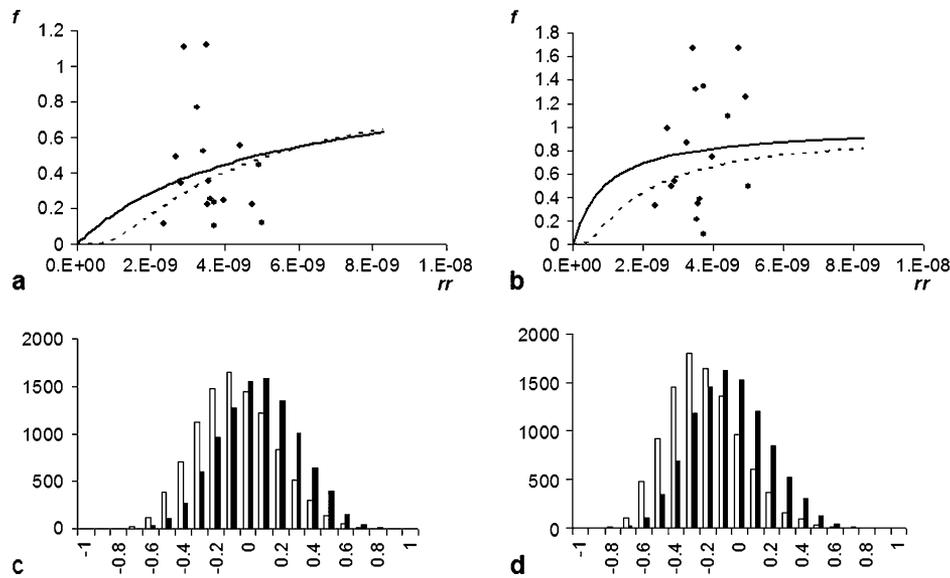
correlation had stronger statistical support ( $\rho = 0.32$ ,  $P < 0.03$  in maize and  $\rho = 0.28$ ,  $P < 0.10$  in teosinte).

#### Testing Correlations for the Hitchhiking or Background Selection Hypothesis

We applied a discriminating test between hitchhiking and background selection (Innan and Stephan 2003) to the 4 observed correlations (for genes on chromosome 2 and for nonstepwise, trinucleotide microsatellites, both in maize and in teosinte).

Nucleotide diversity data on chromosome 2 did not allow us to fit a regression curve with a positive relationship between  $f$  and the recombination rate. We could therefore not obtain suitable estimates for parameters under both models for any  $\theta_{Neu}$  values or range of recombination rates; consequently, we could not obtain expected distributions of the coefficient of correlation under hitchhiking and background selection, and the test was not applicable.

For nonstepwise trinucleotide microsatellites, we observed a correlation coefficient between  $\theta$  and  $\theta/rr$  of 0.97 in maize and 0.93 in teosinte. Best fitting functions for the relationship between  $f$  and the recombination rate under hitchhiking and background selection were found for  $a = 4.9 \times 10^{-9}$  and  $u = 3.6 \times 10^{-9}$  in maize and for  $a = 9 \times 10^{-10}$  and  $u = 1.6 \times 10^{-9}$  in teosinte (Figure 2a and b). Figure 2c and d show the expected distributions of the correlation coefficient under both models. Although the variance of  $\theta$  in our data set seems too high to discriminate between the 2 hypotheses on the basis of observed correlation coefficients between  $\theta$  and  $\theta/rr$ , the observed coefficients in both maize and teosinte were greater than expected under both models ( $P < 0.0001$ ). The same results were obtained with different values of  $\theta_{Neu}$ .



**Figure 2.** Fraction  $f$  of the genetic variation maintained under indirect selection as a function of the recombination rate  $rr$  for the hitchhiking (HH, solid curves) and background selection (BS, dashed curves) models for nucleotide diversity on chromosome 2 (a) and for nonstepwise trinucleotide microsatellite loci (b); (c) and (d) show the distributions of the correlation coefficient between  $\theta$  and  $\theta/rr$  under the hitchhiking (open bars) and the background selection (solid bars) models.

## Discussion

Recombination in plant genome evolution may influence both mutation and selection (for a review, see Gaut et al. 2007). Through this influence, recombination is then often positively correlated with genetic diversity in plants, but the existence of such a correlation in maize is still unclear (Tenaillon et al. 2002). We investigated the maize genome for the relationship between diversity and recombination rate through the study of gene diversity and microsatellites and using 3 different parameters (heterozygosity, reflecting population history; mutation and selection, RH, supposedly affected only by population history; and the variance of allele size, greatly affected by indels). Here we first discuss the representation of the recombination rate across the genome by our sample of loci. Second, we comment on the reliability of RH for not being affected by mutational variations for microsatellites. Finally, we discuss and interpret our findings on the relationship between recombination rate and diversity in our data set.

### Recombination Rates

The physical positions of genetic markers in maize were well predicted from the cytological map based on recombination nodules (Anderson et al. 2004). However, 3 sources of imprecision could have limited our ability to detect significant correlations between diversity and recombination rate.

First, each 0.2- $\mu$ m bin represents about 1.5 Mbp based on an estimated maize genome size of 2500 Mbp (Arumaganthan and Earle 1991). Variation of the recombination rate is likely to exist on a scale of kilobases

due to recombination hot spots in maize (Dooner 1986; Okagaki and Weil 1997; Fu et al. 2002). Their frequency along the chromosome is unknown, but relying on observations in humans, they might occur as frequently as once every 50–200 kb, with each hot spot covering 1 kb (Nachman 2002; McVean et al. 2004; Myers et al. 2005). In other words, there could be as much as 8–30 kb of hot spots within each 1500-kb bin or 0.5–2% of each bin. Hence, we expect a large amount of variation in the local recombination rate across each bin. However, the recombination nodules frequency for a bin should reflect the frequency of recombination hot spots within that bin. Furthermore, the degree to which loci within a bin are influenced by indirect selection should also reflect the frequency of recombination hot spots. Hence, we hope recombination hot spots not to be too much of a confounding factor.

Second, because variation of the recombination rate may also occur between individuals, recombination rate estimates from a single maize line might not be accurate in teosinte. Because the recombination nodules distribution has not yet been studied in teosinte, it is currently not possible to know if there is significant variation in the recombination nodules frequency between maize and teosinte. However, Anderson et al. (2003) reported minor average differences in chiasma frequency between different maize lines, allowing the supposition that the distribution of recombination nodules does not differ greatly, at least among maize inbred lines.

Third, the IBM2 neighbors genetic map, although it is the best map available at this time, is a consensus of multiple genetic maps among which distances between loci as well as gene order sometimes vary. Therefore, some loci might have been assigned to the wrong bin, which would associate their

diversity with the wrong recombination rate estimate. The large size of the physical bins is an advantage in regards to this problem, as it limits the risk of placing a locus in the wrong bin even if its location on the genetic map is not perfectly accurate.

Consequently, we think inaccuracy in the estimation of recombination rates is not a major issue in this study. On the other hand, their representation through our sample of loci, biased toward elevated values compared with the entire range of recombination rates in the maize genome, may interfere with our ability to detect a correlation between diversity and the recombination rate. However, more representative sampling of lower recombination rates on chromosomes 6 and 10 did not lead to better correlations. In addition, other studies that looked at the relationship between diversity and recombination did not differ from our study in the range of recombination rates they considered (Kraft et al. 1998; Stephan and Langley 1998; Nachman et al. 1998; Roselius et al. 2005). It is likely that in these studies, the representation of the recombination rate in the genome is no better than in our study because loci cannot be mapped in regions of very low recombination rate. Consequently, the range of recombination rates sampled in our study should not have prevented us from detecting an effect of indirect selection if it existed.

### Diversity

The diversity maintained in maize from teosinte was much higher for microsatellites than for gene sequences. One reason is that microsatellite data are reported for maize landraces, whereas the sequence data are reported for modern inbreds, which have gone through an additional recent bottleneck. After the domestication bottleneck, a gene like *Adb1*, supposedly neutral regards to the domestication event, is reported to have retained in maize 75% of the diversity that was present in its wild ancestor teosinte (Eyre-Walker et al. 1998). This is still lower than observed at our microsatellite loci taken as a whole but is closer to what we obtain at stepwise loci alone, which retained 80% of the diversity in teosinte. The remaining difference can be explained by the high mutation rate of microsatellites, which may recover diversity to a greater extent than sequences (Vigouroux et al. 2002).

These observations are consistent with the reported effect in our study of the mode of evolution on microsatellite diversity as well as on RH. In addition, they indicate that, although the properties of the RH estimator depend on the mode of microsatellite evolution, it is a good indicator of diversity loss when the assumptions underlying its interpretation are respected. The variance in allele size on the other hand displayed very different results. Although the observed discrepancies between  $H_e$  and allele size variance could be due in part to estimation errors or imprecision, they seem to also indicate that the variance in allele size is not a good indicator of absolute diversity. This measure is likely to be highly affected by indels in the flanking sequence or interrupting the repeat motif.

### Reliability of RH for Being Independent of Mutation Rate Variation

RH, as predicted by theory, was insensitive to interlocus variation of the mutation rate. However, intralocus variation of the mutation rate affected RH and likely blurred the relationship between RH and recombination, especially for dinucleotide loci. But because the relationship between diversity and recombination remained unchanged with or without outlier loci, this does not explain the absence of correlation. Although it could not be tested here, it should be noted that the structure of the repeated region, which can be perfect or interrupted because of punctual mutations, may also vary between individuals and influence the mutation rate (Brinkmann et al. 1998).

The ANOVA also showed that the mutational process had a significant effect on RH values. One may explain this result by noticing that the higher is the actual value of the parameter  $\theta = 4N_e\mu$  the higher is the bias induced by the estimator  $\hat{\theta}_{H_e}$  (Thuillet et al. 2005). Because we calculate RH as  $\hat{\theta}_{H_e}$  in teosinte divided by  $\hat{\theta}_{H_e}$  in maize and because the diversity level is higher in teosinte than in maize, the bias on  $\theta$  is expected to be higher in teosinte as well, leading to higher RH values for nonstepwise loci (rather than lower due to the way RH is calculated) than for stepwise loci, for which no bias is expected.

Contrary to RH, the type of motif affected  $H_e$  values significantly, reflecting differences of the mutation rate for dinucleotide and trinucleotide loci. Unexpectedly, tetranucleotide loci displayed higher  $H_e$  values than trinucleotide loci. Only 7 tetranucleotide loci appear to be stepwise loci. Consequently,  $H_e$  for this group is hard to interpret. The higher  $H_e$  observed for nonstepwise tetranucleotide loci versus nonstepwise trinucleotide loci could be due to selection on trinucleotide loci because most of these are located in coding regions. However, when the nonstepwise trinucleotide loci that were inferred by Vigouroux et al. (2005) to have been subjected to indirect selection were excluded, heterozygosity remained unchanged (data not shown). This seems to indicate that selection was not the main reason explaining why heterozygosity was lower for trinucleotide loci than for tetranucleotide loci. Mutational causes (for instance, unequal proportions of interrupted alleles in both groups) might also have led to the observed difference in heterozygosity between the groups, but we are unable to draw this conclusion on the basis of the present data. Heterozygosity was also affected by the mode of evolution, with higher values for nonstepwise loci than for stepwise loci ( $P < 0.001$ ). An explanation could be that microsatellites with a mutational history involving random indels are expected to have less homoplasmy (as defined by the probability that 2 alleles identical by state are not identical by descent; Estoup et al. 2002) and thus should be more variable than strict stepwise loci.

### Correlations

We tested for the existence of a positive correlation between diversity and recombination rate in maize as a signature of

an extended effect of indirect selection during domestication. Because such a correlation could also have a neutral explanation if the mutation rate is positively associated with the recombination rate (for instance, if recombination is mutagenic), we also checked for a negative correlation between recombination rate and the statistic RH. However, when all data were considered together, we found no correlation between recombination and  $\theta_w$ ,  $H_e$ , or RH.

After partitioning the data set, we found a positive correlation between 1)  $\theta_w$  and the recombination rate on chromosome 2 and 2)  $H_e$  and the recombination rate for nonstepwise trinucleotide loci in both maize and in teosinte. Although the latter 2 were not significant after correction for multiple tests, the correlation between  $H_e$  and the recombination rate for nonstepwise trinucleotides was statistically supported in maize when loci with signatures of past positive selection were excluded. No significant correlations were found for microsatellites between RH and the recombination rate or between allele size variance and the recombination rate. In addition, when the Innan and Stephan (2003) test was applied, observed correlations between  $f$  and the recombination rate were not compatible with expectations under hitchhiking and background selection.

At best, our data suggest a very subtle effect of indirect selection on the maize genome. Whereas the positive correlation on chromosome 2 for genes but not for microsatellites could be consistent with hitchhiking, the absence of correlation with RH on the same chromosome does not support this hypothesis. Furthermore, there is no obvious reason why the effect would be limited to the second chromosome: chromosome 2 lacks low recombination rates and therefore differences in the recombination rate representation among chromosomes do not explain the existence of a correlation for this chromosome in particular. Given the very low statistical support for this correlation, it is possible that it was merely a spurious type I error with no biological meaning. The absence of a genome-wide indirect selective effect on neutral diversity is consistent with previous results in maize (Tenailon et al. 2002), as well as with data suggesting a weak effect of selection during domestication on neutral diversity (Vigouroux et al. 2005) and with the rapid decay of linkage disequilibrium in maize (for a review, see Clark et al. 2004; Flint-Garcia et al. 2003).

The correlation observed for microsatellite loci, on the other hand, might have a selective explanation, although tenuous, if we consider that the trinucleotide loci, but not the other microsatellite loci, were located within genes and might have been particularly subject to purifying selection. The background selection hypothesis, being less likely to explain the effect of domestication on diversity, is consistent with the fact that the correlation was observed in both maize and teosinte. However, diversity in maize is linked through history to that in teosinte, and diversity indices in the 2 species are hence somewhat correlated. Another argument is that, unlike a hitchhiking effect during domestication, the effect of background selection has likely been similar in both species and thus should cancel out in

the calculation of RH; this could explain the absence of a correlation between RH and the recombination rate. Finally, this explanation is also supported by the fact that removing positively selected loci from the correlation improves it rather than weakening it: a better correlation is expected when selection coefficients along the genome are more homogenous, which may be more likely if only loci affected by one type of selection remain.

Two observations are challenging this background selection hypothesis. The first is that stepwise trinucleotide loci did not exhibit the same correlation as did nonstepwise trinucleotide loci, although both types were developed from expressed sequenced tags and were hence located within genes. Given the weakness of the relationship detected on nonstepwise trinucleotide loci, this could merely be a question of power, as only 65 loci were studied for stepwise trinucleotides versus 99 for the nonstepwise trinucleotides. Another explanation could be that nonstepwise mutations are more deleterious than stepwise mutations within genes, given that indels of random size are more likely to disrupt the open reading frame of a gene.

The second observation is that attempts to fit the observed correlations under the hitchhiking or the background selection models failed. Although we observed positive correlations in our data set as predicted under background selection and although the expected distribution under background selection was slightly closer to our observed correlations, these always fell very much outside of the distribution. Problems underlined by Innan and Stephan (2003) regarding their test are the high variance of  $\theta$ , which prevents discrimination of the 2 distributions, and the choice of  $\theta_{\text{Neu}}$ . In our case, we tried several values of  $\theta_{\text{Neu}}$ , but the variance of  $\theta$  seemed to be too large as both distributions always overlapped considerably. Above all, however, the test is applicable only for low recombination rates. Although we chose a range of recombination rates similar to the one chosen by the authors and also tried several other ranges, it could be that our recombination rates are too high to conform to the assumptions of the statistical test.

Tenailon et al. (2002), who found a positive correlation for microsatellite loci but not for sequence diversity, studied microsatellites with different repeat sizes from mononucleotide to 7 repeated bases. A common feature of their microsatellites and ours is that they are all located within genes. To conclude, we favor the background selection hypothesis as the best explanation of our results on nonstepwise trinucleotides. Consistent with rapid decay of linkage disequilibrium in maize, the effect of background selection was quite narrow in scope, limited to genomic regions very tightly linked to target of purifying selection. Finally, we suggest that a stronger purifying selection on our nonstepwise trinucleotide microsatellites located within genes than on nucleotide substitutions in nonrepeated gene sequences could explain that no correlation was observed between nucleotide diversity and recombination in this study or by Tenailon et al. (2002). This would be further consistent with a proposition from Hancock et al. (2001),

who suggest that CAG repeats within genes may end evolving under a higher-than-average purifying selection level.

## Supplementary Material

Supplementary material can be found at <http://www.jhered.oxfordjournals.org/>.

## Funding

The National Science Foundation (DBI-0096033, DBI-0321467, MCB-9728673, MCB-0314644).

## Acknowledgments

We thank Jeff Glaubitz and 2 anonymous reviewers for useful comments on a previous version of this manuscript.

## References

- Anderson LK, Doyle GG, Brigham B, Carter J, Hooker KD, Lai A, Rice M, Stack SM. 2003. High-resolution crossover maps for each bivalent of *Zea mays* using recombination nodules. *Genetics*. 165:849–865.
- Anderson LK, Salameh N, Bass HW, Harper LC, Cande WZ, Weber G, Stack SM. 2004. Integrating genetic linkage maps with pachytene chromosome structure in maize. *Genetics*. 166:1923–1933.
- Arumaganthan K, Earle E. 1991. Nuclear DNA content of some important plant species. *Plant Mol Biol Rep*. 9:208–218.
- Begun DJ, Aquadro CF. 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature*. 356:519–520.
- Brinkmann B, Klintschar M, Neuhuber F, Huhne J, Rolf B. 1998. Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *Am J Hum Genet*. 62:1408–1415.
- Brodhe J, Möller AP, Ellegren H. 2004. Individual variation in microsatellite mutation rate in barn swallows. *Mutat Res*. 545:73–80.
- Buckler IVES, Thornsberry JM, Kresovich S. 2001. Molecular diversity, structure and domestication of grasses. *Genet Res*. 77:213–218.
- Bussel JJ, Pearson NM, Kanda R, Filatov DA, Lahn BT. 2006. Human polymorphism and human-chimpanzee divergence in pseudoautosomal region correlate with local recombination rate. *Gene*. 368:94–100.
- Chakraborty R, Kimmel M, Stivers DN, Davison LJ, Deka R. 1997. Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. *Proc Natl Acad Sci USA*. 94:1041–1046.
- Charlesworth B, Morgan MT, Charlesworth D. 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics*. 134:1289–1303.
- Clark RM, Linton E, Messing J, Doebley JF. 2004. Pattern of diversity in the genomic region near the maize domestication gene *tb1*. *Proc Natl Acad Sci USA*. 101:700–707.
- Cleveland WS. 1981. LOWESS: a program for smoothing scatterplots by robust locally weighted regression. *Am Stat*. 35:54.
- Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB. 1994. Mutational processes of simple-sequence repeat loci in human populations. *Proc Natl Acad Sci USA*. 91:3166–3170.
- Dooner HK. 1986. Genetic fine structure of the *bronzx* locus in maize. *Genetics*. 113:1021–1036.
- Dvorak J, Luo M, Yang Z. 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.
- Estoup A, Jarne P, Cornuet JM. 2002. Homoplasmy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Mol Ecol*. 11:1591–1604.
- Eyre-Walker A, Gaut RL, Hilton H, Feldman DL, Gaut BS. 1998. Investigation of the bottleneck leading to the domestication of maize. *Proc Natl Acad Sci USA*. 95:4441–4446.
- Flint-Garcia SA, Thornsberry JM, Buckler IVES. 2003. Structure of linkage disequilibrium in plants. *Annu Rev Plant Biol*. 54:357–74.
- Fu HH, Zheng ZW, Dooner HK. 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 BS, Wright SI, Rizzon C, Dvorak J, Anderson LK. 2007. Recombination: an underappreciated factor in the evolution of plant genomes. *Nat Rev Genet*. 8:77–84.
- Hancock JM, Worthey EA, Santibáñez-Koref MF. 2001. A role for selection in regulating the evolutionary emergence of disease-causing and other coding CAG repeats in human and mice. *Mol Biol Evol*. 18:1014–1023.
- Hellmann I, Ebersberger I, Ptak SE, Pääbo S, Przeworski M. 2003. A neutral explanation for the correlation of diversity and recombination rate in humans. *Am J Hum Genet*. 72:1527–1535.
- Hellmann I, Prüfer K, Ji H, Zody MC, Pääbo S, Ptak SE. 2005. Why do human diversity levels vary at a megabase scale? *Genome Res*. 15:1222–1231.
- Hudson RR, Kaplan NL. 1995. Deleterious background selection with recombination. *Genetics*. 141:1605–1617.
- Innan H, Stephan W. 2003. Distinguishing the hitchhiking and the background selection models. *Genetics*. 165:2307–2312.
- Kaplan N, Hudson R, Langley C. 1989. The "hitchhiking effect" revisited. *Genetics*. 123:887–899.
- Kauer MO, Dieringer D, Schlötterer C. 2003. A microsatellite variability screen for positive selection associated with the "out of Africa" habitat expansion of *Drosophila melanogaster*. *Genetics*. 165:1137–1148.
- Kraft T, Sall T, Magnusson-Rading I, Nilsson NO, Hallden C. 1998. Positive correlation between recombination rates and levels of genetic variation in natural populations of sea beet (*Beta vulgaris* subsp. *maritima*). *Genetics*. 150:1239–1244.
- Lercher MJ, Hurst LD. 2002. Human SNP variability and mutation rate are higher in regions of high recombination. *Trends Genet*. 18:337–340.
- Liu K, Muse SV. 2005. PowerMarker: integrated analysis environment for genetic marker data. *Bioinformatics*. 21:2128–2129.
- Maynard-Smith J, Haigh D. 1974. The hitchhiking effect of a favorable gene. *Genet Res*. 23:23–35.
- McVean GA, Myers SR, Hunt S, Deloukas P, Bentley DR, Donnelly P. 2004. The fine-scale structure of recombination rate variation in the human genome. *Science*. 304:581–584.
- Myers S, Bottolo L, Freeman C, McVean G, Donnelly P. 2005. A fine-scale map of recombination rates and hotspots across the human genome. *Science*. 310:321–324.
- Nachman MW. 2002. Variation in recombination rate across the genome: evidence and implications. *Curr Opin Genet Dev*. 12:657–663.
- Nachman MW, Bauer VL, Cromwell SL, Aquadro CF. 1998. DNA variability and recombination rates at X-linked loci in humans. *Genetics*. 150:1133–1141.

- Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.
- Ohta T, Kimura M. 1973. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genet Res.* 22:201–204.
- Okagaki RJ, Weil CF. 1997. Analysis of recombination sites within the maize *waxy* locus. *Genetics.* 147:815–821.
- Piperno DR, Flannery KV. 2001. The earliest archaeological maize (*Zea mays* L.) from highland Mexico: new accelerator mass spectrometry dates and their implications. *Proc Natl Acad Sci USA.* 98:2101–2103.
- Roselius K, Stephan W, Städler T. 2005. The relationship of nucleotide polymorphism, recombination rate and selection in tomato species. *Genetics.* 171:753–763.
- Schug MD, Hutter CM, Wetterstrand KA, Gaudette MS, Mackay TF, Aquadro CF. 1998. The mutation rates of di-, tri- and tetranucleotide repeats in *Drosophila melanogaster*. *Mol Biol Evol.* 15:1751–1760.
- Sherman JD, Stack SM. 1995. Two-dimensional spreads of synaptonemal complexes from Solanaceous plants: high-resolution recombination nodule map for tomato (*Lycopersicon esculentum*). *Genetics.* 141:683–708.
- Smith BD. 2001. Documenting plant domestication: the consilience of biological and archaeological approaches. *Proc Natl Acad Sci USA.* 98:1324–1326.
- Sokal RR, Rohlf FJ. 2000. Assumptions of analysis of variance. *Biometry.* 2nd ed. New York: W.H. Freeman and Company. p. 400–453.
- Stephan W, Langley CH. 1998. DNA polymorphism in *Lycopersicon* and crossing-over per physical length. *Genetics.* 150:1585–1593.
- Tenaillon MI, Sawkins MC, Anderson LK, Stack SM, Doebley J, Gaut BS. 2002. Patterns of diversity and recombination along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Genetics.* 162:1401–1413.
- Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS. 2001. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proc Natl Acad Sci USA.* 98:9161–9166.
- Thuillet AC, Bataillon T, Poirier S, Santoni S, David JL. 2005. Estimation of long-term effective population sizes through the history of durum wheat using microsatellite data. *Genetics.* 169:1589–1599.
- Vigouroux Y, Jaqueth JS, Matsuoka Y, Smith OS, Beavis WD, Smith JS, Doebley JF. 2002. Rate and pattern of mutation at microsatellite loci in maize. *Mol Biol Evol.* 19:1251–1260.
- Vigouroux Y, Mitchell S, Matsuoka Y, Hamblin M, Kresovich S, Smith JSC, Jaqueth J, Smith OS, Doebley J. 2005. An analysis of genetic diversity across the maize genome using microsatellites. *Genetics.* 169:1617–1630.
- 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.
- Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS. 2005. The effect of the artificial selection on the maize genome. *Science.* 308:1310–1314.

Received February 7, 2007

Accepted August 28, 2007

Corresponding Editor: John Burke