

# Rate and Pattern of Mutation at Microsatellite Loci in Maize

Yves Vigouroux,\* Jennifer S. Jaqueth,† Yoshihiro Matsuoka,‡ Oscar S. Smith,†  
William D. Beavis,§ J. Stephen C. Smith,† and John Doebley\*

\*Department of Genetics, University of Wisconsin, Madison; †Crop Genetics Research and Development, DuPont Agriculture and Nutrition, Pioneer Hi-Bred International, Johnston, Iowa; ‡Fukui Prefectural University, Matsuoka-cho, Yoshida-gun, Fukui, Japan; and §National Center for Genome Resources, Santa Fe, New Mexico

Microsatellites are important tools for plant breeding, genetics, and evolution, but few studies have analyzed their mutation pattern in plants. In this study, we estimated the mutation rate for 142 microsatellite loci in maize (*Zea mays* subsp. *mays*) in two different experiments of mutation accumulation. The mutation rate per generation was estimated to be  $7.7 \times 10^{-4}$  for microsatellites with dinucleotide repeat motifs, with a 95% confidence interval from  $5.2 \times 10^{-4}$  to  $1.1 \times 10^{-3}$ . For microsatellites with repeat motifs of more than 2 bp in length, no mutations were detected; so we could only estimate the upper 95% confidence limit of  $5.1 \times 10^{-5}$  for the mutation rate. For dinucleotide repeat microsatellites, we also determined that the variance of change in the number of repeats ( $\sigma_m^2$ ) is 3.2. We sequenced 55 of the 73 observed mutations, and all mutations proved to be changes in the number of repeats in the microsatellite or in mononucleotide tracts flanking the microsatellite. There is a higher probability to mutate to an allele of larger size. There is heterogeneity in the mutation rate among dinucleotide microsatellites and a positive correlation between the number of repeats in the progenitor allele and the mutation rate. The microsatellite-based estimate of the effective population size of maize is more than an order of magnitude less than previously reported values based on nucleotide sequence variation.

## Introduction

Microsatellites or simple sequence repeats (SSRs) are composed of a DNA sequence motif of 1–6 bases in length that is repeated tandemly usually five or more times. Microsatellite markers have become widely used in studies of parentage (e.g., Peters et al. 1995), linkage (e.g., Lagercrantz, Ellegren, and Andersson 1993), population genetics (e.g., Goldstein et al. 1999), and evolutionary history (e.g., Bowcock et al. 1994). These loci are also used as aids for selection in breeding programs (e.g., Powell et al. 1996) and for the characterization of inbred lines and varieties of cultivated plants (e.g., Smith et al. 1997).

Numerous studies in animals have calculated the mutation rate of microsatellites, and these studies have shown that the mutation rate varies greatly among species, ranging from  $5 \times 10^{-6}$  in *Drosophila* (Schug, Mackay, and Aquadro 1997; Schlötterer et al. 1998; Schug et al. 1998; Vazquez et al. 2000) to  $10^{-3}$  in human (Brinkmann et al. 1998; Xu et al. 2000). These studies have also outlined several trends of the mutation process that are important for understanding microsatellite evolution. They have shown that the mutation rate varies widely among loci within species (Di Rienzo et al. 1998; Harr et al. 1998) and that the mutation rate increases with the length of the microsatellite (Primmer et al. 1996; Brinkmann et al. 1998). It has also been shown that there is a constraint on the size of microsatellites (Garza, Slatkin, and Freimer 1995) which may simply be the effect of the increased probability of contraction with the size of the microsatellite (Ellegren 2000; Harr and Schlötterer 2000; Xu et al. 2000).

Key words: *Zea mays* subsp. *mays*, microsatellite, mutation rate, maize, teosinte, SSR.

Address for correspondence and reprints: John Doebley, Department of Genetics, University of Wisconsin, Madison, Wisconsin 53706. E-mail: jdoebley@facstaff.wisc.edu.

*Mol. Biol. Evol.* 19(8):1251–1260. 2002

© 2002 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

Mutation rate is a critical parameter in population genetic models because it enables one to relate the variability at microsatellite loci to the history of a population or to the history of a portion of the genome. For example, under the hypothesis of a generalized stepwise model of mutation, knowledge of the mutation rate permits one to estimate the time of divergence between species (Wehrhahn 1975; Goldstein et al. 1995) and the effective population size of species (Slatkin 1995).

In this article, we report the mutation rate for maize (*Zea mays* subsp. *mays*) microsatellites as determined in two different experiments of mutation accumulation involving a total of 142 microsatellites. We also describe the nature of the mutation process, including whether there is a bias toward mutations that increase versus decrease allele size, whether the number of repeats in the progenitor allele is correlated with the mutation rate, and whether the mutation rate differs among loci with dinucleotide versus trinucleotide or higher-repeat motifs. Finally, we apply the mutation rate that we have determined to estimate the effective population sizes of maize and its wild progenitor (*Z. mays* subsp. *parviglumis*).

## Materials and Methods

### Plant Material

Our analyses involved two separate experiments of mutation accumulation. In experiment I, six maize inbred lines were studied (B73, Mo17, Oh43, PHAA0, PH24E, and PHN46), that had been developed through at least seven generations of self-pollination, after the initial breeding cross of their respective parents. Between 55 and 78 microsatellites, which collectively allowed each chromosome arm to be sampled, were used to monitor genetic contamination during two subsequent generations of self-pollination to identify and eliminate any off-types that might have resulted from pollen contamination of the founding seed stocks. The first of these generations was grown in 1998 in Johnston, Iowa. The

second generation was grown in 1999 in each of two locations, Johnston, Iowa, and York, Neb. Using only plants that showed no evidence of pollen contamination, two ears for each of the six inbreds were harvested at both of the 1999 locations and used for this study. Twenty-three progeny from each of the 24 ears (552 total plants) as well as the parents were assayed and compared for evidence of new mutations.

In experiment II, 86 recombinant inbred (RI) lines were studied. These lines were derived from two different crosses, T232 × CM37 (45 lines) and Co159 × Tx303 (41 lines) (Burr et al. 1988). The F<sub>2</sub> progeny of a single F<sub>1</sub> plant from each cross were selfed for 9 to 12 generations. The average number of generations of inbreeding after the F<sub>1</sub> was 11.3 generations for the T232 × CM37 cross and 11.0 generations for the Co159 × Tx303 cross. The 86 RI lines and the four parental inbreds were all genotyped and compared for evidence of new mutations. Our genotyping of these 86 RI lines with 98 microsatellite loci revealed no plants with non-parental alleles at multiple loci, as expected if there had been pollen contamination.

#### Microsatellite Genotyping

For experiment I, DNA was extracted from leaf tissue of 10-day-old seedlings or from freeze-dried tissue of young seedlings (Smith et al. 1997). DNAs of each progeny were allocated into duplicate 96-well liquid handling plates. All PCR amplifications and gel runs were made in duplicate for each progeny, and additional replicates of the parents were amplified and electrophoresed for each microsatellite. Forty-eight microsatellite loci were used and represented a variety of repeat motif types: 6 with dinucleotide repeat motifs, 21 with trinucleotide, 15 with tetranucleotide, 4 with pentanucleotide, 1 with hexanucleotide, and 1 with a di-tetra motif (table 1). Microsatellite genotyping was performed on ABI automated sequencers, using procedures that have been described previously (Smith et al. 1997). DNA samples showing putative mutants were amplified a third time for the microsatellite in question.

For experiment II, DNA extractions were performed as described by Matsuoka et al. (2002a). Ninety-eight microsatellite loci were used, including 83 with dinucleotide repeat motifs, 7 with trinucleotide, 6 with tetranucleotide, and 2 with pentanucleotide (table 1). Genotyping was performed at Celera AgGen (Davis, Calif.), using fluorescent primers on an ABI automated sequencer. A first PCR was performed for the 86 plants and the 4 parents at 98 microsatellite loci. After the first PCR, all nonparental alleles were considered as potential mutations. A second PCR or sequencing was then performed for 179 of the 194 potential mutations (table 2). If the second assay showed a parental allele, then the first PCR was considered inaccurate, and the potential mutant allele was classified as a parental allele. (A discrepancy between the first and second PCR could result because of errors in processing or 1-bp errors during the binning procedure [Matsuoka et al. 2002a]). If the second PCR exactly confirmed the potential mutation, then the

allele was reclassified as a bona fide mutation. In the 15 cases where the second PCR failed or was not done, the allele remained classified as a potential mutant allele. Of these, 10 were in dinucleotide microsatellites and 5 in microsatellites with repeat motifs of 3 bp or more (table 2).

#### Sequencing

In experiment I, putative mutant alleles and their founder parent alleles were reamplified, cloned, and sequenced to reconfirm the mutation and to determine its nature. The purified PCR products were cloned using the TOPO cloning kit (Invitrogen). The founding individual and the putative mutation were each sequenced 34 or more times using the M13 forward and reverse primers. The sequences were aligned, and the number of repeat units in the putative mutant was compared with that of the founding individual.

In experiment II, the mutant alleles and the two parental inbreds for its RI line were sequenced. Microsatellite loci were amplified by PCR using PCR Supermix (BRL) and 10 pmol of each primer, as described by Matsuoka et al. (2002a). Two PCRs were performed for each analysis, and the combined PCR products were purified on Qiagen columns. For lines in which the mutant allele was homozygous, the PCR products were directly sequenced with the primers used for the PCR. For heterozygous individuals, the PCR products were cloned into a plasmid vector, using the TOPO cloning kit (Invitrogen). Plasmid clones (from two to eight) were sequenced for each mutation. The sequencing was performed using M13 forward and reverse primers with a BigDye terminator sequencing kit (ABI) at the University of Wisconsin Biotechnology Center (Madison).

#### Statistics

To estimate the mutation rate, one divides the number of observed mutations by the number of independent generations that the two alleles present at the last generation have experienced (i.e., the number of allele-generations). For experiment I which examined only a single generation, the number of allele-generations is simply equal to 23 kernels × 2 alleles × 4 ears × 6 inbreds × 48 loci (=52,992) minus any missing data. For experiment II, one might consider that the number of allele-generations is simply the number of generations (*g*) times 2 (for a diploid). However, a mutation appearing in the heterozygous state in an early generation has a certain probability to be lost by drift during the successive generation of selfing. So, the number of allele-generations is somewhat less than 2*g*. Thus, the number of allele-generations must be determined using the probability of coalescence of the alleles over the ≈11 generations since the F<sub>1</sub>.

We calculated the number of allele-generations for experiment II as follows. From the last generation, going backward one generation, the probability that the two alleles at this last generation coalesced in the generation before the last is 1/2 because the plants were self-pollinated. In this case, the total number of allele-genera-

**Table 1**  
**Microsatellite Loci Analyzed in Experiments I and II**

Locus <sup>a</sup>	Repeat	Exp.	Locus	Repeat	Exp.	Locus	Repeat	Exp.
bnlg1014 ...	AG	Both	bnlg1605	AG	II	phi061	TTCT-GTAT	II
bnlg1017 ...	AG	II	bnlg1662	AG	II	phi062	ACG	I
bnlg1018* ...	AG	II	bnlg1720	AG	II	phi064	ATCC	Both
bnlg1022 ...	AG	II	bnlg1732*	AG	II	phi065	CACTT	II
bnlg1028 ...	AG	II	bnlg1740	AG	II	phi069	GAC	I
bnlg1043* ...	AG	II	bnlg1782*	AG	II	phi070	AGCTG	I
bnlg1046 ...	AG	II	bnlg1784	AG	II	phi072	AAAC	II
bnlg105* ...	AG	II	bnlg1808*	AG	II	phi073	AGC	I
bnlg1065 ...	AG	II	bnlg1831*	AG	II	phi076	AGCGGG	I
bnlg1070* ...	AG	II	bnlg1834	AG	II	phi078	AG-AAAG	II
bnlg1074* ...	AG	II	bnlg1839*	AG	II	phi079	CATCT	I
bnlg1079* ...	AG	Both	bnlg1866*	AG	II	phi083	AGCT	I
bnlg1094 ...	AG	II	bnlg1890	AG	II	phi085	AACGC	I
bnlg1108 ...	AG	II	bnlg1904	AG	II	phi089	ATGC	II
bnlg1129 ...	AG	II	bnlg1917	AG	II	phi093	AGCT	Both
bnlg1131* ...	AG	II	bnlg1940	AG	II	phi096	AGGTG	II
bnlg1138* ...	AG	II	bnlg2047*	AG	II	phi099*	AC	II
bnlg1160 ...	AG	I	bnlg2086	AG	II	phi101	ACT	II
bnlg1176* ...	AG	II	bnlg2122*	AG	II	phi102228	AGAT	I
bnlg118 ...	AG	II	bnlg2132	AG	II	phi104127	AAGC	I
bnlg1182* ...	AG	II	bnlg2238	AG	II	phi108349	TAAA	I
bnlg1189 ...	AG	II	bnlg2259	AG	II	phi109188	AAAG	I
bnlg1191 ...	AG	II	bnlg2271	AG	I	phi109642	ACGG	I
bnlg1192 ...	AG	I	bnlg2305	AG	II	phi115	AT-ATAC	I
bnlg1194 ...	AG	II	bnlg244	AG	II	phi116	ACTG-ACG	II
bnlg1203 ...	AG	I	bnlg252*	AG	II	phi119*	AG	II
bnlg1208 ...	AG	II	bnlg426*	AG	II	phi120	AAG	II
bnlg1209* ...	AG	II	bnlg589	AG	II	phi121	CCG	I
bnlg1237 ...	AG	II	bnlg615	AG	II	phi127	AGAC	I
bnlg1257 ...	AG	II	bnlg619	AG	II	phi200328	TCGCC	I
bnlg1265 ...	AG	II	dupssr14	CT-CT-CA	II	phi236654	CCG	I
bnlg1287* ...	AG	II	dupssr28	GA	II	phi285358	GAA	I
bnlg1288 ...	AG	II	nc004	AG	II	phi311890	GTT	I
bnlg1302 ...	AG	II	nc009	AG	II	phi314704	CGT	I
bnlg1305* ...	AG	II	phi002	AACG	I	phi314965	GCC	I
bnlg1325* ...	AG	II	phi015	AAAC	I	phi328189	CCG	I
bnlg1329* ...	AG	II	phi017	TAC	II	phi333597	AAG	I
bnlg1360* ...	AG	II	phi021*	AG	II	phi337490	CAG	I
bnlg1371 ...	AG	II	phi024	CCT	II	phi338882	CCA	I
bnlg1429 ...	AG	II	phi033	AAG	II	phi364545	AGC	I
bnlg1449 ...	AG	II	phi034	CCT	I	phi386223	AGC	I
bnlg1456* ...	AG	II	phi037*	AG	II	phi389203	AGC	I
bnlg1484 ...	AG	II	phi041	AGCC	I	phi420701	CCG	I
bnlg149* ...	AG	II	phi050	AAGC	II	phi427434	ACC	I
bnlg1520 ...	AG	II	phi051	AGG	II	phi448880	AAG	I
bnlg1523 ...	AG	II	phi053	ATGT	I	phi96342	ATCC	I
bnlg1526* ...	AG	II	phi059	ACC	I	phi98667	TGCA	I
bnlg1538 ...	AG	II						

NOTE.—Exp. indicates loci used in experiments I, II, or both. An asterisk (\*) identifies the 33 loci used in the estimation of the effective population size of maize and subsp. *parviglumis*.

<sup>a</sup>Primer sequences are available at the MaizeDB ([www.agronomy.missouri.edu](http://www.agronomy.missouri.edu)). Celera AgGen adds a 7-bp clamp (GTGTCCT) on to one primer of each pair to suppress the formation of +A alleles.

tions that the two alleles have experienced is  $g + 1$ . If the two alleles have not coalesced at the generation before the last generation, then the probability that they coalesced two generations before the last (knowing that they have not coalesced one generation before the last generation) is 1/4. In this case, the number of allele-generations is  $g + 2$ . This process can be extended up to the  $F_2$  generation at which point the probability that the two alleles have coalesced, knowing that they have not coalesced elsewhere, is  $(1/2)^{g-1}$ , with the number of allele-generations being  $2g - 1$ . Finally, the probability that the two alleles do not coalesce is  $(1/2)^{g-1}$ , with the

number of allele-generations being  $2g$ . A general formula for the expected number of allele-generations is:

$$E(G) = \sum_{i=1}^g p_i G_i = \sum_{i=1}^{g-1} \left(\frac{1}{2}\right)^i (g + i) + 2g \left(\frac{1}{2}\right)^{g-1}$$

where  $p_i$  is the probability of observing a given coalescence configuration, and  $G_i$  is the length of this coalescence configuration in allele-generations. For more than 10 generations, the expression reduces to  $E(G) \approx g + 2$ .

Two main parameters are important in the mutation process of microsatellites: the mutation rate,  $\mu$ , and the

**Table 2**  
**Results for Experiment II**

	MICROSATELLITE REPEAT TYPE		Total
	Dinu- cleotide	Other Repeats	
Number of loci . . . . .	83	15	98
Number of first PCR performed <sup>a</sup> . . . . .	6,578	1,105	7,683
Potential mutations of 2 bp or more . . . . .	141	22	163
Potential mutations of 1 bp . . . . .	25	6	31
Number of second PCR or sequencing . . . . .	156	23	179
Confirmed mutations of 2 bp or more <sup>b</sup> . . . . .	69	0	69
Confirmed mutations of 1 bp . . . . .	3	0	3
Unconfirmed mutations of 2 bp or more . . . . .	7	2	9
Unconfirmed mutations of 1 bp . . . . .	3	3	6

<sup>a</sup> Excluding missing data resulting from PCR reactions that failed.

<sup>b</sup> Here, we count the six apparently nonindependent mutations at nc009 only once.

variance of change in the number of repeats among mutations,  $\sigma_m^2$  (Slatkin 1995; Zhivotovsky and Feldman 1995). To estimate  $\sigma_m^2$ , one needs to know the change in the number of repeats for each mutation from its progenitor allele. For experiment II, when sequence polymorphisms in the regions flanking the microsatellite repeat enabled us to identify the progenitor allele, we inferred the change in the number of repeat units for the mutation by comparison with the progenitor allele. In cases where the progenitor allele could not be unambiguously identified, the change in the number of repeat units for the mutation was inferred by comparison with the parental allele that was most similar in size to the mutant allele.

Even when no mutations are observed, it is possible to calculate an upper limit of the mutation rate using the Poisson law. The probability of zero mutations is  $P(X = 0) = e^{-G\mu}$ . We can solve this equation for  $P(X = 0) = 0.05$  to obtain the upper limit of the mutation rate (Schug, Mackay, and Aquadro 1997). The 95% confidence interval for the observed mutation rate can also be calculated using the Poisson law. The probability of observing  $k$  or fewer mutations with  $\lambda = G\mu$  is

$$P(X \leq k) = \sum_{i=0}^k e^{-\lambda} \frac{\lambda^i}{i!}.$$

When the mutation rate for an experiment was homogeneous among loci, we determined the 95% confidence interval by solving this function with probabilities of 2.5% and 97.5% for observing  $k$  or fewer mutations (Schug, Mackay, and Aquadro 1997). When the mutation rate was nonhomogeneous, we calculated this confidence interval by resampling the loci, using a bootstrap procedure to create 10,000 random samples.

For experiment II, most mutations have been confirmed by two different assays (either two PCRs or one PCR plus sequencing). However, 15 of 194 putative mutations were not analyzed by a second PCR or sequencing. These may be real mutations or errors in the first PCR. To infer the proportion of the unconfirmed putative mutations that are predicted to be real mutations,

we used the proportion of real mutations among those putative mutations that had been subjected to two assays. Before doing this, we divided these unconfirmed putative mutations into two classes: those that differ by 1 bp from a parental allele and those that differ by more than 1 bp from the parental allele.

## Results

### Identification of the Mutant Alleles

Of the 48 loci and 552 individuals analyzed in experiment I, one individual revealed a mutant allele for bnlg1203 in heterozygous condition with the parental allele. Locus bnlg1203 has a dinucleotide repeat, and consequently, stutters can result in amplified products of different lengths. Sequencing of 36 clones from the founding individual revealed 1 clone of 15 repeat units, 5 clones of 17 repeat units, 27 of 18 repeat units, and 3 of 19 repeat units, indicating that the parental allele contained 18 repeat units. Clones from the individual with the heterozygous parental and nonparental allele included 1 clone of 15 repeat units, 8 clones of 16 repeat units, 17 clones of 17 repeat units, and 8 of 18 repeat units. These numbers indicate that the nonparental allele possesses 17 repeat units or one repeat unit less than the parental allele.

Of the 98 microsatellites analyzed in experiment II, several failed in one of the two RI populations. Loci bnlg1839, bnlg2086, and phi096 failed for the T232 × CM37 population, and bngl1074, bnlg1257, and phi064 failed for the Co159 × Tx303 population. There were also 528 cases where the PCR failed for one or a few plants but worked well for the population as a whole. Excluding these missing data, a total of 7,683 successful PCRs were performed (table 2). Of these, 194 gave a nonparental allele or potential mutation. Of these 194 potential mutations, 31 differed by 1 bp from the parental allele and 163 by 2 bp or more. A second PCR or sequencing was performed on 179 of these potential mutant alleles. Of these, 72 were confirmed and were classified as real mutations. Of these 72 real mutations, 3 differed by only 1 bp from the parental allele and 69 by more than 1 bp (table 2, supplementary material available at MBE web site: www.molbioevol.org).

The nc009 locus gave an unexpected result for six RI lines of the Co159 × Tx303 population, each of which possessed the same nonparental allele of 133 bp. These six nonparental alleles were sequenced, and all have the same number of repeats (AG)<sub>18</sub>. Sequence polymorphism in the 3' flanking region identified the 151-bp parental allele as the progenitor of the 133-bp nonparental allele, indicating that there was an 18-bp deletion or loss of nine repeat units. Because one F<sub>1</sub> plant was used to form the F<sub>2</sub> population, it seems most likely that a premeiotic somatic mutation in either the ear or tassel cell lineage of the F<sub>1</sub> plant gave rise to the 133-bp allele which was then inherited by the six RI lines. This type of event has been reported previously (Jones et al. 1999). For this reason, we have interpreted this result as a single mutation rather than six independent mutations.

**Table 3**  
**Mutation Rate of Dinucleotide and Other Repeat Microsatellites**

	Number of Loci	Number of Mutations	Number of Allele-Generations	Mutation Rate (95% Confidence Interval)
<b>Dinucleotide repeats</b>				
Experiment I . . . .	7	1	7,718	$1.3 \times 10^{-4}$ ( $3.1 \times 10^{-5}$ to $7.2 \times 10^{-4}$ )
Experiment II . . . .	83	71.94 <sup>a</sup>	86,517	$8.3 \times 10^{-4}$ ( $5.6 \times 10^{-4}$ to $1.1 \times 10^{-3}$ ) <sup>b</sup>
Combined . . . . .	88	72.94 <sup>a</sup>	94,235	$7.7 \times 10^{-4}$ ( $5.2 \times 10^{-4}$ to $1.1 \times 10^{-3}$ ) <sup>b</sup>
<b>Other repeats</b>				
Experiment I . . . .	41	0	44,568	Upper 95% limit $6.7 \times 10^{-5}$
Experiment II . . . .	15	0	14,532	$2.1 \times 10^{-4}$
Combined . . . . .	54	0	59,100	$5.1 \times 10^{-5}$

<sup>a</sup> The number of mutations is an expected number that includes 2.94 predicted mutations that were not confirmed by a second PCR (see text for details).

<sup>b</sup> Confidence interval obtained by a bootstrap procedure (see text for details).

To further analyze the nature of the mutations, 54 of the 72 mutations from experiment II were sequenced. All mutations confirmed by two PCRs were again verified by sequencing. Of the three 1-bp mutants confirmed by two PCRs, all were 1-bp changes in the length of a mononucleotide tract flanking the microsatellite.

**Estimation of the Mutation Rate**

In experiment I, a single mutation was observed in one of the seven dinucleotide loci assayed. The number of allele-generations in this case is 7,718 after subtracting 10 missing data points. This yields a mutation rate per generation for dinucleotide microsatellites of  $1.3 \times 10^{-4}$  (table 3). The 95% confidence interval for this rate is  $3.1 \times 10^{-5}$  to  $7.2 \times 10^{-4}$ . For microsatellites with repeat units of greater than 2 bp in length, no mutations were observed, and thus we can only calculate the upper bound of the mutation rate. Here, the number of allele-generations is 44,568 which gives an upper bound of the mutation rate of  $\mu = 6.7 \times 10^{-5}$ .

For calculating the mutation rate at dinucleotide microsatellites with data from experiment II, we considered only mutations in the number of repeats at the microsatellite itself and excluded mutations in the mononucleotide tracts in the flanking region. Three of the 72 mutations were 1-bp mutations in flanking mononucleotide tracts, and thus, the total number of mutations in the number of microsatellite repeats is 69 (table 2). There are an additional 10 potential mutations at dinucleotide microsatellites which were not confirmed by a second assay but need to be considered for calculating the mutation rate. These include seven potential mutations of 2 bp or more. From the cases in which two assays have been performed, we can calculate the percentage of potential mutants from the first PCR that were confirmed by a second PCR or sequencing. For potential mutants that differ by 2 bp or more from the parental allele, 42% (69/166) were confirmed as real mutations by a second assay. This percentage can be used to calculate the expected number of real mutations among the seven potential mutations that were not confirmed by a second PCR. The number of unconfirmed mutations of 2 bp or more expected to be real is  $7 \times$

0.42 or 2.94. Thus, the estimated total number of mutations in the number of repeats at the microsatellites is 71.94. The number of allele-generations is 86,517 for dinucleotide microsatellites, and so the mutation rate per generation is  $8.3 \times 10^{-4}$ , with a confidence interval of  $5.6 \times 10^{-4}$  to  $1.1 \times 10^{-3}$  (table 3). For microsatellites with repeats of more than 2 bp, no mutations were observed; however, given that the number of allele-generations was 14,532, we calculate an upper bound of  $2.1 \times 10^{-4}$  for the mutation rate.

In experiment II, one might also consider mutations that were hidden because one parental allele mutated to the other parental allele. This problem would only be significant when the two parental alleles differ by one repeat because mutations of one repeat are the most common. Twenty-one of the loci-by-RI combinations have a one-repeat difference between the two parental alleles, and among these, we observed three one-repeat mutations. Because 50% of the mutations will be to the other parental allele and thus hidden, we estimate three hidden mutations. This will only increase the mutation rate from  $8.3 \times 10^{-4}$  to  $8.7 \times 10^{-4}$ .

If experiments I and II are grouped together by adding up the number of mutations and the number of allele-generations, the mutation rate for dinucleotide microsatellites is  $7.7 \times 10^{-4}$ , with a confidence interval of  $5.2 \times 10^{-4}$  to  $1.1 \times 10^{-3}$  (table 3). For microsatellites with repeats of more than 2 bp in length, the upper bound of the mutation rate for the combined experiments is  $5.1 \times 10^{-5}$ , considering that 59,100 allele-generations were analyzed.

We also used the data from experiment II to calculate that the variance of change in the number of repeats ( $\sigma_m^2$ ) is 3.17 (table 4). This parameter is strongly influenced by outliers with a large change in the number of repeats. For example, if the mutation in nc009 that caused a decrease of nine repeats is excluded, then  $\sigma_m^2$  drops to 2.03. Accordingly, the value given here should be taken with some caution.

**Trends in the Mutational Process**

To test whether the mutation rate is homogeneous among dinucleotide loci, we compared the distribution

**Table 4**  
**Estimation of the Variance of Change in the Number of Repeats ( $\sigma_m^2$ ) for the Dinucleotide Repeat Loci in Experiment II**

	All Dinucleotide Loci	Dinucleotide Loci Used for Population Size Inferences
Number of mutations. . . . .	69	13
Number of repeats		
1. . . . .	57	10
2. . . . .	9	2
3. . . . .	1	1
4. . . . .	0	0
5. . . . .	0	0
6. . . . .	1	0
7. . . . .	0	0
8. . . . .	0	0
9. . . . .	1	0
$\sigma_m^2$ . . . . .	3.17	2.08

of the number of mutations per locus against a Poisson distribution for the average number of mutations over all loci (fig. 1). Before testing whether the observed data fit a Poisson distribution, we grouped all loci with three or more mutations into a single class to reduce the effect of outliers. This is a conservative adjustment. The  $\chi^2$  test showed a lack of fit of the observed and the Poisson distributions (fig. 1;  $\chi^2 = 15.3$ ,  $P < 0.001$ ), indicating that the mutation rate is not homogeneous across loci.

To test whether there is a bias for mutations to cause either an increase or decrease in size, we classified 71 of the 72 mutations from experiment II (including both 1-bp and larger mutations) as either an increase or a decrease in size relative to their progenitor alleles. The one remaining mutation was exactly midway between the two parental alleles and thus could not be classified. This analysis revealed a significant bias in the mutation process for an increase in size, with 56 causing an increase in size as compared with 15 mutations causing a decrease (fig. 2;  $\chi^2 = 23.7$ ,  $P < 0.001$ ). However, the average size of a decrease was 4.1 bp, compared with 2.3 bp for the average increase in size. For mutations of 2 bp or more at dinucleotide loci (68 events), there is an excess of multiple-repeat-unit mutations relative to single-repeat-unit mutations for mutations that decrease in size as compared with those that increase size ( $\chi^2 = 4.9$ ,  $P = 0.03$ ). If mononucleotide and dinucleotide changes are both considered (71 events), this test is nearly significant ( $\chi^2 = 3.2$ ,  $P = 0.07$ ).

We performed two tests to determine whether there is a correlation between the number of repeats in the progenitor allele and the probability of mutation. First, we observed that the number of repeats in an allele is positively correlated with the number of mutations for that allele (fig. 3a;  $R = 0.53$ ,  $P < 0.001$ ). Second, we asked if there is a bias for alleles of the largest size at a locus to mutate relative to alleles of a smaller size. To do this, we calculated the standardized allele size of the mutant alleles, based on available data on allele frequencies (Matsuoka et al. 2002b). The standardized size is a percentile score for an allele relative to other alleles

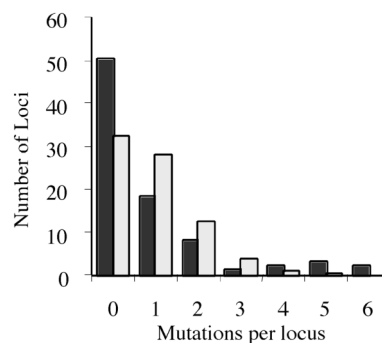


FIG. 1.—Distribution of the dinucleotide microsatellite loci according to the number of observed (dark columns) and expected (gray columns) mutations per locus. The expected distribution is based on the Poisson law, with the same mean as the observed distribution.

at that locus computed as the cumulative frequency from the smallest to the largest allele and using the midpoint between the nearest smaller allele and the allele in question (Ellegren 2000). We observed a strong bias within loci, with 63 of 71 mutating alleles having a standardized allele size greater than 0.5 (fig. 3b;  $\chi^2 = 42.6$ ,  $P < 0.001$ ).

Finally, we tested whether or not the magnitude of the mutations increases with the standardized size of the mutant allele (Ellegren 2000). No significant correlation was observed ( $R = 0.009$ ,  $P = 0.94$ ); however, it is of interest that the three largest contractions are associated with a standardized allele size of more than 0.95.

## Discussion

### Mutation Rate

Our estimate of the mutation rate for maize dinucleotide microsatellites is  $7.7 \times 10^{-4}$  mutations/generation, and our estimate of the upper bound of the rate for microsatellites with repeat motifs of greater than 2 bp is  $5.1 \times 10^{-5}$ . Our estimate for dinucleotide microsatellites is in the middle of the range of values reported for other organisms which vary from  $10^{-2}$  to  $5 \times 10^{-6}$  (Dallas 1992; Dietrich et al. 1992; Schug et al. 1998; Kovalchuk et al. 2000; Vazquez et al. 2000; Udupa and Baum 2001). Multiple factors contribute to the variation

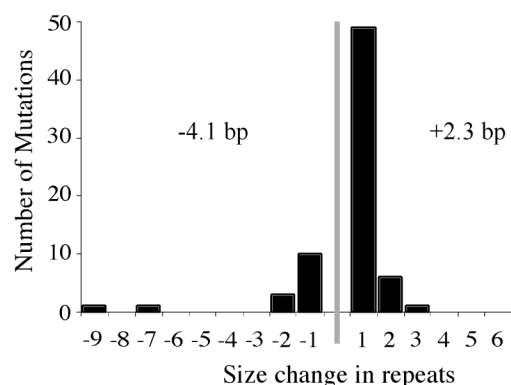


FIG. 2.—Distribution of the observed mutations classified by the change in size of the allele that they produce. The mean values for mutations that increase versus decrease allele size are shown.

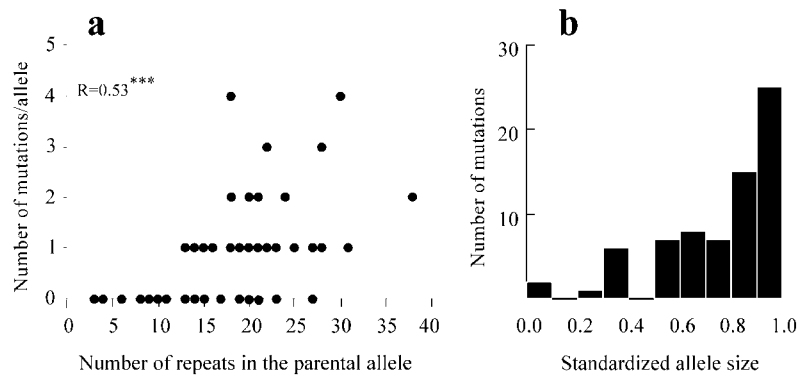


FIG. 3.—*a*, Scatter diagram for the number of repeats in the progenitor allele versus the number of observed mutations for that allele. The correlation is highly significant ( $R = 0.53$ ,  $P < 0.001$ ). *b*, Histogram showing the number of observed mutations over all loci grouped by the standardized allele size (see text). There is a significant excess of mutations among alleles with a standardized size greater than 0.5 (63 vs. 8;  $\chi^2 = 42.6$ ,  $P = 0.001$ ).

among these reported rates, including the average length of the microsatellite alleles, length of the repeat motif, base composition of the repeat motifs, and differences in the fidelity of DNA replication among organisms.

One clear factor governing the rate variation among loci within maize is the length of the repeat motif. We observed no mutations among loci with repeat motifs of more than 2 bp in length, as compared with 70 mutations among loci with dinucleotide repeats. This result is consistent with what has been seen in some other organisms, including *Drosophila* (Schug et al. 1998) and humans (Chakraborty et al. 1997). Nevertheless, there are counterexamples where the opposite relationship has been observed such that loci with dinucleotide repeat motifs mutate at a slower rate than those with larger repeat motifs (Weber and Wong 1993; Eckert and Yan 2000). These conflicting reports highlight the idiosyncratic nature of the mutational process at microsatellites and caution against applying results from one organism to another. It is also possible that the difference in the mutation rate between di-, tri-, and tetranucleotide microsatellites can be explained by a difference in the average number of repeats (Schug et al. 1998; Harr and Schlötterer 2000), with dinucleotide loci having a higher average number of repeats and thus a higher mutation rate. Because we have not determined the average length in size for all the studied loci, we cannot test this hypothesis in maize.

We can also ask whether the mutation rate that we calculated for each locus applies to “natural” populations of maize. If it does, the number of observed mutations at a locus should be positively correlated with the number of alleles and heterozygosity in maize. Matsuoka et al. (2002*b*) have investigated genetic diversity for the 98 loci used in experiment II among a sample of 193 maize plants. Heterozygosity and the number of alleles were estimated using this data for dinucleotide loci. We found that the number of mutations observed in our experiment is correlated with both heterozygosity (nonparametric Spearman correlation,  $R_s = 0.42$ ,  $P < 0.001$ ) and the number of alleles ( $R_s = 0.45$ ,  $P < 0.001$ ).

### Trends in the Mutational Process

In addition to estimating the mutation rate, our data revealed several features of the mutational process for microsatellites in maize. Similar to what has been observed in animals (Amos et al. 1996; Primmer et al. 1996), we found that mutations of a single repeat in length are far more common than mutations of multiple repeats. Of the mutations observed in dinucleotide microsatellites, 83% are one repeat, 13% are two repeats, and 4% are more than two repeats in length (table 4). We also observed heterogeneity in the rate among dinucleotide loci (fig. 1), and because our dinucleotide loci are mostly  $(AG)_n$  repeats, this heterogeneity cannot be explained by diversity in the sequence composition of the repeat motif (see Bachtrog et al. 2000). However, this heterogeneity may be partly attributed to variation in the number of repeats in the parental alleles because we observed a correlation between the number of repeats in the parental allele and the number of observed mutations (fig. 3*a*). Alleles with a greater number of repeats appear to be more mutable.

We have also observed that there is a higher probability to mutate to a larger allele than to a smaller one (fig. 2), with 56 of the 71 observed mutations from experiment II having caused an increase in the size of the allele. This same bias has been observed with other organisms (Amos et al. 1996; Ellegren 2000; Xu et al. 2000). Given this bias, previous authors have asked: why don't microsatellites increase infinitely in size? One possible explanation is that there is an equilibrium between mutations that alter the size of the microsatellite and base substitutions that lead to the degradation of the microsatellite (Kruglyak et al. 1998). In *Drosophila*, however, such an equilibrium process is inadequate to explain the underrepresentation of large microsatellites in the genome (Harr and Schlötterer 2000). Another nonexclusive possible explanation is that the larger the allele, the greater the probability that a mutation will cause a contraction in size. Our observation that mutations causing a decrease in size are on average larger than those that cause an increase is consistent with this

mechanism (fig. 2), partially explaining why microsatellites do not increase infinitely in size (also see Harr and Schlötterer 2000). However, we did not find a significant negative correlation between the standardized size of the progenitor allele and magnitude of the mutation as detected in humans (Ellegren 2000; Xu et al. 2000), although the largest contractions are associated with large alleles.

#### Short-Term Versus Long-Term Mutational Processes

In this study, we evaluated the short-term mutational pattern at microsatellite loci in maize in two different experiments of mutation accumulation. We determined the DNA sequence of the mutant and progenitor alleles for 55 of 73 new mutations. The sequence analysis revealed that all mutations were changes in the number of repeats in the microsatellite or in the length of mononucleotide tracts flanking the microsatellite. We did not observe any indels in the flanking regions, except for the aforementioned changes in mononucleotide tracts.

Contrary to our results, Matsuoka et al. (2002a) observed that microsatellite alleles among lines of maize and teosinte typically differ by indels of 2 to 50 bp (or larger) in the regions flanking the microsatellite repeat. Because we have observed no such mutations in our short-term evolutionary study, this class of indels likely arises only over longer evolutionary periods at a rate far below our estimated rate for dinucleotide microsatellites. Using the Poisson law, the 95% upper bound for the rate for such indels is  $2 \times 10^{-5}$ , given that we have examined a total of more than 153,000 allele-generations without observing any indels in the flanking sequences.

Microsatellites are assayed by screening for length polymorphisms in a DNA region between a pair of primers that flank the microsatellite repeat. Our results combined with those of Matsuoka et al. (2002a) indicate that the observed length polymorphism is the result of several processes that proceed at different rates. First, there can be changes in the number of repeats in the microsatellite that can proceed at an average rate of  $7.7 \times 10^{-4}$  mutations/generation for dinucleotide repeat loci. Second, there can be multistep mutations with the variance of change in the number of repeats among mutations ( $\sigma_m^2$ ) being 3.2. Third, there can be indels of 2 to 50 bp or more in the flanking regions that accumulate at rates below  $2 \times 10^{-5}$  mutations/generation. This mixed mutational pattern cautions against the casual use of models based on a simple stepwise mutation process with a single mutation rate. However, with knowledge of these complexities, it is possible to identify a subset of microsatellites that more closely follow a stepwise process and have a more uniform mutation rate.

#### Estimation of Effective Population Sizes

For estimation of the effective population size of maize, one needs a set of markers that behave in a stepwise manner and have a known mutation rate. Matsuoka et al. (2002b) have investigated genetic diversity for the

98 loci used in experiment II among a sample of 264 maize and teosinte plants. For that sample, 33 of the loci had allelic distributions with less than 10% non-stepwise alleles (table 1). Using the data from our study, we calculated the mutation rate for these 33 loci to be  $4.3 \times 10^{-4}$  and the variance of change in the size of the allele ( $\sigma_m^2$ ) to be 2.08. Using the Poisson distribution, we cannot reject the null hypothesis that the rate is homogeneous among these 33 loci ( $\chi^2 = 0.89$ ,  $P = 0.35$ ), and so this rate can be applied to all 33 loci.

The variance of allele size for this set of 33 microsatellite loci was calculated for 193 maize plants and 34 teosinte (*Z. mays* subsp. *parviglumis*) plants, using the data presented in Matsuoka et al. (2002b). For maize, the average variance of allele size was 23.5 repeats (range 0.88–79.5), and for subsp. *parviglumis*, it was 26.8 (range 0.80–85.1). With these data, we can calculate the effective population size, using a generalized stepwise model that allows for steps of more than one repeat in length (Slatkin 1995). Under this model, the variance in allele size ( $\sigma^2$ ), the effective population size ( $N$ ), the variance of change in the size of the allele ( $\sigma_m^2$ ), and the mutation rate ( $\mu$ ) are related by the following formula:

$$\sigma^2 = 2N\mu\sigma_m^2$$

Accordingly, we calculated that the effective population size for maize is 13,100, and for subsp. *parviglumis* it is 15,000. These estimates could be biased downward if there is a constraint on the variance of microsatellite size (Garza, Slatkin, and Freimer 1995).

One can also estimate effective population size using the equilibrium expectation for heterozygosity ( $H$ ) for microsatellite loci following a strict stepwise model (Kimura and Ohta 1975):

$$H = 1 - (1 + 8N\mu)^{-0.5}$$

Heterozygosity for the data presented in Matsuoka et al. (2002b) ranges from 0.53 to 0.96 for maize, and from 0.73 to 0.95 for subsp. *parviglumis*. The effective population size calculated as the average of the effective size given by each individual locus is 33,000 for maize and 38,500 for subsp. *parviglumis*. This estimate could be biased upwards since it does not incorporate the effect of multi-step mutations or biased downward if there is a constraint on the size of the microsatellites.

This effective population size for maize was previously calculated using polymorphism at *adh1* and estimated to be 660,000 (Gaut and Clegg 1993). In a similar study of *adh1*, Eyre-Walker et al. (1998) reported an estimate of the effective size of subsp. *parviglumis* at 940,000. Our estimates are more than an order of magnitude less than these. The *adh1*-based estimates assume a DNA substitution rate inferred from the amount of DNA sequence divergence accumulated over the 50 to 60-Myr history of the grass family. There is a concern about this rate because it represents a long-term rate over the history of the grasses and may not be appropriate to recent events in the maize lineage, where a lineage-specific rate acceleration can be anticipated because of a generation-time effect (White and Doebley 1999).



In another study, Remington et al. (2001) reported an effective population size of 200,000 for maize, based on the degree of linkage disequilibrium among maize inbred lines and the function  $C = 4Nc$ , where  $C$  is the population recombination parameter, and  $c$  is the recombination rate ([crossovers/bp]  $\times$  generation). However, this report assumes a recombination rate of  $10^{-8}$ , whereas reported rates in maize genes are nearer to  $10^{-7}$  (Patterson et al. 1995; Xu et al. 1995; Dooner and Martinez-Ferez 1997; Okagaki and Weil 1997). Using the latter rate would give a  $4Nc$ -based estimate of 20,000, a value much closer to the estimate based on microsatellites. The value for  $C$  reported by Remington et al. (2001) may also be inappropriate because it is based on a biased sample of the maize germ plasm pool. A value for  $C$  based on a more representative sample of maize is  $\approx 0.02$  (Tenaillon et al. 2001). Using this value of  $C$  and the observed recombination rate for maize genes ( $\approx 10^{-7}$ ), the effective population size for maize would be 50,000. The difficulty with these estimates is the uncertainty surrounding values for  $c$ . The cause of the differences in estimates of effective population size based on  $c$ , sequence polymorphism, and microsatellites will require further exploration.

This study has investigated the mutation rate and process for maize microsatellites. Because these markers are widely used in plants for a variety of purposes, such estimates of the mutation rate and knowledge of the mutation process are needed to clarify the origin and maintenance of genetic diversity at these loci. Whereas the mutation pattern of maize microsatellites is complex, a fuller understanding of these complexities will facilitate their application to a variety of questions in maize genetics and evolution.

### Acknowledgments

We thank Ben Burr for providing the RI lines, Steve Bull for technical assistance, and two anonymous reviewers for thoughtful comments on the manuscript. This work is supported by the U.S. NSF grant DBI-0096033 and by Pioneer Hi-Bred International. Research was completed at Department of Genetics, University of Wisconsin, Madison, Wisconsin, and Crop Genetics Research and Development, Pioneer Hi-Bred International, Johnston, Iowa.

### LITERATURE CITED

- AMOS, W., S. J. SAWCER, R. W. FEAKES, and D. C. RUBINSTEIN. 1996. Microsatellites show mutational bias and heterozygote instability. *Nat. Genet.* **13**:390–391.
- BACHTROG, D., M. AGIS, M. IMHOF, and C. SCHLÖTTERER. 2000. Microsatellite variability differs between dinucleotide repeat motifs—evidence from *Drosophila melanogaster*. *Mol. Biol. Evol.* **17**:1277–1285.
- BOWCOCK, A., J. RUIZ-LINARES, E. TOMFOHRDE, J. MINCH, J. KIDD, and L. CAVALLI-SFORZA. 1994. High resolution human evolutionary trees with polymorphic microsatellites. *Nature* **368**:455–457.
- BRINKMANN, B., M. KLINTSCHAR, F. NEUHUBER, J. HÜHNE, and B. ROLF. 1998. Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *Am. J. Hum. Genet.* **62**:1408–1415.
- BURR, B., F. A. BURR, K. H. THOMPSON, M. C. ALBERSTON, and C. W. STUBER. 1988. Gene mapping with recombinant inbreds in maize. *Genetics* **118**:519–526.
- CHAKRABORTY, R., M. KIMMEL, D. STIVERS, L. DAVISON, and R. DEKA. 1997. Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. *Proc. Natl. Acad. Sci. USA* **94**:1041–1046.
- DALLAS, J. F. 1992. Estimation of the mutation rates in recombinant inbred strains of mouse. *Mamm. Genome* **3**:452–456.
- DIETRICH, W., H. KATZ, S. LINCOLN, H. SHIN, J. FRIEDMAN, N. DRACOPOLI, and E. LANDER. 1992. A genetic map of the mouse suitable for typing intraspecific crosses. *Genetics* **131**:423–427.
- DI RIENZO, A., P. DONNELLY, C. TOOMAJIAN, B. SISK, A. HILL, M. L. PETZL-ERLE, G. H. HAINES, and D. H. BARCH. 1998. Heterogeneity of microsatellite mutations within and between loci, and implication for human demographic histories. *Genetics* **148**:1269–1284.
- DOONER, H., and I. MARTINEZ-FEREZ. 1997. Recombination occurs uniformly within the *bronze* gene, a meiotic recombination hotspot in the maize genome. *Plant Cell* **9**:1633–1646.
- ECKERT, K. A., and G. YAN. 2000. Mutational analyses of dinucleotide and tetranucleotide microsatellites in *Escherichia coli*: influence of sequence on expansion mutagenesis. *Nucleic Acids Res.* **28**:2831–2838.
- ELLEGREN, H. 2000. Heterogenous mutation processes in human microsatellite DNA sequences. *Nat. Genet.* **24**:400–402.
- EYRE-WALKER, A., R. S. GAUT, H. HILTON, D. L. FELDMAN, and B. S. GAUT. 1998. Investigating the bottleneck leading to the domestication of maize. *Proc. Natl. Acad. Sci. USA* **95**:4441–4446.
- GARZA, J. G., M. SLATKIN, and N. B. FREIMER. 1995. Microsatellite allele frequencies in human and chimpanzees, with implications for constraints on allele size. *Mol. Biol. Evol.* **12**:594–603.
- 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.
- GOLDSTEIN, D. B., A. R. LINARES, L. L. CAVALLI-SFORZA, and M. W. FELDMAN. 1995. Genetic absolute dating based on microsatellites and the origin of modern human. *Proc. Natl. Acad. Sci. USA* **92**:11549–11552.
- GOLDSTEIN, D. B., G. ROEMER, D. SMITH, D. REICH, A. BERGMAN, and R. WAYNE. 1999. The use of microsatellite variation to infer population structure and demographic history in a natural model system. *Genetics* **151**:797–801.
- HARR, B., and C. SCHLÖTTERER. 2000. Long microsatellite alleles in *Drosophila melanogaster* have a downward mutation bias and short persistence times, which cause their genome-wide under representation. *Genetics* **155**:1231–1220.
- HARR, B., B. ZANGERL, G. BREM, and C. SCHLÖTTERER. 1998. Conservation of locus-specific microsatellite variability across species: a comparison of two *Drosophila* sibling species, *D. melanogaster* and *D. simulans*. *Mol. Biol. Evol.* **15**:176–184.
- JONES, A. G., G. ROSENQVIST, A. BERGLUND, and J. C. AVISE. 1999. Clustered microsatellite mutations in the pipefish *Syngnathus typhle*. *Genetics* **152**:1057–1063.
- KIMURA, M., and T. OHTA. 1975. Distribution of allelic frequencies in a finite population under stepwise production of neutral alleles. *Proc. Natl. Acad. Sci. USA* **72**:2761–2764.

- KOVALCHUK, O., Y. E. DUBROVA, A. ARKHIPOV, B. HOHN, and I. KOVALCHUK. 2000. Wheat mutation rate after Chernobyl. *Nature* **407**:583–584.
- KRUGLYAK, S., R. T. DURRETT, M. D. SCHUG, and C. F. AQUADRO. 1998. Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. *Proc. Natl. Acad. Sci. USA* **95**:10774–10778.
- LAGERCRANTZ, U., H. ELLEGREN, and L. ANDERSSON. 1993. The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. *Nucleic Acids Res.* **21**:1111–1115.
- 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, J. SANCHEZ, M. GOODMAN, E. S. BUCKLER IV, and J. DOEBLEY. 2002b. Maize domestication and diversification inferred from microsatellite DNA. *Proc. Natl. Acad. Sci. USA* **99**:6080–6084.
- OKAGAKI, R., and C. WEIL. 1997. Analysis of recombination sites within the maize *waxy* locus. *Genetics* **147**:815–821.
- PATTERSON, G., K. KUBO, T. SHROYER, and V. CHANDLER. 1995. Sequences required for paramutation of the maize *b* gene map to a region containing the promoter and upstream sequences. *Genetics* **140**:1389–1406.
- PETERS, J. M., D. C. QUELLER, J. E. STRASSMANN, and C. R. SOLIS. 1995. Maternity assignment and queen replacement in a social wasp. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **260**:7–12.
- POWELL, W., M. MORGANTE, C. ANDRE, M. HANAFEY, J. VOGEL, S. TINGEY, and A. RAFALSKI. 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Mol. Breed.* **2**:225–238.
- PRIMMER, C. R., H. ELLEGREN, N. SAINO, and A. P. MØLLER. 1996. Directional evolution in germline microsatellite mutations. *Nat. Genet.* **13**:391–393.
- REMINGTON, D., J. THORNBERRY, Y. MATSUOKA, L. WILSON, S. WHITT, J. DOEBLEY, S. KRESOVICH, M. GOODMAN, and E. S. BUCKLER. 2001. Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc. Natl. Acad. Sci. USA* **98**:11479–11484.
- SCHLÖTTERER, 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.
- SCHUG, M. D., C. M. HUTTER, K. A. WETTERSTRAND, M. S. GAUDETTE, T. F. C. MACKAY, and C. F. AQUADRO. 1998. The mutation rates of di-, tri-, tetranucleotide repeats in *Drosophila melanogaster*. *Mol. Biol. Evol.* **15**:1751–1760.
- 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.
- SLATKIN, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**:457–462.
- SMITH, J. S. C., E. C. L. CHIN, H. SHU, O. S. SMITH, S. J. WALL, M. L. SENIOR, S. E. MITCHELL, S. KRESOVICH, and J. ZIEGLE. 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.
- TENAILLON, M., M. SAWKINS, A. LONG, R. GAUT, J. DOEBLEY, and B. GAUT. 2001. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays*). *Proc. Natl. Acad. Sci. USA* **98**:9161–9166.
- UDUPA, S. M., and M. BAUM. 2001. High mutation rate and mutational bias at (TAA)<sub>n</sub> microsatellite loci in chickpea (*Cicer arietinum* L.). *Mol. Genet. Genomics* **265**:1097–1103.
- VAZQUEZ, F., T. PEREZ, J. ALBORNOZ, and A. DOMINGUEZ. 2000. Estimation of the mutation rates in *Drosophila melanogaster*. *Genet. Res.* **76**:323–326.
- WEBER, J. L., and C. WONG. 1993. Mutation of human short tandem repeats. *Hum. Mol. Genet.* **2**:1123–1128.
- WEHRHAHN, C. F. 1975. The evolution of selectively similar electrophoretically detectable alleles in finite natural populations. *Genetics* **80**:375–394.
- WHITE, S., and J. DOEBLEY. 1999. The molecular evolution of *terminal ear1*, a regulatory gene in the genus *Zea*. *Genetics* **153**:1455–1462.
- XU, X., A. HSIA, L. ZHANG, B. NIKOLAU, and P. SCHNABLE. 1995. Meiotic recombination break points resolve at high rates at the 5' end of a maize coding sequence. *Plant Cell* **7**:2151–2161.
- XU, X., M. PENG, Z. FANG, and X. XU. 2000. The direction of microsatellite mutations is dependent upon the allele length. *Nat. Genet.* **24**:396–399.
- ZHIVOTOVSKY, L. A., and M. W. FELDMAN. 1995. Microsatellite variability and genetic distance. *Proc. Natl. Acad. Sci. USA* **92**:11549–11552.

BRIAN GOLDING, reviewing editor

Accepted March 5, 2002