Directional Evolution for Microsatellite Size in Maize

Yves Vigouroux, Yoshihiro Matsuoka,† and John Doebley**

*Department of Genetics, University of Wisconsin, Madison; and †Fukui Prefectural University, Matsuoka-Cho, Yoshida-gun, Fukui, Japan

Directional evolution in microsatellites is the tendency for microsatellites either to increase or to decrease in size over time between populations. We analyzed 99 microsatellite loci in a sample of 193 maize plants representing the entire pre-Columbian range of this crop for evidence of directional evolution. We took advantage of the known phylogeographic history of maize with the independent movement of maize from its ancestral location in Mexico to North and South America. We show that there is an increase in the average allele size of microsatellites in the geographically derived North and South American groups relative to the ancestral Mexican group. We also show that there is a negative correlation between average allele size and altitude in all three groups. Directional evolution in maize microsatellites can be explained by changes in the mutation rate over time and space, changes in the degree of mutational bias to a larger allele, demographic differences between the ancestral and geographically derived populations, and/or scenarios involving selection on microsatellite size. The occurrence of directional evolution for microsatellite size indicates that the estimation of population parameters with microsatellite data in maize should be done with caution.

Introduction

Microsatellites are powerful genetic makers that have seen broad application in genetics and evolutionary biology (Jarne and Lagoda 1996). As the understanding of microsatellites has grown, however, it has become increasingly clear that they have a complex mutational process that can generate a variety of biases in microsatellite data. For example, the mutation rate can be heterogeneous among loci (Di Rienzo et al. 1998; Schlötterer et al. 1998), and microsatellite loci may not faithfully follow a simple (or even generalized) mutation model (Colson and Goldstein 1999; Matsuoka et al. 2002*a*). These problems can complicate the application of microsatellites in evolution and population genetics when assumptions of the underlying model of mutation are violated.

One well-documented type of bias in microsatellite mutation is the tendency for new mutations to cause an increase in the size of an allele. This phenomenon has been documented in both plants (Udupa and Baum 2001; Vigouroux et al. 2002) and animals (Amos et al. 1996; Primmer et al. 1996; Cooper et al. 1999). This mutational bias and a differential mutation rate could explain the observed increase in the average size of microsatellites (directional evolution) between humans and nonhuman primates (Rubinsztein et al. 1995). However, this report of directional evolution has been questioned (Amos et al. 1996; Ellegren, Primmer, and Sheldon 1995) because the apparent directional evolution could be an artifact of ascertainment bias during microsatellite discovery (Hutter, Schug, and Aquadro 1998). A more recent study has shown evidence for directional evolution even when ascertainment bias is taken into account (Amos et al. 2003).

In this study, we take advantage of the known phylogeographic history of maize to ask whether maize microsatellites have experienced directional evolution in size. We analyze the evolution of microsatellite size between groups separated by fewer than 10,000 generations. We report both a directional increase in microsatellite size in geographically derived groups and a negative correlation of allele size and altitude that arose independently in North and South America. We discuss possible mechanisms that could generate these patterns and the implication of these biases for the application of microsatellite data to questions surrounding maize evolution and population genetics.

Key words: simple sequence repeat, SSR, genome size, *Zea*, mutation rate.

E-mail: jdoebley@wisc.edu.

Mol. Biol. Evol. 20(9):1480–1483. 2003
DOI: 10.1093/molbev/msg156
Molecular Biology and Evolution, Vol. 20, No. 9,

Microsatellite Data

We have previously analyzed a data set of 193 pre-Columbian maize landraces, genotyped at 99 microsatellite loci (Matsuoka et al. 2002*b*). The data are available online at either http://www.wisc.edu/teosinte or http://statgen.ncsu.edu/panzea/. For each locus, we calculated the average allele size and its standard deviation, and we calculated the standardize size of the alleles for each plant at the locus as (actual size minus the mean)/the standard deviation. Then, for each of the 193 maize plants, we calculated the *average individual size* of its microsatellites as the mean of the standardized size of 99 microsatellite loci. This standardization makes each locus contribute equally to the average individual size.

Maize was domesticated about 7,500 years ago in Mexico, and then spread to North and South America (Matsuoka et al. 2002*b*). A phylogenetic study has shown that North and South American maize are independently derived from the ancestral population in Mexico (Matsuoka et al. 2002*b*). Knowing this phylogeographic structure of maize enables us to ask whether directional evolution in maize microsatellites has occurred. For some of the analysis, we used the phylogeographic data as the basis for dividing our sample between 69 South American plants (SA), 71 Mexican and Guatemalan plants (ME), and 46 United States and Canadian plants (NA). Seven Caribbean plants were not classified in any of these three groups.

Statistical tests were performed using the software SYSTAT (Systat, Inc.).

Results

To test whether directional evolution of microsatellite size occurs in maize, we compared average allele size in

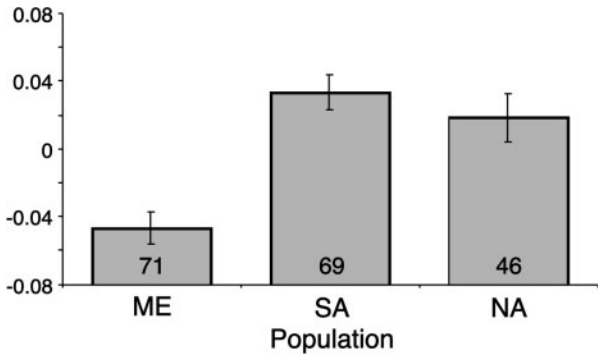

FIG. 1.—Average individual microsatellite allele size for the three maize populations: Mexican (ME), South American (SA), and North American (NA). The mean, the standard error, and the number of plants per population are presented.

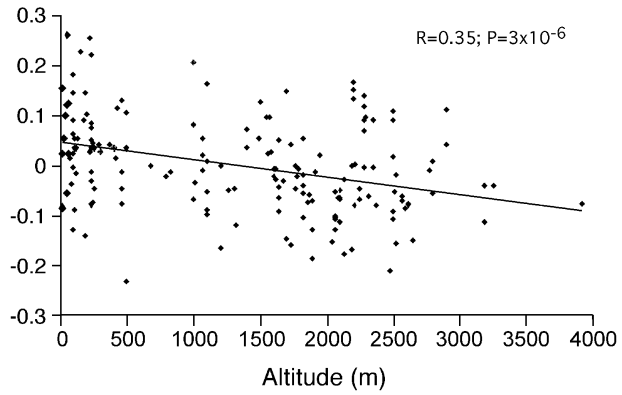

FIG. 2.—Correlation between average individual microsatellite allele size and altitude.

the geographically derived groups (NA and SA) to the ancestral group (ME). The average individual size is higher for the NA ($t = 5.65$, $P < 0.0001$) and SA groups ($t = 3.74$, $P < 0.0003$) than for the ancestral ME group (fig. 1). No difference was detected between the NA and SA groups ($t = 0.82$, $P = 0.41$). These data indicate an increase in allele size in the geographically derived NA and SA groups relative to the ancestral ME group.

Because genome size has been reported to be negatively correlated with altitude in maize (Poggio et al. 1998), and because microsatellite loci are one component of genome size, we examined whether average individual allele size is correlated with altitude and found that it was (fig. 2; $R = 0.35$, $P = 3 \times 10^{-6}$). We found the same correlation if we considered the three groups separately: ME ($R = 0.43$, $P < 0.0002$), SA ($R = 0.30$, $P < 0.016$), and NA ($R = 0.54$, $P < 0.001$). The SA group was derived from low-altitude maize of Guatemala, and the NA group was derived from maize of northern Mexico (Matsuoka et al. 2002*b*). These independent histories argue that the correlations were established independently in ME, SA, and NA.

Next we asked whether the differences between average individual size among groups could result from a difference in the mean altitude for the groups. The average altitude is 1,430 m for the ME group, 1,480 m for the SA group, and 1,030 m for the NA group. First, we determined the regression slopes of altitude onto allele size for each group and found that they are all similar ($F = 2.41$, df = 2, $P = 0.093$). Knowing that the relationship between the average individual size and altitude is similar for all groups, we then asked whether the intercepts are significantly different between groups and found that they are ($F = 19.9$, df = 2, $P < 0.001$). Based on all the tests, we conclude that there are two distinct phenomena: (1) a significant difference between groups in average individual size and (2) a significant correlation between average individual size and altitude.

Discussion

The observation that the average individual allele size for microsatellite loci is larger in North and South American maize than in the ancestral population in Mexico supports directional evolution of microsatellite size in maize. Because the microsatellites were obtained by screening North American maize, one could argue that ascertainment bias resulted in a larger average size in North America (Ellegren, Primmer, and Sheldon 1995). However, one would not expect to observe the same phenomenon for the South American sample. Thus, our data suggest that directional evolution of microsatellite size occurs in maize. The mean difference between the SA and ME, and the NA and ME populations is 4.1 bp and 3.3 bp per locus, respectively. Therefore, we conclude that maize microsatellite loci have not remained at equilibrium for size but have tended to increase in size from the ''ancestral'' to the geographically derived populations.

Directional evolution of this type could be explained in several ways. First, there could be a change in the degree of mutational bias to larger alleles in the derived groups. In our case, this would have had to occur twice, independently in North and South America. Second, given that mutations are more likely to cause an increase in allele size in maize (Vigouroux et al. 2002), a change in the mutation rate with movement into a new environment could cause directional evolution (Rubinsztein et al. 1995). Third, one could also propose a demographic explanation. For example, if the ancestral population is stable in size and at Hardy-Weinberg equilibrium, then the coalescence tree for a sample of alleles could be shorter than the coalescence tree for a similar sample in an expanding, nonequilibrium, derived population. The longer tree implies more opportunity for mutation, and given that mutation tends to increase allele size in maize, the result would be a larger average allele size in the derived population. This explanation implies that the formation of the derived population was not associated with a severe bottleneck. Other demographic scenarios could give the opposite outcome, i.e., a shorter coalescence tree for a derived population.

We have also observed a negative correlation between allele size and altitude in Mexico and North and South America that must have been independently derived given the known phylogeography of maize (Matsuoka et al. 2002*b*). Moreover, the strength of the correlation is similar in all three regions. There is an average decrease of 1.8 bp

per locus per thousand meters elevation. This relationship represents a form of directional evolution. Accordingly, one can propose a set of explanations similar to those stated above. For example, at higher elevation, maize has a shorter generation time (fewer cell divisions). This is expected to result in a reduced mutation rate per generation. Given that mutations tend to increase the size of microsatellites, a lower mutation rate at high elevation is expected to yield a smaller average allele size.

In addition, there is a known negative correlation between genome size and altitude in maize (Poggio et al. 1998). Rayburn et al. (1985) have proposed that this correlation is due to selection for a smaller genome in short-seasoned environments, because a large genome would take longer to replicate at each cell division. To be effective for microsatellites, such selection should affect thousands of loci. With an average estimate of 58.2 dinucleotide microsatellites per Mbp (Morgante, Hanafey, and Powell 2002), a variance of 1.8 bp per locus per thousand meters corresponds to only 0.01% of the maize genome. This value is quite small compared to the genome size variation in heterochromatin of 36% for maize varieties (Poggio et al. 1998). Because variation in microsatellite size contributes very little to genome size variation in maize, it is difficult to imagine that selection is the driving force, although one cannot exclude the possibility that small differences may come under selection (Hughes and Hughes 1995).

A final point is that the occurrence of directional evolution for maize microsatellites cautions against using microsatellites nonchalantly for the estimation of population parameters. For example, the divergence time between maize populations can be estimated using the difference in the average allele size between populations (Wehrhahn 1975; Goldstein et al. 1995). However, directionality between populations in the evolution of microsatellite size will cause an upwardly biased estimate of the mean divergence between populations. Similarly, because of the relationship between mutation rate and the number of repeats, a mutation rate estimated using a population with high average size (e.g., North American maize) would likely give an overestimation of the mutation rate in maize (Amos et al. 1996). Accordingly, we urge caution when using microsatellite data for the estimation of population parameters in maize with simple mutational models. Because we were aware of the problem of directional evolution of maize microsatellites in geographically derived populations, we restricted our prior estimate of the maize-teosinte divergence time (Matsuoka et al. 2002*b*) to a comparison within a single environment (Mexico). Nevertheless, that estimate needs to be viewed with caution because the dynamics of microsatellite evolution in maize are not yet fully understood.

Acknowledgments

This work was supported by U.S. National Science Foundation grant DBI-0096033. We thank C. Schlötterer and an anonymous reviewer for comments on the manuscript and C. Aquadro for sharing unpublished work from his laboratory.

Literature Cited

Amos, W., C. M. Hutter, M. D. Schug, and C. F. Aquadro. 2003. Directional evolution of size coupled with ascertainment bias for variation in Drosophila microsatellites. Mol. Biol. Evol. **20**:660–662.

Amos, W., S. J. Sawcer, R. W. Feakes, and D. C. Rubinsztein. 1996. Microsatellites show mutational bias and heterozygote instability. Nat. Genet. **13**:390–391.

Colson, I., and D. B. Goldstein. 1999. Evidence of complex mutations at microsatellite loci in Drosophila. Genetics **152**:617–627.

Cooper, G., N. J. Burroughs, D. A. Rand, D. C. Rubinsztein, and W. Amos. 1999. Markov chain Monte Carlo analysis of human Y-chromosome microsatellites provides evidence of biased mutation. Proc. Natl. Acad. Sci. USA **96**:11916–11921.

Di Rienzo, A., P. Donnelly, C. Toomajian, B. Sisk, A. Hill, M. L. Petzl-Erler, G. K. Haines, and D. H. Barch. 1998. Heterogeneity of microsatellite mutations within and between loci, and implications for human demographic histories. Genetics **148**:1269–1284.

Ellegren, H., C. R. Primmer, and B. C. Sheldon. 1995. Microsatellite ''evolution'': directionality or bias? Nat. Genet. **11**:360–362.

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 humans. Proc. Natl. Acad. Sci. USA **92**:6723–6727.

Hughes, A. L., and M. K. Hughes. 1995. Small genomes for better flyers. Nature **377**:391.

Hutter, C. M., M. D. Schug, and C. F. Aquadro. 1998. Microsatellite variation in *Drosophila melanogaster* and *Drosophila simulans*: a reciprocal test of the ascertainment bias hypothesis. Mol. Biol. Evol. **15**:1620–1636.

Jarne, P., and P. J. L. Lagoda. 1996. Microsatellites, from molecules to populations and back. Trends Ecol. Evol. **11**:424–429.

Matsuoka, Y., S. E. Mitchell, S. Kresovich, M. M. Goodman, and J. Doebley. 2002*a*. Microsatellites in Zea—variability, patterns of mutations, and use for evolutionary studies. Theor. Appl. Genet. **104**:436–450.

Matsuoka, Y., Y. Vigouroux, M. M. Goodman, J. G. Sanchez, E. Buckler, and J. Doebley. 2002*b*. A single domestication for maize shown by multilocus microsatellite genotyping. Proc. Natl. Acad. Sci. USA **99**:6080–6084.

Morgante, M., M. Hanafey, and W. Powell. 2002. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nat. Genet. **30**:194–200.

Poggio, L., M. Rosato, A. M. Chiavarino, and C.A. Naranjo. 1998. Genome size and environmental correlations in maize (*Zea mays* ssp. *mays*, Poaceae). Ann. Bot. **82**(Suppl. A):107–115.

Primmer, C. R., H. Ellegren, N. Saino, and A. P. Moller. 1996. Directional evolution in germline microsatellite mutations. Nat. Genet. **13**:391–393.

Rayburn, A. L., H. J. Price, J. D. Smith, and J. R. Gold. 1985. C-band heterochromatin and DNA content in *Zea mays*. Am. J. Bot. **72**:1610–1617.

Rubinsztein, D. C., W. Amos, J. Leggo, S. Goodburn, S. Jain, S.-H. Li, R. L. Margolis, C. A. Ross, and M. A. Fergusson-Smith. 1995. Microsatellite evolution—evidence for directionality and variation in rate between species. Nat. Genet. **10**:337–343.

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.

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

Vigouroux, Y., J. S. Jaqueth, Y. Matsuoka, O. S. Smith, W. D. Beavis, J. S. C. Smith, and J. Doebley. 2002. Rate and pattern of mutation at microsatellite loci in maize. Mol. Biol. Evol. **19**:1251–1260.

Wehrhahn, C. F. 1975. The evolution of selectively similar electrophoretically detectable alleles in finite natural populations. Genetics **80**:375–394.

Brian Golding, Associate Editor

Accepted April 24, 2003