DOI: 10.1534/genetics.104.031393

Genetic Diversity and Population Structure of Teosinte

Kenji Fukunaga,*[,1] Jason Hill,[†] Yves Vigouroux,*[,2] Yoshihiro Matsuoka,*[,3] Jesus Sanchez G.,[‡] Kejun Liu,[§,4] Edward S. Buckler and John Doebley*[,5]**

**Laboratory of Genetics, University of Wisconsin, Madison, Wisconsin 53706, [†]Department of Plant Biology, University of Minnesota, St. Paul, Minnesota 55108, [‡]Centro Universitario de Ciencias Biologicas y Agropecuarias, Universidad de Guadalajara, Zapopan, Jalisco, Mexico CP45110, [§]Statistics Department, North Carolina State University, Raleigh, North Carolina 27695 and **USDA-ARS, Institute for Genomic Diversity, Cornell University, Ithaca, New York 14853*

Manuscript received May 19, 2004
Accepted for publication January 3, 2005

ABSTRACT

The teosintes, the closest wild relatives of maize, are important resources for the study of maize genetics and evolution and for plant breeding. We genotyped 237 individual teosinte plants for 93 microsatellites. Phylogenetic relationships among species and subspecific taxa were largely consistent with prior analyses for other types of molecular markers. Plants of all species formed monophyletic clades, although relationships among species were not fully resolved. Phylogenetic analysis indicated that the Mexican annual teosintes divide into two clusters that largely correspond to the previously defined subspecies, *Z. mays* ssp. *parviglumis* and ssp. *mexicana*, although there are a few samples that represent either evolutionary intermediates or hybrids between these two subspecies. The Mexican annual teosintes show genetic substructuring along geographic lines. Hybridization or introgression between some teosintes and maize occurs at a low level and appears most common with *Z. mays* ssp. *mexicana*. Phylogeographic and phylogenetic analyses of the Mexican annual teosintes indicated that ssp. *parviglumis* diversified in the eastern part of its distribution and spread from east to west and that ssp. *mexicana* diversified in the Central Plateau of Mexico and spread along multiple paths to the north and east. We defined core sets of collections of *Z. mays* ssp. *mexicana* and ssp. *parviglumis* that attempt to capture the maximum number of microsatellite alleles for given sample sizes.

TEOSINTE is a wild grass native to Mexico and Central America (Figure 1) and the closest wild relative of cultivated maize (*Zea mays* ssp. *mays* L.). Teosinte represents an important resource for the study of maize genetics (Evans and Kermicle 2001), quantitative genetics (Lukens and Doebley 1999), molecular population genetics (Gaut *et al.* 2000), genome evolution (Sanz-Alferez *et al.* 2003), and crop evolution (Doebley 1990a). Notably, teosinte has become one of the best-characterized systems for plant molecular population genetics, including studies utilizing DNA samples recovered from archeological specimens (Jaenicke-Després *et al.* 2003). The teosintes also represent an important potential resource for maize breeding, although they have not yet been extensively used in this capacity. Given the breadth of use of teosinte in genetic analyses, a refined understanding of its phylogenetics and population structure can help guide further research in all of these areas.

Together, teosinte and maize compose the genus Zea, which has four species (Figure 1): (1) *Z. luxurians* (Durieu and Ascherson) Bird, an annual teosinte from Central America; (2) *Z. diploperennis* Iltis, Doebley and Guzman, a diploid perennial teosinte from Jalisco, Mexico; (3) *Z. perennis* (Hitchc.) Reeves and Mangelsdorf, a tetraploid perennial teosinte from Jalisco, Mexico; and (4) *Z. mays*, a polytypic annual species that includes four subspecies. The four subspecies are (1) ssp. *mays* (maize); (2) ssp. *mexicana* (Schrader) Iltis, a large-spikeleted teosinte adapted to the drier high elevations (~1600–2700 m) of northern and central Mexico; (3) ssp. *parviglumis* Iltis and Doebley, a small-spikeleted teosinte adapted to the moister middle elevation (~400–1800 m) of southwestern Mexico; and (4) ssp. *huehuetenangensis* (Iltis and Doebley) Doebley, an annual teosinte found only in the province of Huehuetenango in western Guatemala (Doebley 1990b). The four species of *Z. mays* have been placed into two sections: section Zea, which contains only *Z. mays*, and section Luxuriantes, which is composed of the other three species. Most of these teosinte species and subspecies have narrow geographic distributions consisting of only a few local populations; however, ssp.

[1]*Present address:* The International Research Center for Japanese Studies, 3-2 Oeyama-cho, Goryo, Nishikyo-ku, Kyoto 610-1192, Japan.

[2]*Present address:* Institut de Recherche pour le Developpement, Montpellier 34730, France.

[3]*Present address:* Fukui Prefectural University, Matsuoka-cho, Yoshida-gun, Fukui 910-1195, Japan.

[4]*Present address:* National Institute of Statistical Sciences, 19 T. W. Alexander Dr., Research Triangle Park, NC 27709-4006.

[5]*Corresponding author:* Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706. E-mail: jdoebley@wisc.edu

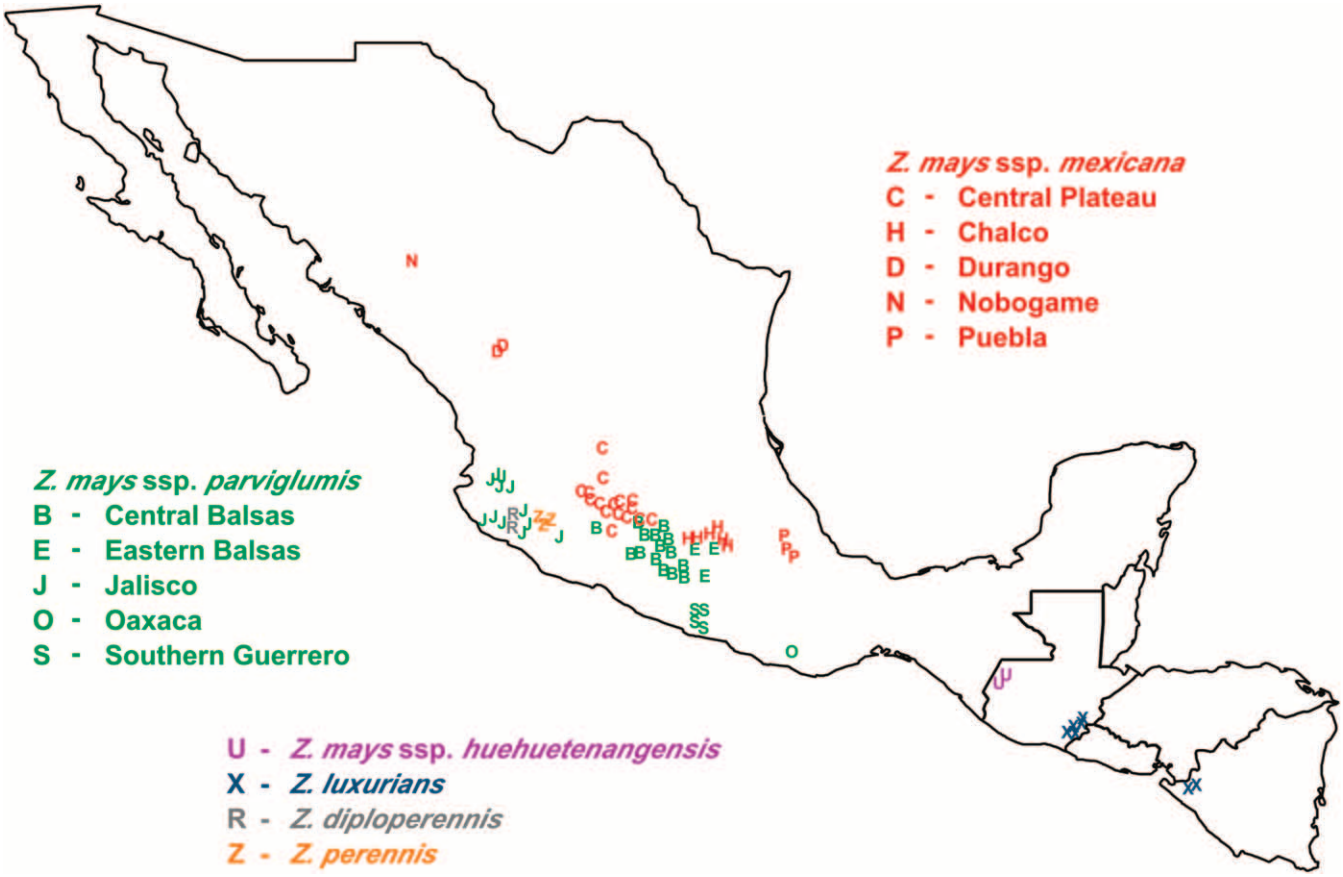


FIGURE 1.—Geographical distribution of the teosinte populations used in this study. Since many accessions come from geographically very close locations, their symbols overlap on the map.

mexicana and ssp. *parviglumis* are exceptions, being widely distributed in Mexico (Figure 1). Recently, ILTIS and BENZ (2000) classified *Z. luxurians* from Nicaragua as a new species, *Z. nicaraguensis* Iltis and Benz. Here, we treat it as one geographical group of *Z. luxurians*.

To clarify the phylogeny and population structure of the teosintes, we used a set of 93 microsatellite or simple sequence repeat (SSR) loci and a sample of 237 individual teosinte plants that cover the entire geographical distribution of the teosintes. We addressed four questions: (1) How are Zea taxa related to each other?, (2) How did the annual teosintes diversify in Mexico?, (3) How is genetic diversity structured in the Mexican annual teosintes?, and (4) Has introgression among species or subspecies played a role in teosinte evolution? We also define core sets of teosinte accessions that best capture the diversity of the teosintes.

MATERIALS AND METHODS

Plant materials: We sampled 237 teosinte plants from 172 accessions representing the entire geographical distribution of teosinte from northern Mexico to western Nicaragua. For each accession, 1–5 individuals were assayed. The sample includes 93 *Z. mays* ssp. *mexicana* individuals (69 accessions), 114 *Z. mays* ssp. *parviglumis* (82 accessions), 7 *Z. mays* ssp. *huehuetenangensis* (3 accessions), 13 *Z. luxurians* (10 accessions), 6 *Z. diploperennis* (5 accessions), and 4 *Z. perennis* (3 accessions) (Figure 1, Table 1). WILKES (1967) divided *Z. mays* ssp. *mexicana* into three races: Central Plateau, Chalco, and Nobogame. Because many new populations have been discovered since Wilkes' seminal work (SANCHEZ *et al.* 1998), we divided *Z. mays* ssp. *mexicana* into five geographical groups: Central Plateau, Chalco, Durango, Nobogame, and Puebla. Further extending Wilkes' analysis, we divided *Z. mays* ssp. *parviglumis* into two races, Balsas and Jalisco, and into five geographical groups, eastern Balsas, Central Balsas, Jalisco, Oaxaca, and southern Guerrero (Table 1, Figure 1). Two individual plants of the genus Tripsacum (one individual each of *T. zopilotense* and *T. peruvianum*) were used as the outgroup in phylogenetic analyses. See supplementary materials at http://genetics.org/supplemental/ for the complete passport data for the plants, including germplasm bank accession numbers and geographical coordinates.

SSR genotypes: Ninety-three SSRs that are evenly distributed throughout the genome were used to genotype all 237 Zea plants and the two Tripsacum individuals. These SSRs were used in the previous analysis of maize and its wild progenitor (MATSUOKA *et al.* 2002b). The plants were genotyped at Celera AgGen (Davis, CA) following procedures published elsewhere (MATSUOKA *et al.* 2002b). See supplementary materials at http://www.genetics.org/supplemental/ for a list of the SSRs, their repeat type, and their genomic locations.

Diversity analyses: Basic statistics, including the number of alleles, observed heterozygosity, gene diversity (or expected heterozygosity), and the number of taxon-specific (private) alleles, were calculated for species, subspecies, and races using PowerMarker (LIU 2002). In these analyses, individual plants of possible hybrid origin as determined by population structure analysis (see below) were excluded.

Phylogenetic trees: We used the FITCH program in the

PHYLIP package (Felsenstein 1993) with the log-transformed proportion-of-shared-alleles distance as implemented in the computer program Microsat (http://hpgl.stanford.edu/projects/microsat/). In FITCH, the J option was used to randomize the input order of samples. We constructed three types of trees: (1) a tree with all individual plants, (2) a tree with individual plants pooled into operational taxonomic units (OTUs) on the basis of their geographic origin and position in the individual tree, and (3) trees rooted with *T. zopilotense* and *T. peruvianum*. Data were pooled into 76 OTUs of two to four individual plants (or five plants in the case of *Z. diploperennis*). Membership of the plants in the 76 OTUs was based on their geographic origin and position in Figure 2 (see supplemental materials at http://genetics.org/supplemental/). Pooling was necessary to reduce the number of entries in the FITCH analysis to a level at which bootstrap replicates could be analyzed. Data for the two Tripsacum species were pooled to create a single synthetic outgroup. This was done because of the high frequency of null phenotypes ("alleles") for the SSRs in the single Tripsacum samples. Once pooled, the synthetic Tripsacum outgroup possessed visible alleles at 61 SSRs and null phenotypes at the other 32 SSRs. In the trees rooted with Tripsacum, we scored null phenotypes in two ways because of the large number of null phenotypes in Tripsacum. First, all null phenotypes at a locus were scored as the same null allele "0" in both Zea and Tripsacum. Second, null phenotypes in teosintes were scored as the allele "0" but as a distinct null allele "00" in Tripsacum, on the basis of the assumption that these alleles had independent origins. Given the high frequency of nulls in Tripsacum (50% nulls in each individual plant), it is highly probable that nulls in Tripsacum are independent of nulls in Zea. To determine the degree of statistical support for different branch points, we evaluated 1000 trees constructed from bootstrap resamplings of the data.

Analysis of phylogeography: We tested geographic (clinal) *vs.* dispersal models for the current distribution of populations of *Z. mays* ssp. *mexicana* and ssp. *parviglumis*. In the geographic model, migration occurs on a small scale and populations are isolated by distance. In this case, genetic distance should be correlated with geographical distance between populations. In dispersal models, long distance migration is allowed, and two geographically close populations may have been derived from distinct distant founder populations. In this case, genetic distance will be less correlated with geographical distance than with distance along the dispersal routes (dispersal distance). To test which scenario best fits the distribution of ssp. *mexicana* and ssp. *parviglumis*, we compared the geographical (clinal) and multiple dispersal distance matrices to the genetic distance between the populations. The Nobogame and Durango populations, which are geographically very distant, were removed from the analysis to prevent these geographic outliers from having undue weight on the analysis.

Various dispersal hypotheses were constructed using the Phylogeographer 1.0 (Buckler 1999). We placed the nodes A–F on the map (Figure 4). These nodes correspond to teosinte population centers as follows: A, eastern Balsas region; B, Central Plateau; C, Valley of Mexico; D, Balsas river drainage; E, southern Guerrero; and F, Jalisco. Each of these centers is ecologically relatively homogeneous with respect to altitude, length of growing season, and annual rainfall, and teosinte populations within each center share morphological similarity (Wilkes 1967; Sanchez *et al.* 1998). Populations within these seven centers were connected to their respective nodes (A–F) as shown in Figure 4. Different dispersal models were then constructed by connecting nodes A–F in various ways. Dispersal distance between two populations for a particular dispersal model is the sum of the distances along paths that connect them. Figure 4 shows the dispersal models we tested.

To calculate genetic distance, we randomly picked one individual from each OTU as a representative. Individual plants of possible hybrid origin as determined by population structure analysis (see below) were excluded. We calculated the log-transformed proportion-of-shared-alleles distance among these individuals and used this distance matrix in all tests. We then calculated correlation coefficients between the genetic distance matrix and dispersal and clinal matrices according to Smouse *et al.* (1986). The significance of the correlations was estimated using matrix permutation tests with 10,000 replicates (Dietz 1983; Smouse *et al.* 1986).

Population structure analysis: For the analysis of population structure and detection of intermediate types (hybrids or ancestral forms), we used a model-based clustering method as implemented in the software program STRUCTURE (Pritchard *et al.* 2000). In this analysis, a number of populations (K) is assumed to be present and to contribute to the genotypes of the sampled individuals. The genotype of each individual is a function of the allele frequencies in these K populations (clusters) and the proportion of its genotype drawn from each of the K populations (q_k). Loci are assumed to be independent, and each K population is assumed to follow Hardy-Weinberg equilibrium.

A Monte Carlo Markov chain method was used to estimate allele frequencies in each of the K populations and the degree of admixture for each individual plant. In the analyses, we did not use any prior information about the geographic origin of the plants. We used STRUCTURE with 1,000,000 iterations and a burn-in period of 30,000. We increased the parameter, ALPHAPROPSD, from 0.05 (the default value) to 0.50 to explore a wide range of possible values of ALPHA, the degree of admixture. At least three independent runs were assessed for each fixed number of populations (K).

For the analysis of introgression between maize and teosinte, we used a sample of 52 Mexican maize landraces (Matsuoka *et al.* 2002b). Separate analyses with maize were made for *Z. luxurians*, *Z. diploperennis*, and *Z. mays* ssp. *huehuetenangensis*. In each case, we assumed two clusters. We analyzed the ssp. *parviglumis* and ssp. *mexicana* plants together with maize (ssp. *mays*), assuming three clusters. We performed this analysis with the three subspecies together since introgression between ssp. *mexicana* and ssp. *parviglumis* may also have occurred. Plants possessing <80% ancestry in their own cluster were considered to be of possible hybrid origin. For the analysis of population structure within ssp. *mexicana* and ssp. *parviglumis*, we used STRUCTURE with $K = 1$–6 on each subspecies separately. For all analyses with STRUCTURE, we used only a subset of 70 SSRs that had <10% null phenotypes (null alleles or missing data).

Core set: To assist in the management of a large germplasm collection, core sets have been defined to represent a large proportion of diversity encompassed in the entire collection (Brown 1989). Recently, core sets in many crops and their relatives have been established (Grenier 2000a,b; Oritz *et al.* 1998). To assist in the use of ssp. *mexicana* and ssp. *parviglumis* germplasm, we defined core sets of accessions that capture the maximum number of SSR alleles using the Core Set function in PowerMarker (Liu 2002). The method is based on a simulated annealing algorithm (Liu 2003). One-hundred replicates with different initial subsets were performed.

RESULTS

Diversity statistics: Observed heterozygosity, gene diversity (expected heterozygosity), number of alleles, and number of private alleles of each taxon are shown in Table 1. *Z. mays* possesses substantially higher values for heterozygosity and gene diversity than the other diploid

TABLE 1

Diversity statistics for teosinte microsatellites

Taxon	No. of accessions	No. of plants	Observed heterozygosity	Gene diversity	No. of alleles	No. of private alleles
			Section Luxuriantes			
Z. diploperennis	5	6	0.33	0.65	340	38
Z. luxurians	10	13	0.33	0.73	690	109
Z. perennis	3	4	—	—	377	52
			Section Zea			
Z. mays	154	214	0.46	0.89	3202	2313
ssp. *huehuetenangensis*	3	7	0.45	0.72	447	64
ssp. *mexicana*	69	93	0.43	0.85	2051	420
Race Central Plateau	32	47	0.42	0.85	1475	163
Race Chalco	33	39	0.46	0.82	1333	156
Race Nobogame	4	7	0.33	0.67	417	24
ssp. *parviglumis*	82	114	0.48	0.89	2609	837
Race Balsas	61	96	0.50	0.89	2456	586
Race Jalisco	17	18	0.41	0.83	966	91

Gene diversity and observed heterozygosity were calculated as the average over plants for a taxon.

species (*Z. luxurians* and *Z. diploperennis*). Within *Z. mays* (*sensu lato*), ssp. *huehuetenangensis* possesses a lower gene diversity value than either ssp. *parviglumis* or ssp. *mexicana*. Subspecies *parviglumis* possesses slightly greater gene diversity than ssp. *mexicana*; however, it possesses twice as many "private alleles" despite roughly equivalent samples of each. Among the races of ssp. *parviglumis* and ssp. *mexicana*, Nobogame is the least diverse. Overall, the taxa with narrow geographic distributions (*Z. luxurians*, *Z. diploperennis*, ssp. *huehuetenangensis*, and race Nobogame) show the least diversity, and the broadly distributed ssp. *parviglumis* is the most diverse taxon.

For all taxa, observed heterozygosity is substantially lower than gene diversity (expected heterozygosity), although under Hardy-Weinberg expectations these values should be the same. This discrepancy may arise for several reasons: (1) failure of one allele to amplify during PCR reactions (so-called "allele drop-out"), (2) population structure, and (3) inbreeding during seed increase in germplasm banks.

Phylogeny: A phylogenetic tree of all 237 individuals was constructed using the Fitch-Margoliash method (Figure 2). In this tree, individuals that are putative hybrids between taxa as shown by population structure analysis (see below) are marked with a large "H." The tree shows that the individual plants of any given species form a monophyletic group, if one excludes one *Z. diploperennis* plant of apparent hybrid origin. The two perennial species (*Z. diploperennis* and *Z. perennis*) are sisters, and *Z. luxurians* is sister to them. Within *Z. mays*, ssp. *huehuetenangensis* is monophyletic and sister to the other two subspecies. Individuals of ssp. *mexicana* and ssp. *parviglumis* are largely but not completely separated. All individuals that failed to cluster with their own subspecies were identified as putative hybrids by population structure analysis and accordingly are marked with a large H in Figure 2. Four of the six individuals of the eastern Balsas geographic group of ssp. *parviglumis* were identified as putative hybrids, containing >20% ssp. *mexicana* ancestry.

Within ssp. *mexicana*, there is clear geographic patterning of the individuals within Figure 2. Plants from Durango, Nobogame, the Valley of Mexico (Chalco), and Puebla mostly cluster near other plants of the same geographic origin. However, plants from the Central Plateau are dispersed throughout the ssp. *mexicana* clade. Within ssp. *parviglumis*, plants from Jalisco, Oaxaca, and southern Guerrero mostly cluster near other plants of the same geographic origin. Plants from the Central Balsas area are more dispersed with plants from the eastern portion of the Central Balsas region separated from those of the western portion. As mentioned above, plants from the eastern Balsas region occur either near or within the ssp. *mexicana* clade.

To further assess the phylogenetic affinities within Zea, we constructed a Fitch-Margoliash tree (Figure 3) for pooled groups (OTUs) of two to four individual plants (or five plants in the case of *Z. diploperennis*). Membership of the plants in the 76 OTUs was based on their geographic origin and position in Figure 2 (see supplemental materials at http://genetics.org/supplemental/). In this tree, there is strong support from the bootstrap procedure for the monophyly of *Z. luxurians* (100%), ssp. *huehuetenangensis* (100%), and ssp. *mexicana* (98%). The two perennial species, *Z. diploperennis* and *Z. perennis*, are also well supported as a monophyletic group (99.5%). *Z. mays* ssp. *parviglumis* is paraphyletic as the eastern Balsas group of this subspecies is basal to ssp. *mexicana*. Since the eastern Balsas group is of putative hybrid origin (see below), we reassessed the

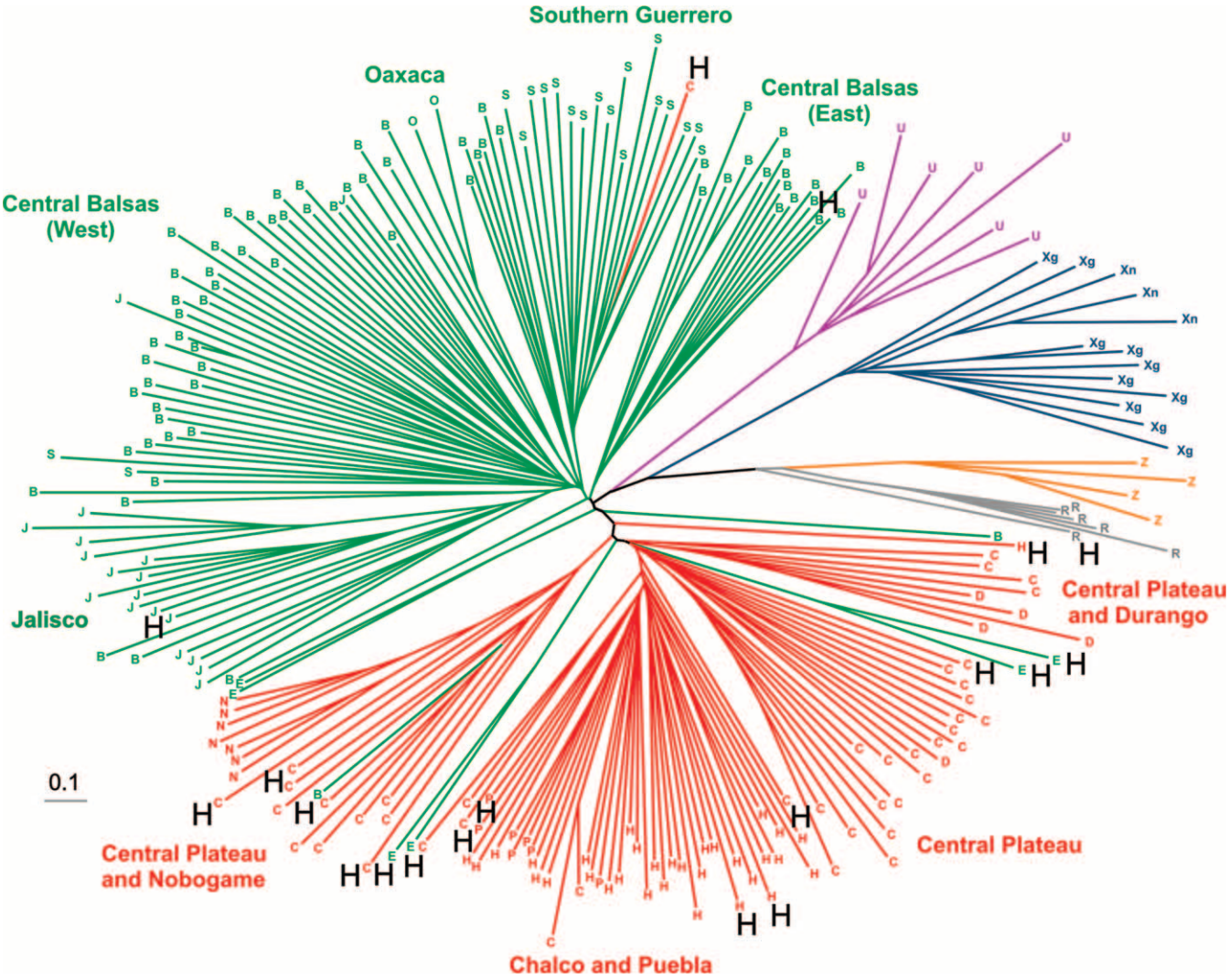
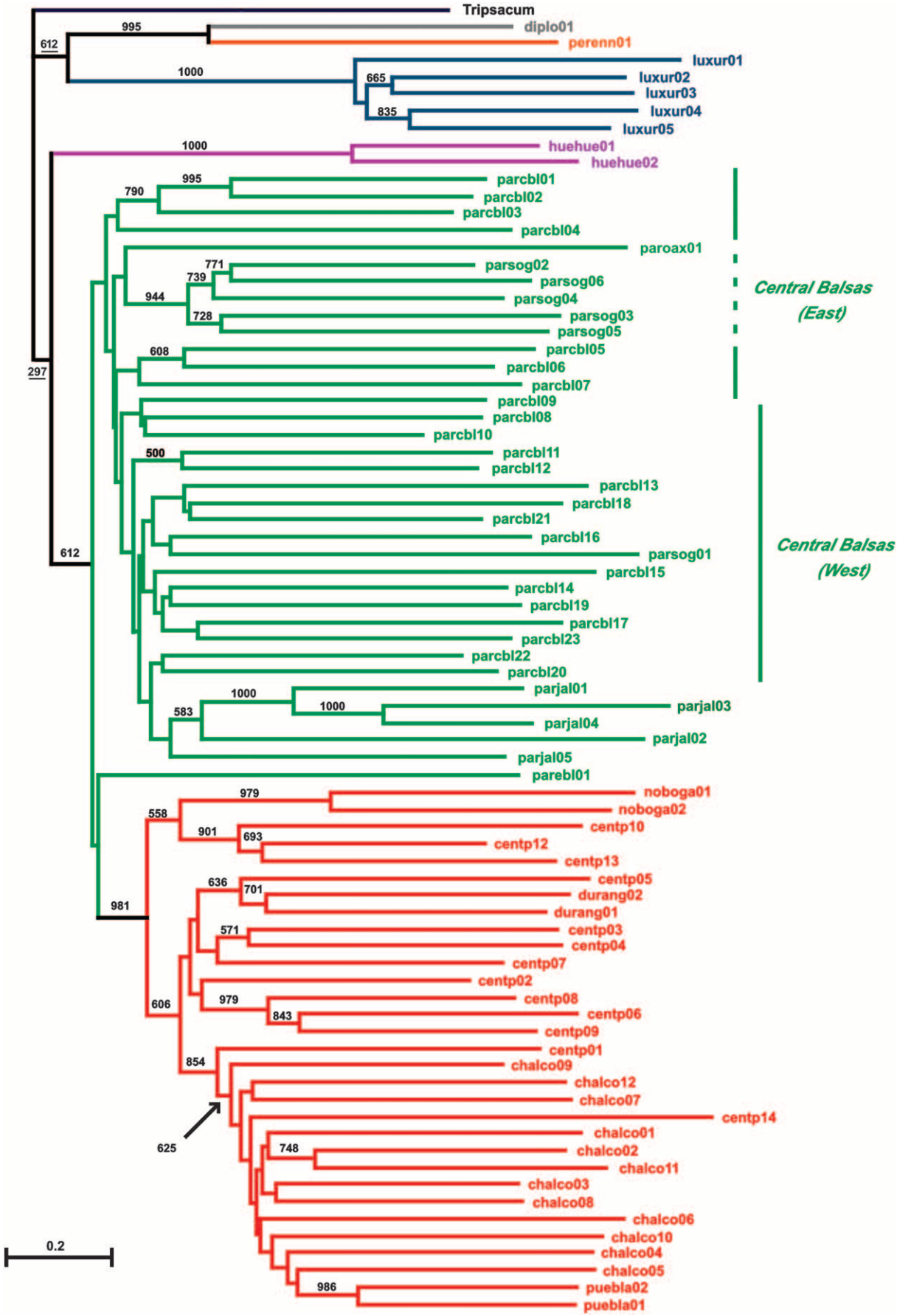


FIGURE 2.—Unrooted phylogeny of individual teosinte plants using the Fitch-Margoliash method and the log-transformed proportion of shared-allele distance among 93 microsatellite loci. The tree contains 237 individuals. A large H indicates plants identified as being of putative hybrid origin by population structure analysis. *Z. mays* ssp. *parviglumis*: (B) Central Balsas, (E) eastern Balsas, (J) Jalisco, (O) Oaxaca, (S) South Guerrero. *Z. mays* ssp. *mexicana*: (C) Central Plateau, (H) Chalco, (D) Durango, (N) Nobogame, (P) Puebla. (U) *Z. mays* ssp. *huehuetenangensis*, (Xg) *Z. luxurians* (Guatemala), (Xn) *Z. luxurians* (Nicaragua), (R) *Z. diploperennis*, (Z) *Z. perennis*.

phylogeny after excluding this OTU, in which case ssp. *parviglumis* is monophyletic, but has only 46% bootstrap support as such.

Figure 3 reveals strong geographic patterning of the OTUs within ssp. *parviglumis*. Excluding eastern Balsas, which is basal to ssp. *mexicana*, OTUs from the eastern portion of the Central Balsas region (parcbl01–parcbl07) are the basal-most samples within this subspecies. Nested within these OTUs is a clade composed of southern Guerrero OTUs plus the Oaxaca OTU, suggesting that the latter groups were derived from the populations of the eastern Central Balsas. OTUs from the western portion of the Central Balsas region (parcbl08–parcbl23) are derived from those of the eastern portion. The Jalisco OTUs fall within the clade of the western Central Balsas OTUs, consistent with their geographic proximity. Only one placement is inconsistent with geography, namely the position of the southern Guerrero OTU (parsog01) within the Central Balsas clade. Many of the basal branches are short within the ssp. *parviglumis* clade and few branch points have strong bootstrap support; however, the consistency of the clades with geography suggest that many of the observed relationships likely reflect the history of populations within this subspecies.

Geographic patterning of OTUs is also seen within ssp. *mexicana* (Figure 3). There is a basal split of this subspecies with OTUs from the Central Plateau geographic region falling in both clades. One clade contains the Nobogame geographic group along with OTUs from the northwestern portion of the Central Plateau (centp10, centp12, and centp13). In the other clade, the



Durango geographic group clusters with OTU centp05 from the core region of Central Plateau populations. OTU centp01 of race Central Plateau, which grows at the highest elevation (2300 m) for this race, is basal to race Chalco, which grows at elevations of ~2300 m and higher in the Valley of Mexico. The Puebla geographic group, which lies to the east of the Valley of Mexico, is nested within the race Chalco of the Valley of Mexico (see Figure 1). The position of OTU centp14 from Jalisco is exceptional in that it is within the Chalco group from the Valley of Mexico and may represent a case of long-distance dispersal. Bootstrap support for most clades within ssp. *mexicana* is generally weak, although the Nobogame (98%) and Puebla (99%) clades are strongly supported.

An interesting question for Zea systematics is the placement of the root of phylogeny for the genus. We rooted the Zea OTU tree using Tripsacum (Figure 3). When nulls in Zea and Tripsacum were treated as the same allele (identical by descent), the root was placed between *Z. luxurians* and the perennial species such that *Z. luxurians* was basal to *Z. mays*; however, this rooting was found in only 35% of the bootstrap samples. When nulls in Zea and Tripsacum were treated as distinct alleles (nonidentical by descent), the root was placed between section Luxuriantes (the perennial species and *Z. luxurians*) and section Zea (*Z. mays*). This rooting has stronger support (61%); however, it is not statistically robust.

Phylogeography of ssp. *mexicana* and ssp. *parviglumis*: We compared multiple dispersal models against a clinal model for the geographic distribution of the Mexican annual teosintes (ssp. *mexicana* and ssp. *parviglumis*) to determine which model best explains the observed pattern of extant variation (Figure 4). Models (1–10 in Figure 4) were chosen to test different linkages between the two subspecies as well as different linkages between the four population centers for ssp. *parviglumis*. Model 1 connects ssp. *parviglumis* through the eastern Balsas region to ssp. *mexicana* through a point between the Central Plateau and Chalco regions. Models 5, 6, 7, and 8 are variants of this connection. Models 5, 6, and 7 could be envisioned as showing an origin in the south (area E) and migration along two paths to the north. Models 2, 3, and 4 are all variants of a connection between ssp. *parviglumis* to ssp. *mexicana* through the central Balsas area (D). Model 9 connects the subspecies through areas C–E and shows migrations out of the south along two paths. Model 10 connects the subspecies through areas B–F and shows migrations from the west along two paths.

The correlation coefficient between geographic (clinal model) and genetic distance is 0.28 ($P < 0.001$), indicating a significant relationship between these distance measures (Table 2). However, the correlation coefficients between dispersal and genetic distance for the different dispersal models are all higher. Model 1 shows the highest correlation of the models tested (0.49; $P < 0.001$). The partial correlation between geographic and genetic distance, while the dispersal distance for model 1 is held constant, is -0.032 ($P > 0.81$); the partial correlation between model 1 dispersal and genetic distance, while geographic distance is held constant, is 0.42 ($P < 0.001$). These results indicate that dispersal model 1 explains considerable variation that is not explained by the clinal model (or by most other dispersal models), while essentially all the variation explained by the clinal model is also explained by the dispersal model.

Population structure: *Hybridization with maize:* Teosinte grows near maize in most locations and is capable of hybridizing with maize, allowing admixture between maize and the teosintes to occur. To measure the degree of admixture, we performed population structure analysis using the software program STRUCTURE (Pritchard *et al.* 2000). *Z. luxurians*, *Z. diploperennis*, and *Z. mays* ssp. *huehuetenangensis* were analyzed separately with maize, assuming two clusters in each analysis. The results indicate that all plants of *Z. luxurians* and ssp. *huehuetenangensis* have a high membership in their own cluster (>99%) with maize contributing <1% to their ancestry. This result argues that there has been little or no gene flow from maize into these teosintes, although different plant samples may produce different results. In contrast, only five of six plants of *Z. diploperennis* showed a high membership (>99%) in their own cluster, and one plant of *Z. diploperennis* (DPCUA01) showed only 77% membership in its own cluster plus 23% membership in the maize cluster. The plant is likely a maize-diploperennis hybrid or an introgressant.

To test for admixture between maize, ssp. *parviglumis*, and ssp. *mexicana*, these three taxa were evaluated together in a single analysis that assumed three clusters. *Z. mays* ssp. *mexicana* and ssp. *mays* showed average membership in their expected clusters of 92% and 96%, respectively, indicating that these taxa maintain distinct gene pools despite growing sympatrically. *Z. mays* ssp. *parviglumis* showed an ancestry in its own cluster of only 68%, the majority of its remaining ancestry (27%) being

Figure 3.—Rooted phylogeny for 76 groups (OTUs) of individual plants using the Fitch-Margoliash method and the log-transformed proportion of shared-allele distance among 93 microsatellite loci. The numbers on the branches indicate the number of times a clade appeared in 1000 bootstrap samples and are shown for all clades with >50% bootstrap support. To locate the root for Zea, separate analyses including the outgroup, Tripsacum, but only 61 SSRs, were performed. The tree as drawn shows the placement of the root when null alleles in Tripsacum are coded as distinct (not identical by descent) from null alleles in Zea (see text). Bootstrap values from this analysis are underlined.

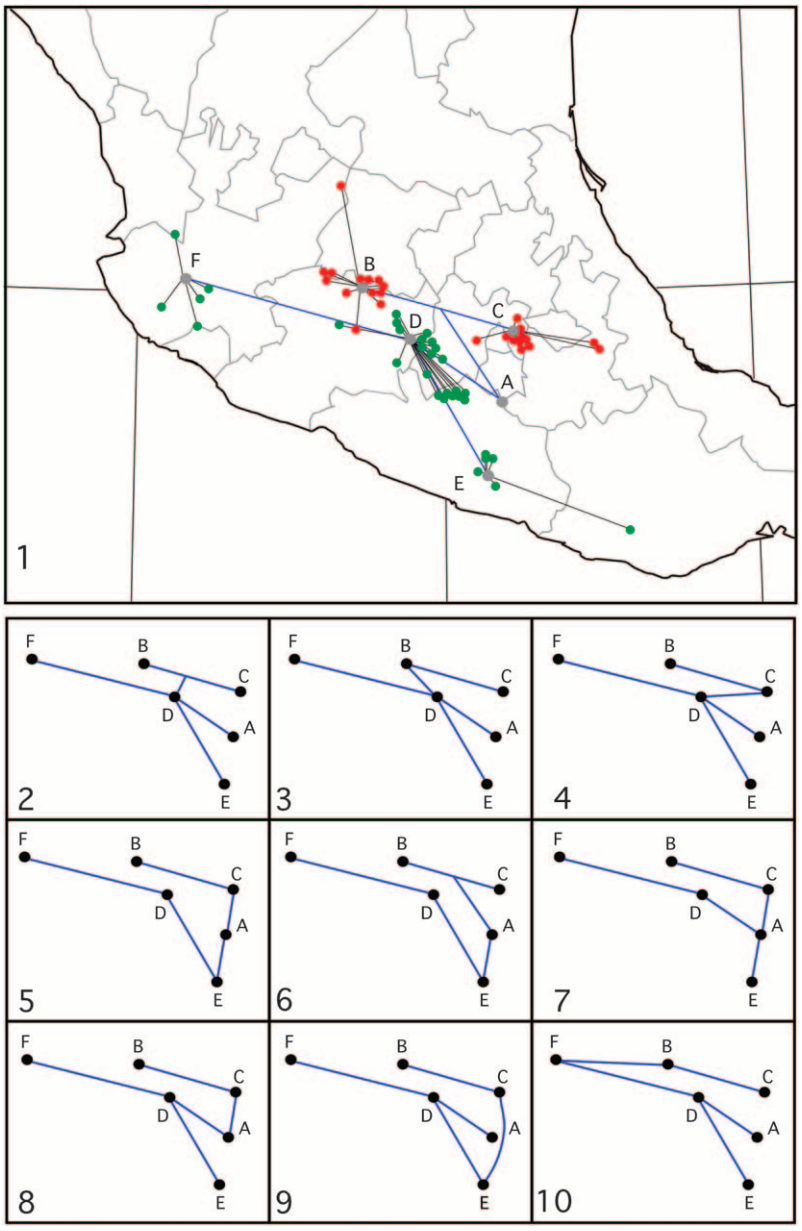


FIGURE 4.—Dispersal models for 63 OTUs of *Z. mays* ssp. *mexicana* and ssp. *parviglumis*. Gray circles indicate nodes of the tree, which are labeled A–F. Model 1 shows the highest correlation with genetic distance.

attributed to the maize cluster. This result suggests some degree of admixture between spp. *mays* and ssp. *parviglumis*; however, it may also be a function of the recent divergence between these two taxa such that their gene pools are not yet completely differentiated.

Examination of the STUCTURE results for individual plants identifies some putative hybrids. Seven ssp. *mexicana* plants have >20% membership in the maize cluster: MXCEN3 (21%), MXCEN5b (33%), MXCHA9 (23%), MXCHA13a (23%), MXCHA16a (55%), MXPUE3 (26%), and MXPUE4 (28%). Three ssp. *mexicana* plants have substantial membership in the ssp. *parviglumis* cluster, all of them from race Central Plateau: MXCEN4b (34%), MXCEN22 (22%), and MXCEN23 (33%). One ssp. *mexicana* plant (MXCEN19d) has 99% membership in ssp. *parviglumis* and may represent a mislabeled individual. Four of six ssp. *parviglumis* plants from the eastern Balsas region have a high membership in the spp. *mexicana* cluster: PREBL3b (28%), PREBL4a (64%), PREBL3a (38%), and PREBL4b (74%). Moreover, both of the remaining two plants had 19% ancestry in the ssp. *mexicana* cluster. These results suggest that the eastern Balsas ssp. *parviglumis* represent either *mexicana-parviglumis* hybrids or perhaps an evolutionarily intermediate form between these two subspecies. Three other ssp. *parviglumis* plants have high membership in the ssp. *mexicana* cluster: PRJAL17 (34%), PRCBL03c (39%), and PRCBL45c (92%).

When the populations were plotted on a map, the results of the structure analysis reveal some geographic trends (Figure 5). First, the ssp. *mexicana* populations that show admixture with maize are all from the eastern portion of the ssp. *mexicana* range (the eastern Central Plateau, Valley of Mexico, and Puebla). Second, the ssp. *mexicana* populations that show admixture with ssp.

TABLE 2

Correlations (*r*) between geographic or dispersal and genetic (SSR) distance matrices

Model	*r*	Model *x*|model *y*	*r*
Geographic	0.283	Geographic|model 1	−0.032
1	0.493	Model 1|geographic	0.423
2	0.390	Model 2|geographic	0.287
3	0.365	Model 3|geographic	0.243
4	0.422	Model 4|geographic	0.327
5	0.431	Model 5|geographic	0.351
6	0.454	Model 6|geographic	0.378
7	0.427	Model 7|geographic	0.336
8	0.464	Model 8|geographic	0.384
9	0.429	Model 9|geographic	0.349
10	0.435	Model 10|geographic	0.384

Models 1–10 are shown in Figure 4. The geographic model is the matrix of geographic or linear distances between the populations. Model *x*|model *y* indicates the partial correlation of model *x* with genetic distance when model *y* is held constant.

parviglumis are from the western portion of the ssp. *mexicana* range. Third, four of the seven ssp. *parviglumis* plants that show admixture with ssp. *mexicana* are from the eastern Balsas region, which is situated near the ssp. *mexicana* populations of the Valley of Mexico, suggesting an opportunity for hybridization.

Population substructure: Both *Z. mays* ssp. *parviglumis* and ssp. *mexicana* are broadly distributed throughout Mexico, raising the question of whether they behave as a single Hardy-Weinberg population or exhibit some degree of geographic substructuring. To address this question, we performed population structure analysis using the software program STRUCTURE (Pritchard *et al.* 2000). For ssp. *parviglumis*, the highest likelihood was obtained for $K = 2$ clusters of plants, dividing this subspecies into a group of 32 plants that come largely from the eastern part of its range (Balsas) and a group of 18 plants that come largely from the western part (Jalisco) (Table 3; supplemental materials at http://genetics.org/supplemental/). Twenty-eight plants were intermediate between these two groups. For ssp. *mexicana*, the highest likelihood was obtained for $K = 3$ clusters of plants. One cluster includes the Nobogame region and allied plants from the Central Plateau, another includes the Chalco-Puebla regions and their allied plants from the Central Plateau, and the third contains other plants of the Central Plateau and the Durango regions. Only two plants were intermediate among these three groups.

Core sets: We have analyzed a large number of accessions (172). For some other bioassays, researchers may need to reduce the number of samples utilized because of time-cost considerations. We have selected core sets of accessions for ssp. *mexicana* and ssp. *parviglumis* that capture the maximum number of SSR alleles for sample sizes of 12 and 25 plants (Table 4). Gene diversity in each core set is equivalent to that found in the entire sample (Tables 1 and 4). Each core set contains a number of accessions from different geographic groups that are proportional to the number of accessions for the different geographic groups in the entire sample, as revealed by nonsignificant goodness-of-fit tests (data not shown).

DISCUSSION

Diversity: Estimates of gene diversity were high for all taxa, reflecting the highly polymorphic nature of SSRs. The gene diversity values reported here for the teosintes exceed those that we have previously observed (Matsuoka *et al.* 2002a). The reasons for this difference likely are that we used mostly dinucleotide repeat loci in the present study, whereas in our previous study we used mostly trinucleotide repeat loci (Matsuoka *et al.* 2002a), and the fact that dinucleotide loci have a higher mutation rate (Vigouroux *et al.* 2002). If one considers relative diversity among taxa, then it is apparent that those taxa with very narrow geographic distributions exhibit lower gene diversity than taxa with broad geographic distributions (Table 1, Figure 1). The lowest value (0.67) is found for race Nobogame, which occurs in a single valley in the state of Chihuahua. *Z. diploperennis*, which exists in only a few local populations on the Sierra de Manatlán of Jalisco, also has a very low value (0.69). *Z. luxurians* (0.73) and *Z. mays* ssp. *huehuetenangensis* (0.72) exhibit the next highest values. *Z. mays* ssp. *huehuetenangensis* exists in multiple populations, but all are within a single province of Guatemala. Similarly, *Z. luxurians* is known mostly from a restricted region of southeastern Guatemala, although it has outlier populations in Honduras and Nicaragua. The taxa that exhibit the highest gene diversity (races Balsas and Central Plateau) have very broad geographic distributions. This overall pattern suggests that small population size or bottlenecks associated with the founding of the narrowly distributed taxa have caused the observed reductions in gene diversity.

A comparison of ssp. *mexicana* and ssp. *parviglumis* indicates that the latter subspecies is more diverse with a slightly higher gene diversity and a far greater number of private alleles. Examination of Figure 3 also shows that branch lengths appear shorter for ssp. *parviglumis* than for ssp. *mexicana*, especially for the internal branches within the clades. Since changes in allele frequency due to genetic drift contribute substantially to genetic distance (branch length), the longer internal branches for ssp. *mexicana* might reflect smaller local population sizes and repeated founder events. Under these circumstances, rare (private) alleles would be readily lost, accounting for the substantially smaller number of private alleles observed in ssp. *mexicana*. Consistent with the idea that ssp. *mexicana* populations are smaller or experienced repeated founder events, previous isozyme data

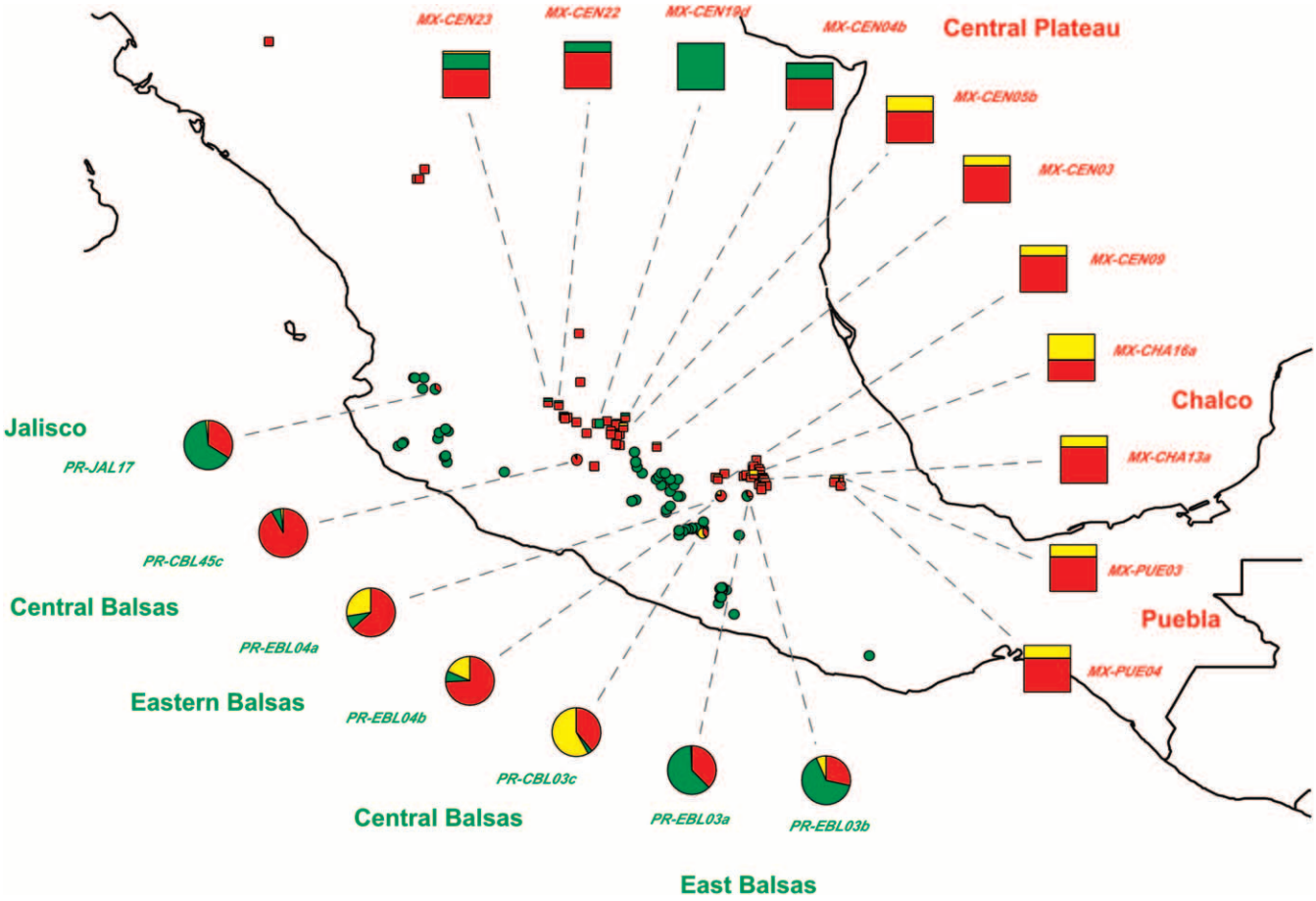


FIGURE 5.—Results of the population structure analysis for ssp. *mexicana* (squares) and ssp. *parviglumis* (circles). The symbols for plants of apparent mixed ancestry based on the arbitrary criterion of possessing <80% membership in their own subspecies are enlarged. For the enlarged symbols, the symbol is color coded proportionally to its degree of ancestry from ssp. *mexicana* (red), ssp. *parviglumis* (green), and ssp. *mays* (yellow).

show that G_{ST}, a measure of population subdivision, is 0.29 in race Central Plateau (ssp. *mexicana*) as compared to 0.16 in race Balsas (ssp. *parviglumis*) (DOEBLEY *et al.* 1984).

Phylogeny: Our analyses present the first comprehensive phylogeny for Zea using a large number of nuclear markers. Despite this large number and comprehensive sampling, many of the branch points within the tree have little statistical support. Thus, we present only a phylogenetic hypothesis that will require verification or revision in the future. Our most strongly supported root within Zea suggests an initial division of the genus into section Luxuriantes (*Z. perennis*, *Z. diploperennis*, and *Z. luxurians*) and section Zea (*Z. mays sensu lato*) (Figure 3). This rooting results when null alleles in Zea and Tripsacum are treated as nonidentical by descent. This assumption seem reasonable since the high frequency of nulls in Tripsacum (50% in each Tripsacum sample) suggests a high probability of multiple independent origins of nulls. This rooting of Zea is consistent with a cpDNA phylogeny that has very strong statistical support (DOEBLEY *et al.* 1984). If one treats the nulls in Zea and Tripsacum as identical by descent, then the SSR data are consistent with the placement of the root for Zea between the perennials and *Z. luxurians* such that *Z. luxurians* is basal to *Z. mays* (BUCKLER and HOLTSFORD 1996). However, this rooting is less well supported than the rooting of Zea between sections Luxuriantes and Zea.

In our phylogeny, the two perennial species are sister taxa (Figure 3), supporting the interpretation that *Z. perennis* is an autotetraploid derived from a *Z. diploperennis*-like ancestor (DOEBLEY *et al.* 1987; KATO and LOPEZ 1990). Our phylogeny also indicates that ssp. *huehuetenangensis* is basal to the Mexican annual teosintes (ssp. *mexicana* and ssp. *parviglumis*), which is in agreement with rDNA sequences (BUCKLER and HOLTSFORD 1996). The subspecies *mexicana* is derived from within ssp. *parviglumis*, rendering the latter subspecies paraphyletic (Figure 3).

The paraphyly of ssp. *parviglumis* hinges upon the status of the eastern Balsas populations, which appear intermediate between the two subspecies in Figure 2. We consider two interpretations:

1. If the eastern Balsas populations are interpreted as *parviglumis-mexicana* hybrids, then both subspecies

TABLE 3

Population structure groups for the Mexican annual teosintes

Cluster	Members
	Z. mays ssp. *mexicana*
Central Plateau	MXCEN05a, MXCEN06, MXCEN07, MXCEN08, MXCEN09, MXCEN10, MXCEN11, MXCEN12, MXCEN13, MXCEN14c, MXCEN15b, MXCEN16, MXCEN17, MXCEN18c, MXCEN19, MXDUR01a, MXDUR02, MXDUR03, MXDUR04
Chalco	MXCEN01, MXCEN02b, MXCEN28b, MXCHA01, MXCHA02, MXCHA03, MXCHA04, MXCHA05, MXCHA06, MXCHA10, MXCHA11, MXCHA12b, MXCHA13, MXCHA14, MXCHA15, MXCHA16b, MXCHA17, MXCHA18, MXCHA19a, MXCHA20, MXCHA21, MXCHA22, MXCHA23, MXCHA24, MXCHA25, MXCHA28, MXCHA29b, MXCHA30, MXCHA32, MXCHA33, MXCHA34, MXCHA7, MXCHA9
Nobogame	MXCEN04a, MXCEN20, MXCEN21, MXCEN25, MXCEN26, MXCEN27d, MXCEN29, MXNOB01a, MXNOB02, MXNOB03c, MXNOB04
Intermediate	MXCEN04a, MXCEN29
	Z. mays ssp. *parviglumis*
Balsas	PRCBL01, PRCBL02a, PRCBL03, PRCBL04a, PRCBL05, PRCBL07d, PRCBL09a, PRCBL11, PRCBL12, PRCBL19, PRCBL20, PRCBL23, PRCBL24, PRCBL25, PRCBL29, PRCBL34, PRCBL36, PRCBL48, PROAX01, PROAX02, PRSOG01, PRSOG02, PRSOG03, PRSOG04a, PRSOG05, PRSOG06, PRSOG07d, PRSOG08, PRSOG09, PRSOG10b, PRSOG11, PRSOG12
Jalisco	PRCBL42, PRCBL44, PRCBL51, PRJAL01, PRJAL02, PRJAL03, PRJAL04, PRJAL05, PRJAL06a, PRJAL07, PRJAL08, PRJAL09, PRJAL10, PRJAL12, PRJAL14, PRJAL15b, PRJAL16, PRJAL17
Intermediate	PRCBL06, PRCBL08, PRCBL10c, PRCBL13, PRCBL14, PRCBL15, PRCBL16b, PRCBL17, PRCBL26b, PRCBL27, PRCBL28, PRCBL30, PRCBL31, PRCBL32, PRCBL33, PRCBL35, PRCBL37, PRCBL38, PRCBL39, PRCBL40, PRCBL41, PRCBL43b, PRCBL45a, PRCBL46c, PRCBL47, PRCBL50, PRJAL11, PRSOG13

would be monophyletic. The population structure analysis is consistent with this interpretation since it identified four of six eastern Balsas plants as being admixed between ssp. *parviglumis* and ssp. *mexicana*, and the remaining two plants are both estimated to possess nearly 20% ssp. *mexicana* germplasm. Geographically, the eastern Balsas populations are in between the Chalco populations of ssp. *mexicana* and the central Balsas population of ssp. *parviglumis*, so an opportunity for hybridization exists.

2. If the eastern Balsas populations are interpreted as the ancestral population out of which ssp. *parviglumis* and ssp. *mexicana* were derived, then ssp. *parviglumis* would be paraphyletic. In this case, the STRUCTURE analysis erroneously identifies eastern Balsas plants as *parviglumis-mexicana* admixtures when in fact they represent the basal (intermediate) population from which both ssp. *mexicana* and other ssp. *parviglumis* were derived. If this interpretation is correct, then the *mexicana-parviglumis* ancestor grew at the middle elevations (~1700 m) and spread from there to both lower elevations (most ssp. *parviglumis*) and higher elevations (ssp. *mexicana*).

Introgression: We performed a model-based analysis of population structure that allows one to infer both the number of populations (clusters) and the degree of membership of each individual in each cluster (admixture or introgression) for a sample (Pritchard *et al.* 2000). This analysis fits a model that minimizes both Hardy-Weinberg and linkage disequilibrium within clusters. We used this analysis to identify the degree of admixture between each of the teosinte taxa and maize. We found no evidence for gene flow or admixture between *Z. luxurians* or ssp. *huehuetenangensis* and maize. This result is consistent with prior field observations that these taxa rarely hybridize with maize or that their hybrids with maize are restricted to plants that invade maize fields (Wilkes 1977). We did identify one *Z. diploperennis* plant that appears to be admixed with maize. This observation is congruent with previous reports from isozymes (Doebley *et al.* 1984) and rDNA (Buckler and Holtsford 1996). It is also consistent with field observations that *Z. diploperennis* grows near maize fields and occasionally forms hybrids with maize (Benz *et al.* 1990).

Subspecies *mexicana* is the teosinte that grows most commonly in maize fields and has been observed to hybridize readily with maize (Wilkes 1977). In some fields, upwards of 10% of "teosinte" plants are actually maize-teosinte hybrids (Wilkes 1967). Despite this record of hybridization, the STRUCTURE analysis indicates that maize and ssp. *mexicana* have very distinct gene pools with an estimate of only 8% of the ssp. *mexicana* gene pool being derived from maize. One ssp.

TABLE 4

Core set accessions of ssp. *mexicana* and ssp. *parviglumis*

Z. mays ssp. *parviglumis*		*Z. mays* ssp. *mexicana*	
Core set of 12	Core set of 25	Core set of 12	Core set of 25
PRCBL07	PRCBL07	MXCEN14	MXCEN05
PRCBL09	PRCBL09	MXCEN15	MXCEN06
PRCBL11	PRCBL14	MXCEN20	MXCEN10
PRCBL16	PRCBL16	MXCEN27	MXCEN14
PRCBL34	PRCBL17	MXCHA02	MXCEN15
PRCBL44	PRCBL18	MXCHA13	MXCEN16
PRCBL45	PRCBL32	MXCHA15	MXCEN20
PRCBL46	PRCBL33	MXCHA18	MXCEN24
PRCBL48	PRCBL34	MXCHA27	MXCEN27
PRJAL16	PRCBL37	MXDUR04	MXCHA02
PROAX02	PRCBL41	MXNOB04	MXCHA11
PRSOG09	PRCBL44	MXPUE1	MXCHA13
	PRCBL45		MXCHA15
	PRCBL46		MXCHA17
	PRCBL48		MXCHA18
	PRCBL50		MXCHA21
	PRCBL51		MXCHA22
	PREBL01		MXCHA23
	PRJAL01		MXCHA24
	PRJAL08		MXCHA27
	PRJAL16		MXDUR04
	PROAX02		MXNOB01
	PRSOG04		MXNOB03
	PRSOG09		MXNOB04
	PRSOG10		MXPUE1
GD = 0.898	GD = 0.895	GD = 0.868	GD = 0.856

GD, gene diversity.

mexicana plant (MXCHA16a) was identified as possessing 55% maize ancestry and may be a maize-teosinte F_1. Six others have ~25% maize, suggesting that some of these may represent BC_1's. If these putative F_1 and BC_1 plants are excluded, then ssp. *mexicana* has only ~4% membership in the maize cluster. Genetic barriers to gene flow may be blocking the more complete homogenization of the maize and ssp. *mexicana* gene pools where they grow sympatrically (Evans and Kermicle 2001).

Three ssp. *mexicana* plants, all from the Central Plateau, were identified as being admixed with ssp. *parviglumis*. This observation is difficult to explain, given that these two subspecies do not grow sympatrically. We consider three possible explanations: (1) there is long-distance dispersal from ssp. *parviglumis* to ssp. *mexicana* populations; (2) the STRUCTURE analysis may erroneously attribute admixture with maize to admixture with ssp. *parviglumis*; or (3) the gene pools of the two subspecies are too recently diverged to be fully differentiated. The first explanation seems unlikely but cannot be discounted. The latter two explanations seem more likely, given that the gene pools of ssp. *mexicana*, ssp. *parviglumis*, and maize differ more in allele frequencies than by allele presence/absence. An interesting fact is that two of these three ssp. *mexicana* plants with ssp. *parviglumis* admixture occur at the low elevation (1520 and 1625 m) typical of ssp. *parviglumis*.

Of 117 ssp. *parviglumis* plants, 56 were identified by the STRUCTURE analysis as being admixed with maize (*i.e.*, possessing 20% or more maize germplasm). Rather than introgression, this result likely reflects the recent origin of maize from ssp. *parviglumis* such that their gene pools have not yet fully differentiated. We suggest recent origin rather than admixture since ssp. *parviglumis* (1) is known as the most "pure, wild" teosinte, (2) frequently grows in natural settings apart from maize fields, and (3) does not commonly hybridize with maize (Wilkes 1977). An additional 7 ssp. *parviglumis* plants were identified as admixed with ssp. *mexicana*. Four of these are from the eastern Balsas region and may represent either *parviglumis-mexicana* introgressants or evolutionary intermediates as discussed above. Another ssp. *parviglumis* plant (PRCBL45c) was assessed as possessing 92% ssp. *mexicana* germplasm. This plant comes from a population located near race Central Plateau populations of ssp. *mexicana*. Two other plants from this population also showed relatively high admixture with ssp. *mexicana* (PRCBL45b at 14% and PRCBL45d at 19%). Given the geographic location of these plants near the Central Plateau and these STRUCTURE results, it seems possible that this population possesses a mixture of *mexicana-parviglumis* germplasm. Another ssp. *parviglumis* plant (PRJAL17) with 34% ssp. *mexicana* ancestry was identified in Jalisco. This Jaliscan population is isolated from ssp. *mexicana*, and all other Jalisco populations showed quite low percentages of ssp. *mexicana* genome (<2%). The origin of this intermediate type could be due to long-distance dispersal.

Phylogeography: The phylogeographic analysis indicates that dispersal model 1 fits the distribution of populations of the Mexican annual teosintes better than a clinal model does. This dispersal model shows a linkage between ssp. *mexicana* and *parviglumis* through the eastern Balsas region (the mountains of Ixcateopan; node A in Figure 4). From the eastern Balsas, ssp. *parviglumis* diversifies in the central Balsas region (node D) and spreads from there along one path into southern Guerrero (node E, Mazatlán and El Salado) and Oaxaca, and along a second path into Jalisco (node F). The subspecies *mexicana* radiates out of a point between nodes B and C, spreading along one path to the Valley of Mexico (node C, Chalco) and Puebla, and along a second path into the Central Plateau through node B.

This phylogeographic model and our phylogenies have several implications (Figures 2 and 3):

1. Subspecies *parviglumis* originated in the eastern Balsas region since populations of this region are basal to other ssp. *parviglumis* populations (Figure 2).
2. Subspecies *parviglumis* originated at middle elevations

(~1500–1800 m) and then diversified into the lower elevations.
3. The ssp. *mexicana* populations from the northwestern part of the Central Plateau may represent the founding populations of this subspecies as shown in Figures 2 and 3. Interestingly, the STRUCTURE analysis indicated that these populations show ~7% admixture with ssp. *parviglumis* as compared to 2% for other ssp. *mexicana* populations. However, rather than admixture, STRUCTURE may be detecting an ancestral similarity between the gene pools of the ssp. *parviglumis* and the ssp. *mexicana* populations of the northwestern Central Plateau.
4. Despite their proximity and status as the two northern-most populations, Nobogame and Durango actually represent independent colonizations of northern Mexico that were derived from distinct ancestral populations in the Central Plateau.
5. Other recent long-distance dispersal events, such as the movement of a race Chalco type from the Valley of Mexico to Puebla, may have occurred. Birds or humans might be the vectors for such events.

Taxonomy: A taxonomy should provide a useful tool for field biologists while reflecting the phylogenetic history of the taxa as accurately as possible. The results reported here generally fit the taxonomy of Zea developed by Iltis and Doebley (Doebley and Iltis 1980; Iltis and Doebley 1980; Doebley 1990b). However, there are a few ambiguities worthy of discussion.

Doebley and Iltis (1980) divided the genus into section Luxuriantes (*Z. diploperennis*, *Z. perennis*, and *Z. luxurians*) and section Zea (*Z. mays*). Our best-supported phylogeny and a strongly supported cpDNA-based phylogeny (Doebley *et al.* 1987) are consistent with this division, but the SSR phylogeny lacks robust statistical support. If further data demonstrate that the root of Zea lies between the perennial and annual species, as suggested by Buckler and Holtsford (1996), then section Luxuriantes would be paraphyletic. Although paraphyletic, the section Luxuriantes/section Zea split would remain a reasonable division of Zea since members of each section share a suite of morphological and genetic features (Doebley 1990b). Other than its annual habit, *Z. luxurians* has little, if anything, in common with *Z. mays*. Paraphyletic taxa are commonly used in taxonomy, since requiring all taxa to be strictly monophyletic would lead to the creation of absurd taxa of no utility to field biologists. Zea already possesses one other paraphyletic taxa, ssp. *parviglumis*, since ssp. *mays* is nested within it.

Iltis and Doebley (1980) divided the Mexican annual teosintes into two subspecies on the basis of ecology and inflorescence morphology. *Z. mays* ssp. *mexicana* grows in the cooler, drier central highlands mostly above 1800 m, while *Z. mays* ssp. *parviglumis* grows in warmer, wetter lower elevations in the river valleys of southern and western Mexico mostly below 1800 m (Wilkes 1967; Iltis and Doebley 1980; Doebley 1983). *Z. mays* ssp. *mexicana* often has red, hairy leaf sheaths, while *Z. mays* ssp. *parviglumis* possesses mostly green and glabrous leaf sheaths. *Z. mays* ssp. *mexicana* typically has larger seed, larger male spikelets, and few tassel branches as compared to ssp. *parviglumis*. As shown in Figures 2 and 3, the SSR data are consistent with this ecological division, although the ssp. *parviglumis* populations of the eastern Balsas region are intermediate. What are the implications of these intermediate populations from a taxonomic perspective?

1. If they are *mexicana-parviglumis* hybrids as discussed above, then the taxonomic division is secure and it is not surprising that subspecies form occasional hybrid populations.
2. If they are evolutionarily intermediate populations, as discussed above, then ssp. *parviglumis* would be paraphyletic, but it is already a paraphyletic taxon since ssp. *mays* was derived from within it. Moreover, the *parviglumis-mexicana* division would still provide a useful taxonomic division of the Mexican annual teosintes into the upland and lowland forms.

Recently, Iltis and Benz (2000) described a new species of Zea from Nicaragua, *Z. nicaraguensis*. These authors noted that *Z. nicaraguensis* is similar to and closely allied with *Z. luxurians*. Our data confirm their observation that *Z. nicaraguensis* and *Z. luxurians* are closely related. Their evidence that it represents a new species is based on differences in ecology and tassel and plant morphology. Our data indicate that *Z. nicaraguensis* is not strongly differentiated from *Z. luxurians* by SSR data. In Figure 2, our three samples of *Z. nicaraguensis* are all nested within *Z. luxurians* and the branch length between these samples and samples of *Z. luxurians* are not particularly long. The status of *Z. nicaraguensis* should be investigated by determining its cross-compatibility with *Z. luxurians*. If they are interfertile, then it would be best to treat *Z. nicaraguensis* as a subspecies of *Z. luxurians*.

Core sets of collections: We employed a method based on a simulated annealing algorithm (Liu 2003) to choose core sets of 12 and 25 accessions for both ssp. *mexicana* and ssp. *parviglumis* to maximize SSR diversity. This algorithm improves upon previously proposed algorithms by avoiding local maxima during the search for the global maximum (Liu 2003). The core sets that we defined cover broad geographic distributions of the Mexican annual teosintes. For applications such as SNP discovery or capturing the maximum amount of allelic diversity for quantitative genetic analyses, these core sets or similar ones should prove useful when available resources do not allow the assay of a larger number of plants.

We thank Hugh Iltis who provided many of the seeds for this study, Jerry Kermicle for comments on the manuscript, and Korise Rasmusson for help with manuscript preparation. This work was supported by National Science Foundation grant DBI-0096033.

LITERATURE CITED

Benz, B. F., L. R. Sánchez-Velásquez and F. J. Santana-Michel, 1990 Ecology and ethnobotany of *Zea diploperennis*: preliminary investigations. Maydica **35:** 85–98.

Brown, A. H. D., 1989 Core collections: a practical approach to genetic resource management. Genome **31:** 818–824.

Buckler, E. S., IV, 1999 Phylogeographer: A tool for developing and testing phylogeographic hypotheses. http://www.maizegenetics.net/bioinformatics/phyloindex.htm.

Buckler, E. S., IV, and T. P. Holtsford, 1996 *Zea* systematics: ribosomal ITS evidence. Mol. Biol. Evol. **13:** 612–622.

Dietz, E. J., 1983 Permutation tests for association between two distance matrices. Syst. Zool. **32:** 21–26.

Doebley, J. F., 1983 The maize and teosinte male inflorescence: a numerical taxonomic study. Ann. Mo. Bot. Gard. **70:** 32–70.

Doebley, J. F., 1990a Molecular evidence and the evolution of maize. Econ. Bot. **44** (Suppl. 3): 6–27.

Doebley, J. F., 1990b Molecular systematics of *Zea* (Gramineae). Maydica **35:** 143–150.

Doebley, J. F., and H. H. Iltis, 1980 Taxonomy of *Zea* species (Gramineae). I. A subgeneric classification with key to taxa. Am. J. Bot. **67:** 982–993.

Doebley, J. F., M. M. Goodman and C. W. Stuber, 1984 Isoenzymatic variation in *Zea* (Gramineae). Syst. Bot. **9:** 203–218.

Doebley, J. F., W. Renfroe and A. Blanton, 1987 Restriction site variation in the Zea chloroplast genome. Genetics **117:** 139–147.

Evans, M. S., and J. L. Kermicle, 2001 *Teosinte crossing barrier1*, a locus governing hybridization of teosinte with maize. Theor. Appl. Genet. **103:** 259–265.

Felsenstein, J., 1993 PHYLIP (Phylogeny Inference Package), Version 3.5c. Department of Genetics, University of Washington, Seattle.

Gaut, B. S., M. Le Thierry d'Ennequin, A. S. Peek and M. C. Sawkins, 2000 Maize as a model for the evolution of plant nuclear genomes. Proc. Natl. Acad. Sci. USA **97:** 7008–7015.

Grenier, C., P. J. Bramel-Cox, M. Noirot, K. E. Prasada Rao and P. Hamon, 2000a Assessment of genetic diversity in three subsets constituted from the ICRISAT sorghum collection using random vs. non-random procedures. A. Using morpho-agronomical and passport data. Theor. Appl. Genet. **101:** 190–196.

Grenier, C., M. Deu, S. Kresovich, P. J. Bramel-Cox and P. Hamon, 2000b Assessment of genetic diversity in three subsets constituted from the ICRISAT sorghum collection using random vs. non-random procedures. B. Using molecular markers. Theor. Appl. Genet. **101:** 197–202.

Iltis, H. H., and B. F. Benz, 2000 *Zea nicaraguensis* (Poaceae), a new teosinte from Pacific coastal Nicaragua. Novon **10:** 382–390.

Iltis, H. H., and J. F. Doebley, 1980 Taxonomy of *Zea* (Gramineae). II. Subspecific categories in the *Zea mays* complex and a generic synopsis. Am. J. Bot. **67:** 994–997.

Jaenicke-Després, V., E. S. Buckler, B. D. Smith, M. T. P. Gilbert, A. Cooper *et al.*, 2003 Early allelic selection in maize as revealed by ancient DNA. Science **302:** 1206–1208.

Kato, T. A., and A. R. Lopez, 1990 Chromosome knobs of the perennial teosintes. Maydica **35:** 125–141.

Liu, K., 2002 PowerMarker: a powerful software for marker data analysis. http://www.powermarker.net.

Liu, K., 2003 Software and methods for analysing molecular genetic marker data. Ph.D. Thesis, North Carolina State University, Raleigh, NC.

Lukens, L. N., and J. F. Doebley, 1999 Epistatic and environmental interactions for quantitative trait loci involved in maize evolution. Genet. Res. **74:** 291–302.

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

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

Oritz, R., E. N. Ruiz-Tapia and A. Mujica-Sanchez, 1998 Sampling strategy for a core collection of Peruvian quinoa germplasm. Theor. Appl. Genet. **96:** 475–483.

Pritchard, J. K., M. Stephens and P. Donnelly, 2000 Inference of population structure using multilocus genotype data. Genetics **155:** 945–959.

Sanchez, J., T. A. Kato, M. Aguilar, J. M. Hernandez, A. Lopez *et al.*, 1998 Distribución y caracterizacion del teocintle. Instituto Nacional de Investigaciones Forestales, Agricolas y Pecuarias, Jalisco, México.

Sanz-Alferez, S., P. SanMiguel, Y. K. Jin, P. S. Springer and J. L. Bennetzen, 2003 Structure and evolution of the Cinful retrotransposon family of maize. Genome **46:** 745–752.

Smouse, P. E., J. C. Long and R. R. Sokal, 1986 Multiple regression and correlation extensions of the Mantel test of matrix correspondence. Syst. Zool. **35:** 627–632.

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

Wilkes, H. G., 1967 *Teosinte: The Closest Relative of Maize.* Bussey Institute, Harvard University, Cambridge, MA.

Wilkes, H. G., 1977 Hybridization of maize and teosinte in Mexico and Guatemala and the improvement of maize. Econ. Bot. **31:** 254–293.

Communicating editor: M. Nordborg