PHYLOGEOGRAPHY OF THE WILD SUBSPECIES OF ZEA MAYS

E.S. Buckler IV1,*, M.M. Goodman2, T.P. Holtsford3, J.F. Doebley4, J. Sanchez G.5

1 USDA-ARS, Dept. of Plant Breeding, Cornell University, Ithaca, New York 14850, USA
2 Dept. of Crop Science, North Carolina State University, Raleigh, North Carolina 27695-7614, USA
3 Division of Biological Science, University of Missouri, Columbia, Missouri 65211, USA
4 Dept. of Genetics, University of Wisconsin, Madison, Wisconsin 53706, USA
5 Dept. of Crop Science, North Carolina State University and CUCBA - Universidad de Guadalajara, km 15.5 Guadalajara-Nogales, Las Agujas, Zapopan, Mexico, CP 45110

Received November 29, 2004

* For correspondence (Fax +1 607 255 6249; e.mail: esb33@cornell.edu)

ABSTRACT - Z. mays ssp. parviglumis and ssp. mexicana are the closest wild relatives to domesticated maize. Using isozyme and chloroplast evidence, this study examined how populations of these subspecies are related to one another, and how geography has structured the relationships between them. As some lines of evidence indicate that ssp. mexicana is a derived clade of ssp. parviglumis, dispersal, isolation by distance (IBD), and altitudinal hypotheses were tested to explain the genetic differentiation of ssp. mexicana and parviglumis populations. Simple dispersal hypotheses explained most of this genetic variation, while IBD and altitudinal models explained very little of the variation. The origin of this dispersal appeared to be the middle and lower elevation regions of Guerrero. These dispersal events are discussed in light of Late Pleistocene and Holocene climate change and maize domestication.

KEY WORDS: Zea mays ssp. parviglumis and mexicana; Maize; Holocene/Pleistocene; Phylogeography; Dispersal; Isolation by distance.

INTRODUCTION

A century’s worth of research clearly documents that the teosintes (Zea) are the wild relatives of domesticated maize (reviewed in Doebley, 1990a). More recently, studies conducted over the last three decades have shown that maize is most closely related to the subspecies of Z. mays ssp. parviglumis and mexicana (Kato Y., 1976; Doebley et al., 1987b, 1984; Smith et al., 1984; Buckler and Holtsford, 1996), and as few as five major genetic loci account for the significant taxonomic morphological differences between them (Beadle, 1939; Doebley and Stec, 1991). To date, however, little research has examined the phylogeographic history of the teosintes. Such research would reveal how the populations of Z. m. ssp. mexicana and parviglumis have been shaped by the geography of Central Mexico and by the major climate changes during the Pleistocene and Holocene periods.

In keeping with original taxonomic delineations based on morphological characters of the male inflorescence (Doebley and Iltis, 1980; Iltis and Doebley, 1980), five species of Zea are currently recognized (Doebley, 1990b; Iltis and Benz, 2000). The two perennial species, Z. perennis and Z. diploperennis, are found in very limited regions of the highlands of Western Mexico. The annual Z. luxuriatus is rare and found in the more equatorial and sometimes more mesic environments of Guatemala, Honduras, and Nicaragua. The Nicaraguan Z. luxuriatus have recently been separated into a new but allied species called Z. nicaraguensis (Iltis and Benz, 2000). The species Zea mays contains four subspecies, including domesticated maize (Z. mays ssp. mays) and three wild taxa (Iltis and Doebley, 1980; Doebley, 1990b): (1) Zea mays ssp. parviglumis can be widespread in the middle and low elevations (500-2000m) of southern and western Mexico, (2) The annual Z. m. ssp. huehuetenangensis is limited to a few highlands of northwestern Guatemala and is differentiated from Z. mays ssp. parviglumis by its isozyme constitution and distinct ecogeography (Doebley, 1990b), and (3) Z. m. ssp. mexicana is restricted to highland populations on the Central Plateau of Mexico and is differentiated from Z. mays ssp. parviglumis by smaller male and female spikelets (Iltis and Doebley, 1980).

Current systematic treatments of the genus Zea are based on a wealth of biological evidence, including isozyme, chloroplast, mitochondrial, chromosomal knob, and nuclear ribosomal markers (Kato Y., 1976; Smith et al., 1982, 1984; Doebley et al.,...
1984; Doebley et al., 1987a; Allen, 1992; Buckler and Holtsford, 1996). These taxonomic and phylogenetic studies have consistently supported the differentiation between the various taxa. The two perennial species generally form a monophyletic clade, with Zea luxurians as a sister species. While this latter relationship is clearly suggested and statistically supported by cpDNA (Doebley et al., 1987a), this topology is not consistent with the gene tree produced by nuclear ribosomal markers (Buckler and Holtsford, 1996; Doebley et al., 1984). Some nuclear markers (isozymes, chromosomal knobs and ribosomal DNA) indicate that Z. mays ssp. buebuetenangensis is basal to ssp. parviglumis, mexicana, and mays. In this study, we used both improved sampling of Z. mays ssp. parviglumis and mexicana and phylogeographic analyses to shed more light on the relationship between these two subspecies.

As the distributions of most plants changed substantially throughout the Pleistocene and Holocene, the major climatic changes that marked these time periods may have played an important role in teosinte distributions and evolution. Though at first underestimated (CLIMAP, 1976), more recent evidence suggests that climate changes in the tropics and Central Mexico were in fact quite dramatic (Colinvaux et al., 1996). Within the last 25,000 years, average temperatures may have dropped by as much as 8°C, causing a 1000m depression in glacial and vegetation lines (Brown, 1985; Heine, 1988; Leyden et al., 1994; Buckler et al., 1998). Around 8,000 years ago, the climate appears to have been substantially warmer and wetter in Mexico than at present, although the data supporting this inference is weaker than that documenting large Pleistocene differences in temperature and moisture would necessarily have affected the distributions and, perhaps, the differentiation of teosinte in Mexico.

The geographical ranges of Z. m. ssp. parviglumis and mexicana are almost parapatric, which suggests that clinal processes such as isolation by distance (IBD) or altitude might have been important in differentiating the two subspecies. In contrast, significant climatic changes over the Late Pleistocene and early Holocene may have been involved, as these provided opportunities for many changes in distribution, adaptation, and dispersal. Thus, both dispersal and clinal hypotheses can be expressed as phylogeographic hypotheses (Sokal et al., 1991). In this study, isolation by distance is defined as possible gene flow in all directions, whereby genetic distance is considered to be proportional to the geographic distance directly between constituent populations. By contrast, the dispersal model holds that migrations occur via a dispersal network along a specific set of paths. We also considered a third possible factor affecting distribution—an altitudinal hypothesis based on differences in elevation and the adaptations such differences require. Using matrix correlations (Sokal and Rohlf, 1995), we evaluated the strength of these alternative phylogeographic hypotheses.

Our research addressed three points: (1) How are the populations of Z. mays ssp. parviglumis and mexicana related to one another? Additionally, how are the newly discovered populations related to those previously described (Sanchez et al., 1998)? (2) Relative to continuous gene flow (differentiation due to IBD or altitude), how important were dispersals in structuring Z. mays populations? (3) How might the phylogeography of these subspecies relate to climate change of the last 20,000 years?

**MATERIALS AND METHODS**

*Data collection*

Starch gel electrophoresis was employed to analyze samples of teosinte populations at the isozyme laboratory, Department of Genetics, North Carolina State University (Stuber et al., 1988). Sixty-one teosinte populations were analyzed in this study, including 16 teosinte populations discovered and collected during the past decade (Sanchez et al., 1998) and 45 populations from the Doebley et al. (1984) study.

Test samples from newly collected populations included: eight populations of Z. mays ssp. parviglumis, seven populations of Z. mays ssp. mexicana, and one population of Z. luxurians. Taxonomic designations were based on morphological characteristics (Jets, 1972; Sanchez et al., 1998). With the exception of Z. luxurians found in Nicaragua, all newly collected populations were from locations in Mexico. More specifically, these populations were from the following areas: Z. luxurians: 1. Nicaragua, 12°54'N, 86°59'W, 8m, Blets 30919; Z. mays ssp. parviglumis: 2. San Cristobal Honduras, 16°20'N, 97°2'W, 1120m, JSG-197; 3. Paso, 18°17'N, 99°11'W, 1225m, JSG-Y-305; 4. Malinalco, 18°57'N, 99°30'W, 1850m, JSG-Y-LOS-159; 5. Amatlán, 18°59'N, 99°2'W, 1700m, JSG-183; 6. Taretan, 19°25'N, 101°55'W, 1180m, JSG-196; 7. Cojumatlán, 20°6'N, 102°25'W, 1700m, JSG-Y-LOS-75; 8. La Cienega, 20°41'N, 104°30'W, 1320m, JSG ET AL.-298; 9. El Tabillo, 20°48'N, 104°33'W, 1090m, JSG-Y-LOS-43; Z. mays ssp. mexicana: 10. Puebla, 19°7'N, 97°38'W, 2425m, JSG ET AL.-316; 11. Toluca, 19°12'N, 99°37'W, 2540m, JSG ET AL.-319; 12. Opozoco, 19°24'N, 101°30'W, 2320m, JSG-Y-194; 13. Chucandiro, 19°54'N, 101°20'W, 1800m, JSG-Y-LOS-53; 14. Copandaro, 19°54'N, 101°10'W, 1825m, JSG-Y-LURD-34; 15. San Jeronimo, 20°24'N, 102°20'W, 1550m, JSG-Y-LOS-4; 16. San Jeronimo, 20°24'N, 102°20'W, 1550m, JSG-Y-LOS-5. Test samples from Doebley et al.

Nine to 50 plants were sampled per population for 21 isozyme loci (DOEBLEY et al., 1984). All isozyme data can be obtained from the web site http://www.panzea.org/. Population designations were based on region, study (Doebley or Sanchez), and sample number, respectively. For example, population AS09 (Fig. 1) is in region A, from the Sanchez sample, and is population number 09 from El Tabillo. A population from the DOELEY et al. (1984) sample is designated with a D instead of an S.

Chloroplast haplotypes were scored in 61 collections of ssp. parviglumis and mexicana. Restriction length polymorphisms were mapped in all populations using the methods previously described (DOEBLEY et al., 1987a; DOEBLEY, 1990a). A complete listing of these data can be obtained from http://www.panzea.org/. The five haplotypes were made into a network by connecting haplotypes that differ by only one restriction site or length polymorphism (TEMPLETON et al., 1992).

Phylogenetic analysis

Continuous character maximum likelihood (ContML) was used for phylogenetic reconstructions (FELSENSTEIN, 1981, 1985b). Since this approach models drift with a Brownian motion model, it is appropriate for the closely related taxa of Zea mays. As ContML does not model mutation, the Cavalli-Sforza chord distances, essentially a square root transformation distance, were also calculated between all pairs of taxa (FELSENSTEIN, 1989), and Fitch clustering (least-squares criterion) was then used to cluster the taxa as implemented by PHYLIP (FELSENSTEIN, 1989).

For both methods, ten random addition orders were tried, and global swapping was used to find the best tree. One hundred bootstrap repetitions of the ContML and Cavalli-Sforza chord distances, with Neighbor Joining (NJ) trees estimated the strength of support for the various branches (FELSENSTEIN, 1985a). In the bootstraps, loci were randomly sampled with replacement. Statistical significance of the phylogeny was further examined by bootstrapping a phylogeny based on average allele frequencies from the major geographic regions of ssp. parviglumis and mexicana. An unweighted average of all accessions within a region was used to estimate the average allele frequency for the region.

Principal Components: Principal component analysis of the isozyme data was used to identify the orthogonal vectors that may represent different dispersal episodes (CAVALLI-SFORZA et al., 1994). Principal component analysis was implemented by the SAS system using the covariance matrix across accessions (SAS).
Allele frequencies were not standardized. The first and second principal components were estimated for each of the ssp. *parviglumis* and *mexicana* populations. Principal component scores for the outgroups (*Z. luxurians* and *Z. m. ssp. huehuetenangensis*) were estimated using the eigenvectors derived from the joint *parviglumis* and *mexicana* analysis. This is a reasonable approach for these outgroups, as they contain relatively few private alleles, and these are almost always in low frequency.

**Geographic Analyses:** Dispersal, isolation by distance (IBD), and altitude hypotheses for the populations of ssp. *mexicana* and *parviglumis* were tested by comparing the genetic distances between the taxa with the geographic distances between the populations. We converted all hypotheses into distance matrices (SOKAL and ROHLF, 1995). In all tests, the genetic distance matrix was the Cavalli-Sforza chord distances based on the ssp. *parviglumis* and *mexicana* isozyme allele frequencies. We then calculated matrix and partial correlation coefficients according to SMouse et al. (1986). The significance of these correlations was estimated through 10,000 matrix permutation tests (Dietz, 1983; SMouse et al., 1986). Since inner correlations may exist, the P values should be considered approximate. Correlations were compared with tests of homogeneity (Sokal and Rohlf, 1995).

The geographic hypotheses were developed with the help of Phylogeographer 1.0 software package (Buckler, 1999). Dispersal distances were calculated as the Great Circle Distance through all nodes by which two taxa are connected. For example: in Fig. 1, the dispersal distance between CS03 and DS05 is the sum of the 73km from CS03 to C, 109km from C to D, and the 44km from D to DS05 for a total of 226km. Distances for the clinal hypotheses are simply the Great Circle Distance between the given two taxa; therefore, the clinal distance between CS03 and DS05 is 79km.

Because a nearly infinite number of dispersal scenarios could be constructed, we used the following guidelines to develop the hypotheses and reduce multiple test issues. Each subspecies was first divided into four arbitrary geographic locations. All taxa within these locations were then connected to a node located in the geographic center of the location. For each subspecies, only those hypotheses that connected adjacent locations were tested (3 for ssp. *parviglumis* and 3 for ssp. *mexicana*), while all possibilities (16 total) were tested between the two subspecies.

**RESULTS**

**Phylogenetic analysis**

Phylogenetic analysis by continuous character maximum likelihood (ContML) and by Fitch clustering of Cavalli-Sforza chord distances produced similar phylogenetic results. Both these analyses clearly indicated that *Z. mays* ssp. *parviglumis* and *mexicana* form a clade (Fig. 2) that was supported by 72% of the bootstraps. Most Chalco (E) and Central Plateau (F and G) ssp. *mexicana* also form a clade, thus supporting a single origin for ssp. *mexicana*. Central Balsas (B), Jalisco (A), and Oaxaca (BS02)
ssp. parviglumis taxa all form a consistent clade. In both analyses, some of the eastern range ssp. parviglumis (regions C and D of the eastern Balsas) were basal to ssp. mexicana, while ContML analysis indicated some C and D taxa were basal to both ssp. mexicana and the A and B clades of ssp. parviglumis. This nesting of ssp. mexicana within ssp. parviglumis populations was an unexpected result. Bootstrapping these 21 loci with all individual taxa did not provide strong support for any single node within the ssp. parviglumis and mexicana clade (<50%).

Prompted by the overall good clustering of taxa from similar geographic regions on the same parts of the phylogeny, we determined the statistical support for the various major branches within this phylogeny by averaging individual accessions from each geographical region. In this simplified ContML phylogeny (Fig. 3), parviglumis is paraphyletic to a mexicana clade. This mexicana clade is well supported by 81% of the bootstrap replications. The clade defined by ssp. parviglumis populations C and D and ssp. mexicana had weak bootstrap support, and non-significant support based on the KISHINO-HASEGAWA test (1989). Preliminary review of the bootstrapped trees indicated that this low bootstrap support was the product of various parviglumis branches swapping positions. During a post hoc examination of the bootstrap data, however, 97% of the bootstrapped trees had parviglumis populations basal and paraphyletic to the mexicana populations. These clade results support the hypothesis that ssp. mexicana is derived from ssp. parviglumis. Virtually the same trees and bootstrap support were found when Cavalli-Sforza distances and Neighbor Joining were used to evaluate significance.

It should also be noted that a Z. luxurians population recently sampled from Nicaragua (S01) appears to be allied with previously known Z. luxurians; it is, however, distinct from other Z. luxurians populations in Guatemala (Fig. 3). If the tree is rooted with Zea diploperennis, S01 forms a clade with the other Z. luxurians (tree not shown). This finding is congruent with the new species designation of Zea nicaraguensis for S01 (Litis and Benz, 2000).

**Principal components analysis**

Although phylogenetic analysis is useful for representing bifurcations between taxa, it is possible that gene flow between populations makes a tree unrepresentative of the true population structure. Principal components analysis was used to examine the two largest principal components involved in teosinte population structure. The first principal component explained 26% of the variation between ssp. parviglumis and ssp. mexicana, separating the populations (Fig. 4). Most parviglumis populations from A and B were discrete from all other taxa, but parviglumis populations from C and D were part of a continuum that included ssp. mexicana. The second principal component explained 12% of the variation and divided mexicana into two main groups: the north and western populations of regions G and H versus the central and eastern populations of regions E and F. Some of the mexicana populations from E and F were intermixed with parviglumis populations from C and D.

By applying principal component scores to the outgroups, we estimated an approximate root for these principal components. The nine outgroups comprised of Z. luxurians and Z. m. ssp. huehuetenangensis populations had first principal component scores that ranged from -0.43 to 0.20, while
second principal component scores ranged from -0.45 to -0.12. These scores suggested, but did not confirm, that the *parviglumis* populations in the C region and some of the *mexicana* populations in the F region were most likely the basal groups of the *mexicana/parviglumis* clade.

**Geographic analysis**

The main alternative to dispersal hypotheses is the IBD hypothesis, which considers ssp. *parviglumis* and *mexicana* populations as parts of a cline that extends from 500m to 2500m high regions in central and southwestern Mexico. In this hypothesis, gene flow occurs in all directions, and thus genetic distance should be proportional to the geographic distance directly between constituent populations, and not via the dispersal network depicted in Fig. 1 where migrations occur along a specific set of paths. Given the tremendous elevation differences between populations, it is also important to consider altitude as part of the cline.

To compare these alternative hypotheses, we chose to relate the genetic distances between ssp. *parviglumis* and ssp. *mexicana* to three specific geographic models, including dispersal, isolation by distance (IBD), and altitudinal models. The dispersal model was converted to a distance matrix by calculating the physical distance along each proposed dispersal route between every pair of taxa. The IBD model was converted to a distance matrix by calculating the linear distance between every taxa. The altitudinal model was converted to a distance matrix by calculating the difference in altitude between every taxa. The altitudinal model was converted to a distance matrix by calculating the difference in altitude between all pairs of taxa. Overall, the data and hypotheses were converted into four matrices: a genetic distance matrix, a dispersal matrix, an IBD matrix, and an altitude matrix.

For this geographic analysis, taxa were grouped into eight regional clusters (A-H) and connected by routes based on geographic position and altitude (Fig. 1). First, dispersal and clinal (IBD and altitude) hypotheses were examined for the populations within each subspecies. Four ssp. *parviglumis* clusters (A – D) were formed, and geographical proximity suggested three possible dispersal patterns. Model A-B-C-D (r=0.38) correlated with genetic distance better than the other models (A-B-D-C, r=0.32; A-B/C/B-D, r=0.31). This A-B-C-D model was also signif-
significantly better than the IBD model \((r=0.22)\); as suggested by a test of homogeneity \((P=0.02)\). For the four \textit{ssp. mexicana} populations, geographic position suggested the dispersal model \textit{E-F-G}, but relationships with the outliers of \textit{H} were unclear. The \textit{E-F-G/F-H} model correlated \((r=0.30)\) marginally better with genetic distance than did the \textit{E-F-G-H} model \((r=0.29)\), and somewhat better than the \textit{H-E-F-G} model \((r=0.24)\). The best dispersal model was not statistically better than the IBD model \((r=0.28)\).

Since \textit{ssp. parviglumis} and \textit{ssp. mexicana} have substantial altitudinal variation and numerous potential points of dispersal, all possible dispersal routes between the populations were tested. All of these dispersal route models \((0.38 \leq r \leq 0.54)\) correlated significantly better with genetic distance than did the IBD model \((r=0.23)\) (Table 1). Altitude explained a significant portion of the variation in genetic distance \((r=0.40)\) (Table 1). The effect of geographically distant outliers was tested by excluding the Nabogame (HD55, HD56), Durango (HD45), and Oaxaca (BS02) populations. These outliers, however, had virtually no effect on the correlations.

To determine the relative importance of the dispersal versus clinal (IBD and altitude) hypotheses, partial regression with three or four matrices was used in evaluating whether the first hypothesis explains genetic variation that is not explained by the alternative hypotheses. The basic dispersal and IBD hypotheses were highly correlated \((r=0.66)\), while the basic dispersal and altitude hypotheses were more weakly correlated \((r=0.21)\). The partial regressions indicated that the dispersal hypothesis explained significant genetic variation in addition to that explained by the IBD hypothesis (Table 1). In contrast, the IBD hypothesis explained only a small portion of variation once the effect of the dispersal hypothesis was removed (Table 1). Thus, the IBD hypothesis explained little not already explained by the dispersal theory. In contrast, the variation in genetic distance explained by altitude was substantially unrelated to the dispersal and IBD models (Table 1).

The optimal dispersal route was explored by ex-

### TABLE 1 - Correlation \((r)\) of geographic models with genetic distances, and the probabilities of obtaining these correlations by chance \((P)\).

<table>
<thead>
<tr>
<th>Model</th>
<th>(r)</th>
<th>(P^2)</th>
<th>Model</th>
<th>Partial</th>
<th>(r)</th>
<th>(P^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{Geographic models for Z. mays ssp. mexicana}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IBD</td>
<td>0.28</td>
<td>0.076</td>
<td>IBD</td>
<td>Dispersal</td>
<td>-0.11</td>
<td>0.166</td>
</tr>
<tr>
<td>Dispersal</td>
<td>0.30</td>
<td>0.054</td>
<td>Dispersal</td>
<td>IBD</td>
<td>0.16</td>
<td>0.032</td>
</tr>
<tr>
<td>\textit{Geographic models for Z. mays ssp. parviglumis}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IBD</td>
<td>0.22</td>
<td>0.040</td>
<td>IBD</td>
<td>Dispersal</td>
<td>-0.24</td>
<td>0.024</td>
</tr>
<tr>
<td>Dispersal</td>
<td>0.38</td>
<td>0.000</td>
<td>Dispersal</td>
<td>IBD</td>
<td>0.39</td>
<td>0.000</td>
</tr>
<tr>
<td>\textit{Geographic models for Z. mays ssp. parviglumis and ssp. mexicana}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Altitude</td>
<td>0.40</td>
<td>0.000</td>
<td>Altitude</td>
<td>IBD</td>
<td>0.39</td>
<td>0.000</td>
</tr>
<tr>
<td>IBD</td>
<td>0.23</td>
<td>0.012</td>
<td>IBD</td>
<td>Altitude</td>
<td>0.22</td>
<td>0.022</td>
</tr>
<tr>
<td>Dispersal</td>
<td>0.51</td>
<td>0.000</td>
<td>Dispersal</td>
<td>IBD</td>
<td>0.49</td>
<td>0.000</td>
</tr>
<tr>
<td>Dispersal</td>
<td>Altitude</td>
<td>0.47</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 Model | Partial indicates the partial correlation of the Model with genetic distance when the Partial model is held constant.
2 Based on 10,000 permutations of the matrix. \(P\)-values are nondirectional.
3 This is the dispersal model pictured in Fig. 1. The \textit{ssp. parviglumis} models examine nodes A, B, C, and D, while the \textit{ssp. mexicana} models only examine nodes E, F, G, and H.

### TABLE 2 - Comparison of various dispersal models between the \textit{ssp. parviglumis} and \textit{ssp. mexicana} ranges, while using partial correlations to hold IBD and altitude constant.

<table>
<thead>
<tr>
<th>(Z. m.) ssp. mexicana</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{Z. m. ssp. parviglumis}</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0.38*</td>
<td>0.26*</td>
<td>0.20*</td>
<td>0.32*</td>
</tr>
<tr>
<td>B</td>
<td>0.42*</td>
<td>0.27*</td>
<td>0.26*</td>
<td>0.38*</td>
</tr>
<tr>
<td>C</td>
<td>0.45*</td>
<td>0.38*</td>
<td>0.36*</td>
<td>0.40</td>
</tr>
<tr>
<td>D</td>
<td>0.45*</td>
<td>0.38*</td>
<td>0.36*</td>
<td>0.40</td>
</tr>
</tbody>
</table>

* indicates the partial correlation was significantly worse than the optimal Fig. 1 dispersal model at \(P<0.05\) by a test of homogeneity.
Table 3 - Correlation between outgroup distance and dispersal distance.

<table>
<thead>
<tr>
<th>Node</th>
<th>Z. luxurians</th>
<th>Z. m. ssp. huehuetenangensis</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.54</td>
<td>0.41</td>
</tr>
<tr>
<td>B</td>
<td>0.47</td>
<td>0.33</td>
</tr>
<tr>
<td>A</td>
<td>0.28</td>
<td>0.22</td>
</tr>
<tr>
<td>D</td>
<td>-0.03*</td>
<td>-0.03*</td>
</tr>
<tr>
<td>E</td>
<td>-0.11*</td>
<td>-0.09*</td>
</tr>
<tr>
<td>F</td>
<td>-0.42*</td>
<td>-0.28*</td>
</tr>
<tr>
<td>G</td>
<td>-0.43*</td>
<td>-0.30*</td>
</tr>
</tbody>
</table>

* indicates the correlation was significantly less than the Node C correlation (P<0.05) by a test of homogeneity.

To identify the origin of this dispersal, an island hopping model was considered where genetic distance from an outgroup should be proportional to geographic distance from the origin of the dispersal. This assumption is based on the concept that drift associated with migration from the center of origin would add to genetic distance, while populations remaining near the origin would experience less drift and thus exhibit shorter genetic distances from the outgroup. Correlations were calculated between (1) the genetic distance between each taxa and the outgroup, and (2) the corresponding geographic distance between each taxa and the different nodes tested (A, B, etc.). Outgroups Z. m. ssp. huehuetenangensis and Z. luxurians were used to estimate the source of this dispersal. All central geographic nodes (Fig. 1) were tested as possible nodes of origin. Nodes C and B had a strong positive correlation with genetic distance and the geographic point of origin, while the remaining nodes were significantly weaker. If these 800-1600m altitude populations persist near their pre-dispersal ranges, this analysis suggests the point of dispersal origin was in the lower and middle altitude regions of the central Balsas valley.

Chloroplast haplotypes
Five chloroplast haplotypes were identified among ssp. parviglumis and mexicana populations (Table 4). A haplotype network can be developed in which each haplotype is differentiated by a single substitution (Fig. 5). Haplotype 1 appears to be basal according to outgrouping with Z. luxurians, Z. diploperennis, and Z. perennis (Doebley et al., 1987a; Doebley, 1989). Haplotype 2 is only found in the outgroup Z. mays ssp. huehuetenangensis. Based on connections with other parts of the tree, haplotypes 1 and 2 are most likely ancestral, while haplotypes 3, 4, and 5 were derived more recently. A Fisher’s Exact test indicates a significant difference (p=0.018) in frequency of ancestral haplotypes 1 and 2 between ssp. mexicana (7 of 29) and ssp. parviglumis (1 of 32). Derived haplotypes 4 and 5 are only found in ssp. parviglumis, while haplotype 3 is ubiquitous.

Discussion
How are ssp. parviglumis and mexicana related to one another? In general terms, the results of our analyses confirm previous reports that these two

Table 4 - Distribution of chloroplast haplotypes among clades.

<table>
<thead>
<tr>
<th>Geographic Regions and Taxa</th>
<th>Z. mays ssp. parviglumis</th>
<th>Z. mays ssp. mexicana</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haplotype</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>1</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>2</td>
<td>–</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>–</td>
<td>3</td>
</tr>
</tbody>
</table>
subspecies can be differentiated and are largely non-overlapping (Doebley, 1990a; Doebley et al., 1984, 1987b). The increased sampling of *parviglumis* from the eastern side of its range provided more examples of populations that are genetically intermediate between central Balsas *parviglumis* (C) and ssp. *mexicana*.

The resulting clade structure may reflect a consistent gene flow pattern or be the product of divergence and monophyly.

While previous analyses could be interpreted as suggesting monophyly for both taxa, our analyses indicate that ssp. *parviglumis* is paraphyletic to ssp. *mexicana*. The present phylogenetic analyses using ContML and Fitch indicate that most *mexicana* populations are more diverged from the outgroup than are most *parviglumis* populations, suggesting that ssp. *mexicana* is an offshoot of the eastern *parviglumis* populations (regions C and D). This discrepancy may stem from the fact that previous studies were mostly taxonomic in approach, employing methods such as average linkage cluster analysis that are sensitive to differences in rates and thus inappropriate for phylogenetic reconstruction. Average linkage clustering forces the long branches of ssp. *mexicana* to be basal. For example, when average linkage clustering is applied to the current data set, the long branch taxon of AD35 and FD54 become basal, eventually followed by the split of ssp. *mexicana* and *parviglumis* (results not shown). As there is no reason to believe that population size, and thus the rate of drift and branch lengths, should be the same among populations, rate-insensitive approaches (ContML and Fitch) are more appropriate for phylogenetic analyses.

Our phylogenetic results also suggest that ssp. *mexicana* should be considered a lineage emerging from the variation of ssp. *parviglumis*. Bootstrapping of the isozyme data found 97% support for ssp. *parviglumis* being basal and paraphyletic to the ssp. *mexicana* clade, although it was unclear which ssp. *parviglumis* population in particular was immediately basal. Subspecies *mexicana* appeared to be most closely allied with eastern ssp. *parviglumis* populations. If this topology is correct, these eastern populations may show the widest range of molecular and phenotypic diversity, and thus warrant more focus in future analyses.

As many phylogenetic situations can be influenced by reticulate evolution, this possibility must be considered in regards to the analysis presented above. For example, extensive introgression of maize into ssp. *mexicana* might have produced the observed pattern of ssp. *mexicana* derivation from ssp. *parviglumis*. Such introgression will remain a possible explanation for this new phylogenetic position of ssp. *mexicana*.

What historical events and/or processes produced the current distribution? The current distribution of *parviglumis* and *mexicana* populations suggests...
that these populations at one time may have covered a contiguous area of Mexico, with *mexicana* at the higher end of a cline and *parviglumis* occupying the lower end of the cline. When the IBD hypothesis was tested, however, only a modest correlation with genetic distances resulted, thus countering the argument that clines or continuous gene flow were important factors in shaping teosinte variation. In contrast, the dispersal model explained a great deal of genetic variation, indicating that dispersal played the dominant role in the distribution of teosinte populations.

While the geographic matrix comparisons robustly indicate that dispersal, rather than clinal (IBD and altitudinal) models, are responsible for teosinte distribution, certain limitations inherent in using geographic matrix correlations should be acknowledged. First, the simple spoke design within each node is probably unrealistic; rather the IBD model is most likely more accurate at the smallest scales. These differences in scale most likely contribute to the slight negative correlations obtained from partial regressions (Table 1). Secondly, these different multiple scales of migration can produce inner matrix correlations, which means the statistical significance must be considered approximate.

If dispersal events, and not gene flow, resulted in the current distribution of these populations, then where and when did these populations disperse? Phylogenetic and geographic analysis indicate various *parviglumis* populations were basal to *mexicana*, while principal component analysis and some chloroplast data suggest that eastern *parviglumis* and perhaps a couple *mexicana* populations were most similar to the outgroup. This implies that present-day *parviglumis* and *mexicana* populations dispersed from the low and middle elevations to the highland areas now occupied by ssp. *mexicana*. Below we develop three hypotheses of when and how *mexicana* may have diverged from *parviglumis*.

The first hypothesis (non-adaptive model) theorizes that ssp. *mexicana* dispersed from lowland populations during the Late Pleistocene-Early Holocene. At the time of the last glacial maximum (18,000 BP), central Mexico was much colder and drier than it is now. These colder temperatures (between 5-8°C lower than present) were accompanied by 1000m depressions in vegetation zones (Brown, 1985; Heine, 1988; Colinvaux et al., 1996; Buckler et al., 1998). Based on current elevations of between 2600 and 400m for ssp. *mexicana* and *parviglumis* populations, a 1000m-vegetation depression would place most of these populations between 1600m and sea level. Since many modern populations grow above this 1600m line (Fig. 1), dispersal events during the early Holocene could have moved populations from lowland to more highland sites. This suggests the most hospitable region for teosinte was probably deep in the Balsas Valley or along the Guerrero coast in a lowland refuge. As the early part of the Holocene (10,000-7000BP) was probably warmer and wetter than present, conditions may have been apt for ssp. *mexicana* and *parviglumis* to leave these areas and spread throughout Central Mexico. This lowland refuge theory is supported by the isozyme data, which suggests that teosinte dispersed from the lowland nodes B or C (Table 3). These regions also have some of the highest isozyme diversity (Doebly et al., 1987b), which supports the lowland refuge concept.

An alternative hypothesis (adaptive model) proposes that ssp. *mexicana* has adaptations to the cooler upland regions, thus enabling it to persist in the uplands during full glacial periods. Under this hypothesis, ssp. *mexicana* divergence could be older than the early Holocene warming. Support for this hypothesis includes: (1) the most primitive cpDNA type (S cytoplasm) is known only from ssp. *mexicana* (Doebly et al., 1987a; Doebly and Nabhan, 1989; Doebly and Sisco, 1989; Allen, 1992), with other ancestral haplotypes significantly more common in ssp. *mexicana* as well. Such a pattern is inconsistent with a simple origin of *mexicana* from within *parviglumis*. This pattern could also be produced if the ancestral haplotypes were adaptive to cooler regions, but there is currently no evidence to support this; (2) ssp. *mexicana* exhibits darker anthocyanin coloration and dense pubescence, both likely adaptations to cooler temperatures (Doebly, 1984). These morphological adaptations may have permitted *mexicana*’s ancestors to exist in the highlands during the colder Late Pleistocene; (3) domesticated maize is grown in the high Andes of Peru, where maximum temperatures are about 6°C colder than those found in highland Mexico today. This suggests that some *Zea* taxa can rapidly adapt to cooler regions; and (4) although crossing barriers between subspecies of *Zea mays* have not been fully elucidated, some evidence exists for partial crossing barriers between ssp. *mexicana* and *parviglumis* (Kermicle, 1997).

Finally, a mixture of these hypotheses might also be true. For example, some populations may have diverged early and colonized the highland regions.
To survive and persist, these highland populations necessarily adapted to cold climate during various periods of the Pleistocene and thus came to exhibit a range of cold-adapted morphologies. Then, during the early Holocene, \textit{parviglumis}-like lowland germplasm could have dispersed into the highlands. Introggression of the highland germplasm by the lowland and mid-elevation germplasm during the early Holocene could have produced the \textit{ssp. mexicana} of today. This combined model could explain the cold adapted morphologies of \textit{mexicana}, the ancestral chloroplast haplotypes of \textit{mexicana}, and the phylogenetic result of \textit{mexicana} being nested within \textit{ssp. parviglumis}.

Evaluation of these differing hypotheses will require high-resolution paleobotanical evidence from the Late Pleistocene. In addition, extensive population-level nucleotide data from numerous populations could be used in coalescent simulations to compare the three hypotheses.

The Nabogame, Durango, and Oaxaca populations are all at least 500km from the central part of the range. The mechanism of dispersal to these regions is unclear, but humans and waterfowl should be considered as possible dispersal agents. The Nabogame and Durango populations are both on the Western flight path for waterfowl, and both of these taxa are related to \textit{mexicana} populations in the Central Plateau - an area with many lakes. Oaxaca has an extensive history of human habitation and \textit{Zea} exploitation. While feasible, these unique dispersal events will be difficult to evaluate. The distribution of these isolated populations might also be due to population fragmentation, a hypothesis that could be evaluated using paleobotanical surveys.

\textbf{How does current distribution relate to maize domestication?} Although humans had traveled to the Pacific coast of South America by 13,500-12,500 BP (Dillehay, 1989), the earliest archaeological remains for maize date between 7,000 and 5,500BP for Mexico and South America (Long et al., 1989; Peaarsall, 1995). This timeline suggests that maize was being domesticated at roughly the same time as major climatic changes were occurring in Mexico. If the non-adaptive dispersal model is correct and domestication began before 10,000 BP, then early humans likely encountered the teosintes in lowland and perhaps even coastal locations. In keeping with this model, archaeological evidence for domestication should be found in lower reaches of the Balsas or the coast of Guerrero. However, if the adaptive hypothesis is correct or if domestication occurred in the early Holocene, then archaeological evidence is more likely to be found in mid-elevation regions of Mexico. Paleobotanical surveys examining the distribution of teosintes might greatly assist archaeological surveys attempting to unlock when and where maize was domesticated.

\section*{Conclusions}

The analyses presented suggest that \textit{Z. m. ssp. mexicana} dispersed from eastern \textit{ssp. parviglumis} populations. Geographic dispersal hypotheses were clearly supported over competing IBD and altitudinal hypotheses using distance matrix correlation approaches. Isozyme evidence is congruent with this dispersal occurring during the warm early Holocene, while chloroplast evidence and ssp. \textit{mexicana} adaptations suggest that earlier Pleistocene divergences are clear alternatives.

\section*{Acknowledgements}

This research was supported by National Science Foundation grant DBI-0321467 and by the United States Department of Agriculture-Agricultural Research Service. The authors would like to thank N. Stevens, M. Fabiano and C. Keith Buckler for their comments on various versions of this manuscript.

\section*{References}


Buckler E.S. IV, 1999 Phylogeographer: A tool for developing and testing phylogeographic hypotheses, 0.3 ed, http://statgen.ncsu.edu/buckler/.


