Early Allelic Selection in Maize as Revealed by Ancient DNA

Viviane Jaenicke-Després,1 Ed S. Buckler,2 Bruce D. Smith,3 M. Thomas P. Gilbert,4 Alan Cooper,4 John Doebley,5 Svante Pääbo*1

Maize was domesticated from teosinte, a wild grass, by ~6300 years ago in Mexico. After initial domestication, early farmers continued to select for advantageous morphological and biochemical traits in this important crop. However, the timing and sequence of character selection are, thus far, known only for morphological features discernible in corn cobs. We have analyzed three genes involved in the control of plant architecture, storage protein synthesis, and starch production from archaeological maize samples from Mexico and the southwestern United States. The results reveal that the alleles typical of contemporary maize were present in Mexican maize by 4400 years ago. However, as recently as 2000 years ago, allelic selection at one of the genes may not yet have been complete.

The wild grass, teosinte (Zea mays ssp. parviglumis), from which maize (Zea mays ssp. mays) was domesticated, is endemic to southern and western Mexico (J). The earliest undisputed archeological evidence of domesticated maize is 6250 years old (2). However, recent molecular data suggest that domestication could have begun as early as 9000 years ago and that the Balsas River Valley in southern Mexico is the likely geographical origin of domestication (3). The early history of character selection in maize is documented in the archaeological record by morphological features discernible in cobs. For example, an increase in the number of rows of kernels and a reduction in glume size have been noted in early maize cobs (4). By 5500 years ago, kernel size had also increased (5). However, nothing is currently known about when characters not observable from the morphology of cobs, such as plant architecture and starch properties, were selected by early farmers.

Recently, a number of genetic loci associated with phenotypic differences between maize and teosinte have been identified (6–9), and three genes involved in such differences have been cloned and relatively well characterized in function (7, 9, 10). In each of these genes, the allelic diversity in maize compared to teosinte has been shown to be reduced, presumably as a result of selection by early farmers. The first gene, teosinte branched 1 (tb1), carries a maize variant that represses the growth in axillary meristems, leading to the unbranched plant architecture typical of maize. It also contributes to the presence of female cobs on the primary branches in maize rather than male tassels as in teosinte (11, 12). The second gene encodes the prolamalin box binding factor (pbf), which is involved in the control of expression of seed storage proteins in the kernel (13–15). Whereas the third gene, sugary 1 (su1), encodes a starch debranching enzyme expressed in kernels (16). Together with branching enzymes, this enzyme determines the structure of amylopectin (16, 17). The chain length of amylopectin, as well as the ratio of amylose to amylopectin, is important for the gelatinization properties of starch (9) and, thus, affects the textural properties of tortillas (18, 19).

Because DNA in archaeological remains is generally degraded to small sizes (20), we sequenced data in each gene that are short enough to allow amplification from ancient corn cobs yet distinguish among the spectrum of gene variants (alleles) found in present-day maize and teosinte (10). For tb1, the allelic variation in contemporary maize and teosinte is well described (10). This allowed us to choose a fragment of 56 base pairs (bp) for which maize carries a single allele, Tbl-M1; this allele has a frequency of 36% in teosinte, where a total of six additional alleles exist. In order to characterize the contemporary variation in pbf and su1, we sequenced a longer segment of each gene in 66 maize landraces from South, Middle, and North America as well as 23 teosinte parviglumis lines (10). The estimated number of alleles segregating in maize is reduced about threefold at pbf and su1 in comparison to teosinte (Fig. S1, B and C). At pbf, we selected a 25-bp fragment in which the alleles Pbf-M1 and Pbf-M2 are carried in 97% and 3% of maize, whereas the same alleles are carried in 7% and 83% of teosinte, respectively (Fig. 1). At su1, we selected a 60-bp fragment in which two major alleles, Su1-M1 and Su1-M2, are carried in maize at a frequency of 30 and 62%, respectively, whereas they both are carried in teosinte at a frequency of about 7% (Fig. 1; table S2) (10).
Each cob was dated directly using tandem accelerator mass spectroscopy (10). Two cobs from the Ocampo Caves are around 4300 years old, whereas the other three are between 2300 and 2800 years old. Two cobs from Tularosa Cave are around 1900 years old, and the remaining four are between 650 and 900 years old (Fig. 3). From 140 to 200 mg of each cob, we extracted DNA and amplified the fragments containing the diagnostic nucleotide for the alleles at \( tb1 \), \( pbf \), and \( su1 \) (10). DNA sequences were reconstructed from multiple clones derived from several amplification products from each sample, according to established procedures (21), and were compared to modern maize and teosinte alleles (Fig. 3). To verify the results, samples from two cobs (NMNH246279, 314984 cob2) were sent to our Oxford laboratory where DNA was extracted and DNA sequences from all three genes were independently determined. In all cases, the results in Oxford were identical to those in our Leipzig laboratory (10).

At \( tb1 \), the allele \( Tbl-M1 \) common in contemporary maize was found in all 11 cobs analyzed. At \( pbf \), all ancient samples analyzed carried the \( Pbf-M1 \) allele, which occurs in an estimated 97% of modern maize. At \( su1 \), all cobs from Mexico carried the allele \( Su1-M2 \). Among the cobs from the southwestern United States, the four younger cobs carried the \( M1 \) allele, whereas the two 2000-year-old cobs were heterozygous. One of these cobs carried the alleles \( M1 \) and \( M2 \), and the other carried \( M1 \) and the allele \( Su1-T1 \), which has not been seen in contemporary maize but occurs at an estimated frequency of 4% in modern teosinte (Fig. 1). In the case of samples found to be homozygous, heterozygosity cannot be rigorously ruled out because low template copy numbers and primer failure due to unknown alleles are impossible to exclude.

The results show that alleles known to occur in modern maize at the genes \( tb1 \), \( pbf \), and \( su1 \) were already present in maize 4400 years ago in Mexico. This shows that plant morphology as well as biochemical properties of the protein and starch were selected early in the history of maize, and before maize was introduced to the southwestern United States.

It has been suggested that genes which make harvesting of grasses easier may have been selected before genes increasing yield (22). Because the allele \( Tbl-M1 \) at the \( tb1 \) gene causes the female cobs to be attached close to the main axis of the plant, the presence of \( Tbl-M1 \) in maize 4400 years ago is consistent with this suggestion. However, the morphology of corn cobs, ranging from 6250 years old (2) to 5500 years old (5) to the 4400-year-old cobs from the Ocampo Caves analyzed here (23) (Fig. 2), shows that cob size increased continuously during the first 2000 years of selection by early farmers. The fact that \( pbf \) and \( su1 \), involved in protein and starch quality, respectively, carried the alleles typical of modern maize 4400 years ago suggests that kernel quality along with cob size was an early target of selection. However, the lack of available early maize samples means it is not possible to know how complete the selection process was by 4400 to 4300 years ago. In fact, the observation that a 2000-year-old cob from New Mexico carries a \( su1 \) allele which occurs today in teosinte but is very rare or absent in maize indicates that the selection process at \( su1 \) was not complete at that time. It is interesting that this gene may influence the pasting properties of maize starch and that this property, which is of importance for the suitability of the corn for making tortillas, may thus have been selected long after the initial domestication of maize. Analyses of more maize samples are necessary to assess to what extent this is indeed the case.

Maize appears in the archaeological record of the southwestern United States by ~3200 years ago (24). It appears infrequently in the eastern United States from 2100 years ago onward, but it did not become a dominant crop there until ~1200 years ago (25). Today, the allele \( Su1-M1 \) is found only in inbred maize lines that have a substantial contribution from Northern Flint, one of the two parents of modern Com Belt maize (9), whereas \( Su1-M2 \) is present in other maize inbred lines. The absence of \( Su1-M1 \) in the Mexican archaeological samples and the presence of \( Su1-M1 \) at a high frequency in New Mexico 1000 years ago (Fig. 1) suggests that this allele may be closely connected to maize that
later became Northern Flint before it was introduced into the eastern United States. The fact that both 2000-year-old cobs were heterozygous for Sul-M1 may indicate that the predominance of Sul-M1 in North America was established sometime between 2000 and 1000 years ago in the southwestern United States. Analysis of maize remains from additional archaeological sites of this time period will be required to determine when Sul-M1 became predominant in the southwestern United States.

In conclusion, by 4400 years ago, early farmers had already had a substantial homogenizing effect on allelic diversity at three genes associated with maize morphology and biochemical properties of the corn cob. Thus, selection by farmers had profound genomic effects relatively early in the history of this crop. As more genes involved in selected features become identified in maize as well as other crops, the ability to determine nuclear gene sequences from domesticated plants recovered from archaeological excavations will make it possible to follow comprehensively the genetic consequences of domestication over time.

References and Notes

10. Material and Methods are available as supporting material on Science Online.
26. We thank S. E. Nash, J. Haas, D. Martin-Ross, and G. Feinman, Field Museum; E. W. Wood, D. Pfister, and W. Wetterstrom, Harvard University Herbaria, for access to and permission to sample their archaeologi-