Mapping the genes that made maize

JOHN DOEBLEY

George Beadle proposed that the striking morphological differences between cultivated maize and its probable wild progenitor (teosinte) were initiated by a small number of mutations with large effects on adult morphology. Recent genetic analyses using molecular markers provide some support for this view and show where in the maize genome the putative loci are likely to be located. This work sets the stage for fine-scale linkage mapping of these genomic regions and the eventual cloning of the genes involved in this remarkable evolutionary transformation.

Maize or 'Indian Corn' and its wild relatives, the teosintes (*Zea* species), differ profoundly in both ear morphology and plant growth form (Figs 1, 2). Nevertheless, substantial genetic evidence suggests that maize and the Mexican annual teosintes (*Z. mays* ssp. *parviglumis* and ssp. *mexicana*) are members of the same biological species and share a recent common ancestor. In 1939, George Beadle[1,2] reconciled these discrepant observations by proposing (1) that maize is a domesticated form of teosinte and (2) that a small number of major genes selected by the prehistoric peoples of Mexico may have transformed teosinte into maize within the past 10 000 years.

During the next 50 years, Beadle's proposal that maize is a domesticated form of teosinte received strong support from many independent researchers[3], and an alternative hypothesis[4] that maize evolved from a hypothetical wild maize did not. However, comparatively little progress was made toward determining the genetic basis of the morphological differences between these plants. In this article, I will review attempts to determine the number of genes controlling the morphological differences between maize and teosinte and the chromosomal locations of these genes. As will be discussed, recent developments in molecular biology and quantitative genetics have provided some new insights into these old issues.

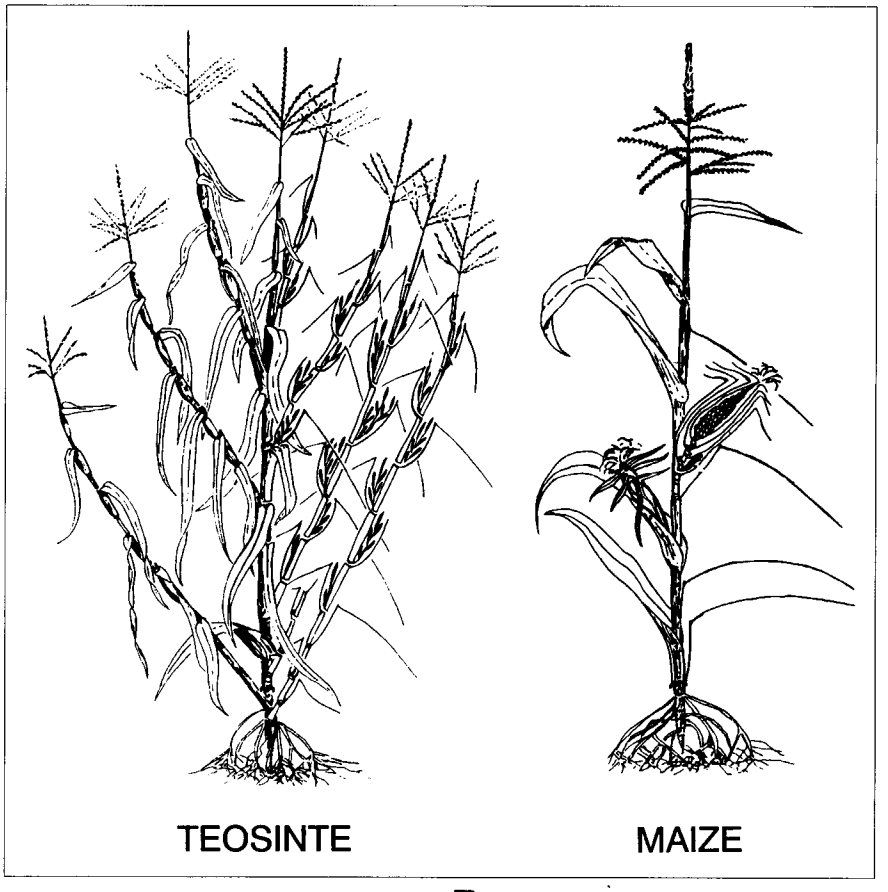


FIG 1

Mexican annual teosinte (left), with its candelabra-like growth form, differs substantially from maize (right) in plant architecture. The long lateral branches of teosinte are each tipped by a tassel, while the short lateral branches of maize are tipped by ears. Adapted from Ref. 19.

The 'key' traits distinguishing maize and teosinte

The morphological differences between maize and teosinte, although complex, can be boiled down to five key traits. First, teosinte plants have elongated lateral branches that are terminated by male inflorescences (tassels; Fig. 1). The teosinte plant therefore looks like a candelabra. In maize, the lateral branches are short and terminated by female inflorescences (ears). Second, the teosinte ear is composed of a series of triangular units called cupulate fruitcases (Fig. 2A, E). Each teosinte cupulate fruitcase contains a single spikelet in which a single kernel develops (Fig. 2C). The corresponding structure in maize, called the cupule, has two (or paired) spikelets from which two kernels develop (Fig. 2D). Third, the teosinte ear bears grains on only two sides of its axis, a condition called two-ranked or distichous (Fig. 2C). In contrast, the maize ear bears its grains in at least four ranks (four sets of cupules), the polystichous condition (Fig. 2D). Fourth, the cupulate fruitcases of teosinte are separated by abscission layers that enable the fruitcases to separate (disarticulate) at maturity for dispersal (Fig. 2A, E). In maize, abscission layers are absent and the ear remains intact at maturity, which allows easy harvest. Fifth, the cupulate fruitcase of teosinte is sealed shut by the outer glume of the spikelet, obscuring the kernel from view (Fig. 2A, E). Both the cupulate fruitcase and this glume become highly indurated (hardened) at maturity, protecting the kernel from predation. In maize, this glume is much softer, and it is smaller

relative to the size of the kernel (Fig. 2F). Thus, the maize kernels are visible on the ear once the husks are removed.

Inheritance of morphological traits in maize–teosinte hybrids

During the past 70 years, several authors have investigated the inheritance of the individual morphological traits that distinguish maize and teosinte. These analyses typically involved crossing teosinte to US maize lines with marker loci, such as sugary kernels (*su1*) or glossy leaves (*gl1*), and then analysing the F_2 or backcross progeny.

Mangelsdorf and Reeves[4] took this approach and identified four chromosomal segments that have major effects on the differences between the maize and teosinte ears. They suggested that each of the four chromosomal segments contained a group of morphological genes rather than a single gene, although they presented no evidence to support this view. Mangelsdorf[5] reported that these four segments probably map to chromosomes 1, 3, 4 and 9, but again he did not show convincingly whether they contain a single or multiple genes.

Langham[6] studied F_2, F_3 and backcross progeny of maize–teosinte hybrids and reported that two traits were each under the control of a single major locus. He named *Tr* as the gene controlling the switch from the two-ranked ears of teosinte to the four-ranked ears of maize and *pd* as the gene converting the single spikelets of teosinte to the paired spikelets of maize. Neither of these genes showed clear linkage to his marker loci, although *pd* was loosely linked to liguleless (*lg2*), a marker on the long arm of chromosome 3. Mangelsdorf[7] and his student, Rogers[8], analysed the inheritance of four-ranked ears, paired spikelets, ear disarticulation and glume induration. These authors were unable to confirm Langham's[6] results regarding the inheritance of paired spikelets and four-ranked ears. They reported that each of the traits is controlled by multiple unlinked genes, and they specifically rejected unifactorial inheritance. The chromosomal locations identified by Mangelsdorf[7] do not always agree with those of Rogers[8]; however, both authors detected a factor(s) on chromosome 4 with a 'strong' effect on glume induration. This result was consistent for different types of teosinte and maize.

The results of these linkage analyses often conflict (Table 1). In my view, there are several reasons for these conflicts. First, different authors used different maize and teosinte parents. The genes differentiating a 12-rowed maize from teosinte are unlikely to be the same as those distinguishing a 20-rowed variety. Second, most authors grew their plants at northern latitudes where, because teosinte is a short-day plant, development is severely affected by improper day length and some genotypes fail to mature. Third, some authors categorized plants as maize-like or teosinte-like for each trait so that 'mendelian' ratios could be calculated, despite the fact that variation for the traits is continuous (in some cases bimodally continuous) and not discrete[7–10]. Forcing such variation into a mendelian pattern is apt to introduce a bias. Finally, various authors did not measure or score the traits in the same fashion, and several authors clearly did not completely understand the morphology of the plant.

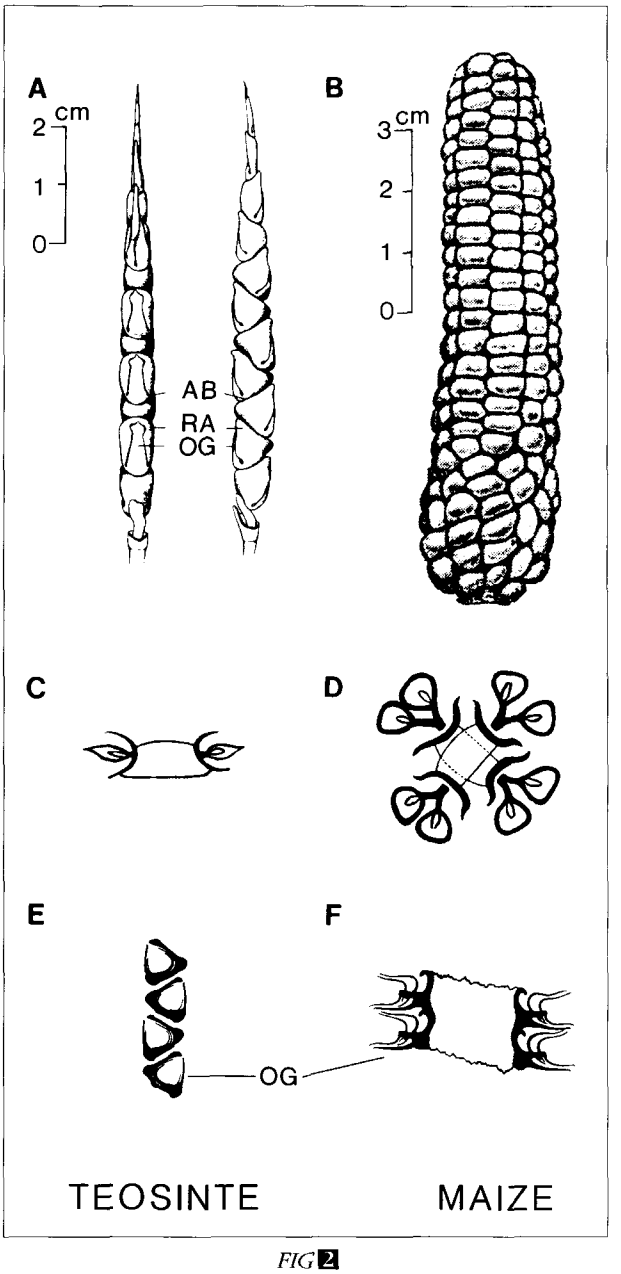


FIG 2

The most striking differences between maize and teosinte are in the architectures of their female inflorescences or ears. (A) Teosinte ear: AB, abscission layer that allows the teosinte ear to shatter when mature; OG, outer glume; RA, rachis internode. (B) Maize ear. (C, D) Schematic transverse cross-sections. (C) Teosinte, showing two ranks of cupules with one kernel per cupule. (D) Maize, showing four ranks of cupules with two kernels per cupule. (E, F) Longitudinal cross-sections. (E) Four teosinte cupulate fruitcases, showing the outer glume oriented upward, parallel to the axis of the ear. (F) Maize, showing the outer glume oriented outward, perpendicular to the axis of the ear. Adapted from Refs 19, 20.

Beadle's experiment: how many genes?

Beadle[2,11] took a wholly different approach and simply attempted to determine the number of independent genes differentiating the maize and teosinte ear morphologies. He grew 50 000 F_2 progeny of a cross of

TABLE 1. Linkages between the traits differentiating maize and teosinte and the marker loci on maize chromosomes

Trait	Chromosome[a]										Refs
	1	2	3	4	5	6	7	8	9	10	
Two- versus four-ranked ears											21
Kempton	+										
Mangelsdorf	+	+	+	+		+		+	+		7
Rogers	+	+	+			+		+		+	8
Galinat		+									17
Single versus paired spikelets											
Kempton	+		+								21
Langham			+								6
Mangelsdorf				+				+			7
Rogers			+				+				8
Mangelsdorf			+								18
Hard versus soft glumes											
Mangelsdorf			+	+		+	+	+	+	+	7
Rogers				+		+	+				8
Mangelsdorf				+							18
Shattering versus solid ears											
Mangelsdorf			+	+		+		+			7

[a]+, Significant linkage reported.

a primitive maize (race Chapalote) and Mexican annual teosinte (ssp. *mexicana* race Chalco). He classified each of the F_2 plants as either maize-like, teosinte-like or intermediate and reported that one in 500 F_2 plants resembled the parental types. Then, applying simple mendelian segregation ratios, he inferred that 'five major and independently inherited gene differences' distinguish primitive maize and teosinte. Beadle recognized that linkage and dominance complicated his interpretation; nevertheless, he believed that he had established that genetic control of the morphological differences was not so complex as to render untenable his hypothesis that maize was derived from teosinte by human selection under domestication.

Beadle's experiment is notable because he avoided two of the pitfalls of previous studies. First, he grew his plants in Mexico, where their development under the proper day-length regime would be normal. Second, he used a primitive Mexican maize, while previous experiments all used elite US agronomic lines as their maize parents. Beadle[2,11] realized that the latter are not appropriate for mapping genes involved in the origin of maize since their second generation hybrids with teosinte will simultaneously segregate for the genes differentiating primitive from advanced maize (Fig. 3).

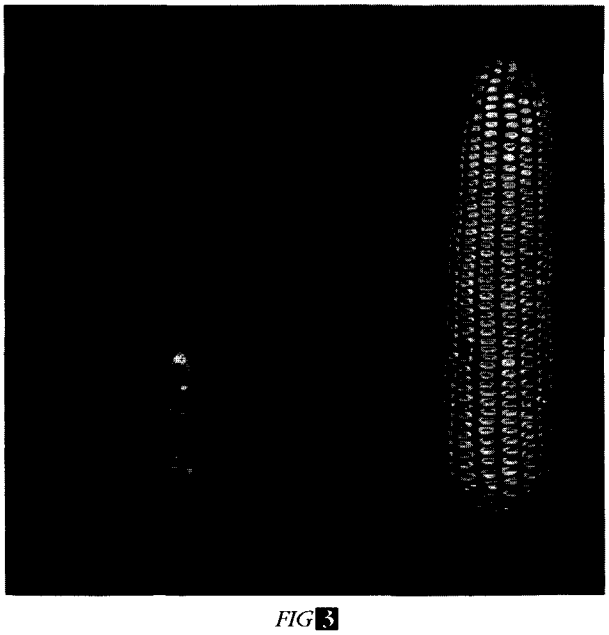


FIG 3

The monstrous ears of elite agronomic varieties (right) are much more strongly differentiated from teosinte (left) than the primitive maize land races Chapalote (center right) and Pollo (center left) of Latin America. Presumably, this reflects continued selective pressure, resulting in more complex genetic differences.

Beadle[2] also attempted to identify a likely candidate for one of his putative five genes. He proposed that the *Tunicate* locus (*Tu1*) of maize, which causes elongation of the glumes, might be the locus that controls the difference between hard and soft glumes. The evidence he cites for this hypothesis is that the oldest archaeological maize was reported to have elongate glumes[12]. He may also have been influenced by the fact that *Tu1* is located on chromosome 4, and Mangelsdorf[7] had shown that a factor affecting this trait in maize–teosinte hybrids maps close to the centromere on this chromosome.

A quantitative genetic approach using molecular markers

Two recent developments have greatly enhanced the ability to map the genes controlling morphological traits. First, the development of saturated linkage maps using restriction fragment length polymorphisms (RFLPs) provides a virtually unlimited number of marker loci throughout the genome[13]. Because RFLP differences exist between virtually all varieties of maize and teosinte, a primitive maize–teosinte population can be used, thus avoiding the complications associated with the use of elite US maize marker stocks. Second, new statistical methods[14,15] for mapping and characterizing quantitative trait loci (QTLs) allow traits to be scored in a quantitative fashion, so continuous variation does not have to be forced into discrete mendelian categories. These statistical methods essentially estimate the probability that there is a QTL at or near each of the marker (RFLP) loci (see Box 1). QTL mapping also provides an estimate of the proportion of the phenotypic variance controlled by

different regions of the genome. By this means, both putative major and minor loci can be identified.

Taking this approach, my colleagues and I essentially repeated Beadle's 1972 experiment with the addition of RFLP markers that enabled us to follow segregation of specific chromosomal segments in our F_2 population[10,16]. We used the same cross as Beadle (race Chapalote maize × race Chalco teosinte). The principal results of our experiment are summarized in Fig. 4. The height of each column corresponds to the percentage of the phenotypic variance explained by each region of the genome. For some traits, such as the number of ranks of cupules around the circumference of the ear, one chromosomal region explains a large portion of the variance (>40%), while several other regions explain much smaller portions. A similar pattern is seen for glume induration and length of the lateral branch (Fig. 4). This is the expected pattern if these traits are controlled by a major locus plus modifiers, as predicted by Beadle[1]. For two other traits, the proportion of single versus paired spikelets in an ear and ear disarticulation, no single region explains more than 24% of the variance and there is gradation from larger to smaller effects. This pattern better fits a multigenic mode of inheritance.

Concordance between our results and those of previous authors is far from complete; however, there are some clear areas of agreement. Our results indicate that chromosome arm 2S has a large effect on the switch from two- to four-ranked, as shown by Galinat[17]. Similarly, our results reveal that chromosome 3L has a large effect on single versus paired spikelets, as shown by Langham[6] and Mangelsdorf[18]. Our results also suggest that chromosome 4 (near *su1*) has a very strong effect on soft versus hard glumes, as previously shown by Mangelsdorf[7] and Rogers[8]. This last result is of interest because we mapped this factor to the short arm of chromosome 4, while *Tu1*, which Beadle thought controlled this trait, is in the middle of the long arm. Thus, our results clearly indicated that *Tu1* is not the gene responsible for controlling the switch from hard to soft glumes.

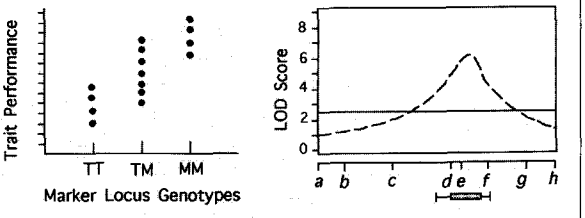
There is also an interesting result concerning a known maize mutant called teosinte-branched (*tb1*) on chromosome arm 1L. This mutant has long lateral branches that are tipped by tassels instead of ears, thus making maize resemble teosinte. In our population, the principal genes controlling these traits also mapped to chromosome arm 1L (Fig. 4). Perhaps in teosinte these traits are controlled by an allele of *tb1*.

Pleiotropy or linkage

An examination of Fig. 4 reveals that chromosomes 1, 2, 3, 4 and 5 are

Box 1. QTL mapping

There are two principal approaches to mapping quantitative trait loci (QTLs). One method[14] uses linear regression to examine the relationship between the performance for the quantitative trait and the genotypes at the marker locus (Fig., left). In the present case, the marker locus



genotypes would be either homozygous teosinte (*TT*) or maize (*MM*), or heterozygous (*TM*). In the figure on the left, each black dot represents a single F_2 plant. If there is a statistically significant association between the trait performance and the marker locus genotypes, one infers that a QTL is located near the marker locus. A more recent method[15], called 'interval mapping', involves calculating the ratio of the likelihood (odds) that there is a QTL to the likelihood that there is no QTL at each position along the length of a chromosome (e.g. from marker loci *a–b* in Fig., right). These values are normally given as the $\log_{10}$ of the 'odds' ratio or LOD score. On the basis of the size of the genome and the number of marker loci analysed, a threshold value (e.g. 2.5 in Fig., right) or significance level is determined. Where the LOD score (dashed line) exceeds the threshold, a QTL is likely.

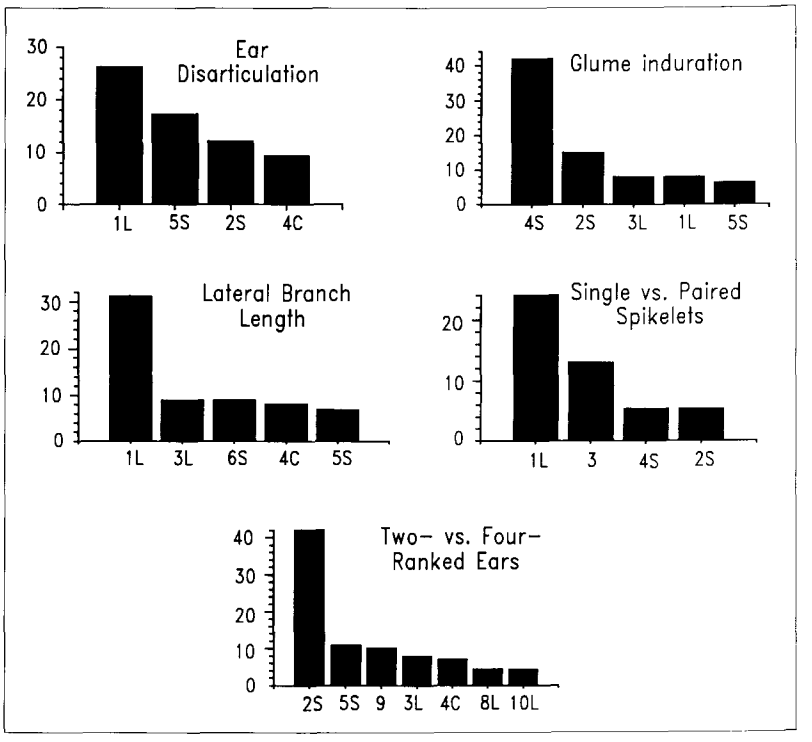


FIG 4

These histograms show the numbers and magnitudes of significant associations between the molecular marker loci and the key morphological traits differentiating maize and teosinte. The heights of the columns represent the percentage of the phenotypic variance explained by a specific region of the genome that is identified below the column by the chromosome number and region (*L*, long arm; *S*, short arm; *C*, near the centromere). The detected QTLs do not explain all the phenotypic variance for any trait since some of the variance will be due either to environmental effects or to QTLs too small to be detected with the number of F_2 plants we analysed.

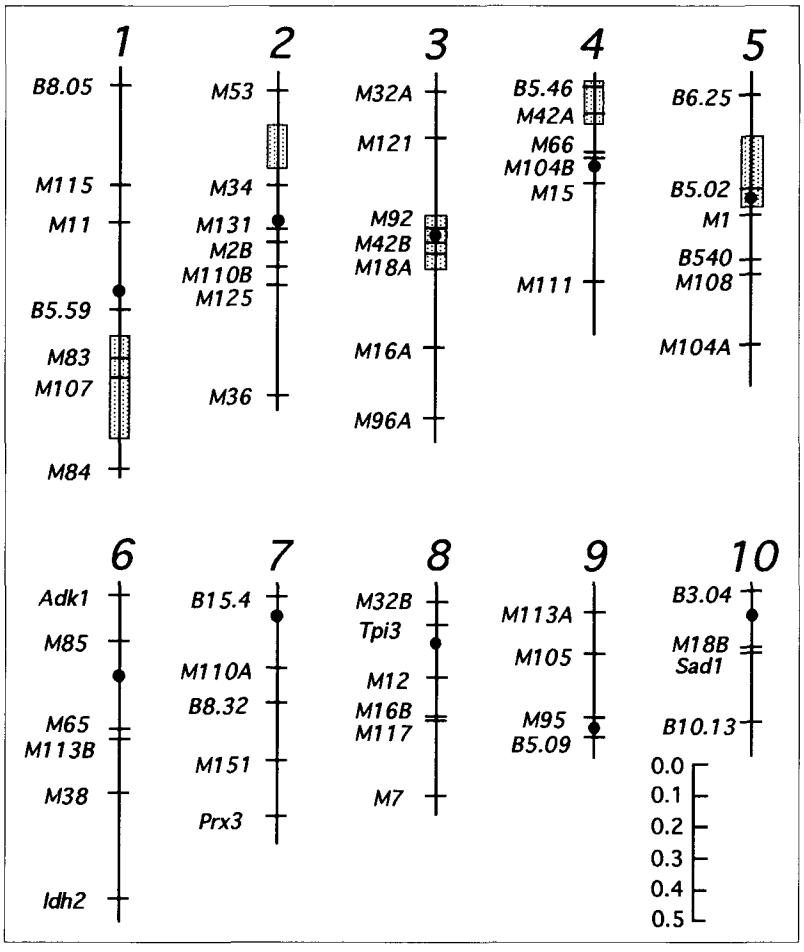


FIG 5

Diagram of the ten teosinte–maize chromosomes showing the five regions (stippled blocks) of the genome with major effects on the morphological differences between maize and teosinte. The cross-lines indicate the positions of molecular marker loci used to map the genes controlling the morphological traits. Distances between the marker loci represent recombination fractions (see scale). Solid circles indicate the approximate positions of the centromeres.

frequently involved in governing the differences between maize and teosinte. When all significant effects are plotted on the maize map, it is clear that the major effects (and most minor effects) map to five restricted regions (Fig. 5). This result helps to explain why Beadle recovered maize-like and teosinte-like segregants at frequencies suggesting that 'five independently inherited gene differences' differentiate maize from teosinte. Three of these regions (1L, 3L and 4S) also correspond to regions identified by Mangelsdorf[5]. Mangelsdorf's fourth region was on chromosome 9, where we found virtually no effect on the differences between the primitive race Chapalote maize and teosinte. Perhaps the region of chromosome 9 identified by Mangelsdorf represents a region distinguishing primitive and his elite US maize parent.

The clustering of most of the effects in five regions of the genome may be interpreted in two ways: either these regions contain single major loci with pleiotropic effects on several traits, or they contain several tightly linked genes each affecting only a single trait. Beadle[2,11] believed the former to be true, and Mangelsdorf[5] the latter. Neither had the data to prove his point. However, this question may be investigated by creating a series of isogenic lines for these regions and then fine mapping the regions. My suspicion is that when this issue is resolved, both linkage and major genes with pleiotropic effects will be shown to be involved.

Conclusions and future prospects

There have been several attempts to identify the genes involved in the morphological evolution of maize. The result has been a series of conflicting reports that have even led one author to conclude that such genes do not exist[19]. My view is that these conflicts result not from the nonexistence of the genes, but rather from the limitations of the attempts to identify them. A lack of adequate marker loci, the use of elite US agronomic lines as the maize parent, and attempts to force continuous variation into discrete mendelian categories have limited the success of these efforts. Recent advances in molecular biology and quantitative genetics have overcome these problems.

Our first efforts with these new approaches seem promising. The picture that has emerged agrees reasonably well with that envisaged by Beadle[1]. We have identified five regions of the genome that control most of the differences between maize and teosinte. Three of these regions contain candidates for major genes controlling three of the five key traits, namely long versus short lateral branches, soft versus hard glumes, and two-ranked versus four-ranked ears. For two other traits, single major loci may not be involved and inheritance may be more aptly described as multigenic; however, I do not believe this damages the essence of Beadle's hypothesis.

Our recent analyses of the genetic differences between maize and teosinte are only first steps. First, it remains to be shown that our results are repeatable. In this regard, we have analysed a second F_2 population derived from different maize and teosinte parents. Again, we found that the same five regions of the genome (Fig. 5) control most of the differences (J. Doebley and A. Stec, unpublished). Second, it needs to be determined whether the five regions contain single major loci with pleiotropic effects or several linked loci. To address this question, we have begun to create isogenic lines for each of these regions and, ultimately, we will fine map them. Third, we need to examine how the genes we have identified alter inflorescence development. This can be accomplished by comparing the course of morphogenesis for the isogenic lines we are creating. Finally, these genes need to be cloned and characterized at the molecular level. In maize, this can be attempted via transposon tagging. If all of this can be accomplished, our understanding of

the morphological evolution of maize may well surpass that of any other plant species.

Acknowledgements

This work was supported by National Science Foundation grants (BSR-88-06889 and BSR-91-07175) and by Pioneer Hi-Bred, Int. of Johnston, Iowa.

References

1 Beadle, G.W. (1939) *J. Hered.* 30, 245–247
2 Beadle, G.W. (1972) *Field Mus. Natl Hist. Bull.* 43, 2–11
3 Doebley, J. (1990) *Econ. Bot.* 44 (3 Suppl.), 6–27
4 Mangelsdorf, P.C. and Reeves, R.G. (1939) *Texas Agric. Exp. Sta. Bull.* 574
5 Mangelsdorf, P.C. (1974) *Corn: its Origin, Evolution and Improvement*, Harvard University Press
6 Langham, D.G. (1940) *Genetics* 25, 88–107
7 Mangelsdorf, P.C. (1947) *Adv. Genet.* 1, 161–207
8 Rogers, J.S. (1950) *Genetics* 35, 541–558
9 Collins, G.N. and Kempton, J.H. (1920) *J. Agric. Res.* 19, 1–37
10 Doebley, J., Stec, A., Wendel, J. and Edwards, M. (1990) *Proc. Natl Acad. Sci. USA* 87, 9888–9892
11 Beadle, G.W. (1980) *Sci. Am.* 242, 112–119, 162
12 Mangelsdorf, P.C., MacNeish, R.S. and Galinat, W.C. (1967) in *The Prehistory of the Tehuacan Valley I* (Byers, D.S., ed.), pp. 178–200, University of Texas Press
13 Helentjaris, T. and Burr, B., eds (1989) *Development and Application of Molecular Markers to Problems in Plant Genetics*, Cold Spring Harbor Laboratory Press
14 Edwards, M.D., Stuber, C.W. and Wendel, J.F. (1987) *Genetics* 116, 113–125
15 Lander, E.S. and Botstein, D. (1989) *Genetics* 121, 185–199
16 Doebley, J. and Stec, A. (1991) *Genetics* 129, 285–295
17 Galinat, W.C. (1973) *Evolution* 27, 644–655
18 Mangelsdorf, P.C. (1952) in *Heterosis* (Gowen, J.W., ed.), pp. 175–198, State College Press, Ames, Iowa
19 Iltis, H.H. (1983) *Science* 222, 886–894
20 Galinat, W.C. (1969) *Mass. Agric. Exp. Sta. Bull.* 577, 1–19
21 Kempton, J.H. (1924) *J. Agric. Res.* 27, 537–596

J. DOEBLEY IS IN THE DEPARTMENT OF PLANT BIOLOGY, UNIVERSITY OF MINNESOTA, ST PAUL, MN 55108, USA.

Master genes in mammalian repetitive DNA amplification

PRESCOTT L. DEININGER, MARK A. BATZER, CLYDE A. HUTCHISON, III AND MARSHALL H. EDGELL

The analysis of species-specific subfamilies of both the LINE and SINE mammalian repetitive DNA families suggests that such subfamilies have arisen by amplification of an extremely small group of 'master' genes. In contrast to the master genes, the vast majority of both SINEs and LINEs appear to behave like pseudogenes in their inability to undergo extensive amplification.

A significant percentage of all mammalian genomes consists of interspersed repetitive DNA sequences. These are generally classified as SINEs (short interspersed elements) or LINEs (long interspersed elements). The SINEs range in size from 90 to 400 bp, while LINEs can be as large as 7000 bp. New copies of both types of element find their way into the genome via reverse transcription of an RNA intermediate, a process called retroposition or retrotransposition (reviewed in Ref. 1). The RNA intermediate involved in the retroposition of SINEs is transcribed by RNA polymerase III, while that for LINEs is thought to be produced by RNA polymerase II. Both SINEs and LINEs are present in the mammalian genome in copy numbers in excess of 100 000. Many of the major SINE and LINE families can be divided into subfamilies, which are defined in terms of common nucleotide variations at specific locations. All current evidence indicates that the subfamilies arose via amplification, rather than through any process acting to reduce sequence divergence among pre-existing members of the family within a species.

Subfamilies of SINEs and LINEs

Although comparison of a large array of their properties shows that these two repetitive families belong to different classes, they both share a common feature that seems unexpected for a family of selfish elements: the lineages of both families seem to be dominated by a very small number of 'master' elements. It is this shared feature of these two disparate families that is discussed below.

Alu subfamilies

A given SINE family is usually present in only a moderate number of related species[2]. In primates, the most abundant SINE is the Alu family. It has long been known that the Alu family members exhibit species-specific polymorphisms within various primate species[3]. More recently, a number of laboratories have identified a series of subfamilies of Alu elements within some primate species[4–9]. Each of these subfamilies has one or more diagnostic differences in their consensus sequence relative to the Alu consensus sequence. However, there are significant differences in the amount of sequence divergence seen in the different subfamilies[8,9], suggesting that they arose at a variety of evolutionary times. Furthermore, analysis of the diagnostic changes in the different subfamilies suggested that subfamilies could be placed in a sequential order[7]. Analysis of the times of insertions, by sequencing and polymerase chain reaction (PCR) amplification of orthologous primate loci, confirms that the youngest subfamily members have the most diagnostic changes[8], but the least sequence divergence overall. These data, along with those from the recent insertion events to be described below, suggest that as a new subfamily is established, the older subfamily stops amplifying.