Copyright © 1990 by the Genetics Society of America

DNA Sequence Variation Within Maize and Melon: Observations From Polymerase Chain Reaction Amplification and Direct Sequencing

Donna M. Shattuck-Eidens, Russell N. Bell, Susan L. Neuhausen and Tim Helentjaris

Native Plants Incorporated, Salt Lake City, Utah 84108

Manuscript received July 13, 1989
Accepted for publication June 4, 1990

ABSTRACT

While compiling genetic linkage maps in several plant species based upon restriction fragment length polymorphisms (RFLPs), it was noted that the incidence of polymorphism differs among species. The basis of this disparity was investigated in this study by examining the nucleotide sequence at homologous loci among distinct cultivars within two species which exhibit considerably different levels of RFLPs. Using the polymerase chain reaction, homologous regions from different cultivars were first amplified and the nucleotide sequence of the products were determined. Four genomic regions of seven maize cultivars and three genomic regions of eight melon cultivars were examined to compare the respective levels of sequence variation between the two species. Levels of variation for both base substitutions and insertions/deletions varied widely among the maize sequences and between maize and melon for base substitutions. Estimates of theta (a measure of polymorphism) ranged from 0 to 0.002 in melon and from 0.006 to 0.040 for base substitutions and from 0.002 to 0.023 for insertions/deletions in maize. Critical value tests and chi-squared tests suggested that in maize the underlying processes generating and maintaining neutral mutations differ among the regions. The results not only suggest that several mechanisms are necessary to explain the variation seen in these two species, but also point to some basic dissimilarities in the organization and maintenance of the genomes of different plant species.

THE detection of variation in plant DNA sequence by restriction fragment length polymorphism (RFLP) analysis has shown that the apparent level of variation detected as RFLP within a species differs greatly from one species to another (HELENTJARIS *et al.* 1985; BERNATZKY and TANKSLEY 1986; APUYA *et al.* 1988). For instance in maize (*Zea mays*), we have found that greater than 95% of all unique sequence clones (ranging in size from 500 to 2000 base pairs) are able to reveal a RFLP even when tested with only a few restriction enzymes. The number of fragments of different lengths detected by one clone within Corn Belt germplasm is also quite high; up to 13 alleles per locus and with an average greater than six (HELENTJARIS *et al.* 1985).[1] With other species the level of polymorphism is much lower. We found, for instance, that only approximately 5% of unique sequence clones are able to detect RFLPs in the domesticated tomato, *Lycopersicum esculentum*, even when up to a dozen enzymes are used in the analysis. Furthermore, an informative clone usually reveals only two alleles, one of which often predominates within the germplasm of interest (HELENTJARIS *et al.* 1985).

In RFLP analysis, the differences in DNA sequence between individuals of a species are detected by Southern analysis of restriction enzyme digested genomic DNA using randomly selected genomic or cDNA cloned sequences as hybridization probes. These RFLPs can result from base changes, deletions, additions or rearrangements in the DNA sequence homologous to the cloned segment itself or in the area surrounding the cloned sequence but between the nearest recognition sites for the restriction enzyme used in the analysis. The number of enzymes that reveal RFLPs at a locus and the average number of alleles detected by individual clones can indicate whether the polymorphism is due to a single base change or to a rearrangement. A polymorphism revealed by a single enzyme with only two alleles is probably the result of a sequence difference which occurs in a recognition site for that enzyme. On the other hand, polymorphisms at a single locus that are revealed by several enzymes and which possess multiple alleles may reflect more complex rearrangements or, alternatively, a very high level of base substitutions. RFLP analysis demonstrates only that there are differences in the DNA sequence without yielding unequivocal information as to the nature of those differences. Before initiating this investigation we speculated that the differing rates of RFLPs among species might simply reflect dissimilar occurrences of rearrangements while the frequency of single nucleotide substitutions would be comparable. The objec-

[1] Although these are not genetically characterized regions, for the purpose of this discussion we will use the terms "locus" to define a region of the genome and "allele" to define a DNA sequence variant at a locus.

tives of this study were to examine different regions within the genomes of several cultivars of both maize and melon and to compare the incidence of nucleotide sequence variation within and between the two species to determine if a basis for the level of RFLPs could be identified. *Zea mays* represents those species which exhibit a high level of variation detected by RFLP analysis, while melon, *Cucumis melo*, represents those species which exhibit little detectable variation by the same approach. For this study the determination of nucleotide sequence was performed on DNA segments amplified by the polymerase chain reaction (PCR). The direct sequencing of PCR-amplified DNA obviates the need to isolate and sequence homologous sequences from libraries made from each individual. Direct sequencing of PCR product also mitigates artifacts caused by nucleotide misincorporation inasmuch as such misincorporation, even in the first few cycles of amplification, will affect only a small fraction of the final product.

Previous comparisons of sequence variation between plant species or members of a species have utilized characterized genes (WERR *et al.* 1985; ZACK, FERL and HANNAH 1986; FURTEK *et al.* 1988; RALSTON, ENGLISH and DOONER 1988), while the analysis presented here focuses upon uncharacterized clones. Although we used different criteria to select the individual regions to examine, the types and frequency of variation detected in maize in this study are consistent with what has previously been reported. No sequence comparison of any kind among melon cultivars has been reported. The maize germplasm examined herein represents a cross section of members of U.S. and Canadian maize varieties and could be considered a limited collection of the entire species since it includes only yellow dent varieties and none with any significant flint or tropical germplasm in their background. The melon isolates represent a broad collection of breeding germplasm from several countries including Indian, Iranian, Mexican, Spanish and U.S. lines. Our results clearly indicate that the narrower maize collection examined displays much more DNA sequence variability than the broader collection of melons.

MATERIALS AND METHODS

Biological materials: The clones were selected from pUC19 libraries of size-fractionated (0.5–2 kb) *Pst*I-digested genomic DNA from maize or melon. The DNA for constructing the original maize library was prepared from the Co159 × Tx303 hybrid (COTX) supplied by C. STUBER and M. EDWARDS, North Carolina State University. The Mangelsdorf tester line (MT) was supplied by D. WEBER, Illinois State University. The other maize lines examined are publicly available commercial inbreds. The melon clones and genomic DNA were from a NPI collection. Both maize and melon clones were selected from libraries by their representation of unique genomic sequences and their ability to reveal RFLPs (HELENTJARIS 1987; S. NEUHAUSEN, unpublished results). The maize clones used in this study have all been assigned to a chromosomal origin by a variety of methods. The melon clones have not been assigned to chromosomes but have been shown to reveal polymorphisms between melon varieties (S. NEUHAUSEN, unpublished results).

DNA sequencing of cloned segments: All cloned segments were sequenced using a Sequenase Kit (United States Biochemical Corp., Cleveland, Ohio) using [α-*thio*^{35}S]dATP according to instructions provided for plasmid sequencing. Sequencing of cloned DNA was initiated with universal primers and completed by preparing internal primers to extend the sequenced regions. Sequencing products were separated on 6% polyacrylamide field-gradient (wedge) gels containing 7 M urea (TABOR and RICHARDSON 1987).

Selection of primers for PCR: The sequences of the primers to be used in PCR amplification were derived from the sequence of the cloned segments. The criteria used for determining the sequence of the primers were based upon location in the cloned sequence, length and estimated dissociation temperature (T_d, calculated as described in MEINKOTH and WAHL 1984). The primer pairs were chosen so as to be 400–800 bases apart; a size range which we have found to produce high yields of product. Primers were 20–29 nucleotides in length with a calculated T_d of 60–65° (see Table 3). Primers were prepared by Genetic Designs Inc. (Houston, Texas) and were further purified by centrifugation to remove insoluble material with subsequent ethanol precipitation. In cases where a pair of primers failed to amplify, the substitution of one of the primers with a closely neighboring primer usually proved effective.

PCR reaction conditions: Conditions used for PCR amplification were as described in SAIKI *et al.* (1988), except that temperature cycling was accomplished in a Programmable Cyclic Reactor (Ericomp, San Diego, California). Denaturation, annealing and extension temperatures were 93°, 55°, and 72°, respectively. Thirty cycles of 30-sec denaturation, one minute of annealing and two minutes of extension were sufficient to amplify segments of less than 1 kb length. The last cycle utilized an eight minute extension to insure complete polymerization of all products begun in earlier cycles.

Purification and sequencing of the amplified segment: After amplification in a 300 μl reaction volume, the products were extracted with phenol then chloroform, precipitated with ethanol and dissolved in water. This DNA was electrophoresed in a 1.2% agarose gel, visualized by ethidium bromide staining and dissected out of the gel. The amplified DNA was isolated from the agarose using a GeneClean kit (Bio101, LaJolla, California) as described in the instructions. The DNA was then dried down and dissolved in water. The recovery was usually sufficient for two to three sequencing reactions. A portion of the product, in a volume of 9 μl, was denatured by heating 2 min in a boiling water bath. Immediately after denaturation 2–5 pmol of sequencing primer and 2 μl of 5× Sequenase buffer were added and the mixture cooled to room temperature. The primer used for sequencing was either one of the primers used in amplification or a primer located internally in the sequence. Most sequences could be read from within approximately 30 nucleotides of the primer extending to 200–300 bp. Sequence information near the amplifying primer was obtained by using an internal primer and reading from the opposite direction to the end of the amplified fragment.

Calculations for comparing sequence variation: Tests were performed to determine if there was significant heterogeneity among the maize sequences and among maize

and melon regions. The model tested was the "infinite sites neutral" model. The assumptions are that (1) mutations occur at random and every nucleotide change can occur with equal probability; (2) the number of sites is assumed to be large and the number of mutations small, so that sites are mutated no more than once in the population; and (3) mutations are selectively neutral (EWENS 1979). We are also assuming that there is an underlying mutation rate for base substitutions and a different rate for insertions/deletions.

The parameter theta was used as a measure of polymorphism under the infinite sites model: $\Theta = 4N\mu$ where N is the population size and μ is the mutation rate per site per generation. The estimates are per site so that differences in number of nucleotides and number of lines sequenced per region are accounted for in the calculations.

The following equations were used for estimating theta and its variance (EWENS 1979). An unbiased estimator of theta is given by: $\hat{\Theta} = s_n a_{n-1}^{-1} m^{-1}$; where $a_{n-1} = \sum_{i=1}^{n-1} 1/i$, s_n = number of segregating sites and m = number of sites in the region. The variance is bounded by:

$$\hat{\Theta} a_{n-1}^{-1} m^{-1} \text{ and } \hat{\Theta} a_{n-1}^{-1} m^{-1} + \Theta^2\{1 + 1/4 + \cdots + (n-1)^2\} a_{n-1}^{-2} m^{-1};$$

the values for free recombination and no recombination between sites, respectively.

For each region a lower and an upper critical value were computed for each estimated theta by calculating values that yield a five percent probability of observing a number of polymorphisms less than or equal to those actually observed (or greater than or equal to observed, depending on the one-sided test of interest). The critical values were determined by using the following recursive equation for the probability, $P_n(s)$, of observing s segregating sites in a sample of size n to 0.05 (HUDSON 1990): $P_n(s) = \sum_{i=0}^{s} P_{n-1}(s-1)Q_n(i)$ where

$$Q_n(s) = \left(\frac{\Theta}{\Theta + n - 1}\right)^s \left(\frac{n-1}{\Theta + n - 1}\right),$$

and solving for theta. The distribution of the number of polymorphic sites can then be quickly calculated using computer iteration.

Two different chi-squared tests were utilized to determine heterogeneity among the estimated thetas. An average theta value was calculated as weighted by the variance for no recombination.

Conservative-test: (AGUADE, MIYASHITA and LANGLEY 1989)

$$\chi^2 = \sum_{i=1}^{k} \frac{\{s(i)_{\text{observed}} - s(i)_{\text{expected}}\}^2}{\text{Var}\{s(i)_{\text{expected}}\}}$$

Standard test:

$$\chi^2 = \sum_{i=1}^{k} \frac{\{s(i)_{\text{observed}} - s(i)_{\text{expected}}\}^2}{s(i)_{\text{expected}}}$$

RESULTS

RFLP data from cultivars: Four regions of the maize genome and three regions of the melon genome were selected for comparison among six to eight cultivars of each species. Maize and melon cultivars were selected which exhibited distinct RFLP patterns with the clones used in this study. The maize lines selected for this study and their characteristics are described in Table 1. Although the lines differ in their pedigrees and vary for days to maturity, they are all yellow dent inbreds and are considered members of U.S.-developed Corn Belt germplasm. The exceptions are Co159, Tx303 and the Mangelsdorf tester which are yellow dent but not U.S. Corn Belt varieties. Tx303 and Co159 were selected by Stuber and Goodman as being the most distinctive inbreds by isozyme analysis in a comparison of over 400 maize lines at 23 loci (STUBER and GOODMAN 1983). The maize inbreds in Table 1 do not include any exotic, tropical or northern flint types. Both commercial and exotic melon germplasm were examined and descriptions of these melons are also found in Table 1. The melon cultivars examined are a subset of a collection from many parts of the world including muskmelon, casaba, crenshaw, and two unclassified plant introduction lines. The striking difference in RFLP diversity between the two species is illustrated in Table 2 where the RFLP data for the combination of probe, restriction enzyme and DNA substrate are summarized for the regions examined. The four regions selected in maize are four cloned sequences which were known to reveal RFLPs among maize inbreds. They represent sequences from distinct chromosomes (indicated in parenthesis in Table 2) and reveal various numbers of alleles among the cultivars. Both the fragment size and the number of alleles detected diminish when *Pst*I is used for

TABLE 1

Characteristics of plant material

Plant	Days to maturity	Type
Maize[a]		
A619Ht[b]	110	Yellow dent
B14A	118	Yellow dent
B73	116	Yellow dent
H99[b]	110	Yellow dent
Mo17	118	Yellow dent
Oh51	118	Yellow dent
W153R	95	Yellow dent
Co159[c]	80	Yellow dent
Tx303[c]	NA[d]	Yellow dent
Manglesdorf tester[e]	NA	Yellow dent
Melons		
Birdsnest	NA	Unclassified (Iran)
Crenshaw	110	Crenshaw
Itsy Bitsy	70	Noncommercial muskmelon
Juanes des Canary	105	Casaba (Spain)
PI414723	NA	Unclassified (India)
PI190184	NA	Unclassified (Mexico)
PMR45	90	Commercial muskmelon
Topmark	90	Commercial muskmelon

[a] Seedsman Handbook, Mike Brayton Seeds and Maize Research and Breeders Manual No. X, Illinois Foundation Seeds.
[b] Related lines derived from Oh43.
[c] These two inbreds were crossed to produce the COTX hybrid.
[d] NA, not available.
[e] Line created as a marker stock with a genetic marker on each of the ten chromosomes.

TABLE 2

Comparison of Corn and Melon RFLP diversity

	238 (1)[a]		451 (4)		288 (5)		445 (10)	
Maize	*Hin*dIII	*Pst*I	*Hin*dIII	*Pst*I	*Eco*RI	*Pst*I	*Hin*dIII	*Pst*I
A619Ht	2.7	0.9	7.0	0.7	3.9	1.3	3.6	1.0
B14A	2.7	0.9	9.7	0.7	3.9	1.6	3.6	1.0
B73	6.5	0.9	9.4	0.7	3.4	1.3	3.6	1.0
H99	3.8	0.9	7.0	0.7	3.7	1.3	3.6	1.0
Mo17	8.5	0.9	7.0	0.7	3.4	1.3	10.0	1.0
Oh51	4.0	1.0	9.4	0.7	3.7	1.3	4.0	1.0
W153R	3.5	1.2	9.7	0.7	3.4	1.3	4.0	1.0
Total No. alleles	6	3	3	1	3	2	3	1

	M2H5		M6B6		M6C11	
Melon	*Eco*RI	*Pst*I	*Hin*dIII	*Pst*I	*Eco*RV	*Pst*I
Birdsnest	10.0	2.0	12.5	1.8	7.3	1.5
Crenshaw	1.3	2.0	12.5	1.8	7.3	1.5
Itsy Bitsy	10.0	2.0	12.5	1.8	7.3	1.5
Juanes des Canary	1.3	2.0	12.5	1.8	7.3	1.5
PI414723	1.3	2.0	12.5	1.8	3.5	1.5
PI190183	10.0	2.0	12.5	1.8		1.5
PMR45	10.0	2.0	7.8	1.8	3.5	1.5
Topmark	10.0	2.0	12.5	1.8	3.5	1.5
Total No. alleles	2	1	2	1	2	1

The molecular size (kilobases) is shown for each of the fragments detected in the probe/substrate combinations. Number of alleles refers to the number of different-sized fragments observed with each combination.

[a] The numbers in parentheses indicate the chromosome to which the cloned sequence was assigned.

TABLE 3

Primers used for PCR amplifications

Region	Primer 1	Primer 2
Maize		
238	GGGGGGTGGCAGAGGAAGAG	GCGCCCGGCCAAGGAAGCGG
451	GGCTTCTAACGGATCCTCCT	CCAAAGCTTGTGTCTTCTCTGACTCACGC
288	AGTCCTCCTCCTCGATCGCG	GTATCGCGAACGTCATGATC
445	AAACAAGCAGTGACAGTTGCCG	CATGTTGTAGCTGCACCGAA
Melon		
M2H5	CTCCTCAAACATATCCTGTG	CTATTCTTCCCTCAAGGAAC
M6B6	TTTGGCATGAGAGAGGCACC	TCCTTCCCCCATAACAATGG
M6C11A	GTTCACACCCCCAACGTTGA	CATCTGGATCTGATACTGGG
M6C11B	CCTATCAAGCTCCCAGAGTG	TCCAATGGACCTAGGAACAG

The sequence of each of the pairs of primers used to amplify the target DNA during the PCR reactions prior to direct sequencing is shown in 5′ to 3′ orientation from left to right.

digestion. The amplified regions in this investigation are confined within the cloned *Pst*I sequences, so some information is in a sense "discarded" insofar as the sequences flanking the cloned segment are not examined.

Much less is known about the cloned sequences selected for analysis of melon. The three clones reported here were chosen because they reveal at least two alleles among the eight lines tested. There was insufficient variability detected between any two melon isolates to generate a linkage map, but the clones chosen for these experiments are assumed to be unlinked. Although the maize clones were assigned to chromosomes (Helentjaris, Weber and Wright 1986), no such means was available for the melon clones. As seen in Table 2, none of these clones revealed polymorphisms among melon cultivars when genomic DNA was digested with *Pst*I.

Nucleotide sequence determination: The nucleotide sequences were determined for the clones and examined for open reading frames consistent with coding regions for genes. Significant open reading frames were found in two of the melon clones (290 bases for M2H5 and 440 bases for M6B6) but none were found in the maize clones. No RNA hybridization analyses were performed to determine if homologous sequences are transcribed. This sequence information was used to identify segments for the preparation of primers to be used in PCR amplification of these regions in the individual cultivars. One pair of

TABLE 4

Sequence variation

Clone/region	No. of bases	Base change[a]	Deletions/insertions	Percent variation[b]
238	320	29 (12/17)	17	14.4
288	675	44 (19/25)	15	8.7
445	462	7 (1/6)	3	2.2
451	352	12 (7/5)	9	6.0
M2H5	470	0	0	0
M6B6	600	0	0	0
M6C11	502	2 (2/0)	0	0.4

[a] The numbers in parentheses indicate the nature of the base change, (transitions/transversions).

[b] Percent variation was calculated by summing the base changes and deletions/insertions for each clone then dividing by the number of bases compared.

primers was selected from each region in order to amplify 400–800 base pairs of DNA. Attempts to amplify larger segments resulted in diminished yields of product. M6C11 was too large to amplify in one interval, so two portions of this sequence, one from each end, were amplified and sequenced separately. The two regions, 200 bp and 302 bp long, are separated by about 500 bp. A list of the final primer pairs used as amplifiers for the maize and melon regions is shown in Table 3.

Characteristics of the sequence variation in maize: The collection of maize inbreds, while not the widest cross-section of the species available, did exhibit extensive and diverse nucleotide changes (Table 4). Many of the single base changes among the maize inbreds occur in sequences which are directly repeated. Examples of this type of change found in the 238 region are listed in Table 5. Another type of variation observed in maize is the variable number of nucleotides in a repeating unit such as a monotonous string of nucleotides (*e.g.*, AAA. . .) or a string of alternating nucleotides (*e.g.*, CTCT. . .). Examples of each of these can be found in the 445 region where there are seven A residues from nucleotide 369–375 in COTX, A619Ht, B14A and Mo17; while in B73, Oh51 and W153R there is an additional A for a total of eight A residues. Similarly, beginning with position 548 six repetitions of CT are seen in COTX; seven in Mo17 and B14A but only four in A619Ht, B73, H99, Oh51 and W153R (data not shown).

In Figure 1, the sequence around nucleotide 600 of the 288 region is represented and discontinuities of homology are depicted as insertions of DNA. The largest insert of 266 bp, found in line B14A, is flanked by a duplication of five bases, GATAA (site 2). Near the ends of the 266 bp insertion is a region of inverted repeat with a short five base perfect inverted repeat and a longer inverted repeat with 15 bases matched (Figure 1C). In the other seven lines, where this insertion is absent, the site 2 sequence GATAA is found only once at nucleotides 653 to 657. From nucleotides 591 to 599 in line B14A, the sequence TATGTTCTAA (site 1) is present while in the other seven lines this segment is repeated with one change so that nucleotides 588 to 607 read TATGTTCTA-ATCTGTTCTAA (Figure 1B). At the position corresponding to the point of insertion in B14A, a segment of 37 bases is observed in all other lines. These 37 bases are not present in B14A and therefore can be considered an insertion as compared to the B14A sequence. There are short imperfect inverted repeats in this insertion and it is flanked by a direct repeat of AAATTAGA (site 3) from 609 to 616 and AAA-TAAGA from 656 to 663 (Figure 1B). The sequence AAATAAGA is found only once in B14A, four bases away from the 266-bp insertion.

A measure of sequence variation is shown in Table 4 as "percent variation." This quantity was derived by summing the base changes and insertions/deletions for each clone and then dividing by the number of bases compared. The fewest and simplest changes were found in the region defined by clone 445. There were ten changes in 462 nucleotides or 2.2% sequence variation: seven single nucleotide changes, two single base insertions and one 4-base insertion. The variation seen in the other regions is more complex: a large insertion of 266 bases in the 288 region of B14A and several small insertions of 2–17 bases in several regions of the other lines. Comparison of the cultivars' nucleotide sequences in the 451 region with that of the cloned COTX segment shows 6.0% variation. There are twelve single-base changes and nine more complex changes, including insertions ranging in length from 1 to 17 nucleotides (Figure 2). Similar calculations result in 8.7% sequence variation in region 288 and 14.4% sequence variation in region 238.

Tests were performed to determine if the differences among the maize sequences were statistically significant. If the levels of variation are heterogeneous, then the underlying processes generating and maintaining mutations are probably different in the various regions (because population size is assumed the same). Estimates of theta, a parameter used as a measure of polymorphism under the infinite sites model, are given in Table 6, a and b, for base substitutions and insertions/deletions, respectively. In maize, the estimates for theta ranged from 0.006 to 0.040, a sevenfold range for base substitutions; and from 0.009 to 0.023 for additions/deletions. For both base substitutions and additions/deletions, clone 238 exhibited the most variation and clone 445 the least.

Chi-squared tests were performed to test the hypothesis that there was an underlying mutation rate to neutral alleles for the maize regions for base substitutions and for additions/deletions. A neutral value of theta was calculated as the weighted average of the

TABLE 5

Sequence variation in 238 region of maize

Nucleotide sequence[a]	Maize inbred
68 ACG	A619Ht and Mo17
ACG GCG	COTX
114 CCGG CCGA	B73, B14A, A619Ht and Mo17
CCGG CAGA	COTX
159 TCTTTAGTC	W153R
TCTTAGTC TCTTTAGTC	B73, B14A, A619Ht and Mo17
TCTTTAGTC TTTACTC	COTX
304 ATT	B73, B14A, A619Ht, Mo17 and W153R
ATT AGT	COTX

[a] The number of first nucleotide in the sequence is given first.

four regions. Two different chi-square tests were performed, the standard test and a conservative test which may be overly conservative (M. KREITMAN, personal communication). Based on the results the actual chi-squares are bounded between 5.02 and 30.4 for base substitutions and 5.04 and 18.3 for addition/deletions (Table 7). Because the actual chi-squared is bounded between two values, one of which is significant and the other not, the hypothesis cannot be firmly rejected. However, based on the range of the boundary and the level of significance for the chi-squared values, there is a strong indication that there is heterogeneity and that the regions are not evolving in the same manner.

A more precise way to test whether there is an underlying neutral parameter value would be to determine whether there is any theta value for which the observed polymorphism levels are likely (M. KREITMAN, personal communication). However, no such test is yet available. The extent of the differences can be measured by comparing an observed theta value with the appropriate one-sided Θ_U or Θ_L for the other regions (Table 6a and 6b). Clone 288 has an intermediate theta value ($\hat{\Theta} = 0.027$) relative to the other maize regions for base substitutions. One hypothesis is that the actual value of theta is 0.027. This hypothesis is accepted when comparing the observed theta value of 288 to the Θ_L value for 238 (0.018) and to the Θ_U value for 445 (0.038). However, it must be rejected when comparing it to the Θ_U value for 451 (0.010). Based on this critical value test and the chi-squared tests, the differences in levels of polymorphisms are most likely a function of differences in the rate of mutation to neutral alleles for both base substitutions and additions/deletions.

Sequence comparison among melon lines: The melon DNA sequence comparisons exhibited very little sequence variation in the regions examined (Table 4). Within regions of M2H5 and M6B6, there were no sequence differences within the collection of germplasm. In the M6C11 region there were two single base changes, both C–T transitions. These single base substitutions were the only changes observed in a combined total of 1572 bases sequenced in melon; no rearrangements of any kind were observed. One of the base changes in M6C11 distinguishes Birdsnest, Itsy Bitsy, Juanes des Canary, and PI190184 (all Ts) from Topmark, PMR45, and PI414723 (all Cs). This classification agrees with the RFLP analysis of these lines, although the observed nucleotide changes are not directly related to these polymorphisms. The other base change groups Birdsnest, PI190184 and PMR45 (all Ts) as one allele and Itsy Bitsy, Juanes des Canary and PI414723 (all Cs) as another. Therefore this base change distinguishes Birdsnest from Itsy Bitsy and Juanes des Canary as well as distinguishing PI414723 from PMR45. The sequence information facilitates the assignment of four alleles at this locus rather than two based on RFLP alleles. Allele assignment could be made as follows: Birdsnest and PI190184 as allele "A" (T–T), Itsy Bitsy and Juanes des Canary as allele "B" (T–C), PI414723 as allele "C" (C–C) and PMR45 as allele "D" (C–T). There is incomplete sequence information for Topmark (C–X) so it can not be assigned.

Sequence comparison between maize and melon: The difference in variation between melon and maize was very large. Since no additions/deletions were observed in the melon sequences, only base substitution polymorphisms can be compared. Estimates of theta ranged from 0.002 in clone M6C11 in melon to 0.040 for clone 238 in maize. Comparing the observed theta for M6C11 to the Θ_L for each of the maize regions reveals that the value is only at the critical value for 445, the maize clone with the least amount of variation. The average Θ for the three melon regions, 0.0007, however, is less than the lower critical value of the 445 region of maize ($\Theta_L = 0.002$). The significant difference in levels of polymorphism between maize and melon is a function of rates of mutation to neutral alleles and of population size differences between the species. It is not possible to determine what proportion of the variation is due to differences in

A

B14A

2 2

1 3

All other lines

3 2 3

1 1

B

B14A

588 site 2 site 2 **633**

TTGTATGTTCTAATGATAA...(266 base pair insert)...GATAAATAAGA

site 1 site 3

All other lines

588 site 2 **663**

TATGTTCTAATCTGTTCTAAGAAATTAGA...(37 base pair insert)...GATAAATAAGA

site 1 site 1 site 3 site 3

C

B14A insert

GATAACTAATATATAGAATTAGATCCAAATAGCATAAACCTAGTGCAATCACTCAATTTT
GGTAAAAAGAGCATTATCCTGTGCAATCGCCATCCTATGTCTAAATCCAACGGTTTCATA
TGGAAGTAGGTTAAATGTAAAATAGACACTAGGTTTCAGGTGGAAGGGGTTAAATGTAAA
ATAACAGTTACCATTAGATCTAGATCGGATGCACCAGATCGCTTGATTGCATGCTTGCAC
CAGGATTAGTGATTGNATAGGATATTTTCCC**GATAA**

All other lines

AAATTAAGATCCAAGTGGCATAAAAGGGNATAAAACATTTGCTATGAT**AAATAAGA**

FIGURE 1.—(A–C) Placement and sequence composition of the insertions in region 288. In (A) the line drawing shows the relative arrangement of sites 1–3 and their orientation with respect to the insertion. In (B) the nucleotide sequence flanking the insertions is displayed and sites 1–3 are indicated. In (C) the nucleotide sequences of the insertions are displayed with bold underlined type indicating the direct repeats and arrows indicating the inverted repeats.

underlying mutation rate versus differences in population sizes.

DISCUSSION

The construction of RFLP genetic linkage maps in several plant species has demonstrated that some species, such as maize, Brassica and Arabidopsis, exhibit a high rate of polymorphism among individuals, while the level of polymorphism among individuals of other species, such as tomato and melon, is considerably lower. The reason for this observed difference in genetic variation is not well understood. There appears to be no correlation between this disparity and genome size or ploidy levels. Maize, Brassica and Arabidopsis differ in genome size, with 1C values of 3.8, 0.9 and 0.2 picograms (pg), respectively (BENNETT and SMITH 1976), yet all three display significant polymorphism. The 1C values for tomato and melon are within the same range at 1.83 and 0.95 picograms respectively (BENNETT and SMITH 1976; INGLE, TIMMIS and SINCLAIR 1975), but demonstrate very little

```
Mol7        AGAAAACTGTTTTACCAAATNAATTTT---------CTGACTTTTT
COTX    CACAAGAAAACTGTTTTACCAAATGAATTTTAAAGTGGTTCTGATTTTTT    50
A619                                        CTCT-ACTTTTT
B73                                           CTGACTTTTT

Mol7    AGAGAAACCAAAGANCCAGGGGAGTCAGTGCTGAAACTGTTTTTAGAGGA
COTX    AGAGAAACCAAAGAGCCAGGGGAGTCAGTGCTGAAACTGTTTTTAGAGGA    100
A619    AGAGAAACCAAAGAGCCAGGTGAGTCAGTGCTGAAACTGTTTTTAGAGGA
B73     AGAGAAACCAAAGAGCCAGGGGAGTCAGTGCTGAAACTGTTTTTAGAGGA

Mol7    ACC-----------------GGAGCGGCGGCAGACCTACCAAACNGCCCT
COTX    ACCCTATTTTTAGAGGAACCGGAGC---------CCTACCAAAGGGCCCT    150
A619    ACCCTGTTTTTAGAGGAACCGGAGC---------CCTACCAAAGGTCCCT
B73     ACC-----------------GGAGC--GCGGAGACCTACCAAACGGCCCT

Mol7    TATTCTAGATCGGGCCTATAAGTCAGTGACCCAAATGCA-AGTGAATTGC
COTX    TATTCTAGATCGGGCCTATAAGTCAGTGATCCAAATGCA-AGTGAATTGC    200
A619    TATTCTAGATCGGGCCTATAAATCAGTGACCCAAATGCAAAGTGAATTGC
B73     TATTCTAGATCGGGCCTATAAGTCAGTGACCCAAATGCAAAGTGAATTGC

Mol7    AGGTGAATTAATTGTAGATGATGATCTTGTTGTGTCTAGGA----GGAGG
COTX    AGGTGAATTAATTGTA---GATGATCTTGTTGTGTCT--GA----GGAGG    250
A619    AGGTGAATTAACTGTAGATGATGATCTTGTTGTGTCT--GAAGAGGGAGG
B73     AGGTGAATTAATTGTAGATGATGATCTTGTTGTGTCT--GAAGAGGGAGG

Mol7    GCTCAATTGTAATTAACATGTTTAGAGC--------GACTCCAAAAGACT
COTX    GCTCAATTGTAATTAACATGTTTAGAGC--------GACTCCAAAAGACT    300
A619    GCTCAATTGTAATTAACATGTTTAGAGCTTTAGAGCGACTCCAAAAGACT
B73     GCTCAATTGTAATTAACATGTTTAGAGC--------GACTCCAAAAGACT

Mol7    GCTATAAAATTGTTCCCCAAAACTTAATATTAGGGGCTGATGTAAAAAAN
COTX    GCTATAAAATTGTTCCCCAAAACTTAATATTAGGGGCTGATGTAAAAAAN    350
A619    GCTATAAAATTGTTCCCCAAAACTTAATATTAGGGGCTGATGTAAAAAAG
B73     TCTATAAAATTGTTCCCCAAAACTTAATATTAGGGGCTGATGTAAAAAAG

Mol7    GTTTCCTAAAAAAT
COTX    GTTTCCTAAAAAAATTCTAAATTCACAACAGACTACTAATAAATTAACCCT
A619    GTTTCCTAAAAAACTCTAAATACACAACAGACTA                    400
B73     GTTTCCTAAAAAACTCTAAATTCACAACAGACTA
```

FIGURE 2.—Sequence of the amplified 451 region of several cultivars is compared to the cloned COTX sequence. Vertical lines are used to indicate identity. Where mismatches are detected the changed base is shown. Insertions are typed out with a gap, depicted by dashes, indicating the lack of the sequence in other cultivars or the cloned sequence. The sequence of H99 and Mt are the same as A619Ht in this region. The sequence of OH51 and B14A are the same as B73 in this region.

polymorphism. All of these species are diploid. Wheat is hexaploid with a large genome (1C equals 17.3 pg (BENNETT and SMITH 1976)), and does not exhibit significant isozyme or RFLP variation (HART 1987). Alfalfa, a tetraploid with 1C value of 1.7 pg (BENNETT and SMITH 1976), displays abundant polymorphism (G. KING, personal communication). A relationship has been noted, however, between variation assessed by RFLP analysis and reproductive strategy (HELENTJARIS *et al.* 1985). Those species which reproduce primarily by out-crossing also display abundant RFLP variation, while species which utilize self-pollination more extensively exhibit much lower levels of variation. This disparity in apparent genetic variation has also been reported using isozymes. Within the *Lycopersicon* species, those species which reproduce primarily by self pollination including *L. esculentum* (tomato) are monomorphic for most isozymes. In *L. chmielewskii*, which has larger flowers and is primarily allogamous, there is considerably more isozyme variation. Even higher levels of isozyme variability were seen in the self-incompatible *L. pennellii* (RICK 1983).

In the evolution of a species a reduction in population size could result in a lack of genetic variation. This may have occurred in the U.S. commercial melon germplasm since Hale's Best Jumbo is an early progenitor of many current melon cultivars. In this study melon germplasm collected from around the world was compared so that the effects of a breeding bottleneck should not be a factor.

In plants where isozyme variability can be detected there are usually 2 to 4.8 alleles per locus (GOTTLIEB 1981). In contrast, there are 5.17 to 7.09 alleles per isozyme locus in maize (GOODMAN and STUBER 1983; DOEBLEY, GOODMAN and STUBER 1985). Interestingly the level of RFLPs or isozyme variation does not necessarily correlate with the level of phenotypic variability among isolates. For example *C. melo*, though highly polymorphic in phenotype, shows little or no allelic variation at most isozyme loci (DANE 1983). While phenotypic variation is expected to be reflected in DNA diversity, extensive genotypic polymorphism has not been observed in the investigations of melon made thus far.

DNA sequence examination of regions of the melon genome revealed a paucity of sequence variation. This observation is consistent with the lack of polymorphism detected by RFLP analysis. The melon region M6C11, which has no open reading frame, showed two base changes, both transitions (C–T) and therefore 0.4% variation. The M6B6 sequence has an open reading frame of 440 bp and M2H5 has an open reading frame of 290 bp; neither of these regions revealed base changes among the cultivars examined. Our results are consistent with those of ZAMIR and TANKSLEY (1988) who reported that the DNA sequence in coding regions from tomato are conserved throughout the family Solanaceae. Their study used hybridization conditions which detect 80% sequence homology, making their criteria less stringent than sequence comparison.

Knowing the nucleotide sequence of the small segment represented by clone M6C11 yields more information about variation among the melon lines at this locus than is afforded by RFLP analysis using restriction enzymes with 6-base recognition sites (*i.e.*, the number of resolvable alleles increases from two to four). Conversely, with the other melon regions there was a loss of information.

Contrary to the prediction that maize and melon would share similar rates of base substitution, the data

TABLE 6

Estimates of Θ, Θ_L and Θ_U are lower and upper critical values of Θ at $P = 0.05$

a. For base substitutions

Clone	No. lines	No. base subs.	$(\hat{\Theta})$/site	Var$(\hat{\Theta})$/site $(\times 10^5)$	$\Theta_{L(0.05)}$	$\Theta_{U(0.05)}$
238	6	29	0.040	5.0–54.0	0.018	0.121
288	7	44	0.027	2.0–19.0	0.013	0.073
451	8	12	0.013	1.4–5.0	0.006	0.038
445	8	7	0.006	0.5–1.3	0.002	0.018
M6C11	6	2	0.002	0.1–0.2	0.000	0.009

b. For insertions/deletions

Clone	No. lines	No. insertions/ deletions	$(\hat{\Theta})$/site	Var$(\hat{\Theta})$/site $(\times 10^5)$	$\Theta_{L(0.05)}$	$\Theta_{U(0.05)}$
238	6	17	0.023	3.0–18.0	0.010	0.074
288	7	15	0.009	0.5–2.6	0.004	0.027
451	8	9	0.010	1.9–4.1	0.004	0.030
445	8	3	0.002	0.2–0.3	0.001	0.010

TABLE 7

Probability of observed levels of polymorphism based on a weighted average of theta using χ^2 test

Clone	S_{obs}	S_{exp}	Var$(s)_{exp}$	χ^2_{con}	χ^2_{std}
a. For base substitutions: $\overline{\Theta} = 0.029$					
238	29	21.2	146.9	0.41	2.87
288	44	48.0	618.9	0.03	0.33
451	12	26.4	185.9	1.12	7.86
445	7	34.7	305.8	2.51	22.11
				4.07†	33.17*
b. For insertions/deletions: $\overline{\Theta} = 0.019$					
238	17	14.0	68.0	0.13	0.64
288	15	31.4	276.5	0.97	8.57
451	9	17.3	84.8	0.81	3.98
445	3	22.7	139.0	2.79	17.10
				4.70‡	30.29*

$$\chi^2_{con} = \sum_{i=1}^{n} \frac{(\text{observed} - \text{expected})^2}{\text{Var (expected)}}$$

$$\chi^2_{std} = \sum_{i=1}^{n} \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

* $P < 0.01$.
† $0.50 < P < 0.20$.
‡ $0.1 < P < 0.05$.

clearly show a much lower incidence of base substitution in melon. Theta estimates were 0.002 for M6C11, the one melon clone exhibiting variation, and averaged 0.29 for the maize regions. These estimates are not compatible with a hypothesis that the two species evolved with the same underlying mutation rate. However, we can not account for differences in population size between maize and melon which could account for significantly different theta estimates. It is important to recognize that the criteria for selecting the RFLP loci for this investigation may impose some poorly understood limitations on their accurate representation of the genome. Because *Pst*I is a methylation-sensitive restriction enzyme, the clones probably represent undermethylated DNA within the genome. Furthermore, the clones are subjected to further bias by culling unique sequences from repetitive sequences. A more successful strategy for detecting variation in plants such as melons might be to include sequences that represent the genome more uniformly. Repetitive sequences, for example, have been shown to reveal higher rates of variability in other systems (ZAMIR and TANKSLEY 1988; JEFFREYS, WILSON and THEIN 1985).

In this study abundant sequence variation was detected in homologous segments of the maize genome by comparing several representatives of the species. Although a limited sample of the genome was analyzed, the percentage of variation was relatively high for all regions and ranged from 2.2% to 14.8%. Estimates of theta ranged from 0.006 to 0.040 for base substitutions and from 0.002 to 0.023 for insertions/deletions. Based on the critical value and the chi-squared tests, the regions appear to be evolving at different rates.

The observed pattern of rearrangements among maize isolates suggest possible mechanisms by which the variation occurs. One mechanism to consider for maize is transposable element activity. Previous comparisons of DNA sequence variation among maize isolates have been made primarily on identified genes (WERR *et al.* 1985; ZACK, FERL and HANNAH 1986; FURTEK *et al.* 1988; RALSTON, ENGLISH and DOONER 1988). In such studies of several maize genes (*Wx*, *Adh1*, *Bz* and *Sh1*), small tandem repeats have been identified as footprints left behind following the insertion and subsequent excision of transposable elements. This was established by the sequential examination of the wild type, the transposable element-disrupted and revertant genes. In those studies the sequence diversity in the revertant genes, which appeared to be the result of transposable element activity was seen only

in the noncoding regions associated with those genes. The revertant alleles, by definition, conferred the wild type phenotype for that gene. It is not possible, therefore, to determine if transposable elements preferentially avoid inserting in coding regions or whether, because of the resulting loss of function, such insertion would not be observed with such an experimental design. Differences located within the exons were found to occur primarily in the third position in the codon and therefore represent silent changes. A very high level of sequence variation was observed in the *Shrunken-1* locus (*Sh1*): 16 changes in 540 bases within the coding regions, all of them silent, and 10 changes for 270 bases for the transcribed but untranslated 3′ region (WERR et al. 1985). These levels of variation, 3.0% in the coding and 3.7% in the noncoding, are within the range we report for maize, but are still higher than the level we report for regions within the melon genome.

The insertions near nucleotide 600 in the maize 288 region (Figure 2) have characteristics of transposable elements: they are flanked by short duplications and have inverted repeats at the ends. These insertions, however, show very limited sequence homology. If these insertions are the result of transposable elements, the clustering of the insertions may indicate that transposable elements preferentially insert in specific locations in the genome or that once inserted, they tend to excise and reinsert at closely neighboring sites. It has been suggested that in maize the size of the flanking duplication indicates the class of transposable element that generated it (DORING and STARLINGER 1984). The duplications in this region, however, vary in length and share no obvious homology. These changes may be the result of the activity of two unrelated transposable elements, since in maize the characterized transposable elements leave a footprint of a consistent size. It should be noted, however, that the *Antirrhinum* transposable element *Tam3* has been shown to generate different length footprints at different insertion sites (DENNIS et al. 1988).

As predicted by the RFLP analysis, the sequence variation in maize is much higher than in melons, both in the number of base changes and DNA rearrangements. It is not known what accounts for the difference in RFLP and nucleotide sequence variation between maize and melon. Whatever the source(s) of diversity is in maize, it appears not to operate in melon. It has been proposed that transposable element activity plays a role in generating insertion-deletion based diversity in maize (SCHWARZ-SOMMER et al. 1985). The results presented here are consistent with that proposal to the extent that some of the changes in maize lines resemble transposable element-mediated rearrangements, whereas no changes of this type were observed in melon. It is possible that species which are primarily self-pollinated generally lack significant levels of transposable element activity. For example, in *Lycopersicum esculentum*, a self-pollinating species, there is no evidence for active transposable elements (YODER et al. 1988). However, mutable alleles have been reported in soybean, another self-pollinated species (CHANDLEE and VODKIN 1988; GROOSE et al. 1988). It has been reported that self-pollination of maize lines harboring active Mu transposable elements often leads to significant loss of Mu transposition (ROBERTSON 1983). The transmission advantage of a transposable element is reduced in self-pollinating lineages because the opportunity to infect new chromosomes through genetic exchange is reduced by the inbreeding process (M. CLEGG, personal communication).

In Table 5, there is a summary of the sequence variation which occurs in what appear to be duplicated sequences. Sequence duplications resulting from transposable element insertion are often found to be altered (SUTTON et al. 1984; SCHWARZ-SOMMER et al. 1985), although only deletions of nucleotides at the junction of the two duplicated sequences have been observed previously. The base changes shown in Table 5 could also have resulted from transient misalignment during DNA synthesis as proposed by KUNKEL and SONI (1988). Changes such as the addition or loss of one or more bases in a repeating unit may also be the result of unequal crossover in recombination or errors in DNA replication as has been proposed elsewhere (SMITH 1976).

As a first approximation, the incidence of point mutations is expected to be uniform over the noncoding sequences in the genome. There are definite differences, however, in the percentage of sequence variation detected in the four different regions of maize. Based on the estimates of theta and the critical value and chi-squared tests, it appears that the clones examined represent regions of the genome that differ in their susceptibility to mutation and/or to the fixation of mutations. It has been reported that the rate of nucleotide substitution varies considerably depending on the region examined, and that the more important the function of the region, the lower the rate of polymorphism (LI, LUO and WU 1985). If these are noncoding regions, transposable element activity, which may not be uniform over the genome, may account for some of the differences. An alternative explanation has recently been presented for mammalian genomes. WOLFE, SHARP and LI (1989) proposed that differences arise because mutation patterns vary with the timing of replication of different chromosomal regions. In connection with this it is interesting to note that region 445, which displays the least variation, is located near the centromere of chromosome 10. The other three regions; 238, 288 and 451; map

to the ends of chromosome 1, 5 and 4, respectively (HELENTJARIS 1987). The greater sequence variation in these regions may be due to the timing of replication as WOLFE, SHARP and LI suggest, or may reflect greater recombination or transposable element activity at the ends of chromosomes. More regions from various locations along the chromosomes must be examined to test the validity of these proposals.

We wish to thank MARTIN KREITMAN for his help in analyzing the sequence variation data. We are indebted to CHARLES HUSSEY for his guidance with DNA sequencing and electrophoresis. We appreciate the contribution of CAROL COPINGA-CRANE and JAN TIVANG in preparing the figures and tables.

LITERATURE CITED

AGUADE, M., N. MIYASHITA and C. LANGLEY, 1989 Reduced variation in the *yellow-achaete-scute* region in natural populations of *Drosophila melanogaster.* Genetics **122:** 607–615.

APUYA, N. R., B. L. FRAZIER, P. KEIM, E. J. ROTH and K. G. LARK, 1988 Restriction fragment length polymorphism as genetic markers in soybean, *Glycine max* (L.) merrill. Theor. Appl. Genet. **75:** 889–901.

BENNETT, M. D., and J. B. SMITH, 1976 Nuclear DNA amounts in angiosperms. Proc. R. Soc. Lond. Ser. B **274:** 28–274.

BERNATZKY, R., and S. D. TANKSLEY, 1986 Toward a saturated linkage map in tomato based on isozymes and random cDNA sequences. Genetics **112:** 887–898.

CHANDLEE, J. M., and L. O. VODKIN, 1988 Identification and characterization of active transposable element systems in soybean, p. 364 in *Plant Transposable Elements*, edited by O. NELSON. Plenum Press, New York.

DANE, F., 1983 Cucurbits pp. 369–390, in *Isozymes in Plant Genetics and Breeding, Part B*, edited by S. D. TANKSLEY and T. J. ORTON. Elsevier Science Publishers B.V., Amsterdam.

DENNIS, E. S., E. J. FINNEGAN, B. H. TAYLOR, T. A. PETERSON, A. R. WALKER and W. J. PEACOCK, 1988 Maize transposable elements: structure, function and regulation, pp. 101–113 in *Plant Transposable Elements*, edited by O. NELSON. Plenum Press, New York.

DOEBLEY, J. F., M. M. GOODMAN and C. W. STUBER, 1985 Isozyme variation in the races of maize from Mexico. Am. J. Bot. **72:** 629–639.

DORING, H.-P., and P. STARLINGER, 1984 Barbara McClintock's controlling elements: now at the DNA level. Cell **39:** 253–259.

EWENS, W. J., 1979 *Mathematical Population Genetics.* Springer-Verlag, New York.

FURTEK, D., J. W. SCHIEFELBEIN, F. JOHNSTON and O. E. NELSON, JR., 1988 Sequence comparisons of three wild-type *Bronze-1* alleles from *Zea mays.* Plant. Mol. Biol. **11:** 473–481.

GOODMAN, M. M., and C. W. STUBER, 1983 Maize, pp. 1–33 in *Isozymes in Plant Genetics and Breeding, Part B*, edited by S. D. TANKSLEY and T. J. ORTON. Elsevier Science Publishers B.V., Amsterdam.

GOTTLIEB, L. D., 1981 Electrophoretic evidence and plant populations. Prog. Phytochem. **7:** 1–46.

GROOSE, R. W., S. M. BLOMGREN, P. KEIM, R. C. SHOEMAKER and R. B. PALMER, 1988 A mutable allele of the *w4* locus for anthocyanin pigmentation in soybean, p. 367 in *Plant Transposable Elements*, edited by O. NELSON. Plenum Press, New York.

HART, G. E., 1987 Genetic and biochemical studies of enzymes. In *Wheat and Wheat Improvement*, p. 199, edited by E. G. HEYNE. American Society of Agronomy, Madison, Wisc.

HELENTJARIS, T., 1987 A genetic linkage map for maize based on RFLPs. Trends Genet. **3:** 217–221.

HELENTJARIS, T., D. F. WEBER and S. WRIGHT, 1986 Use of monosomics to map cloned DNA fragments in maize. Proc. Natl. Acad. Sci. USA **83:** 6035–6039.

HELENTJARIS, T., G. KING, M. SLOCUM, C. SIEDENSTRANG and S. WEGMAN, 1985 Restriction fragment polymorphisms as probes for plant diversity and their development as tools for applied plant breeding. Plant. Mol. Biol. **5:** 109–118.

HUDSON, R. R., 1990 Gene genealogies and the coalescent process. Oxf. Surv. Evol. Biol. (in press)

INGLE, J., J. N. TIMMIS and J. SINCLAIR, 1975 The relationship between satellite deoxyribonucleic acid ribosomal ribonucleic acid gene redundancy and genome size in plants. Plant Physiol. **55:** 496–501.

JEFFREYS, A. J., V. WILSON and S. L. THEIN, 1985 Individual-specific 'fingerprints' of human DNA. Nature **316:** 76–79.

KUNKEL, T. A., and A. SONI, 1988 Mutagenesis by transient misalignment. J. Biol. Chem. **263:** 14784–14789.

LI, W.-H., C.-C. LUO and C.-I. WU, 1985 Evolution of DNA sequences, pp. 1–94 in *Molecular Evolutionary Genetics*, edited by R. J. MCINTYRE. Plenum Press, New York.

MEINKOTH, J., and G. WAHL, 1984 Hybridization of nucleic acids immobilized on solid supports. Anal. Biochem. **138:** 267–284.

RALSTON, E. J., J. J. ENGLISH and H. K. DOONER, 1988 Sequence of three *bronze* alleles of maize and correlation with the genetic fine structure. Genetics **119:** 185–197.

RICK, C. M., 1983 Tomato, pp. 147–166 in *Isozymes in Plant Genetics and Breeding, Part B*, edited by S. D. TANKSLEY and T. J. ORTON. Elsevier Science Publishers B.V., Amsterdam.

ROBERTSON, D. S., 1983 A possible dose-dependent inactivation of mutator (Mu) in maize. Mol. Gen. Genet. **191:** 86–90.

SAIKI, R. K., D. H. GELFAND, S. STOFFEL, S. J. SCHARF, R. HIGUCHI, G. T. HORN, K. B. MULLIS and H. A. ERLICH, 1988 Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239:** 487–491.

SCHWARZ-SOMMER, Z., A. GIERL, H. CUYPERS, P. A. PETERSON and H. SAEDLER, 1985 Plant transposable elements generate the DNA sequence diversity needed in evolution. EMBO J. **4:** 591–597.

SMITH, G. P., 1976 Evolution of repeated DNA sequences by unequal crossover. Science **191:** 528–535.

STUBER, C. W., and M. M. GOODMAN, 1983 Allozyme genotypes for popular and historically important inbred lines of corn, *Zea mays* L. ARS, Agricultural Research Results, Southern Series, No. 16. Raleigh, N.C.

SUTTON, W. D., W. L. GERLACH, D. SCHWARTZ and W. J. PEACOCK, 1984 Molecular analysis of *Ds* controlling element mutations at the *Adh1* locus of maize. Science **223:** 1265–1268.

TABOR, S., and C. C. RICHARDSON, 1987 DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA **84:** 4767–4771.

WERR, W., W.-B. FROMMER, C. MAAS and P. STARLINGER, 1985 Structure of the sucrose synthase gene on chromosome 9 of *Zea mays* L. EMBO J. **4:** 1373–1380.

WOLFE, K. H., P. M. SHARP and W.-H. LI, 1989 Mutation rates differ among regions of the mammalian genome. Nature **337:** 283–285.

YODER, J. I., J. PALYS, K. ALPERT and M. LASSNER, 1988 *Ac* transposition in transgenic tomato plants. Mol. Gen. Genet. **213:** 291–296.

ZACK, C. D., R. J. FERL and L. C. HANNAH, 1986 DNA sequence of a *shrunken* allele of maize: evidence for visitation by insertional sequences. Maydica **31:** 5–16.

ZAMIR, D., and S. D. TANKSLEY, 1988 Tomato genome is comprised largely of fast-evolving, low copy-number sequences. Mol. Gen. Genet. **213:** 254–261.

Communicating editor: B. BURR