Insertion/deletion mutations in the *Zea* chloroplast genome

J. F. Doebley, D. P. Ma, and W. T. Renfroe*

Department of Biology, Texas A&M University, College Station, TX 77843, USA

Summary. The chloroplast (cp) genomes of *Zea* species are distinguished by at least four restriction fragment length (insertion/deletion) mutations. All four mutations occur in the large unique sequence region of the genome. Restriction fragments containing three of these mutations were cloned. The large and small forms of two of the mutated fragments were sequenced. This revealed 80 and 83 bp insertion/deletions. The inserted/deleted segments are not parts of tandem repeats nor were they flanked by direct repeats. Two other insertion/deletion mutations were not sequenced, but their sizes were estimated to be 150 and 250 bp by size fractionation on agarose gels. Use of *Tripsacum pilosum* and *Sorghum bicolor* as outgroups suggests that three of the fragment length mutations arose via deletions. The fourth could not be polarized. The three species of section *Luxuriantes* of *Zea* were identical to one another for each of the four length mutations, and they were consistantly distinguished from the taxa of section *Zea* by these mutations. These data support the division of *Zea* into the above named sections.

Key words: *Zea mays* – Chloroplast DNA – Length mutations – Taxonomy

Introduction

The chloroplasts of flowering plants contain circular DNA molecules that range in size from 120 to 217 kb (Whitfeld and Bottomley 1983; Palmer 1985). These molecules evolve by base pair substitutions and struc-

tural mutations such as inversions and insertions/deletions (Palmer 1985). The insertion/deletion or length mutations can be classified into small (1–50 bp), medium (50–1,200 bp) and large (> 1,200 bp) forms (Palmer 1985). The medium and large mutations can make significant contributions to the range of size variation that has been observed among higher plants. The mechanism(s) generating length mutations in chloroplast genomes are not fully understood. Three possibilities are: (1) unequal crossing over within arrays of tandem repeats (Palmer et al. 1985), (2) deletion of segments between direct repeats (Palmer 1985) and (3) insertion/ deletion events which create or destroy short direct repeats (Takaiwa and Sugiura 1982; Zurawski et al. 1984). Some authors have noted that small length mutations occur preferentially in AT rich regions (Zurawski et al. 1984), and that, in some cases, they map close to the origin of replication (Koller and Delius 1982).

Medium-sized fragment length mutations have been reported in the chloroplast genomes of several angiosperm genera. Gordon et al. (1982) detected 20 length mutations among members of the genus *Oenothera*. Bowman et al. (1983) reported 14 length mutations among members of *Triticum* (including *Aegilops*). These mutations are spread over a 60 kb portion of the large unique sequence region of the genome. Palmer et al. (1983) reported a number of fragment length mutations among species of *Brassica*. Finally, Palmer et al. (1985) described 17 fragment length mutations in *Pisum* that are distributed among four small regions ("hot spots") of the genome.

In this paper, we report four length mutations found in the large unique sequence region of the *Zea* (Gramineae) chloroplast genome. An attempt was made to determine the directionality (insertion or deletion) of each, and two of these mutations were analyzed by DNA sequencing.

Present address: Department of Biology, James Madison University, Harrisonburg, VA 22807, USA

Abbreviations. bp = base pairs; kb = 1,000 bp; cpDNA = chloroplast DNA; SDS = sodium dodecyl sulfate

Table 1. Accessions analyzed

Taxa	Locality	Source	Collection
Zea Section *Luxuriantes*			
1. *Z. perennis* (Zp)	Piedra Ancha, Jalisco, Mexico	Iltis	1050
2. *Z. diploperennis* (Zd)	La Ventana, Jalisco, Mexico	Guzman	777
3. *Z. luxurians* (Zl)	Ipala, Chiquimula, Guatemala	Iltis	G-42
Zea Section *Zea*			
4. *Z. mays* subsp. *parviglumis* var. *huehuetenangensis* (Zmh)	San Antonio Huista, Huehuetenango, Guatemala	Iltis	G-120
5. *Z. mays* subsp. *parviglumis* var. *parviglumis* (Zmp)	Teloloapan Guerrero, Mexico	Iltis	80
6. *Z. mays* subsp. *mexicana* Race Central Plateau (Zmx)	Durango, Mexico	Doebley	625
7. *Z. mays* subsp. *mexicana* Race Chalco (Zmx)	Texcoco, Mexico, Mex.	Doebley	479
8. *Z. mays* subsp. *mays* Race Northern Flint (Zmm)	USA	USDA	214,195
9. *Tripsacum pilosum* (Tp)	Jalisco, Mexico	Doebley	467
10. *Sorghum bicolor* (Sb)	–	K. Schertz	BTx623

Materials and methods

The genus *Zea* includes maize and its wild relatives, the teosintes. As treated by Iltis and Doebley (1980), *Zea* contains four species divided between two sections. Section *Zea* contains one species, *Z. mays*. This species is highly polymorphic, containing three subspecies: subsp. *mays* (corn or maize), subsp. *mexicana* (Chalco, Central Plateau and Nobogame teosintes) and subsp. *parviglumis* (Balsas and Huehuetenango teosintes). Within the subspecies, varietal and racial designations are given to some populations (Wilkes 1967; Iltis and Doebley 1980). Section *Luxuriantes* contains three wild species (teosintes), *Z. perennis*, *Z. diploperennis* and *Z. luxurians*. For the present study, eight accessions of *Zea* were chosen to represent the array of taxonomic variation in *Zea* (Table 1). For each accession, seed from a single plant was used for chloroplast isolation. For comparative purposes, cpDNA was isolated from *Tripsacum pilosum* and *Sorghum bicolor* both of which belong to the same tribe (Andropogoneae) of the grass family as *Zea* (Table 1).

Chloroplasts were isolated from 2–3 week old seedlings by the sucrose gradient method (Palmer 1986). After isolation, the chloroplasts were resuspended in 50 mM Tris-25 mM EDTA pH 8.0 and frozen (−20 °C) for later DNA extractions. DNA was isolated from the chloroplasts by phenol and chloroform extractions (Zimmer and Newton 1982).

Restriction enzymes were purchased from Bethesda Research laboratories or New England Biolabs and the restrictions carried out according to manufacturer's recommendations. Restricted DNAs were electrophoresed in 0.5, 0.8, or 1.5% agarose gels with a running buffer of 100 mM Tris-Acetate, 1 mM EDTA pH 8.1. DNA fragment sizes were estimated using the maximum likelihood method with *Hind*III-digested bacteriophage λDNA and/or *Hae*III-digested φX174 DNA as standards (Schaffer and Sederoff 1981). DNA fragments in the gels were denatured and transferred to a nylon hybridization membrane (Gene Screen Plus) according to manufacturer's (New England Nuclear) recommendations. Nylon filters were prehybridized overnight in a hybridization buffer of 10% dextran sulfate, 1 M NaCl and 1% SDS. Cloned DNA fragments were labeled with [32]P dATP by nick translation as described by Maniatis et al. (1982). Nick translated probes were separated from unincorporated ^{32}P-dATP on spun columns (Maniatis et al. 1982), and then added to the hybridization buffer. Hybridizations were carried out at 65 °C overnight. Hybridization membranes were washed according to manufacturer's instructions and exposed to x-ray film (Kodak XAR-5) at −80 °C for 6 h to 1 week using Dupont Cronex intensifier screens.

Chloroplast DNA was restricted with *Sma*I or *Pst*I and ligated into pUC9 as described by Maniatis et al. (1982). Ligation products were transformed into *E. coli* strain JM83 (Dagert and Ehrlich 1979). Recombinant plasmids were isolated from *E. coli* by the NaOH/SDS method (Birnboim and Doly 1979), and those clones containing fragments with length mutations were identified on agarose gels. A *Bam*HI library of *Sorghum bicolor* (IS1112C) cpDNA in pUC8 (Dang and Pring 1986) was used in some experiments.

DNA sequences were obtained using the dideoxynucleotide chain termination method (Sanger et al. 1977, 1980) after ligation into M13 and transfection in *E. coli* JM101 (Messing and Vieira 1982). The single-stranded recombinant phage DNAs were isolated by phenol extraction of the polyethylene glycol concentrated phage and sequenced using a universal primer (Anderson et al. 1980). Sequential series of overlapping clones for sequencing were produced in both orientations using single-stranded recombinant M13 (mp8 and mp9) DNA and complementary 22- or 29-mer deoxyoligonucleotides as described by Dale et al. (1985). The complete sequence of contiguous regions was assembled from individual clones using programs from the University of Wisconsin Genetics Computer Group (Devereux et al. 1984).

Results

Restriction endonuclease analysis

Each of the *Zea* cpDNAs (Table 1) was digested with a series of restriction enzymes. *Sal*I, *Xho*I, *Hpa*I, *Bgl*I, *Bst*EII, *Sac*II, and *Pvu*II showed no differences in the sizes and numbers of bands among the samples. *Sma*I, *Pst*I, and *Kpn*I showed the same numbers of bands in each sample, although one or two bands varied slightly
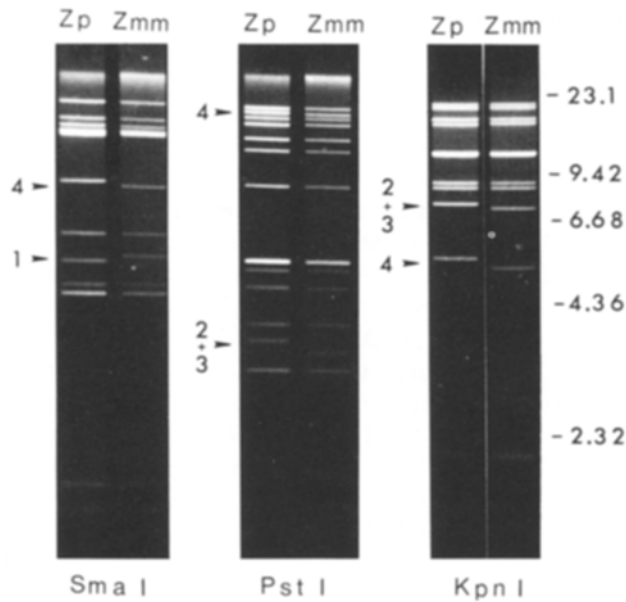
Fig. 1. Restriction digests of *Zea perennis* (Zp) and *Z. mays* subsp. *mays* (Zmm) with *Sma*I, *Pst*I, and *Kpn*I. *Arrows* indicate fragments that differ in size between the two cpDNAs and the numbers refer to specific length mutations (see text). Size markers are shown on right in kb

in size (Fig. 1). Several other enzymes (*Eco*RI, *Bam*HI, *Hin*dIII, *Sac*I, *Bgl*II, *Nco*I, *Dra*I and *Cla*I) showed more complex differences in both size and numbers of bands among the samples, suggesting that both restriction site and length mutations distinguished the samples.

For *Sma*I, *Z. perennis*, *Z. diploperennis* and *Z. luxurians* cpDNAs contained a 5.18 kb fragment, while cpDNAs of *Z. mays* (including all subspecies) had a 5.34 kb fragment (Fig. 1). The 5.18 kb *Sma*I fragment was

cloned in pUC9 from *Z. perennis,* and the 5.34 kb *Sma*I fragment from subsp. *mays.* The cloned fragments were digested separately with *Eco*RI, *Bam*HI, *Sal*I and *Pst*I. Because a map for each of these enzymes has been published for maize (Larrinua et al. 1983), it was possible to place these *Sma*I fragments on the maize map simply by examination of the size and number of fragments produced with each of the four enyzmes. Both *Sma*I fragments lie between positions 41,000 and 46,500 on the maize map (Fig. 2). Restrictions of the two cloned *Sma*I fragments also demonstrated that they differ by a 150 bp length mutation, which we have designated LM-1. This mutation occurs in the *Eco*RI-1$'$ fragment or *Bam*-HI-15$'$ (Fig. 2). Fragment number (or letter) designations are those of Larrinua et al. (1983).

To determine whether LM-1 represents an insertion or deletion, we compared the size of *Eco*RI-1$'$ and *Bam*HI-15$'$ of our *Zea* cpDNAs to the homologous fragments of *Sorghum bicolor.* Dang and Pring (1986) have shown that the sorghum and maize maps are remarkably similar. Phylogeneticists refer to the use of a related taxon in this manner as *outgroup analysis.* The logic applied in this comparison is that the genetic mechanism which requires the fewest events is the mechanism most likely involved in the mutation. For example, if the sorghum fragment is the same size as the larger *Zea* fragment, then a single deletion occurring in one of the *Zea* cpDNA phylads is implicated. Under these circumstances, the alternative explanation (insertion) would require that identically-sized insertions occurred independently in *Eco*RI-1$'$ in sorghum and some *Zea.* This is a far less likely occurrence. The sorghum recombinant plasmid pLD18 (Dang and Pring 1986) contains sorghum *Bam*HI-18 which is homologous to *Zea* *Bam*HI-15$'$. This plasmid was used to probe a filter con-
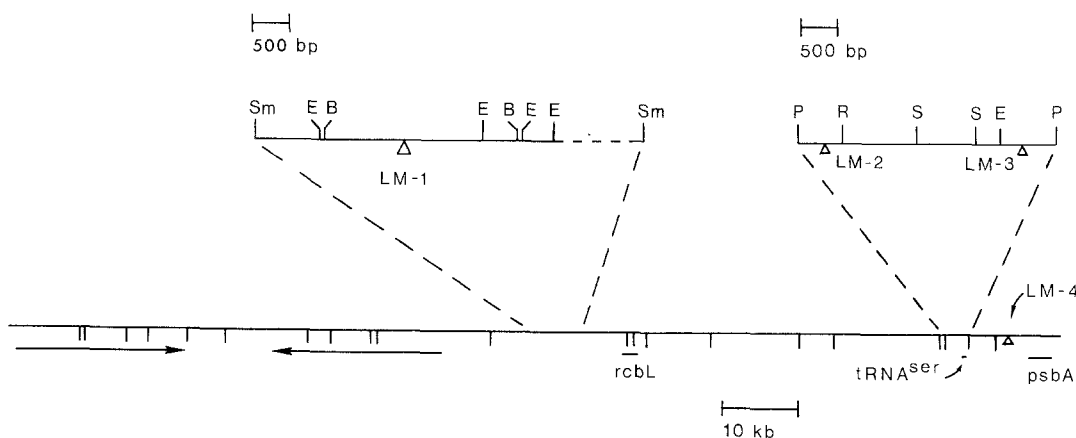
Fig. 2. Restriction map of *Zea mays* subsp. *mays* chloroplast genome showing the positions of the four length mutations (LM-1 to 4). *Pst*I sites, *rcbL*, tRNA[ser] (UGA), *psbA* and inverted repeats (*large arrows*) are shown. On the enlarged *Sma*I and *Pst*I fragments, *Bam*HI (B), *Eco*RI (E), *Pst*I (P), *Rsa*I (R), *Sal*I (S) and *Sma*I (Sm) sites are shown. *Dashed line* on the enlarged *Sma*I fragment represents a region with several *Bam*HI and *Eco*RI sites that were not mapped. The enlarged *Pst*I fragment contains additional *Rsa*I sites which were not mapped

Fig. 3. Autoradiograph used to determine directionality of LM-1. *Bam*HI and *Eco*RI digests of *Zea perennis* (Zp), *Z. mays* (Zmm) and *Sorghum bicolor* (Sb) cpDNA. Sorghum clone pLD18 was used as the probe. Zmm and Sb have 2.85 kb *Bam*HI fragments, but the homologous fragment in Zp is 2.7 kb. Zmm and Sb have 2.45 kb *Eco*RI fragments, but the homologous fragment in Zp is 2.30 kb

Fig. 4A, B. Autoradiographs used to determine the polarity of LM-2. **A** *Pst*I-*Sal*I double digests of *Zea perennis* (Zp), *Z. mays* (Zmm) and *Sorghum bicolor* (Sb) cpDNA using the 1.6 kb *Sal*I-*Pst*I fragment in which LM-2 occurs as the probe. **B** *Pst*I-*Rsa*I double digests of Zp, Zmm and *Tripsacum pilosum* (Tp) cpDNA using the 600 bp *Pst*I-*Rsa*I fragment in which LM-2 occurs as the probe

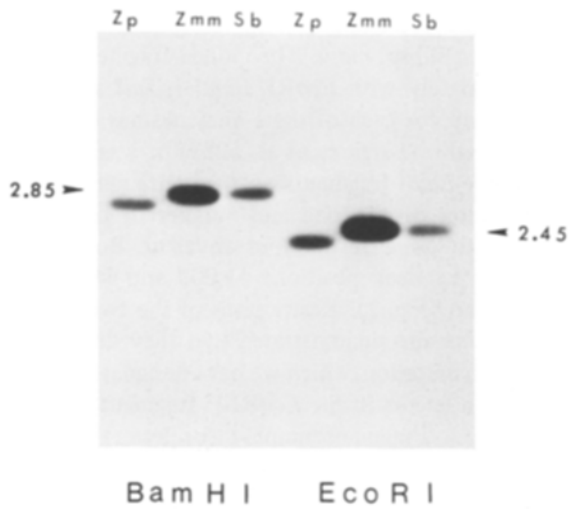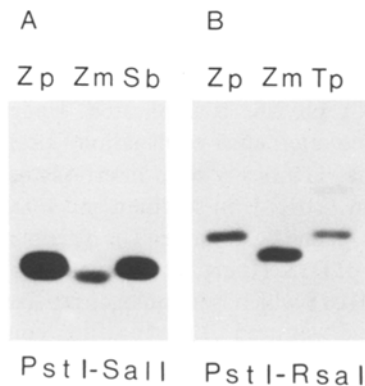taining *Bam*HI and *Eco*RI digests of *Zea perennis, Z. mays* subsp. *mays* and sorghum cpDNAs (Fig. 3). Maize *Eco*RI-1' and *Bam*HI-15' are identical in size to the homologous fragments of sorghum, but both are approximately 150 bp larger than their counterparts from *Z. perennis* (Fig. 3). By outgroup analysis as discussed above, this suggests that LM-1 was produced by a deletion in the phylad leading to section *Luxuriantes*. This experiment was repeated using *Tripsacum pilosum* as the outgroup, and we obtained the same result.

For *Pst*I, a 3.45 kb fragment found in *Z. perennis, Z. diploperennis* and *Z. luxurians* cpDNAs is replaced by a 3.25 kb fragment in *Z. mays* cpDNAs. This fragment has been designated *Pst*I-12 by Larrinua et al. (1983). The two size variants of *Pst*I-12 were cloned from *Z. perennis* and *Z. mays* subsp. *parviglumis* cpDNAs. Each was restriction mapped, which revealed that the fragment contains two length mutations, LM-2 and LM-3 (Fig. 2). Each of these mutations was subsequently sequenced (see below).

To determine if LM-2 and LM-3 caused other fragment length variants, the pUC clone of *Pst*I-12 was hybridized to a filter containing *Kpn*I digests of the cp-DNAs (not shown). This demonstrated that the small size difference between two approximately 7.5 kb *Kpn*I fragments (Fig. 1) occurred in the same portion of the genome as *Pst*I-12. The small size difference between the *Kpn*I fragments is probably a manifestation of LM-2 and LM-3.

The directionality of LM-3, which is located in the 800 bp *Eco*RI-*Pst*I, was investigated by DNA sequencing (see below). The directionality of LM-2, which occurs in the 1.6 kb *Pst*I-*Sal*I fragment (Fig. 2), was investigated by comparison of restriction fragment lengths to *Sorghum* and *Tripsacum*. Maize, *Z. perennis* and sorghum cpDNAs were double digested with *Sal*I and *Pst*I. It is known from the work of Dang and Pring (1986) that *Zea* and *Sorghum* have identical *Sal*I and *Pst*I sites in this region of the genome. Digests were electrophoresed and transferred to Gene Screen. The pUC clone of *Pst*I-12 was similarly digested and the 1.6 kb *Sal*I-*Pst*I fragment excised from a gel after electrophoresis. The 1.6 kb *Sal*I-*Pst*I fragment was radioactively labelled and hybridized to the Gene Screen filter (Fig. 4A). This shows that sorghum contains a fragment similar in size to the larger *Zea* fragment. By outgroup analysis as discussed above, one can infer that LM-2 arose as a deletion in the phylad leading to section *Zea* of the genus. A similar experiment was performed using *Tripsacum pilosum* as the outgroup and double digesting with *Pst*I-*Rsa*I (Fig. 4B). Again, the outgroup and *Z. perennis* have identically sized *Pst*I-*Rsa*I fragments, which are larger than the homologous maize fragment. This confirms the polarization made using sorghum as an outgroup and does so with a smaller (600 bp) fragment.

In Fig. 1, another small size difference is shown between a 17.0 kb fragment (*Pst*I-2) in *Z. perennis* and maize cpDNAs. Dang and Pring (1986) reported a 720 bp difference between maize and sorghum for this fragment and noted that the length mutation occurs in sorghum *Bam*HI-7. Our estimate of the size difference between maize *Pst*I-2 and sorghum *Pst*I-2 is only 250 bp, nearly 500 bp less than the estimate of Dang and Pring (1986). To further investigate this mutation (LM-4),
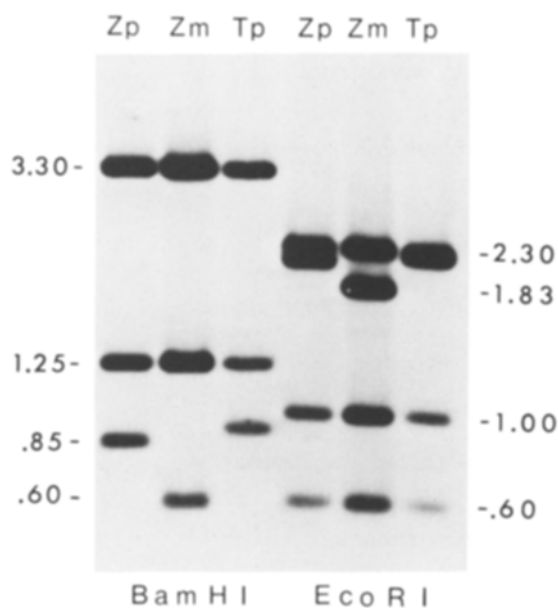
Fig. 5. Autoradiograph used to determine directionality of LM-4. *Bam*HI and *Eco*RI digests of *Zea perennis* (Zp), *Z. mays* (Zmm), and *Tripsacum pilosum* (Tp) cpDNA. The 5.36 kb sorghum *Bam*HI-7 fragment cloned in pUC8 (pLD7) was used as the probe

Fig. 6. Ethidium bromide stained gel and autoradiograph of Gene Screen filter with *Kpn*I and *Sma*I digests of *Zea perennis* (Zp) and *Z. mays* (Zm) cpDNAs. The 5.36 kb sorghum *Bam*HI-7 fragment (pLD7) was used as the probe. LM-4 produces the size differences seen in the autoradiographs

cpDNAs of maize, *Z. perennis* and *Tripsacum* were digested with *Bam*HI, electrophoresed and transferred to Gene Screen. pLD7, a pUC clone of sorghum *Bam*HI-7, was hybridized to the Gene Screen filter (Fig. 5). pLD7 hybridized to 3.30 and 1.25 kb *Bam*HI fragments in all three cpDNAs (Fig. 5). *Zea perennis* has an additional 850 bp fragment, *Tripsacum* a 930 bp fragment and maize a 600 bp fragment. This suggests that *Z. perennis* and *Z. mays* differ by a 250 bp length mutation (LM-4), but that additional alterations distinguish *Zea* from *Tripsacum*. Sorghum lacks the two *Bam*HI sites that create the three fragments seen in Fig. 5, thus, pLD7 hybridizes only to itself (not shown).

pLD7 was also hybridized to *Eco*RI digests of *Z. perennis*, maize, and *Tripsacum* (Fig. 5) and sorghum cpDNAs (not shown). Here pLD7 hybridizes to 2.30, 1.0 and 0.6 kb fragments in each cpDNA. Maize has an additional 1.83 kb fragment, *Z. perennis* a 2.07 kb fragment and sorghum and *Tripsacum* a 2.15 kb fragment. LM-4 accounts for the approximately 250 bp difference between maize and *Z. perennis*, however additional event(s) are necessary to explain the difference between *Zea* and the other genera. For this reason, LM-4 could not be polarized. LM-4 occurs in *Eco*RI-p′ of maize. The short form of LM-4 is found in all section *Zea* cpDNAs, and the long form in all section *Luxuriantes*.

LM-4 is located in a region of the genome that Larrinua et al. (1983) did not map for *Bam*HI, but referred to as a 1.9 kb gap. Our data show that this gap is occupied by 1.25 and 0.60 kb *Bam*HI fragments in maize, and that the length mutation occurs in the 0.60 kb *Bam*HI piece. We do not know the order of these fragments within the gap. Dang and Pring (1986) did not detect this 0.6 kb *Bam*HI fragment, and thus they concluded that the gap was occupied only by the 1.25 kb piece and they judged that sorghum and maize differ by a 720 instead of a 250 bp length mutation in this region.

Two other fragment length variants are apparent, one involving a 5.4 kb *Kpn*I fragment and another in a 8.0 kb *Sma*I fragment (Fig. 1). A southern hybridization of pLD7 to *Kpn*I and *Sma*I digests demonstrated that both of these fragments lie in the same region of the genome as sorghum *Bam*HI-7 (Fig. 6). Thus, the variation in length observed for these two fragments could be caused by LM-4 which lies in sorghum *Bam*HI-7.

In summary, all variation in fragment lengths can be explained by four length mutations. Other length mutations of similar size (> 50 bp) may exist among these cpDNAs, but could not be detected because they do not occur in restriction fragments of small size (< 10 kb). Another confounding factor is that two length mutations may occur in the same fragment such as reported

```
1    CTGCAGCAGCGTTTGAAAAGGGAATTGATCGCGATTTGGAACCTGTTCTTTACATGACCCCTCTTAACTAAGATTTTTTATTTATACCTGTTCTAGTTTT
                                    .T

101  TTTCTGTTCTGGCTCGGTTATTCCATCTAGCCGAGC.ATTCATGCTTTTTTATGAAAGAATGATATAAGGGGGGCTGCACAAAGAAAAACATAGATAGAA
                                     A               (.)      A

201  AGAAGCAAATGTATTCAATATACAAAAGGAGAGAGAGGGATTCGAACCCTCGATAGTTCCTAGAACTATACCGGTTTTCAAGACCGGAGCTATCAACCAC
     A

301  TCAGCCATCTCTCCACAGCCCAATCCCTATTTTATTCCTACAAATAGAACATAGCTATATGAAATGATCTACTAA.....CTCATCTCAGATGCAAGTCC
                                                                          CTCAT

401  ACTTTCAATATATCTCTGTATACTGTATACACGGATACAGAATCCGCTATATCCGTTGTGAAATAAA.GCTAAATCCCCTCCCCTCAACTCCATATCTAA
                                                           (.)      G                                C

501  ATAAAAGCGGTAAGGAAAAAGTTTTAATAAAAAGAAGAATCAATGGATTCATGATTAAACCCCTCCTACTTCTTGTATTTTCATACAATTTTGATTAAGT
                                                                              G

601  GAGGGATCAAATAGAAATATGTAGTCAACTTTATTTGAAGGTAGTTTGGAGGATTATAAATATGAATATTGCTTTCCAATTAGCTGTTTTTGCATTAATT
                                                    C      C

701  GCGAACTTCCTCAGTTTTAGTAATTAGTGTACCCCTTGTATTTGCTTCTCCTGATGGTTGGTCAAATAATAAAAACGTTGTATTTTCCGGTACATCATTA
     (.)

801  TGGATTGGATTAGTCTTTCTGGTAJCTATTCTGAATTC
```

Fig. 7. DNA sequence of the 831 bp *Eco*RI-*Pst*I fragment containing LM-3 from *Zea perennis* shown in Fig. 2. The *blocked region* shows the 80 bp segment deleted in *Z. mays* cpDNA. Differences between the *Zea perennis* and sorghum sequence are shown beneath the *Zea* sequence. *Arrows* show small inverted or direct repeats. tRNA^ser is underscored

above for maize fragment *Pst*I-12. Thus, an accurate assessment of the size and number of length mutations can only be obtained by detailed restriction mapping or DNA sequencing.

DNA sequence analysis

To investigate the mechanism generating the length mutations, two mutations were chosen for DNA sequence analysis. First, the 0.8 kb *Eco*RI-*Pst*I fragment, which contains LM-3 (Fig. 2), was cloned into M13 from cpDNA of *Z. perennis* and *Z. mays* subsp. *parviglumis*. This fragment was also subcloned from *Sorghum bicolor* pLD4 into M13. DNA sequencing revealed an 80 bp deletion in subsp. *parviglumis* relative to *Z. perennis* and sorghum (Fig. 7). The fragment was 831 bp in *Z. perennis*, 751 bp in subsp. *parviglumis* and 835 bp in sorghum. By outgroup analysis, these data suggest that LM-3 arose as a deletion in the phylad leading to subsp. *parviglumis* and other section *Zea* cpDNAs. The sequences from *Z. perennis* and subsp. *parviglumis* are identical for this fragment with the exception of the 80 bp deletion. The sorghum sequence differs from *Z. perennis* by five single bp deletions (three in sorghum, two in *Zea*), one 5 bp deletion in *Zea* (part of a small direct repeat) and seven separate single bp substitutions (Fig. 7). The seqeunce was analyzed with the "Wordsearch" program (Devereux et al. 1984) which revealed a tRNA^ser gene previously sequenced in maize (Krebbers et al. 1984). Our analysis demonstrated that the tRNA^ser sequence is the same in sorghum as in maize. The sequences were also analyzed for direct and inverted repeats using the "stem-loop" and "repeat" programs (Devereux et al. 1984). Numerous small direct and inverted repeats are scattered throughout the sequence. Those occurring near the end-points of LM-3 are shown (Fig. 7), however, there is no obvious relationship between these repeats and the deletion end-points.

The 1.6 kb *Pst*I-*Sal*I portion of *Pst*I-12 contains LM-2 (Fig. 2). This 1.6 kb fragment was cloned from cpDNAs of *Z. perennis* and subsp. *parviglumis* in pUC9. Restriction analysis of these clones revealed that LM-2 resides in a 600 bp *Pst*I-*Rsa*I fragment (Fig. 2). This 600 bp fragment was subcloned into M13 and sequenced. This demonstrated that subsp. *parviglumis* has an 83 bp deletion relative to *Z. perennis* (Fig. 8). The two sequences also differ by two single nucleotide substitutions (Fig. 8). The sequences contain a number of small direct and indirect repeats, but they bear no obvious relationship to the deleted segment. Those occurring near the deletion end-points are shown (Fig. 8). No genes or open reading frames were found in this fragment.

The two sequences containing the LM-2 and LM-3 were compared using the "Dot-Plot" procedure (Devereux et al. 1984) to check for any similarity in sequence

```
1    GTACAGTTCCCATAAATCCCTTTCTTCCTTTTCTTTTTTGTTCAGAATTGAACAAAGAAATTGCGAAAGATGTTTTCTTCCTCCACGTATCATGAAGTGC
     C .

101  GAGCCATAGGGAAGGAGTGAGATGACTTTCCAAAATTTTTCATAGACTTCGTCTATCGCTTGAGAGAAGCAACAAAGAGTAATAACTTAAAAAGAAAAGC

201  AGATACGAATCG ACAGATTTACCTGATGAAAATTTACATCAGAGTCCTCTGATGAGGATTCCTCAAACTCTTCATAAAGAAGATCCGGAAAGTCT CTGGT

301  TAGTCGATCCTCTCCATTTCCTAATTTCCTCTCTTCTTTTCCGCTCAATTCTAGTTTATTAGATTCTTGTTTAAAAGAATCAAAGAAGATGAATAGAACT

401  AAGAACACATAAAAAAGAGCATAGAGAACCATTACCAAACGTTCCTCCTAAGAATCATATTGGGTATCTGTCCCCTTCCTTTTAACCCTAGGATCGAAAA
                                                                                         A

501  TCTAGAATCCTCCCTTTTCCTAGTGTTCGGAAACGGAAAACCCTGAACCTGGAGGAGTTAGGTATGTAGGATATGCTTTTTATTTTTTGTTGGGCAGGCA

601  GTAGTATAATTTTGATACTCTGCAG
```

Fig. 8. DNA sequence of the 625 bp *PstI-RsaI* fragment containing LM-2 from *Zea perennis* shown in Fig. 4. The *blocked region* shows the 83 bp segment deleted in *Z. mays* cpDNA. Other differences between the *Z. perennis* and *Z. mays* sequences are shown below the *Z. perennis* sequence. *Arrows* show small inverted or direct repeats

that might occur in or near the deleted segments. No significant similarity could be detected between the two sequences. There is no similarity between the end points of the two deleted segments (Fig. 7 and 8). The GC content of LM-3 is 40% and the GC content of LM-2 is 37%. These percentages do not differ appreciably from the estimated 37–39% GC content for the chloroplast genomes of angiosperms.

Discussion

Medium-sized length mutations are well-documented in chloroplast genomes of angiosperms. These mutations have been analyzed primarily by restriction mapping and little is known about their directionality (insertions vs. deletions) or the mechanisms which generate them. Length mutations occur most commonly in the large unique sequence region of the genome. Some evidence suggests that they occur preferentially in small regions of the genome ("hot spots") prone to this type of mutation (Kung et al. 1982; Palmer et al. 1985).

We have described four fragment length mutations among eight cpDNAs representing the range of taxonomic variation in *Zea*. The directionality of three of these (LM-1, LM-2, LM-4) was investigated by comparison of restriction fragment lengths using *Tripsacum pilosum* and *Sorghum bicolor* as outgroups. The fourth (LM-3) was polarized by DNA sequencing using sorghum as the outgroup. These experiments suggested that LM-1, LM-2, and LM-3 all arose via deletions. Our evidence is most secure for LM-3 which was sequenced and is least secure for LM-1 which was polarized by comparison of restriction fragments greater than 2 kb. LM-4

could not be polarized because the 600 bp *Bam*HI fragment in which it occurs differs from the outgroups by an additional mutation(s). Once length mutations are polarized in other genera, it will be of interest to determine if insertions or deletions occur with equal frequency or if one of these predominates.

Each of these length mutations occurs in the large unique sequence region of the genome. They occur in the same regions of the chloroplast genome in which Bowman et al. (1983) reported numerous length mutations in *Triticum*. Further, LM-2, 3 and 4 occur in a region of the genome in which several inversion endpoints in the cereal genome have been found (Quigley and Weil 1985). It appears this region should be considered a *hot spot* for length mutations in the *Zea* chloroplast genome.

LM-1 is a deletion in all taxa of section *Luxuriantes* relative to the taxa of section *Zea*. LM-2 and 3 are deletions in the taxa of section *Zea* relative to section *Luxuriantes*. LM-4 may be either a deletion or insertion. The long form of LM-4 is found in all section *Luxuriantes*, and the short form in all section *Zea*. Thus, the four length mutations support previous taxonomic judgement that *Zea* could be divided into these two sections (Doebley and Iltis 1980; Doebley et al. 1984).

DNA sequence analysis of two of the length mutations failed to elucidate the mechanism which generated them. The deleted regions are not members of arrays of tandem repeats nor are they flanked by direct repeats. The two deleted segments do not share any apparently significant sequence homology. The same may be said of their flanking regions. The GC content of the deleted segments approximates the typical values for angiosperm chloroplast genomes. Bowman and Dyer (1985) reported

that *Triticum* has several small repeats which have one segment in the large inverted repeat and the other in the region where length mutations are frequent. Such repeats could play a role in the length mutations described in this paper. We have not examined the *Zea* chloroplast genome for such features.

References

Anderson S, Gait MJ, Mayol L, Young IG (1980) Nucleic Acids Res 8:1731–1743

Birnboim HC, Doly J (1979) Nucleic Acids Res 7:1513–1524

Bowman CM, Bonnard G, Dyer TA (1983) Theor Appl Genet 65:247–262

Bowman CM, Dyer TA (1985) Structural variation and evolution of the chloroplast DNA of cereals. First International Congress of Plant Molecular Biology Abstracts, Athens, Georgia, pp 124

Dagert M, Ehrlich SD (1979) Genetics 6:23–28

Dale RMK, McClure BA, Houchins JP (1985) Plasmid 13:31–40

Dang LH, Pring DR (1986) Plant Mol Biol 6:119–123

Devereux J, Haeberli P, Smithies O (1984) Nucleic Acids Res 12:387–395

Doebley J, Goodman MM, Stuber CW (1984) Systematic Botany 9:203–218

Doebley J, Iltis HH (1980) Am J Bot 67:982–993

Gordon KHJ, Crouse EJ, Bohnert HJ, Hermann RG (1982) Theor Appl Genet 61:373–384

Iltis HH, Doebley J (1980) Am J Bot 67:994–1004

Koller B, Delius H (1982) EMBO J 1:995–998

Krebbers E, Steinmetz A, Bogorad L (1984) Plant Mol Biol 3:13–20

Kung SD, Zhu YS, Shen GF (1982) Theor Appl Genet 61:73–79

Larrinua IM, Muskavitch KMT, Gubbins EJ, Bogorad L (1983) Plant Mol Biol 2:129–140

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory, New York

Messing J, Vieira J (1982) Genetics 19:269–276

Palmer JD (1985) Evolution of chloroplast and mitochondrial DNA in plants and algae. In: MacIntyre RJ (ed) Molecular evolutionary genetics. Plenum, New York, pp 131–240

Palmer JD (1986) Isolation and structural analysis of chloroplast DNA. In: Weissbach A, Weissbach H (eds) Plant molecular biology. Methods in enzymology, Academic, Orlando, Florida, 118:167–186

Palmer JD, Jorgensen RA, Thompson WF (1985) Genetics 109:195–213

Palmer JD, Shields CR, Cohen DB, Orton TJ (1983) Theor Appl Genet 65:181–189

Quigley F, Weil JH (1985) Theor Appl Genet 9:495–503

Sanger F, Coulson AR, Barrell BG, Smith AJH, Roe BA (1980) J Mol Biol 143:161–178

Sanger F, Nicklen S, Coulson AR (1977) Proc Natl Acad Sci USA 74:5463–5467

Shaffer HE, Sederoff RR (1981) Anal Biochem 115:113–122

Takaiwa F, Sugiura M (1982) Nucleic Acids Res 10:2665–2676

Whitfeld PR, Bottomley W (1983) Annu Rev Plant Physiol 34:279–310

Wilkes HG (1967) Teosinte: the closest relative of maize. Bussey Inst, Harvard Univ, Cambridge, Massachusetts

Zimmer EA, Newton KJ (1982) A simple method for the isolation of high molecular weight DNA from individual maize seedlings and tissues. In: Sheridan WF (ed) Maize for biological research. Plant Molecular Biology Assoc, Charlottesville, Virginia, pp 165–168

Zurawski G, Clegg MT, Brown AHD (1984) Genetics 106:735–749