The EMBO Journal vol.5 no.11 pp.2781-2788, 1986

DNA sequence analysis of a 5.27-kb direct repeat occurring adjacent to the regions of S-episome homology in maize mitochondria

J.P.Houchins, H.Ginsburg, M.Rohrbaugh, R.M.K.Dale, C.L.Schardl[1,2], T.P.Hodge[1] and D.M.Lonsdale[1]

Botany Department, University of Minnesota, St. Paul, MN 55108, USA, and [1]Department of Molecular Genetics, Plant Breeding Institute, Cambridge, UK.

[2]Present address: Department of Plant Pathology, University of Kentucky, Lexington, KY 40546, USA

Communicated by C.J.Leaver

The DNA sequence of the 5270-bp repeated DNA element from the mitochondrial genome of the fertile cytoplasm of maize has been determined. The repeat is a major site of recombination within the mitochondrial genome and sequences related to the R1(S1) and R2(S2) linear episomes reside immediately adjacent to the repeat. The terminal inverted repeats of the R1 and R2 homologous sequences form one of the two boundaries of the repeat. Frame-shift mutations have introduced 11 translation termination codons into the transcribed S2/R2 URFI gene. The repeated sequence, though recombinantly active, appears to serve no biological function.

Key words: plant/maize/mitochondria/recombination/repeat

Introduction

The mitochondrial genomes of higher plants range from ~200 kb to >2400 kb (Ward *et al.*, 1981). The mitochondrial genomes of *Brassica campestris* (turnip) and *Zea mays* (maize, fertile cytoplasm) have been determined to be 218 kb (Palmer and Shields, 1984) and 570 kb (Lonsdale *et al.*, 1984a), respectively. Both of these genomes can be organised as a single circular molecule, which has been termed the 'master circle'. The turnip master circle contains two copies of a 2-kb sequence which have the same relative orientation. The genome can be represented as either a single circular molecule (master circle) or as two subgenomic circles of 135 kb and 83 kb. These three forms of the genome can be postulated to interconvert by homologous recombination between the 2-kb repeat elements. In maize, six pairs of repeated DNA sequences have been identified on the master circle of 570 kb (Lonsdale *et al.*, 1984a). Five of these repeated sequences, of approximately 1 kb, 2 kb, 3 kb, 10 kb and 12 kb, are directly orientated and all, except the 10-kb repeat, appear to be involved in homologous recombination events. This leads to a predicted complex multipartite genome structure with subgenomic circles ranging in size from 47 kb to 523 kb (Lonsdale *et al.*, 1984a).

In maize, in addition to the normal (N; fertile) cytoplasm, three genetically different cytoplasmic types can be identified on the basis of pollen fertility restoration characteristics: C, S and T (Duvick, 1965). These cytoplasms in combination with particular nuclear genotypes give rise to plants that are unable to produce functional pollen (male-sterile).

The mitochondrial genome of the S cytoplasm, in addition to the high mol. wt mtDNA, contains two self-replicating linear DNA molecules of 6397 bp (S1; Paillard *et al.*, 1985) and 5453 bp (S2; Levings and Sederoff, 1983). Like many other linear replicons they have blocked 5′ termini due to the covalent attachment of a polypeptide which is thought to be involved in the priming of DNA replication (Kemble and Thompson, 1982; Sederoff and Levings, 1985). The S1 and S2 episomes have identical terminal inverted repeats of 208 bp. A sequence homologous to the terminal 186 bp of the terminal inverted repeat is found at two locations in the mitochondrial master circle. Recombination appears to occur between the terminal inverted repeats of the episomes and the 186-bp homologous sequences in the mitochondrial genome creating a number of linear chromosomal molecules possessing terminal S1 or S2 sequences (Schardl *et al.*, 1984). The mitochondrial genomes of the C and T cytoplasms lack substantial sequence homology to the S1 and S2 episomes (Thompson *et al.*, 1980), though significant sequence homology is present in the mitochondrial genomes from normal fertile lines (McNay *et al.*, 1983). Two of the four unique sequences flanking the 3-kb repeat have substantial sequence homology to the S episomes and these four unique sequences exist in four paired combinations present in equimolar proportions (Lonsdale *et al.*, 1983b). Homologous recombination between the 3-kb repeat elements can therefore be predicted. Recombination of plant mtDNAs has been demonstrated in interspecific fusion hybrids (Nagy *et al.*, 1983) and can be inferred to be an active process involving mitchondrial genomes leading to the predicted complex structures and the multimeric forms of the small plasmids (Dale, 1981).

We have determined the nucleotide sequence of the two regions of the Wf9-N maize mitochondrial genome that contain the two copies of the 3-kb repeat in order to try and shed more light on recombination in relation to the adjacent S1 and S2 sequences. The repeat was found to extend to 5.27 kb and hereafter will be referred to as the 5-kb repeat.

Results and Discussion

The maize mitochondrial genome contains 570 kb of DNA on a circular molecule, as derived from a series of overlapping cosmid clones (Lonsdale *et al.*, 1984a). The mitochondrial genomic map in Figure 1 shows the location of the alpha and beta copies of the 5-kb repeat along with the positions of several mitochondrial genes. Lonsdale and colleagues (Lonsdale *et al.*, 1983b, 1984a) proposed a recombination system which operates on some of these repeated sequences giving a complex multipartite structure. Homologous recombination between the 5-kb direct repeats gives rise to subgenomic circles of 248 kb and 322 kb (Figure 1). The relative molar proportions of the recombinant products appear similar suggesting that recombination between the 5-kb repeats plays a major role in determining the overall organisation of this genome (Lonsdale *et al.*, 1983b).

S1- and S2-related sequences

Figure 2 shows an expanded map of the regions containing the

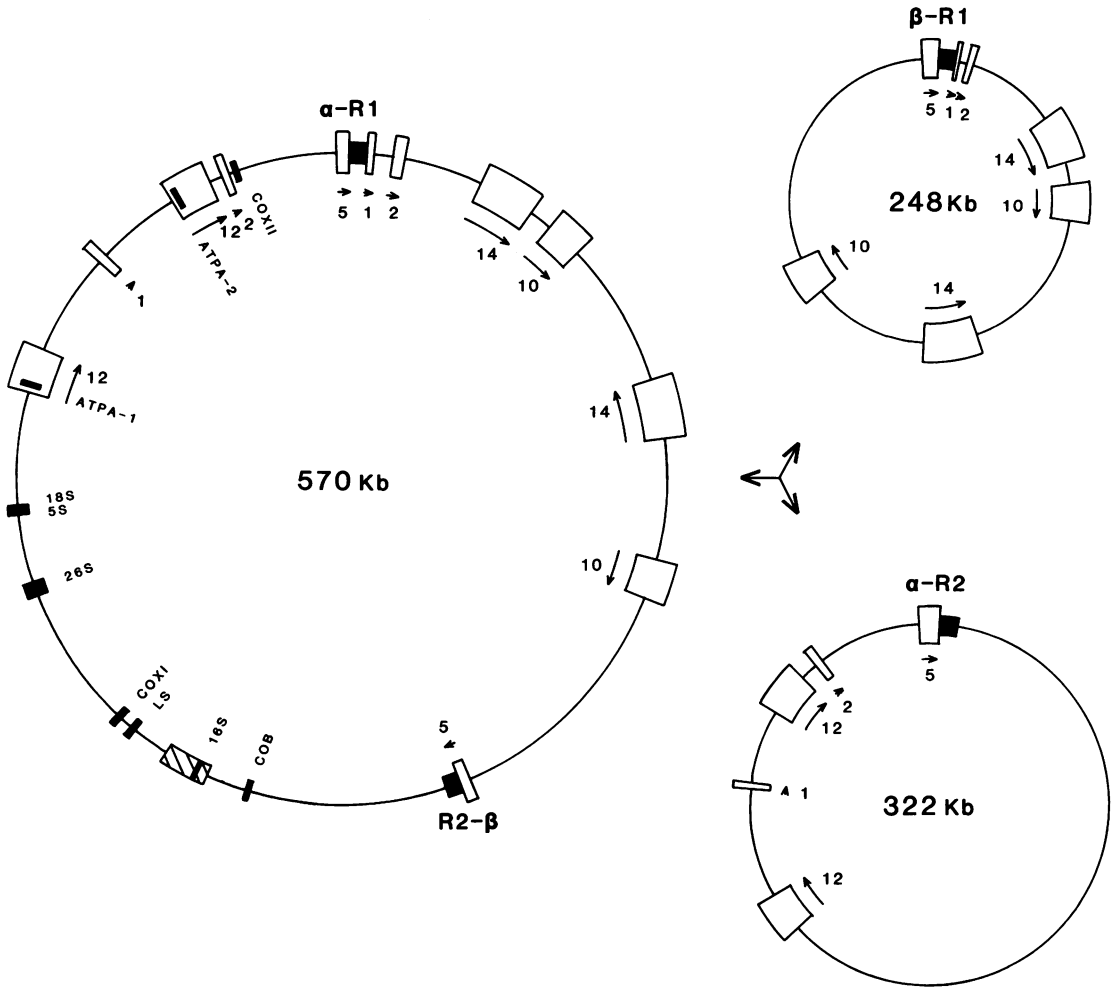

Fig. 1. The mitochondrial genome of the Wf9-N fertile cytoplasm of maize. The approximate size and orientation (arrows) of repeated DNA sequences are shown (open boxes). The black boxes represent sequences whose origin and function have been identified. These include the known mitochondrial genes (COXI, COXII, COB, ATPA-1, ATPA-2 and the 26S, 18S and 5S rRNAs). Sequences homologous to chloroplast DNA are also shown (see Lonsdale, 1985); LS: gene for the large subunit of ribulose-1,5-bisphosphate carboxylase; hatched box: 12-kb sequence homologous to the inverted repeat of the chloroplast genome which contains the 16S rRNA gene. The positions of α-R1, β-R2 sequences and the recombinant forms β-R1 and α-R2 flanking the 5-kb repeat are indicated.

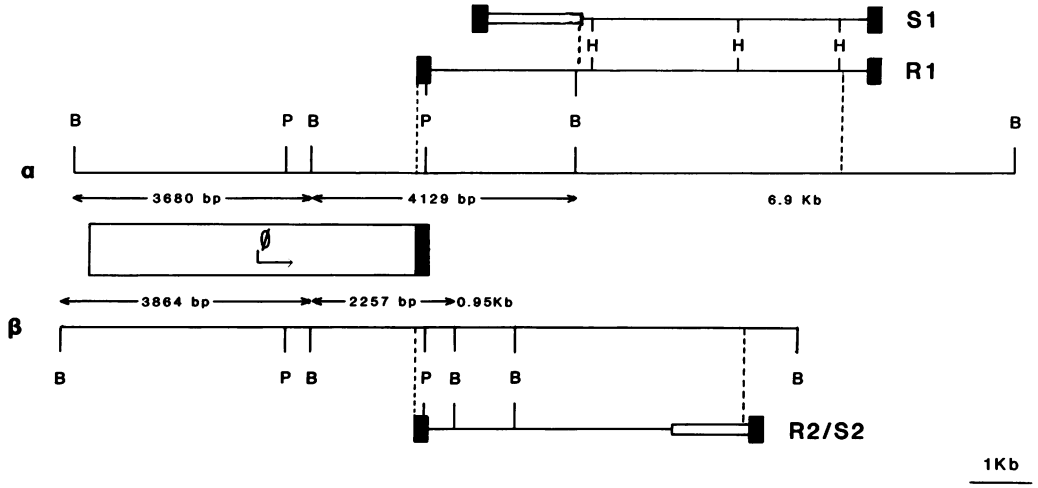

Fig. 2. Regions of the mitochondrial genome having homology to the R1, R2, S1 and S2 mitochondrial episomes. The sequenced *Bam*HI (B) fragments containing the α-R1 and β-R2 copies of the 5-kb repeat are shown, as are the *Bam*HI fragments with homology to the R1, S1, R2 and S2 episomes; the terminal inverted repeat homology is shown (black boxes). The dotted lines indicate the ends of the homologous sequences. S1 is suggested to have arisen from an intermolecular recombination event between R1 and R2 (Levings *et al.*, 1979), this replaced the terminal region of R1 with part of the R2 sequence (open box). Other than this homology to R2, S1 is homologous to R1 and to the integrated R1 homology sharing a minimum of the two internal *Hin*dIII (H) fragments. The position of the 5-kb repeat and the zero coordinate of the restriction map located at the *Sst*II cleavage site is shown (ϕ). The positions of the two *Pst*I sites (P) delimiting the 2173-bp fragment used in these studies are shown.

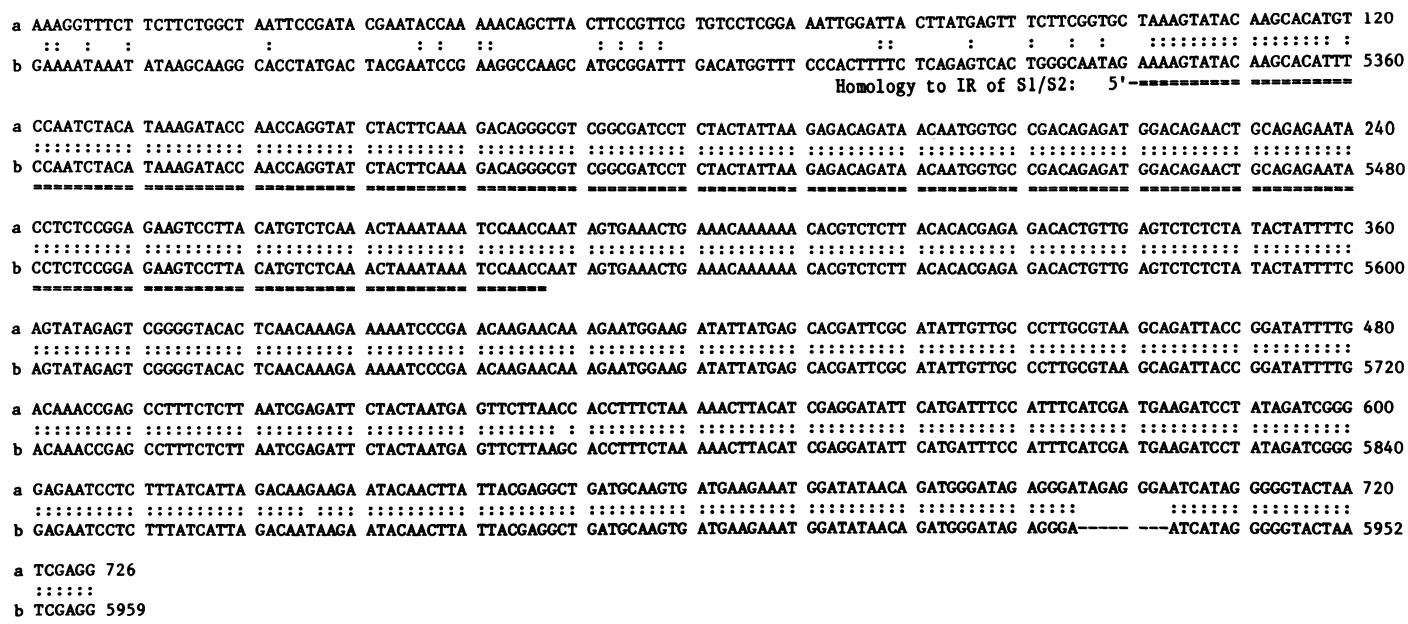

Fig. 3. Comparison of the α-R1 sequence to the $\delta-\delta'$ sequence of the S mitochondrial genome (Schardl *et al.*, 1985). (**a**) $\delta-\delta'$ sequence base pairs 1–726; (**b**) α-R1 sequence, base pairs 5121–5952 which contain the homology to the R1 episome. The sequence homology to the terminal 187 bp of the S1 and S2 terminal inverted repeat is indicated (=).

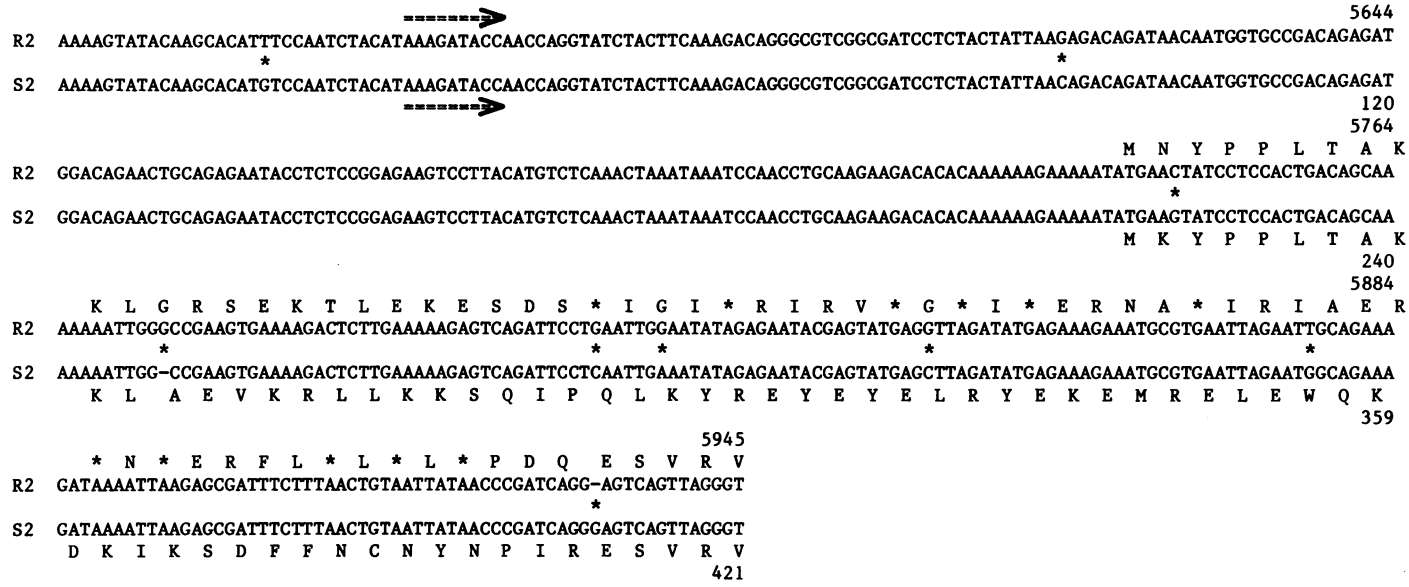

Fig. 4. Comparison of the β-R2 sequence, base pairs 5525–5945, with the first 421 bp of the S2 episome (Levings and Sederoff, 1983). The transcription start site of the S2 ORFI is indicated (= →, Traynor and Levings, 1986). Differences between the two sequences and the effects on the open reading frame (ORFI) translation are indicated by (*).

alpha and beta copies of the 5-kb repeat, as well as the regions homologous to the linear replicons of maize mitochondria.

In addition to S1 and S2 episomes of S cytoplasm which are mainly associated with North American maize accessions, numerous South American maize accessions contain two linear replicons, designated R1 and R2 (Weissinger *et al.*, 1982). Both R1 and R2 can be distinguished from S1 and S2 on the basis of size and restriction site polymorphisms (Levings *et al.*, 1979; Sederoff and Levings, 1985).

In North American male-fertile (N or Normal) cytoplasms, including Wf9-N, for which a restriction map is known, sequences homologous to S1 and S2 reside adjacent to the 5-kb repeat (Figure 2). The homology to S1 and R1 is contained within a 6.9-kb *Bam*HI fragment (Figure 2; Levings *et al.*, 1983; Lonsdale *et al.*, 1983a; McNay *et al.*, 1983). The adjacent 4129-bp *Bam*HI fragment, hybridized only weakly to S1 probe, to S2 and to the 2257-bp *Bam*HI fragment (Spruill *et al.*, 1980; McNay *et al.*, 1983). This weak homology results directly from the terminal 187-bp homology to the inverted repeat of the S1 and S2 episomes. However, the 4129-bp *Bam*HI fragment showed strong hybridization to R1 and heteroduplex analysis revealed ~2.4 kb of homology with R1 (Levings *et al.*, 1983). The DNA sequence of the 4129-kb *Bam*HI fragment immediately 3′ to the terminal inverted repeat homology extending to the *Bam*HI site is therefore derived from the R1 episome.

In the mitochondrial genome of the S cytoplasm two sequences

```
b GGATCCGATT CAATACTTCC TAGATTTACG GGTTGTTCTT CGCTGAGAGC TTCTTCACTC AATTACCAAG ATTCAATCGA CTTTCTTTTC CGCTCGCTGC CCGATGCGTT CTTTATTAAA 120

a                                                                                                   GGATCC AGTGGATAGA TACTTTAAAC GATGGTAGGA ACGGGAGCTA CCCTGGATCA 56
b GCCCAGCTAT ACGTGGAAGG ATAGGTGGAA GGCACATGCT TCTCTTTACG ACGTAAAGGA AAGACCTCTT TTATAATTCT ATACTAGTTC CTATGGCGTA AAGGAAAAGA GTTTCATTTC 240

a CTTGAGGAAG TAAGAACGCA AGTCTCTCAG TTGAAATTAG ATCGTTCCAG GATAATAGCT GAAGCGCTGA GAACTCCCAA AGAGAAGAGA ACCAGCTTAA GGAGCAGGTA TCCGGAGTAT 176
b GGCTATTACA TCTATGGGTC AACGAGTTTC TGGAGTCTCT TAAAGAGGGG ATCGAAGCAA GGTCCAATCC CATTATTCTA AATTATTCTA ATACGAGTAT CTCACAGATG AAGAAGTGAG 360

a CAGCAAGCTA TAAATAGCTT TAGGATTGTA TGAATAGAGT TGAGGATCAA GTATCCCTCA TATTTAGATC CCCCGTTGGG AAAAGCCCCT TTTTTAAATC CTTTCACGCT TTCCCTTTGC 296
b CAAGCCTTTC TCCAATGGAT TCTAACAGCG CAGACTAGGC ATCTTTATCA GATTTCCATT TCTATATTAG TTTAGTTTAG TAAAGCCCCT TTTTTAAATC CTTTCACGCT TTCCCTTTGC 480

b CGAAGGCCGG CCTCTTTAGT TCCTACGTTT TCGGAAAAGA AGGTGAGACG CTAGCTAGCA ACGGGAAAAA ACGAGTAGTT AAAGATGTGA ATGTAAGACT ACTCGAAAGT ACTTCCATTT 600

b TTATTATATT ATATAGTTTC TAATAGAGAG TACAGTTGTA TCTATGTGCG AAACGACGAA AGAAAGATGG AAGGCCCAAA GAGCTCTACA GTAGTGCCCC GGAATGGGGT CGTAGAGCCG 720

b GTAACCCACT CTGTTTAGTA TTTTGATTTC GTTCAAGGTG TTGCTCTGTC AAATAGAGAT TGTGTGGGTG TTCAGTCTAC CACTCATGTT CACGTTCTAC AAATCTAAGG GAGTGCTCCT 840

b AGACTGGACG ACTTCCCGTA AATAACCTAG TTCATGTTCC TTTGAAAGAA ATAAAATGAA AGCTTTTGAA GCGAAGATGT CAACGCGGTG CTAGGTTTGC GCCGGCCACG ATTATTTATA 960

b AGTATATTGG AAAGAAGAGT GACTAGGCTA CGGTCTAACT TTTTCTTCCT AGCGGAGTTC AATACGAAAA TAAAGGCGCT CCGCGTTGGG AATGGCGTAC GTAGTACGGG TGGTATTTCC 1080

b GGTATAACTG ATCAGGGTTG TTCTGAAAGC AAGTCAATTG GTGCTTTGGA AAGTGAGTGC CAGAGGCTAG TCACTGCTTG GTTTACAGAA GGGAAGGAAG AAAGAGATAC CGAAGCAGAT 1200

b TCAGTATAGA AAAGAAGAAA TTGATACCAT TCATTCCAGC TTATTTGATA CCCACTTAAA GTTTCTATCA AACCATGTCT TTTTCTTCGA ACGTCAATCT CGTAGGAATT TAGATTGGAA 1320

b GTGTGGTATT GAAGGGCGAA GCCTTTACTC TTTTACTTGT TTCTTTGAAT ATAATATAAC ATAGTATTGA CTACGCACTT CGAAAGATAA TGGTGTATTC CGAGGATCAG TCCGTTTTTT 1440

b ATTGAATTCG CAAATTGATA GATCTGGATG ATTAAAAGAA TGATTCTTCC CTGGGTATTT CCCCGCTACC ATTATGGAAA TTGGAGCTTT TCAATTCCAT TTTGGAGATG CTCGTAACGA 1560

b TTTTTTTCTT TTGGGTGGGC AAGAAAAGGA TTTGACTCTT GAAGATATCA ACTCATACGC CTGCTCCATA CATTAAAAAA CATTTTGGGT ACTTTCTATT AGAATGGATA AATAGAAATG 1680

b GGTACATTCC TCTTTCTCTT TCCTGGCCAA ACTACCAAAA AAAAAGAAGA GAGAGAGCTG TAAGAGCGGT AAAAGCAAGC TCGGTGGCCC GGTTCACTAC AGGTTGATGT ATCCTTTTGA 1800

b AAGAAAAGTC AATCTCTCTA AAAAATAAGC ACTTCAACAA GTCAGGGCCC TTTTTTATAC AACTCCTATG CGTAATTGAT AGCTTTTTGG TATTCTCCAT ATATACTACT AGGTACGAAA 1920

b TAACAGGAAA GATAGAAATA GTGACTTTGG CTTCAAATAG AGAGCAGCTT AAAGGGAAGG AGGACAGCTC TGAAAAAATA ACTGAACACT GGCTAGGAAC TGGGCTGCAC TAAGCCCGGC 2040

b ACTTTTTGGG CTTCTTCATG AATCTCGTTT TCAGGTAAGG TCAGACTTAT TGAATATATA TCTGAAAGAA AGCAAAGGAA TGAATCGTTT TTCCAATTAC CTTCTCTCGA AAACCATAGA 2160

b TCTTCTGGTT ACCATTGACT ATCCGGCACT TTCGTCTGTA GAGCGAAGCA AGAAATTGCA TGGAAGGGCT TGCCCTGCTA ATATCAGCCT TGTTTGGAAC AATAGATGCA ATCCATTTCA 2280

b GTGTATGTCA CTGAGTTCAC CATTGCAGTT GAGACGTCAG GCAATTGAGC GTTCCCTAGA GTTTGGTTCA AAGGCTAAGG ACTCCCCGCC GCCTTCATAA GGGCACTAAG GCGGAGCACG 2400

b GAAAAGAAAA TGGGGCACCC ATGTTTTACC AGTCCCGGAT CTCAGGGCCG TTTTGCACAT GAACGAAATC GAATTTCATT GCGGGAAATT TGGATTTCAT GAGGGAGGAA AAGGCAAGGC 2520

b GGAGGGTGGA GAGTCGTACA ATGCATTGGT GGGATGAAGG TGCTAGCTTG GCTGAGATAG GAAGTTTCGA TGTGTATGTA GATTTGGTTC TAGTGCCCCC ACCCTTTTCG AGATCTGTCA 2640

b GGATTTAAAA GATGTTATAG TGTAAACTAT AGAGAAATGG AATGAATGAA ATGTATCTAA GAAGAAAGGA AGCCATGCTC CATCCTGTTT ATAAAAAGAA CATCACCTCC CTTCCCTTTT 2760

b TGGTTAGATA CCGCTCCGTT AGGTACTAAT GCTTCTACCT ATCACCCTCC GGGTGAGTTT GAGAGCTGTG TTTTTTCCAA CGTTAATAGC ATTCCGCGAG AAAAGAGAAA AGTCATCTAC 2880

b TGATGTTGTC AGCGGAGTCC CTCGTCCATC CATGTATGAA TAGCGGTATC CCCCATTTTG GACAGTGATG AGGTAGTCGA CGCAATTTGC ATGTGTATTT GCGTTCGGCT TTGCTACTTT 3000

b CCTTCTAGAC TATAAAAGAA TGTCCGCGGA AGGGAATAGT CTACGTGGCC CGGGCCCGGG GTAGTCTTTC AATTCCTTTA TCGGGTGGGA GAGTTTAGAA CCTGTTAAGC CAATAGCTGC 3120

b AGCTTTAGTC ACACTAGCTA CATCAGTGGA TCATGGATTT AGCATACCAC CTCAGAATAG TCTACGTGGC CCCTACTTTT ATTTTACTAC TTTTTTTTCC AGTAATGCAG ACAGCCCTTT 3240

b TTAAAGCGCT AGGCCGGGCA GTTGGTACGA AAACAGAACC ATTAGTGGAG TAAGATCCCG ATCAAAGCAG AGCGGTCTAA TCGAAATAAT CATATCTAAG AGCAGGCAGG CACACGACTG 3360

b GATATCAAAG GTGGGGCTCC CCTCCTCTAC TCTACTCTAC TCTACTAAAT GAAATGAAAT AGTATTTGTA TTTGGGTCGT CTTTGTGTTC GATTCGAATA GAAGCGCCAT TGTCATCCGC 3480

b TGCAGTTCTT CATAGTCCGT CCTAGTAAGC TGGTAGGTTG GTGGATGTGA ACATAAACTC CTCATATTTC CTACTGTTGC TTCCTGTATA AAGGAAGAGT TGTCAGACAA TGATACTCTT 3600

b TCCAAATCTG CAATCTTCGT CGTTGTAATG CCTCAGTCAT CAATAGTAGA TAGTTGCTCA ACTATGCAAT AAGGATTTTC CTCTCCTCAC AGATTGGAGG ACTTTCACTT GACTCACTAG 3720

b TGGGTTTTGC CTCGGGAAAC AAGAATATGG ATTTGAGCAA GTCTGTTTTC CCATGGGCCT ACTACCTGCT ATTTTCAAAG CGTAGAGGAC CCGCCTTTCT GATCAAAGTA TGCTTCCCTT 3840

b ATGGGCGAAA CCACTGAATC GGACGGATCC TTTTTTCTTG GATTGACCAG GCCAGCTTCT CCACCCAATT CTTTTCTTAC GAGACGAGAT AGGCACAGTG TCAGGGTCAT GTGATGTGAA 3960

b TGGTTTGTTT GATCTGTTTC TTGTCCCTAT GGGTTCGATT GGCAATCAGG ATATCGGTCG CTTGGCACTA GTCGAATAGG CGATTCAGTC GTGATAGTGC CATCCCTTGC CCCTTGGAGT 4080

b GAGTTGAAAG CACTCTTCCA GCTCTTGATT CGCGGAATTC ATTACGAAGC AAGATTTCAT TTCATTTCCA TTTACGTCAT TAGCAAAACA AGCAAGATTC ACCTCAGGAT ACGACGAATT 4200

b CTGGAACACT TCCTTGATTG ATATATGCAT TTTCTAGAGC AGGCACCATT GACTGCTTCT TCAACGTCTA TACTATACGA GACGAGACGA GTTATTGTAC GAGTCAAGCT TGCTCCAGAC 4320

b CAAGCTCACT TTACACGGGC TTAGAACTCA ACCTGACCCA ACAAGAAAAC GGATAGACGC AGCGCAACGG CTTGATAGCC TGACTCAGCA AGAGGCGAGG ATGCACGTGA CTAACTAAAA 4440

b CACGTTTCGC GCTATCAACT TCTTTGATCG AACCTGGCCT GGCGATTGAA GCATTCGAGA TATAGAAGAA ACTGCCCAAG TCAACCGGAT ATTCCAAAGC AAGGGTTATG AATCAAAAGA 4560

b GAAGTGCCAG CTGGATCTCT GGTTCAAAGT TCAGGGCATG GAATTGATTC GTATGCTCGT TCTCTCGCTG AGCAGAACCA TCTCTGTCTG AGAAGACCCA GTCTCCCAGT GTTTCTAGAA 4680

b AAGGAAATCC CATTCGTCAA CCTAAATGGA AGGGCAGGTA AGGGAAGGCC AGCAGGAAAT TCAGTAAGTA AGGTATCTCA GTGCCAGTGC TCTCCGGTTT CAAGTTGATG CGGGCAGTCG 4800

b TGGGGACAGG CAAAGCTGGG ATGCCTTACT AGCAGGAATG AAAAGGTATG TATGAACCAA CAAAGGTCAA GTAGACGAAC ACATATCCAC TCCACTAGAG AAAGCGTCGA ACCAGGGAAG 4920

b AATGAATGGT TAGAACCAGT CAAGGGTGCC GTACTCATGC CTACGGGGAA AGAGACTGAT CCGTATCTGA TACGCCTTCA CTTCTTTCGA GCTTCTTATT CGATTTCTTA AATGACTTCT 5040
```

```
b TCGGGCCTTA CCCTAGACTA CCTTCACTTA ACTTAATGCA GAAGCCAGAG TTCGAGGAAA GCGGTAGTTG GTTATTTCGG GATGAATTTC TATTTCTAAA ATGTGGATGA ACATAAATGG 5160

b ACGGAAACGA GAAAAATGCA TAATTCATAT AAGTTCATTC GAAAGAATGA GGATTCATGA TGTCTCTTCC CAACCATTCT TCCTGAGGAA GGCACCTCCG ATCGAGTCCT TAGCGCCATG 5280

b GATATTCTTC GCAGTGCCTT TGGATGCTTC TCAGAAACCG CCTCTTTCCA CTGGCTATAT GAAATTCCAA ATGAGACTTC CCTGCGTCGC CAAAGGAATG GGAAACTTCT TGGTTCGAAA 5400

b CCTTTGATCC TATGCCGCCT GTCTGAAAAT AAATATAAGC AAGGCACCTA TGACTACGAA TCCGAAGGCC AAGCATGCGG ATTTGACATG GTTTCCCACT TTTCTCAGAG TCACTGGGCA 5520

b ATAGAAAAGT ATACAAGCAC ATTTCCAATC TACATAAAGA TACCAACCAG GTATCTACTT CAAAGACAGG GCGTCGGCGA TCCTCTACTA TTAAGAGACA GATAACAATG GTGCCGACAG 5640

a AGATGGACAG AACTGCAGAG AATACCTCTC CGGAGAAGTC CTTACATGTC TCAAACTAAA TAAATCCAAC CAATAGTGAA ACTGAAACAA AAAACACGTC TCTTACACAC GAGAGACACT 5576
  :::::::::: :::::::::: :::::::::: :::::::::: :::::::::: :::::::::: :::::::::: :   :   :  ::  : : :: :: : :  :  :     :  :    :  ::
b AGATGGACAG AACTGCAGAG AATACCTCTC CGGAGAAGTC CTTACATGTC TCAAACTAAA TAAATCCAAC CTGCAAGAAG ACACACAAAA AAGAAAAATA TGAACTATCC TCCACTGACA 5760

a GTTGAGTCTC TCTATACTAT TTTCAGTATA GAGTCGGGGT ACACTCAACA AAGAAAAATC CCGAACAAGA ACAAAGAATG GAAGATATTA TGAGCACGAT TCGCATATTG TTGCCCTTGC 5696
  :   :      :       :  :   :  :              :  :  :      :::          :  :  :     :: :::          :: :::: : ::    :: :     :  :   :
b GCAAAAAAAT TGGGCCGAAG TGAAAAGACT CTTGAAAAAG AGTCAGATTC CTGAATTGGA ATATAGAGAA TACGAGTATG AGGTTAGATA TGAGAAAGAA ATGCGTGAAT TAGAATTGCA 5880

a GTAAGCAGAT TACCGGATAT TTTGACAAAC CGAGCCTTTC TCTTAATCGA GATTCTACTA ATGAGTTCTT AAGCACCTTT CTAAAAACTT ACATCGAGGA TATTCATGAT TTCCATTTCA 5816
  : :::   :       ::    ::     :  :    ::            : :    : :  : :   ::    :       :   :  :           :  :    :
b GAAAGATAAA ATTAAGAGCG ATTTCTTTAA CTGTAATTAT AACCCGATCA GGAGTCAGTT AGGGTATCAA CGGATAGAGA CCCTGATTTA GAGGATGAAA AAAGAGAGCA GCTAGGAGAG 6000

a TCGATGAAGA TCCTATAGAT CGGGGAGAAT CCTCTTTATC ATTAGACAAT AAGAATACAA CTTATTACGA GGCTGATGCA AGTGATGAAG AAATGGATAT AACAGATGGG ATAGAGGGAA 5936
  :: ::: ::      :: :    :   :  :   :    :       :    : : :   : : :          :  : :    :          :    :  ::              :
b TCTATGCAGG CTGAATGGGA GAGACTAAGG CGGGGTGTAG AAGTCGGTAC TTCAGAAGAT ATCAATGTTG ACACAGTCAA ACGTTGGGGT TTGCAAAATA ATAAGTACAA CGCTGAACAC 6120

a TCATAGGGGG TACTAATCGA GGCCACAACG GTGGTGGCGA CGGTGAAGCT CTATTCATTC AAATAAGCAG GTTACTTATG AAATATATAA AGTATTATAT ACTCAACATA ACACATACCT 6056
  :
b TGGATCC 6240

a ATGATGGCCA TGCCTTGGTG GGTAAGCGGG ATACCATGGT GGATGGCTGA GAGTATCCTC ACACTCTTAT CGTTAGTATG TATACATTCT ATACTTCAAT TTAATCCTCT TATCTATAAG 6176

a GGCAATCGAT GTGATATTAT TTATAAGGTG CCGGATATTA CATTAATAGG AATAACCAAC CATATAACAA AGGAACCAAT TAAACAGCTA CAATGGCTCT ATTCTCTGAC ACCTGATAAC 6296

a TATCCTACTA AATGGATCTA TAGACTAAGA AATCTCTTTA GTTATGATGA GAATATTCCA TATAGGATAG GATTTCTAAG AATAAAGTGT TCTTTAGGCT TACGGAGTAT AAAAGGCTTA 6416

a GTATTAAGTG ATTCTGTATT ATATCAATAT GCCCATTTGG GAGCTGGCCT CAAATCTCTC ATTCCGCATC TAGAGTAGCC ATTTCATGCA GTGCACTCGT CATTTTGATG AAAGGCATCG 6536

a TAAGTATCTC TACGGTTTCT TTAATATACT TGTGTATTCT GGCGTAGGAT TAGTTGTTTG GTACTATATC ACACCACACC TTATACCTGA TTTAGGATTC TTCCGCGGGG TAAACCTAGA 6656

a TTCTTTTAAG GACCCAGCTT CGGTTGTTGA TGGGTACAAG AGGGTTTCGT ACATATTTAC AAAAGACGTC GTTAAGACTG TGTTAAGAAC TTGTCGAAAT GATACTTTTT TTCACAATTT 6776

a ACTAAATGAG GAATCAATCC AATTTTTTGA TAACGAAATT ATCAGCATAA CCGATTTGAA TTCAAACAGA CACCTTAATC TAGATCACTT CGAGGAGGTT GAAGCTAACA AGCTAGATCT 6896

a AGTAGCGTGT ATTGCAGTAG GTATTATGAG TGCTACCTTT ATAGCTACCA ATCTTCTTCC TCCTGGTACT ACTACGATAG TCCAATGAAT AAGGAGAAGG TCGTTATTTG TCAGTATTGA 7016

a GCCTGATAGC ATAATTCAGT ACCCGTGATG ATGGGCTCTA GCAAGGAGGA AGGGAAGCAT CCTCTTCAAT AGCTTGAAAT CAACTACAGA GTGTTGAGTA TCAAATTACG TATTCTATTT 7136

a ATTCAGAAAT CGCACAGGGG AAGCTCCTGA ATTCATTAGT TGAGTATGTG ATCTCATAGT AAGAAGTGAT CACAAACGGT GATGGATGCC GAGAGTATGT TGATGGGGTG AATCTGTAGA 7256

a ACAATACTAC AGTAGAGATT GGGTAGCCCC CTCAACTATA AAATAAGCAA ATCCTGAAAG GCGAGTAAAA GCACCTACAG TTGTATCTGT CTAAGCTATT ATAAGCTTAG AGAAGGGGGG 7376

a GAAACCCCAT CGGTAGATGC AACTATACAA AATACCAACA CACACAAAAA CAAGGGACAC ACTATTATAT CATGAAAAAC ACTGTCGAGT CATTAGAACA GTTGAATACC ACACTTTCTA 7496

a ATACAGTCAA CGATATACCA GAATCATCTG CAATGGGTGC CTTAAGCCAT TCGGCTAATT CGAGTATTAA TGCAGCTCTT AACAGTACGG AAGAGGTAAA AAGGGGTGAA GATGTCGCGG 7616

a GGGATTCAAT TCATGCAAAT ATTAGTGATA TGTTTCATAT AGGTAACCCA TTCAACCCAT TTCAGTGTTT CATTGATAAC CACATGCGTA AAACCATCTT TGCAAAAAAT CCATATGTGC 7736

a AATACGCTGT CGAAAGTTCT TATTACACAC CTGCGGTTAA TCGTCTGGTT AAAGGGTTGA GTTTGAAAGA TATGGATCC                                                  7815
```

Fig. 5. DNA sequence of the 5-kb repeat. The entire sequence of the β-R2 (b; *Bam*HI fragments 3864 and 2257 bp) and α-R1 (a; *Bam*HI fragments 3680 and 4129 bp) are shown. Only the β-R2 copy of the repeat sequence is shown. All sequence coordinates in the test refer to the β-R2 sequence.

have been identified which have, apart from the terminal A, 186 bp of homology to the termini of the inverted repeats of the S1 and S2 episomes. These are flanked by four unique sequence combinations which have been identified as δ, δ', ψ, ψ' (Isaac *et al.*, 1985b; Schardl *et al.*, 1985). Comparison of this sequence to the 5-kb repeat sequence showed it to be virtually identical to the integrated R1 homologous sequence (Figure 3), apart from an 8-bp duplication, which may have arisen from polymerase slippage during replication. The full extent of the R1 sequence from the S mitochondrial genome was not determined, though it has been estimated to be at least 1 kb and forms a repeat located at both the $\delta-\delta'$ and $\psi-\psi'$ loci, the positions at which S1 and S2 integrate to form the linear mitochondrial DNA molecules (Schardl *et al.*, 1984). This homology to the R1 episome in the S mitochondrial genome supports the suggestion that S1 arose from an illegitimate intermolecular recombination event between R1 and R2 (Levings *et al.*, 1983; Sederoff and Levings, 1985; Elmore-Stamper and Levings, 1986) and provides evidence that the S cytoplasms originally contained episomes related to R1 and R2.

The region homologous to S2 in Wf9-N is located adjacent to the beta copy of the 5-kb repeat as shown in Figure 2. The right hand junction of the 5-kb repeat is formed by the first 187 bases of the S1/S2 terminal inverted repeat except for a single base change, T to G, at base 19 in the S1/S2 homology. The remaining 416 bases extending from the end of the 5-kb repeat to the *Bam*HI site are 97% homologous to the S2 episome sequence (Levings and Sederoff, 1983). In addition to the T → G change at base pair 19, other base alterations include additions/deletions, transitions and transversions. The main result is a series of frame-shift mutations introducing 11 translational termination codons into the transcribed open reading frame I (ORFI) sequence of S2 (Figure 4), making this transcript non-functional.

Because the region homologous to S1 in the Wf9-N cytoplasm represents an integrated R1 episome relic, it would be more correct to assume that the region homologous to S2 represents a

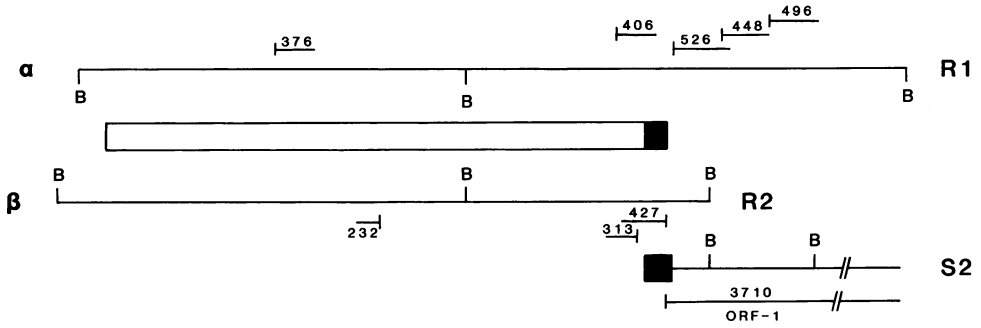

Fig. 6. ORFs associated with the sequenced *Bam*HI (B) fragments of the 5-kb repeat (open box). The homology to the terminal inverted repeat of S2 is indicated (black box) as is the position of the 5′ end of the 3710 nucleotide ORF of S2. The location and lengths in bp of the largest ORFs, 5′ ends are represented by a vertical bar, of α-R1 and β-R2 sequences are shown.

relic sequence of the R2 episome. Therefore this sequence homology will be referred to as R2 hereafter in the text (see Figure 1). Restriction endonuclease studies and heteroduplex studies have demonstrated that neither the integrated R1 nor integrated R2 sequences are complete copies of the free episomes; deletions of their terminal inverted repeats and adjacent unique sequence distal to the right junction of the 5-kb repeat have occurred (Levings *et al.*, 1983; Lonsdale *et al.*, 1983a; McNay *et al.*, 1983).

The RU cytoplasm of South American maize contain the mitochondrial free replicating R1 and R2 mitochondrial episomes and have mtDNA restriction profiles almost indistinguishable from those of the current North American fertile lines (Timothy *et al.*, 1983; Weissinger *et al.*, 1983). It is probable that the South American RU cytoplasms are the direct progenitors of the current North American fertile lines. The R1 and R2 episomes have integrated into the mitochondrial master chromosome and become fixed there by deletion of one of the terminal inverted repeat sequences and the loss of the free replicating episomes. Such a process has been observed in the male-sterile M825 accessions of the S cytoplasm, where reversion to fertility is associated with the loss of the free replicating S1 and S2 episomes and deletion of one of the terminal inverted repeats of the integrated S2 sequence (Schardl *et al.*, 1985).

DNA sequence analysis of the 5-kb repeat

The entire DNA sequence of the four *Bam*HI fragments which encompass both copies of the repeat is shown in Figure 5. The alpha and beta copies of the 5-kb repeat were found to be exact duplicates of 5270 bases. Searches for sequence homology between the two regions of the Wf9-N mitochondrial genome presented herein and the current versions of the N.I.H. GenBank and EMBL data bases were carried out in the hope of shedding some light on any biological function associated with the repeat. The only significant homologies to known genes include the above-mentioned homology to the terminal inverted repeats of R1 and R2 and a stretch of 54 bp (bases 758–811) that are 95% homologous to the 5′-flanking region and putative transcription start site of the maize mitochondrial 26S rRNA gene (Dale *et al.*, 1984).

Computer analysis did reveal several alternating purine/pyrimidine tracts which have the potential to shift easily into a Z-DNA conformation (Rich *et al.*, 1984). Two of these tracts are each 14 bp in length (starting at bases 2590 and 5529, respectively) with one base in each tract deviating from alternating purine/pyrimidine. Four additional tracts of alternating purine/pyrimidine are 20 bp long. One of these (starting at base 2965) has two bases deviating from alternating purine/pyrimidine. The three other stretches (starting at bases 630, 2903 and 4871) have three bases each deviating from the pattern. A single deviation in a 14-base tract does not prevent a shift to a Z-DNA conformation; two or three deviations in a longer tract also may not prevent a Z-DNA shift (Nordheim *et al.*, 1982; Wang *et al.*, 1985). Evidence in other systems suggests that Z-DNA stretches may provide a starting point for recombination with a nearby homologue (Kmiec and Holloman, 1984; Kmiec *et al.*, 1985). If the potential Z-DNA tracts found here are critical for recombination to occur, and if recombination at other repeated sequences in the maize mitochondrial genome involves a similar mechanism, similar tracts of alternating purine/pyrimidine should also appear in other mitochondrial repeats. Confirmation of this hypothesis awaits sequence determination of other repeated sequences including those which do not recombine or recombine only at a low frequency, for example the '10-kb repeat'. It should be noted that one of the Z-DNA tracts (starting at base 5529) lies within the segment of the 5-kb repeat that is homologous to the terminal inverted repeat of the S1/S2 episomes, i.e. the part of the episome that has been proposed to recombine with the maize mitochondrial genome (Schardl *et al.*, 1984).

The perfect identity in the base sequence over the entire length of the 5-kb repeat suggests that direct comparison of the two copies of the repeat and correction of any mutational changes (mismatch repair) that appear in either copy are ongoing processes. The apparent recombination between homologous sequences leading to the formation of subgenomic circles, as well as the complete absence of any sequence differences between the individual copies of the repeat, can be explained by a general recombination mechanism involving the formation of heteroduplex DNA consisting of one strand from each copy of the homologous sequence.

The formation of a *chi* structure in a region of hybrid DNA and subsequent resolution by endonucleolytic cleavage would allow the division of a single large genomic circle into two small circles, each containing one copy of the repeated DNA (Figure 1). Formation of the same structure, followed by branch migration, could create large stretches of heteroduplex DNA permitting correction of any mispairs that existed between two copies of the repeat. For example, the T → G transversion at base pair 19 in the integrated R1/R2 terminal inverted repeat homology is not found in the S mitochondrial DNA sequence (Figure 3), presumably as a direct result of sequence correction. It is likely that within the maize mitochondrial genome other general recombination events, such as those involving the terminal inverted repeats of the S episomes or any of the other repeats found in the mitochondrial genome, proceed by a similar mechanism. This mechanism of recombination between two loci may be the in-

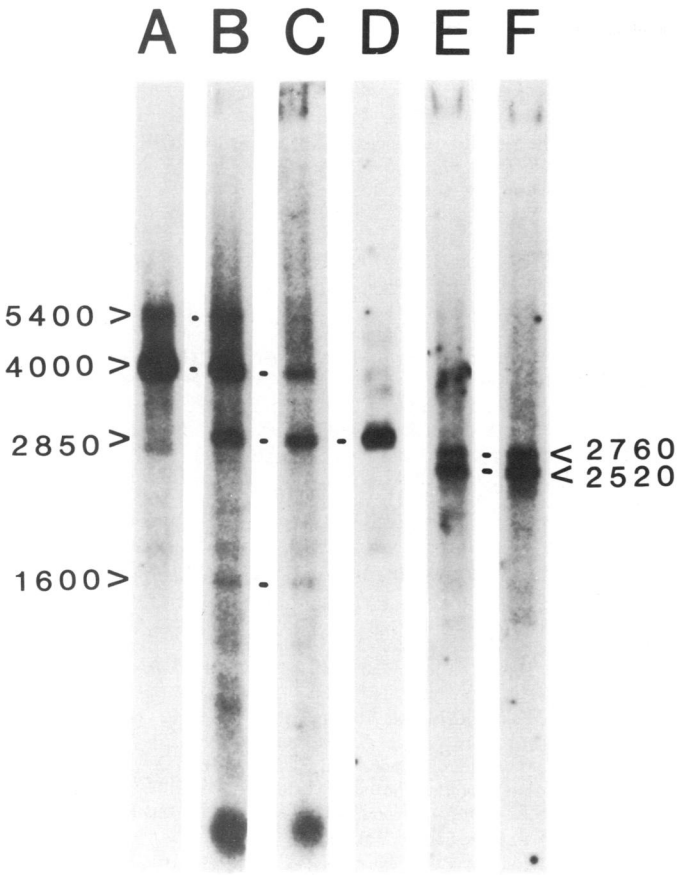

Fig. 7. Transcripts of α-R1 and β-R2. Mitochondrial RNA was probed with **A**, 0.95-kb *Bam*HI fragment; **B**, the 2257-bp *Bam*HI fragment; **C**, the 2173-bp *Pst*I fragment; **D**, 4129-bp *Bam*HI fragment; **E**, 3680-bp *Bam*HI fragment; **F**, 3864-bp *Bam*HI fragment. The positions of these fragments are given in Figure 2.

itial event in the formation of the large repetitive DNA sequence elements. In the present example, perhaps the illegitimate integration of an R1/R2 replicon initiated genomic recombination. Mutation of the adjacent sequences and copy correction have expanded the terminal inverted repeat sequence of the R1/R2 replicon to its present day size of 5270 bp.

Several other statistically significant features were found in the sequence, notably a perfect 18-bp direct repeat starting at positions 3034 and 3175, respectively, and a 4-fold repeat of the pentanucleotide CTCTA starting at position 3385. Their biological significance, if any, is unknown. It should also be noted that the 5-kb repeated sequences have none of the characteristics of transposable elements such as terminal inverted repeats and flanking direct repeats.

Reading frames and transcription

Analysis of the DNA sequence for open ORFs reveals many small ORFs but only few of these are of any significant length, >300 bp. The positions of the longest ORFs are shown in Figure 6. Probing mitochondrial RNA with cloned sequences of the *Bam*HI fragments revealed six discrete transcripts (Figure 7). The 3680- and 3864-bp *Bam*HI fragments hybridised to two transcripts of 2760 and 2520 nucleotides. Neither of these transcripts were detected by probing with the 2173-bp *Pst*I fragment within the 5-kb repeat, thereby placing them within the first 3000 bp of the repeat. At least one of these two transcripts may well originate within the 54-bp homology to the 5′-flanking sequence of the 26S rRNA which contains the putative transcription start site (Dale *et al.*, 1984). The 2257-bp *Bam*HI fragment with homology to the R2 sequence hybridised to four transcripts of 5400, 4000, 2850 and 1600 nucleotides. The 0.95-kb *Bam*HI fragment, which is internal to R2 sequence and internal to the transcribed reading frame ORFI of the S2 episome, hybridizes to both the 5400- and 4000-nucleotide transcripts. Previous transcript studies of the S2 ORFI sequence, in normal fertile, S and RU cytoplasms (Traynor and Levings, 1986) identified a 4100-nucleotide transcript initiating at nucleotide position +32 within the terminal inverted repeat (Figure 4). It can be assumed that this is the 4000-nucleotide transcript detected in these studies. The lack of significant homology of the 2173-bp *Pst*I and the 4129-bp *Bam*HI fragments to both the 5400- and 4000-nucleotide transcripts would suggest that the 5400-nucleotide transcript initiates at a similar position to the 4000-nucleotide transcript. The length of this transcript would span the entire length of the integrated and deleted form of the R2 sequence. It may result from a low level of transcription through the transcription termination signal marking the end of the 4000-nucleotide transcript or, alternatively, the 4000-nucleotide transcript is a 3′ processed form of the 5400-transcript. If the latter, then it may well be that in the RU and S cytoplasms the entire sequence of the R1, S1, R2 and S2 replicons are transcribed. Evidence for such a large transcript has been obtained in earlier studies (see Traynor and Levings, 1986).

It is probable that the sequence within the R2 terminal inverted repeat promoting transcription of the R2 (S2) ORFI, giving the 4000-nucleotide transcript, is also promoting transcription into the integrated R1 homologous sequence. The 2850-nucleotide transcript appears to be a possible candidate for such a transcript being homologous to the 4129-bp *Bam*HI fragment. However, the relatively high homology of the 2850-nucleotide transcript to the 4129-bp *Bam*HI fragment, the 2173-bp *Pst*I fragment and to the 2257-bp fragment suggests that the transcript contains sequences homologous to the 5-kb repeat as well as to the R1 sequence. Therefore the transcript's 5′ terminus appears within the R1 sequence with transcription proceeding towards and into the repeat.

The 1600-nucleotide transcript and the small abundant transcripts (Figure 7, tracks B and C) correlate with probes having sequence homology to the terminal inverted repeat bases 1−186. Other than this their position cannot be assigned. The localisation of the 5′ and 3′ ends of the identified transcripts has not been determined. However, a consideration of their origins was undertaken. Homology to other transcribed mitochondrial and chloroplast gene sequences is a possibility. Hybridisation of the *Bam*HI fragments to cosmid clones spanning the entire mitochondrial genome (Lonsdale *et al.*, 1984a) was performed with the 2257- 3860- and 4129-bp *Bam*HI fragments. They hybridized as predicted except for the 3860-bp *Bam*HI fragment which showed an additional hybridisation to a 3.65-kb *Xho*I fragment. This *Xho*I fragment contains sequence homology to two chloroplast *Hin*dIII fragments of 1269 bp and 740 bp which contain the 3′ end of the 23S rRNA, the 4.5S and 5S RNAs and the tRNAs for arginine, asparagine-2, and glycine-1 (Lonsdale, 1985). However, the 3860-bp *Bam*HI fragment, as with the other fragments of the 5-kb repeat, demonstrated no homology to the cosmid clones spanning the entire chloroplast genome of maize (Brears *et al.*, 1986).

Genes necessary for mitochondrial biogenesis appear to be conserved even between unrelated plant species. The presence of

sequences related to the 5-kb repeat was investigated in other species. Restriction digests of *Brassica campestris* (Chinese cabbage) and *Beta vulgaris* (sugar beet — fertile and male-sterile cytoplasms) mtDNAs were probed. Weak homology to the 5-kb repeat sequences was detected in the *Brassica campestris* mitochondrial genome though not to the fertile or male-sterile sugar beet mitochondrial genomes. The transcripts observed from this region of the maize mitochondrial genome can therefore be postulated to have no essential translational products. The 2750- and 2520-nucleotide transcripts may either result from fortuitous transcription initiation within the sequence homologous to the 5′ region of the 26S rRNA gene or may have been detected because of sequence homology of this region of the repeat to an as yet unidentified transcribed region within the 3.65-kb *Xho*I mitochondrial fragment.

In conclusion, it would appear that the 5-kb repeat of the maize mitochondrial genome serves no vital coding function. The recombinational activity of the repeat cannot be linked to any sequence motifs at the present time, though the results do suggest that the origins of the repeat began with the integration of a R1/R2 replicon. Recombination between the terminal inverted repeats, mutation of the adjacent sequences and copy correction have expanded the repeat to its present day size.

Materials and methods

Four *Bam*HI restriction fragments encompassing both copies of the 5-kb direct repeat, as well as their flanking regions, were obtained from a cosmid library of maize mitochondrial DNA from line Wf9-N (Lonsdale *et al.*, 1981, 1984a, 1984b). Two *Bam*HI fragments of 3680 bp (identified as a 3.5-kb fragment in earlier references (Lonsdale *et al.*, 1981, 1983b; McNay *et al.*, 1983) and of 4129 bp (previously 3.9 kb) were obtained from cosmid clone 2c11 (Lonsdale *et al.*, 1981). These fragments contain the complete alpha copy of the repeat occurring adjacent to the site of the S1 homology (Figure 2). A *Bam*HI fragment of 3864 bp (previously 3.6 kb) was obtained from cosmid 8-3E6 and a fragment of 2257 bp (previously 2.1 kb) was derived from cosmid 2c44. The latter two fragments contain the beta copy of the repeat found next to the S2-homologous DNA (Figure 2). All fragments except the 3864-bp fragment have been cloned into plasmids pBR322 or pUR2 (Ruther, 1980). The 3864-bp fragment from cosmid 8-3E6, as well as the *Bam*HI fragments from the three plasmids, were isolated on low melting point agarose and cloned symmetrically into the *Bam*HI site of M13 mp19 (Norrander *et al.*, 1983). Subclones of each fragment were prepared using the single-stranded sequential deletion method (Dale *et al.*, 1985). Determination of the size of subcloned fragments and preparation of single-stranded template DNA was carried out as described previously. Sequencing reactions were carried out using the dideoxy chain termination method (Sanger *et al.*, 1977).

DNA sequence analysis

Most computer analyses were performed using the Staden, Wisconsin and IntelliGenetics software packages either on the BIONET National Molecular Biology Computer Resource (a DEC 2060), a SUN 2/120 computer or a Vax II/750. Sequence comparisons to the current EMBL and NIH GenBank (release 39) database were performed. In addition recently published mitochondrial DNA sequences were screened for homology to the 5-kb repeat, these included: cytochrome oxidase subunit I (Isaac *et al.*, 1985b), the α subunit of the F_1-ATPase (Isaac *et al.*, 1985a; Braun and Levings, 1985), subunit 6 (Dewey *et al.*, 1985a) and subnut 9 of the F_0-ATPase (Dewey *et al.*, 1985b). Both strands of the repeat were used for all analyses.

Mitochondrial RNA

RNA was isolated from DNase- and RNase-treated mitochondria using the procedure of Kirby (1968). RNA samples (10 μg) were electrophoresed in 1% agarose–6% formaldehyde gels prior to transfer to nitrocellulose membrane.

Acknowledgements

We thank Pioneer Hybrid International for supplying seed. This work was supported by a DeKalb–Pfizer Genetics Corporation grant to C.L.S. Support for Bionet provided by NIH Division of Research Resources (Grant #IV41RR01685-02). Also supported by grants from NIH (GM 32113) and the McKnight Foundation.

References

Braun,C.J. and Levings,C.S., III (1985) *Plant Physiol.*, **79**, 571–577.
Brears,T., Schardl,C.L. and Lonsdale,D.M. (1986) *Plant Mol. Biol.*, **6**, 171–178.
Dale,R.M.K. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 4454–4457.
Dale,R.M.K., Mendu,N., Ginsburg,H. and Kridl,J.C. (1984) *Plasmid*, **11**, 141–150.
Dale,R.M.K., McClure,B.A. and Houchins,J.P. (1985) *Plasmid*, **13**, 31–40.
Dewey,R.E., Levings,C.S., III and Timothy,D.H. (1985a) *Plant Physiol.*, **79**, 914–919.
Dewey,R.E., Schuster,A.M., Levings,C.S., III and Timothy,D.H. (1985b) *Proc. Natl. Acad. Sci. USA*, **82**, 1015–1019.
Duvick,D.N. (1965) *Adv. Genet.*, **13**, 1–56.
Elmore-Stamper,S. and Levings,C.S.,III (1986) *Maize Genet. Coop. Newslett.*, **60**, 110–111.
Kemble,R.J. and Thompson,R.D. (1982) *Nucleic Acids Res.*, **10**, 8181–8190.
Kirby,K.S. (1968) *Methods Enzymol.*, **12B**, 87–99.
Kmiec,E. and Holloman,W.K. (1984) *Cell*, **36**, 593–598.
Kmiec,E., Angelides,K. and Holloman,W. (1985) *Cell*, **40**, 139–145.
Isaac,P.G., Brennicke,A., Dunbar,S.M. and Leaver,C.J. (1985a) *Curr. Genet.*, **10**, 321–328.
Isaac,P.G., Jones,V.P. and Leaver,C.J. (1985b) *EMBO J.*, **4**, 1617–1623.
Levings,C.S.,III and Sederoff,R.R. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 4055–4059.
Levings,C.S.,III, Shah,D.M., Hu,W.W.L., Pring,D.R. and Timothy,D.H. (1979) In Cummings,D.J., Borst,P., Dawid,I.B., Weissmann,S.M. and Fox,C.F., (eds), *Extrachromosomal DNA*, Academic Press, NY, pp. 63–73.
Levings,C.S.,III, Sederoff,R.R., Hu,W.W.L. and Timothy,D.H. (1983) In Cifferri,O. and Dure,L. (eds), *Structure and Function of Plant Genomes*. Plenum Press, NY, pp. 363–371.
Lonsdale,D.M. (1985) In Hohn,B. and Dennis,E.S. (eds), *Plant Gene Research*, Springer-Verlag Wein, NY, Vol. **2**, pp. 51–60.
Lonsdale,D.M., Thompson,R.D. and Hodge,T.P. (1981) *Nucleic Acids Res.*, **9**, 3657–3669.
Lonsdale,D.M., Fauron,C.M.-R., Hodge,T.P., Pring,D.R. and Stern,D.B. (1983a) In Chater,K.F. *et al.*, (eds) *Genetic Rearrangement, The Fifth John Innes Symposium*. Croom Helm, London and Canberra, pp. 183–205.
Lonsdale,D.M., Hodge,T.P., Fauron,C.M.-R. and Flavell,R.B. (1983b) In Goldberg,R.B. (ed.), *UCLA Symposia on Molecular and Cellular Biology*. Alan R.Liss Inc., NY, New Series, Vol. **12**, pp. 445–456.
Lonsdale,D.M., Hodge,T.P. and Fauron,C.M.-R. (1984a) *Nucleic Acids Res.*, **12**, 9249–9261.
Lonsdale,D.M., Hodge,T.P. and Stoehr,P.J. (1984b) *Nucleic Acids Res.*, **12**, 429–436.
McNay,J.W., Pring,D.R. and Lonsdale,D.M. (1983) *Plant Mol. Biol.*, **2**, 177–187.
Nagy,F., Lazar,G., Menczel,L. and Maliga,P. (1983) *Theor. Appl. Genet.*, **66**, 203–207.
Nordheim,A., Lafer,E., Peck,L., Wang,J., Stollar,B.D. and Rich,A. (1982) *Cell*, **31**, 209–218.
Norrander,J., Kempe,T. and Messing,J. (1983) *Gene*, **26**, 101–106.
Palmer,J.D. and Shields,C.R. (1984) *Nature*, **307**, 437–440.
Paillard,M., Sederoff,R.R. and Levings,C.S.,III (1985) *EMBO J.*, **4**, 1125–1128.
Rich,A., Nordheim,A. and Wang,A.H.-J. (1984) *Annu. Rev. Biochem.*, **53**, 791–846.
Ruther,U. (1980) *Mol. Gen. Genet.*, **178**, 475–477.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Schardl,C.L., Lonsdale,D.M., Pring,D.R. and Rose,K.R. (1984) *Nature*, **310**, 292–296.
Schardl,C.L., Pring,D.R. and Lonsdale,D.M. (1985) *Cell*, **43**, 361–368.
Sederoff,R.R. and Levings,C.S.,III (1985) In Hohn,B. and Dennis,E.S. (eds), *Plant Gene Research*, Springer-Verlag Wein, NY, Vol. **2**, pp. 91–109.
Spruill,W.M.,Jr., Levings,C.S.,III and Sederoff,R.R. (1980) *Genet.*, **1**, 363–378.
Thompson,R.D., Kemble,R.J. and Flavell,R.B. (1980) *Nucleic Acids Res.*, **8**, 1999–2008.
Timothy,D.H., Levings,C.S.,III, Hu,W.W.L. and Goodman,H.H. (1983) *Maydica*, **28**, 139–149.
Traynor,P.L. and Levings,C.S.,III (1986) *Plant Mol. Biol.*, in press.
Wang,H.-J., Gessner,R., van der Marel,G., van Boom,J. and Rich,A. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 3611–3615.
Ward,B.L., Anderson,R.S. and Bendich,A.J. (1981) *Cell*, **25**, 793–803.
Weissinger,A.K., Timothy,D.H., Levings,C.S.,III, Hu,W.W.L. and Goodman,M.M. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 1–5.
Weissinger,A.K., Timothy,D.H., Levings,C.S.,III and Goodman,M.M. (1983) *Genetics*, **104**, 365–379.

Received on 30 June 1986; revised on 8 August 1986