Curr Genet (1988) 13:343–349

Current Genetics

Common features of three inversions in wheat chloroplast DNA

Christopher J. Howe[1], Richard F. Barker[2*], Catherine M. Bowman[2], and Tristan A. Dyer[2]

[1] Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QW, UK
[2] Plant Breeding Institute, Maris Lane, Trumpington, Cambridge, CB2 2LQ, UK

Summary. We have determined the DNA sequences of regions involved in two of the three inversions known to have occurred during the evolution of wheat chloroplast DNA. This establishes the extent of the second largest of the three inversions. Examination of these sequences suggests that although short repeated sequences are present, the endpoints of the second and third inversions are not associated with repeated sequences as long as those associated with the first inversion. However the endpoints of all three inversions are all adjacent to at least one tRNA gene, and there is evidence that three of the tRNA genes have been subjected to partial duplication, possibly at the time of inversion. This suggests that tRNA genes might be involved with rearrangements of chloroplast DNA, as has also been postulated for mitochondrial DNA.

Key words: Chloroplast DNA – tRNA genes – Gene duplication – Inversion

Introduction

The chloroplast genomes of most land plants and many aquatic ones are very similar in organisation, being circular molecules composed of two large inverted repeats separated by single copy region of different sizes. Typically the inverted repeated sequences are of the order of 20 kbp each, with single copy regions of 20 kbp and 80 kbp (Palmer 1985). With the exception of a number of legumes which have lost one arm of the inverted repeat, the chloroplast genomes of the other plants can be derived from an ancestral form (probably resembling the tobacco genome) by a combination of two main processes (Palmer and Thompson 1982; Palmer 1985). One is fluctuation in the size of the inverted repeats, which may expand or contract, gaining genes which would otherwise be in a single copy region, or losing them to a single copy region. The mechanism of this is not clear, but may involve incorporation of regions of weak homology at the edges of the repeats (Shen et al. 1982). The other major process of large-scale change is alteration in gene order. In most plants, the order of genes can be derived from the ancestral form by a relatively small number of inversions and other simple recombinational rearrangements (Palmer and Thompson 1982; Palmer 1985). Inversion of the relative orientation of the two single copy regions also occurs. This is a much more rapid process, being frequent within the lifetime of the plant, and is brought about by recombination within the arms of the inverted repeats (Palmer 1983). Intermolecular recombination also occurs, to form multimers of the genome (Kolodner and Tewari 1979).

What particular sequences are involved in these recombination events is not known, although it is striking that similar regions of the genomes of several different species have been involved in independent rearrangements, suggesting that there may be recombinational hotspots around the genome (Kung et al. 1982; Howe 1986). Analysis of the effects of deletions within the inverted repeats on their recombination indicated that this depended simply on homologous recombination or else on specific but widely scattered sites (Palmer et al. 1985), but mapping recombination sites within such large repeats is a priori difficult. Therefore it is easier to identify those recombination sites which have led to rearrangements in the single copy region. Wheat and a

* *Present address:* Sequencing Systems Ltd., Unit 184, Cambridge Science Park, Milton Road, Cambridge CB4 4GN, UK

Offprint requests to: C. J. Howe

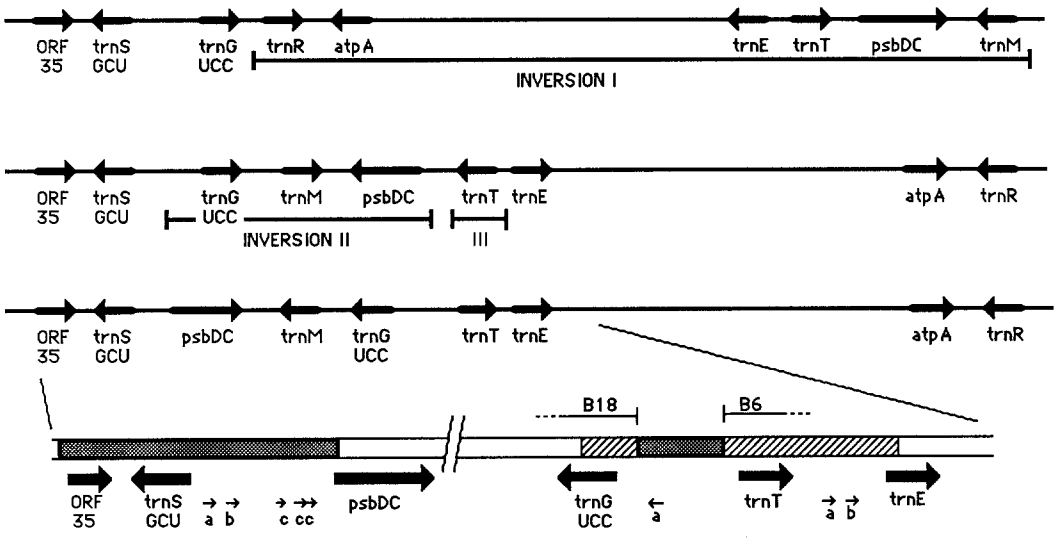

Fig. 1. Diagram (not to scale) showing the evolution of part of wheat chloroplast DNA from an "ancestral" form (*top line*) through an intermediate form (*second line*) to the present form (*third line and expanded in bottom line*). Selected tRNA and other genes are marked (for clarity the *trn*S(UGA), URF62 and trnG(GCC) genes 3′ to *psb*DC, and the genes *trn*Y, *trn*D and *trn*C 3′ to *trn*E have been omitted), with the directions of transcription. Together with the short intergenic repeats "a", "b" and "c". The *hatched area* indicates previously published sequence used in these comparisons, and the *shaded area* sequence reported in this paper. B18 and B6 indicate the extent of the respective Bam HI fragments

number of other monocots provide a good example to study in this context. Mapping and sequencing studies have shown that the genome has been subjected to at least three inversions (Courtice et al. 1985; Howe 1985; Quigley and Weil 1985). These are found in some, but not all, monocots (Oliver and Poulsen 1984; de Heij et al. 1983), and are therefore believed to have occurred recently in evolutionary terms. This suggests that although a number of small insertions and deletions have also occurred near some of the endpoints (Bowman et al. 1983) it may still be possible to identify the squences responsible for these changes. The largest inversion (referred to here as inversion I) covers more than 20 kbp and includes, among others, the cluster of ATP synthase genes which contains the alpha subunit gene *atp*A (Howe et al. 1983). The next largest inversion (inversion II) includes the genes *psb*D and *psb*C, together with a number of tRNA genes and at least one unidentified open reading frame. The precise extent of this inversion is not known, however (Quigley and Weil 1985; Howe 1985). The third inversion (inversion III) includes simply a tRNA Thr (GGU) gene (Quigley and Weil 1985). Se-

A

```
GGATCCTATTCTACGGTTAC GACTACAATAATAATTATTT ACTTTGCAAGGACGTAAGTG TGTCAGACTGCTAAATTGTT 80

TTATTATTCCTACAATATAA ACTAGGGGATATAGGGGGTA ACACTGTCCCTATAAATTCC TTTTTTCCTTTTCTTTTTTG 160

TCCAGAATTGAACAAAAAAG AAATTCTGGAAGATGTTTTC TTCCCCCATTAAGTCCGAGC CATAGAGTAAGAGTGAGATG 240

AGTTTTCCAAATTCATCATA GACTTTCCCTATGGCTTGAT AGAAGTAAGCAGCAAAGAGT CGTAACTTAAAGAAAAAAAC 320
                                                                                  a

GGACAGGGCCGACAGATTTA CCTGATGTAAAGAAGATCCT AAAAGTCTCTGATTAGTCGA CTCTCTCCATTTAACACTTT 400

CCTCTCTCTCCTTTTTTCCA CTGACTCAATTCTAGTTTAT TAGATTCTTGTTTAAAAGAA TGAAAAAAGTTCAATAGAAC 480

TAAGAACAAATAAAAAATAG CATAGAGGATCAAGGCCATT ACCAAAAGTTCCTCCCAAGG ATCC
```

Fig. 2A.

B

```
GGATCCGTGTGGTAAGGAAA AACTGGTAATCTATTCCTTA AAAAAAAATCTTGGAGATTA TGTAATGCTTACTCTCAAAC 80
                                                                M  L  T  L  K

TTTTTGTTTATACAGTAGTG ATATTCTTTGTTTCACTCTT TATCTTTGGATTCTTATCTA ATGACCCAGGACGGAATCCT 160
L  F  V  Y  T  V  V   I  F  F  V  S  L  F   I  F  G  F  L  S   N  D  P  G  R  N  P

GGACGCGAGGAGTAATAATT CCGTTTTTTTTTTTTTTCT TATTGGATTTGTTTCGTACA TTTATCTATGAGAAAATCCG 240
 G  R  E  E  *

GGGGTCAGAATTCTTTCCAA TTCGAAAGTCCCAAATGATC CAAGGGAGCGGAAAGAGAGG GATTCGAACCCTCGGTACAA 320
                                              |<<tRNA Ser

AAAAATTGTACAACGGATTA GCAATCCGCCGCTTTAGTCC ACTCAGCCATCTCTCCACGT TCCAAAGCGAAAGGTTTCCG 400
                                                <<tRNA Ser|

TGATATGATATAGGCAAGAA ATAAGAAATAATGGTTGCAA AAAAACCCCTTTTTTTCTT TCAAAAGTCTATAAAAATTA 480
                                                              a

TATTGCCAATTCCATTTTAG TTATATTCTTTTTTCTTAAT GTTAATAAAAAAAGAAGAAA ATTCTTGTTTTTTCTTTCTA 560
      b

AAAATCGATATTGGCCGAGA GACAATCAAATAGATTTTCT CTTTAGCGGGCATTTCCATA TAGGACTTGTTATAATTATA 640

ATAAAACAAGCAGGTTATAT AAAAAAAACGCTTTTTTTGT CGATTATTTATCAAGAAAGC AAAAAGGGTTCTTATCAAAT 720

CCACCATAAAATTGGAAAGA AGCATAAAGTAAGTAGACCT GACTCCTTTAATGATGCCTC TATCCGCTATTCTGATATAT 800

AAATTCGATGTAGATGAAAT TGTATAAGCGAATTTTTTTT ATTTCCTTAGACTTAGACCG CGCAAGACAAGAATTTTTCG 880

CTATTTACGATTTCATATTC TTGTTACTAGATGTTCTATA GGAATAAGAAGAAATCGCAA CTCCTTTGCGCTACACATAA 960

AATTGGATTTCGAAGGTCCT TTTTTTTTTCAGAATCCTCT ATTTTAGTTCTTCCACCCAT GCAATAGAGAGGGAATGGGA 1040

AAAGAAGGGTTACTTTTATT TTCATTTTTCCCTTAAAAGA TAGACTTTGAAATAGGAGTC TTGGAATAATGCTGAATTCA 1120

AAGGTTTATTTCTTTCTATA GTATAAGAAAGAAAAAAAAT TTATTTCTATTTCTATAGTA TAAGAAAAACAAATAGTATA 1200
                    c                                             c

AGAAAAACAAATCGAATCAA ATTCATGGATTTACCACGAC CTCGGTTGTGACTCCATAGA TAAAAATGGGAAATTTCTCT 1280
 c

CTTCGAGACCATTGAAAAAG GGCATTGAACGAGAAAGAAA TCGTCCACAGATGATAAAAC TATCATATGCCTTGGAAAGT 1360

GATATGAGGTGCTCGGAAAT GGTTGAAGTAATTGAATAGG AGGATCACTATGACTATAGC CCTTGGTAGAATTCCTAAAG 1440
                                              |psbD>
```

Fig. 2A, B. Sequences of (**A**) the BamHI fragment between BamHI fragments B18 and B6 and (**B**) the region upstream from *psb*DC. The occurrence of repeats "a", "b" and "c" is marked, with the translation of ORF35 and the positions of *trn*S(GCU) and *psb*D. A possible ribosome binding site for ORF35 is *underlined*

quence analysis of the endpoints of inversion I had revealed the presence of repeated sequences, recombination across which might have generated the inversion. The repeat included a sequence homologous to the attachment site of bacteriophage lambda. There had also been some duplication of a tRNA gene at one endpoint (Howe 1985). We have carried out DNA sequence analysis of regions of wheat chloroplast DNA which allows the determination of the extent of inversion II, and which together with previously published data

(Quigley and Weil 1985) covers the endpoints of inversions II and III. We have analysed these sequences to search for repeats or other features which may have been involved in the rearrangement.

Materials and methods

The 544 bp *Bam*HI fragment situated between *Bam*HI fragments B18 and B6 (Quigley and Weil 1985) was isolated from a clone containing the *Pst*I fragment P10 inserted into pBR322, and subcloned into pUC12. The *Bam*HI fragment and *Hae*III and *Sal*GI subfragments were isolated from gels, cloned into M13 vectors mp8 and mp9 and sequenced using the chain termination procedure (Biggin et al. 1983). *Sal*GI and *Sal*GI-*Alu*I subfragments of P10 were likewise sequenced to ascertain the orientation of the 544 bp *Bam*HI fragment with regard to the B18 and B6 fragments. The sequence of the region upstream of the *psb*D gene as far as the next *Bam*HI site was determined by the method of Maxam and Gilbert (1980) from a *Sst*I-*Bam*HI fragment digested with the enzymes *Eco*RI, *Cla*I, *Acc*I and *Nde*I. Sequence was determined for both strands, and verified the restriction sites used for sequencing and cloning. Sequences were compared using the programme DIAGON (Staden 1982) on a BBC microcomputer.

Results

Identification of endpoints

The positions of the DNA sequences determined are shown in Fig. 1, along with the position of previously published sequence relevant to this study, and the regions involved in inversions I, II and III. The DNA sequence of the 544 bp region between *Bam*HI fragments 18 and 6 is shown in Fig. 2A and that of the region upstream from *psb*D, together with the first 10 codons of *psb*D (S. Hird, pers. comm.) is shown in Fig. 2B. No genes were identified in the 544 bp *Bam*HI fragment, but the sequence in Fig. 2B contains a sequence (nucleotides 289–376) highly homologous and in a similar position to the tRNA Ser (GCU) gene of tobacco. Its putative secondary structure is shown in Fig. 3. There are five nucleotide differences from the spinach sequence, and they are all located in the variable stem and loop. Two of them represent complementary transition mutations in the stem, retaining the same degree of base pairing, but substituting a GC pair in tobacco for an AT pair in wheat. Mubumbila et al. (1985) have reported a tRNA Ser gene in this region (Mubumbila et al. 1985), and it is highly likely that this sequence is the wheat gene.

The identification of the tRNA Ser (GCU) gene allows a more accurate delineation of the inversion II in wheat, since it was not clear until now whether or not this gene had been involved (Quigley and Weil 1985). Clearly it has not, as it would otherwise be adjacent to the trnT gene (Fig. 1). Rather it is in the "ancestral" position, and one inversion breakpoint must therefore be in the region between *trn*S(GCU) and *psb*DC of wheat. The second breakpoint must be in spacer DNA adjacent to *trn*T, although it is not clear on which side, since *trn*T has itself been inverted (inversion III), and depending on the actual evolutionary order (and extent within spacer) of inversions, the inversion III may have moved the position of the endpoint sequence of inversion II.

Beyond *trn*S(GCU) is an open reading frame, extending from the start of the sequence determined to nucleotide 172. It is possible that, if functional, it starts at nucleotide 65 (although this would require initiation at an AUG codon other than the first after the putative ribosome binding site). This would then generate a 35 residue polypeptide which would be identical to the product of the corresponding region in tobacco (Shinozaki et al. 1986) and 94% homologous to that in *Marchantia,* a liverwort (Ohyama et al. 1986). This high conservation across a wide evolutionary range suggests that the ORF is functional, particularly as the nucleotide sequence is less highly conserved than the amino acid sequence – there are 13 nucleotide sequence differences between the tobacco and wheat sequences, and the flanking sequence is less conserved still. Such a polypeptide would be highly hydrophobic, with a possible membrane spanning segment (residues 6–27) and a C-terminal region able to interact with extrinsic or soluble components.

Sequence comparison of endpoints

The three regions between *trn*S(GCU) and *psb*DC, *trn*G(UCC) and *trn*T, and *trn*T and *trn*E were compared with each other for repeated sequences (using both strands separately) and other structural features. In contrast to the observations for inversion I, no repeats of more than a few base pairs were found flanking inversions II or III. A number of rather short repeats were found, and are marked on Figs. 1 and 2. Repeat "a" (TTTTTTTCTTT) occurs in all the intergenic regions compared. The copy in *trn*G-*trn*T is in inverse orientation. Repeat "b" (TATATTGACAATTCCA) occurs in the *psb*DC-*trn*S and the *trn*T-*trn*E regions as a direct repeat. Within the *trn*S(GCU)-*psb*DC region there are two tandem copies of an 18 bp sequence. (TAGTATAAGAAAAACAAA), repeat "c", and a third partial copy (16/18) separated by 19 bp. The interrupting sequence is similar to part of the repeated sequence found flanking inversion I. Perhaps more strikingly, the *trn*T and *trn*E genes have been involved in duplication events. The *trn*E gene is immediately preceded by a direct duplication of 19 bp (18 of which are identical) and the *trn*T gene is followed by one of 68 bp (53 of which are identical). These repeats are shown in Fig. 4. The barley sequence covering *trn*T also shows the repeat, on inspection, but neither repeat is found in tobacco, spinach,

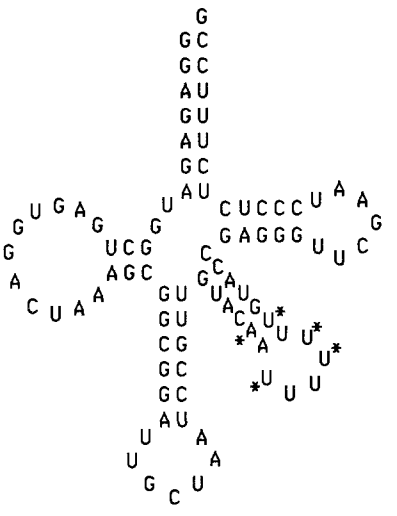

Fig. 3. Putative secondary structure of tRNA Ser (GCU). The -CCA sequence, which is not encoded in the DNA, is omitted. *Asterisks* indicate bases that are different from the tobacco sequence

broad bean or pea, which have not undergone inversion (Oliver and Poulsen 1984; Shinozaki et al. 1986; Holschuh et al. 1984; Kuntz et al. 1984; Rasmussen et al. 1984). Many of the limited differences between the spacer regions of wheat and barley are due to short direct duplications typically of 3–4 bp (or deletion of one of a pair of short direct repeats). This presumably represents the results of "slippage replication". It is not apparent when the wheat and tobacco sequences are compared, probably owing to the much greater degree of divergence between the two. This process has been identified before as a significant cause of sequence alteration in chloroplasts as well as in other systems (Zurawski et al. 1984; Farabaugh et al. 1978; Efstratiadis et al. 1980).

Discussion

The identification and mapping of *trn*S(GCU) and the open reading frame ORF35 allow the extent of inversion II to be determined. Previously, it was not known if these genes had been included in the inverted region. The sequences reported, together with previously published data, therefore cover the endpoints of the three inversions known to have occurred during the evolution of the chloroplast genome of wheat and a number of other monocots. Earlier studies had shown the endpoints of inversion I to be associated with repeated sequences, and it was postulated that a rare homologous recombination event might have given rise to this inversion. The analysis described here, however, did not reveal similar pairs of substantial repeat sequences at the endpoints of inversions II and III. The largest set of such repeats was a direct repeat ("b") of 16 bp, one member of which is in the spacer between *trn*S(GCU) and *psb*DC, and the other in the *trn*T-*trn*E spacer. It is possible that before the inversion of *trn*T the latter repeat was in the *trn*T-*trn*G(UCG) spacer (and thus in the required position and orientation with regard to the other copy to have caused inversion II by recombination), but it seems unlikely that a repeat as short as this would

```
TTAAATAAAAAGCCCTTTTAACTCAGTGGTAGAGTAATGCCATGGTAAGGCATAAGTCATCGGTTCAAATCCGATAAAG
                                                                    trnT

GGCTTTTTAAATTAGTGGTAGAGTAATCTCGTGCTAAGACGTAAGTCGTTGGTTCGAATCTGATAGAGTACTTTTCTACTAA
    ****** * ***************  * ** **** * ****** * ***** **** **** **  **
    TTTTAACTCAGTGGTAGAGTAATGCCATGGTAAGGCATAAGTCATCGGTTCAAATCCGATAAAGGGCT
```

A

```
TGTATATGGCCCTATCGTCTAGTGATGCCCCTATCGTCTAGTGGTTCAGGACATCTCTCTTTCAAGGAGGCAGCGGG
    *** ****************                                    trnE
    ATGCCCCTATCGTCTAGTG
```

```
GATTCGACTTCCCCTGGGGGTAGGGAGTATTATGAAAGGAGGTT
```

B

Fig. 4. Partial duplications of (4A) *trn*T and (4B) *trn*E. The *bold sequence* is the sense DNA sequence (Quigley and Weil 1985). The *underlined sequence* has been repeated on the line below, and matches with the flanking sequence are indicated by *asterisks*. The sequences shown correspond to **A** positions 340–500 and **B** 770–890 of the sequence determined by Quigley and Weil (1985)

be able to act as a substrate for homologous recombination. The 11 bp repeat "a" is the only one occurring in all three intergenic regions, and the probability of this happening by chance is rather less than 1%. However, this repeat is also unlikely to be long enough to act as a substrate for homologous recombination, and neither it nor repeat "b" appears at the endpoints of inversion I, so they probably do not represent widespread sites for site-specific recombination systems. Their significance, if any, is unclear.

There are, however, certain features common to all three inversions. One is the association of the endpoints with tRNA genes. In the case of inversion I, one endpoint is between two adjacent tRNA genes (*trn*G(UCC) and *trn*R), and the other is adjacent to one (*trn*M). Inversion II also has one endpoint between two adjacent tRNA genes (*trn*G(UCC) and *trn*S(GCU)) and the other adjacent to one (*trn*T), while inversion III has one or both endpoints between pairs of tRNA genes (*trn*G(UCC) or *psb*DC and *trn*T, and *trn*T and *trn*E). The same appears to be true for the major inversion found in the chloroplast genome of *Marchantia,* both endpoints of which are adjacent to tRNA genes. In addition, the gene *mbp*X of *Marchantia,* (which has not been found in other chloroplast DNAs) is flanked by tRNA genes (Ohyama et al. 1986). Remarkably, the endpoints of the 23.5 kbp inversion in sunflower chloroplast DNA appear to be exactly the same as the two extreme endpoints of wheat inversions II and III, located between *trn*S and *trn*G, and *trn*E and *trn*T (Heyraud et al. 1987). The association with tRNA genes does not extend to all evolutionary rearrangements, however, since the endpoint of the rearrangement which in pea separates the *pet*A and other genes from *rbc*L is not, in the ancestral form, adjacent to any tRNA gene (Willey et al. 1983; Shinozaki et al. 1986). Analysis of a range of other rearrangements is needed to establish just how widespread is the association with tRNA genes.

The involvement of tRNA genes in these rearrangements is also suggested by the observation that partial duplications of at least three have taken place (including *trn*M in inversion I (Howe 1985)). The extent of these duplications ranges from some 20 bp to nearly 70 bp, but even 20 bp is much greater than the normal size of duplication seen in simple slippage replication (when wheat and barley sequences are compared). Furthermore, while these larger duplications can be found (where sequence is available) in the barley chloroplast genome, which has undergone the inversion (Oliver and Poulsen 1984), they are not found in those genomes that retain the ancestral form, suggesting that the duplication is in some way closely involved with the inversion. Partial duplication of a tRNA gene which has been involved in an evolutionary rearrangement has also been reported in *Vicia faba.* The *trn*L(UAA) gene which has been brought adjacent to the *trn*L(CAA) gene in *Vicia* has suffered a partial duplication including 35 bp of the 5′ exon as well as some intron and flanking sequence (Bonnard et al. 1985). It is interesting to note also the general occurrence in the chloroplast genome of two adjacent highly homologous *psa*A and *psa*B genes (Ohyama et al. 1986; Shinozaki et al. 1986), and to speculate that this duplication might have arisen in a similar way in the early evolution of the chloroplast.

It is difficult to propose a simple model which would account for the involvement of the tRNA genes and their subsequent duplication. They show only rather limited homology to each other, mainly around the -GTTC- sequence found in all tRNA genes. They do not contain sequences closely related to known "recombinogenic" sequences, such as Chi or the 16 bp sequence associated with recombinational hotspots in human minisatellite sequences (Smith et al. 1980; Jeffreys et al. 1985). It is possible that some feature of their possible secondary structure or transcription might be involved. Possible models to account for the duplication might include the generation of a staggered gap, or some replication-linked event. Clearly, though, duplication cannot be associated with all chloroplast recombination events, or the frequent recombination across the inverted repeat would cause a significant increase in size of the genome during the lifetime of a single plant. Perhaps only aberrant recombination events, bringing about rearrangements, also bring about duplications. An association of tRNA genes with genome rearrangements has also been noted for mitochondria. In *Neurospora crassa* mitochondria, recombination across repeated *trn*M genes generates chromosome heteromorphism (Gross et al. 1984). In mitochondria of *Podospora anserina,* tRNAs are associated with sites of excision of senDNA sequences (Turker et al. 1987), and in sea urchins duplication and "remoulding" of tRNA genes is associated with the evolutionary rearrangement of mitochondrial genomes (Cantatore et al. 1987).

It should be borne in mind that the fact that some of these regions of chloroplast DNA have served as hotspots of rearrangement may itself make the sequences involved susceptible to loss or scrambling in insertions and deletions occurring in addition to the inversions. It is therefore still possible that substantial repeated sequences could have caused inversions II and III, and been lost subsequently. It will be necessary to examine as many chloroplast genome rearrangements as possible before the general principles governing the sites and the mechanism become apparent. More rigorous tests will require the transformation of chloroplasts to determine the recombinogenic activity of particular constructs in vivo, or the development of chloroplast extracts able to catalyse recombination in vitro. An understanding of the requirements for recombination will help to clarify

the evolution of the chloroplast genome and perhaps also the transposition of sequences to other organelles and the rearrangements of DNA within other organelles.

Acknowledgmenets. This work was supported by the A.F.R.C. We thank J. C. Gray and S. Hird for helpful discussions and C. J. B. Brookes for advice on probability calculations.

References

Biggin MD, Gibson TJ, Hong GF (1983) Proc Natl Acad Sci USA 80:3963–3965

Bonnard G, Weil JH, Steinmetz A (1985) Curr Genet 9:417–422

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

Cantatore P, Gadaleta MN, Roberti M, Saccone C, Wilson AC (1987) Nature 329:853–855

Courtice GRM, Bowman CM, Dyer TA, Gray JC (1985) Curr Genet 10:329–333

Efstratiadis A, Posakony JW, Maniatis T, Lawn RM, O'Connell C, Spritz RA, DeRiel JK, Forget BG, Weissman SM, Slightom JL, Blechl AE, Smithies O, Baralle FE, Shoulders CC, Proudfoot NJ (1980) Cell 21:653–668

Farabaugh PJ, Schmeissner U, Hofer M, Miller JH (1978) J Mol Biol 126:847–863

Gross SR, Hsieh T, Levine PH (1984) Cell 38:233–239

Heij HT de, Lustig H, Moeskops D-JM, Bovenberg WA, Bisanz C, Groot GSP (1983) Curr Genet 7:1–6

Heyraud F, Serror P, Kuntz M, Steinmetz A, Heizmann P (1987) Plant Mol Biol 9:485–496

Holschuh K, Bottomley W, Whitfeld PR (1984) Plant Mol Biol 3:313–317

Howe CJ (1985) Curr Genet 10:139–145

Howe CJ (1986) In: Mantell SH, Chapman GP, Strett PFS (eds) The chondriome. Longman, London
driome. Longman

Howe CJ, Bowman CM, Dyer TA, Gray JC (1983) Mol Gen Genet 190:51–55

Jeffreys AJ, Wilson V, Thein SL (1985) Nature 314:67–73

Kolodner R, Tewari KK (1979) Proc Natl Acad Sci USA 76: 41–45

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

Kuntz M, Weil JH, Steinmetz A (1984) Nucleic Acids Res 12: 5037–5047

Maxam AM, Gilbert W (1980) Methods Enzymol 65:499–560

Mubumbila M, Bowman CM, Droog F, Dyer TA, Kuntz M, Weil JH (1985) Plant Mol Biol 4:315–320

Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, Sano S, Umesono K, Shiki Y, Takeuchi M, Chang Z, Aota S, Inokuchi H, Ozeki H (1986) Plant Mol Biol Reporter 4:148–175

Oliver RP, Poulsen C (1984) Carlsberg Res Commun 49:647–673

Palmer JD (1983) Nature 301:92–93

Palmer JD (1985) Annu Rev Genet 19:325–354

Palmer JD, Thompson WF (1982) Cell 29:537–550

Palmer JD, Boynton JE, Gillham NW, Harris EH (1985) Evolution and recombination of the large inverted repeats in *Chlamydomonas* chloroplast DNA. In: Steinback KE, Arntzen CJ, Bogorad L (eds) Molecular biology of the photosynthetic apparatus. Cold Spring Harbor

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

Rasmussen OF, Stumman BM, Henningsen KW (1984) Nucleic Acids Res 12:9143–9153

Shen GF, Chen K, Wu M, Kung SD (1982) Mol Gen Genet 187:12–18

Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayshida N, Matsubayasha T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kata A, Tohdoh N, Shimada H, Sugiura M (1986) Plant Mol Biol Reporter 4:110–147

Smith GR, Schultz DW, Crasemann JM (1980) Cell 19:785–793

Staden R (1982) Nucleic Acids Res 10:2951–2961

Turker MS, Domenico JM, Cummings DJ (1987) J Mol Biol 198:171–185

Willey DL, Huttly AK, Phillips AL, Gray JC (1983) Mol Gen Genet 189:85–89

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

Communicated by C. J. Leaver

Received August 21 / December 18, 1987