Curr Genet (1988) 14:127–136

Current Genetics

In wheat ctDNA, segments of ribosomal protein genes are dispersed repeats, probably conserved by nonreciprocal recombination

Catherine M. Bowman, Richard F. Barker*, and Tristan A. Dyer

Institute of Plant Science Research, Cambridge Laboratory, Maris Lane, Trumpington, Cambridge CB2 2LQ, UK

Summary. Some dispersed repeated sequences and their flanking regions from wheat and maize ctDNAs have been characterized. Two sets of wheat ctDNA repeats were found to be the chloroplast ribosomal protein genes *rpl2* and *rpl23*, plus nonfunctional segments of them, designated *rpl2'* and *rpl23'*. Pairwise comparisons were made between the wheat *rpl23* and *rpl23'*, and the maize *rpl23'* sequences. The precise patterns of homology suggest that the divergence of the wheat and maize nonfunctional (*rpl23'*) sequences is being retarded by nonreciprocal recombination, biased by selection for individuals with functional (*rpl23*) sequences. The implied involvement of these sequences in mechanisms of homologous recombination, and therefore in the creation and spread of new ctDNA variants, is discussed.

Key words: Wheat chloroplast DNA – Repeated sequences – Ribosomal protein genes – Evolution

Introduction

The multicopy genomes of higher plant chloroplasts are small (120–160 kbp) circular molecules that contain a large inverted repeat structure (large IR), the two segments of which are separated by a large single-copy region (LSCR) and a small single-copy region (SSCR). Comparison of chloroplast DNAs (ctDNAs) representative of the lower and higher plants, has shown that the evolving chloroplast genome has not accumulated mutations randomly. In general, the large IR has been remarkably conserved and so is a characteristic feature of the molecule (Palmer 1983) while most ctDNA sequence divergence has occurred in the single-copy regions (for review, see Palmer 1985). In one section of the legumes however, chloroplast genome rearrangement has uncharacteristically deleted one segment of the inverted repeat. In most but not all such cases, this loss of the large IR correlates with even more extensive rearrangement of the remainder of the genome. It has therefore been proposed that the presence of the ctDNA large IR may confer evolutionary stability on the molecule (see Palmer et al. 1987).

Higher plant ctDNA evolution is thus characterized by conservation of the large IR and divergence of the single-copy regions. It is also believed that in comparison with nuclear genomes, ctDNA is evolving at a conservative rate, in terms of its structure, gene organisation and primary sequence (see Zurawski and Clegg 1987; Wolfe et al. 1987). Evolution of any genome involves mutation, and the spreading of that mutation through a population or species. The latter can be caused by selection, genetic drift, or by a variety of processes known collectively as mechansisms of "genome turnover" (see Dover and Tautz 1986). These mechanisms include transposition, unequal exchange and gene conversion. While transposition often involves site-specific recombination, unequal exchange and gene conversion are caused by homologous recombination between repeated sequences. Each chloroplast contains multiple copies of its genome, and each copy contains small repeats in addition to the large IR (Bowman and Dyer 1986); Michalowski et al. 1987; Palmer et al. 1987). Therefore, mechanisms of genome turnover involving repeated DNA, may be important in the evolution of chloroplast genomes.

In this paper, two of the longer wheat ctDNA dispersed repeats (previously known as repeats 3 and 9) and their flanking sequences, have been examined. They have been identified as functional and nonfunctional sequences

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

Offprint requests to: C. M. Bowman

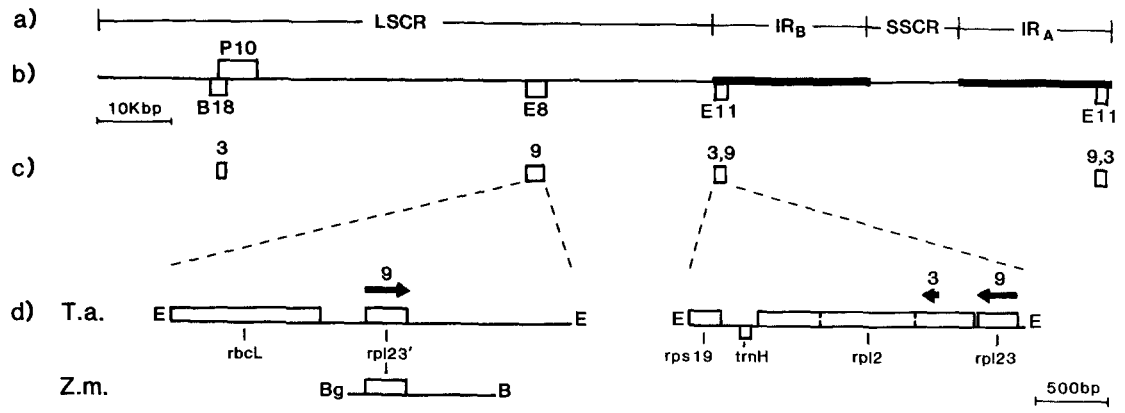

Fig. 1a–d. Location and identity of wheat ctDNA dispersed repeat sets 3 and 9. a Chloroplast genome regions. Large single-copy region (LSCR), small single-copy region (SSCR) and segments of the large inverted repeat (IR_A and IR_B). Although circular, the map is drawn in linear form with the two large IR segments (*thickened lines*) *on the right.* b Positions of restriction enzyme fragments (B18, P10, E8 and E11) that contain repeats 3 and/or 9. c Positions of dispersed repeat copies. d Maps of the wheat (Ta.) ctDNA EcoRI fragments E8 (*upper left*) and E11 (*upper right*) and of a maize (Zm.) ctDNA BglII-BamHI fragment (*lower left*) related to E8. Relevant previously-mapped genes are indicated, and named according to Hallick and Bottomley (1983). In *rpl2, dotted lines* mark the intron boundaries. The *arrows* show the extent and orientation of the repeats. The *rpl23'* pseudogene is identified in this paper as a copy of repeat 9

coding for chloroplast ribosomal proteins. A comparison of the repeats has allowed conclusions to be drawn about the role of homologous recombination in chloroplast genome evolution.

Materials and methods

Recombinant plasmids pTacE11 and pTacE8 contain wheat (*Triticum aestivum*) ctDNA insert fragments, which in turn contain copies of two wheat ctDNA dispersed repeats: repeat 3 and repeat 9 (Bowman and Dyer 1986). Caesium chloride purified plasmid DNA was prepared, and both strands of the insert fragments E11 and E8 were sequenced by the method of Maxam and Gilbert (1980).

Homology to wheat ctDNA repeat 9 had previously been detected in a 4.3 kbp maize (*Zea mays*) ctDNA BamHI fragment, B9 (Bowman and Dyer 1986). This maize ctDNA fragment also contains *rbcL,* the gene for the large subunit of ribulose bisphosphate carboxylase/oxygenase, and has been cloned in a recombinant plasmid pZmcB1B (Gatenby et al. 1981). A clone harbouring pZmcB1B was a gift from Dr. A. A. Gatenby. Plasmid DNA was prepared from it, and homology to repeat 9 was further traced to a 1.0 kbp BglII-BamHI fragment. This fragment was cloned into the BamHI-site of pUC19. Using minipreparations of overlapping subclones, both strands of the fragment were sequenced by the dideoxy chain-termination method (Sanger et al. 1977) modified for double-stranded DNA (Chen and Seeburg 1985; Murphy and Kavanagh 1988).

Results

Identity of repeats

Figure 1 shows the location of the two repeat sets, previously called repeat 3 and repeat 9, on the wheat chloroplast genome. Each set contains 3 copies of the repeat, two in the large IR, and one in the LSCR. The copy of repeat 3 in the LSCR is in the segment overlapped by PstI fragment P10 and BamHI fragment B18 (Fig. 1). The sequence of this wheat ctDNA region has already been obtained (Quigley and Weil 1985; Howe et al. 1988). The LSCR copy of repeat 9 is about 0.2 kbp downstream of *rbcL.* Also shown in Fig. 1 are the wheat and maize ctDNA fragments that were analysed to obtain sequences representing copies of the repeats from the large IR or the LSCR.

Comparison with known gene sequences from tobacco ctDNA (Shinozaki et al. 1986) revealed that repeats 3 and 9, respectively, contained sequences related to chloroplast ribosomal protein coding genes *rpl2* and *rpl23.* The two copies of *rpl2* and *rpl23* in the large IR (fragment E11), are assumed to be functional, because of their intactness and uninterrupted open reading frames (Figs. 2, 3). They map close to the border between the large IR and the LSCR just as they do in tobacco ctDNA, but in wheat there has been a rearrangement at the large IR border (for details, see Barros et al. 1988). The third copy, or segment of both *rpl2* and *rpl23* is assumed for several reasons to be nonfunctional: both segments contain in-frame stop codons and insertions/deletions resulting in frameshifts (Table 1), the nonfunctional segment of *rpl2* is incomplete, and the nonfunctional segment of *rpl23* has been deleted in the wild-wheat lineage that contains *Aegilops sequarrosa* Bowman and Dyer 1986). The nonfunctional segments of these genes, present in the LSCR, have been designated *rpl2'* and *rpl23'*.

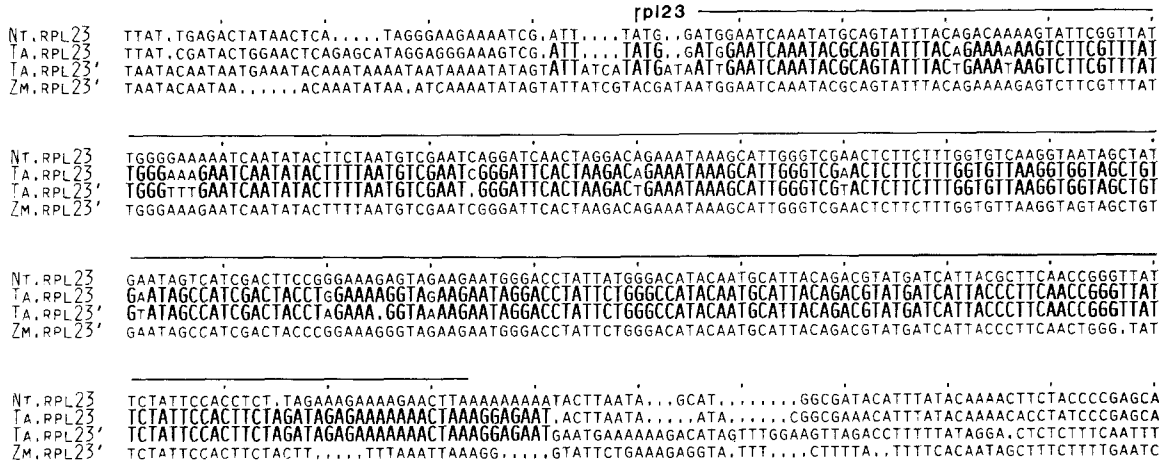

Fig. 2. Alignment of wheat, tobacco and maize *rpl23* homologues. Nucleotide sequences containing the *rpl23* or *rpl23'* homologue from wheat (Ta.), tobacco (Nt.) (Shinozaki et al. 1986) and maize (Zm.) ctDNAs were aligned to maximise homology between all four sequences. Gaps introduced for alignment are represented by *dots*. The *rpl23* gene sequence is *overlined*. The wheat ctDNA repeated sequence is printed in *large characters*

Table 1. Nature of divergence between pairs of homologues. Each pair of sequences was aligned to maximise homology. The type of each base-change: substitution (Sub.), deletion (Del.) or insertion (Ins.), and its effect on the derived amino acid sequence: synonymous change (Syn.), non-synonymous change (Non.) or frame-shift (FS), was assigned relative to the nucleotide sequence and codon positions of homologue (a). Where appropriate the distribution of changes between the 3 codon positions was tested for deviation from randomness using the χ^2 test. In the Ta.rpl23 x Ta.rpl23' comparison, the small number of base changes invalidates the test

Homologues compared	Length of homology (bp)	Extent of divergence (%)	Distribution of base changes in homologue (b)								
			Type			Codon position			Effect on amino acid sequence		
			Sub.	Del.	Ins.	1	2	3	Syn.	Non.	FS
a) Ta.rpl23 x b) Nt.rpl23	281	11.0	29	1	1	6	7	18	14	17	–
a) Ta.rpl23 x b) Ta.rpl23'	290	5.5	12	2	2	5	5	4	4	10	4
a) Ta.rpl23' x b) Zm.rpl23'	281	10.0	25	1	1	10	9	8	6	21	2
a) Ta.rpl2 x b) Ta.rpl2'	105	22.0	20	3	–	9	6	8	2	21	3

Divergence between rpl23 homologues

In Fig. 2, nucleotide sequences homologous with *rpl23* have been aligned for cross-referencing. The functional *rpl23* sequence from wheat has been aligned with the functional *rpl23* sequence from tobacco (Shinozaki et al. 1986), the nonfunctional (*rpl23'*) sequence from wheat, and the nonfunctional (*rpl23'*) sequence from maize. The nature of the divergence between pairs of these sequences is examined in Figs. 3 and 4, and Table 1. The distribution of base-changes between the 3 codon positions, between different regions of the gene, and between gene sequences and flanking regions, reveals some of the different evolutionary mechanisms that are contributing to the divergence of *rpl23* and its homologues.

First, comparing the wheat and tobacco functional *rpl23* sequences, reveals a pattern that is often seen

Fig. 3. Alignment of wheat and tobacco *rpl23* homologues. Nucleotide sequences containing the *rpl23* or *rpl23'* homologues from wheat (Ta.) and tobacco (Nt.) (Shinozaki et al. 1986) ctDNAs were aligned to maximise homology. Gaps introduced for alignment are represented by *dots.* The sequence *Ta.rpl23* is shown *in full,* with the gene *rpl23 overlined,* and the *single-character* translation shown above it. The aligned sequences *Nt.rpl23* and *Ta.rpl23'*, show only those nucleotides that differ from *Ta.rpl23.* The wheat ctDNA repeated sequence is printed in *large characters*

among divergent functional gene sequences (Fig. 3, Table 1) (for review see Kimura 1986). The nucleotide sequences are 11% divergent, but the 31 base-changes are not randomly distributed between the different codon positions ($P < 0.02$). More than half are at the third codon position, and of these, most are synonymous changes. The proportion of synonymous to nonsynonymous changes is about equal, but there is a cluster of nonsynonymous changes in the 3′ region of the gene that contains the last 8 codons.

Divergence between the wheat functional (*rpl23*) and nonfunctional (*rpl23'*) sequences shows a strikingly different pattern (Fig. 3, Table 1). Wheat *rpl23* and *rpl23'* are only 5% divergent, but the 14 base changes are equally distributed between the three codon positions (the number of base changes is too small to test whether this represents a random distribution). Consequently, two-thirds of the changes are nonsynonymous, and the coding sequence contains four frameshifts and four stop-codons. There is also no clustering of divergence at the 3′ end of the sequence, rather, complete homology even extends 8 bp downstream of the stop codon. This pattern implies that in the wheat chloroplast genome, divergence of *rpl23'* has not been constrained by function, and that while accumulating these deleterious mutations *rpl23'* has been evolving as a pseudogene.

Finally, the nonfunctional (*rpl23'*) sequences and their flanking regions from wheat and maize ctDNA have been compared (Fig. 4). The two *rpl23'* sequences themselves are only 10% divergent. Their extreme similarity, particularly at their 5′ ends, implies that they do represent two sequences that are descended from an original duplicative insertion event in the wheat/maize progenitor (see discussion). (The precise extent of the original inserted sequence is not known, since there is some slight divergence at the 5′ end and complete divergence at the 3′ end).

When the homology between the *rpl23'* sequences and flanking regions in wheat and maize is examined (Figs. 4a, b) it is clear that different blocks of sequence are diverging at very different rates. The first block of conserved sequence is the 3′ region of *rbcL* (the only functional gene in the comparison) and this is highly conserved until the last 12 codons, when there is an abrupt breakdown in homology, up to and including the stop-codon. The 87 bp block of sequence immediately downstream of *rbcL* is not conserved, and was therefore probably selectively neutral during the wheat/maize divergence. However, further downstream of *rbcL* and up to 200 bp 5′ of *rpl23'* there is a further conserved block of sequence. This includes the region of dyad symmetry (here designated RDS1) that may terminate or stabilise *rbcL* transcripts (see Stern and Gruissem 1987), plus the sequences that flank that region. Conservation of RDS1 and its 5′ flanking sequence is also seen in most other ctDNAs that have been examined, including tobacco (Shinozaki et al. 1986). These are therefore assumed to be regulatory sequences conserved by selection.

The apparent 5′ endpoint of the inserted *rpl23'* sequence lies within one of the conserved segments (Fig. 4a). The segment is a 16 bp palindromic region of dyad symmetry (designated RDS2), in which there have been compensatory mutations in wheat or maize that preserve the palindrome. The distance between RDS2

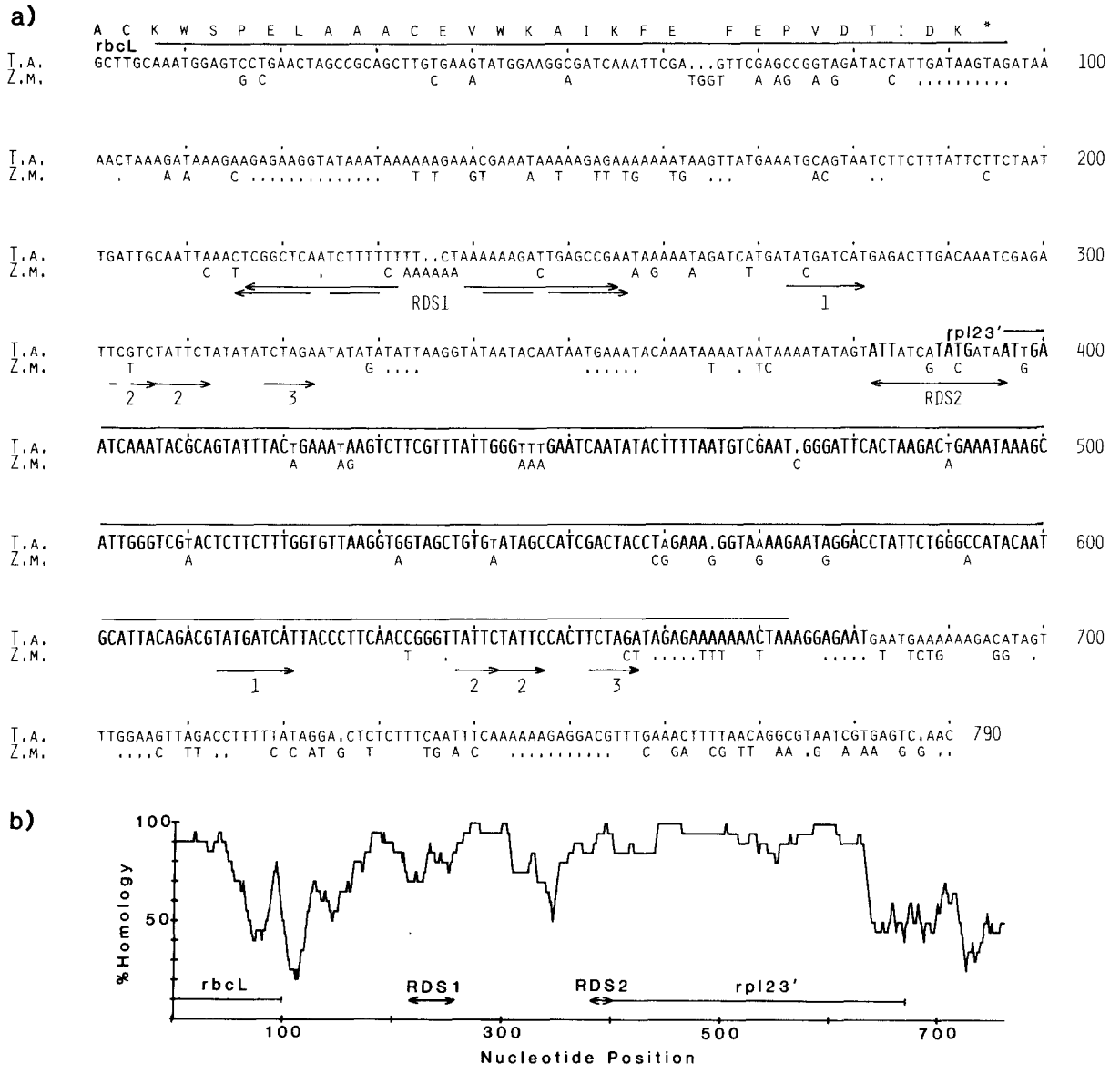

Fig. 4a, b. Alignment and conservation of wheat and maize *rpl23'* homologues and flanking sequences. **a** Nucleotide sequences containing the homologue *rpl23'* plus flanking sequences, from wheat (Ta.) and maize (Zm.) ctDNAs were aligned to maximise homology. Gaps introduced for alignment are represented by *dots*. The wheat sequence is shown in *full*. The aligned maize sequence shows only nucleotides that differ from the wheat sequence. The first 269 nucleotides of the maize sequence are from McIntosh et al. (1980). The 5' region of the gene *rbcL*, and the complete *rpl23'* sequence are *overlined*. The *single-character* translation is shown above *rbcL*. Regions of dyad symmetry (RDS) and short repeats are *underlined* and *numbered*. The wheat ctDNA repeated sequence is printed in *large characters*. **b** Graph to show regions of conservation between the diverging wheat and maize ctDNA sequences aligned in a. Percentage homology was calculated within a moving window of 20 bp. The 5' region of the gene *rbcL*, the homologue *rpl23'* and the two regions of dyad symmetry RDS1 and RDS2 are indicated

and RDS1 is not conserved. In wheat and maize respectively, RDS2 is 124 bp and 113 bp 5' of RDS1. This type of structure may also be conserved in dicots. Tobacco ctDNA contains a region of dyad symmetry, 139 bp 5' of RDS1, but it is not a palindrome, and there is no sequence homology with RDS2. In the remainder of the *rpl23'* sequence, the wheat/maize divergence follows an interesting pattern. The distribution of changes between the three codon positions is random ($P > 0.80$), and the majority of these are nonsynonymous changes (Table 1). Neither of these observations is unexpected if both sequences are diverging pseudogenes. However, in the 5' half of *rpl23'*, there are 10 substitutions and 1 deletion/insertion. All but two of these base changes result from the wheat *rpl23'* sequence diverging from its wheat functional *rpl23* homologue, when the maize *rpl23'* has not (Figs. 2, 4a). In the 3' half of the sequence, both wheat and maize *rpl23'*

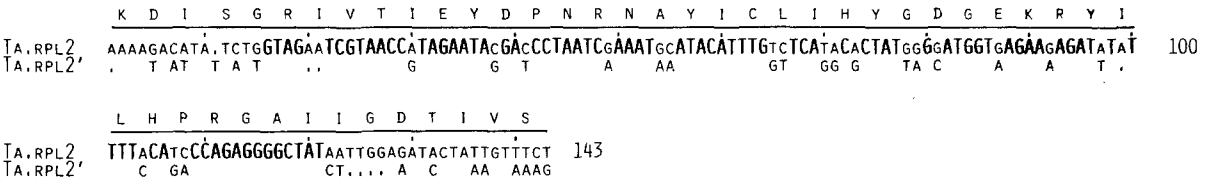

Fig. 5. Alignment of wheat *rpl2* homologues. A region of the wheat (Ta.) ctDNA *rpl2* nucleotide sequence was aligned with previously-published sequence containing the homologous wheat ctDNA segment *rpl2'* (Quigley and Weil 1985; Howe et al. 1988). The gaps introduced to maximise homology between them are represented by *dots*. The region of *rpl2* sequence is shown in *full* and is *overlined*, with the *single-character* translation above it. The repeated segment of *rpl2* is printed in *large characters*. The aligned sequence *Ta.rpl2'* shows only those nucleotides that differ from *Ta.rpl2*

diverge from wheat *rpl23*, and from each other. The wheat/maize *rpl23'* homology breaks down 6 codons away from the 3' end of the sequence, and that complete divergence extends at least 200 bp downstream of the stop-codon (only 110 bp shown). These downstream sequences were therefore probably selectively neutral during the wheat/maize divergence.

During the divergence of wheat and maize *rpl23'*, the pattern of base changes (Table 1) and their impact on the potential translation of the messenger RNA, implies that the wheat and maize *rpl23'* sequences are both diverging as pseudogenes. However, they are diverging from one another much more slowly than other blocks of sequence that surround them (Fig. 4b).

Divergence between rpl2 homologues

Figure 5 shows alignment between part of the gene *rpl2*, and the homologous nonfunctional segment *rpl2'*, that was found in wheat ctDNA fragments B18 and P10 (Bowman and Dyer 1986). The complete nonfunctional *rpl2'* segment is only 105 bp long, and lies 155 bp downstream of the gene *trnG-UCC* (see Fig. 3 of Quigley and Weil 1985). The homologous region in *rpl2* is only 236 bp downstream of *rpl23* (Fig. 1) beginning 221 bp from the start codon and ending 53 bp from the intron. Since duplication, the *rpl2* and *rpl2'* sequences have diverged by 22%, and the pattern of divergence is similar to that seen between *rpl23* and *rpl23'* (Table 1). Relative to *rpl2*, the 23 base changes are randomly distributed between the three codon positions ($P > 0.70$). The changes are almost all nonsynonymous, and the short coding sequence homologue contains 3 frameshifts and 2 stop codons.

Discussion

Identity and origin of repeats

The two sets of wheat ctDNA dispersed repeats examined in this paper, were identified as sequences related to the two chloroplast ribosomal protein coding genes *rpl2* and *rpl23*. The segment *rpl23'*, an almost complete but nonfunctional copy of *rpl23*, could be categorised as an *rpl23* pseudogene. The nonfunctional *rpl2'* segment is probably too short to be categorised as such. This is the first report of protein coding sequences repeated in higher plant ctDNA, but many other higher plant ctDNA repeats have already been characterised. Published examples include repeated simple sequences generated by slippage replication, other tandem repeats (see Zurawski et al. 1984), dispersed repeats that are segments of tRNA coding sequence (Oliver and Poulsen 1984; Bonnard et al. 1985) and also a non-coding sequence, repeated at the ends of a wheat ctDNA inversion, that contains homology with the bacteriophage lambda attachment site (Howe 1985).

The location of the repeats and their identity as segments of coding sequences *rpl2* and *rpl23*, offers no simple explanation for the origin of the nonfunctional segments *rpl2'* and *rpl23'*. Their location may be relevant because they map close together in the large IR, near its border with the LSCR (Fig. 1). Three-quarters of the dispersed repeats so far detected in wheat ctDNA have copies in the large IR (Bowman and Dyer 1986). There is plenty of evidence that the two segments of the large IR recombine and that its borders have also receded and extended considerably during ctDNA divergence (see Palmer 1985). Therefore, duplicative insertion of small segments of the large IR in the single-copy regions of the genome could be a rare outcome of its observed recombinogenic behaviour. On the other hand, that behaviour could simply mean that once inserted elsewhere in the genome, sequences from the large IR are more likely to be conserved. There are also several ways that transcribed sequences may be directly involved in DNA duplication and recombination. The transcripts themselves may mediate transposition (for review, see Weiner et al. 1986). Also, it has recently been shown in yeast, that homologous recombination is stimulated many-fold between sequences showing enhanced transcription (Voelkel-Meiman et al. 1987). Although this specific yeast example involves rDNA regulatory sequences directing and enhancing transcrip-

tion by RNA polI, the authors believe that transcriptional activity may be a general feature controlling the frequency and distribution of recombination in eukaryotic cells.

Mechanisms of sequence conservation

While it is possible only to speculate on the mechanism by which the nonfunctional sequences *rpl2'* and *rpl23'* arose, the antiquity of one of the events may be approximated. The extreme similarity between the *rpl23'* segments in wheat and maize ctDNA (Fig. 4) suggests that this copy was generated by a duplicative insertion at or before the divergence of wheat and maize. Furthermore, hybridisation experiments have apparently revealed the same repeat in sorghum ctDNA (Dang and Pring 1986).

If this initial assumption is made, it means that the functional (*rpl23*) and nonfunctional (*rpl23'*) homologues in wheat ctDNA have been diverging from one another for at least as long as the nonfunctional (*rpl23'*) sequences and their flanking regions in wheat and maize have been diverging. Without the maize data, the 95% homology between the functional and nonfunctional *rpl23* homologues in wheat would imply that the duplication of *rpl23'* was relatively recent. However, the sequence data presented in this paper is consistent with the hypothesis that divergence of the nonfunctional *rpl23'* sequences has in fact been retarded by nonreciprocal homologous recombination, biased by selection.

Analysis of base changes between diverging functional *rpl23* sequences in wheat and tobacco, and diverging functional (*rpl23*) and nonfunctional (*rpl23'*) sequences in wheat (Figs. 2, 3 and Table 1), shows that they have been under different evolutionary constraints. Base changes tolerated by the wheat and tobacco functional sequences must, almost certainly, have been selectively neutral (Table 1), while many of the changes accumulated in wheat *rpl23'* would make it impossible to translate the functional mRNA. Therefore, while accumulating these deleterious changes, wheat *rpl23'* has been diverging as a pseudogene.

The wheat *rpl23'* pseudogene sequence has been diverging from its functional homologue for at least as long as it has been diverging from its pseudogene homologue in maize ctDNA. Comparison with maize *rpl23'* and flanking sequences shows that these two *rpl23'* pseudogene sequences are diverging from one another much more slowly than other, surrounding sequences (Fig. 4b). If the sequence conservation between *rpl23'* and *rpl23* in wheat ctDNA is not due to shortage of time, nor, as *rpl23'* is a pseudogene, to selection, then divergence of *rpl23'* has been retarded. One mechanism that can retard divergence between homologues is that of nonreciprocal recombination, also known as gene conversion (for review, see Dover and Tautz 1986).

If gene conversion has been responsible for the high degree of *rpl23'* sequence conservation, then irrespective of precise mechanism (see Szostak et al. 1983), it has operated by using the functional copies of *rpl23* for the nonreciprocal transfer of sequence to the nonfunctional *rpl23'* sequence in the same genome. Several details in the *rpl23* sequence analysis support this. One of the most compelling is that complete homology between the *rpl23* and *rpl23'* copies in wheat ctDNA extends 8 bp beyond the stop-codon, to include the sequence –AGGAGAAT–. Since this sequence was shown to have diverged in the wheat/maize comparison (Fig. 4a), it is unlikely to have been conserved by selection in wheat. This particular pattern of homology is however typical of sequences that have been conserved by gene conversion. According to current understanding (for review see Dover and Tautz 1986), when gene conversion is involved, the length of the conserved sequence domain depends on the signals that initiate and terminate conversion. The nature of these signals is unknown, but they exist irrespective of the boundaries of the gene as a unit of function (see Dover and Tautz 1986). Since the 5' endpoint of the wheat ctDNA *rpl23/rpl23'* repeat homology is so distinct (Fig. 2) a strong signal may exist that initiates or terminates conversion between the two GAAT sequences.

The differing patterns of sequence divergence between the different pairs of wheat and maize *rpl23* and *rpl23'* homologues are also consistent with the presence of conversion domains. In the wheat *rpl23'* sequence, divergence from the wheat functional homologue is greater toward the 5' end, while the 3' end scarcely diverges (Fig. 3). In contrast, the 5' half of the maize *rpl23'* sequence has diverged from the wheat functional homologue by only two base changes (Fig. 2, positions 87 and 192). These observations could be explained if the conversion domains in the wheat and maize chloroplast genomes have been different. If, in wheat, the 5' region of *rpl23'* has been converted less often than the 3' region, while in maize, the 5' region of *rpl23'* has been converted frequently enough almost to prevent divergence.

Divergence is however considerable between the wheat and maize *rpl23'* sequences in the 3' region (Fig. 4a). This divergence is also seen in the wheat/tobacco *rpl23* comparison (Fig. 3). This is still consistent with gene conversion of maize *rpl23'*, because the homologue converting maize *rpl23'* would be the maize functional homologue, *rpl23.* Therefore, divergence of maize *rpl23'* from the wheat *rpl23* homologues could easily reflect divergence of the functional maize and wheat homologues. The sequence of maize *rpl23* is at present unknown. Thus, the pattern of *rpl23'* conservation in

wheat and maize can be explained if it is attributed to gene conversion, biased by selection not for *rpl23'* itself, but for the wheat and maize functional homologues.

Without additional data from divergent species, the timing of the duplication of the nonfunctional *rpl2'* segment cannot be established. While the nature of the base changes (Fig. 5, Table 1) and the shortness of the sequence make it extremely unlikely that the divergence of *rpl2'* from its functional homologue has been constrained by selection, the extent of that divergence may simply reflect elapsed time. It is also possible of course that the duplication was indeed an ancient event, and that *rpl2'* divergence has been retarded in the same way as that of *rpl23'*.

In most analyses of evolving DNA sequences, there is more than one plausible explanation for observed changes. In the case of *rpl2'* and *rpl23'*, one can never ignore the possibility of conservation due to selection for an unknown and perhaps newly-acquired function. However, the length of the *rpl2* and *rpl23* repeats, and their relative degree of homology (Table 1), is consistent with their involvement in gene conversion. It has been shown that in eukaryotes, only 200 bp of homology is required for efficient gene conversion, and the rate of conversion is proportional to the length of homology (Liskay et al. 1987). In prokaryotes, less than 20 bp is required (Albertini et al. 1982). Further, biased gene conversion could also explain the maintenance of other duplicated gene segments in ctDNA, such as the fairly common segments of tRNA genes (e.g., Oliver and Poulsen 1984).

Evolutionary implications

The nature of the evolutionary conservation of the *rpl23'* pseudogene sequences implies that homologues of *rpl23'*, in wheat and maize are separately undergoing cycles of gene conversion that are frequent enough to retard divergence severely. The occurrence of nonreciprocal homologous recombination between small ctDNA repeats has important implications concerning ctDNA evolution.

Each chloroplast contains a multicopy genome that itself contains large and small repeated sequences. In other genetic systems, the evolutionary importance of recombination between homologues is now recognised (Dover 1982). The ways in which homologous recombination can influence the creation and spread of new variants has been most extensively studied in both genic and nongenic families of nuclear DNA repeats (Maeda and Smithies 1986; Dover and Tautz 1986). By analogy, a stochastic mechanism of recombination operating between ctDNA repeats, has a similar potential to influence the evolution of the chloroplast genome. The consequences of homologous recombination will depend on its frequency, whether it is reciprocal, whether it occurs within or between molecules, and on the orientation of the homologues. For detailed discussion, see Dover and Tautz (1986), Smithies and Powers (1986), Maeda and Smithies (1986) and Baltimore (1981). In summary, when homologous recombination is reciprocal (crossing over), it can generate new variants by deletion and inversion of DNA segments within a molecule, or by deletion and insertion of segments between molecules (unequal crossing over). Such intermolecular crossing over can also influence the spread of new variants among molecules. When homologous recombination is nonreciprocal (gene conversion), it can generate new variants by mixing segments of near-homologous sequence from the same or from different molecules. Intramolecular gene conversion between homologues can spread a new variant among those homologues, while intermolecular gene conversion can similarly influence the spread of new variants between genomes.

When it comes to chloroplast genome evolution, the importance of homologous recombination in the creation and spread of new variants will clearly be governed by the mechanism(s) by which homologues in ctDNA recombine. Two important properties to consider are the stochasticity, and the frequency of (intramolecular and intermolecular) homologous recombination in ctDNA.

The stochasticity of ctDNA homologous recombination is important because if recombination between *rpl23* homologues were stochastic in outcome (i.e., each event is unpredictable and could be reciprocal or nonreciprocal) then conversion of *rpl23* would imply that crossing over also occurs. If the mechanism were not stochastic, *rpl23* conversion could imply only the conversion of other repeats.

How confidently may conversion of *rpl23* homologues in wheat ctDNA be interpreted as a nonreciprocal recombination product of a stochastic mechanism? The functional *rpl23* genes are part of the large IR, and most of the available clues come from the recombinational behaviour of the large IR itself. The sequence conservation of the large IR is attributed to "copy correction", which is synonymous with nonreciprocal homologous recombination (see Palmer 1985 for review). The "head-to-head" ctDNA dimers detected by electron microscopy were proposed by Kolodner and Tewari (1979) to be products of crossing over between large IR segments of two ctDNA monomers. In higher plants, large IR inversion is frequent enough to be detected in a single individual (Palmer 1983), and in *Chlamydomonas* ctDNA, it cannot be abolished by deleting any part of the large IR. Inversion must therefore be due either to a site-specific mechanism with multiple sites (see Palmer 1985) or, to intramolecular reciprocal homologous recombination anywhere within the inverted repeat. Multiple sites of intermolecular recombination involving the ctDNA large IR have also been demonstrated in interspecific crosses of *Chlamydomonas* (Lemieux and Lee 1987). Therefore, although it is not

the only interpretation, all the above examples of large IR recombination can be reconciled as reciprocal or non-reciprocal outcomes of a stochastic mechanism of homologous recombination.

If for the purposes of dicussion one accepts this view, then the contribution of intramolecular homologous recombination to ctDNA evolution, could be analogous to its contribution to nuclear DNA evolution. The diverse evolutionary implications can be illustrated by considering the possible outcomes of hybrid DNA formation between the *rpl23* homologues in wheat ctDNA. If the two functional homologues (in the large IR) recombine, all outcomes are viable. Nonreciprocal and/or reciprocal recombination would respectively cause conversion of *rpl23* and/or inversion of the large IR. If the *rpl23'* pseudogene recombines with either functional homologue, conversion of the pseudogene by the functional homologue (as described in this paper) is the least harmful outcome. A lethal variant could be created by conversion of either functional homologue by the pseudogene. Reciprocal recombination between the pseudogene and either functional sequence would also create lethal variants: deletion of the DNA segment between directly repeated homologues would remove vital genes, including *rbcL*, while inversion of the DNA between inverted *rpl23* homologues would interrupt presumed cotranscription of *rpl23* and *rpl2* (e.g., Tanaka et al. 1986), and could also destabilise the molecule, by transposing inverted repeat sequences into the single copy region. (Lethal variants can be supported at low frequency in heterogeneous ctDNA populations, for example in *Chlamydomonas* (Spreitzer and Chastain 1987) and apparently in rice (Moon et al. 1987). They are also maintained as sectors in variegated plants.)

If *rpl23* homologues are typical repeats, then the possible outcomes of intramolecular homologous recombination between ctDNA repeats fit the observed pattern of ctDNA evolution. Recombination between homologues repeated within the large IR, will normally perpetuate the large IR. Recombination involving a homologue in a single-copy region of the genome can influence the divergence of the homologues, and also create a new variant.

Homologous recombination may also contribute to another characteristic feature of chloroplast genome evolution; its slow rate of primary sequence divergence (e.g., Wolfe et al. 1987). However, its contribution will depend on the frequency of recombination between the multiple genomes in a chloroplast. When a new genome variant arises, there is a high probability that it will be converted to wild-type, by the high-frequency wild-type genomes. Therefore, intermolecular gene conversion would tend to accelerate the elimination of variant genomes or alleles, and reduce the overall rate of primary sequence divergence. However, in changing the frequency of alleles in diverging plastid lineages, gene conversion would be interacting with another stochastic process, random genetic drift. Random drift is so important in this role (see Birky 1983 and Gillham et al. 1985 for review), that in order to make a significant contribution to the spread of variants between chloroplast genomes, intermolecular conversion would have to occur at high frequency.

Information on the actual frequency of intermolecular ctDNA recombination in higher plants is scarce because in natural crosses parental chloroplasts do not fuse. However, given the opportunity, it can occur. Extensive intermolecular ctDNA recombination was seen in a rare regenerant tobacco plant, recovered by selection following protoplast fusion (Medgyesy et al. 1985). There are certainly times during normal plant development when opportunities for intermolecular ctDNA recombination might exist. It has been shown that *recA* protein aligns homologous DNA duplexes most rapidly in vitro by diffusion through coaggregates of DNA (see Kowalczykowski 1987). Such an environment may be created in proplastids and young chloroplasts during phases of rapid DNA replication, when many hundreds of DNA molecules are generated per plastid (e.g., Boffey and Leech 1982). Even in mature chloroplasts the DNA is packaged into nucleoids, each containing multiple genomes (e.g., Sellden and Leech 1981). Some idea of the natural frequency might be deduced from electron microscopy of self-renatured ctDNA. It can be calculated that about 1.0% of the lettuce and spinach ctDNA molecules observed by Kolodner and Tewari (1979) were thought to be "head-to-head" dimers, and therefore products of intermolecular ctDNA recombination.

In summary, conservation of *rpl23'* pseudogene sequences in wheat and maize ctDNA implies that *rpl23* homologues are undergoing nonreciprocal recombination. That is, gene conversion between *rpl23* homologues could be an evolutionarily stable outcome of a stochastic recombination mechanism. Such a mechanism operating between ctDNA repeats would have diverse evolutionary potential, many details of which fit the observed pattern of ctDNA evolution.

Acknowledgements. We thank Professor R. B. Flavell for much helpful discussion, and M. G. Jarvis for assistance in sequencing.

References

Albertini A, Hofer M, Calos MP, Miller JH (1982) Cell 24:319–328

Baltimore D (1981) Cell 24:592–594

Barros MDC, Barker RF, Dyer TA (1988) Plant Mol Biol (in press)

Birkey CW Jr (1983) Science 222:468–475

Boffey SA, Leech RM (1982) Plant Physiol 69:1387–1391

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

Bowman CM, Dyer TA (1986) Curr Genet 10:931–941
Chen EJ, Seeburg PH (1985) DNA 4:165–170
Dang LH, Pring DR (1986) Plant Mol Biol 6:119–123
Dover G (1982) Nature 299:111–117
Dover GA, Tautz D (1986) Philos Trans R Soc Lond [Biol] 312:275–289
Gatenby AA, Castleton JA, Saul MW (1981) Nature 291:117–121
Gillham NW, Boynton JE, Harris EH (1985) In: Cavalier-Smith T (ed) The evolution of genome size. Wiley, New York, pp 299–251
Hallick RB, Bottomley W (1983) Plant Mol Biol Rep 1(4):38–43
Howe CJ (1985) Curr Genet 10:139–145
Howe CJ, Barker RF, Bowman CM, Dyer TA (1988) Curr Genet 13:343–349
Kimura M (1986) Philos Trans R Soc Lond Ser B 312:343–354
Kolodner R, Tewari KK (1979) Proc Natl Acad Sci USA 76: 41–45
Kowalczykowski SC (1987) Trend Biochem Sci 12:141–145
Lemieux C, Lee RW (1987) Proc Natl Acad Sci USA 84:4166–4170
Liskay RM, Letsou A, Stachelek JL (1987) Genetics 115:161–167
Maeda N, Smithies O (1986) Annu Rev Genet 20:81–108
Maxam AM, Gilbert W (1980) Methods Enzymol 65:499–560
McIntosh L, Poulsen C, Bogorad L (1980) Nature 288:556–560
Medgyesy P, Fejes E, Maliga P (1985) Proc Natl Acad Sci USA 82:6960–6964
Michalowski C, Breunig KD, Bohnert HJ (1987) Curr Genet 11:265–274
Moon E, Kao T-H, Wu R (1987) Nucleic Acids Res 15:611–630
Murphy G, Kavanagh A (1988) Nucleic Acids Res (in press)
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, Osorio B, Aldrich J, Thompson WF (1987) Curr Genet 11:275–286
Quigley F, Weil JH (1985) Curr Genet 9:495–503
Sanger F, Nicklen S, Coulsen AR (1977) Proc Natl Acad Sci USA 74:5463–5467
Sellden G, Leech RM (1981) Plant Physiol 68:731–734
Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi 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, Kato A, Tohdoh N, Shimada H, Sugiura M (1986) EMBO J 5:2043–2049
Smithies O, Powers PA (1986) Philos Trans R Soc Lond [Biol] 312:291–302
Spreitzer RJ, Chastain CJ (1987) Curr Genet 11:611–616
Stern DB, Gruissem W (1987) Cell 51:1145–1157
Szostak JW, Orr-Weaver TL, Rothstein R, Stahl FW (1983) Cell 33:25–35
Tanaka M, Wakasugi T, Sugita M, Shinozaki K (1986) Proc Natl Acad Sci USA 83:6030–6034
Voelkel-Meiman K, Keil RL, Roeder GS (1987) Cell 48:1071–1079
Weiner AM, Deininger PL, Efstradiatis A (1986) Annu Rev Biochem 55:631–661
Wolfe KH, Li W-H, Sharp PM (1987) Proc Natl Acad Sci USA 84:9054–9058
Zurawski G, Clegg MT (1987) Annu Rev Plant Physiol 38:391–418
Zurawski G, Clegg MT, Brown AHD (1984) Genetics 106:735–749

Communicated by C. J. Leaver

Received January 25, 1988 / Revised March 25, 1988