Current Genetics (1985) 9:495–503

Current Genetics

Organization and sequence of five tRNA genes and of an unidentified reading frame in the wheat chloroplast genome: evidence for gene rearrangements during the evolution of chloroplast genomes

F. Quigley* and J. H. Weil

Institut de Biologie Moléculaire et Cellulaire du CNRS, Université Louis Pasteur, 15 Rue Descartes, F-67084 Strasbourg, France

Summary. The genes for the initiator $tRNA^{Met}_{CAU}$, $tRNA^{Gly}_{UCC}$, $tRNA^{Thr}_{GGU}$, $tRNA^{Glu}_{UUC}$ and $tRNA^{Tyr}_{GUA}$ and an open reading frame of 62 codons have been identified by sequencing a 2,358 bp BamHI and a 1,378 bp BamHI-Sst2 DNA fragments from wheat chloroplasts. A comparison of the organization of these five tRNA genes and of the open reading frame on the wheat, tobacco and spinach chloroplast genomes suggests that at least three genomic inversions must have occurred during the evolution of the wheat chloroplast genome from a spinach-like ancestor genome. Furthermore, it seems that in wheat the 91 bp intergenic region between the genes for the initiator $tRNA^{Met}$ and the gene for $tRNA^{Gly}_{UCC}$ is one end-point of the 20 kbp genomic inversion proposed by Palmer and Thompson in the case of maize (Palmer and Thompson 1982). A 119 bp duplication is located at this junction: the first copy comprises the 91 bp of the intergenic region and the first 28 bp of the $tRNA^{Met}$ gene, the second copy is found downstream of the $tRNA^{Met}$ gene.

Key words: Chloroplast tRNA genes – URF – Chloroplast genome evolution

Introduction

In the higher plants studied so far, the chloroplast tRNA genes are dispersed throughout the chloroplast genome and are helpful markers to study the evolution of the chloroplast genome. By mapping the tRNA genes on the physical map of the chloroplast genome of various plants (Bohnert et al. 1982) and by heterologous DNA-DNA hybridization studies, it has become clear that chloroplast genomes have been rearranged during evolution (Palmer and Thompson 1982; Howe et al. 1983). The chloroplast DNA of maize, in particular, has been shown to have a 20 kbp inversion as compared to the chloroplast DNAs of spinach and petunia (Palmer and Thompson 1982).

In wheat (*Triticum aestivum*), cross-hybridization studies of chloroplast DNA fragments containing part of the ATP synthase α subunit gene (*atp*A) has confirmed the occurence of a 20 kbp inversion within the large single copy of the genome, as compared to spinach (Howe et al. 1983). Our previous studies on the localization and sequencing of the wheat chloroplast genes for $tRNA^{Asp}$ (*trn*D) and $tRNA^{Cys}$ (*trn*C), have provided evidence for such a genomic rearrangement (Quigley et al. 1985).

We have now sequenced, in the same region of the wheat chloroplast genome, an open reading frame of 62 codons (URF-62) and the genes for five tRNAs, namely initiator $tRNA^{Met}_{CAU}$ (*trn*M-CAU), $tRNA^{Gly}_{UCC}$ (*trn*G-UCC), $tRNA^{Thr}_{GGU}$ (*trn*T-GGU), $tRNA^{Glu}_{UUC}$ (*trn*E-UUC) and $tRNA^{Tyr}_{GUA}$ (*trn*Y-GUA). In this report we describe the organization of these genes on the wheat chloroplast genome and suggest that, starting from a spinach-like ancestral chloroplast genome, at least three genomic inversions are necessary to explain the present organization of these genes in the wheat chloroplast genome.

Material and methods

Recombinant plasmids pTA29 and pTA90 containing respectively the 2,358 bp and 6,500 bp BamHI fragments B18 and B6

* *Present address:* Université de Grenoble I, Laboratoire de Physiologie Cellulaire Végétale, B.P. 68, F-38402 Saint-Martin-d'Héres, France

Offprint requests to: F. Quigley

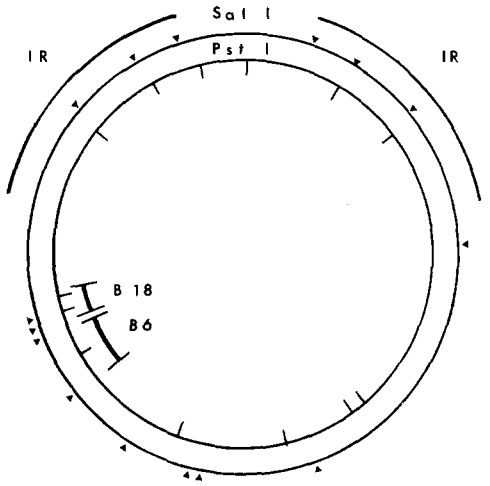

Fig. 1. Localization of two BamHI fragments B18 and B6 on the physical map of wheat chloroplast DNA (Bowman et al. 1981). *IR* indicates the two inverted repeat regions

from wheat chloroplast DNA were constructed as described previously (Quigley et al. 1985). Plasmid DNA was purified from bacterial lysates by ethidium bromide-CsCl density gradient centrifugation (Clewell and Helinski 1969).

Nucleotide sequencing. The nucleotide sequences of B18 and of a 1,378 bp BamHI-Sst2 subfragment of B6 were determined by the dideoxy chain termination method (Smith 1980), after subcloning DNA fragments into suitable restriction sites of the replicative forms of M13 derivatives mp8, mp9, mp10 and mp11.

Results

Location of the sequenced tRNA genes

BamHI fragment B18 and B6 are located on the wheat chloroplast genome as shown in Fig. 1. They are separated by about 450 bp. The two fragments had previously been shown to hybridize to wheat chloroplast tRNAs and, upon sequencing of a 398 bp SalI-PstI subfragment of B18, it was established (Quigley et al. 1985) that this subfragment contains the gene for $tRNA^{Gly}_{GCC}$ (*trn*G-GCC). The genes for $tRNA^{Asp}_{GUC}$ (*trn*D-GUC) and $tRNA^{Cys}_{GCA}$ (*trn*C-GCA) have been located, respectively, on a 742 bp AvaI-PvuI subfragment and on a 420 bp EcoRI subfragment of fragment B6 (Quigley et al. 1985).

We have now completed the sequence of the 2,358 bp BamHI fragment B18 and of the 1,378 bp BamHI-Sst2 subfragment of B6 which is located upstream of *trn*D. The sequencing strategies are detailed in Fig. 2.

Examination of the sequence of fragment B18 (Fig. 3) shows that it contains an open reading frame of 62 codons (URF-62) (positions 223 to 411) located 277 bp upstream on the same strand as the gene coding for $tRNA^{Gly}_{GCC}$ (*trn*G-GCC) which has been previously described (Quigley et al. 1985).

On the opposite strand, fragment B18 contains two additional tRNA genes: the gene for the initiator $tRNA^{Met}_{CAU}$ (*trn*M-CAU) between positions 1,208 and 1,281 and the gene for the $tRNA^{Gly}_{UCC}$ (*trn*G-UCC) between positions 1,373 and 2,120 (this gene contains a 676 bp intron).

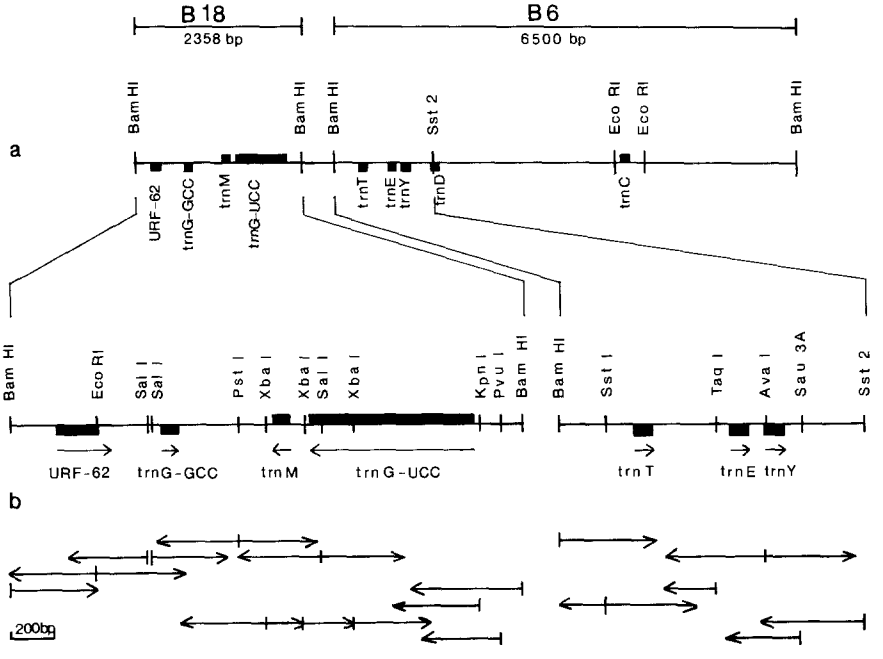

Fig. 2. Position of wheat chloroplast URF-62, initiator *trn*M, *trn*G-UCC, *trn*T, *trn*E and *trn*Y on BamHI fragments B18 and B6. The position of three previously sequenced genes *trn*G-GCC, *trn*D and *trn*C (Quigley et al. 1985) is also indicated. a Partial restriction endonuclease map of B18 and B6. b Sequencing strategy for the regions containing URF-62, initiator *trn*M, *trn*G-UCC, *trn*T, *trn*E and *trn*Y. *Thick arrows* indicate the fragments which have been sequenced. The genes are shown by *black boxes* and *thin arrows* indicate the direction of transcription

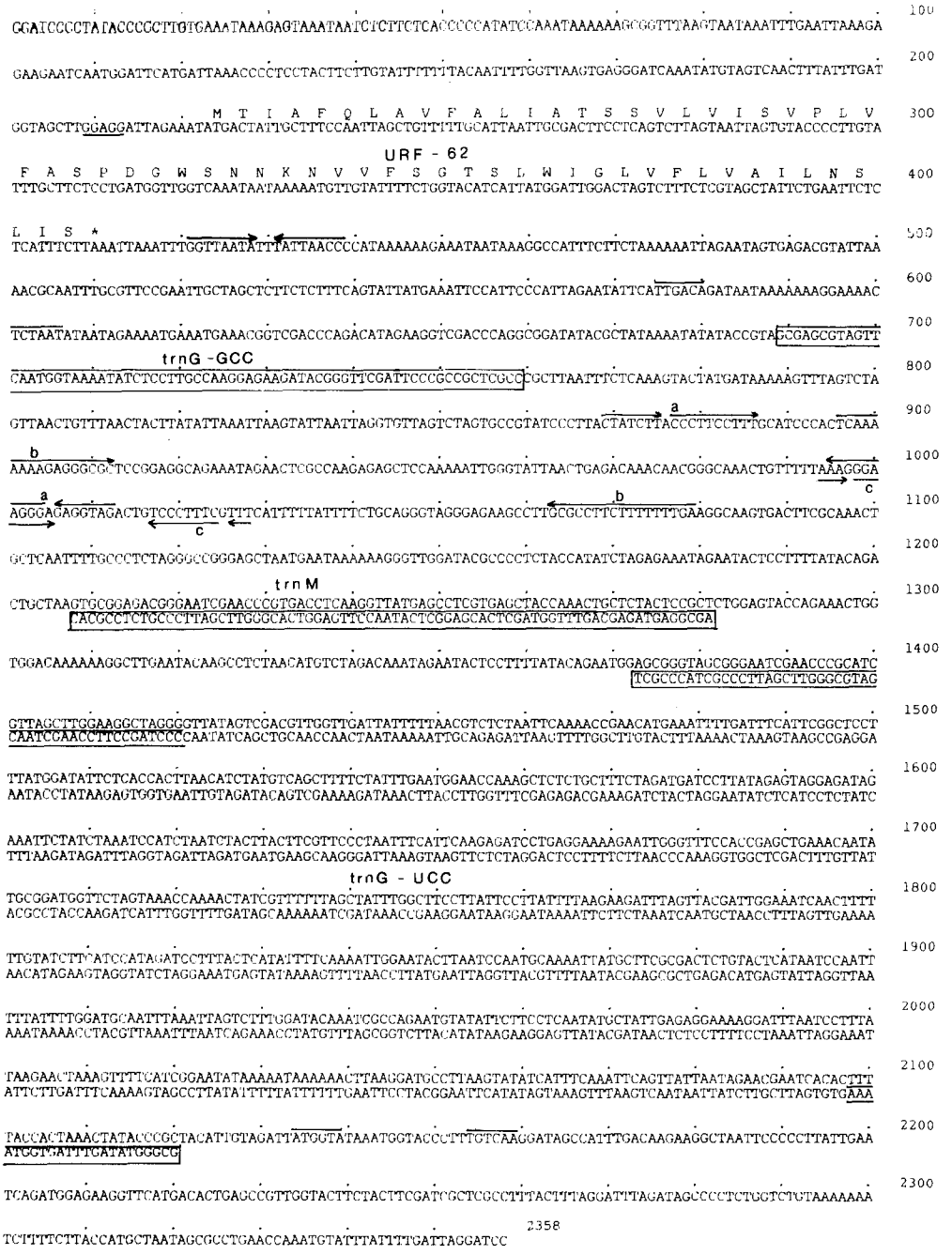

Fig. 3. Nucleotide sequence of fragment B18 (2,358 bp long) from wheat chloroplast DNA. The sequence is written in the 5′ to 3′ direction and the tRNA-like strand is shown. For *trn*M and *trn*G-UCC, the complementary strand is also shown as these two tRNAs are transcribed from the complementary strand. The aminoacid sequence deduced for URF-62 is shown in single letter code. The tRNA-like sequences are boxed. The putative promoter-like sequences ("−10" and "−35") found upstream of the tRNA genes are overlined. A potential ribosome binding site upstream of URF-62 is underlined. Sequences of dyad symetry are shown by *inverted arrows*

Three tRNA genes were identified upon sequencing the B6 subfragment·BamHI-Sst2 (Fig. 4): the genes for $tRNA^{Thr}_{GGU}$ (*trn*T-GGU) between positions 514 and 585, for $tRNA^{Glu}_{UUC}$ (*trn*E-UUC) between positions 958 and 1,031, and for $tRNA^{Tyr}_{GUA}$ (*trn*Y-GUA) between positions 1,093 and 1,174. These three tRNA genes are located on the same strand as *trn*D which is found 366 bp downstream of *trn*Y (Figs. 2 and 4).

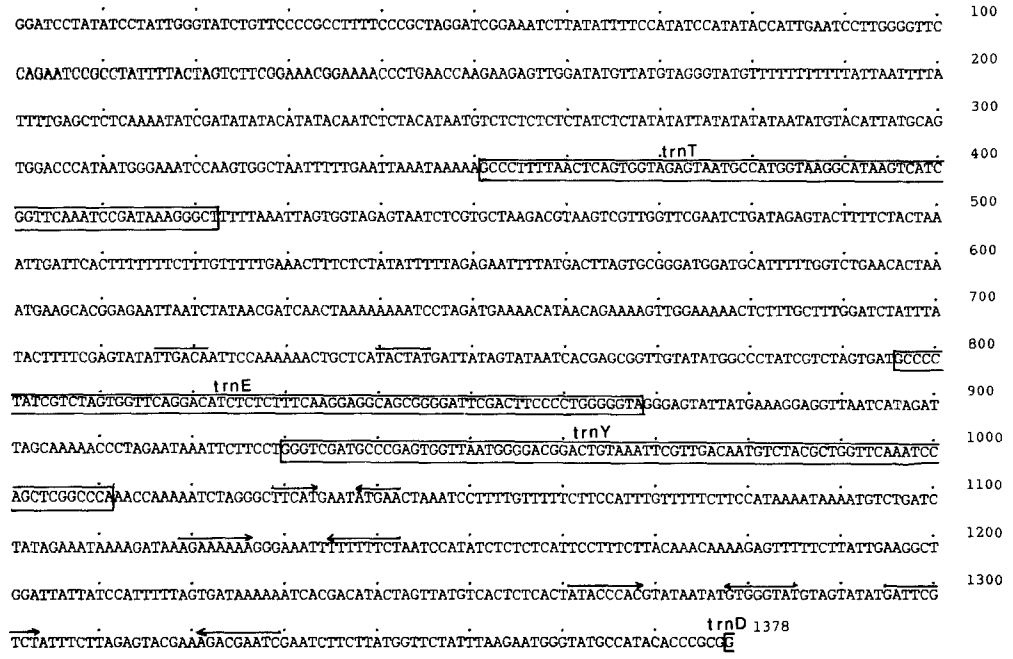

Fig. 4. Nucleotide sequence of 1,378 bp BamHI-Sst2 B6 subfragment of wheat chloroplast DNA. The sequence is written in the 5′ to 3′ direction and only the tRNA-like strand is shown. The tRNA genes are boxed. Overlines show potential promoter-like sequences ("−10" and "−35"). Sequences of dyad symetry are shown by *inverted arrows*

Table 1. Percentage of homology between wheat chloroplast tRNAs and chloroplast tRNAs from other higher plants, as deduced from the gene sequences (this report) and tRNA or tRNA gene sequences (previous studies)

	$tRNA^{Met}$	$tRNA^{Gly}_{UCC}$	$tRNA^{Thr}$	$tRNA^{Glu}$	$tRNA^{Tyr}$
tobacco[a,b]	89.2	100 (exons)			
spinach[c,d]	89.2		94.5	94.5	95.3
bean[c,e]	86.5				
broad bean[f]			95.8	94.5	95.3
pea[g]				94.5	95.3

[a] Ohme et al. 1984
[b] Deno and Sugiura 1984
[c] Sprinzl and Gauss 1984
[d] Holschuh et al. 1984a
[e] Canaday et al. 1980
[f] Kuntz et al. 1984
[g] Rasmussen et al. 1984

Structural features of the sequenced tRNA genes

These five newly localized tRNA genes (two on fragment B18, namely *trn*M-CAU and *trn*G-UCC, and three on the B6 subfragment, namely *trn*T-GGU, *trn*E-UUC and *trn*Y-GUA) were identified by their anticodon and by comparison with the corresponding chloroplast tRNA or tRNA gene sequences from other species.

The homology between the five tRNA gene sequences described here and the tRNA or tRNA gene sequences from other higher plant chloroplasts ranges from 86.5 to 100% (Table 1). Such a high degree of homology is usually found when the sequences of higher plant chloroplast isoaccepting tRNAs are compared.

The 5′ terminal nucleotide of the wheat initiator $tRNA^{Met}$ is not base-paired to the 3′ side. This feature is common to procaryotic and plant chloroplast initiator $tRNAs^{Met}$ where a 5′ terminal C is located opposite to an A (Sprinzl and Gauss 1984). However, in the case of wheat chloroplast initiator $tRNA^{Met}$ there is an A.A mismatch (instead of a C.A mismatch). Another mismatch (C.C) has been found in the same position, in the initiator $tRNA^{Met}$ of the blue-green alga *Anacystis nidulans* (Ecarot-Charrier and Cedegren 1976).

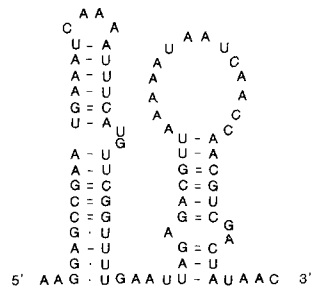

Fig. 5. Secondary structure model of the 3′-terminal region of $tRNA^{Gly}_{UCC}$ intron

The intron of trnG-UCC

The fact that *trn*G-UCC, found on B18 between positions 1,373 and 2,120, is unusually long, is explained by the presence in this gene of a 676 bp intron located in the D stem. A long intron (691 bp) is also present in *trn*G-UCC of tobacco chloroplasts (Deno and Sugiura 1984). The tRNA coding regions (exons) of both wheat and tobacco genes are completely identical (Table 1), whereas the wheat intron shows only a 69% sequence homology with the tobacco intron. Like most of the introns found in chloroplast tRNA genes (for instance maize and tobacco *trn*I and *trn*A, tobacco *trn*V), wheat and tobacco *trn*G-UCC introns show sequence and structural homologies with the 2nd family of introns as defined by Michel and Dujon (1983). One particular feature of this intron family is the conservation of the nucleotide sequences at the 5′ and 3′ ends. In wheat and tobacco *trn*G-UCC introns, the 5′ end reads (in the deduced RNA sequence): GUGUG (consensus GUGCG). The last 87 nucleotides at the 3′ end of the intron are 93% homologous (when wheat and tobacco *trn*G-UCC introns are compared) and this sequence can be folded into the two loops characteristic of class II introns (Fig. 5). These data confirm the location of the intron between nucleotides 24 and 25 in the D stem of $tRNA^{Gly}_{UCC}$, as already suggested by the comparison with the *E. coli* and *B. subtilis* $tRNA^{Gly}_{UCC}$ sequences (Deno and Sugiura 1984). Usually the introns contained in chloroplast tRNA genes are located in the anticodon loop. This is the second example of an intron located in the D stem (the first case was *trn*G-UCC of tobacco chloroplasts, reported by Deno and Sugiura 1984).

Promoter-like sequences upstream of trnG-UCC and trnE-UUC

When looking for the "−35" and "−10" putative consensus promoter sequences TTGaNa and TataaT (Crouse et al. 1984) upstream of the five tRNA genes localized ans sequenced in this work, we find that only two of these genes, namely *trn*G-UCC and *trn*E, are preceeded by typical promoter sequences. A "−35" sequence (TTGACA) and a "−10" sequence (TACCAT) are found, respectively, 34 bp and 14 bp upstream of *trn*G-UCC. This 39 bp DNA region shows a 70% sequence homology with the corresponding region upstream of tobacco chloroplast *trn*G-UCC (Deno and Sugiura 1984). A "−35" (TTGACA) and a "−10" (TACTAT) sequences are found, respectively, 75 bp and 51 bp upstream of *trn*E. The alignement of the coding and flanking regions of *trn*E and *trn*Y from wheat (this work), spinach (Holschuh et al. 1984a), broad bean (Kuntz et al. 1984) and pea (Rasmussen et al. 1984) chloroplast DNAs shows that a strong sequence homology between these four chloroplast DNAs is found not only for the *trn*E and *trn*Y coding regions, but also for the intergenic region

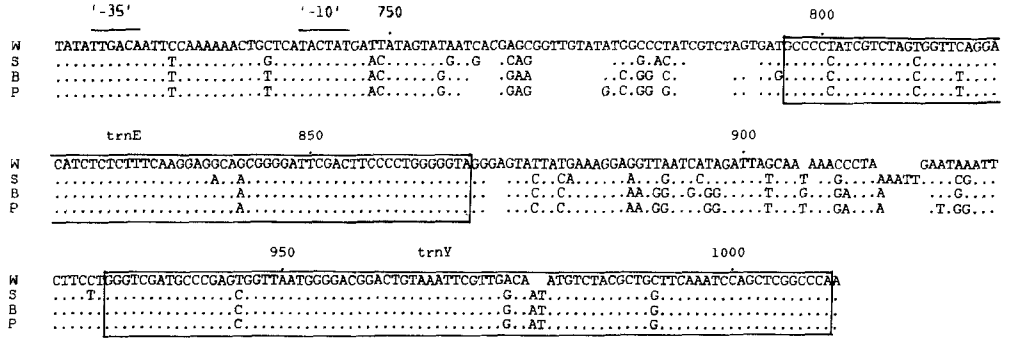

Fig. 6. Comparison of the nucleotide sequences of the wheat (*W*), spinach (*S*) (Holschuh et al. 1984a), broad bean (*B*) (Kuntz et al. 1984) and pea (*P*) (Rasmussen et al. 1984) chloroplast DNA fragments containing *trn*E and *trn*Y. The tRNA-like strands are shown. Numbering refers to the wheat sequence and is the same as in Fig. 4. *Dots* indicate nucleotides which are identical to the wheat sequence (and are therefore common to the four sequences). *Gaps* (shown as *blanck spaces*) have been introduced to maximize homology. The "−35" and "−10" promoter sequences upstream of *trn*E are overlined, while the tRNA genes themselves (*trn*E and *trn*Y) are boxed

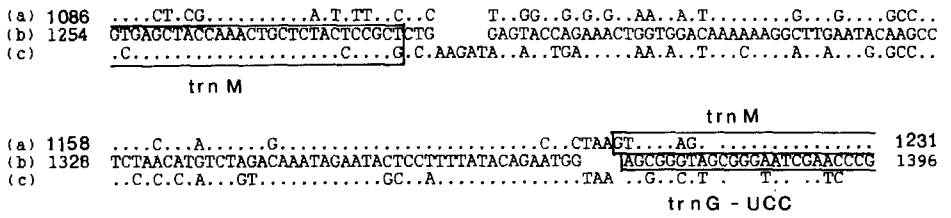

Fig. 7. Sequence comparison between the region downstream of wheat initiator *trn*M (*a*), the region upstream of wheat initiator *trn*M (which is the intergenic region between initiator *trn*M and *trn*G-UCC) (*b*), and the region upstream of tobacco initiator *trn*M (*c*) (Ohme et al. 1984). The numbering refers to the wheat sequence (*a* and *b*) and is the same as in Fig. 3. The sequences are written from the 5′ to 3′ direction, in analogy to Fig. 3 and only the coding strands are shown. *Dots* indicate nucleotide which are identical in the three regions

between these two tRNA genes and for the region upstream of *trn*E which contains the "–35" and "–10" sequences (Fig. 6).

The "–35" and "–10" sequences upstream of *trn*E are identical in the four chloroplast DNAs and they are in each case separated by 18 bp. The homology in the promoter region extends from 4 bp upstream of the "–35" TTGACA sequence (the four sequences are quite different further upstream) to 14 bp downstream of the "–10" TACTAT sequence (Fig. 6), suggesting that longer regions of DNA may be involved in the interaction with RNA polymerase (and not only the two hexanucleotide stretches known as "–35" and "–10" sequences).

Although two putative "–35" and "–10" promoter sequences TTAATC (889–894) and TAGAAT (913–918) are found in the 62 bp intergenic region between wheat *trn*E and *trn*Y, at approximatively the same position as in spinach (Holschuh et al. 1984a), no such sequence is found in front of broad bean or pea *trn*Y (Kuntz et al. 1984; Rasmussen et al. 1984), and it is likely that wheat chloroplast *trn*E and *trn*Y are co-transcribed, using the classical promoter sequences located upstream of *trn*E.

The flanking regions of trnM

Looking for homologies between the potential promoter regions upstream of the genes coding for the initiator $tRNAs^{Met}$ from wheat (this work) and tobacco (Ohme et al. 1984), we noticed that both the regions located upstream and downstream of wheat *trn*M could be aligned with the region upstream of tobacco *trn*M (Fig. 7). The first 28 nucleotides of wheat and tobacco $tRNA^{Met}$ coding sequences plus 91 nucleotides (in wheat) and 96 nucleotides (in tobacco) can be aligned with the sequence downstream of wheat *trn*M. The intergenic region between *trn*M and *trn*G-UCC shows 68% homology when wheat (b) and tobacco (c) sequences are compared. The 122 bp stretch downstream of wheat *trn*M (a) has 73% homology with the region upstream of tobacco *trn*M (c) as shown in Fig. 7. The region (shown in a) which has a high degree of homology with the first 28 nucleotides of *trn*M (b) can therefore be considered as a pseudogene which includes the D stem and loop of *trn*M. The homology between the regions upstream and downstream of wheat *trn*M extends into the terminal part of *trn*G and includes its Tψ loop. There is also a high degree of homology between *trn*M and *trn*G-UCC last 25 nucleotides (only 5 nucleotides differ). In fact, tRNAs accepting different amino acids often show a high degree of homology and it has been suggested that this might be due to a common origin of these tRNAs (Steinmetz et al. 1983).

These data suggest that a duplication involving part of the $tRNA^{Met}$ gene and approximately 90 nucleotides directly upstream of this gene has been duplicated, possibly during the rearrangements which have occurred in this part of the wheat chloroplast genome (see Discussion).

Despites the high homology displayed by the regions upstream of wheat and tobacco *trn*M, no typical "–35" or "–10" promoter sequences could be identified, in contrast to what has been observed upstream of wheat *trn*E. Wheat *trn*M is probably co-transcribed with *trn*G-UCC. The transcription of these two genes probably ends about 130 bp downstream of *trn*M where the presence of several inverted repeats (Fig. 3) may confer to this region a high potential for secondary structure. This region containing several inverted repeats may also be a signal for the end of transcription of *trn*G-GCC which is located about 110 bp upstream of this region, on the strand opposite to that coding for *trn*M and *trn*G-UCC (Fig. 3).

The open reading frame

The open reading frame of 62 codons (URF-62) found between positions 223 and 412 of fragment B18 is located on the physical map of wheat chloroplast DNA approximately 22 kbp upstream of *atp*A, the gene cod-

```
                    20                  40                  60
                    .                   .                   .
wheat     MTIAFQLAVFALIATSSVLVISVPLVFASPDGWSNNKNVVFSGTSLWIGLVFLVAILNSLIS
maize     .N..........................................................
spinach   ................I.L....V.........S...I........L......G.......
```

Fig. 8. Comparison of the aminoacid sequences of the URF-62 (deduced from their nucleotide sequences) from wheat, maize (Krebbers 1983) and spinach (Holschuh 1984a). Sequence identity is marked by *dots*

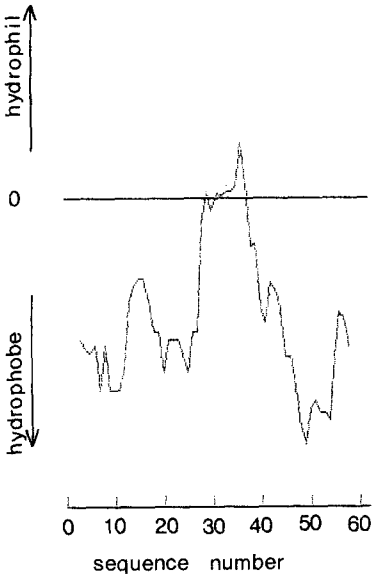

Fig. 9. Hydrophylicity plot (according to Hopp and Woods 1981) of wheat URF-62 using a six aminoacid window. The aminoacid numbering is as in Fig. 8

ing for the α subunit of ATP synthase (Howe et al. 1983). Likewise, an open reading frame of 62 codons has been identified on the maize chloroplast genome, about 25 kbp upstream of *atp*A (Krebbers 1983). The maize URF-62 is located 270 bp upstream, on the opposite strand, of a tRNA gene, *trn*S-UGA (Fig. 10). A comparison of the wheat and maize URF-62 nucleotide sequences shows that there are 98.8% homologous. Upon looking at the DNA sequence upstream of *trn*-UGA in spinach chloroplast DNA (Holschuh et al. 1984b), we found an open reading frame of 62 codons, which has 87.6% homology with wheat URF-62 and is located 362 bp upstream of *trn*S-UGA, but on the opposite strand. The three URFs-62 found in wheat, maize and spinach chloroplast DNA are preceeded by a potential ribosome binding site GGAGG, found in all three plants 9 bp upstream of the sart codon ATG.

The fact that these three URFs-62 have a similar location on the chloroplast genome, the presence of a ribosome binding site 9 bp upstream of the start codon, and the high sequence homology found for these three URFs, suggest that this URF is coding for a functional polypeptide in the three plants.

The amino acid sequences deduced from the nucleotide sequences of the wheat, maize and spinach URFs are compared in Fig. 8. The wheat and maize sequences are identical except for one aminoacid (the second residue is a threonine in wheat, and an asparagine in maize), while the wheat and spinach sequences differ by 7 amino acids.

The small protein coded by the wheat URF-62 would have a molecular weight of 6.5 kd, would be highly hydrophobic (Fig. 9), and is likely to be an intrinsic membrane protein. The two regions on each side of the small hydrophylic cluster would be large enough (20 aminoacids or more) to span accross the thylakoïd membrane.

A high homology is found between the wheat and maize DNA sequences in the 160 bp region upstream of URF-62 and in the 200 bp region downstream of URF-62 (the maize chloroplast DNA was not sequenced further downstream; between wheat and spinach some homology also exists in the upstream region, but it is not as high and it extends only 130 bp upstream of URF-62.

In particular, no region corresponding to the consensus "–35" and "–10" promoter sequences could be found in wheat, maize or spinach DNA upstream of the open reading frame, suggesting that URF-62 is perhaps co-transcribed together with the 2 protein genes, *psb*C and *psb*D, which have been found 500 bp upstream in maize and spinach chloroplast DNA (Krebbers 1983, Holschuh et al. 194b).

Discussion

The data reported here have allowed to locate precisely, within a 4.2 kbp region of the wheat chloroplast genome, five tRNA genes (*trn*M-CAU, *trn*G-UCC, *trn*T-GGU, *trn*E-UUC, *trn*Y-GUA) and an open reading frame of 62 codons. We have previously reported (Quigley et al. 1985) the location in the same region of three other tRNA genes (*trn*G-GCC, *trn*D-GUC and *trn*C-GCA).

The arrangement of these eight tRNAs and of URF-62, is as followed (Fig. 2): URF-62 – 277 bp – *trn*G-GCC – 449 bp – *trn*M – 91 bp – *trn*G-UCC – 1.2 kbp – *trn*T – 373 bp – *trn*E – 62 bp – *trn*Y – 366 bp – *trn*D – 2.2 kbp – *trn*C. The open reading frame (URF-62) and *trn*G-GCC, *trn*T, *trn*E, *trn*Y, *trn*D are on the same strand as *atp*A (coding for the α subunit of ATP synthase). The three other tRNA genes, namely *trn*M, *trn*G-UCC and *trn*C are on the opposite strand.

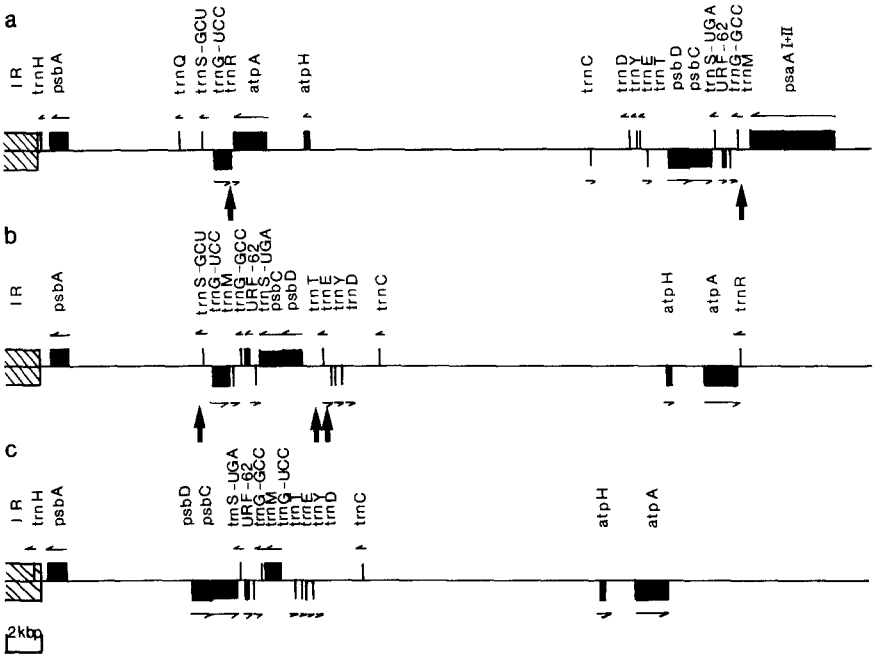

Fig. 10. Possible evolution of the chloroplast genome region located between the inverted repeat (*IR*) and the gene coding for the α subunit of ATP synthase (*atp*A). a Map of this region in the spinach/tobacco chloroplast genome showing the organization of the sequenced tRNA genes (Deno and Sugiura 1983; Holschuh et al. 1983, 1984a, b; Deno and Sugiura 1984; Ohme et al. 1984; Sugita et al. 1984; Zurawski et al. 1984) and of the protein genes either mapped or sequenced (Westhoff et al. 1981; Alt et al. 1983, 1984; Deno et al. 1983; Holschuh et al. 1984b; Sugita and Sugiura 1984). b Map of this region in a putative intermediate chloroplast genome. c Map of this region in the wheat/maize chloroplast genome showing the organization of the sequenced tRNA genes (this study; Schwarzt et al. 1981; Krebbers et al. 1984; Quigley et al. 1985), and of either mapped or sequenced protein genes (Bedbrook et al. 1978; Howe et al. 1982, 1983; Krebbers 1983). *Thick arrows* indicate the proposed ends of genomic inversions by which a can be transformed into b, then into c. The genes are indicated by *black boxes* and *thin horizontal arrows* show the direction of transcription of these genes

The organization of the chloroplast genome of cereals (such as wheat and maize) has been shown to be essentially the same (Vedel et al. 1980). We have indeed found that wheat chloroplast DNA probes (cloned fragments) containing either *trn*G-GCC, or *trn*M or *trn*C hybridized to wheat and maize chloroplast DNA fragments located in the same region of the chloroplast genome (unpublished results). The wheat gene for $tRNA^{Ser}_{UGA}$ is probably located like in maize (Krebbers et al. 1984) just upstream of URF-62, on the opposite strand. In maize this gene (*trn*S-UGA) is preceeded by *psb*C and *psb*D (coding for two photosystem II proteins) which are transcribed in the same direction as URF-62 (Krebbes 1983). These two genes are most likely to be found at the same location on the wheat chloroplast genome. The organization of the tRNA genes sequenced in this study and of the protein genes either mapped or sequenced in this region of the wheat/maize chloroplast genome is represented diagramatically in Fig. 10c.

When the organization of the maize chloroplast genome is compared to that of the spinach or tobacco chloroplast genomes (spinach and tobacco chloroplast genomes are thought to be essentially colinear), (Fluhr and Edelman 1981), it appears that an inversion of a 20 kbp fragment must have occurred during evolution (Palmer and Thompson 1982). The relative location of the gene coding for the α subunit of ATP synthase in the wheat (Howe et al. 1983), spinach (Westhoff et al. 1981) or tobacco (Deno et al. 1983) chloroplast genomes has confirmed this hypothesis. Our finding that *trn*D and *trn*C are transcribed in wheat in the opposite direction as compared to the spinach genes is also in agreement with an inversion of part of the chloroplast genome (Quigley et al. 1985). Cross-hybridization studies have suggested that regions near *atp*A and wheat chloroplast fragment B18 were involved in the rearrangements of the wheat chloroplast genome relatively to spinach (Howe et al. 1983). It is now possible to define more precisely the regions involved in the rearrangements by looking at the organization of the genes in this part of the wheat/maize (Fig. 10c) and spinach/tobacco (Fig. 10a) chloroplast genomes.

The gene for the initiator $tRNA^{Met}$ on one hand and the genes for $tRNA^{Gly}_{UCC}$ and *atp*A on the other hand are located on the tobacco chloroplast genome about 30 kbp apart (Ohme et al. 1984; Deno and Sugiura 1984). On the wheat genome these two genes are separated by only 91 bp, but there is a distance of 22 kbp between

*atp*A on one hand and *trn*G-UCC and *trn*M on the other hand. These data could be explained by an inversion of a 20–30 kbp tobacco fragment, one end of the inversion being near the upstream end of *trn*M, the other end being close to the downstream end of *trn*G-UCC (Fig. 10). The 20–30 kbp inversion would correspond to the 20 kbp inversion proposed by Palmer and Thompson and is compatible with the inverse organization of *trn*T, *trn*E, *trn*Y, *trn*D and *trn*C on the wheat and tobacco/spinach chloroplast genomes (Fig. 10). The high nucleotide sequence homology observed in the region upstream of the wheat and tobacco *trn*M and the apparent duplication of the same region in wheat may be linked with this inversion. In order to explain the orientation of URF-62, *trn*G-GCC, *trn*M and *trn*G-UCC on wheat chloroplast fragment B18, it is necessary to postulate a second (smaller) inversion of about 6 kbp. It is not possible to locate precisely the ends of this inversion, however one end must lye upstream of *psb*C, in order to explain the position of this gene on the maize genome. As the distance between *trn*T and *trn*G-UCC is about 1.2 kbp in wheat, only a relatively short DNA fragment upstream of *trn*G-UCC, possibly including *trn*S-GCU may have been carried along during this inversion.

Finally, as wheat *trn*T is transcribed in the opposite direction as compared to spinach *trn*T, the region containing this gene must have been inverted at some time during the evolution of the chloroplast genome. The region upstream of *trn*T is one of the two regions of the wheat chloroplast genome where heterogeneity has been observed (Bowman et al. 1983). Such clustering of DNA alterations in plant chloroplast genomes has been taken as an evidence of the existence of "hot spots" of DNA rearrangements (Kung et al. 1982). Our data show that at least two DNA inversion ends map in this region.

In order to get a better understanding of the inversions/rearrangements which have led from a spinach-like ancestor chloroplast genome to the wheat chloroplast genome, it would be necessary to compare the sequences of other junction regions with the wheat chloroplast *trn*M – *trn*G-UCC junction, as they are likely to have been involved in the inversions/rearrangements events.

Acknowledgements. We thank Ms. A. Klein for her excellent technical assistance, Dr. L. Bogorad and Dr. E. Krebbers for communicating maize chloroplast DNA sequence data prior to publication, Dr. T. A. Dyer for helpful discussions, the "Ministère de l'Industrie et de la Recherche" for a fellowship to F. Quigley and Ms. C. Schaeffer for help in the preparation of this manuscript.

References

Alt J, Winter P, Sebald W, Moser JG, Schedel R, Westhoff P, Herrmann RG (1983) Curr Genet 7:129–138

Alt J, Morris J, Westhoff P, Herrmann RG (1984) Curr Genet 8:597–606

Bedbrook JR, Link G, Coen DM, Bogorad L, Rich A (1978) Proc Natl Acad Sci USA 75:3060–3064

Bohnert HJ, Crouse EJ, Schmitt JM (1982) In: Parthier B, Boulter D (eds) Encyclopedia of plant physiology, nucleic acids and proteins in plant II. Springer, Berlin Heidelberg, New Series 14B:475–530

Bowman CM, Koller B, Delius H, Dyer TA (1981) Mol Gen Genet 183:93–101

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

Canaday J, Guillemaut P, Weil JH (1980) Nucleic Acids Res 8:999–1008

Clewell DB, Helinski DR (1969) Proc Natl Acad Sci USA 62: 1159–1166

Crouse EJ, Bohnert HJ, Schmitt JM (1984) Chloroplast RNA synthesis. In: Ellis RJ (ed) Chloroplast biogenesis, Seminar Series of the Society for Experimental Biology. Cambridge University Press, 21:83–136

Deno H, Sugiura M (1983) Nucleic Acids Res 11:8407–8414

Deno H, Sugiura M (1984) Proc Natl Acad Sci USA 81:405–408

Deno H, Shinozaki K, Sugiura M (1983) Nucleic Acids Res 11: 2185–2191

Ecarot-Charrier B, Cedergren RS (1976) FEBS Lett 63:287–290

Fluhr R, Edelman M (1981) Nucleic Acids Res 9:6841–6853

Holschuh K, Bottomley W, Whitfeld PR (1983) Nucleic Acids Res 11:8547–8554

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

Holschuh K, Bottomley W, Whitfeld PR (1984b) Nucleic Acids Res 12:8819–8834

Hopp TP, Woods KR (1981) Proc Natl Acad Sci USA 78:3824–3828

Howe CJ, Auffret AD, Doherty A, Bowman CM, Dyer TA, Gray JC (1982) Proc Natl Acad Sci USA 79:6903–6907

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

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

Krebbers ET (1983) Characterization of some *Zea mays* chloroplast genes. PhD thesis, Harvard University, Cambridge, Mass

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

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

Michel F, Dujon B (1983) EMBO J 2:33–38

Ohme M, Kamogashira T, Shinozaki K, Sugiura M (1984) Nucleic Acids Res 12:6741–6749

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

Quigley F, Grienberger JM, Weil JH (1985) Plant Mol Biol 4:305–310

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

Schwarz Z, Jolly SO, Steinmetz AA, Bogorad L (1981) Proc Natl Acad Sci USA 78:3423–3427

Smith AJH (1980) Methods Enzymol 65:560–580

Sprinzl M, Gauss DH (1984) Nucleic Acids Res 12:r59–r132

Steinmetz AA, Krebbers ET, Schwarz Z, Gubbins EJ, Bogorad L (1983) J Biol Chem 258:5503–5511

Sugita M, Sugiura M (1984) Mol Gen Genet 195:308–313

Sugita M, Kato A, Shimada H, Sugiura M (1984) Mol Gen Genet 194:200–205

Vedel F, Lebacq P, Quetier F (1980) Theor Appl Genet 58: 219–224

Westhoff P, Nelson N, Bunemann H, Herrmann RG (1981) Curr Genet 4:109–120

Zurawski G, Bottomley W, Whitfeld PR (1984) Nucleic Acids Res 12:6547–6558

Communicated by H. Kössel

Received February 13, 1985