Curr Genet (1993) 23:166–174

Current Genetics

A transcription map of the chloroplast genome from rice (*Oryza sativa*)

Akira Kanno and Atsushi Hirai

Graduate Division of Biochemical Regulation, School of Agricultural Sciences, Nagoya University, Nagoya 464-01, Japan

Received April 14, 1992/Accepted June 9, 1992

Abstract. The complete nucleotide sequence (134,525 bp) of the chloroplast genome from rice (*Oryza sativa* L.) contains four rRNA genes, 30 tRNA genes, and over 100 genes that encode proteins. However, expression of only a few of these genes, namely *psbA*, *rbcL*, and *atpB/E*, has been detected. We constructed the complete transcription map of rice chloroplast DNA by Northern hybridization of total RNA from rice seedlings, using subclones from a clone bank of rice chloroplast DNA as probes. Approximately 90% of the chloroplast genome was transcribed, as detected by a non-radioactive hybridization system. Most of the genes on the chloroplast DNA are organized as clusters and are co-transcribed as long primary transcripts. We identified 16 polycistronic transcripts from the rice chloroplast genome. Furthermore, the genes for components of photosystems I and II, the gene for the large subunit of RuBisCO, the genes for ATPase, the genes for components of the cytochrome complex, and the rRNA genes were expressed at the highest levels.

Key words: Rice – *Oryza sativa* – Chloroplast genome – Transcription map

Introduction

The chloroplast DNA (ctDNA) of higher plants is 120–160 kb long and encodes over 100 genes (Palmer 1985). The transcription system of ctDNA is very similar to that of prokaryotic genomes. The ctDNA genes have prokaryote-type promoters and terminators and some are organized as clusters and are transcribed polycistronically. There are many reports of the patterns of transcription of ctDNA but they are limited to important genes; for example, the light-regulated expression of the *psbB* operon and the *psbD/C* operon (Tanaka et al. 1987; Westhoff and Herrmann 1988; Yao et al. 1989). The expression of the entire ctDNA has been examined in a few organisms: transcription of the ctDNA of *Euglena* by heteroduplex mapping (Koller and Delius 1984) and that of mustard by hybridization to ctDNA fragments of radiolabelled chloroplast RNA (Link 1984). In the ctDNA of pea, many transcription units were detected by Northern hybridization of total pea RNA with pea ctDNA clones that cover approximately 90% of the genome (Woodbury et al. 1988). However, the genes encoded in the transcripts were not precisely defined because the entire sequence of pea ctDNA has not been determined.

The complete nucleotide sequence of rice ctDNA was published by Hiratsuka et al. (1989). There are four rRNA genes, 30 tRNA genes, and over 100 genes that encode proteins, in the ctDNA. However, transcription of only a few of these genes has been detected; namely, *psbA* (Kanno and Hirai 1989), *rbcL* (Nishizawa and Hirai 1987) and *atpB/E* (Nishizawa and Hirai 1989). In this study, we have determined the complete transcription map of rice ctDNA using a bank of overlapping clones.

Materials and methods

Plant material and preparation of total RNA from green leaves. Seedlings of rice (*Oryza sativa* L. cv. Nipponbare) were grown in a growth chamber at 28 °C for 10 days under fluorescent light. Rice leaves were blended in liquid nitrogen, and total RNA was prepared by the method of Kawakami and Watanabe (1988).

Clones of rice chloroplast DNA. For Northern analysis, subclones from the clone bank of rice chloroplast DNA (Hirai et al. 1985) were used as probes. Each probe is defined by a capital letter and a number.

Agarose gel electrophoresis and Northern hybridization with non-radioactive probes. Aliquots of 5 μg and 20 μg of total RNA were denatured by formaldehyde and separated in a denaturating gel. After electrophoresis, the RNA was transferred to a nylon membrane by a vacuum-transfer method using a Vacugene system (Pharmacia LKB Biotechnology, Sweden). After drying, the membrane was exposed to UV light for cross-linking. Hybridizations were carried out using a non-radioactive DNA labelling and detection kit

Correspondence to: A. Hirai

(Boehringer Mannheim, Germany). All buffers, except for the 'colour solution' of nitroblue tetrazolium salt and 5-bromo-4-chloro-3-indolyl phosphate, were autoclaved at 120 °C for 20 min to inactivate RNases. RNA molecular weight marker I (Boehringer Mannheim) was used as a size marker.

Results

General characteristics of the transcription of rice chloroplast DNA

Almost all genes identified by sequencing were transcribed. Many are organized as gene clusters and their expression patterns were polycistronic. However, transcripts of some genes for ribosomal proteins and RNA polymerase were detected at very low levels, if at all, probably because of the low sensitivity of our non-radiolabelled hybridization system. The bands around 1.7 kb and 2.6 kb are assumed to be artifacts due to cytoplasmic ribosomal RNAs since we used total RNA. Almost all transcripts corresponded to the genetic map derived from the sequencing data. We found none that did not cover regions of mapped genes but as described below, one transcript, which includes the gene for ORF29, did not correspond to the length of the corresponding gene.

Transcription around the rbcL and atpB/E genes

Northern hybridization patterns with probes from around the *rbcL* gene, which encodes the large subunit of RuBisCO, are shown in Fig. 1. With P14, transcripts of 2.6 kb and 1.9 kb were detected. Since transcripts of 2.6 and 1.9 kb hybridized with F3 and P15, respectively, these transcripts correspond to *atpB/E* and *rbcL*, respectively; *atpB* and *atpE* encode the β and ε subunits of ATP synthase. With F7, transcripts of 7.0, 2.6, 0.8 and 0.1 kb were detected. Transcripts of 0.8 and 0.1 kb were assumed to correspond to *trnV* and *trnM*, respectively, since only the 0.8-kb transcript hybridized with F10. *trnV* and *trnM* are the genes for $tRNA^{Val}$ and $tRNA^{Met}$, respectively. The *trnV* gene contains an intron, but we did not detect the spliced transcript (mature tRNA) of this gene. The 7.0-kb transcript hybridized less strongly with F7, F2 and F6, indicating that this transcript encodes all the genes between *atpB/E* and ORF159. For unknown reasons, however, we failed to detect this transcript with P14, F3, F10, F14 and F12. Three transcripts of 3.0, 2.7 and 2.6 kb, detected with F2 but not with F7, also hybridized with F6. These transcripts probably correspond to *ndhC*, *ndhK* and ORF159, suggesting that these genes are expressed polycistronically. *ndhC* and *ndhK* encode the subunits of NADH dehydrogenase. Transcripts of 0.8 and 0.1 kb hybridized with F5 and F12 and the transcript of 0.1 kb hybridized with F14. Thus, the transcript of 0.8 kb corresponds to *trnL*, the gene for $tRNA^{Leu}$, and the transcript of 0.1 kb corresponds to both *trnF*, the gene for $tRNA^{Phe}$, and spliced *trnL*. The expression of tRNA genes is discussed below.

The complete transcription pattern of rice ctDNA was revealed by the use of probes that covered the entire ctDNA.

The complete transcription map of rice ctDNA

The complete map is shown in Fig. 2. The transcription of the genes is indicated in order of their locations within the ctDNA. The numbering starts from the junction between the inverted repeat and the large single-copy region near the *psbA* gene and proceeds counterclockwise around the ctDNA (Fig. 3) as indicated in an earlier study (Hiratsuka et al. 1989):

(1) 1–1 200; *psbA*. *psbA* encodes the 32-kDa protein of the PS II reaction center. It was previously reported that the *psbA* mRNA was 1.2 kb in length and the site of initiation of transcription was mapped 77 bp upstream of the ATG codon by primer extension analysis (Kanno and Hirai 1989). In almost complete agreement with this study, the 1.1-kb transcript hybridized with A1. We did not detect the polycistronic transcript of *psbA* and *trnH* which was identified after in vitro transcription in spinach chloroplasts (Stern and Gruissem 1987).

(2) 1 200–4 500; *trnK*, ORF542. An open reading frame (ORF542) is found in the intron of rice *trnK*, which encodes $tRNA^{Lys}$ (Hiratsuka et al. 1989), and seems to encode a maturase-like protein, as described in mustard (Neuhaus and Link 1987). A transcript of 2.7 kb hybridized with A2 and A3, and seemed to be the transcript of *trnK* and ORF542. However, we did not detect the spliced transcripts of both genes. Except for *trnF*, the spliced transcripts (mature tRNA) of other tRNA genes that contain introns were not detected.

(3) 4 500–7 000; *trnQ*, *rps16*. *trnQ* and *rps16* encode $tRNA^{Gln}$ and a ribosomal protein, respectively. Transcripts of 2.0, 1.6, 1.4, 1.2, 0.4 kb hybridized with A7 and A8 and an 0.1-kb transcript hybridized only with A8. Thus, *trnQ* and *rps16* were co-transcribed as a transcript of 2.0 kb. Three transcripts of 1.6–1.2 kb were intermediates in the splicing process and the 0.1-kb transcript was mature $tRNA^{Gln}$. The transcript of 0.4 kb, which did not hybridize with A10, was probably the spliced *rps16* mRNA.

(4) 7000–14000; *psbK*, *psbI* ORF100, *psbD*, *psbC*, ORF62, *trnG*, *trnS*, *trnfM*. *psbK* and *psbI* encode two small polypeptides, K and I, of PS II. *psbD* and *psbC* encode the D2 protein and the 43-kDa polypeptide of PS II. *trnG*, *trnS*, and *trnfM* encode $tRNA^{Gly}$, $tRNA^{Ser}$, and $tRNA^{fMet}$, respectively.

We detected nine transcripts using A5. Since the transcript of 6.1 kb was long enough to include all the genes between *psbK* and *trnG*, the complicated pattern of transcription is probably caused by polycistronic transcription of genes in this region. We prepared some shorter probes (A9, A6, B1, B3, B7, B6, B9) and carried out Northern hybridization (data not shown) to construct the transcription map of this region (Fig. 2).

In tobacco ctDNA, the locations of *psbK/psbI* and *psbD/C* are different and transcription of these regions is

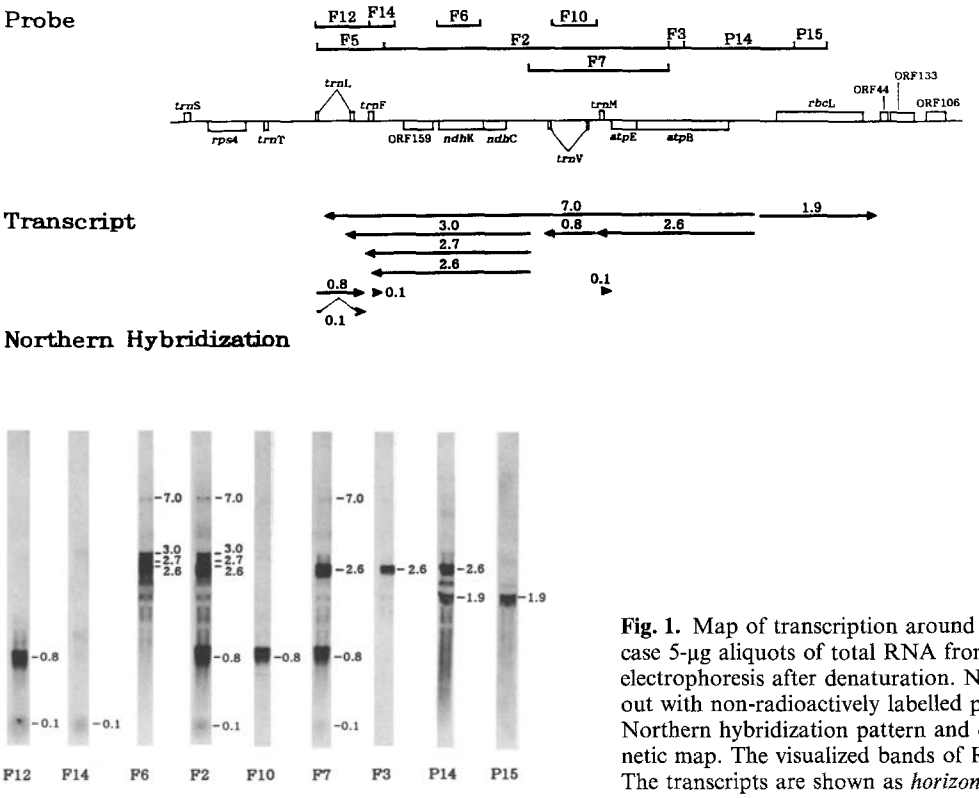


Fig. 1. Map of transcription around the *rbcL* and *atpB/E* genes. In each case 5-μg aliquots of total RNA from rice seedlings were subjected to electrophoresis after denaturation. Northern hybridizations were carried out with non-radioactively labelled probes, as indicated under each Northern hybridization pattern and depicted by *bold lines* above the genetic map. The visualized bands of RNA are labelled by their sizes in kb. The transcripts are shown as *horizontal arrows* under the genetic map

independent (Yao et al. 1989; Meng et al. 1991). In rice, ctDNA has a large inversion and *psbK/psbI* and *psbD/C* are located close to one another and transcribed together. However, it is not clear whether the promoter of *psbD/C* is functional. ORF100 was co-transcribed with *psbK/psbI* and *psbD/C* but the function of the protein encoded by ORF100 is unknown.

There are two *trnS* genes and two *trnG* genes in this region. We detected the transcript of the *trnS*(GCU) gene downstream from the *psbI* gene (transcript of 0.1 kb) but did not detect the other *trnS*(UGA) gene, which is located downstream from the *psbC* gene, with A6 and B3. We did, however, detect the transcript of the *trnG* genes with B9 and B8: the *trnG*(UCC) gene has an intron while the *trnG*(GCC) gene does not. Transcripts of the *trnG*(UCC) and *trnG*(GCC) gene were 0.8 kb and 0.1 kb long, respectively. However, we failed to detect the mature transcript of the *trnG*(UCC) gene which contains an intron, as does *trnK*. The transcript of *trnfM* was detected with B10.

(5) 14000–17000; ORF91, ORF70, *trnT*, *trnE*, *trnY*, *trnD*, *psbM*. *trnT*, *trnE*, *trnY*, and *trnD* encode $tRNA^{Thr}$, $tRNA^{Glu}$, $tRNA^{Tyr}$, and $tRNA^{Asp}$, respectively. *psbM* encodes the polypeptide M of PS II.

B13, B15, B16 and B14 hybridized to 0.1-kb transcripts, an indication that the four tRNA genes in this region are expressed.

The transcripts corresponding to ORF91, ORF70 and *psbM* were not detected with B4, B8 and B11, respectively. The product of the *psbM* gene is considered to be a component of the PSII reaction center, so *psbM* should be expressed even though it was not detected under our conditions.

(6) 17000–19000; ORF29, *trnC*. With C5, three transcripts (0.7, 0.3 and 0.1 kb) were detected. The transcript of 0.1 kb hybridized with C5 and C8 but not with C7. Thus, this transcript corresponds to *trnC*, which encodes $tRNA^{Cys}$. A transcript of 0.7 kb hybridized with C7, but not with B11 and C8, indicating that the transcript includes almost the entire sequence of the C7 probe. The 0.3-kb transcript hybridized with C7, but not with C8 and C9, indicating that the transcript is a result of the processing of the 0.7-kb transcript. These two transcripts may encode ORF29 but they are too large to encode ORF29 since the coding region of this gene is only 90 bp. These transcripts correspond either to the ORF29 gene or to an undetected gene near ORF29. The amounts of these transcripts are comparatively high, so they may be functional. However, we found no ORF apart from ORF29 and the sequence around this region has no homology to any reported genes. It is now necessary to determine which strand is actually transcribed.

(7) 19000–29500; *rpoB*, *rpoC1*, *rpoC2*. These three genes are considered to encode the subunits of a putative chloroplast RNA polymerase. A transcript of 6.4 kb hybridized weakly with C3 and C6, and a transcript of 2.7 kb also hybridized weakly with C3 but not with C6. Using C4, we did not detect any transcripts. Thus, the 6.4-kb transcript corresponds to the *rpoC1* and *rpoC2* genes and the 2.7-kb transcript to the *rpoC1* gene. The transcript of 6.4 kb is a little too short to cover the *rpoC1* and *rpoC2* genes and the transcript of 2.7 kb is a little too large to cover only the *rpoC1* gene. If these genes are expressed, the amounts of the transcripts of the *rpoB*, *rpoC1* and *rpoC2* genes appear to be very low.

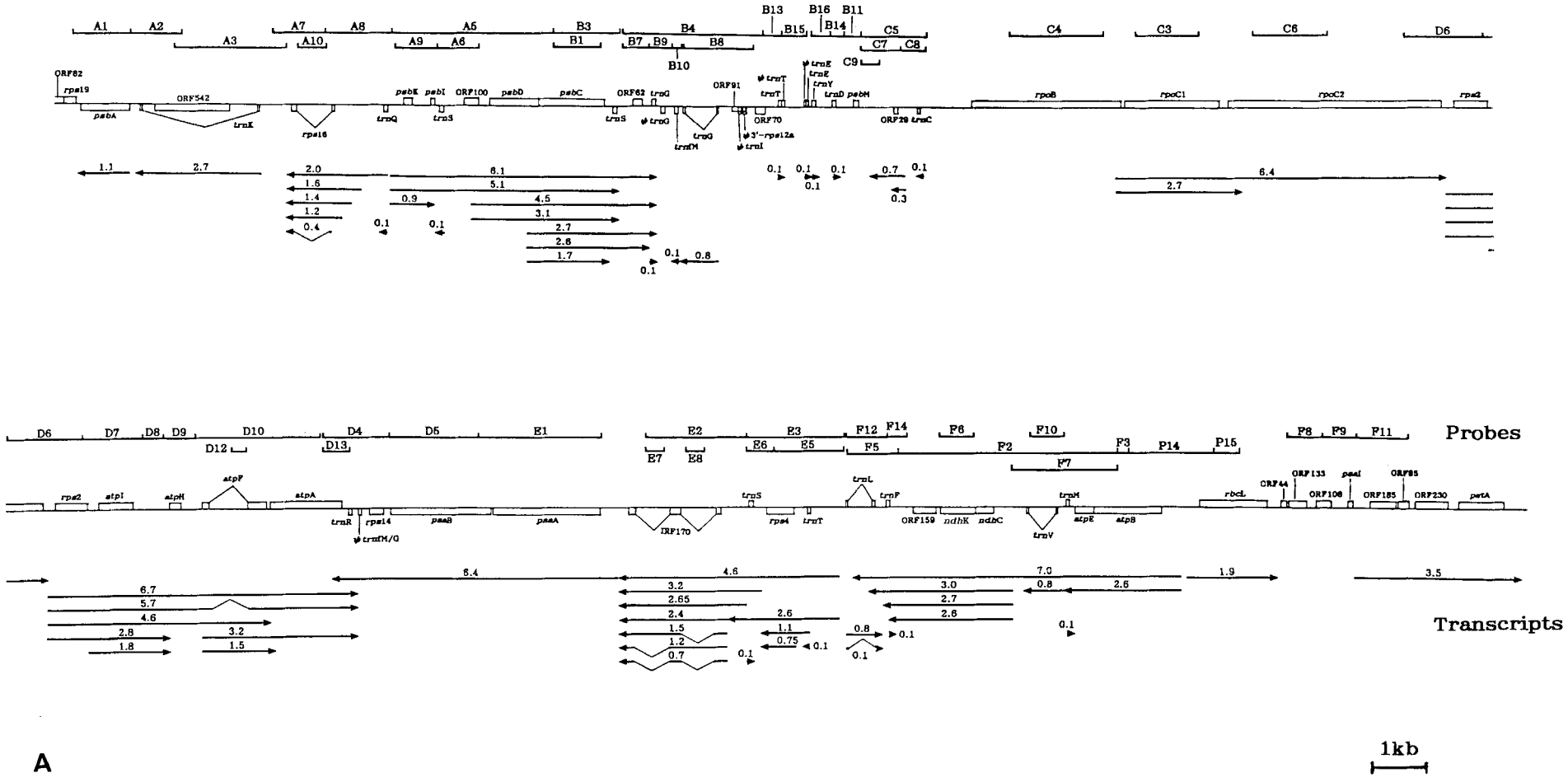
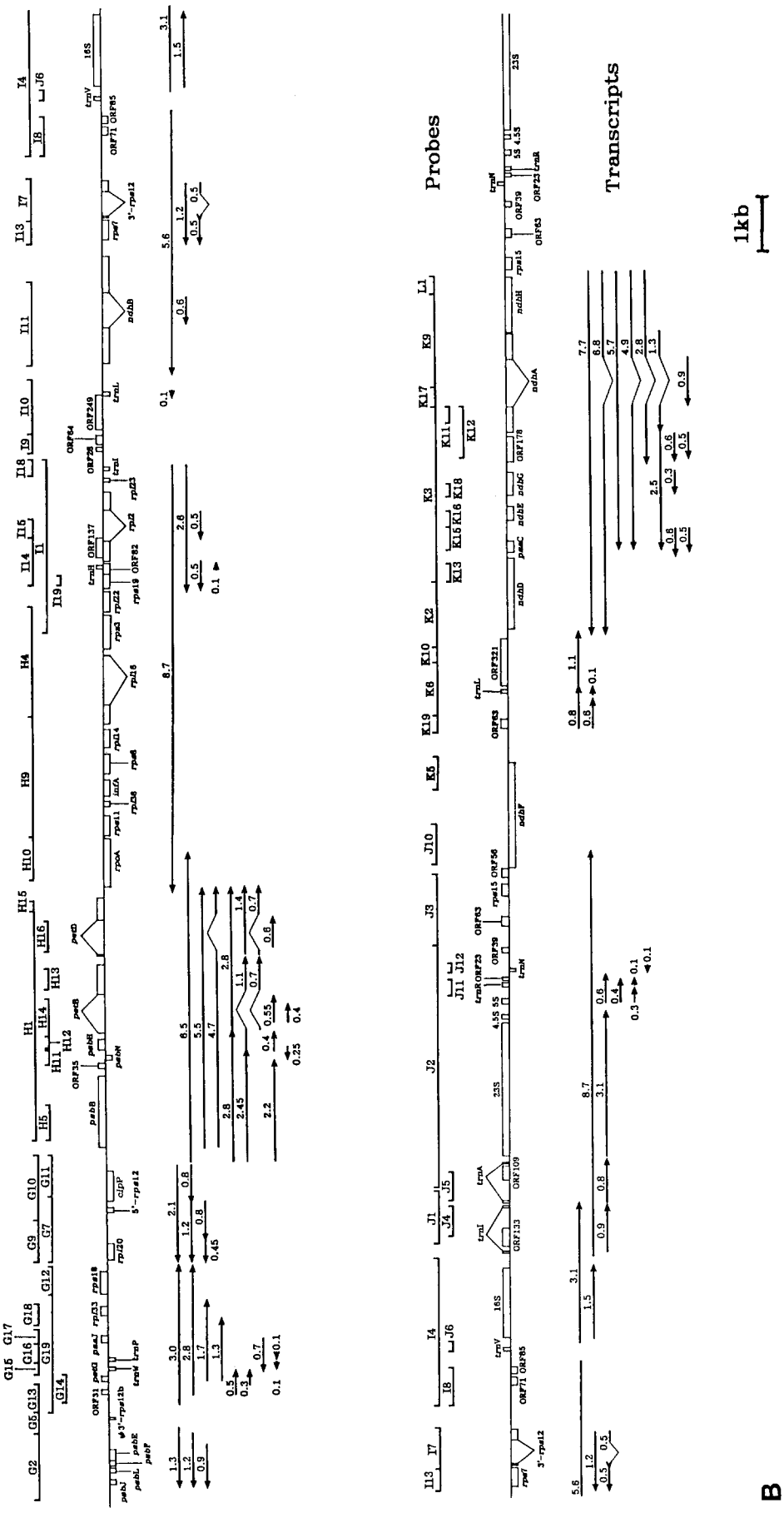


Fig. 2A, B. The complete map of the transcription of rice chloroplast DNA. As in Fig. 1, the probes are depicted as *bold lines* above the genetic map and the transcripts are shown as *horizontal arrows* under the genetic map



B

Fig. 2A, B. (Continued)

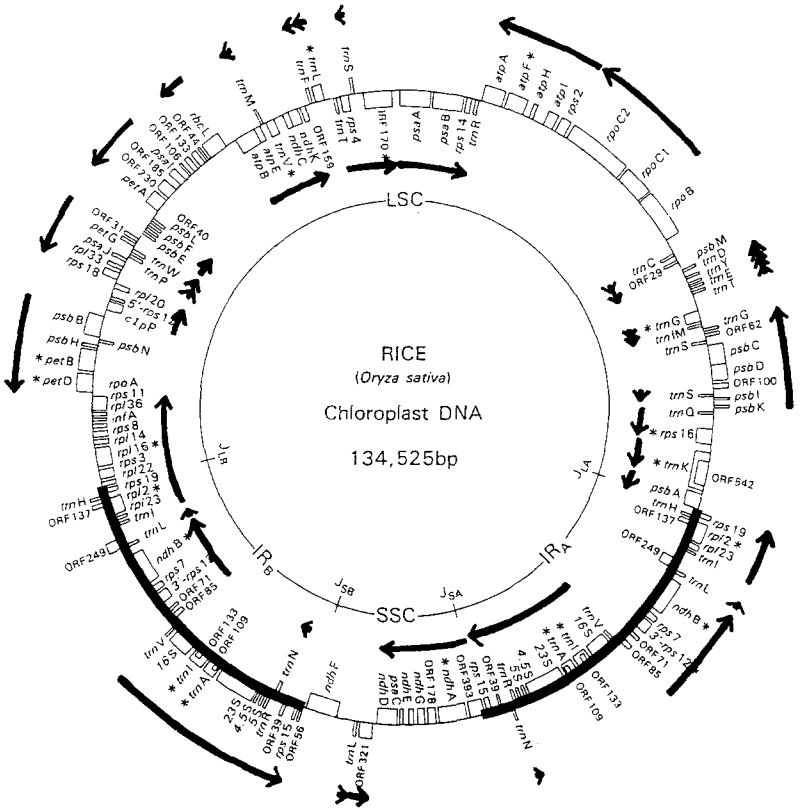


Fig. 3. Circular map of the transcription of rice chloroplast DNA. Genes shown on the outside of the circle are transcribed counter-clockwise and those on the inside of the circle are transcribed clockwise. Polycistronic and monocistronic transcripts are shown as *arrows* around the circle

(8) 29 500–35 800; *rps2*, *atpI*, *atpH*, *atpF*, *atpA*. *rps2* encodes a ribosomal protein, and *atpI*, *atpH*, *atpF* and *atpA* encode subunits of ATPase. A transcript of 6.7 kb hybridized very strongly with D6, D7, D8, D9, D10 and D13, indicating that *rps2*, *atpI*, *atpH*, *atpF*, *atpA* were co-transcribed. With D6, transcripts of 5.7, 4.6, 2.8 and 1.8 kb were detected. Since the 5.7-kb transcript did not hybridize with D12, it must be a spliced transcript in which the intron of the *atpF* gene has been excised. The transcript of 4.6 kb did not hybridize with D13. Thus, this is the transcript that encodes the sequence from *rps2* to *atpF* but is not spliced. Using D12, four transcripts of 6.7, 4.6, 3.2 and 1.5 kb were detected. The transcript of 3.2 kb hybridized with D10 and D13, but the transcript of 1.5 kb hybridized only with D10. Thus, the 3.2-kb transcript corresponds to the unspliced transcripts of both the *atpF* and *atpA* genes, while the transcript of 1.5 kb represents the unspliced transcript of the *atpF* gene.

In pea, *atpI*, *atpH*, *atpF* and *atpA* are co-transcribed as a transcript of 6.0 kb, and spliced and/or processed to form shorter transcripts (Woodbury et al. 1988). The expression pattern in this region is very similar to that in rice.

(9) 35 800–41 500: *trnR*, *rps14*, *psaB*, *psaA*. *trnR* and *rps14* encode $tRNA^{Arg}$ and a ribosomal protein, respectively. *psaB* and *psaA* are the genes for the P700 apoproteins A2 and A1 of PS I, respectively.

Using D4, D5 and E1, we detected only one band of 6.4 kb. Thus, *psaA*, *psaB*, *rps14*, *trnR* were co-transcribed in that order. We did not detect any processing intermediates.

(10) 41 500–46 500; IRF170, *trnS*, *rps4*, *trnT*. IRF170 encodes an intron-containing reading frame and the function of the putative product has not been determined; *trnS* and *trnT* encode $tRNA^{Ser}$ and $tRNA^{Thr}$, respectively; *rps4* encodes a ribosomal protein.

A transcript of 4.6 kb hybridized with E2 and E3. Thus, three genes, *trnT*, *rps4* and IRF170, were co-transcribed in that order. Many shorter transcripts were also detected with the E2 and E3 probes. Six transcripts hybridized with E2, and four with E3. These transcripts are intermediates due to splicing. Using E7 and E8, which are specific for the introns of IRF170, we determined the precise transcription map of this region, as shown in Fig. 2. The product(s) of IRF170 should be important, since the level of transcripts from this gene was high. We detected many transcripts that were intermediates due to splicing of IRF170. Thus, splicing of this gene appears to be very slow.

Transcripts of 0.1 kb hybridized with E6 and E5. Thus, these transcripts correspond to the *trnS* and *trnT* genes.

(11) 55 700–61 000; ORF44, ORF133, ORF106, *psaI*, ORF185, ORF85, ORF230, *petA*. *psaI* and *petA* encode the small peptide of PS I and the apoprotein of cytochrome f, respectively. A transcript of 3.5 kb hybridized strongly with F11 but no transcripts were detected with F8 and F9. Thus, the transcript of 3.5 kb includes the sequences from ORF185 to *petA*.

(12) 61 000–62 500; *psbJ*, *psbL*, *psbF*, *psbE*. *psbJ* and *psbL* encode a component of PS II and the L protein of PS II, respectively. *psbF* and *psbE* are the genes for the two components of the apocytochrome b-559 protein. Using G2, we detected a transcript of 0.9 kb with a very distinct band and two transcripts of 1.2 and 1.3 kb with very weak signals. These data indicate that *psbE*, *psbF*,

psbL and *psbJ* were co-transcribed in that order. Two low-level RNAs (1.2 and 1.3 kb) were the pre-mature transcripts and were cleaved to 0.9 kb.

(13) 62500–66200; ORF31, *petG*, *psaJ*, *rpl33*, *rps18*, *trnW*, *trnP*. *petG* and *psaJ* encode subunit 5 of the cytochrome b/f complex and a small peptide of PS I. *rpl33* and *rps18* are genes for ribosomal proteins.

Two transcripts of 3.0 and 2.8 kb hybridized with G19 and G12 but no transcript was detected with G5. Thus, these transcripts correspond to the genes between ORF31 and *rps18*. The 2.8-kb transcript is assumed to be a product generated by cleavage of the 5′ or 3′ end of the 3.0-kb transcript. With G19, other transcripts of 1.7, 1.3, 0.7, 0.5, 0.3 and 0.1 kb were detected. The 1.7-kb transcript hybridized with G14, G15, G16, G17 and G18, but not with G13 and G12. Thus, this transcript corresponds to the *petG*, *psaJ* and *rpl33* genes. The transcript of 1.3 kb, which hybridized to the same probes as the 1.7-kb transcript is the processed product of the 1.7-kb transcript, and it probably corresponds to *petG* and *psaJ*.

The transcript of 0.5 kb hybridized with G13 and G14 and the 0.3-kb transcript hybridized with only G14, an indication that the 0.5-kb transcript corresponds to the ORF31 and *petG* genes, and the 0.3-kb transcript to only the *petG* gene.

The 0.7-kb transcript hybridized with G15 and G16, but not with G14 and G17, indicating that this transcript encodes two tRNAs but does not correspond to the *petG* and *psaJ* genes. The transcript of 0.1 kb hybridized with G15 and G16. G15 is a specific probe for *trnW*, which encodes $tRNA^{Trp}$, and G16 is specific for *trnP*, which encodes $tRNA^{Pro}$. Thus, both these genes are expressed. However, the 0.1-kb transcript hybridized less strongly with G16 than with G15, suggesting that the amount of each tRNA is different.

(14) 66200–68500; *rpl20*, 5′-*rps12*, *clpP*. *rpl20* and *rps12* are genes for ribosomal proteins. 5′-*rps12* is the 5′ end of the *rps12* gene. The 3′ end of this gene is located on inverted repeats and is expressed independently. The transcripts of 5′- and 3′-*rps12* are spliced in trans (Koller et al. 1987; Zaita et al. 1987; Hildebrand et al. 1988).

A transcript of 2.1 kb hybridized with G7, G9, G10 and G11. Thus, *clpP*, 5′-*rps12* and *rpl20* are co-transcribed in that order. With these probes, some shorter transcripts were found in this region. Three transcripts of 1.2, 0.8 and 0.45 kb hybridized with G9 and an 0.8-kb transcript also hybridized with G11. Thus, the 0.8-kb transcripts detected by G9 and G11 appear to be different. These results indicate that the transcript of 2.1 kb was cleaved first between 5′-*rps12* and *clpP*, so generating transcripts of 0.8-kb and 1.2-kb. Then the 1.2-kb transcript was cleaved upstream of the *rpl20* gene resulting in the formation of the 0.8-kb and 0.45-kb transcripts.

(15) 68500–74200; *psbB*, ORF35, *psbH*, *petB*, *petD*, *psbN*. *psbB*, *psbH* and *psbN* encode proteins of PS II. *petB* and *petD* are the genes for cytochrome b6 and the cytochrome b/f complex, respectively.

The transcripts from this region (*psbB* operon) were resolved into 18 major RNA species in spinach (Westhoff and Herrmann 1988). Similarly, we detected 13 major transcripts in this region using H1. The order of the genes and the length of this operon are quite similar in rice and spinach. However, the splicing is somewhat different.

Specific probes were constructed for the coding region of each gene in this operon and the introns of *petB* and *petD*. The precise transcription map of this operon is shown in Fig. 2. The primary transcript of 6.5 kb is slightly longer than that in spinach and tobacco (Sugiura 1991). The RNA of 4.7 kb includes the spliced transcript of *petD*, but we did not detect the spliced transcript of *petB* which was detected in spinach. With H14, *petB* intron-specific transcripts (0.55 and 0.4 kb) were detected. They were too short to be the excised intron. Thus, they may not be by-products formed during splicing but, instead, may be functional transcripts.

(16) 74200–83300; *rpoA*, *rps11*, *rpl36*, *infA*, *rps8*, *rpl14*, *rpl16*, *rps3*, *rpl22*, *rps19*, ORF82, *rpl2*, *rpl23*, *trnI*, *trnH*, ORF137. *rps11*, *rpl36*, *rps8*, *rpl14*, *rpl16*, *rps3*, *rpl22*, *rps19*, *rpl2*, and *rpl23* encode ribosomal proteins. *infA* is thought to encode a chloroplast transcription initiation factor analogous to the initiation factor IF-1 of *E. coli*.

rpoA encodes the α subunit of chloroplast RNA polymerase. As described below, this gene appears to be expressed. With H10, some faint bands of large RNAs were detected, suggesting that the *rpoA* gene was co-transcribed with some other genes. The RNA in the largest faint band was 8.7 kb, which is long enough to include the entire sequence between *rpoA* and *trnI*, which encodes $tRNA^{Ile}$. This transcript hybridized with H9, H4, I1, and I18, indicating that it encodes all these genes.

A transcript of 2.6 kb hybridized with I1 but not with H4, H9 and H10. Thus, this transcript corresponds to the genes between *rps19* and *trnI*. The 0.5-kb transcript hybridized with I19 and I15, which are specific for the *rps19* gene and the intron of the *rpl2* gene, respectively. Thus, there are two kinds of 0.5-kb transcript in this region, one corresponds to *rps19* and ORF82 and the other to the sequence of the *rpl2* intron.

The transcript of 0.1 kb hybridized with I14. Thus, this is the transcript of *trnH*, which encodes $tRNA^{His}$. However, we failed to detect mature tRNA from *trnI* with I18.

(17) 83300–85000; ORF28, ORF64, ORF249, *trnL*. A transcript of 0.1 kb hybridized with I10. From the size of the transcript, this is $tRNA^{Leu}$. No other transcripts were detected with I10 and no transcript hybridized with I9. Thus, ORF28, ORF64 and ORF249 are not expressed or else are expressed at only very low levels.

(18) 85000–91000; *ndhB*, *rps7*, 3′-*rps12*, ORF71, ORF85. A transcript of 5.6 kb hybridized with I11, I7 and I8. Thus, these five genes are co-transcribed. With I11, only an 0.6-kb transcript could be detected in addition to a 5.6-kb transcript. From its size, this small transcript covers the intron of the *ndhB* gene, which encodes the ND2 protein of NADH dehydrogenase. We did not detect specific transcripts of the *ndhB* gene.

A transcript of 1.2 kb hybridized with I13 and I7. Thus, this transcript corresponds to both the *rps7* and 3′-*rps12* genes, which encode ribosomal proteins. In addition to a 1.2-kb transcript, an 0.5-kb transcript was detected with I13 and I7. Thus, these are the transcripts of

rps7 and spliced 3′-*rps12*. We did not detect the transcript of unspliced 3′-*rps12*. As described above, the transcript of the 3′-*rps12* gene should be linked to that of 5′-*rps12* by trans-splicing. The levels of transcripts of 5′- and 3′-*rps12* seemed to be similar.

(19) 91 000–101 500; *trnV*, 16*s* rRNA, *trnI*, ORF133, *trnA*, ORF109, 23*s* rRNA, 4.5*s* rRNA, 5*s* rRNA, *trnR*, ORF23, *trnN*, ORF39, ORF63, *rps15*, ORF56 (truncated *ndhH*). *trnV*, *trnI*, *trnA*, *trnR*, and *trnN* encode $tRNA^{Val}$, $tRNA^{Ile}$, $tRNA^{Ala}$, $tRNA^{Arg}$, and $tRNA^{Asn}$, respectively. *rps15* encodes a ribosomal protein. ORF56 encodes a truncated *ndhH* protein.

A transcript of 3.1 kb hybridized with I4 and J4 but not with J6 and J5. Thus, it corresponds to both 16*s* rRNA and the unspliced *trnI* gene. A transcript of 1.5 kb was also detected with I4 as a very distinct band, and was identified as 16*s* rRNA. We did not detect the mature transcript of the *trnV* gene. Transcripts of 0.9 kb and 0.8 kb hybridized with J4 and J5, respectively. They are unspliced tRNAs. However, we failed to detect mature tRNA with J1. Besides the 0.9-kb and 0.8-kb transcripts, an 8.7-kb transcript was detected with J4 and J5. This transcript also hybridized with J2, J3, and J10. Thus, it covers all genes between *trnI* and ORF56. A transcript of 3.1 kb hybridized with J2, giving a very heavy band, and was identified as 23*s* rRNA.

Transcripts of 0.6, 0.4, 0.3, and 0.1 kb hybridized with J11. The 0.6-kb transcript is probably expressed from the promoter located upstream of the 5*s* rRNA gene. Two transcripts of 0.4 kb and 0.3 kb were the cleaved products of the 0.6-kb transcript. The transcript of 0.1 kb detected with J11 was mature $tRNA^{Arg}$. The 0.1-kb transcript also hybridized with J12, indicating that it is $tRNA^{Asn}$.

(20) 101 500–106 300; *ndhF*, ORF63, *trnL*, ORF321. No transcripts specific for the *ndhF* gene, which encodes the ND5 protein of NADH dehydrogenase, were detected with J10 and K5.

Transcripts of 0.8 kb and 0.6 kb hybridized with K19. A transcript of 0.1 kb also hybridized with K6. Thus, the 0.8-kb transcript corresponds to ORF63 and *trnL*, the gene for $tRNA^{Leu}$, and the 0.6-kb and 0.1-kb transcripts were the cleaved products of the 0.8-kb transcript. A transcript of 1.1 kb hybridized with K10. Thus, it is the transcript of ORF321.

(21) 106 300–114 000; *ndhD*, *psaC*, *ndhE*, *ndhG*, ORF178, *ndhA*, *ndhH*. *ndhD*, *ndhE*, *ndhG*, *ndhA*, *ndhH* encode ND4, ND4L, ND6, ND1 and a 49-kDa protein, respectively. ORF178 encodes a 18-kDa protein related to a subunit of NADH dehydrogenase.

Eleven transcripts were detected with the K9 and K3 probes. Thus, the transcription pattern of the genes in this region is very complex and these genes appear to be co-transcribed, as are those in the *psbB* operon.

A transcript of 7.7 kb hybridized with K2, K3, and K9. This transcript, thus, covers all genes between *ndhD* and *ndhH*. Probes specific for each gene were prepared: K13 for *ndhD*; K15 for *psaC*; K16 for *ndhE*; K18 for *ndhG*; K11 for the coding region of *ndhA*; K12 for ORF178 and the coding region of *ndhA*; K17 for the intron of *ndhA*; and L1 for *ndhH*. Using these probes, we constructed a transcription map for this region, as shown in Fig. 2. Major transcripts were a 2.5-kb transcript covering genes between *psaC* and ORF178, 0.6- and 0.5-kb transcripts of ORF178, and 0.6- and 0.5-kb transcripts of the *psaC* gene.

Discussion

Transcription of the genes on ctDNA

Woodbury et al. (1988) failed to detect transcripts of less than about 0.5 kb because they lost them during preparation of total pea RNA. However, we detected mature tRNAs (of about 0.1 kb) with very distinct bands by Northern hybridization.

Most of the genes on ctDNA form clusters and are transcribed polycistronically. We found 16 such clusters in rice ctDNA (counting the genes on inverted repeats only once) and 22 monocistronic transcripts were observed, as shown in Fig. 2 and 3. Transcription of a total of 116.2 kb was identified, equivalent to approximately 86% of rice ctDNA (134,525 bp). Thirty-three sites of initiation of transcription were found in pea ctDNA (Woodbury et al. 1989), and the number in rice ctDNA is similar.

The genes related to PS I and II, the gene for the large subunit of RuBisCO, the genes for ATPase, the genes related to the cytochrome complex, and the rRNA genes are all expressed at high levels. However, transcripts of the *psbM* and *psaI* genes, which encode constituents of PS I or II, were not detected in our study. It is noteworthy that the levels of these latter transcripts were quite different from those of other related genes.

The transcript of ORF178, which encodes an 18-kDa protein related to NADH dehydrogenase, was detected at high levels. The levels of transcripts of other genes related to NADH dehydrogenase were not as high, indicating that the product of ORF178 may have a specific function.

Transcription of genes for RNA polymerase and tRNAs

We detected a very low level of transcripts from the *rpoA*, *rpoC1* and *rpoC2* genes and none from the *rpoB* gene. Initially, chloroplast RNA polymerase was considered to be encoded in the nuclear genome (Lerbs et al. 1985). However, the sequences of *rpoA*, *rpoB*, and *rpoC*, which potentially encode proteins homologous to the α, β, and β' subunits of RNA polymerase in *E. coli*, have been detected in ctDNA from several plants. Furthermore, synthesis of chloroplast RNA was inhibited by antibodies raised against fusion proteins derived from chloroplast *rpoA*, *rpoB* and *rpoC1/C2* (Little and Hallick 1988), suggesting that the chloroplast *rpo* genes are indeed functional. We failed to detect these transcripts unambiguously, perhaps because of the limited sensitivity of our system or the growth conditions of the plants.

Transcripts of almost all the tRNA genes were identified. We found an interesting pattern of transcription of tRNA genes that contain an intron. Transcripts of 0.8 kb

and 0.1 kb were detected with both of the probes specific for *trnL* and *trnV*. The amount of 0.8-kb (unspliced) transcript was much larger than that of the 0.1-kb mature tRNA. However, unspliced tRNA is assumed to have no function since the intron does not encode any protein. Then why is there so much unspliced tRNA? One possibility is that the expression of these tRNAs is regulated by splicing and unspliced tRNAs are accumulated for protein synthesis.

Transcription of the genes for the chloroplast ribosomal proteins

In *E. coli*, control of the transcription of genes for ribosomal proteins is simplified by their organization into several different operons. The primary control of ribosomal protein synthesis occurs at the level of translation of the mRNA because the ribosomal proteins are repressors of their own translation. However, some genes for ribosomal proteins on ctDNA are organized as clusters with other genes. For example, *rps2* is co-transcribed with *atpA* and *atpH*, *rps14* with *psaA* and *psaB*, and *rpl33* and *rps18* with *petG* and *psaJ*. All these genes are expressed at high levels. However, low levels of the transcripts that correspond to ribosomal protein genes such as *rps3* and *rpl16* are found in chloroplasts. Thus, control of the synthesis of ribosomal proteins in chloroplasts must involve post-transcriptional regulation.

Acknowledgements. This work was supported by Grants-in-Aid from the Ministry of Education, Science, and Culture of Japan.

References

Hildebrand M, Hallick RB, Passavant CW, Bourque DP (1988) Proc Natl Acad Sci USA 85:372–376
Hirai A, Ishibashi T, Morikami A, Iwatsuki N, Shinozaki K, Sugiura M (1985) Theor Appl Genet 70:117–122
Hiratsuka J, Shimada H, Whittier R, Ishibashi T, Sakamoto M, Mori M, Kondo C, Honji Y, Sun CR, Meng BY, Li YQ, Kanno A, Nishizawa Y, Hirai A, Shinozaki K, Sugiura M (1989) Mol Gen Genet 217:185–194
Kanno A, Hirai A (1989) Plant Sci 59:95–99
Kawakami N, Watanabe A (1988) Plant Cell Physiol 29:33–42
Koller B, Delius H (1984) Cell 36:613–622
Koller B, Fromm H, Galun E, Edelman M (1987) Cell 48:111–119
Lerbs S, Brautigam E, Parthier B (1985) EMBO J 4:1661–1666
Link G (1984) Plant Mol Biol 3:243–248
Little MC, Hallick RB (1988) J Biol Chem 263:14302–14307
Meng BY, Wakasugi T, Sugiura M (1991) Curr Genet 20:259–264
Neuhaus H, Link G (1987) Curr Genet 11:251–257
Nishizawa Y, Hirai A (1987) Jpn J Genet 62:389–395
Nishizawa Y, Hirai A (1989) Jpn J Genet 64:223–229
Palmer JD (1985) Annu Rev Genet 19:325–354
Stern DB, Gruissem W (1987) Cell 51:1145–1157
Sugiura M (1991) In: Bogorad L, Vasil IK (eds) The molecular biology of plastids. Academic Press, San Diego, pp 125–137
Tanaka M, Obokata J, Chunwongse J, Shinozaki K, Sugiura M (1987) Mol Gen Genet 209:427–431
Westhoff P, Herrmann RG (1988) Eur J Biochem 171:551–564
Woodbury NW, Roberts LL, Palmer JD, Thompson F (1988) Curr Genet 14:75–89
Woodbury NW, Dobres M, Thompson WF (1989) Curr Genet 16:433–445
Yao WB, Meng BY, Tanaka M, Sugiura M (1989) Nucleic Acids Res 17:9583–9591
Zaita N, Torazawa K, Shinozaki K, Sugiura M (1987) FEBS Lett 210:153–156

Communicated by R. W. Lee