Volume 15 Number 2 1987 Nucleic Acids Research

Rice chloroplast DNA molecules are heterogeneous as revealed by DNA sequences of a cluster of genes

Eunpyo Moon, Teh-Hui Kao[1] and Ray Wu*

Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, NY 14853, USA

Received August 26, 1986; Revised and Accepted December 12, 1986

ABSTRACT

We describe the isolation of two rice chloroplast HindIII fragments (9.5 kb and 5.3 kb) each containing a gene cluster coding for the large subunit of ribulose-1,5-bisphosphate carboxylase (rbcL), β and ε subunits of ATPase (atpB and atpE), $tRNA^{met}$ (trnM) and $tRNA^{val}$ (trnV). All five genes contained in the 9.5 kb fragment are potentially functional, whereas in the 5.3 kb fragment, rbcL is truncated and atpB is frame-shift mutated. The copy number of the 9.5 kb fragment is 10 times that of the 5.3 kb fragment, indicating that the two fragments are probably located on different chloroplast genomes and represent two different (major and minor) genomic populations. Thus, the rice chloroplast genome appears to be heterogeneous, contrary to general belief. We also describe the isolation of a rice mitochondrial HindIII fragment (6.9 kb) which contains an almost complete transferred copy of this chloroplast gene cluster. In this transferred copy, the coding sequences of rbcL, atpE and trnM contain perfectly normal reading frames, whereas atpB has become grossly defective and trnV is truncated.

INTRODUCTION

In higher plants, each cell contains a few thousand copies of circular chloroplast genomes distributed within many chloroplasts. It is generally assumed that these populations of chloroplast DNA (ctDNA) molecules are homogeneous and indistinguishable from one another within each cell[1]. The main reasons for this assumption are uniparental inheritance of the chloroplast genome (predominantly maternal), lack of detectable ctDNA recombination, and rapid vegetative segregation of ctDNA[2,3]. So far, studies of chloroplast genomes from a variety of plants at the molecular level, including their physical mapping, renaturation kinetics, isolation and gene sequencing, all support the notion that there is little or no heterogeneity among the population of ctDNA molecules within each cell[1,4].

In contrast to chloroplasts, mitochondrial genomes of higher plants are heterogeneous, existing in several physical forms within each cell presumably due to frequent DNA recombination[5,6]. This recombination is not

limited to exchanges between mitochondrial DNA (mtDNA) molecules. It has been shown that mitochondrial genomes of several plant species contain DNA sequences with extensive homology to their ctDNA sequences[7-10]. This homology suggests that exchange occurs between the chloroplast and mitochondrial genomes as well, although the direction of transfer appears to be unilateral, from the chloroplast to the mitochondria, not the reverse. These transferred sequences in mitochondrial genomes have been found to have strongest homology to chloroplast ribosomal RNA genes and chloroplast genes coding for the large subunit of ribulose-1,5-bisphosphate carboxylase (rbcL) and β subunit of ATPase (atpB)[7,9,10]. Neither their precise sequences nor the possible functional roles (if any) within the mitochondria has been determined.

During the course of our study of rice chloroplast genes, we found two different rice chloroplast DNA fragments which contain the same gene cluster, rbcL-atpB-atpE-trnM-trnV. This gene cluster was also found in the rice mitochondrial genome, indicating that gene transfer from chloroplast to mitochondrial genome also occurred in rice. After extensive sequencing of all three DNA fragments, two chloroplast copies and the transferred copy in mitochondria, we report here the results of our sequence comparison.

MATERIALS AND METHODS

The conditions for germinating rice (Oryza sativa L., variety: Labelle) seeds in the dark and preparation of mtDNA were the same as described previously[11]. To prepare ctDNA, rice seeds were grown in a growth chamber at 30°C under a 12 hour light/12 hour dark cycle. Ten-day-old seedlings were homogenized in a blender using the following homogenization buffer: 0.4 M sucrose, 30 mM Hepes (pH 7.2), 5 mM EDTA, 50 mM KH_2PO_4, 1 mM DTT, 0.1% bovine serum albumin. The homogenate was filtered through 3 layers of Miracloth and centrifuged at 170 x g to remove nuclei. The supernatant was centrifuged at 1,000 x g, and the chloroplast pellet was then resuspended in the homogenization buffer and layered on top of a stepwise 0.58 M/1.32 M/1.75 M sucrose gradient made in the above buffer. Centrifugation was performed at 15,000 x g for 1 hour (10K in a Beckman SW27 rotor). The chloroplast band at the interphase between 1.32 M and 1.75 M sucrose was collected and diluted with 3 volumes of the homogenization buffer minus bovine serum albumin and sorbitol. Chloroplasts were recovered by centrifugation at 1,000 x g for 10 minutes. The chloroplast pellet was lysed with 0.012% (w/v) proteinase K in 50 mM Tris-HCl (pH 8),

20 mM EDTA, 2% sarkosyl at 37°C for 1 hour. The lysate was extracted with phenol twice and the ctDNA was precipitated from the aqueous phase with the addition of 1/10 volume of 3 M sodium acetate and 2 volumes of ethanol. Construction of mitochondrial and chloroplast genomic libraries, restriction mapping and DNA sequence analysis were carried out as described previously[11].

A rapid method was developed by M. McCann for isolation of total rice DNA. Rice seeds were grown in soil in a greenhouse under a 16 hour light/8 hour dark cycle at 30°C. Individual 30- to 40-day-old plants (approximately 5 g plant tissue) were cut into small pieces and ground in a blender with liquid nitrogen. The pulverized rice tissue was transferred to a mortar and ground for 5 minutes in 10 ml of grinding buffer which was warmed to 65°C. (Grinding buffer is 4 parts homogenization buffer + 1 part phage lysis buffer and is prepared just prior to use. Homogenization buffer: 0.1 M NaCl, 0.2 M sucrose, 0.01 M EDTA, 0.03 M Tris-HCl, pH 8.0. Phage lysis buffer: 0.25 M EDTA, 2.5% (w/v) SDS, 1% sarkosyl, 0.5 M Tris-HCl, pH 9.2.) The homogenate was transferred to a centrifuge tube and incubated at 65°C for 30 minutes. One-third volume potassium acetate, pH 4.7, was added and the mixture incubated on ice for 30 minutes. The mixture was centrifuged 5 minutes at 1,600 x g to pellet the plant material. The supernatant was transferred to a new tube and centrifuged at 10,000 x g for 10 minutes. RNase A was added to a final concentration of 10 µg/ml. The tube was incubated at 37°C for 30 minutes. The mixture was then extracted once with phenol and the DNA precipitated from the aqueous phase with the addition of 2 volumes of ethanol. The DNA was pelleted by centrifugation at 12,000 x g for 30 minutes and resuspended in 0.5 ml of TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA). The resulting suspension was extracted once each with phenol and chloroform and the DNA precipitated with the addition of 1/10 volume of 3 M sodium acetate, pH 6.0, and 2 volumes of ethanol. After centrifugation for 15 minutes at 10,000 x g, the DNA pellet is washed twice with 70% ethanol, dried in a desiccator, and resuspended in 0.25-0.50 ml of TE buffer. The average yield is 1 mg of DNA per 5 g of rice tissue.

RESULTS

Identification of rbcL in the rice chloroplast genome and its homologous sequences in the rice mitochondrial genome

To identify ctDNA fragments containing rbcL and determine the possible presence of homologous sequences in the mitochondrial genome, we used

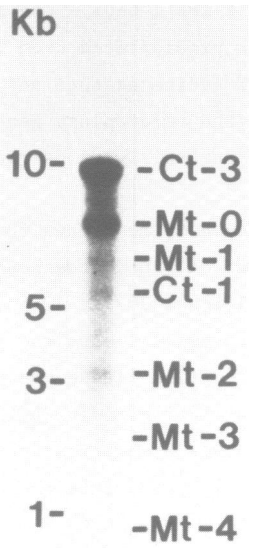


Fig. 1. Hybridization analysis of rice DNA. A mixture of ctDNA and mtDNA (5 μg) was digested with HindIII restriction enzyme, fractionated on a 0.9% agarose gel, and then transferred to a Nytran filter. The Nytran filter was hybridized with the nick-translated PstI-PstI fragment (561 bp) of rbcL of Ct-1. The same band pattern was obtained as when the filter was probed with spinach rbcL.

spinach rbcL[12] to probe a Southern blot of HindIII-digested rice DNA containing both ctDNA and mtDNA. About ten hybridizing bands were obtained, and seven of these were studied: 9.5 kb, 6.9 kb, 5.8 kb, 5.3 kb, 3.0 kb, 2.5 kb and 1.5 kb in length (Figure 1).

To determine the organelle origin of each band, we carried out a Southern blot experiment using a highly purified ctDNA (HindIII-restriction pattern is shown in Figure 2A). Using the rice rbcL gene as the probe, a similar multiple band pattern was observed (Figure 2B, Lane 1) as probing the mixture of ctDNA and mtDNA (Figure 1). To examine whether any band is due to the contamination of either mtDNA or nuclear DNA, we probed the same filter using a rice mitochondrial gene, COI (Moon et al., in preparation)

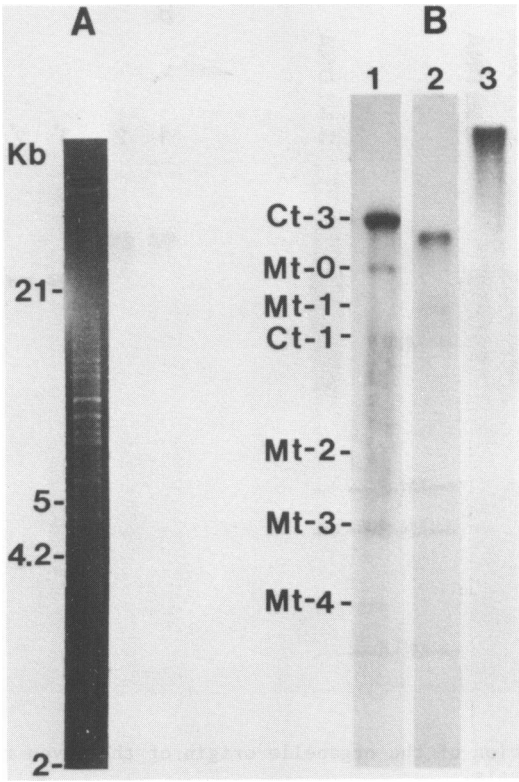


Fig. 2. Hybridization analysis of a highly purified ctDNA preparation.
A. HindIII restriction pattern of ctDNA (1 µg) visualized with ethidium bromide-staining after fractionation on a 0.9% agarose gel.
B. The same HindIII digested DNA shown in A. was transferred to a Nytran filter.
Lane 1: The Nytran filter was probed with the nick-translated PstI-PstI fragment of rbcL of Ct-1.
Lane 2: The rbcL probe was removed from the filter by boiling for 30 min in 0.01 x SSC, 0.01% SDS, then the filter was probed with the nick-translated rice mitochondrial gene, COI.
Lane 3: The filter was boiled again under the same conditions, and then probed with the repetitive rice nuclear sequence, pOS-N1.
The filter represented by Lane 1 was exposed for 5 h, that of Lane 2 for 20 h, and that of Lane 3 for 50 h.

(Figure 2B, Lane 2), and a middle repetitive nuclear sequence (pOS-N1) from rice (Wu et al., in preparation) (Figure 2B, Lane 3). From the intensity of the hybridizing bands (after tracing with a densitometer), it was estimated that our ctDNA preparation was contaminated with about 5% of mtDNA and 1% of nuclear DNA. Thus, even when the ethidium bromide-stained

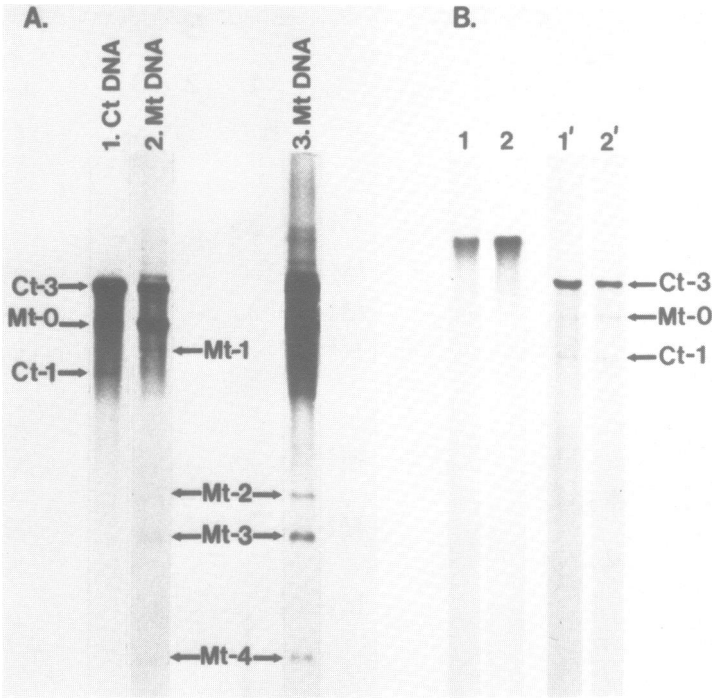


Fig. 3. Determination of the organelle origin of the seven fragments by hybridization analysis.

A. Equal amounts of HindIII-digested mitochondrial and chloroplast DNA were separately fractionated on a 0.9% agarose gel, and then transferred to a Nytran filter. The filter was probed with the PstI-PstI fragment (561 bp) of rbcL of Ct-1. Lane 3 was the same as Lane 2 except that the X-ray film was exposed 5X longer to better visualize Mt-2, Mt-3 and Mt-5 bands.

B. Approximately equal amounts of HindIII-digested highly purified ctDNA (Lanes 1 and 1') and total DNA (Lanes 2 and 2') were separately fractionated on a 0.9% agarose gel, and then transferred to a Nytran filter.

Lanes 1 and 2: The filter was hybridized to the nick-translated pOS-N1. Lane 1 was exposed 10 times longer than Lane 2.

Lanes 1' and 2': The filter was boiled and probed with the nick-translated PstI-PstI fragment of rice rbcL. The exposure time was the same for Lanes 1' and 2'.

restriction pattern looked reasonably clean, the ctDNA was not entirely free of any cross-contamination from mtDNA and nuclear DNA. As a result, it was difficult to determine the organelle origin of each fragment using only ctDNA preparation. To circumvent this problem, we carried out a Southern blot analysis using both ctDNA and mtDNA preparations. To begin

with, we determined the extent of cross-contamination of ctDNA and mtDNA in each preparation using chloroplast genes, psaA1 (coding for the 67 kd p700 chlorophyll a apoprotein of photosystem I) and petA (coding for cytochrome f) (Côtè, J.-C.; Wu, N.H., unpublished results), and two rice mitochondrial genes, COII[11] and COI (Moon et al., in preparation). We found that hybridization of the mitochondrial probe with the mtDNA was 6 times more intense than with the ctDNA (data not shown). These results indicated that this mtDNA sample was also relatively pure, but contained 16% of ctDNA contamination. Using a set of ctDNA and mtDNA of known purity, we compared the intensity of each fragment hybridized to the rice rbcL probe (Figure 3A). The 9.5 kb (Ct-3) and 5.3 kb (Ct-1) bands were more intense in the ctDNA preparation, whereas the 6.9 kb (Mt-0), 5.8 kb (Mt-1), 3.0 kb (Mt-2), 2.5 kb (Mt-3) and 1.5 kb (Mt-4) were more intense in the mtDNA preparation. Thus, we assumed the 9.5 kb and 5.3 kb bands to be of chloroplast origin and the other bands to be of mitochondrial origin.

To rule out the possibility of the nuclear origin of the minor copy, Ct-1, we carried out the Southern analysis of a set of a ctDNA preparation and a total DNA preparation with the different amounts of nuclear DNA contamination. The extent of nuclear DNA contamination of two preparations was estimated by the intensity of hybridization to the rice nuclear repetitive sequence, pOS-N1. The ctDNA preparation contains about 1% nuclear DNA contamination, while the total DNA contains over 50% nuclear DNA. When the same filter was reused for hybridization with the rice rbcL probe, the relative intensity of the two bands, Ct-3 and Ct-1, remained constant at 10 to 1. These results confirm that Ct-3 and Ct-1 are both of chloroplast origin.

The finding that two different ctDNA fragments contain the same rbcL raises the question of whether these two DNA fragments are located in the same or different chloroplast genomes. We observed that the intensity of the hybridizing band of the 5.3 kb fragment was only one-tenth that of the 9.5 kb fragment, indicating that these two fragments are not present in equal copy number. The simplest interpretation therefore is that rice chloroplasts contain two populations of genomic DNA molecules: a major one containing the 9.5 kb fragment and a minor one containing the 5.3 kb fragment. This interpretation implies that the rice chloroplast genome is heterogeneous. Since ctDNA was isolated from a population of several hundred rice seedlings, the heterogeneity we observed could occur either within each individual plant or within a given population of plants. To

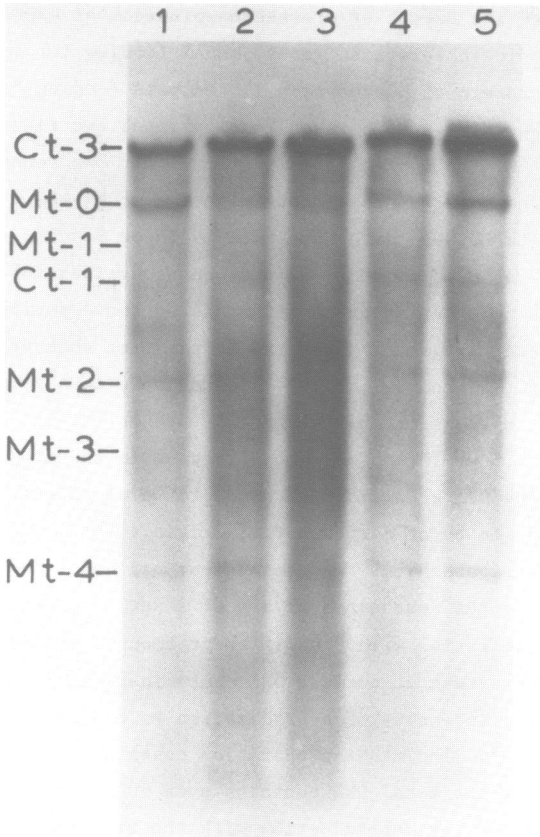


Fig. 4. Hybridization analysis of rice DNA from a population (Lane 1) and four individual plants (Lanes 2, 3, 4 and 5). DNA prepared by the rapid method was digested with HindIII restriction enzyme, fractionated on a 0.9% agarose gel, and then transferred to a Nytran filter. The Nytran filter was hybridized with the nick-translated PstI-PstI fragment (561 bp) of rbcL of Ct-1.

distinguish between these two possibilities, we performed Southern blot analysis on total DNA isolated from four individual plants. The results (Figure 4, Lanes 2-5) show that two bands, corresponding to Ct-1 and Ct-3, appear for each plant. The relative intensity of the two bands in each plant is approximately the same as was found in the entire population (Figure 4, Lane 1).

Isolation and characterization of ctDNA and mtDNA fragments containing rbcL or its homologs

Chloroplast and mitochondrial genomic libraries were constructed in

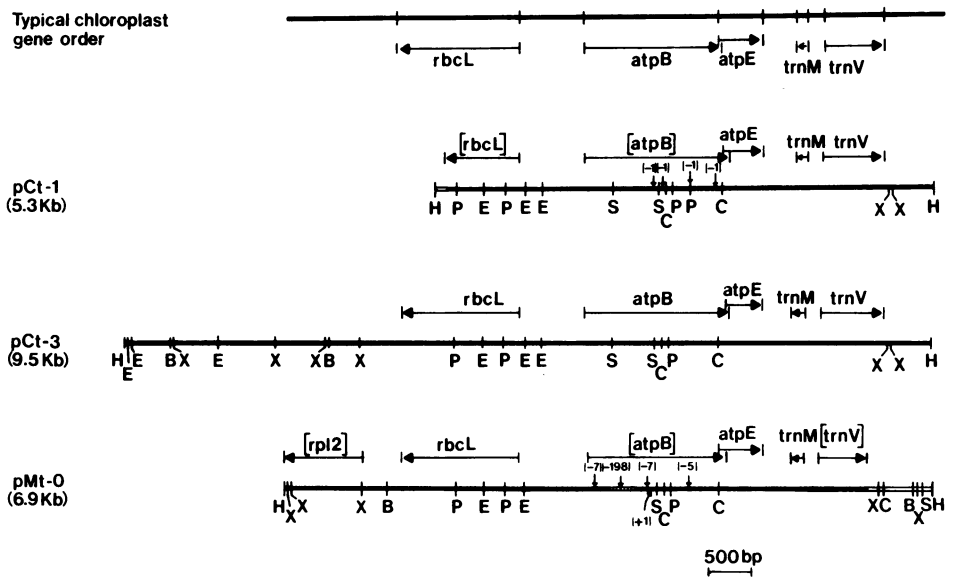


Fig. 5. Restriction maps and gene locations for Ct-1, Ct-3, Mt-0 clones. Restriction maps were determined by gel electrophoretic analysis of several single or double restriction enzyme digestions of each clone. Southern transfers of these digests were hybridized with nick-translated spinach rbcL and wheat atpB to locate the corresponding genes. atpE, trnV and trnM were identified using sequencing data. rpl2 was identified by searching against the MicroGenie data bank for homology. Numbers above arrows indicate number of bases inserted or deleted. Brackets indicate non-functional genes. The open box at the left end of Ct-1 represents the 156 bp sequence which is not found in Ct-3 or Mt-0. The open box at the right end of Mt-0 represents the 930 bp sequence which is not present in Ct-1 or Ct-3. Abbreviations: B, BglII; C, ClaI; E, EcoRI; H, HindIII; P, PstI; S, SstI; X, XbaI.

plasmid vectors pBR322 or pUC13, and clones containing rbcL or its homologous sequences were identified by colony hybridization using spinach rbcL as a probe. Clones containing each of the seven HindIII fragments have been isolated and are designated pCt-3 (9.5 kb), pMt-0 (6.9 kb), pMt-1 (5.8 kb), pCt-1 (5.3 kb), pMt-2 (3.0 kb), pMt-3 (2.5 kb) and pMt-4 (1.5 kb). We have characterized all these clones and found pMt-0, despite its mitochondrial origin, to be very similar in gene organization and nucleotide sequence to the two chloroplast clones, pCt-1 and pCt-3. Therefore, we report the results of our studies of these three clones in this communication. The other four mitochondrial clones show much less sequence homology to the chloroplast clones, and probably represent chloroplast

sequences which were transferred to the mitochondrial genome much earlier during the process of evolution than pMt-0. Findings on these four mitochondrial clones will be reported elsewhere.

The HindIII insert fragments in pCt-1, pCt-3 and pMt-0 were subjected to restriction mapping with results shown in Figure 5. At the top is the typical organization and direction of transcription of chloroplast genes of higher plants in the vicinity of rbcL[1]. Various restriction fragments of the three HindIII fragments were separately probed with atpB and atpE of wheat[13] to determine the locations of these two genes. In addition, sequence analysis revealed the presence of the genes coding for $tRNA^{met}$ (trnM) and $tRNA^{val}$ (trnV) in all three HindIII fragments. In the 6.9 kb mtDNA fragment, a sequence homologous to the chloroplast gene encoding the ribosomal protein L2 (rpl2) was also identified immediately to the left of rbcL. In the chloroplast genome, two rpl2 genes are located much farther from rbcL, 20 kb and 50 kb, in the two inverted repeats. In summary, all three HindIII fragments contain an rbcL-atpB-atpE-trnM-trnV gene cluster arranged in similar order as has been found in other plant species. However, the 6.9 kb mtDNA fragment also contains an rpl2 homolog.

When we further compared restriction sites among the three clones, we found all the sites of the 5.3 kb ctDNA fragment, except its left-hand HindIII site, to be present in the 9.5 kb ctDNA fragment. The left-hand HindIII site of the 9.5 kb fragment occurs 4.2 kb beyond that of the 5.3 kb fragment. The 6.9 kb mtDNA fragment contains many restriction sites similar to those in the other two ctDNA fragments in the region between rbcL and trnV. However, the restriction sites in the region to the left of rbcL and to the right of trnV are completely different.

Nucleotide sequence analysis of pCt-1, pCt-3 and pMt-0

The results of restriction mapping indicate that the two chloroplast clones, pCt-1 and pCt-3, may contain very similar sequences and that the mitochondrial clone, pMt-0, also resembles them. In order to analyze in more detail the degree of sequence similarity among these three clones, we have determined their DNA sequences.

Appropriate restriction fragments of each clone were subcloned into M13 mp11 and mp10 vectors, and their nucleotide sequences were determined[14]. To date, the major portion of each of the three HindIII fragments has been sequenced. The nucleotide sequences were then compared across the three clones. For the sake of simplicity, we shall refer hereafter to both the

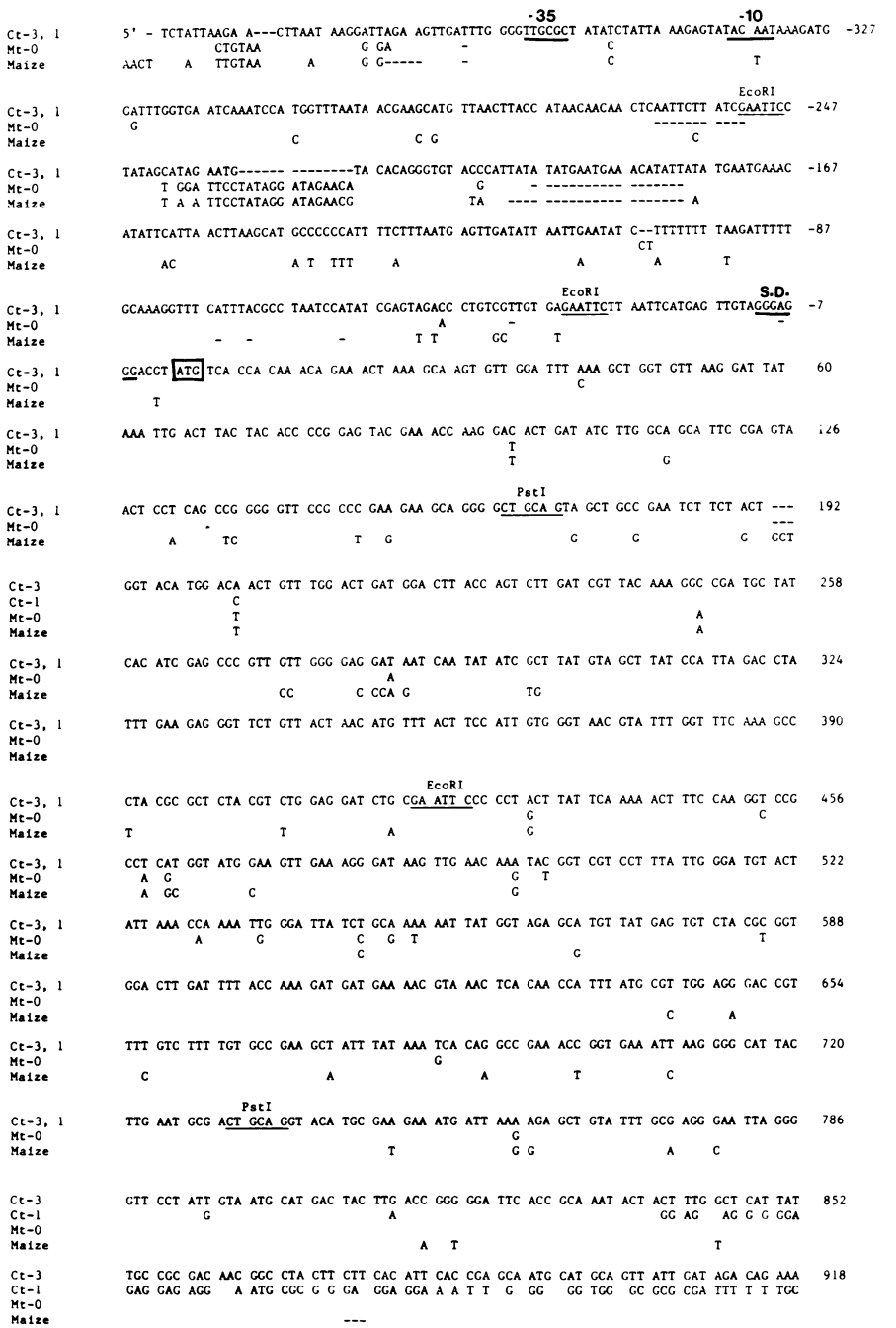
chloroplast genes and their mitochondrial homologs by the name of the gene itself.

Sequence comparison of rbcL from three rice clones

RbcL has been sequenced from a variety of plant species[12,15-17]. Here, we chose maize rbcL[15] to align with the three rice rbcL sequences, since maize and rice are evolutionarily closely related. This alignment led to assignment of the coding, 5'- and 3'-noncoding regions of the three rice sequences as shown in Figure 6.

The coding regions of rbcL in pCt-1 and pCt-3 are extremely homologous within 834 bp from the translation initiation codon ATG, a sequence coding for 278 amino acids. Only 3 bp are different and all are in the third position of their codon resulting in no change in the amino acid residue. However, the nucleotide sequence beyond the 278th codon and continuing to the left-hand HindIII site of pCt-1, a total of 156 bp (indicated by an open box in Figure 5), shows no homology at all with the corresponding region in pCt-3. When comparing these divergent sequences with the maize rbcL sequence, we found that pCt-3 resembles maize, while pCt-1 does not. The 156 bp sequence of pCt-1 contains an open reading frame of 34 amino acids continuing after the preceding 278 amino acid sequence before a stop codon is encountered. Thus, the entire coding region of rbcL in pCt-3 codes for a complete large subunit of ribulose-1,5-bisphosphate carboxylase (Rubisco) of 477 amino acids, while that of pCt-1 appears to be truncated and codes for only 278 amino acids of the large subunit of Rubisco, plus 34 amino acids of unknown origin. On the contrary, rbcL of pMt-0, a chloroplast gene acquired by the mitochondrial genome, still retains a perfect coding sequence with the coding capacity for 475 amino acids of the large subunit of Rubisco. This sequence exhibits very high homology with that of pCt-3. Except for 12 bp gaps required for alignment to maximize the homology of these two sequences, there are only 37 bp differences out of 1422 bp compared; 19 are in the third position of the codon, resulting in only four replacement substitutions in the amino acid sequence, and nine each are in the first and second positions of the codon, resulting in 18 replacement substitutions. The overall sequence homology between the chloroplast rbcL (pCt-3) and the transferred version in the mitochondria (pMt-0) is 97.4%, which is higher than the homology between chloroplast rbcL of rice and maize, 93.3%.

In all plant species studied, rbcL and atpB are transcribed in opposite


-35
-10
EcoRI
EcoRI
S.D.
PstI
EcoRI
PstI
Ct-3, 1
Mt-0
Maize
Ct-3
Ct-1
Mt-0
Maize

```
Ct-3      AAT CAT GGT ATG CAT TTC CGT GTA TTA GCT AAA GCA TTG CGT ATG TCT GGG GGA GAT CAT ATC CAC   984
Ct-1      CAT GGA ATT GCT TTT CG  T A TGG  GC T   TT  CTC GCC ACC  GC AAA AAA CAC A   TT  CAT T G
Mt-0
Maize                                                                 G

Ct-3      GCT GGT GCA GTA GTA GGT AAG TTA GAA GGG GAA CGC GAA ATG ACT TTA GGT TTT GTT GAT TTA TTG  1050
Ct-1      CGG AAG CTT - 3'
Mt-0
Maize     T C

Ct-3      CGC GAT GAT TTT ATT GAA AAA GAT CGT --- GCT CGC GGT ATC TTT TTC ACT CAG GAC TGG GTA TCC  1116
Mt-0                                    TCG TTG TTC GCG  T
Maize                                           T

Ct-3      ATG CCA GGT GTT ATA CCG GTG GCT TCA GGG GGT ATT CAT GTT TGG CAT ATG CCA GCT CTG ACC GAA  1182
Mt-0                        C   C                                       --- --- ---
Maize                                 T

Ct-3      ATC TTT GGA GAT GAT TCT GTA TTG CAA TTT GGT GGA GGA ACT TTA GGA CAT CCT TGG GGT AAT GCA  1248
Mt-0          C     G
Maize         C                 A       A                                           A

Ct-3      CCT GGT GCA GCA GCT AAT CGG GTG GCT TTA GAA GCC TGT GTA CAA GCT CGT AAC GAA GGG CGC GAT  1314
Mt-0
Maize      A                            A

Ct-3      CTT GCT CGT GAA GGT AAT GAA ATT ATC CGA TCA GCT TGC AAA TGG AGT CCT GAA CTA GCC GCA GCT  1380
Mt-0                                          G
Maize       A       C   G  TA --- C           AA  G                     G C

Ct-3      TGT GAA ATA TGG AAA GCG ATC AAA TTC GAG TTC GAG CCG GTA GAT AAA CTA GAT ACC TAG ACTAAGTGGA  1450
Mt-0            G         G                                                          ----------
Maize       C               G  A                T GGT TTC AAA  CG ATG G T ACC ATA ---   A  A   AAAA

Ct-3      TAAAATTAGA TAGAAAAAAG GTCTAAATAA AAAAGAAGAG AAATAGAAAG ATCAAAAATC AGTTACGAAA ATGCAGTAAT      1530
Mt-0      - T  G      A                              C                              -
Maize     A GC AA T   GA GTG  A AAA   G T  TG  ATGA A TG A TG C  TAATTCTT A TTCCT T  T TGATT C

                                                                                          BglII
Ct-3      TCTTCTTTTT TCTTCTAATT GATTGCAATT AAACTCGGCT CAATCT--GA AAAAAGATTG AGCCGAGTTT AAATAGATTT - 3'  1610
Mt-0                                                      TTTT TTTC                        C  - 3'
Maize       AAT CGGC   A   TT C T AAAA  AA    GA T AG  CGAAAA-          CT - 3'
```

Fig. 6. Nucleotide sequences of rbcL from three rice clones and a maize clone. Sequences were aligned to maximize homology. Only those nucleotides which differ from the Ct-3 sequence are shown for Ct-1, Mt-0 and maize. Dashes indicate gaps that are necessary to align sequences. The number at the right of each row indicates the position of the last base in that row from the first base of the initiation codon. The ATG sequence enclosed in the box is the initiation codon of rbcL. The Shine-Dalgarno sequence is marked as S.D., and the promoter sequence is marked as -35 and -10 on top of the sequences. Restriction sites are shown above the nucleotide sequences.

directions, and there is an intergenic space between their coding regions of approximately 800 bp. Promoter sequences ("-10" and "-35" sequences) for both rbcL and atpB have been identified in this region[18]. We have not yet determined the 5'-terminus of the rice rbcL transcript. However, because of the high degree of homology between rice and maize rbcL in the 5'-noncoding sequence, we were able to identify the "-35" sequence, 5'-TTGCGC-3', and "-10" sequence, 5'-TACAAT-3', which are present in all three rice rbcL in perfect conformance with those found in rbcL genes of all other plants. A 7 bp sequence 5'-GGGAGGG-3' located immediately upstream from the translation initiation codon ATG of all rbcL genes has been proposed to be the ribosome binding site[4] similar to the "Shine-

Dalgarno" sequence found in the prokaryotic mRNA[19]. The two rice chloroplast rbcL do contain this exact sequence, while the mitochondrial copy is missing the A and contains only 6 G's. For the purpose of comparing the degree of overall sequence homology in the 5'-noncoding region of the three rbcL and maize rbcL, we arbitrarily set the sequence alignment beginning at 40 bp upstream from the "-35" sequence and began our analysis from that point. The two rice chloroplast rbcL in pCt-1 and pCt-3 contain completely identical sequences in this region. Their sequences are 83% homologous to the 5'-noncoding sequence of rbcL in pMt-0, and 78% to that of maize. Surprisingly, the 5'-noncoding sequence of pMt-0 rbcL is actually more similar to maize (85% homology) than the two rice chloroplast rbcL are.

Nucleotide sequence comparison of the region between rbcL and the right-hand HindIII site of three rice clones

The region between rbcL and the right-hand HindIII site has been completely sequenced for pCt-1 and pCt-3, and 95% sequenced for pMt-0 (sequences not shown, will be reported elsewhere). In all three rice clones, we have identified the coding sequences of atpB, atpE, trnM and trnV sequence alignment with the corresponding maize or barley genes (20,21). In this entire region the two chloroplast clones differ by only a single base pair substitution in the 5'-noncoding region of atpB and four single base pair deletions in the coding region of atpB in pCt-1 (Figure 5). The deletions result in premature termination in the deduced protein sequence, yielding an incomplete protein of 246 amino acids. On the other hand, the atpB coding region of pCt-3 is perfect, and can code for a polypeptide of 498 amino acids. The single base pair substitution occurs right in the "-35" region of the consensus promoter sequence for atpB. pCt-3 contains the "-35" consensus sequence, 5'-TTGACA-3', while pCt-1 has a variant "-35" sequence, 5'-TTGGCA-3', due to the substitution of a G for the first A. Both pCt-1 and pCt-3 contain an identical "-10" sequence, 5'-TAGTAT-3', which is also found in atpB of maize[20]. The coding region of atpB in pMt-0, unlike its rbcL, is grossly defective, with one 198 bp deletion, two 7 bp deletions, one 5 bp deletion and one single base-pair insertion (Figure 5). Excluding these deletions and insertion, there are 19 bp differences in the atpB coding region between pMt-0 and pCt-3; 11 are in the third position of the codon resulting in only one replacement substitution in the amino acid sequence, and four each are in the first and second positions of the codon resulting in six replacement substitution. Notwithstanding the numerous deletions in the coding sequence, pMt-0

surprisingly contains perfectly normal "-10" and "-35" promoter sequences. The entire 5'-noncoding sequence also shows a very high degree of homology to that of pCt-3 and pCt-1, 91.3%. However, a 5 bp sequence, 5'-GTGAT-3', which has been proposed to be the ribosome binding site[20] is present in pCt-3 and pCt-1, but absent in pMt-0.

A 6 bp sequence, 5'-AAATGA-3', consisting of the last codon and the translation termination codon of atpB overlaps with the translation initiation codon ATG of atpE. This overlapping of atpB and atpE has also been found in other plants[4]. The atpE in all three rice clones contains an open reading frame of 137 amino acids. The coding sequences of atpE in pCt-3 and pCt-1 are identical, and pMt-0 differs from them by only 3 bp, two in the third position and the other in the first position of the codon. As a result, there is one replacement substitution in the amino acid sequence.

The trnM gene is located downstream from atpE in all three rice clones, and is transcribed in the opposite direction to atpB and atpE as is true in all other plants. In both pCt-1 and pCt-3, the last base pair of trnM is located at the 112th bp from the termination codon of atpE, while in pMt-0, it is located at the 104th bp. The intergenic sequences between atpE and trnM in pMt-0 and pCt-3 or pCt-1 are 87.4% homologous. The coding region of trnM is identical in all three clones.

The trnV gene in pCt-1 is interrupted by an intervening sequence of 597 bp which shows 95.7% sequence homology with that of barley. The coding sequence of the two trnV genes are completely identical. The corresponding region in pCt-3 has only been partially sequenced (45% complete).

The restriction map of pMt-0 in this region, particularly toward the right-hand HindIII site, is very different from that of pCt-1 or pCt-3. Therefore, we have focused our sequencing effort on this segment. An approximately 1.1 kb sequence beginning with the right-hand HindIII site has been determined. Most of this sequence, 930 bp, from HindIII site to 278 bp past the second XbaI site, indicated by an open box in Figure 5, shows no homology with the sequence of either pCt-1 or pCT-3. The rest of this 1.1 kb sequence, from the 279th bp past the XbaI site and continuing to the left for 130 bp, is 90% homologous to a region in the trnV intron of pCt-1. Presumably, if we continued our sequencing of this region in pMt-0, we would find it to be homologous to the rest of the intron upstream from this point, and to the first exon of the trnV gene. The last base pair before the point of divergence between pMt-0 and pCt-1 (the 279th bp past the XbaI site) is the 416th bp of the intron, and it most likely defines

one terminus of the chloroplast segment which was transferred to the mitochondrial genome.

Sequence comparison of regions downstream from rbcL of two rice clones, pCt-3 and pMt-0

The region downstream from the translation termination codon of rbcL to the left-hand HindIII site has been almost completely sequenced for both pCt-3 and pMt-0. The first 430 bp immediately downstream from the termination codon of rbcL in both clones are very similar. However, all homology ends beyond the 24th bp to the left-hand side of the first XbaI site, the end of the 430 bp sequence. In order to determine whether this point of sequence divergence might represent the other terminus of the transferred chloroplast DNA fragment, we have searched all published chloroplast and mitochondrial sequences included in the GeneBank for homology with the sequence of this region using the Microgenie sequence analysis program. Unexpectedly, we found the sequence of pMt-0 in this region to be 88.7% homologous to the tobacco chloroplast gene coding for the ribosomal protein L2 (rpl2)[17]. This homology extends all the way to the left-hand HindIII site. These and other results on rpl2 will be reported elsewhere.

DISCUSSION

Rice ctDNA population is heterogeneous

The conclusion that Ct-1 is located on the chloroplast genome but not from the nucleus is supported by the fact that: (a) chloroplast DNA is present in several thousand copies in each rice cell, and nuclear DNA only in one copy for the single copy genes; (b) from other published work, when chloroplast-like sequences were found in the nucleus, they were present in low copies[22]; thus, the copy number of the Ct-1 (present in several hundred copies per cell) is still too high to be accounted for by nuclear DNA contamination in the ctDNA preparation; (c) from the constant ratio of hybridization intensity of bands corresponding to Ct-3/Ct-1, we found that the ratio of Ct-3/Ct-1 to be close to 10 in all cases: in highly purified ctDNA (containing around 1% nuclear DNA) and in total rice DNA preparation (containing over 50% nuclear DNA). Thus, we conclude that Ct-1 was of chloroplast origin.

The fact that the rice chloroplast gene cluster rbcL-atpB-atpE-trnM-trnV was found in two different HindIII fragments strongly suggests the existence of a heterogeneous population of rice ctDNA molecules. An alternative explanation which would support the general assumption that

ctDNA is homogeneous might be that at some point during evolution this gene cluster was duplicated, and both copies are thus present in the same ctDNA molecule. However, we can rule out this explanation on the basis that the copy numbers of these two HindIII fragments are unequal.

CtDNA heterogeneity might theoretically occur either within each individual plant or within a given population. However, DNA sequence analysis reveals that only the 9.5 kb HindIII fragment contains all potentially functional genes, while the 5.3 kb fragment has defects in two essential genes, rbcL and atpB. Therefore, if the latter case were true, that is, heterogeneity occurs within the population, then one would expect that the plants of one subpopulation would have defective chloroplast genomes. We did not consider this likely and favored the former interpretation that rice ctDNA is heterogeneous within each individual plant. This has been shown to be the case by running Southern blot analysis on total DNA of individual plants. Results in Figure 1 indicate that each plant contained the 9.5 kb and the 5.3 kb fragments in comparable intensity as in the mixed population.

The existence of a heterogeneous population of mtDNA molecules within the individual has been well documented in animals (for review, see 23). In the case of Holstein cows, a population of mtDNA molecules with varying lengths of the D-loop cytosine homopolymer has been found within the tissues of each individual[24]. However, to the best of our knowledge, no heterogeneity in the chloroplast genome has been observed in any of the higher plant species studied so far. We do not know whether what we have found in rice is unique among plants or not.

pCt-1 has been entirely sequenced, and its overall homology with the corresponding sequenced region in pCt-3 (5.1 kb), except for the 156 bp sequence near one end of pCt-1, is 98% (9 bp difference over 4.6 kb compared). We believe that the 9.5 kb HindIII fragment represents the original functional copy of the rbcL gene cluster. We speculate that the later generation of the 5.3 kb HindIII fragment might be the result of insertion of a DNA fragment into a small fraction of ctDNA molecules containing the 9.5 kb fragment. This insertion sequence is unrelated to the cluster of the five genes in pCt-3, is of undetermined length, and contains a HindIII restriction site 156 bp from one of its terminus.

This interpretation poses a question as to how the 5.3 kb-containing ctDNA molecules can be maintained along with the 9.5 kb-containing molecules in view of the fact that two of its genes are defective. Perhaps the

Table 1

NUCLEOTIDE SEQUENCE COMPARISONS (% HOMOLOGY) BETWEEN RICE pCt-3 CLONE AND OTHER SEQUENCES

GENE	rbcL			atpB		atpE		trnM	trnV
	5' FLANKING	CODING	3' FLANKING	5' FLANKING	CODING	CODING	3' FLANKING		
# OF B.P. COMPARED	407	1422	170	397	1496	414	112	73	873
RICE pCt-1	100	99.6[a]	-	99.7	99.7	100	100	100	100
RICE pMt-0	83	97.4	89.3	91.3	85.0	99.3	87.4	100	90[b]
MAIZE	78	93.3	47.6	88.9	95.7	96.6	91.1	100[c]	95.7[c]

[a] The total sequence is only 834 bp long, due to truncation.
[b] Only 750 bp of the total have been sequenced.
[c] Barley sequence was used for comparison since maize sequence is not available.

lack of frequent recombination between ctDNA molecules has allowed the two populations to co-exist without converting the defective molecules through intermolecular recombination with the functional molecules. Another possible explanation might be that there are other mutations somewhere else in the 9.5 kb-containing chloroplast molecule, whereas the corresponding region in the 5.3 kb-containing molecule is normal. Thus, both populations would be required to provide all the necessary functional genes. On the other hand, the fact that the copy number of the 5.3 kb fragment is only 1/10 that of the 9.5 kb fragment may suggest that this minor population of ctDNA molecules may already be in the process of being slowly selected against, and may ultimately be eliminated altogether.

The 6.9 kb mtDNA fragment contains a transferred chloroplast gene cluster

We present in this report another example of ctDNA transfer into the mitochondrial genome, reinforcing the earlier conclusion that interorganellar DNA transfer is a general phenomenon in plants[10]. Previous studies focused mainly on comparison of restriction maps of ctDNA fragments with their homologous sequences in the mitochondrial genome. At this level, it is impossible to reveal the fine differences between them[7,9,10]. We are able to examine the differences in their sequences in much greater detail through extensive sequence analysis of a rice ctDNA fragment and its transferred copy in the mitochondria. A summary of the nucleotide sequence comparisons between the five genes of pCt-3 and those of the transferred fragment in pMt-0, as well as those of pCt-1 and of maize or barley, is given in Table 1.

One of the interesting results of our sequence analysis is the finding

that there is a very high degree of sequence conservation between the ctDNA fragment containing the gene cluster rbcL-atpB-atpE-trnM-trnV and its transferred copy in the mitochondrial genome. The entire sequence transferred from chloroplasts to mitochondria is a minimum of 6.0 kb in length, but the exact size is not known since only one terminus falls within the 6.9 kb mitochondrial fragment isolated in this study. Of the 5.5 kb of the transferred segment we have sequenced so far, 4.6 kb was found to be 91.2% homologous to the original chloroplast gene cluster rbcL-atpB-atpE-trnM-trnV, and the rest of the sequence, 0.9 kb, showed 88.7% homology with the chloroplast rpl2 gene of spinach (rice rpl2 has not yet been sequenced). Perhaps even more striking is the finding that three of the six transferred chloroplast genes, rbcL, atpE and trnM, still retain perfectly normal open reading frames in their coding sequences, despite the occurrence of a number of nucleotide substitutions between them and the chloroplast genes (data not shown). The nucleotide substitutions also occur most frequently in the third position of the codon, and many of them result in synonymous substitutions in amino acid residues.

Two factors might explain the high degree of nucleotide sequence conservation which exists between the chloroplast DNA and its transferred copy in the mitochondria. In the first place, interorganellar DNA transfer is an ongoing process, occurring throughout the course of evolution. This particular fragment was probably transferred from the chloroplast to the mitochondrial genome relatively recently, and, because of the slow rate of sequence evolution in the mitochondria, has not changed significantly. Another factor might be the action of correction processes in mitochondria which tends to return mutated sequences back to the original ones through various stochastic recombination processes as suggested by Lonsdale et al.[9].

Our data as of now does not definitively prove which factor takes precedence; however, we favor the first explanation for the following reason. We have also isolated five other mitochondrial clones which all contain transferred chloroplast rbcL and atpB genes (Moon, unpublished results). Their sequences are not only much less homologous to the chloroplast genes than is the mitochondrial clone reported here (73% vs. 91%) but also contain numerous deletions and frame-shift mutations in the coding regions of both rbcL and atpB. In addition, the rate of nucleotide substitution is almost equal at all three positions of the codon. These results suggest that it is not necessary that the transferred chloroplast genes be

corrected if mutated, as the second explanation implies. They also imply that these transferred sequences have evolved under virtually no functional constraints. We believe that the difference between these other clones and the one we are reporting (pMt-0) lies in the time at which the transfer took place.

ACKNOWLEDGEMENTS

We thank Chris Howe for the wheat atpB and atpE clones, and Tom Fox and Margaret McCann for helpful comments. This work was supported by research grants RF84066, Allocation No. 3 from the Rockefeller Foundation, and GM29179 from the National Institutes of Health, U.S. Public Health Service.

*To whom reprint requests should be sent

[1]Present address: Department of Molecular and Cell Biology, 403 Althouse Laboratory, The Pennsylvania State University, University Park, PA 16802, USA

REFERENCES

1. Palmer, J.D. (1985) In MacIntyre, R.J. (ed.), Molecular Evolutionary Genetics, Plenum Press, New York, pp. 373-402.
2. Birky, C.W., Jr. (1978) Annu. Rev. Genet. 12, 471-512.
3. Birky, C.W., Jr. (1983) Science 222, 468-475.
4. Whitfeld, P.R. and Bottomley, W. (1983) Annu. Rev. Plant Physiol. 34, 279-310.
5. Leaver, C.J. and Gray, M.W. (1984) Annu. Rev. Plant Physiol. 33, 373-402.
6. Lonsdale, D.M. (1984) Plant Mol. Biol. 3, 201-206.
7. Stern, D.B. and Lonsdale, D.M. (1982) Nature 299, 698-702.
8. Ellis, J. (1982) Nature 299, 678-679.
9. Lonsdale, D.M., Hodge, T.P., Howe, C.J. and Stern, D.B. (1983) Cell 34, 1007-1014.
10. Stern, D.B. and Palmer, J.D. (1984) Proc. Natl. Acad. Sci. USA 81, 1946-1950.
11. Kao, T.H., Moon, E. and Wu, R. (1984) Nucleic Acid. Res. 12, 7305-7315.
12. Zurawski, G., Perrot, B., Bottomley, W. and Whitfeld, P.R. (1981) Nucleic Acids Res. 9, 3251-3270.
13. Howe, C.J., Fearnley, I.M., Walker, J.E., Dyer, T.A. and Gray, J.C. (1985) Plant Mol. Biol. 4, 333-345.
14. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
15. McIntosh, L., Poulsen, C. and Bogorad, L. (1980) Nature, 288, 556-560.
16. Shinozaki, K. and Sugiura, M. (1982) Gene 20, 91-102.
17. Zurawski, G., Bottomley, W. and Whitfeld, P.R. (1984) Nucleic Acids Res. 12, 6547-6558.
18. Mullet, J.E., Orozco, E.M., Jr. and Chua, N.H. (1985) Plant Mol. Biol. 4, 38-54.
19. Shine, J. and Dalgarno, L. (1975) Nature 254, 34-38.
20. Krebbers, E.T., Larrinua, I.M., McIntosh, L. and Bogorad, L. (1982) Nucleic Acids Res. 10, 4985-5002.
21. Zurawski, G. and Clegg, M.T. (1984) Nucleic Acids. Res. 12, 2549-2559.
22. Timmis, J.N. and Scott, N. (1983) Nature (London) 305, 65-67.
23. Brown, W.H. (1985) In MacIntyre, R.J. (ed.), Molecular Evolutionary Genetics, Plenum Press, New York, pp. 95-130.
24. Hauswirth, W.W., Walle, M.J.V.D., Laipis, P.J. and Olivo, P.D. (1984) Cell 37, 1001-1007.