J. nat. Rubb. Res., **8(2)**, 146 – 153

Nucleotide Sequence of the Chloroplast Gene for the Large Sub-unit of Ribulose-1,5-bisphosphate Carboxylase/Oxygenase from Hevea brasiliensis

CHEONG KAY FONG* AND KOH CHONG LEK**

A shotgun genomic library of the chloroplast DNA (ctDNA) of Hevea brasiliensis *(clone RRIM 600) was constructed with the plasmid cloning vector pUC19 in* Escherichia coli. *The gene,* rbcL, *for the large sub-unit (LS) of ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) was isolated by first screening the library by colony hybridisation and then confirmation by Southern hybridisation with a radiolabelled DNA probe that contains the internal sequence of* rbcL *from tobacco,* Nicotiana tabacum. *Sequence data indicated that the* Hevea rbcL *gene contains an open reading frame of 1425 nucleotides encoding a protein of 475 amino acid residues. Conserved sequences in the gene and LS protein were detected. Palindromic sequences which may form intrastrand stem-loop structures for the termination of transcription were encountered at the 3' non-coding region of the gene.*

In plants, ribulose-1,5-bisphosphate carboxylase/oxygenase (Fraction I protein; Rubisco; EC4.1.1.39) is an important bifunctional enzyme located in the chloroplast stroma. This enzyme, through its carboxylase activity, is responsible for the cardinal step in photosynthetic CO_2 fixation: it also catalyses the process of photorespiration with its oxygenase activity[1].

Rubisco is a holoenzyme, consisting of eight large sub-unit (LS) polypeptides (M_r 53 000) and a similar number of small sub-unit polypeptides (M_r 12 000–14 000)[2]. The LS protein is known to contain the active sites of the enzyme. In higher plants, the LS protein gene (*rbcL*) is located in the chloroplast genome[3], whereas the small sub-unit protein gene (*rbcS*) is in the nuclear genome[4]. Therefore, the synthesis of functional Rubisco enzyme clearly depends on the expression of two distinct compartmentalised genetic systems.

Complete nucleotide sequences of *rbcL* have been determined from plants of diverse taxonomic groups: the algae such as *Chlorella ellipsoidea*[5], *Astasia longa*[6], *Olisthodiscus luteus*[7], and *Bryopsis maxima*[8]; the bryophyte *Marchantia polymorpha*[9]; the pteridophyte *Angioteris lygodiifolia*[10]; the gymnosperm *Pseudotsuga menziesii*[11]; and the spermatophytes such as *Zea mays*[12], *Spinacia oleracea*[13], *Nicotiana tabacum*[14], *Oryza sativa*[15], and *Ipomoea purpurea*[16]. Sequence data of *rbcL* have been used in the study of plant systematics (including the extinct genera[17,18]) for the establishment of phylogenetic relationships[19,20].

Hitherto, there is no sequence information on the *rbcL* gene of *Hevea brasiliensis*. In view of its importance in photosynthesis, the cloning and nucleotide sequence of its coding and 3' non-coding regions are reported.

MATERIALS AND METHODS

Leaves of Rubber Trees, Plasmids, Bacteriophage and Bacterial Strains

Fresh young leaves were harvested from *H. brasiliensis* (RRIM 600) clonal source bushes

*Rubber Research Institute of Malaysia, P.O. Box 10150, 50908 Kuala Lumpur, Malaysia
**Department of Genetics and Cellular Biology, University of Malaya, 59100 Kuala Lumpur, Malaysia

at the Rubber Research Institute of Malaysia Experiment Station, Sungai Buloh, Selangor Darul Ehsan. Recombinant plasmid pTB27 harbouring *Bam*HI-fragment 22c that contains an internal sequence of the tobacco (*N. tabacum*) *rbcL* gene[21] was a generous gift from Pro M. Sugiura, Nagoya University, Japan. Plasmid pUC19 and phage M13mp18 were used as vectors for cloning and DNA sequencing respectively, and were grown in *Escherichia coli* DH5α and DH5αF' respectively.

Enzymes, Antibiotic, and DNA Markers

Restriction endonucleases *Bam*HI, *Cla*I, *Eco*RI, *Hin*cII, and *Hin*dIII were from New England Biolabs, USA. Ampicillin, from Sigma Chemical Company, USA, was used at a final concentration of 50 µg/ml. DNA size reference markers, 1 kilobase pair (kb) DNA ladder, were from Life Technologies, Inc., USA.

General Methods

Media and growth conditions used for the propagation and selection of *E. coli*, with and without plasmid or phage, were as described by Sambrook *et al.*[22]

DNA Techniques

Chlorophast DNA (ctDNA) was prepared from young leaves of RRIM 600 by the methods of Palmer[23] and Herrmann *et al.*[24], with modifications. Standard methods described elsewhere[22] were used in the preparation of plasmid DNA, restriction endonuclease digestions, agarose gel electrophoresis in 1x TBE buffer (89 mM Tris, 89 mM boric acid, and 2.5 mM EDTA, pH 8.3), ligations, transformations, transfections, labelling of DNA with [α–^{32}P] dCTP by the random primed method[25], colony and Southern[26] hybridisations, autoradiography, preparation of single-stranded DNA templates from recombinant M13mp18 for sequencing, and electrophoresis on 6% acrylamide-7 M urea gels to resolve the products of sequencing reactions.

Chloroplast DNA Genomic Library

A shotgun genomic library of the ctDNA of RRIM 600 was constructed by ligating together *Bam*HI-generated ctDNA fragments and *Bam*HI-digested pUC19 and transforming *E. coli* DH5α with the ligation mixture to ampicillin resistance.

Isolation of *Hevea rbcL* Gene

The ctDNA genomic library was screened by probing *E. coli* DH5α transformant colonies harbouring recombinant pUC19 plasmids with radiolabelled *Bam*HI-fragment 22c excised from pTB27 with *Bam*HI. The *Hevea rbcL*-containing insert excised from the recombinant pUC19 from the positive clone was then confirmed by Southern hybridisation.

DNA Sequencing and Sequence Analysis

The *Hevea rbcL*-containing insert from the recombinant pUC19 was sub-cloned into M13mp18 in two orientations. The nucleotide sequences of both strands were determined by Sanger's dideoxyribonucleotide chain termination method[27] using the M13 universal primer or synthetic primers, *Taq* DNA polymerase, and [α–^{35}S] dATP. Synthetic oligonucleotides were designed based on sequence generated earlier and used as primers to continue sequencing. Nested deletions generated with Cyclone Biosystem I kit (International Biotechnologies, Inc., USA) were also used to obtain the double-strand sequence data which were then analysed with the DNA Inspector™ IIe software (Textco, Inc. USA).

RESULTS AND DISCUSSION

Cesium chloride density gradient purified ctDNA from young leaves of RRIM 600 was digested separately with *Bam*HI, *Cla*I, *Eco*RI, *Hin*cII, and *Hin*dIII. Restriction fragments generated from each of these digestions were probed by Southern hybridisation with radiolabelled *Bam*HI-fragment 22c excised from pTB27. In each digestion, a restriction fragment hybridised with the probe (data not shown), indicating that the heterologous probe from tobacco could detect the *rbcL* gene from *Hevea* ctDNA.

Results from colony hybridisation indicated that an ampicillin resistant *E. coli* transformant in the shotgun genomic library of the ctDNA of RRIM 600 harboured a recombinant pUC19

carrying the *rbcL* gene. The recombinant plasmid, designated pHBB24, was purified and digested with *Bam*HI to release an insert of 3.1 kb (*Figure 1a*, lane 1). The result from Southern hybridisation confirmed the presence of the *rbcL* gene on this insert (*Figure 1b*, lane 1).

When pHBB24 was digested with *Eco*RI or *Bam*HI and *Eco*RI, it released a 1.7-kb fragment (*Figure 1a*, lanes 2 and 3) that hybridised with fragment 22c excised from pTB27 (*Figure 1b*, lanes 2 and 3) that hybridised with fragment 22c excised from pTB27 (*Figure 1b*, lanes 2 and 3). This 1.7-kb *Eco*RI fragment was eventually sub-cloned in the *Eco*RI site of M13mp18 in two orientations for sequencing.

Figure 2 shows the coding and the 3' non-coding nucleotide sequences of the rbcL gene of RRIM 600. Like the sequences of maize[12], spinach[13], tobacco[14], liverwort[9], rice[15], and *I. purpurea*[16], the *Hevea rbcL* has a 5-base pair (bp) consensus sequence (GGAGG), the putative ribosome binding site or the Shine-Dalgarno[28] sequence, at 10 bp upstream of the translation initiation codon (AUG). The coding region of the *rbcL* gene has a length of 1425 nucleotides and terminates with a UAA codon. In the 3' non-coding region, palindromic sequences which may form two intrastrand stem-loop structures[14,29] (*Figure 3*) are present. Such structures are important for the termination of transcription of chloroplast and other prokaryotic genes.

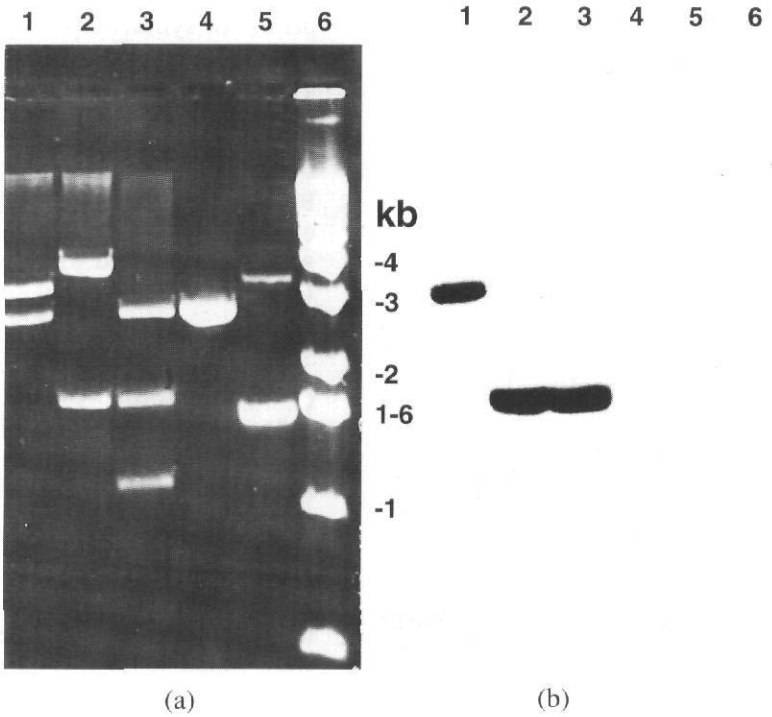

Figure 1. (a) Ethidium bromide-stained 0.7% agarose gel of BamHI-digested pHBB24 (Lane 1), *EcoRI-digested pHBB24* (Lane 2), *BamHI and EcoRI-digested pHBB24* (Lane 3), *EcoRI-digested pUC19* (Lane 4), *undigested pUC19* (Lane 5), *and 1 kb DNA ladder* (Lane 6). (b) *Results of Southern blot hybridisation and autoradiography of the gel in (a), with ^{32}P-labelled BamHI-fragment 22c excised from pTB27 as the probe.*

Figure 2. Nucleotide sequence of the Hevea rbcL *gene. Restriction endonuclease recognition sites are shown in the sequence as bold prints. Palindromic sequences capable of forming transcription stem-loop termination structures are indicated by broken arrows. The translation initiation and termination sites are indicated by italics. The putative ribosome binding site is double-underlined.*

Figure 4 shows the deduced amino acid sequence of the *Hevea* LS protein encoded by the *rbcL* gene. This protein has a length of 475 amino acid residues and has a molecular weight of 52 657. It is smaller than its counterpart in tobacco by two amino acid residues[14]. The amino acid residues between positions 117–132 of the protein form the sequence for the small sub-unit binding site. This sequence is very similar to the corresponding sequences in the LS proteins from maize[12], spinach[13], tobacco[14], and *I. purpurea*[16].

When the LS proteins of *Hevea*, maize[12], spinach[13], tobacco[14], and *I. purpurea*[16] were compared, the amino acids present in or at the vicinity of the proposed catalytic sites were found to be well conserved. In *Hevea*, these amino acid sequences correspond to positions 165–177, 321–339, and 451–466 of its LS protein. A number of amino acid substitutions were noted in these conserved sequences. For example, in *Hevea* LS protein, at position 328 alanine replaced serine in the LS proteins of maize, spinach, and tobacco, and at position

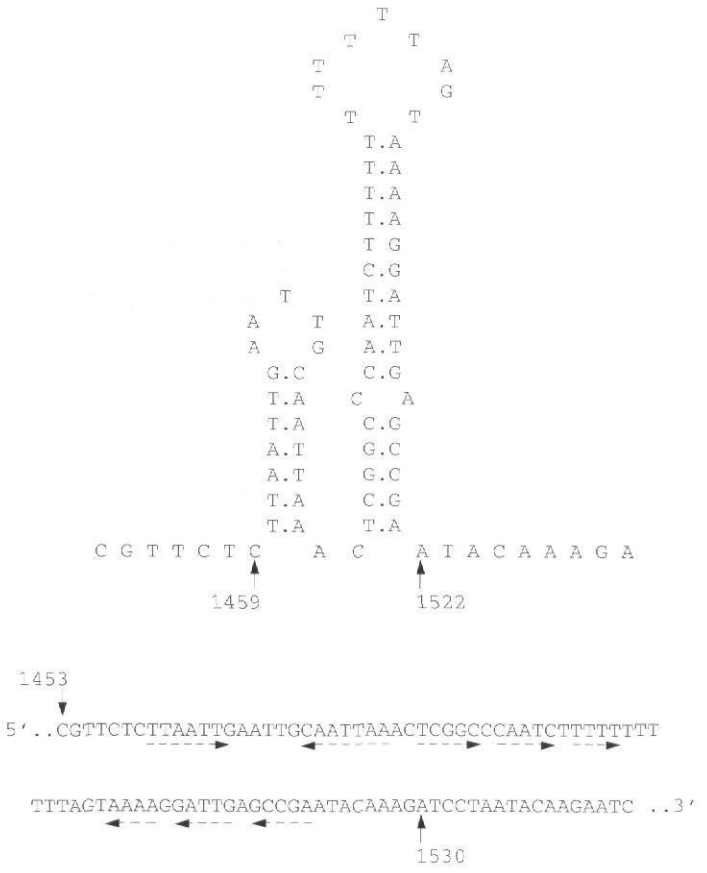

Figure 3. Possible stem and loop secondary structures formed by palindromic sequences (horizontal arrows) at the 3' non-coding region (transcription termination) of the Hevea rbcL *gene.*

```
MSPQTETKAS  VGFKAGVKDY  KLTYYTPEYQ  TKDTDILAAF  RVTPQPGVPP      50
EEAGAAVAAE  SSTGTWTTVW  TDGLTSLDRY  KGRCYDIEPV  PGEENQYIAY     100
VAYPLDLFEE  GSVTNMFTSI  VGNVFGFKAL  RALRLEDLRI  PPAYSKTFQG     150
PPHGIQVERD  KLNKYGRPLL  GCTIKPKLGL  SAKNYGRAVY  ECLRGGLDFT     200
KDDENVNSQP  FMRWRDRFLF  CAEAIYKAQA  ETGEIKGHYL  NATAGTCEEM     250
IKRAVFAREL  GVPIVMHDYL  TGGFTANTSL  AHYCRDNGLL  LHIHRAMHAV     300
IDRQKNHGMH  FRVLAKALRL  SGGDHIHAGT  VVGKLEGERD  ITLGFVDLLR     350
DDFIEKDRSR  GIYFTQDWVS  LPGVIPVASG  GIHVWHMPAL  TEIFGDDSVL     400
QFGGGTLGHP  WGNAPGAVAN  RVALEACVQA  RNEGRDLARE  GNDIIREASK     450
WSPELAAACE  VWKEIKFEFE  AVDTL                               475
```

Figure 4. Amino acid sequence of the Hevea *LS protein as deduced from its* rbcL *gene nucleotide sequence. (Single and double underlines indicate the regions containing the small sub-unit binding site and the active sites, respectively. Amino acid residues to which RuBP binds are in bold print).*

466 lysine replaced arginine in the LS protein of *I. purpurea.*

The cluster of nine histidine residues with four of them in two histidine-isoleucine-histidine sequences as reported in the LS protein of maize[12] is also present in the *Hevea* LS protein. In *Hevea,* this histidine cluster is within positions 267–327, and as in maize, is located nearer to the N-terminal of the active

TABLE 1. CODON USAGE FOR THE TRANSLATION OF THE *HEVEA* LS PROTEIN FROM ITS *rbcL* GENE

Codon (Amino acid)	No. of usage	Codon (Amino acid)	No. of usage	Codon (Amino acid)	No. of usage	Codon (Amino acid)	No. of usage
AAA (lys)	19	AAC (asn)	3	AAG (lys)	5	AAU (asn)	12
ACA (thr)	6	ACC (thr)	7	ACG (thr)	0	ACU (thr)	17
AGA (arg)	6	AGC (ser)	3	AGG (arg)	1	AGU (ser)	3
AUA (ile)	1	ACU (ile)	9	AUG (met)	8	AUU (ile)	13
CAA (gln)	10	CAC (his)	5	CAG (gln)	2	CAU (his)	9
CCA (pro)	5	CCC (pro)	2	CCG (pro)	2	CCU (pro)	14
CGA (arg)	6	CGC (arg)	4	CGG (arg)	0	CGU (arg)	12
CUA (leu)	10	CUC (leu)	0	CUG (leu)	3	CUU (leu)	9
GAA (glu)	22	GAC (asp)	6	GAG (glu)	11	GAU (asp)	22
GCA (ala)	15	GCC (ala)	7	GCG (ala)	1	GCU (ala)	22
GGA (gly)	14	GGC (gly)	2	GGG (gly)	8	GGU (gly)	22
GUA (val)	15	GUC (val)	1	GUG (val)	4	GUU (val)	12
UAA (STOP)	1	UAC (tyr)	5	UAG (STOP)	0	UAU (tyr)	13
UCA (ser)	2	UCC (ser)	4	UCG (ser)	0	UCU (ser)	6
UGA (STOP)	0	UGC (cys)	3	UGG (trp)	8	UGU (cys)	5
UUA (leu)	8	UUC (phe)	8	UUG (leu)	10	UUU (phe)	13

site between positions 321–339. This cluster may play a role in maintaining enzyme activity, perhaps by chelating Mg^{2+}. Similar histidine clusters are also present in the LS proteins of spinach[13], tobacco[14], and *I. purpurea*[16].

Table I shows the codon usage of the *Hevea rbcL* gene. Of the sixty-one codons encoding the twenty amino acids, four were not used. Codon usage was not random for a number of amino acids. For instance, the codons AGG and CGG for arginine, CUC and CUG for leucine, and UCA and UCG for serine were rarely used. On the other hand, the codon GCU for alanine, GAU for aspartic acid, AAU for asparagine, GAA for glutamic acid, AAA for lysine, CCU for proline, and ACU for threonine were frequently used. Like the *rbcL* of *Z. mays*[12], there is a strong bias towards the adenine and uracil as the rhird base in the codons used. This results in a lower G + C content of the gene. Just like the *rbcL* of *N. tabacum*[14], the termination codon UAA was used.

CONCLUSION

The plastid-encoded *rbcL* gene of *H. brasiliensis* (clone RRIM 600) has been isolated and sequenced. The Hevea gene shows great similarity with the nucleotide sequences reported from the *rbcL* gene of other plants[9,12,16].

ACKNOWLEDGEMENTS

We thank the Director, Rubber Research Institute of Malaysia, for permission to publish this paper. We also thank Prof M. Sugiura, Nagoya University, Japan, for pTB27; Mr Chew Nyu Ping and Mr M. Gopalsami for excellent technical assistance; Miss Lim Moo Eng for computer assistance; and Mr Tsan Fan Kui for his assistance in photography. This work was supported by a grant, IRPA Project/Sub-project No. 4-03-04-014/J02, from the Ministry of Science, Technology and the Environment, Malaysia, to the Rubber Research Institute of Malaysia.

Date of receipt: January 1993
Date of acceptance: April 1993

REFERENCES

1. JENSEN, R G. AND BAHR, J T (1977) Ribulose 1,5 biphosphate Carboxylase-oxygenase *A Rev Pl Physiol*, **28,** 379

2. KUNG, S D SAKANO, L AND WILDMAN, S.G (1974) Multiple Peptide Composition of the Large and Small Sub-units of *Nicotiana tabacum* Fraction I Protein Ascertained by Fingerprinting and Electrofocusing. *Biochim biophys. Acta,* **365,** 138

3. CHAN, P H AND WILDMAN, S G (1972) *Chloroplast DNA Codes for the Primary Structure* of the Large Sub-unit of Fraction I Protein *Biochim. biophys Acta,* **277,** 677

4. HIGHFIELD, P E AND ELLIS, R J. (1978) Synthesis and Transport of the Small Sub-unit of Chloroplast Ribulose Bisphosphate Carboxylase *Nature,* **271,** 420

5. YOSHINAGA, K , OHTA T., SUZUKI Y AND SUGIURA, M. (1988) *Chlorella* Chloroplast DNA Sequence Containing a Gene for the Large Sub-unit of Ribulose-1,5-bisphosphate Carboxylase/oxygenase and a Part of a Possible Gene for the β' Sub-unit of RNA Polymerase *Pl. Mol Biol.,* **10,** 245

6. SIEMEISTER, G AND HACHTEL, W (1990) Structure and Expression of a Gene Encoding the Large Sub-unit of Ribulose-1,5-bisphosphate Carboxylase (*rbcL*) in the Colourless Euglenoid Flagellate *Astasia longa. Pl Mol. Biol*, **14,** 825

7 HARDISON, L.K., BOCZAR, B A , REYNOLDS, A.E AND CATTOLICO, R A (1992) A Description of the Rubisco Large Sub-unit Gene and its Transcript in *Olisthodiscus luteus Pl Mol Biol* **18,** 595

8 KONO, M , SATOH, H , OKABE, Y , ABE, Y KAKAYAMA, K AND OKADA, M. (1991) Nucleotide Sequence of the Large Sub-unit of Ribulose-1 5-bisphosphate Carboxylase/oxygenase from the Green Alga *Bryopsis maxima. Pl. Mol. Biol.*, **17,** 505

9 OHYAMA K , FUKUZAWA, H KOHCHI, T , SHIRAI, H , SANO, T , SANO, S , UMESONO, K , SHIKI, Y , YAKEUCHI, M , CHANG Z . AOTA, S., INOKUCHI, H. AND OZEKI, H (1986) Chloroplast Gene Organisation Deduced from Complete Sequence of Liverwort *Marchantia polymorpha* Chloroplast DNA *Nature,* **322,** 572

10. YOSHINAGA, K., KUBOTA, Y., ISHII, T. AND WADA, K (1992). Nucleotide Sequence of *atpB, rbcL, trnR, dedB and psaI* Chloroplast Genes from a Fern *Angiopteris lygodiifolia* a Possible Emergence of Spermatophyta Lineage Before the

Separation of Bryophyta and Pteridophyta. *Pl. Mol. Biol.*, **18,** 79.

11. HIPKINS, V., TSAI, C-H AND STRAUSS, S.H. (1990) Sequence of the Gene for the Large Subunit of Ribulose 1,5-bisphosphate Carboxylase from a Gymnosperm, Douglas Fir. *Pl. Mol. Biol.*, **15,** 505.

12. MCINTOSH, L., POULSEN, C. AND BOGORAD, L. (1980) Chloroplast Gene Sequence for the Large Sub-unit of Ribulose Bisphosphate-carboxylase of Maize. *Nature,* **288,** 556.

13. ZURAWSKI, G., PERROT, B., BOTTOMLEY, W. AND WHITFELD, P.R. (1981) The Structure of the Gene for the Large Sub-unit of Ribulose 1,5-bisphosphate Carboxylase from Spinach Chloroplast DNA. *Nucleic Acids Res.*, **9,** 3251.

14. SHINOZAKI, K. AND SUGIURA, M. (1982) The Nucleotide Sequence of the Tobacco Chloroplast Gene for the Large Subunit of Ribulose-1,5-bisphosphate Carboxylase/oxygenase. *Gene,* **20,** 91.

15. HIRATSUKA, J., SHIMADA, H., WHITTIER, R., ISHIBASHI, T., SAKAMOTO, M., MORI, M., KONDO, C., HONJI, Y., SUN, C.R., MENG, B.Y., LI, Y.Q., KANNO, A., NISHIZAWA, Y., HIRAI, A., SHINOZAKI, K. AND SUGIURA, M. (1989) The Complete Sequence of the Rice (*Oryza sativa*) Chloroplast Genome: Intermolecular Recombination Between Distinct tRNA Genes Accounts for a Major Plastid DNA Inversion During the Evolution of the Cereals. *Mol. Gen. Genet.*, **217,** 185.

16. HABERHAUSEN, G. AND ZETSCHE, K. (1992) Nucleotide Sequence of the *rbcL* Gene and the Intergenic Promoter Region between the Divergently Transcribed *rbcL* and *atpB* Genes of *Ipomoea purpurea* (L.). *Pl. Mol. Biol.*, **18,** 823.

17. GOLENBERG, E.M., GIANNASI, D.E., CLEGG, M.T., SMILEY, C.J., DURBIN, M., HENDERSON, D. AND ZURAWSKI, G. (1990) Chloroplast DNA Seqence from a Miocene *Magnolia* Species. *Nature (Lond.),* **344,** 656.

18. SOLTIS, P., SOLTIS, D.E. AND SMILEY, C.J., (1992) An *rbcL* Sequence from a Miocene *Taxodium* (Bald Cypress). *Proc. Natl. Acad. Sci. USA,* **89,** 449.

19. PALMER, J.D., JANSEN, R.K., MICHAELS, H.J., CHASE, M.W. AND MANHART, J. (1988) Chloroplast DNA Variation and Plant Phylogeny. *Ann. Missouri Bot. Gard.,* **75,** 1180.

20. HASEBE, M., ITO, M., KOFUJI, R., IWATSUKI, K. AND UEDA, K. (1992) Phylogenetic Relationships in Gnetophyta Deduced from *rbcL* Gene Sequences. *Bot. Mag. Tokyo,* **105,** 385.

21. SUGIURA, M., SHINOZAKI, K., ZAITA, N., KUSUDA, M. AND KUMANO, M. (1986) Clone Bank of the Tobacco (*Nicotiana tabacum*) Chloroplast Genome as a Set of Overlapping Restriction Endonuclease Fragments: Mapping of Eleven Ribosomal Protein Genes. *Pl. Sci.*, **44,** 211.

22. SAMBROOK, J., FRITSCH, E.F. AND MANIATIS, T. (1989) *Molecular Cloning: A Laboratory Manual,* 2nd edition. New York: Cold Spring Harbor Laboratory Press.

23. PALMER, J.D. (1986) Isolation and Structural Analysis of Chloroplast DNA. *Meths. Enzymol.*, **118,** 167.

24. HERRMANN, R.G., BOHNERT, H.J., KOAWLLICK, K.V. AND SCHMITT, J.M. (1975) Size, Conformation, and Purity of Chloroplast DNA of Some Higher Plants. *Biochim. Biophys. Acta,* **378,** 305.

25. FEINBERG, A.P. AND VOGELSTEIN, B. (1983) A Technique for Radiolabelling DNA Restriction Endonuclease Fragments to High Specific Activity. *Analyt. Biochem.*, **132,** 6.

26. SOUTHERN, E.M. (1975) Detection of Specific Sequences among DNA Fragments separated by Gel Electrophoresis. *J. Mol. Biol.*, **98,** 503.

27. SANGER, F., NICKLEN, S. AND COULSON, R. (1977) DNA Sequencing with Chain Terminating Inhibitors. *Proc. Natl. Acad. Sci. USA,* **74,** 5463.

28. SHINE, J. AND DALGARNO, L. (1974) The 3'-Terminal Sequence of *Escherichia coli* 16S Ribosomal RNA: Complementarity to Nonsense Triplets and Ribosome Binding Sites. *Proc. Natl. Acad. Sci. USA,* **71,** 1342.

29. ROSENBERG, M. AND COURT, D. (1979) Regulatory Sequences Involved in the Promotion and Termination of RNA Transcription. *A. Rev. Genet.*, **13,** 19.