

An Open Reading Frame (*ycf11*) is Evolutionary Conserved from Cyanobacteria to the Plastid DNAs of Archegoniates and Gymnosperms, is Modified in the Plastid DNAs of Dicots, and is not Plastome Encoded in Monocots

S. KRUSE¹, W. MARTIN², M. WEHE¹, and R. RESKI^{1,*}

¹ Institut für Allgemeine Botanik, Ohnhorststr. 18, 22609 Hamburg, Germany

² Institut für Genetik, Spielmannstr. 7, 38023 Braunschweig, Germany

Received November 24, 1994 · Accepted January 25, 1995

Summary

To study molecular evolution of plants, the plastid encoded *rbcL* sequences are widely used. In most plastid DNAs, an open reading frame (ORF) designated *ycf11* can be found next to the highly conserved *rbcL* gene. This ORF appears to be only loosely conserved and its function is a matter of debate: On the one hand it is the only gene in plastid DNA of land plants suspected to encode a regulatory zinc finger protein. On the other hand it was postulated to encode the β -subunit of an acetyl-CoA-carboxylase. Accordingly, this ORF has been previously described as *zfpA* or *accD*, respectively. Phylogenetic analysis reveals evolutionary conservation of two *ycf11*-domains from bacteria to the plastids of dicots. We show that in dicots *ycf11* has gained additional sequences through insertions, whereas it has been lost from the plastid DNA in monocots. These findings may reflect physiological differences between major groups of land plants. Furthermore, we show that *ycf11* may be a useful molecular marker in the study of plant evolution.

Key words: *Angiopteris lygodiifolia*; *Astasia longa*; *Cuscuta reflexa*; *Epifagus virginiana*; *Euglena gracilis*; *Escherichia coli*; *Marchantia polymorpha*; *Nicotiana tabacum*; *Oryza sativa*; *Physcomitrella patens*; *Pinus thunbergii*; *Pisum sativum*; *Synechocystis PCC6803*; *accD*; acetyl-CoA-carboxylase; molecular evolution; *zfpA*; zinc finger protein.

Abbreviations: *accD* = gene possibly encoding a β -subunit of acetyl-coenzyme A carboxylase, now designated *ycf11*; CPGN = Commission on Plant Gene Nomenclature; ORF = open reading frame; ptDNA = plastid DNA; *rbcL* = gene encoding the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase; *ycf* = mnemonic for plastome encoded open reading frames with yet uncertain function; *zfpA* = gene possibly encoding a zinc finger protein, now designated *ycf11*.

Introduction

Chloroplast gene sequences are widely used to study the evolution of plants at the molecular level. Especially the highly conserved *rbcL* gene has been sequenced from numerous higher plants for this purpose (Clegg, 1993; Chase et al., 1993). However, when studying plant evolution at a higher

taxonomical level, genes showing more variation than *rbcL* may be more advantageous (Olmstead and Palmer, 1994).

During our molecular analysis of chloroplast division (Reski et al., 1991; 1994) we identified a plastid open reading frame (Kasten et al., 1991; 1992) that appeared to be only loosely conserved throughout the plant kingdom. A homolog has been described by Sasaki during her search of light-regulated plastid genes in *Pisum sativum* as a gene possibly encoding a zinc finger protein, the *zfpA*-gene (Sasaki et al.,

* Correspondence.

1989). Sequence comparison showed *zfpA*-related ORFs in *E. coli* (*dedB*) and in the plastid DNAs of *Nicotiana* (ORF512) and *Marchantia* (ORF316). However, some homology between the C-terminal region of *dedB*, *zfpA* and the β -subunit of rat mitochondrial propionyl-CoA carboxylase revealed that *zfpA* might code for an enzyme of C_1 metabolism (Smith et al., 1991). Meanwhile, evidence was presented that *dedB* from *E. coli* and *zfpA* from pea plastids encode a β -subunit of acetyl-coenzyme A carboxylase (Li and Cronan, 1992a; Li et al., 1992; Sasaki et al., 1993). Therefore, the gene was renamed *accD* and it was argued that *zfpA* might be a misleading designation, especially as the region of homology to carboxylases is present in the ORF106 of rice whereas the zinc finger motif is not (Li and Cronan, 1992b). However, the Commission on Plant Gene Nomenclature (CPGN) in their latest release neither sanctions *zfpA* nor *accD* as plant-wide designations for these ORFs until the function of the gene product has been established (Hallick and Bairoch, 1994). In the meantime, the CPGN recommends the designation *ycf11* (Hallick, pers. communication). As the complete intact ORF can be found even in the drastically reduced plastome of the nonphotosynthetic parasitic plant *Epifagus virginiana*, *ycf11* should code for a protein that is essential in the maintenance of plastid structures (Wolfe et al., 1992). However, compared to other plastome-encoded genes, *ycf11* exhibits an unusually high amount of variation throughout the plant kingdom. These observations prompted us to establish the first moss *rbcL* sequence, to identify all known *ycf11* sequences from the literature and to perform analyses of the evolutionary conservation of *ycf11*-gene products. We constructed trees for several land plants and compared these to trees generated from *rbcL* sequences from the same taxa. The results indicate that *ycf11* may serve as a molecular marker for plant evolution.

Materials and Methods

The *rbcL* gene of *Physcomitrella* was cloned as a *Bgl* II fragment and was subsequently sequenced after combined exonuclease III/S1 nuclease treatment according to standard procedures (Sambrook et al., 1989). For comparison of the *ycf11* (*zfpA/accD*) genes, we analysed open reading frames of the following organisms: *Angiopteris lygodiiifolia* ORF314 (Yoshinaga et al., 1992, EMBL X58429), *Cuscuta reflexa* ORF497 (Haberhausen, unpublished, SWISS-PROT P31562), *Epifagus virginiana* ORF505 (Wolfe et al., 1992, GenBank M81884), *Marchantia polymorpha* ORF316 (Ohya et al., 1988, EMBL X04465, Y00686), *Nicotiana tabacum* ORF512 (Shinozaki et al., 1986, EMBL Z00044), *Oryza sativa* ORF106 (Hiratsuka et al., 1989, EMBL X15901), *Physcomitrella patens* ORF315 (Kasten et al., 1992, EMBL X59787), *Pinus thunbergii* ORF321 (Wakasugi et al., 1994, DDBJ 17510), and *Pisum sativum* ORF590 (Nagano et al., 1991, EMBL X56315). As outgroups we used the *Synechocystis* PCC6803 ORF326 (Ogura et al., 1991, no accession number) and the *Escherichia coli* ORF304 (Nonet et al., 1987, Swiss Prot P08193). Note, that the *Pisum ycf11* was first described as ORF371 (Sasaki et al., 1989), and revised as ORF587 (Smith et al., 1991). We took it as ORF590 according to the latest revision of the original describer (Nagano et al., 1991). For comparison of *rbcL* genes we selected the following sequences: *Angiopteris lygodiiifolia* (Yoshinaga et al., 1992, EMBL X58429), *Cuscuta reflexa* (Haberhausen et al., 1992, SWISS-PROT P30401), *Marchantia polymorpha* (Ohya et al., 1988, EMBL X04465, Y00686), *Nicotiana tabacum* (Shinozaki et al., 1986,

EMBL Z00044), *Oryza sativa* (Nishizawa and Hirai, 1987, EMBL D00207), *Physcomitrella patens* (this paper, EMBL X74156), *Pinus thunbergii* ORF321 (Wakasugi et al., 1994, DDBJ 17510), and *Pisum sativum* (Zurawski et al., 1986, EMBL X03853). As an outgroup we used the *Synechocystis* PCC6803 *rbcL* (Yaguchi et al., 1993, EMBL D13539). Nucleotide sequences were translated into protein sequences by DNASIS (Pharmacia, Freiburg, Germany). Standard sequence analysis was performed with the GCG-Package (Devereux et al., 1984). Sequences were aligned with CLUSTAL V (Higgins and Sharp, 1988) and by eye with LINEUP (GCG-package, Genetics Computer Group Inc., Madison, USA). Regions of uncertain positional homology and those positions not occupied in all OTUs (operational taxonomical unit) were excluded from alignments that were used for phylogenetic inference (available upon request). Distance between sequences was measured as numbers of amino acid substitutions per site corrected for multiple substitutions by assuming a gamma distribution for the variability of substitution rate across positions (Ota and Nei, 1994). For this, a neighbour-joining tree (Saitou and Nei, 1987) was constructed using the proportion of amino acid differences between sequences, from which the gamma parameter was estimated. The value thus determined (2.9648 for *ycf11* and 0.4469 for *rbcL*) was used to estimate numbers of substitutions per site between sequences using the gamma correction. The resulting distance matrix yielded the final neighbour-joining tree (calculated with the programme NJBOOT kindly provided by M. Nei, Penn State University). The reliability of branches was estimated by bootstrapping (100 replications) using the same gamma parameter.

Results

Ycf11 (*zfpA/accD*)-sequences are available from a cyanobacterium and from the plastid DNAs of a moss, a liverwort, a fern, a gymnosperm, a monocotyledonous plant, and four dicotyledonous plants. The latter include the dramatically reduced plastome of the nonphotosynthetic parasitic plant *Epifagus virginiana*. Alignment of the deduced proteins revealed two conserved domains (Fig. 1); the N-terminus harbouring the sequence capable to encode a C4-type zinc finger motif and the C-terminus exhibiting homology to carboxylases (data not shown). Both domains comprise nearly the whole putative polypeptide in the cyanobacterium

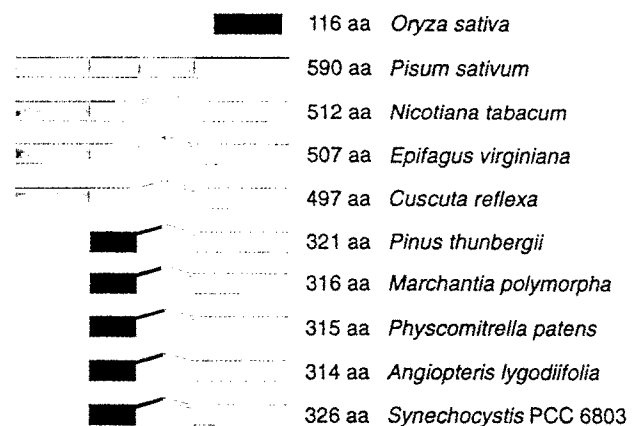


Fig. 1: Schematic representation of sequence alignment of all known *ycf11* (*zfpA/accD*)-proteins. Homologous regions are represented by black boxes, non-homologous regions by shaded boxes. aa: amino acids.

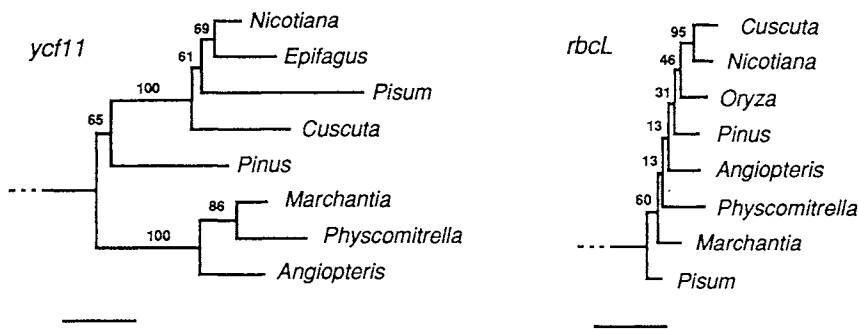


Fig. 2: Phylogenetic trees inferred by the neighbour-joining method of *ycf11* (*zfpA/accD*)-protein sequences and *rbcL*-protein sequences, respectively. Numbers attached to internal branches contain the results of bootstrap parsimony analysis for the corresponding protein and indicate the number of times the branch was found out of 100 replicates. Branch lengths are drawn to scale. The scale bars represent 0.1 substitution per site in both trees. *ycf11*: 253 sites compared, gamma parameter 2.96475 determined on p-value tree. *rbcL*: 474 sites compared, gamma parameter = 0.44690 determined on p-value tree.

(ORF326), the three archegoniates (moss ORF315, liverwort ORF316, fern ORF314), and the gymnosperm (black pine ORF321). The respective ORFs of the three dicots are larger (*Cuscuta* ORF497, *Epifagus* ORF505, tobacco ORF512, pea ORF590) and have no extensive homology in the additional sequences, although there are some between *Nicotiana* and *Epifagus* (data not shown). The ORF of the monocot is smaller (rice ORF106) and encodes solely part of the C-terminal region. Therefore, it was not considered for further analysis. The aligned proteins were taken to construct neighbour-joining trees with the *Synechocystis* sequence and the *E. coli dedB* gene (ORF304) as outgroups (Fig. 2 *ycf11*). Here, the archegoniates are clearly separated from the seed plants. Moreover, liverwort and moss group as bryophytes distinct from the fern. Within the seed plants there is a deep bifurcation between the gymnosperm and the four dicotyledonous angiosperms.

Molecular evolution in plants is widely studied with the plastid gene encoding the large subunit of the carbon dioxide fixing enzyme ribulose-1,5-bisphosphate-carboxylase/oxygenase, the *rbcL* gene. In most plant species, *rbcL* and *ycf11* are adjacent plastid genes. In order to compare *rbcL* sequences from the same plants as above, we established the first *rbcL* sequence of a moss (*Physcomitrella patens*, EMBL X74156). The nonphotosynthetic parasitic plant *Epifagus* was not included as its *rbcL* is a pseudogene (Wolfe et al., 1992). The aligned *rbcL*-proteins were taken to construct neighbour-joining trees with the *Synechocystis* sequence as an outgroup (Fig. 2 *rbcL*). Obviously, the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase has not evolved fast enough to be useful in molecular phylogenies; i.e. the *Pisum* sequence is erroneously placed at the deepest bifurcation, putatively due to stochastic similarity to the *Synechocystis* outgroup.

Discussion

One striking feature of the plastome encoded *ycf11* is its enormous variability in length (ORF106 to ORF590), which is unknown for other plastid genes. Our alignment of the polypeptide sequences showed two regions of conserved homology in all species except the monocot, indicating that the protein comprises two domains; an N-terminal domain

suspected to fold to a zinc finger (Sasaki et al., 1989) and the C-terminal one, exhibiting homology to non-plant carboxylases (Smith et al., 1991). The *Oryza* ORF106 covers exactly the hot-spot region of length mutations in the wheat group (Ogihara et al., 1992). Therefore, in *Triticum* and *Aegilops* it is even more divergent from other *ycf11* sequences than it is in *Oryza* itself. In light of the high degree of conservation of typical plastid genes, it seems very unlikely that a functional plastid gene exists in one monocot while it is absent in others, as Li and Cronan (1992b) implied. Therefore, monocots may lack the *ycf11* gene product and this deficiency may be the cause of sensitivity of gramineae to certain herbicides (Konishi and Sasaki, 1994).

In most plant species *ycf11* (*zfpA/accD*) and *rbcL* are adjacent plastid genes and *rbcL* sequences are widely used to study molecular evolution, especially of higher plants (Clegg, 1993; Chase et al., 1993). For comparison, we constructed neighbour-joining trees based on *rbcL* and thus had to establish the first *rbcL* sequence of a moss. Using this data set, the slow evolution of *rbcL* once more became evident. Because of the limited substitution of amino acids, one single sequencing error can significantly alter the position of the sequence within the tree. This may be the case with the *Pisum* sequence currently available in databases. Therefore, analysis of plastid genes with a higher substitution rate than *rbcL* are necessary (Olmstead and Palmer, 1994).

The plastid DNA region covering *ycf11* was recently used to trace the origin of seed plant evolution (Yoshinaga et al., 1992; Raubeson and Jansen, 1992). We now present data on all *ycf11* sequences available so far, spanning the entire range of plant evolution from cyanobacteria to flowering plants. These data reveal evolutionary conservation of sequences encoding a carboxylase function as well as a zinc binding motif from cyanobacteria to the plastids of dicots. The plastome encoded dicot genes gained additional N-terminal sequences with no significant homologies and yet unknown functions. In *Pisum*, however, *ycf11* and *rbcL* are uncoupled due to a legume-specific rearrangement in the ptDNA (Nagano et al., 1991), affecting transcription of the *ycf11*-*petA* operon during evolution (Kruse et al., 1995). Furthermore, *ycf11* gained additional internal sequences separating the two putative domains of its protein in pea plastids. Regarding *ycf11*, monocotyledonous plants occupy a special position: During the monocot-dicot split this gene was either transferred from the

prokaryon (= plastid DNA) to the eukaryon (= nuclear DNA), or this plant group is the only one that lost an essential bacterial type protein, as argued by Konishi and Sasaki (1994). However, the latter appears to be unlikely, as *ycf11* is under a strong selection pressure even in the reduced plastome of *Epifagus virginiana*, a nonphotosynthetic parasite, indicating that it encodes a protein essential for maintenance of plastid structures (Wolfe et al., 1992). Therefore, *ycf11*, unlike most other plastid genes (Olmstead and Palmer, 1994), seems to be appropriate to include non-photosynthetic plants into a phylogenetic analysis based on plastid DNA sequences. On the other hand, *ycf11*-homologous sequences are not plastome-encoded, neither in *Euglena gracilis* (Hallick et al., 1993) nor in *Astasia longa* (Gockel et al., 1994). However, the phylogenetic origin of euglenoid plastids still remains uncertain (Martin et al., 1992).

Schimper (1883) postulated the idea of endosymbiosis followed by the hypothesis of Mereschkowsky (1905) that cyanobacteria are the progenitors of chloroplasts. Today, these ideas are widely accepted and it became evident that the prokaryotic ancestor lost its autonomy by losing most of its genes (Gray, 1992). Some of these prokaryotic genes were transferred into the eukaryotic background of the nucleus. The problem of achieving coordinated expression of these genes and subsequent reimport of their products into the ancestral organelle are obvious, and require coevolution of three different genomes in one single cell (Herrmann et al., 1992). Gene transfer may be an ongoing process for *ycf11*, and thus, despite the debate about its possible function, *ycf11* sequences appear to be suitable for the study of plastid evolution.

Acknowledgements

Financial support by the Deutsche Forschungsgemeinschaft (Re837/2-2) is gratefully acknowledged. This article is based on a doctoral study of S. K. at the University of Hamburg, Faculty of Biology.

References

- CHASE, M. W., D. E. SOLTIS, R. G. OLNSTEAD, D. MORGAN, D. H. LES, B. D. MISCHLER, M. R. DUVAL, R. PRICE, H. G. HILLS, Y.-L. QIU, K. A. KRON, J. H. RETTIG, I. CONTI, J. D. PALMER, J. R. MANHARDT, K. J. SYSTMA, H. J. MICHEALS, W. J. KRESS, K. G. KAROL, W. D. CLARK, M. HEDREN, B. S. GAUT, R. K. JANSSEN, K.-J. KIM, C. F. WIMPEE, J. F. SMITH, J. R. FURNIER, S. H. STRAUSS, Q.-Y. XIANG, G. M. PLUNKETT, P. S. SOLTIS, S. M. SWENSEN, S. E. WILLIAMS, P. A. GADEK, C. J. QUINN, L. E. EGUIARTE, E. GOLENBERG, G. H. LEARN, S. GRAHM, S. C. H. BARRETT, S. DAYANDAN, and V. A. ALBERT: DNA sequence phylogenetics of seed plants: An analysis of nucleotide sequences from the plastid gene *rbcl*. *Annals Missouri Bot. Gard.* 80, 528–580 (1993).
- CLEGG, M. T.: Chloroplast gene sequences and the study of plant evolution. *Proc. Natl. Acad. Sci. USA* 90, 363–367 (1993).
- DEVEREUX, J., P. HAEBERLI, and O. SMITHIES: A comprehensive set of sequence analysis programmes for the VAX. *Nucl. Acids Res.* 12, 387–395 (1984).
- GOCKEL, G., W. HACHTEL, S. BAIER, C. FLISS, and M. HENKE: Genes for components of the chloroplast translational apparatus are conserved in the reduced 73-kb plastid DNA of the nonphotosynthetic euglenoid flagellate *Astasia longa*. *Curr. Genet.* 26, 256–262 (1994).
- GRAY, M. W.: The endosymbiont hypothesis revisited. *Int. Rev. Cyt.* 141, 233–357 (1992).
- HABERHAUSEN, G., K. VALENTIN, and K. ZETSCHKE: Organization and sequence of photosynthetic genes from the plastid genome of the holoparasitic flowering plant *Cuscuta reflexa*. *Mol. Gen. Genet.* 231, 154–161 (1992).
- HALLICK, R. B. and A. BAIROCH: Proposals for the naming of chloroplast genes. III. Nomenclature for open reading frames encoded in chloroplast genomes. *Plant Mol. Biol. Rep.* 12, 29–30 (1994).
- HALLICK, R. B., L. HONG, R. G. DRAGER, M. R. FAVREAU, A. MONFORT, B. ORSAT, A. SPIELMANN, and E. STUTZ: Complete sequence of *Euglena gracilis* chloroplast DNA. *Nucl. Acids Res.* 21, 3537–3544 (1993).
- HERRMANN, R. G., P. WESTHOFF, and G. LINK: Biogenesis of plastids in higher plants. In: HERRMANN, R. G. (ed.): *Plant Gene Research. Cell Organelles*, 275–349. Springer Verlag, Wien, New York (1992).
- HIGGINS, D. G. and P. M. SHARP: CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* 73, 237–244 (1988).
- HIRATSUKA, J., H. SHIMADA, R. WHITTIER, T. ISHIBASHI, M. SAKAMOTO, M. MORI, C. KONDO, Y. HONJI, C.-R. SUN, B.-Y. MENG, Y.-Q. LI, A. KANNO, Y. NISHIZAWA, A. HIRAI, K. SHINOZAKI, and M. SUGIURA: 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, 185–194 (1989).
- KASTEN, B., M. WEHE, R. RESKI, and W. O. ABEL: *trnR*-CCG is not unique to the plastid DNA of the liverwort *Marchantia*: gene identification from the moss *Physcomitrella patens*. *Nucl. Acids Res.* 19, 5074 (1991).
- KASTEN, B., M. WEHE, S. KRUSE, K. REUTTER, W. O. ABEL, and R. RESKI: The plastome-encoded *zfpA* gene of a moss contains prokaryotic as well as eukaryotic promoter consensus sequences and its RNA abundance is modulated by cytokinin. *Curr. Genet.* 22, 327–333 (1992).
- KONISHI, T. and Y. SASAKI: Compartmentalization of two forms of acetyl-CoA carboxylase in plants and the origin of their tolerance toward herbicides. *Proc. Natl. Acad. Sci. USA* 91, 3598–3601 (1994).
- KRUSE, S., S. QUAST, A. IPSEN, K. REUTTER, and R. RESKI: Cytokinin promotes maturation of complex plastid transcripts in cytokinin-sensitive moss mutants. submitted (1995).
- LI, S.-J. and J. E. CRONAN: The genes encoding the two carboxyl-transferase subunits of *Escherichia coli* acetyl-CoA carboxylase. *J. Biol. Chem.* 267, 16841–16847 (1992a).
- Putative zinc finger protein encoded by a conserved chloroplast gene is very likely a subunit of a biotin-dependent carboxylase. *Plant Mol. Biol.* 20, 759–761 (1992b).
- LI, S.-J., C. O. ROCK, and J. E. CRONAN: The *dedB* (*usg*) open reading frame of *Escherichia coli* encodes a subunit of acetyl-coenzyme A carboxylase. *J. Bac.* 174, 5755–5757 (1992).
- MARTIN, W., C. C. SOMERVILLE, and S. LOISEAUX-DE GOER: Molecular phylogenies of plastid origins and algal evolution. *J. Mol. Evol.* 35, 385–404 (1992).
- MERESCHKOWSKY, C.: Über Natur und Ursprung der Chromatophoren im Pflanzenreiche. *Biol. Centralbl.* 25, 593–604 (1905).
- NAGANO, Y., R. MATSUNO, and Y. SASAKI: Sequence and transcriptional analysis of the gene cluster *trnQ-zfpA-psaI-ORF231-petA* in pea chloroplasts. *Curr. Genet.* 20, 431–436 (1991).
- NISHIZAWA, Y. and A. HIRAI: Nucleotide sequence and expression of the gene for the large subunit of rice ribulose-1,5-bisphosphate carboxylase. *Jpn. J. Genet.* 62, 389–395 (1987).

- NONET, M. L., C. C. MARVEL, and D. R. TOLAN: The *hisT-purF* region of the *Escherichia coli* K-12 chromosome. *J. Biol. Chem.* **262**, 12209–12217 (1987).
- ÖGIHARA, Y., T. TRACHI, and T. SASAKUMA: Structural analysis of length mutations in a hot-spot region of wheat chloroplast DNAs. *Curr. Genet.* **22**, 251–258 (1992).
- OGURA, Y., T. YOSHIDA, Y. NAKAMURA, M. TAKEMURA, K. ODA, and K. OHYAMA: Gene encoding a putative zinc finger protein in *Synechocystis* PCC6803. *Agric. Biol. Chem.* **55**, 2259–2264 (1991).
- OHYAMA, K., H. FUKUZAWA, T. KOHCHI, T. SANO, S. SANO, H. SHIRAI, K. UMEZOMO, Y. SHIKI, M. TAKEUCHI, Z. CHANG, S. AOTA, H. INOKUCHI, and H. OZEKI: Structure and organization of the *Marchantia polymorpha* chloroplast genome. Cloning and gene identification. *J. Mol. Biol.* **203**, 281–298 (1988).
- OLMSTEAD, R. G. and J. D. PALMER: Chloroplast DNA systematics: A review of methods and data analysis. *Am. J. Bot.* **81**, 1205–1224 (1994).
- OTA, T. and M. NEI: Estimation of the number of amino acid substitutions per site when the substitution rate varies among sites. *J. Mol. Evol.* **38**, 642–643 (1994).
- RAUBESON, L. A. and R. K. JANSEN: Chloroplast DNA evidence on the ancient evolutionary split in vascular land plants. *Science* **255**, 1697–1699 (1992).
- RESKI, R., M. WEHE, B. HADELER, J. R. MARIENFELD, and W. O. ABEL: Cytokinin and light quality interact at the molecular level in the mutant PC22 of the moss *Physcomitrella*. *J. Plant Physiol.* **138**, 236–243 (1991).
- RESKI, R., M. FAUST, X.-H. WANG, M. WEHE, and W. O. ABEL: Genome analysis of the moss *Physcomitrella patens* (Hedw.) B.S.G. *Mol. Gen. Genet.* **244**, 352–359 (1994).
- SAITOU, N. and M. NEI: The neighbour-joining method: A new method for the reconstruction of phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).
- SAMBROOK, J., E. F. FRITSCH, and T. MANIATIS: Molecular cloning. A laboratory manual, 2nd edn. Cold Spring Harbour Laboratory Press, Cold Spring Harbour, USA, 1989.
- SASAKI, Y., Y. NAGANO, S. MORIOKA, H. ISHIKAWA, and R. MATSUNO: A chloroplast gene encoding a protein with one zinc finger. *Nucl. Acids Res.* **17**, 6217–6227 (1989).
- SASAKI, Y., K. HAKAMADA, Y. SUAMA, Y. NAGANO, I. FURUSAWA, and R. MATSUNO: Chloroplast-encoded protein as a subunit of acetyl-CoA carboxylase in pea plant. *J. Biol. Chem.* **268**, 25118–25123 (1993).
- SCHIMPER, A. F.: Über die Verteilung der Chlorophyllkörner und Farbkörper. *Bot. Ztg.* **41**, 105–112, 121–131, 137–146, 153–162, Erwiderung: 803–817 (1883).
- SHINOZAKI, K., M. OHME, M. TANAKA, T. WAKASUGI, N. HAYASHIDA, T. MATSUBAYASHI, N. ZAITA, J. CHUNGWONGSE, J. OBOKATA, K. Y. SHINOZAKI, C. OHTO, K. TORAZAWA, B. Y. MENG, M. SUGITA, H. DENO, T. KAMOGASHIRA, K. YAMADA, J. KUSUDA, F. TAKAIWA, A. KATO, N. TOHDO, H. SHIMADA, and M. SUGIURA: The complete nucleotide sequence of the tobacco chloroplast genome: Its gene organization and expression. *EMBO J.* **5**, 2043–2049 (1986).
- SMITH, A. G., R. M. WILSON, T. M. KAETHNER, D. L., WILLEY, and J. C. GRAY: Pea chloroplast genes encoding a 4 kDa polypeptide of photosystem I and a putative enzyme of C₁ metabolism. *Curr. Genet.* **19**, 403–410 (1991).
- WAKASUGI, T., J. TSUDZUKI, S. ITO, M. SHIBATA, and M. SUGIURA: A physical map and clone bank of the black pine (*Pinus thumbergii*) chloroplast genome. *Plant Mol. Biol. Rep.* **12**, 227–241 (1994).
- WOLFE, K. H., C. W. MORDEN, and J. D. PALMER: Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. *Proc. Natl. Acad. Sci. USA* **89**, 10648–10652 (1992).
- YAGUCHI, T., S. CHUNG, Y. IGARASHI, and T. KODAMA: Cloning, sequence and overexpression of the thermophilic cyanobacterium gene for the ribulose-1,5-bisphosphate carboxylase/oxygenase. *J. Ferment. Bioeng.* **75**, 1–8 (1993).
- YOSHINAGA, K., Y. KUBOTA, T. ISHII, and K. WADA: Nucleotide sequence of *atpB*, *rbcL*, *trnR*, *dedB* and *psaI* chloroplast genes from a fern *Angiopteris lygodifolia*: a possible emergence of *Spermatophyta* lineage before the separation of *Bryophyta* and *Pteridophyta*. *Plant Mol. Biol.* **18**, 79–82 (1992).
- ZURAWSKI, G., P. R. WHITFIELD, and W. BOTTOMLEY: Sequence of the gene for the large subunit of ribulose-1,5-bisphosphate carboxylase from pea chloroplasts. *Nucl. Acids Res.* **14**, 3975–3976 (1986).