Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biochemical and Biophysical Research Communications 299 (2002) 57–61
BBRC
www.academicpress.com

Non-AUG translation initiation of mRNA encoding plastid-targeted phage-type RNA polymerase in *Nicotiana sylvestris*☆,☆☆

Yuki Kobayashi,[a] Yuko Dokiya,[a] Yoshinori Kumazawa,[b] and Mamoru Sugita[a,*]

[a] *Center for Gene Research, Nagoya University, Chikusa, Nagoya 464-8602, Japan*
[b] *Department of Earth and Planetary Sciences, Graduate School of Science, Nagoya University, Nagoya 464-8602, Japan*

Received 15 October 2002

Abstract

A third nuclear gene encoding a bacteriophage T7-type RNA polymerase, NsRpoT-C, was isolated and characterized from *Nicotiana sylvestris*. The gene, *NsRpoT-C*, consists of 21 exons and 20 introns and encodes a polypeptide of 977 amino acid residues. The predicted NsRpoT-C protein shows the highest identity (72% amino acid identity) with *Arabidopsis thaliana* RpoT;3 which is a plastid-targeted protein. Surprisingly, comparison of the deduced amino acid sequence of NsRpoT-C with that of *A. thaliana* RpoT;3 predicted that the *NsRpoT-C* starts at a CUG triplet, a rare translation initiation codon. Transient expression assays in protoplasts from tobacco leaves demonstrated that the putative N-terminal transit peptide of NsRpoT-C encodes a targeting signal directing the protein into chloroplasts. This strongly suggests that NsRpoT-C functions as an RNA polymerase transcribing plastid-encoded genes. We have designated this protein NsRpoTp.

Keywords: Non-AUG translation initiation; Chloroplast; Bacteriophage T7-type RNA polymerase; *Nicotiana sylvestris*; Plastid localization; Tobacco; Moss; *Physcomitrella patens*

Single subunit bacteriophage-like RNA polymerases (RpoT) are widely distributed among eukaryote lineages and are most commonly localized in the mitochondria [1]. In land plants, most plastid genes are transcribed by a eubacterial-type plastid-encoded RNA polymerase although some plastid genes are transcribed by a nucleus-encoded RpoT [2,3].

A number of nuclear RpoT genes have been identified from humans [4], fungi [5,6], the malaria parasite *Plasmodium falciparum* [7], and land plant species [8–16]. Plant RpoTs are encoded by small gene families and localized in either the mitochondria or plastids. For example, *Arabidopsis* RpoT;1, RpoT;3, and RpoT;2, respectively, target mitochondria [9], plastids [9], and both [10]. To date, only two RpoT genes have been identified in monocotyledonous plants [12,13] and the moss *Physicomitrella patens* [16]. We previously identified two *Nicotiana sylvestris* RpoT genes and demonstrated that one gene product, NsRpoT-A, is targeted to mitochondria [14] and another one, NsRpoT-B, to both mitochondria and plastids [15]. This raises the question of whether a set of three RpoT genes exists in general in dicotyledonous plants or if it is unique to *Arabidopsis* plant.

In this study, we isolated and characterized a third RpoT gene, *NsRpoT-C*, from *N. sylvestris*. Surprisingly, a CUG triplet of the *NsRpoT-C* mRNA was shown to function as a translation initiation codon. The assay for targeting properties of the putative transit peptide of NsRpoT-C clearly indicated that NsRpoT-C is targeted to chloroplasts.

Materials and methods

Plant materials. N. sylvestris and *N. tabacum* were grown at 28 °C in a growth chamber and the *N. sylvestris* mature leaves, young leaves, stems, roots, and cotyledons were harvested as described previously

☆ *Abbreviations:* RpoT, bacteriophage T7-type RNA polymerase; RACE, rapid amplification of cDNA ends; PCR, polymerase chain reaction; RT-PCR, reverse transcription-PCR; GFP, green fluorescence protein; sGFP, synthetic GFP.

☆☆ The nucleotide sequences reported in this paper have been deposited in the DDBJ/GenBank/EMBL databases under Accession Nos. AB084950 (*NsRpoT-C* gene) and AB084951 (*NsRpoT-C* cDNA).

* Corresponding author. Fax: +81-52-789-3080.
E-mail address: sugita@gene.nagoya-u.ac.jp (M. Sugita).

0006-291X/02/$ - see front matter © 2002 Elsevier Science (USA). All rights reserved.
PII: S0006-291X(02)02579-2

[14]. The moss *P. patens* subsp. *patens* was grown at 25 °C under continuous illumination from white fluorescent lamps at 40 μmol/m^2/s [17].

Isolation and sequencing of genomic clone. Among the previously isolated λ clones [14], a clone λNS7 was sequenced and characterized as described previously [14,15]. Phylogenetic tree was constructed using the neighbor-joining method [18].

Isolation of cDNAs and rapid amplification of cDNA ends. Total cellular RNA prepared from the *N. sylvestris* leaves was treated with DNase I, and reverse transcribed using an oligo$(dT)_{12-18}$ as described previously [14]. A cDNA fragment (2538 bp) was amplified by polymerase chain reaction (PCR) using the first-strand cDNA as template and a pair of primers: C-L (5′-GTTAAAGAAGTAGGGCGACTCGAGAAC-3′) and C-56 (5′-AGAGAAATTGAGAGAAGGAGAAGATATG-3′). The 3′-terminal portion of the mRNA (cDNA) was isolated by 3′-rapid amplification of cDNA ends (3′-RACE), as described [14]. The first-strand cDNAs were synthesized using oligo(dT) adapter primer, 5′-GG CCACGCGTCGACTAGTAC$(T)_{17}$. They were used as the templates for 3′-RACE with the primers 3′-anchor (5′-GATTGGCCACGCGTC GACTAGTAC-3′) and 3′-C1 (5′- ATGTGCAACCTCCAGTGGATC AGTC-3′), and subsequently using the primers 3′-anchor and 3′-C2 (5′-A TTCAGAATTGCTACAAGAGATCCG-3′). Finally, cDNA amplification was performed using 3′-anchor and 3′-C3 (5′-TGTGATCCATT GGTTGTTGCGGGAG-3′) and the PCR product was obtained.

Transient expression of GFP fusion proteins in protoplasts. DNA fragment encoding the N-terminal 171 amino acids of NsRpoT-C was amplified from the cDNA as above with primers 5′-CTG (5′-ATCGTC GACATTAAAGCTGGCTTCCACA-3′) and 3′-GFP (5′-GATCCAT GGTCTCACACATTTCCCTCTCCAACT-3′). DNA fragment encoding the N-terminal residues 59–171 of NsRpoT-C was amplified with primers 5′-ATG (5′-GGGTCGACCATGCCAATTAATAATAATAT TCAGTCCCAA-3′) and 3′-GFP. *Sal*I–*Nco*I-digested amplified DNA fragments were introduced into *Sal*I–*Nco*I-cleaved CaMV35S-sGFP(S65T)-nos3′ [19] and plasmids CTG-GFP and ATG_{59}-GFP were obtained. The putative translation initiation codon CTG of NsRpoT-C was mutated to CTC or ATG using the Quick Change site-directed mutagenesis kit (Stratagene) using plasmid CTG-GFP and the following oligonucleotides; CTC-F (5′-GTCGACATTAAAGCTCGCTTCC ACAGCTTC-3′) and CTC-R (5′-GAAGCTGTGGAAGCGAGCTTT AATGTCGAC-3′) for construction of plasmid CTc-GFP, and ATG-F (5′-GTCGACATTAAAGATGGCTTCCACAGCTTC-3′) and ATG-R (5′-GAAGCTGTGGAAGCCATCTTTAATGTCGAC-3′) for construction of plasmid aTG-GFP. As control, a construct carrying the *P. patens* chloroplast σ factor transit peptide of PpSIG1 [20] was used. These reporter constructs were introduced into the protoplasts prepared from *N. tabacum* leaves or *P. patens* protonemata by polyethyleneglycol-mediated transformation [14,20]. One day after transformation, fluorescence of GFP was monitored using a BX-50 fluorescence microscope (Olympus).

Measurement of mRNA levels by RT-PCR. Total cellular RNA was prepared from the different *N. sylvestris* organs and treated with DNase I, and then cDNAs were synthesized as described above. RT-PCR was performed using primers CP1 (5′-GTTAGAGCTAAACAGGTAC AA-3′) and CP2 (5′-ATTACAGCCATTTTATCCGCA-3′) and a 326-bp fragment was amplified by the indicated number of cycles (each of 30 s at 94 °C, 30 s at 50 °C, and 30 s at 72 °C). To compare the relative amounts of mRNA in each sample, an α-tubulin gene sequence was used as a control for the PCRs [14]. PCR products were separated by agarose gel electrophoresis [21].

Results and discussion

Gene structure and coding sequence of NsRpoT-C

In the present study, the clone λNS7 was found to encode a polypeptide homologous to NsRpoT-A [14] and NsRpoT-B [15]. Therefore, we tentatively designated this gene *NsRpoT-C*. Comparison of the genomic (AB084950) and cDNA sequences (AB084951) revealed that the gene consists of 21 exons and 20 introns, and encompasses a region of 9.9 kb. All plant RpoT genes identified to date are composed of 18 introns and 19 exons. The *NsRpoT-C* gene has two additional introns. The first intron (601 bp) is positioned within the first exon of the other plant RpoT genes and the additional 20th intron (154 bp) in the 3′ untranslated region. The positions of the remaining introns (82–816 bp) in *NsRpoT-C*, however, are identical to those of the other plant RpoT genes.

The reading frame starting from the first ATG encodes 919 amino acid residues with 72% amino acid identity with *Arabidopsis* RpoT;3 [9] and only 25% identity with T7 RNA polymerase [22]. However, functional motifs identified in T7 RNA polymerase [23,24] are well conserved in the *NsRpoT-C* gene. To further investigate the phylogenetic relationships of these proteins, we constructed a phylogenetic tree with 13 RpoTs from various plant species (Fig. 2). NsRpoT-C clusters with orthologs of *Arabidopsis* RpoT;3 and maize RpoTp, which are experimentally shown to be plastid-targeted proteins.

Possible translation initiation at CTG codon

Interestingly, the *NsRpoT-C* gene encodes an N-terminal extension of 74 amino acids upstream from the first in-frame ATG codon that lacks the canonical translation initiation codon (Fig. 1). The N-terminal 18 residues following the first leucine residue (encoded by CTG) of NsRpoT-C are highly conserved with N-terminal *Arabidopsis* RpoT;3 (Fig. 3). Thus, the CTG triplet was considered to be the translation initiation codon. Non-AUG translation codons are well known to function in plant virus RNAs [25] and the *Arabidopsis AGAMOUS* gene [26]. *NsRpoT-C* is therefore the second example in plants of a non-viral mRNA initiated from a non-AUG codon. If *NsRpoT-C* does start at the CUG, the deduced polypeptide will consist of 977 amino acid residues.

Plant and animal phage-type RNA polymerases contain N-terminal transit peptide sequences targeting the protein to the mitochondria or plastids. There are two possible translation initiation sites for the *NsRpoT-C* mRNA, either CUG or AUG codon (Fig. 1). We used the TargetP [27] and MitoProt II [28] programs for protein sorting to predict whether the N-terminal 145 residues of NsRpoT-C starting from the CUG codon specify targeting to plastids or mitochondria (Fig. 3). The TargetP program produced a score of 0.856 for localization to plastids and 0.116 for mitochondrial localization. In contrast, the MitoProt II did not predict mitochondrial localization (a score of

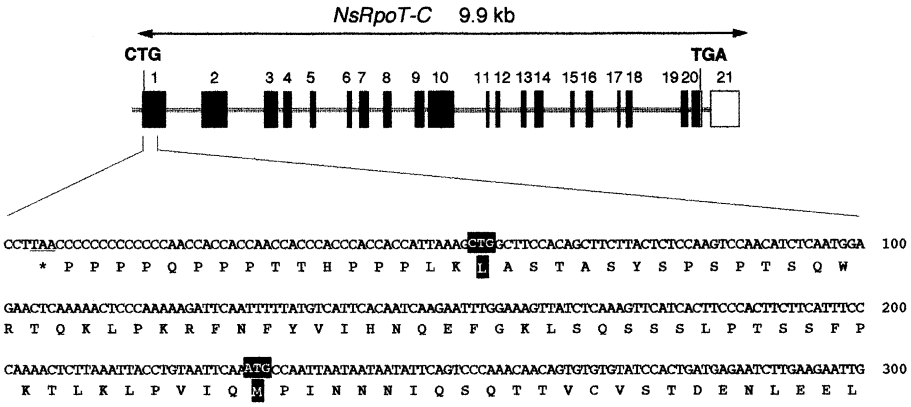

Fig. 1. (Top) Structure of the *N. sylvestris NsRpoT-C* gene as deduced from the genomic and cDNA sequences. The translated and untranslated regions are indicated by filled and opened boxes, respectively. (Bottom) 5′ part of the *NsRpoT-C* gene sequence together with the corresponding predicted amino acid sequence. The first methionine (M) and leucine (L) and their triplets are indicated as filled squares.

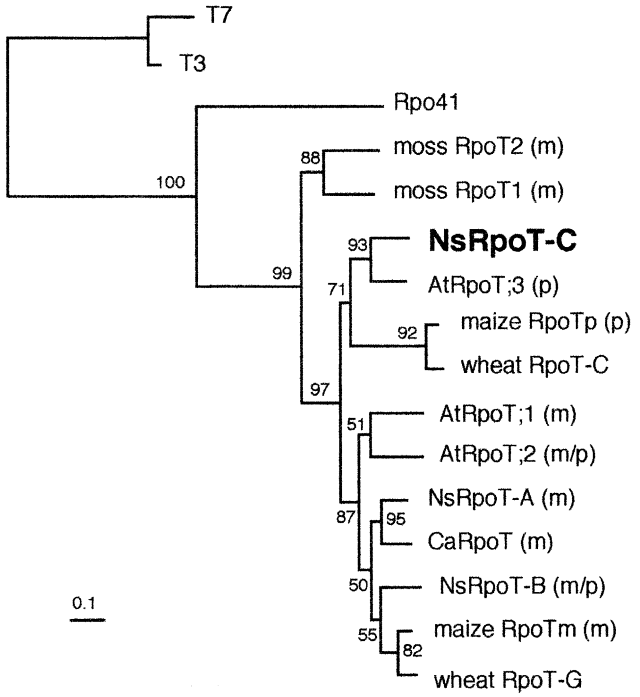

Fig. 2. Phylogenetic tree constructed from the sequences of RpoTs from various plant species. The tree was constructed using the neighbor-joining method [18] from aligned amino acid sequences (668 sites after gaps were removed). Numbers at each node represent bootstrap values from 1000 replications. NsRpoT-C is indicated in bold-face letters. The other RpoTs are as follows: AtRpoT;1, *Arabidopsis thaliana* RpoT;1 (Y08137); AtRpoT;2, *A. thaliana* RpoT;2 (AJ278248); AtRpoT;3, *A. thaliana* RpoT;3 (Y08463); maize RpoTp, *Zea mays* RpoTp (AF127022); maize RpoTm, *Z. mays* RpoTm (AF127021); CaRpoT, *Chenopodium album* RpoT (Y08067) and wheat-G, wheat RpoT-G (AF091838); wheat-C, wheat RpoT-C (U34402); NsRpoT-A (AB058954); NsRpoT-B (AB058957); moss RpoT1, *P. patens*RpoT1 (AB055214); moss RpoT2, *P. patens* RpoT2 (AB055215); T7, T7 RNA polymerase (M38308); T3, T3 RNA polymerase (P07659); Rpo41, *Saccharomyces cerevisiae* mitochondrial RNA polymerase (P13433). Cellular localization demonstrated experimentally is indicated as mitochondria (m), plastids (p), or both (m/p).

0.457). Thus, a clear prediction of plastid localization was made that coincides with the result from the phylogenetic tree (Fig. 2). In contrast, a short N-terminal sequence of 87 residues starting from the first methionine, the 59th residue, did not exhibit any organelle targeting properties.

Plastid targeting by NsRpoT-C

We further investigated the cellular localization of chimeric proteins comprised of the NsRpoT-C N-terminal sequence translationally fused to sGFP in protoplasts of tobacco leaf cells. The construct encoding the CTG-starting N-terminal 171 residues of NsRpoT-C produced GFP fluorescence clearly associated with chloroplasts (Fig. 4a, CTG-RpoT). A similar pattern was obtained with a positive control plasmid expressing plastid σ factor protein (Fig. 4e, Sig1). In contrast, the construct mutated at the CTG to CTC gave strong GFP fluorescence in the cytoplasm and nucleus but not in chloroplasts or mitochondria (Fig. 4b, CTc-RpoT). A similar fluorescence pattern was observed using the construct ATG_{59}-RpoT encoding a shortened transit peptide that begins at the first ATG codon (Fig. 4d). Mutation at the CTG to ATG resulted in chloroplast GFP fluorescence (Fig. 4c, aTG-RpoT). These results indicated that the CUG codon functions in translation initiation and the N-terminal sequence of NsRpoT-C functions as a plastid-targeting signal. We therefore concluded that initiation at the CUG is essential for specific targeting to the chloroplasts. Interestingly, the construct encoding the CTG-starting N-terminal sequence failed to give GFP fluorescence within chloroplasts from the moss *P. patens* (Fig. 4i). This implies that the translation initiation at the CUG did not occur in the moss. The *P. patens* RpoT mRNAs are known to possess a canonical AUG translation initiation codon [16].

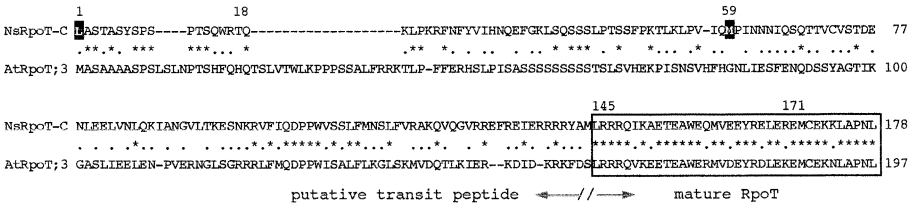

Fig. 3. Amino acid sequence alignment of N-terminal part of NsRpoT-C with that of *Arabidopsis* RpoT;3. Identical and conserved residues are indicated as asterisks and dots, respectively. The residues utilized for construction of plasmids or in silico prediction are numbered above the amino acid sequence.

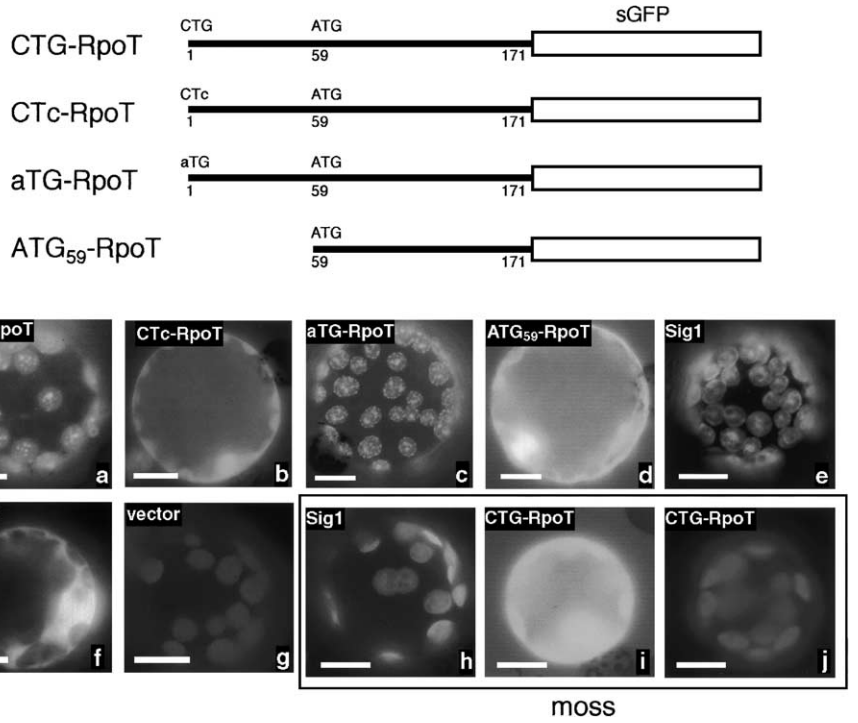

Fig. 4. Intracellular localization of the NsRpoT-C–GFP fusion proteins. (Top) Constructs used in transient expression of GFP fusion proteins in protoplasts. The length of N-terminal NsRpoT-C is indicated by amino acid numbers. Mutations introduced in the CTG codon of NsRpoT-C are indicated as lower-case letters. (Bottom) The constructs were expressed in the protoplasts of tobacco leaf cell (a–g) or moss protonemata (h–j). Images of green fluorescence pattern of NsRpoT-C–GFP fusions are shown for CTG-RpoT (a and i), CTc-RpoT (b), aTG-RpoT (c), and ATG_{59}-RpoT (d). Fluorescence images obtained with plastid σ factor transit peptide–GFP fusion protein (Sig1; e and h) and expression vector CaMV35S-sGFP(S65T)-nos3′ (vector; f) are indicated. GFP fluorescence was detected using a U-MWIBA/GFP cube (a–f, h, and i). The chloroplast pigments were detected using a U-MWIG cube for excitation and emission (g and j). Protoplasts were observed at a magnification of 600× with a fluorescence microscope. Bars are 10 μm.

During the preparation of this manuscript, Hedtke et al. [29] reported a result the same as ours showing that *N. tabacum* RpoT3 mRNA is initiated at a CUG codon.

Transcript levels of various plant organs

To compare the relative mRNA levels in different organs, semi-quantitative RT-PCR was performed. *NsRpoT-C* transcripts were detected in all organs examined and accumulated at the highest level in the young leaves (Fig. 5). This indicates that the *NsRpoT-C* gene is actively expressed in photosynthetic organs containing chloroplasts and NsRpoT-C protein could

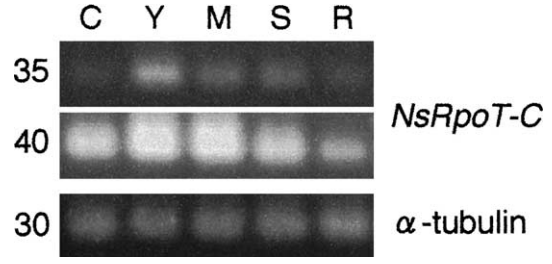

Fig. 5. Semi-quantitative RT-PCR analysis of *NsRpoT-C* transcripts in cotyledons (C), young leaves (Y), mature leaves (M), stems (S), and roots (R). Numbers on the left of each panel indicate the number of PCR cycles. PCR products amplified with primers for the *NsRpoT-C* transcript. Control PCR products for transcripts of the α-tubulin gene.

function as one of nucleus-encoded plastid RNA polymerases.

The present observations support that a set of three RpoT genes normally exists in dicotyledonous plants. NsRpoT-C probably specifies transcription of some plastid-encoded genes, but not photosynthesis-related genes.

Acknowledgments

This work was supported by Grant-in-Aid for Scientific Research on Priority Areas "Genome Biology" from the Ministry of Education, Culture, Sports, Science and Technology of Japan, from the Japan Society of the Promotion of Science (0927103), and from the Ministry of Agriculture, Forestry and Fisheries of Japan (BDP 02-I-1-7).

References

[1] N. Cermakian, T.M. Ikeda, P. Miramontes, B.F. Lang, M. Gray, R. Cedergren, On the evolution of the single-subunit RNA polymerases, J. Mol. Biol. 45 (1997) 671–681.
[2] P. Maliga, Two plastid RNA polymerases of higher plants: an evolving story, Trends Plant Sci. 3 (1998) 4–6.
[3] A. Weihe, T. Börner, Transcription and the architecture of promoters in chloroplasts, Trends Plant Sci. 4 (1999) 169–170.
[4] V. Tiranti, A. Savoia, F. Forti, M.F. D'Apolite, M. Centra, M. Rocchi, M. Zeviani, Identification of the gene encoding the human mitochondrial RNA polymerase (h-mtRPOL) by cyberscreening of the expressed sequence tags database, Hum. Mol. Genet. 6 (1997) 615–625.
[5] B.S. Masters, L.L. Stohl, D.A. Clayton, Yeast mitochondrial RNA polymerase is homologous to those encoded by bacteriophages T3 and T7, Cell 51 (1987) 89–99.
[6] B. Chen, A.R. Kubelik, S. Mohr, C.A. Breitenberger, Cloning and characterization of the *Neurospora crassa cyt-5* gene: a nuclear-coded mitochondrial RNA polymerase with a polyglutamine repeat, J. Biol. Chem. 271 (1996) 6537–6544.
[7] J. Li, J.A. Maga, N. Cermakian, R. Cedergren, J.E. Feagin, Identification and characterization of a *Plasmodium falciparum* RNA polymerase gene with similarity to mitochondrial RNA polymerase, Mol. Biochem. Parsitol. 113 (2001) 261–269.
[8] A. Weihe, B. Hedtke, T. Börner, Cloning and characterization of a cDNA encoding a bacteriophage-type RNA polymerase from the higher plant *Chenopodium album*, Nucleic Acids Res. 25 (1997) 2319–2325.
[9] B. Hedtke, T. Börner, A. Weihe, Mitochondrial and chloroplast phage-type RNA polymerases in *Arabidopsis*, Science 277 (1997) 809–811.
[10] B. Hedtke, T. Börner, A. Weihe, One RNA polymerase serving two genomes, EMBO Rep. 1 (2000) 435–440.
[11] D.A. Young, R.L. Allen, A.J. Harvey, D.M. Lonsdale, Characterization of a gene encoding a single-subunit bacteriophage-type RNA polymerase from maize which is alternatively spliced, Mol. Gen. Genet. 260 (1998) 30–37.
[12] C.C. Chang, J. Sheen, M. Bligny, Y. Niwa, S. Lerbs-Mache, D.B. Stern, Functional analysis of two maize cDNAs encoding T7-like RNA polymerase, Plant Cell 11 (1999) 911–926.
[13] T.M. Ikeda, M.W. Gray, Identification and characterization of T3/T7 bacteriophage-like RNA polymerase sequences in wheat, Plant Mol. Biol. 40 (1999) 567–578.
[14] Y. Kobayashi, Y. Dokiya, M. Sugiura, Y. Niwa, M. Sugita, Genomic organization and organ-specific expression of a nuclear gene encoding phage-type RNA polymerase in *Nicotiana sylvestris*, Gene 279 (2001) 33–40.
[15] Y. Kobayashi, Y. Dokiya, M. Sugita, Dual targeting of phage-type RNA polymerase to both mitochondria and plastids is due to alternative translation initiation in single transcripts, Biochem. Biophys. Res. Commun. 289 (2001) 1106–1113.
[16] Y. Kabeya, K. Hashimoto, N. Sato, Identification and characterization of two phage-type RNA polymerase cDNAs in the moss *Physcomitrella patens*: implication of recent evolution of nuclear-encoded RNA polymerase of plastids in plants, Plant Cell Physiol. 43 (2002) 245–255.
[17] Y. Miyata, C. Sugiura, Y. Kobayashi, M. Hagiwara, M. Sugita, Chloroplast ribosomal S14 protein transcript is edited to create a translation initiation codon in the moss *Physcomitrella patens*, Biochim. Biophys. Acta 1576 (2002) 346–349.
[18] N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees, Mol. Biol. Evol. 4 (1987) 406–425.
[19] Y. Niwa, T. Hirano, K. Yoshimoto, M. Shimizu, H. Kobayashi, Non-invasive quantitative detection and applications of non-toxic, S65T-type green fluorescent protein in living plants, Plant J. 18 (1999) 455–463.
[20] K. Hara, M. Sugita, S. Aoki, Cloning and characterization of the cDNA for a plastid sigma factor from the moss *Physcomitrella patens*, Biochim. Biophys. Acta 1517 (2001) 302–306.
[21] J. Sambrook, E.F. Fritsch, T. Maniatis, Molecular Cloning: A Laboratory Manual, second ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.
[22] M.A. Grachev, A.G. Pletnev, Phage T7 RNA-polymerase: gene cloning and its structure, Bioorg. Khim. 10 (1984) 824–843.
[23] W.T. McAllister, C.A. Raskin, The phage RNA polymerases are related to DNA polymerases and reverse transcriptases, Mol. Microbiol. 10 (1993) 1–6.
[24] R. Sousa, Structural and mechanistic relationships between nucleic acid polymerases, Trends Biochem. Sci. 21 (1996) 186–190.
[25] F. Fütterer, T. Hohn, Translation in plants—rules and exceptions, Plant Mol. Biol. 32 (1996) 159–189.
[26] J.L. Riechmann, T. Ito, E.M. Meyerowitz, Non-AUG initiation of *AGAMOUS* mRNA translation in *Arabidopsis thaliana*, Mol. Cell. Biol. 19 (1999) 8505–8512.
[27] O. Emanuelsson, H. Nielsen, S. Brunak, G. von Heijne, Predicting subcellular localization of proteins based on their N-terminal amino acid sequence, J. Mol. Biol. 300 (2000) 1005–1016.
[28] M.G. Claros, P. Vinvens, Computational method to predict mitochondrially imported proteins and their targeting sequences, Eur. J. Biochem. 241 (1996) 779–786.
[29] B. Hedtke, J. Legen, A. Weihe, R.G. Herrmann, T. Börner, Six active phage-type RNA polymerase genes in *Nicotiana tabacum*, Plant J. 30 (2002) 625–637.