Vol. 43, No. 2, October 1997 BIOCHEMISTRY *and* MOLECULAR BIOLOGY INTERNATIONAL
Pages 433–442

CHARACTERIZATION OF THE *orf*31-*pet*G GENE CLUSTER FROM THE PLASTID GENOME OF *Populus deltoides*

Sushma Naithani, Prabodh K. Trivedi* and Prafullachandra.V. Sane

Centre for Plant Molecular Biology, National Botanical Research Institute, Lucknow, INDIA-226001.

Received June 24, 1997
Received after revision July 18, 1997

SUMMARY: The *orf*31-*pet*G gene cluster is located approximately 1.2 kb away from the *psb*EFLJ operon in the chloroplast genome of *Populus deltoides*. The *orf*31 (*ycf*7) encodes an unidentified polypeptide while the *pet*G gene encodes subunit V of an important component, cytochrome b_6/f complex, involved in photosynthetic electron transport. We have determined the nucleotide sequence of the *orf*31-*pet*G gene cluster from the plastid genome of *a tree, Populus deltoides*. Our sequence analysis suggests that these genes possess high homology with the published sequences of these genes from other plants. Northern analysis suggests development dependent transcription of the *orf*31-*pet*G cluster in leaves.

Key words: *orf*31, *pet*G, nucleotide sequence, substitutions, transcription, *Populus deltoides*

INTRODUCTION

The cytochrome b_6/f complex transfers electrons from reduced plastoquinone to the soluble electron carrier, plastocyanin, in plant chloroplasts and cyanobacteria and to cytochrome$_{553}$ in most algae and some cyanobacteria (1,2,3). In mitochondria, nonphotosynthetic and purple photosynthetic bacteria, a structurally and functionally equivalent complex, cytochrome bc1 complex, acts as ubiquinol: cytochrome c oxidoreductase. However, in cyanobacteria, the cytochrome b_6/f complex is the sole complex which participates in both photosynthetic (noncyclic) and respiratory electron transport (3). Cytochrome b_6/f complex is composed of atleast five polypeptide species namely cytochrome f in two variant forms (33/34 kDa), cytochrome b_6 (23kDa), the Rieske-iron-sulfur protein (19kDa), subunit IV (17kDa) and subunit V (4kDa) (4,5). The genes encoding these components of cyt b_6/f complex in most higher plants and cyanobacteria are located on

*For correspondance
Fax: 91-522-282881/282849
Email: manager@nbri.sirnetd.ernet.in

1039–9712/97/020433–10$05.00/0

three separate operons of the chloroplast genome namely, *pet*A, *psb*B-*psb*H-*psb*T-*pet*B-*pet*D and *orf*31-*pet*G-*orf*42 (3,5).

In our earlier report (6), we had shown that the *orf*31-*pet*G gene cluster in *Populus deltoides* is located adjacent to *psb*EFLJ operon and transcribes divergently. This gene organization is conserved in *M. polymorpha* (7), *N. tabacum* (8), *O. sativa* (9), *Z. mays* (10) and in most higher plants. Exceptions are *Chlamydomonas eugametos* and *C. reinhardtii*, where *pet*G forms the *psb*F-*psb*L-*pet*G-*rps*3 operon (11). However, in *Pinus thunbergii* an open reading frame of 62 amino acids (*orf*62b) is located upstream of *pet*G and one of the transcripts originates from *orf*62b (12,13). Wakasugi *et al.* (1996) have detected three editing events in this transcript; one creates an initiation codon by changing ACG to AUG while the other produces a stop codon by changing CAA to UAA, thereby leading to the formation of a reading frame consisting of 33 codons (*orf*33). The predicted product of *orf*33 is homologous to that of *ycf*7 (*orf*31) from other plant species (13). The third editing event restores an internal codon (UUG) for the conserved Leu from the Ser codon (UCG) (13).

The *orf*31-*pet*G-*orf*42 operon has been shown to be transcribed as a polycistronic as well as a monocistronic message in maize and increased levels of primarily monocistronic *pet*G transcripts accumulate from an internal promoter upstream of the *pet*G gene (5). Most of the information available on this cluster is based on the nucleotide sequence and expression from different plants but not much is known from the tree species. In this paper, we present the complete nucleotide sequence of *orf*31-*pet*G gene cluster and its expression at transcript level from a tree species, *Populus deltoides,* commonly known as poplar.

MATERIALS AND METHODS

Cloning and sequencing of the *orf*31-*pet*G gene cluster

As reported earlier (6), the recombinant clone pPCYT from *Populus deltoides* (L. clone D121) containing three gene clusters namely *pet*A, *psb*EFLJ and *orf*31-*pet*G was identified and complete nucleotide sequence for *psb*EFLJ operon was established. General molecular genetic methods were employed as described by Sambrook *et al.* (1989) (14). The subclones and deletions of the clone pPCYT containing the *orf*31-*pet*G gene cluster were sequenced completely. Sequencing reactions of double stranded plasmids were carried out using the T7 sequencing kit (Pharmacia) and [^{35}S]-dATP (Amersham) as per manufacturer's instructions. The reaction products were resolved on a urea-acrylamide gel and autoradiographed for determining the sequence. Alternatively, the automated DNA sequencing system (373, ABI) was used for obtaining the sequence using the dye terminator cycle sequencing kit (Perkin Elmer).

Nucleotide substitution analysis

All gene sequences of *pet*G and *orf*31(*ycf*7) from other plants, except poplar, were extracted from the EMBL database and pairwise multiple alignments of nucleotide sequences in coding regions were carried out using the Clustal software in PC Gene release 16.0 (Intelligenetics Inc., USA). Sites of nucleotide substitutions were analysed as described earlier (6). Nucleotide substitutions in the coding regions are classified into synonymous and nonsynonymous substitutions; synonymous substitutions (SS) do not lead to amino acid substitutions, whereas nonsynonymous substitutions (NS) result in amino acid replacement.

RNA isolation and northern hybridizations

Total leaf RNA was prepared from leaves of *Populus deltoides*. Leaves were harvested and frozen in liquid nitrogen and RNA was extracted by guanidine isothiocyanate method (15). RNA was separated on a denaturating 1.3% agarose MOPS-formaldehyde gel and transferred to Zeta-probe membrane (Biorad). The RNA gel blots were hybridized with strand specific DNA probe in 250 mM sodium phosphate (pH 7.2), 7% SDS, 1mM EDTA, and 50% formamide at 42°C for 16 h. The probe used for northern hybridization was the homologous probe from poplar.

RESULTS AND DISCUSSION

Sequence analysis of the *orf*31-*pet*G gene cluster

We have earlier (6) shown the proximity of the *orf*31-*pet*G gene cluster with the *psb*EFLJ operon in poplar. This gene organization appears to be conserved in higher plants. In *Pinus thunbergii,* however, an orf of 62 amino acids (*orf*62b), as deduced from nucleotide sequence, is present upstream of *pet*G gene. This *orf*62b gets converted to *orf*33 post-transcriptionally by RNA editing. This *orf*33 is highly homologous to *orf*31 (*ycf*7) of higher plants (12). Chlamydomonas is an exception where *pet*G is a part of the *psb*F-*psb*L-*pet*G-*rps*3 operon (11).

Analysis of the sequence data of *orf*31-*pet*G gene cluster in poplar revealed the presence of two open reading frames of 93 bp and 111 bp corresponding to *orf*31 and *pet*G genes, respectively. These two genes are separated by an AT rich intergenic region of 182 bp as shown in figure #1. The intergenic region between *orf*31 and *pet*G is highly AT rich and contains a stem-loop secondary structure which is unique in poplar. This structure may play a role in the regulation of the *pet*G gene expression. In the *orf*31, putative ribosome binding site (RBS) is not present. However, on the other hand, Shine-Dalgarno-like sequence is present 43 nucleotide upstream from the initiation codon ATG of *pet*G (figure #1). The absence of SD-like sequence before the *orf*31 suggests the presence of other sequence motifs serving the purpose of the ribosome binding site as has already been shown

```
                                        orf31 -> M    L    T    I
            AAAAAAAAATAGTTTATTAGACCTTC      ATG  CCC  ACT  CTA

 T    S    Y    F    G    F    L    L    V    A    L    T
ACT  AGT  TAT  TTC  GGT  TTT  CTG  TTA  GTA  GCT  TTA  ACT

 I    T    S    A    L    F    I    S    L    S    K    I
ATA  ACT  TTA  GTT  CTA  TTT  ATT  AGT  CTG  AGC  AAG  ATA

 R    L    I
CGA  CTT  ATT TGA AATTAATTGAATGAACAATTCATAAAAAAAAAAAAAAGAA

TTTTTTTTTTTTTTGCAGTATTCCATTTTCTAGTTCCTTACCGTGTCAATTTCCAATT

CTTGGTTATTGAGATTTATGGGCAATATGCATTAATATTTAAGGATAGATATTACCTC
                                                RBS
                                  petG -> M    I    V   V    F    T
CTTTTTCCTCTTTTCAAATCAAATTGAA     ATG  ATT  GAA GTT TTT  CTA

 F    G    I    V    L    G    L    I    P    I    T    L
TTT  GGA  ATC  GTC  TTA  GGT  CTA  ATT  CCT  ATT  ACT  TTG

 A    G    L    F    V    T    A    Y    L    Q    Y    S
GCT  GGA  TTA  TTC  GTA  ACT  GCT  TAT  TTA  CAA  TAC  AGA

 R    G    N    Q    L    D    L
CGT  GGT  GAC  CAG  TTG  GAT  CTT   TGA TTAATTAATACCCTTTTTT

TTTGACCTCCTCCTTTTTTTAATCCACAGGAGGTCAAATTAAGATTGATGTTCAAGCTT
```

Figure #1. Nucleotide Sequence of *orf*31(*ycf7*)-*pet*G gene cluster. The deduced amino acid sequence is shown above the nucleotide sequence. The underlined portion of the sequence is the putative ribosome binding site (RBS) and the bold letters represent putative promoter sequence ("-10" and "-35") motifs.

in the case of *psb*A (16). The typical "-10" and "-35 " motifs are present upstream of the *pet*G coding region (highlighted in figure #1) and may act as internal promoter sequences for this gene.

Analysis of the nucleotide substitutions in the coding regions

To understand molecular evolution, it is important to know the number and pattern of nucleotide substitutions between homologous sequences. An analysis of the SS and NS in *orf*31 (*ycf7*) and *pet*G coding regions among the various reported sequences is shown in figure #2. In both cases, SS are distributed throughout the entire length of the genes. The rate of SS is reported to be related to codon usage (17-19) and sequence composition (20). Codon usage in *E. coli* as well as in *Saccharomyces cerevisae* is correlated with tRNA abundance (21). The level of expression of a gene has been suggested to correspond to the degree to which codon usage is biased in that gene. Selection at codon level is presumed to influence translational efficiency and accuracy (22). The variation in the rate of SS is shown to be

pet G

*orf*31*(ycf7)*

Figure #2. Distribution of synonymous (SS) and nonsynonymous (NS) nucleotide substitutions in the *orf*31(*ycf7*) and *pet*G genes based on pairwise alignments of nucleotide and amino acid sequences. For analysis of *pet*G, sequences of poplar (this study), tobacco(Z00044), spinach(X74430), sugarbeet(X87637), maize(J04502), pea (X05395), rice (X15901), marchanctia(X04465), cyanophora (X16974) and chlamydomonas (L29282) and of parasitic plant cuscuta (X61698) were used from EMBL database. In case of *orf*31 sequences of poplar, tobacco, maize, beet, rice, and pine(D17510) were used from EMBL database.

related with the level of gene expression (17,19). Highly expressed genes of the *M. polymorpha* chloroplast genome have been reported to show bias for certain synonymous codons (23).

NS were further analysed for replacement with similar residues viz. those that can interchange with each other without modifying the biological activity of a protein (24). The following groups have been considered as similar residues (S), (i) A,S,T,G and P, (ii) D and E, (iii) N and Q, (iv) R and K, (v) I,L,M and V, (vi) F,Y, and W. Residues which belong to different groups but still do not differ in charge were considered semi-conserved (Sc). Those differing in group as well as charge were considered as non-conserved (N). The analysis of NS with reference to the location on the polypeptide(s) is shown in table 1. In *orf*31 (*ycf7*), out of 16 NS only one is N, seven are S and eight are Sc. Similarly, in *pet*G out of 16 NS, six are N, seven are S and three are Sc. In both cases, N are located either at amino and/or at the carboxy-terminal end(s) of polypeptide(s) and, therefore, can tolerate such

Table 1: List of nonsynonymous substitutions in *ycf7* (*orf31*) and *petG* genes

Gene	Amino acid position in putative polypeptide	Amino acid substitutions	Nature of substituttions
ycf7 (*orf31*)	2	L,P	Sc
	4	I,L	S
	5	T,I	Sc
	9	G,L	Sc
	12	L,I	S
	13	A,G,V	Sc
	14	A,S	S
	15	L,I	S
	16	T,I	Sc
	17	I,L	S
	18	T,A	S
	19	S,L,P	Sc
	20	A,V	Sc
	24	G,S	S
	26	N,S	Sc
	29	R,Q	N
petG	2	I,V	S
	4	L,A,V	Sc
	5	F,L	N
	6	F,S	N
	14	I,V	S
	16	I,V	S
	18	L,I	S
	19	A,I,L	Sc
	20	G,P	S
	23	V,L	S
	24	T,A	S
	33	N,D	Sc
	34	Q,A	N
	35	L,F	N
	36	D,T,E	N
	37	L,F	N

Translational frames of *ycf7* (*orf31*) and *petG* genes from different plants(refer figure #2) were compared to score the amino acid positions with different substitutions. For S, Sc and N refer text.

changes without affecting functional integrity of the polypeptide. As we have reported earlier (6), most of the N in PSII (photosystem II) genes are restricted to the amino and/or carboxy-terminal ends and therefore the location of N sites seems to be under functional-structural and evolutionary constraints. Most of the NS in *petG* and *ycf7* are more or less silent ones and do not cause any major change in protein structure.The Sc type of NS in *petG* specially separates higher plants from the lower phototrophs.

Development-dependent transcription of *orf*31-*pet*G gene cluster

Leaves from different nodal positions were taken as representatives of different developmental stages for northern analysis using total RNA. Northern blot hybridization with *orf*31-*pet*G strand-specific probe shows differential transcript pattern (figure #3). The young leaves corresponding to first internode position show only a single transcript of 220 nucleotides while the mature leaves at the 4^{th} internode position show two transcripts of 220 nucleotides and 1.7kb. The old leaves at 12^{th} internode position show only the longer transcript of 1.7 kb (figure #3). The 1.7 kb transcript is a primary polycistronic transcript for this gene cluster while the smaller one could be either a processed transcript or may have arisen from an internal promoter located upstream of the *pet*G gene. The presence of different sized transcripts in young, mature and old leaves indicates development-dependent differential accumulation of these transcripts. The presence of only a 220 nucleotide transcript in young leaves may be attributed to the functioning of

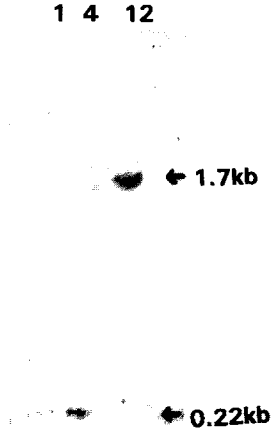

Figure #3. Northern analysis using total RNA isolated from different developmental stages of the leaves (from nodal position 1, 4 and 12). 22μg of total RNA isolated from each developmental stage of the leaves was electrophoresed on a denaturating agarose gel and transferred onto Zeta-probe membrane. Homologous probe from poplar was used for hybridization. Sizes of the transcripts are shown in kb.

an internal promoter for selective transcription of *pet*G whose product may be needed in early stages of chloroplast development. The results also suggest that the *orf*31 product is apparently not required in the early stages of leaf development. It is known that in the early stages of chloroplast development in light, electron transport complex proteins (PSI, PSII and $cytb_6/f$ complex) are synthesized rapidly (25,26). To constitute these complexes in chloroplasts, selective promoters may be used initially to transcribe the essential genes at this stage. However, in the mature leaf, the *orf*31 gene product may be required for some unknown function and this requirement may be met in later stages through polycistronic transcript.

The absence of any other smaller transcript except a transcript of 220 nucleotides suggests that this polycistronic transcript is not processed further as in case of the *psb*EFLJ operon (6). The 220 nucleotide transcript is probably generated through the internal promoter of *pet*G. As shown in figure #1, "-10" and "-35" sequence elements are present 5' upstream of the *pet*G coding region and may act as an internal promoter. A similar situation has also been observed in the case of *psb*D/C operon in tobacco, where the internal promoter for *psb*C is active along with the *psb*D/C promoter present upstream of *psb*D gene (27). The presence of a polycistronic and a monocistronic transcript for *orf*31-*pet*G-*orf*42 operon has been reported in maize (5). The polycistronic transcript in maize is further processed into smaller transcripts. The initiation of monocistronic *pet*G transcript from internal *pet*G promoter was also experimentally demonstrated by mapping the 5' terminus of the transcript in maize (5). Development-dependent change in RNA polymerase activity may lead to initiation of transcription from new loci with the help of some developmentally induced transcription factors and may also be responsible for differential accumulation.

In rice, six transcripts of sizes 3.0 kb, 2.8 kb, 1.7kb, 1.3 kb, 0.5 kb and 0.3 kb, hybridizing to *orf*31-*pet*G region were reported (28). Development-dependent changes in transcription of various plastid genes is well studied in various systems. In tomato, the *rbc*S and *cab* transcript levels were high in 4cm (young) leaves, while transcripts of *rbc*L and *psb*A accumulated to high levels in 6cm (mature) leaves (29). In *Phaseolus vulgaris,* dot blot and northern hybridization studies indicated that the transcript levels for *cab* and *rbc*S decreased with maturation of leaves while the *psb*A transcript was present in higher proportion of total RNA of 14, 21 and 28-day old tissue than in 7-day old tissue (14). During maturation of chloroplasts, a decrease in overall transcriptional activity was detected in barley (30), spinach (31), tobacco (32) and wheat (33). Earlier, we have reported development-dependent differential accumulation of the *psb*EFLJ tetracistronic transcript in poplar (6).

Our studies on *orf*31(*ycf*7)-*pet*G operon suggest that its organization, nucleotide and deduced amino acid sequence in *Populus deltoides* is highly homologous to other land plants studied. However, the non-coding spacer region is variable with respect to other plants. The northern analysis suggests a development induced transcript pattern for *orf*31-*pet*G transcripts. In young leaves, the *pet*G transcript is preferentially synthesized probably to meet the demands of the *pet*G product while in mature leaves the polycistronic transcript is preferentially accumulated. It is tempting to suggest that the product of *orf*31(*ycf*7) is specifically required in mature leaves. It is not clear whether the product is a specific kinase, or a protease/phosphatase which are found to be active in mature leaves leading to senecence but sequence analysis (of the deduced amino acid sequence) does not show presence of signatures of any of these. The small polypeptide encoded by the gene may be associated with some membrane-bound complex as it has been predicted that *orf*31(*ycf*7) gene product possesses a single trans-membrane domain.

ACKNOWLEDGEMENTS

We thank Dept. of Biotechnology, New Delhi for the financial support to carry out this work and the Council of Scientific and Industrial Research for Senior Research fellowship to SN.

References

1. Bendall, D.S. (1982) Biochim. Biophys. Acta 683,119-151.
2. Hurt, E. and Hauska, G. (1981) Eur. J. Biochem. 117,591-599.
3. Koike, H. and Satoh, K. (1996) J. Sci.& Ind. Res.55, 564-582.
4. Hurt, E. and Hauska, G. (1982) J. Bioenerg. Biomembr. 14, 405-424.
5. Haley, J. and Bogorad, L. (1990) Plant Cell 2, 323-333.
6. Naithani, S., Trivedi, P.K., Tuli, R. and Sane, P.V. (1997) J. Genet. (in press).
7. Ohyama, K., Fukyzawa, H., Konchi, T., Shirai, H., Sano ,T., Sano, S., Umesono, K., Shiki,Y., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H., Ozeki, H. (1986). Nature 322, 572-574.
8. Shinozaki, K., Ohme, M.,Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi,T., Zaita, N., Chunwongse J., Obokata,J., Shinozaki, K.Y., Ohta, C., Torazawa, K., Meng, B.Y., Sugita, M., Deno, H., Kamogashira T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh N., Shimada, H., Sugiura, M.(1986) EMBO J. 5, 2043-2049.
9. Hiratsuka, J., Shimada, H., Whitter, R.F., Ishibashi, T., Sakamato, M., Mori M., Kando, C., Honjii Y., Sun, C.R., Meng, B.Y., Li,Y., Kanno, A., Nishizawa Y., Hirai A., Shinozaki, K., Sugiura, M.(1989) Mol. Gen. Genet 217, 185-194.
10. Maier, R.M., Neckermann, K., Igloi, G.L., Kossel, H. (1995) J. Mol. Biol. 251,614-628.
11. Turmel, M. and Otis, C.(1994) Curr. Genet. 27, 54-61.
12. Wakasugi, T., Tsudzuki, J., Ito,S., Nakashima, K., Tsudzuki,T., and Sugiura, M. (1994) Proc. Natl. Acad. Sci. USA. 91, 9794-9798.
13. Wakasugi,T., Hirose,T., Horihata, M.,Tsudzuki,T., Kossel, H. and Sugiura, M.(1996) Proc. Natl. Acad. Sci. USA. 93, 8766-8770.

14. Sambrook,J., Fritsch, E.F., Maniatis, T.(1989) Molecular cloning: a laboratory manual 2nd ed.(New york, Cold Spring Harbor Laboratory Press.) Cold Spring Harbor.
15. Bate, N.J, Rothstein, S.J., and Thompson, J.E (1991) J.Exp. Bot. 42, 801-811.
16. Hirose,T. and Sugiura, M. (1996) EMBO J. 15, 1687-1695.
17. Sharp, P.M. and Li, W.H. (1987) Nucleic Acids Res.15, 1281-1295.
18. Shields, D.C., Sharp, P.M., Higgins, D.G. and Wright, F.(1988) Mol. Biol. Evol.5, 704-716.
19. Sharp, P.M. (1991) J. Mol. Evol.33, 23-33.
20. Moriyama, E.N. and Gojobori, T. (1992) Genetics 130, 855-864.
21. Ikemura, T.(1985) Mol. Biol. Evol. 2, 13-34.
22. Dix, D.B. and Thompson, R.C.(1989) Proc.Natl.Acad. Sci. USA. 86, 6888-6892.
23. Morton, B.R (1994) Mol. Biol. Evol. 11, 231-238.
24. Meyers, E.W. and Miller, W.(1988) Comp. App. Biosci. 4, 11-17.
25. Klein, R.R., Gamble, P.E., and Mullet, J.E.(1988) Plant Physiol. 88, 1246-1256.
26. Mullet, J.E. (1988) Annu.Rev. Plant Physiol. plant Mol. Biol. 39, 475-502.
27. Yao, W.B., Meng, B.Y., Tanaka M. and Sugiura, M. (1989) Nucleic Acids Res.17, 9583-9591.
28. Kanno, A. and Hirai, A. (1993) Curr. Genet. 23,166-174.
29. Piechulla, B.(1998) Plant Mol. Biol.11, 345-353.
30. Krupinska, K.(1992) Planta 186, 294-303.
31. Deng, X.W. and Gruissem W (1987) Cell 49, 379-387.
32. Wollgiehn, R., Lerbs, S. and Munche, D.(1976) Biochem. Physiol. Pflanzen. 170, 381-387.
33. Patterson, B.D. and Smillie R.M.(1971) Plant. Physiol. 47,196-198.