Mol Gen Genet (1991) 226:345–352
0026892591001156

MGG

Duplication of the *psbA* gene in the chloroplast genome of two *Pinus* species

Jonas Lidholm[1], Alfred Szmidt[2], and Petter Gustafsson[1]

[1] Department of Plant Physiology, University of Umeå, S-901 87 Umeå, Sweden
[2] Department of Forest Genetics and Plant Physiology, Faculty of Forestry, Swedish University of Agricultural Sciences, S-901 83 Umeå, Sweden

Summary. The *psbA* gene, encoding the D1 protein of photosystem II, was found to be duplicated in the chloroplast genome of two pine species, *Pinus contorta* and *P. banksiana*. Analysis of cloned overlapping restriction fragments of *P. contorta* chloroplast DNA showed that the two *psbA* genes have the same orientation and are separated by approximately 3.3 kb. The nucleotide sequences of the coding and the upstream regions of the two *psbA* copies were found to be identical, whereas the downstream sequences diverged from a point 20 bp 3′ of the stop codons. Downstream of the gene copy designated *psbAII*, a dyad symmetry which allows the formation of a strong mRNA hairpin structure, and a *trnH* gene were found. No such elements, which are characteristic of *psbA* downstream regions, were found 3′ of *psbAI*. This suggests that *psbAII* is the ancestral gene copy in *P. contorta*. Upon comparison with *psbA* from other plants, the pine 353-codon sequence appeared almost as distant from the angiosperm as from the liverwort counterpart. As compared to tobacco, 14 substitutions in the predicted amino acid sequence were found, most of which were located in the terminal regions of the protein.

Key words: Chloroplast genome – Conifer – Gene duplication – Gymnosperm – *psbA* gene

Introduction

The chloroplast genomes of the vast majority of vascular plants are circular molecules in the size range of 120–180 kb and are composed of two inverted repeat segments separating two single-copy regions. The genetic content of chloroplast genomes is highly conserved and includes genes for protein components involved in photosynthetic processes, RNA polymerase subunits, ribosomal and transfer RNAs and ribosomal proteins. Many of the chloroplast genes are organized into phylogenetically conserved, cotranscribed clusters. Most of the genes verified as encoding photosynthetic components are located in the large single-copy region whereas the rRNA genes are confined to the inverted repeats (reviewed by Palmer 1985a). The genes for tRNAs, ribosomal proteins and approximately 40 open reading frames (ORFs), encoding putative gene products of unknown or uncertain identity and function, are scattered throughout the genome (Shinozaki et al. 1986; Ohyama et al. 1988a; Hiratsuka et al. 1989). Sequence homologies to other known genes, e.g. human mitochondrial NADH dehydrogenase and bacterial iron-sulfur ferredoxin (Shinozaki et al. 1986; Ohyama et al. 1986, 1988b; Hiratsuka et al. 1989) have been found among these ORFs.

The exceptions to the consensus gross structure known, to date, among vascular plants include the chloroplast genomes of a few legume genera (reviewed by Palmer 1985b) and conifers (Lidholm et al. 1988; Strauss et al. 1988; White 1990), which lack the inverted repeat organization, and *Geranium hortorum*, in which the inverted repeats are significantly enlarged and encompass several genes that are normally located in the single-copy regions (Palmer et al. 1987). In general, however, it appears that no sequence information has actually been gained or lost as a result of the rearrangements producing these aberrant genomes.

Apart from genes contained within the large inverted repeats, duplications or multiple copies of coding sequences are virtually absent from the chloroplast genomes. The presence in the chloroplast genome of small repeated elements has been reported for a few species, e.g. wheat (Howe 1985; Bowman and Dyer 1986; Howe et al. 1988), clover (Palmer 1985b) and Douglas-fir (Tsai and Strauss 1989). Copies of such short repeated sequences (Howe 1985; Tsai and Strauss 1989) as well as tRNA genes, chimeric or partially duplicated (Quigley and Weil 1985; Howe et al. 1988; Shimada and Sugiura

Offprint requests to: J. Lidholm
The sequence presented in this manuscript appears in the EMBL Data Library under the accession numbers X53721 (*psbAI*) and X53722 (*psbAII*)

1989), have been found at the endpoints of large inversions. Hence, both these kinds of elements have been suggested to play a role in the generation of inversions by acting as substrates for recombination.

In this paper we describe a duplication in *Pinus contorta* and *P. banksiana* of a DNA segment which contains the entire *psbA* gene, encoding the D1 reaction center protein of photosystem II. We report the cloning and sequencing of the two gene copies from *P. contorta* and their positions relative to each other.

Materials and methods

Plant material, bacterial strains, vectors and probes. Seedlings of *P. contorta* Dougl. (lodgepole pine), *P. banksiana* Lamb. (jack pine), *P. aristata* Engelm. (bristlecone pine), *P. strobus* L. (eastern white pine), *P. monticola* Dougl. (western white pine), *P. flexilis* James (limber pine), *P. sylvestris* L. (Scots pine) and *P. mugo* Mill. (Swiss mountain pine) were grown in a climate chamber to 4 months of age before harvesting. The vector used for the libraries as well as for subcloning was pUC19 (Yanisch-Perron et al. 1985) and the *Escherichia coli* strain used was DH5α, a derivative of DH1 (Hanahan 1983). The heterologous *psbA*-specific probe used in this study was the 769 bp *Pst*I-*Xba*I insert from the spinach clone pPSII32/1, equivalent to amino acids 87–341 of the D1 polypeptide (a gift of Dr. R.G. Herrmann, Munich, FRG).

DNA analysis, cloning and sequencing. Chloroplast DNA was extracted from pines as described (Szmidt et al. 1987). Plasmid preparations were carried out using the alkaline lysis method (Sambrook et al. 1989). Restriction endonucleases and modifying enzymes were used under the conditions recommended by the supplier (Pharmacia, Sweden). DNA fragments were separated by electrophoresis in TAE buffer (Sambrook et al. 1989) through agarose gels containing 0.5 μg/ml ethidium bromide. Blot transfer of DNA fragments from agarose gels and UV crosslinking to nylon membranes were performed according to the manufacturer's recommendations (Amersham, UK). Southern blots and colony membranes were hybridized to radioactive probes at 65° C in 6 × SSC, 2 × Denhardt's solution, 2 mM EDTA, 0.2% SDS (Sambrook et al. 1989). All filters were subjected to high-stringency washing at 60° C in 0.1 × SSC, 0.1% SDS. In order to remove bound probe prior to re-hybridization, membranes were boiled in 0.1 × SSC, 0.1% SDS for 5–7 min in a microwave oven. DNA fragments were prepared from agarose gels by capture on DEAE cellulose (NA45, Schleicher and Schuell, FRG) as described (Sambrook et al. 1989). Radiolabelled probes were prepared by the random priming procedure (Feinberg and Vogelstein 1983) and subsequently purified by chromatography through Sephadex G-50 spun columns (Sambrook et al. 1989). Libraries of *Bam*HI and *Hin*dIII fragments of chloroplast DNA from *P. contorta* were made by ligating 0.8 μg of digested chloroplast DNA to 0.4 μg of linearized, phosphatase-treated vector and transforming *E. coli* cells made competent by the RbCl method (Hanahan 1985). For sequencing, both random fragments and defined restriction fragments were subcloned. Random fragments for sequencing were produced by sonication (20 × 10 s at setting 2, using a Branson Sonifier B15 fitted with a microtip) of pre-ligated fragments covering the regions to be sequenced. After repairing the ends with 2 units of mung bean nuclease per μg of DNA at 15° C for 30 min, fragments in the size range of 250–500 bp were prepared and cloned into the *Sma*I site of pUC19. Sequencing of double-stranded plasmid templates, using α-^{35}S-dATP, was performed essentially according to Chen and Seeburg (1985). Sequencing extension products were separated in 6% polyacrylamide gels containing 8 M urea. The DNA sequences obtained were processed by computer programs included in the GENEUS software package (Harr et al. 1986).

Results

Detection and confirmation of a psbA gene duplication

Southern blot analysis of the eight pine species, using a *psbA*-specific probe, resulted in two different kinds of hybridization patterns (Fig. 1). In each of the *Eco*RI, *Sac*I and *Bam*HI digests of chloroplast DNA from *P. contorta* and *P. banksiana*, two equally intense hybridization signals occurred. In contrast, only one band could be seen in samples from the other species. The weak additional band in the *Eco*RI digest of *P. mugo* chloroplast DNA was later found to be the result of incomplete digestion. In the *Hin*dIII digests, the probe detected three fragments in the *P. contorta* and *P. banksiana* samples, whereas either one or two bands were seen in the other species. The hybridization patterns observed suggested a duplication of the *psbA* gene in *P. contorta* and *P. banksiana*, but could also be due to an inversion isomerism, a polymorphism or heteroplasmy in these two pine species.

In order to verify the gene duplication, we decided to clone and analyze overlapping restriction fragments hybridizing to the *psbA* probe. For this purpose, libraries of *Bam*HI and *Hin*dIII fragments of *P. contorta* chloroplast DNA were constructed, from which we isolated *psbA*-containing clones, as identified by colony hybridization. The clones obtained from the library of *Bam*HI fragments harbored inserts corresponding to either of the two 3.8 and 2.8 kb *Bam*HI fragments previously seen in the Southern blots. Two clones, designated pPCB121 and pPCB932, which contained the larger and the smaller of these fragments, respectively, were chosen for further analysis. From restriction maps of these clones and from the results of re-probing the *P. contorta* Southern blot with appropriate subfragments of the pPCB121 and pPCB932 inserts, it was established that the 4.3 kbp *Hin*dIII fragment detected by the spinach *psbA* probe overlapped the two *Bam*HI fragments. A clone, designated pPCH157, containing this fragment was isolated from the library of *Hin*dIII fragments. After

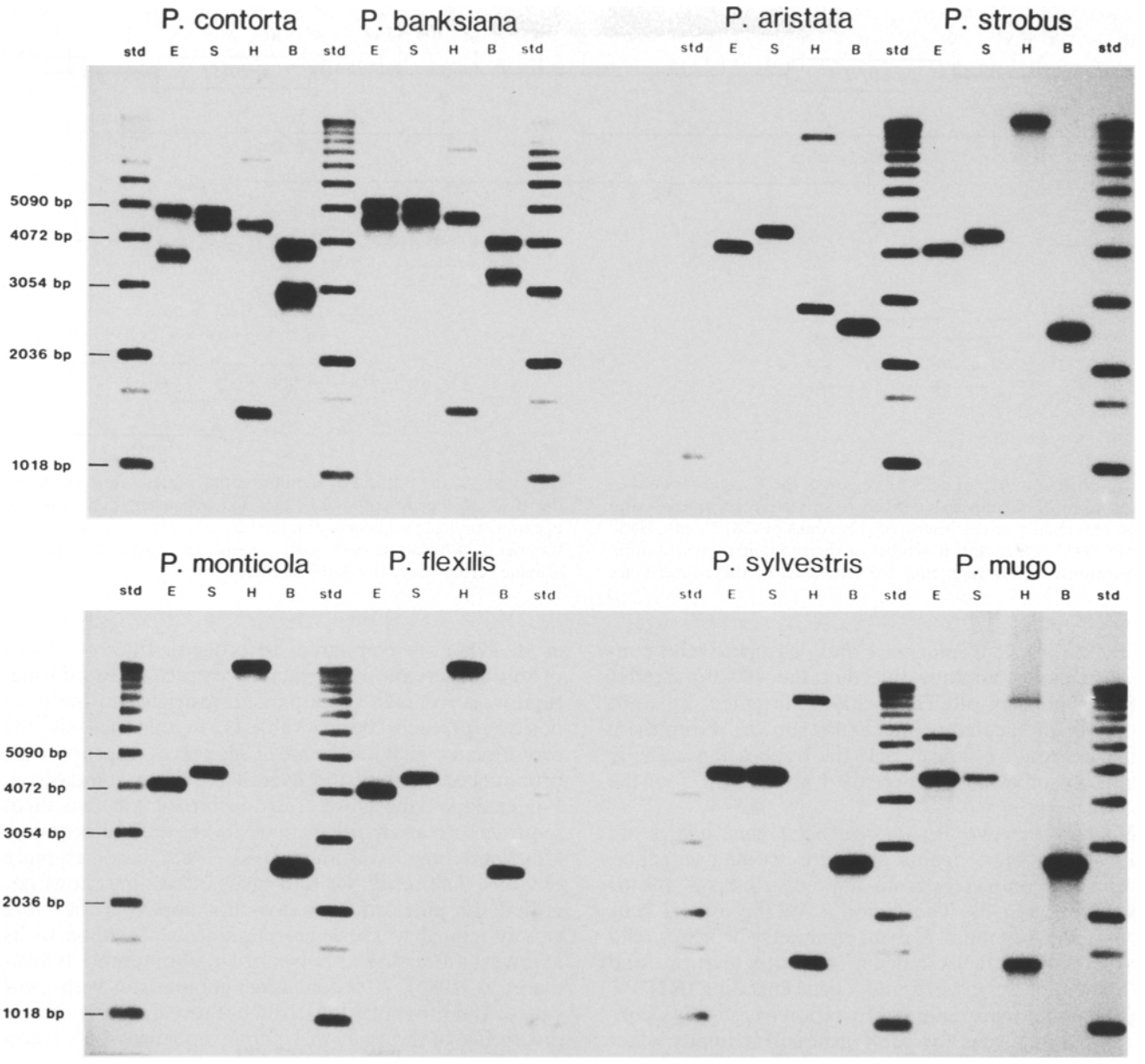

Fig. 1. Southern blot analysis of chloroplast DNA from eight *Pinus* species, using a spinach *psbA*-specific probe. The DNA samples were digested with restriction enzymes *Eco*RI (E), *Sac*I (S), *Hin*dIII (H) and *Bam*HI (B). The molecular size standard (std) is the BRL 1 kb ladder

restriction mapping and Southern blot analysis of pPCH157 it was possible to align the three clones to derive a restriction map of a 7.4 kb region of the chloroplast genome (Fig. 2). On the basis of this map we could conclude that the observed hybridization pattern for *P. contorta* was due to a duplication of a segment containing at least part of the *psbA* gene. As judged from hybridization data, this is also true for *P. banksiana* (not shown).

Sequence analysis

A number of cloned restriction and sonication subfragments of pPCB121 and pPCB932 were used for sequencing the regions indicated by arrows in Fig. 2. It emerged that the inserts of pPCB121 and pPCB932 contained identical open reading frames of 353 codons (Fig. 3), highly homologous to the tobacco (*Nicotiana tabacum*) *psbA* gene (Shinozaki et al. 1986). It was also found that the 5′ flanking sequences, extending to the proximal *Kpn*I sites, were identical. In addition to confirming the *psbA* gene identity of the ORFs, the sequence analysis established that the duplication includes all of the *psbA* gene. The position on the restriction map of the two *psbA* genes, designated *psbAI* and *psbAII*, is shown in Fig. 2. They are located on the same strand at an intergenic distance of approximately 3.3 kb. Further hybridization experiments showed that the 700 bp and 500 bp *Bam*HI/*Kpn*I subfragments from pPCB121 and

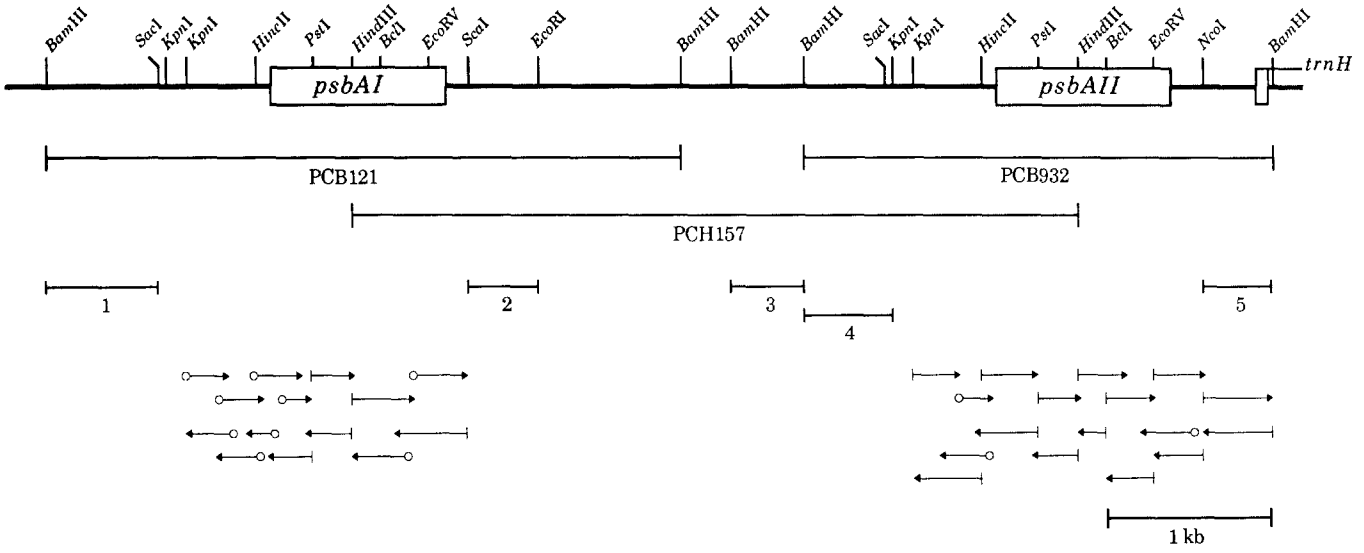

Fig. 2. Restriction map of a 7.4 kb region of the *P. contorta* chloroplast genome. The map was derived from restriction analysis and partial sequencing of the inserts of the clones pPCB121, pPCH157 and pPCB932, which are shown below the genomic representation. The positions of subfragments 1–5 (see text) of these inserts are indicated. At the bottom of the figure, the strategy for sequencing the two *psbA* gene copies is illustrated. *Open circles* at the end of *arrows* indicate sequences obtained from random sonication subfragments, whereas *arrows* with *vertical bars* indicate sequences obtained from restriction subfragments

pPCB932 (Fig. 2, fragments 1 and 4, respectively) contain sequence homology but that the 400 bp *Bam*HI subfragment from pPCH157 (Fig. 2, fragment 3) is not part of the duplicated segment (not shown). From these results we could establish that the duplication actually extends to somewhere between 0.7 and 1.2 kb 5′ of the *psbA* start codon.

Whereas the two *P. contorta psbA* gene copies are identical over the 5′ region and entire coding sequence, the diverge completely from a point 20 bp 3′ of the stop codon (Fig. 3). The region 3′ of the *psbAII* gene was found to contain a dyad symmetry which would allow a stable mRNA hairpin structure to form, and 400 bp further down we found a gene encoding $tRNA^{His}$ (Fig. 3). The downstream organization of *psbAII* is similar to that reported for *psbA* genes from many other plants. No such structures were found downstream of *psbAI*, suggesting that *psbAII* is the ancestral member of the gene pair, *psbAI* being the duplication product. This was substantiated by the results from rehybridizations of the Southern blot of chloroplast DNA from *P. sylvestris*, which lacks the *psbA* duplication (Fig. 1), using subfragments covering the regions downstream of the *psbA* coding sequences as probes. We used the 420 bp *Sca*I-*Eco*RI fragment from pPCB121 (Fig. 2, fragment 2) as a *psbAI*-specific probe and the 430 bp *Nco*I-*Bam*HI fragment from pPCB932 (Fig. 2, fragment 5) as a probe specific for *psbAII*. Whereas the *psbAII*-specific probe hybridized to all of the restriction fragments detected by the spinach *psbA* probe, the *psbAI*-specific probe hybridized to none of these (not shown).

A DNA sequence comparison of the *psbA* coding sequence from *P. contorta* and a number of other plants is shown in Table 1. The overall similarity is very high, almost 80%, even when such distant species as tobacco and the green alga *Chlamydomonas reinhardtii* (Erickson et al. 1984) are compared. In order to filter out some of this conservation and thereby extract more information, we also made a comparison restricted to the third position of each codon (Table 1). In this analysis, the evolutionary distance between sequences appears more pronounced than in the overall sequence comparison. For example, the closer relationship of the two dicot sequences to each other than to the barley sequence (*Hordeum vulgare*, Efimov et al. 1988) becomes more obvious. The result we find most interesting, however, is that the pine sequence does not appear to be more closely related to the angiosperm sequences than to its liverwort (*Marchantia polymorpha*) counterpart (Ohyama et al. 1986). A codon usage comparison with *psbA* genes from other plant taxa did not reveal any significant divergence of the pine *psbA* coding sequence and is therefore not shown.

The predicted amino acid sequences of the D1 proteins from different species show extremely high similarity. The pine and the tobacco sequences, for example, differ at only 14 positions (Fig. 3). The variability that does exist generally occurs in the C-terminal region which is removed during protein maturation (Takahashi et al. 1988). It is also in this region where most of the substitutions in the deduced pine D1 amino acid sequence, as compared to other species, were found (Fig. 3). None of the substitutions that yield resistance to the *s*-triazine and urea classes of herbicides (see Svensson et al. 1990 for a compilation) were found in the derived pine D1 sequence.

The 52 bp dyad symmetry found downstream of *psbAII* in *P. contorta* (Fig. 3) contains no mismatches and gives a value for the free energy of dissociation (ΔG) of -108 kcal/mol, as calculated by the algorithm of Tinoco et al. (1973). This implies the formation of an RNA hairpin structure which is significantly more stable

Fig. 3. Nucleotide sequence between the *psbA*-proximal *Kpn*I site and the *Sca*I site of clone pPCB121 and between the corresponding *Kpn*I site of clone pPCB932 and the *Bam*HI site at the end of the insert. The two sequences are shown as one where they are identical. The deduced amino acid sequence of the D1 protein of *P. contorta* is shown above the *psbA* coding sequence. The amino acid residues of the tobacco D1 polypeptide (Shinozaki et al. 1986) are shown at positions where the two sequences differ; at all other positions they are identical. The two *underlined* residues Leu5 and Glu231 in the tobacco sequence are virtually invariant in the derived D1 sequences from other species but are substituted by Ile and Gln, respectively, in *P. contorta*. The *head-to-head arrows* below the PCB932 sequence indicate a dyad symmetry which allows the formation of an RNA stem-loop structure with a calculated free energy of −108 kcal/mol. Near the end of the PCB932 sequence (downstream of *psbAII*), the gene for $tRNA^{His}$ (*trnH*) is indicated

than the corresponding structures of *psbA* messages from other plants, e.g. that of tobacco with a calculated ΔG of −33 kcal/mol. Also, we found no sequence similarity between the dyad symmetries of *P. contorta* and tobacco; only the structure is conserved.

The regions upstream of the *psbA* genes in *P. contorta* were found to contain sequence elements identical or very similar to the *psbA* −35 and −10 promoter boxes and the ribosome binding site in other species. An alignment of the *psbA* upstream sequences of four species, representing distant taxa, revealed considerable similarities also outside these elements (Fig. 4). This is illustrated by the consensus sequence shown at the bottom of the figure. The function of the conserved sequences outside the promoter boxes upstream of the *psbA* gene has not been systematically studied. The distance between the −10 box and the ribosome binding site is 55 and 50 bp in tobacco and barley, respectively, whereas it is only 23 bp in *P. contorta* and 24 bp in liverwort.

Discussion

Chloroplast genomes characteristically show a low rate of evolution, both in terms of nucleotide substitutions

Table 1. Percentage similarities in overall sequence comparisons and at third codon positions of the *psbA* genes from various species

	Pinus	Tobacco	*Amaranthus*	Barley	Liverwort	*Chlamydomonas*
Pinus		88.4	87.6	88.7	87.5	78.3
Tobacco	71.4		97.1	92.8	87.7	79.5
Amaranthus	69.4	92.1		92.5	87.4	80.8
Barley	72.5	81.6	80.5		87.3	79.8
Liverwort	70.3	68.3	66.6	66.0		81.5
Chlamydomonas	54.5	55.7	59.1	57.4	62.8	

Overall

3 rd position

Sources of *psbA* sequences: *Pinus contorta,* this work; tobacco, Shinozaki et al. (1986); *Amaranthus hybridus,* Hirschberg and McIntosh (1983); barley, Efimov et al. (1988); liverwort, Ohyama et al. (1986); *Chlamydomonas reinhardii,* Erickson et al. (1984)

and, in particular, with respect to large-scale rearrangements. We have found a duplication of a segment carrying the entire *psbA* gene on the chloroplast genome of two closely related pine species, *P. contorta* and *P. banksiana*. This is the only major duplication we have been able to find; the chloroplast genomes of conifers do not have the inverted repeat organization (Lidholm et al. 1988; Strauss et al. 1988; White 1990). To our knowledge, this is the first report of a duplication of a protein-encoding gene in the chloroplast genome of a vascular plant. Kushel and Kowallik (1987) reported a putative duplication of *psaA* in the brown alga *Dictyota dichotoma*, as judged by results of heterologous filter hybridization, but this was not verified by further analysis.

Several factors might contribute to the low frequency of duplications observed in chloroplast genomes. Apart from negative selection in the case of impaired gene function, these include instability of duplications in a recombination-proficient system and inefficiency of segregation and stable fixation, due to the high level of polyploidy of chloroplasts. Only true tandem repeats, however, will generate a high frequency of reversion by intermolecular homologous recombination. In the case of the *psbA* duplication in *P. contorta*, the gene copies are separated by at least 2.1 kb of unique DNA. In a homologous recombination, which would eliminate the duplication, this unique DNA would be concomitantly lost. If it carries an essential genetic element, which is likely considering the high informational density of chloroplast DNA, the recombination would be lethal. This appears a conceivable explanation for the retention of the *psbA* duplication, although the reason and mechanism for its initial occurrence are more obscure.

As to the time of the duplication event, the absolute sequence identity of the two copies of the duplicated segment does not necessarily provide any meaningful information. It might reflect the recent occurrence of the duplication but it could as well be due to the continuous action of the copy-correction mechanism shown to operate in chloroplasts (Wolfe et al. 1987). The only conclusion that can be drawn is that the rearrangement arose after the phylogenetic separation of the *Pinus* subsection *Contortae*, since it is confined to this group.

Some potentially important implications of a gene duplication can be discerned. First, two gene copies can evolve independently to acquire altered or new functions, either in terms of gene product properties or in terms of regulation of expression. In fact, a gene duplication is believed to be a prerequisite for divergent gene evolution. An example of this can be found in the cyanobacterium *Anacystis nidulans*, which has three differentially regulated copies of the *psbA* gene (Schaefer and Golden 1989). One of these gene copies encodes a D1 polypeptide with a modified amino acid sequence (Golden et al. 1986), which might therefore have somewhat altered functional characteristics. However, as evident from the DNA sequence identity, no such divergence of the two *psbA* copies has occurred in *P. contorta.*

Secondly, a duplication can lead to a gene dosage effect. Since the promoters as well as the coding regions of the two *psbA* gene copies of *P. contorta* are identical, one would expect that they are also transcribed at the same level. However, preliminary results from northern blot analyses indicate that the accumulation of transcripts originating from *psbAI* is significantly lower than that from *psbAII* (not shown), i.e., most of the total pool of *psbA* transcripts is produced from *psbAII.* The reason for this is unclear, although it may be related

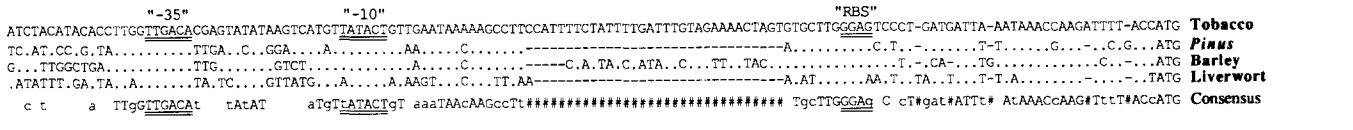

Fig. 4. Alignment of the *psbA* upstream sequences from tobacco (Shinozaki et al. 1986) with that of *P. contorta*, barley (Efimov et al. 1988), and liverwort (Ohyama et al. 1986). The *Escherichia coli*-like promoter elements and ribosome binding sites are indicated by "−35", "−10" and "RBS", respectively. *Periods* indicate nucleotides identical to the tobacco sequence whereas *hyphens* indicate positions where nucleotides are "missing". The consensus sequence shown at the bottom is derived from these four sequences. Positions where all four sequences are identical are indicated by *upper case letters*; *lower case letters* have been used at positions where one of the four sequences deviates; *gaps* mark non-conserved positions. Redundant nucleotides, i.e. positions where nucleotides are missing in one or more of the compared sequences, are indicated by #

to the fact that *psbAI* lacks the 3′ dyad symmetry present in *psbAII* as well as in *psbA* from other plants. It was demonstrated by Stern and Gruissem (1987) that such plastid 3′ structures stabilized synthetic RNAs *in cis* upon incubation is a chloroplast extract and that they also served as processing sites. It was suggested that downstream hairpin structures are important determinants for regulation of mRNA stability and accumulation in plastids. Indirect evidence for RNA stability as an important factor in the post-transcriptional control of plastid gene expression has been given in several reports (reviewed by Gruissem 1989).

The sequence conservation of chloroplast genes is known to be generally very high. This appears particularly pronounced for genes which encode polypeptides involved in complex interactions with other components, e.g. allosteric groups, other proteins and membrane lipids. This is illustrated by the overall sequence comparison of the *psbA* genes, encoding the D1 reaction center protein of photosystem II, from different plants (Table 1). In cases of highly conserved genes, overall sequence homology is not a very useful measure for assessing evolutionary distance between species. Ideally, the sequence comparison should be restricted to positions at which nucleotide substitutions are either silent or yield amino acid substitutions that are tolerated. Such a comparison should also take the codon usage into account. However, if the comparison is simply restricted to the third position of each codon, the resolution of the comparison is greatly improved. The third position comparison of *psbA* from different species provides an interesting molecular perspective on the early divergence of the gymnosperm and the angiosperm lineages (Mirov 1967). As can be seen in Table 1, the *Pinus*, tobacco and liverwort third position sequences are all, in fact, almost equally divergent from each other. However, we believe that any further conclusions concerning evolutionary relationships to other plant taxa, should be based on more sequence information than is reported here.

Acknowledgements. We thank Dr. R.G. Herrmann, Munich, FRG for providing the spinach *psbA* probe. This work was supported by grants from the Swedish Council for Forestry and Agricultural Research, The Swedish Natural Science Research Council, The Bo Rydin Foundation for Scientific Research and The Nordic Council for Agricultural Research (NKJ).

References

Bowman CM, Dyer TA (1986) The location and possible evolutionary significance of small dispersed repeats in wheat ctDNA. Curr Genet 10:931–941

Chen EY, Seeburg PH (1985) Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. DNA 4:165–170

Efimov VA, Andreeva AV, Reverdatto SV, Jung R, Chakhmakhcheva OG (1988) Nucleotide sequence of the barley chloroplast *psbA* gene for the Q_B protein of photosystem II. Nucleic Acids Res 16:5685

Erickson JM, Rahire M, Rochaix J-D (1984) *Chlamydomonas reinhardii* gene for the 32000 mol. wt. protein of photoystem II contains four large introns and is located entirely within the chloroplast inverted repeat. EMBO J 3:2753–2762

Feinberg AP, Vogelstein B (1983) A technique for radiolabelling DNA endonuclease fragments to high specific activity. Anal Biochem 132:6–13

Golden SS, Brusslan J, Haselkorn R (1986) Expression of a family of *psbA* genes encoding a photosystem II polypeptide in the cyanobacterium *Anacystis nidulans* R 2. EMBO J 5:2789–2798

Gruissem W (1989) Chloroplast gene expression: how plants turn their plastids on. Cell 56:161–170

Hanahan D (1983) Studies on the transformation of *Escherichia coli* with plasmids. J Mol Biol 166:557–580

Hanahan D (1985) Techniques for transformation of *E. coli*. In: Glover DM (ed) DNA cloning, vol 1, a practical approach. IRL Press, Oxford, pp 109–135

Harr R, Fällman P, Häggström M, Wahlström L, Gustafsson P (1986) GENEUS, a computer system for DNA and protein sequence analysis containing an information retrieval system for the EMBL data library. Nucleic Acids Res 14:273–284

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, 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 217:185–194

Hirschberg J, McIntosh L (1983) Molecular basis of herbicide resistance in *Amaranthus hybridus*. Science 222:1346–1349

Howe CJ (1985) The endpoints of an inversion in wheat chloroplast DNA are associated with short repeated sequences containing homology to *att*-lambda. Curr Genet 10:139–145

Howe CJ, Barker RF, Bowman CM, Dyer TA (1988) Common features of three inversions in wheat chloroplast DNA. Curr Genet 13:343–349

Kuhsel M, Kowallik KV (1987) The plastome of a brown alga, *Dictyota dichotoma*. II. Location of structural genes coding for ribosomal RNAs, the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase and for polypeptides of photosystem I and II. Mol Gen Genet 207:361–368

Lidholm J, Szmidt AE, Hällgren J-E, Gustafsson P (1988) The chloroplast genomes of conifers lack one of the rRNA-encoding inverted repeats. Mol Gen Genet 212:6–10

Mirov NT (1967) The genus *Pinus*. The Ronald Press, New York

Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, Sano S, Umesono K, Shiki Y, Takeuchi M, Chang Z, Aota S, Inokuchi H, Ozeki H (1986) Chloroplast gene organization deduced from complete sequence from liverwort *Marchantia polymorpha* chloroplast DNA. Nature 322:572–574

Ohyama K, Fukuzawa H, Kohchi T, Sano T, Sano S, Shirai H, Umesono K, Shiki Y, Takeuchi M, Chang Z, Aota S, Inokuchi H, Ozeki H (1988a) Structure and organization of *Marchantia polymorpha* chloroplast genome. I. Cloning and gene identification. J Mol Biol 203:281–298

Ohyama K, Kohchi T, Sano T, Yamada Y (1988b) Newly identified groups of genes in chloroplasts. Trends Biochem Sci 13:19–22

Palmer JD (1985a) Evolution of chloroplast and mitochondrial DNA in plants and algae. In: MacIntyre RJ (ed) Monographs in Evolutionary Biology: Molecular Evolutionary Genetics. Plenum Press, New York, pp 131–240

Palmer JD (1985b) Comparative organization of chloroplast genomes. Annu Rev Genet 19:325–354

Palmer JD, Nugent JM, Herbon LA (1987) Unusual structure of geranium chloroplast DNA: a triple-sized inverted repeat, extensive gene duplications, multiple inversions, and two repeat familes. Proc Natl Acad Sci USA 84:769–773

Quigley F, Weil JH (1985) Organization and sequence of five tRNA genes and of an unidentified reading frame in the wheat chloroplast genome: evidence for gene rearrangements during the evolution of chloroplast genomes. Curr Genet 9:495–503

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Schaefer MR, Golden SS (1989) Differential expression of members of a cyanobacterial *psbA* gene family in response to light. J Bacteriol 171:3973–3981

Shimada H, Sugiura M (1989) Pseudogenes and short repeated sequences in the rice chloroplast. Curr Genet 16:293–301

Shinozaki K, Ohme M, Tanaka M, Wagazugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expressions. EMBO J 5:2043–2049

Stern DB, Gruissem W (1987) Control of plastid gene expression: 3′ inverted repeats act as mRNA processing and stabilizing elements, but do not terminate transcription. Cell 51:1145–1157

Strauss SH, Palmer JD, Howe GT, Doerksen AH (1988) Chloroplast genomes of two conifers lack a large inverted repeat and are extensively rearranged. Proc Natl Acad Sci USA 85:3898–3902

Svensson B, Vass I, Cedergren E, Styring S (1990) Structure of donor side components in photosystem II predicted by computer modelling. EMBO J 9:2051–2059

Szmidt AE, Aldén T, Hällgren J-E (1987) Paternal inheritance of chloroplast DNA in *Larix*. Plant Mol Biol 9:59–64

Takahashi M, Shiraishi T, Asada K (1988) COOH-terminl residues of D1 and the 44 kDa CPa-2 at spinach photosystem II core complex. FEBS Lett 240:6–8

Tinoco I, Borer PN, Dengler B, Levine MD, Uhlenbeck OC, Crothers DM, Grotta J (1973) Improved estimation of secondary structure in ribonucleic acids. Nature New Biol 146:40–41

Tsai C-H, Strauss SH (1989) Dispersed repetitive sequences in the chloroplast genome of Douglas-fir. Curr Genet 16:211–218

White EE (1990) Chloroplast DNA in *Pinus monticola*: 1. Physical map. Theor Appl Genet 79:119–124

Wolfe KH, Li W-H, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci USA 84:9054–9058

Yanisch-Perron C, Vieira J, Messing J (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33:103–119

Communicated by R.G. Herrmann