Plant Molecular Biology **19**: 355–365, 1992.

Conserved relationship between *psbH* and *petBD* genes: presence of a shared upstream element in *Prochlorothrix hollandica*

Karen L. Greer and Susan S. Golden*
*Department of Biology, Texas A&M University, College Station, TX 77843-3258, USA (*author for correspondence)*

Received 8 November 1991; accepted in revised form 23 January 1992

Key words: petB, *petD*, prochlorophyte, *Prochlorothrix hollandica*, *psbB*, *psbH*

Abstract

Prochlorophytes are an unusual group of prokaryotic oxygenic photoautotrophs that morphologically appear to bridge the gap between cyanobacteria and the chloroplasts of eukaryotic plants. Molecular data place this group evolutionarily within the cyanobacteria, but they have a photosynthetic apparatus that is very similar to that found in chloroplasts. We have sequenced from the prochlorophyte *Prochlorothrix hollandica* a set of genes (*psbB*, *psbH*, *petB* and *petD*) that has a conserved organization in chloroplast genomes that is different from the organization in cyanobacterial genomes. The four genes are linked as an operon in chloroplasts, but only *petB* and *petD* are closely linked and cotranscribed in cyanobacteria. Although the prochlorophyte gene arrangement resembles that of cyanobacteria, one feature suggests the coordinated regulation of the unlinked genes. A 93 bp region of absolute conservation occurs upstream of the *psbH* gene and the *petBD* operon, near the site of transcription initiation in each gene set. This conserved element may indicate an alternative to cotranscription for achieving co-regulation of the *psbH* and *petBD* genes in the prochlorophyte.

Introduction

Prochlorothrix hollandica is a prokaryotic oxygenic photoautotroph that belongs to a group known as the prochlorophytes [7]. Distinguishing prochlorophytes from cyanobacteria is their system of light harvesting: prochlorophytes have an antenna that contains chlorophylls *a* and *b* as do chloroplasts and lack the phycobilisomes that are characteristic of cyanobacteria [7, 16]. Pigment composition and thylakoid membrane structure indicate that *P. hollandica* and other prochlorophytes have characteristics of both chloroplasts and cyanobacteria [5, 6], and it has been proposed that they might represent extant relatives of the prokaryotic endosymbiont that evolved into the chloroplast [17, 22, 23]. We have examined the sequences and organization of several genes that encode components of the photosynthetic apparatus in *P. hollandica* [22, 23, 24] with the

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession numbers X60313 (*petB*, *petD*) and X60314 (*psbH*).

goals of providing clues to the evolutionary position of this organism as well as revealing conserved mechanisms for regulation of the photosynthetic apparatus.

In the current study we have focused on four genes: *psbB*, *psbH*, *petB* and *petD*. These genes encode proteins of photosystem II (*psbB*, *psbH*) and of the cytochrome b_6f complex of the photosynthetic electron transport chain (*petB* and *petD*) [3, 12]. In chloroplast genomes all four genes are in an operon in the order *psbBHpetBD*, and *petB* and *petD* have class II introns [26, 30]. A number of transcripts arise from the operon by processing from a longer primary transcript, and these events are thought to be involved in regulation of gene expression [3, 4, 30, 33]. The relative level of the transcripts from these four genes is similar in light- and dark-grown leaves, but the *psbB* gene product shows a different pattern of accumulation from that of the *psbH* and *petBD* gene products under certain light conditions [4, 13].

The photosystem II and cytochrome b_6f complexes of chloroplasts contain components synthesized in the cytoplasm from nuclear genes as well as the chloroplast-encoded polypeptides. Proper assembly of the complexes depends on coordinate expression of products from both compartments in the cell. In prokaryotes, coordinate expression is not complicated by the partitioning of genes between two genomes, and the organization and numbers of cyanobacterial genes for photosynthesis proteins are different than in chloroplast genomes [1, 10, 24, 28]. In the cyanobacterium *Nostoc* sp. strain PCC 7906, only the *petB* and *petD* genes are linked in an operon; *psbB* and *psbH* are not found upstream from them [14]. Additionally, there are no introns in the *Nostoc petB* and *petD* genes [14]. Linkage of the four genes may be evolutionarily recent in the streamlining of the chloroplast genome as a response to developing into an organelle. The gene products contribute to different complexes of the photosynthetic pathway, and seem to be unrelated; thus there is no obvious need for their linkage in chloroplasts as a regulatory mechanism.

We have identified the *psbB*, *psbH*, *petB* and *petD* genes of *P. hollandica* to compare their organization to that observed in cyanobacteria and chloroplasts. The *P. hollandica* genes have a cyanobacterial organization, with *petB* and *petD* linked and cotranscribed but *psbB* and *psbH* elsewhere in the genome. However, there is a suggestion of coordinate expression of *psbH* and *petBD* in *P. hollandica*: a 93 bp perfect repeat sequence precedes the two cistrons. The function of the repeated sequence is not known, but transcripts for both *psbH* and *petBD* begin in close proximity to the element. The *psbB* gene does not have the same upstream element, suggesting that the presumed regulatory link between *psbH* and *petBD* is not shared by *psbB*.

Materials and methods

An axenic isolate of *P. hollandica* was provided by R.A. Lewin and grown in modified BG-11 medium [8] at 22 °C and at a photon flux density of 25 μ Em^{-2} s^{-1}. Genomic DNA and total RNA were isolated from 500 ml cultures bubbled with CO_2 as described for *Synechococcus* sp. strain PCC 7942 [27].

Approximately 1 μg of genomic DNA from axenic *P. hollandica* was digested with restriction enzymes for Southern analysis. Electrophoresis was performed on gels of 0.7% (w/v) agarose in Tris-acetate buffer [18] at 20 V for 18 h. Gels were treated for 15 min per treatment with a depurinating solution (0.125 M HCl), a denaturing solution (1.5 M NaCl, 0.5 M NaOH), and a neutralizing solution (1.5 M NaCl, 0.5 M Tris pH 7.4) [21]. DNA was transferred from the gel to charged nylon membranes (MagnaGraph, Micron Separations, Westboro, MA) using 10 × SSPE (20 × SSPE = 3.6 M NaCl, 200 mM NaH_2PO_4, 20 mM EDTA [18]) as a transfer medium with a Vacublot VAC-1000 Transfer System (American Bionetics, Hayward, CA). The resulting blots were dried under vacuum at 80 °C for 1–2 h.

Total RNA from *P. hollandica* was separated on 1.2% agarose gels buffered with 3-[N-morpholino]-propanesulfonic acid (MOPS) con-

taining 1% formaldehyde (v/v) [2]. RNA was transferred by capillary blotting onto MSI MagnaGraph membranes and blots were dried as described above.

Gel-purified DNA fragments were radiolabeled using a random-primer labeling kit (Boehringer Mannheim Biochemicals, Indianapolis, IN) and [α^{32}P]dCTP. Radioactive antisense RNAs were transcribed from promoters in pBluescript (Stratagene, LaJolla, CA) with either T7 or T3 polymerase in the presence of [α^{32}P]UTP (RNA transcription kit, Stratagene). Northern blots were hybridized with antisense RNA probes in a solution of 50% formamide (v/v), 5 × SSPE [18], 1% sodium dodecyl sulfate (SDS), and 100 μg/ml salmon sperm DNA at 50 °C overnight, and washed at 65 °C in 0.5 × SSPE, 0.1% SDS [27]. Radiolabeled *P. hollandica* DNA fragments were hybridized with Southern or northern blots at 55 °C in 5 × SSPE, 1% SDS, 100 μg/ml salmon sperm DNA, and were washed at 65 °C in 0.5 × SSPE, 0.1% SDS. Blots incubated with probes made from heterologous DNA were hybridized at 50 °C in 5 × SSPE, 1% SDS, 100 μg/ml salmon sperm DNA, and were washed at 50 °C in 5 × SSPE, 0.1% SDS.

The following heterologous probes were used to identify *P. hollandica* genes: a 270 bp *Eco* RV-*Sal* I fragment from pSMH10 which includes part of the *Synechocystis* sp. strain PCC 6803 *psbH* gene [20]; a 400 bp *Bam* HI-*Eco* RI fragment from p119petB which is internal to the maize *petB* gene [3]; and a 250 bp *Bam* HI fragment from pAB6 which carries a portion of the maize *petD* gene [3]. An existing library of *P. hollandica* *Eco* RI fragments in λgt10 [22] and a new library of 6–14 kb *Bam* HI fragments in λL47.1 [18] were screened for the genes. Restriction fragments from positive recombinant phage were subcloned in pBluescript (Stratagene) plasmids, or, in the case of a 3 kb *Bam* HI fragment that spans the *P. hollandica psbH* gene, in pBGS19 [29]. Subclones for DNA sequence analysis were generated from the *psbH*-bearing fragment by cloning smaller restriction fragments, and from a 2.2 kb *Bam* HI fragment that contains all of *petB* and approximately half of *petD* by exonuclease III digestion (Erase-A-Base kit, Promega Biotec, Madison, WI). Plasmid DNA was prepared as previously described [24] for DNA sequencing from double-stranded plasmid templates by the dideoxy chain-termination method (Sequenase kit, United States Biochemical Corporation, Cleveland, OH). Analysis of DNA sequences was performed using the Genetics Computer Group programs [9].

S1 nuclease protection analyses were performed as described by Tumer *et al.* [31]. Primer extension analyses were as described by Golden *et al.* [10], modified by slowly cooling the annealing mix from 65 °C to 42 °C over a 2.5 h period. Approximately 20 μg total RNA from *P. hollandica* was used for each assay. Two fragments were used for S1 nuclease analysis of *petBD* transcripts: an 840 bp *Acc* I-*Hin*d III fragment (extends leftward from nucleotide 630, Fig. 3A) and a 225 bp *Rsa* I-*Hae* III fragment (extends leftward from nucleotide 350, Fig. 3A). Primer extension reactions used a 17 bp oligonucleotide complementary to the RNA at positions 517 to 533 (Fig. 3A). A 280 bp *Dde* I-*Hin*d II fragment (extends leftward from nucleotide 550, Fig. 3B) and a 270 bp *Hin*d II to *Bam* HI fragment (extends leftward from nucleotide 270, Fig. 3B) were used for S1 nuclease protection of *psbH* transcripts. Primer extension used a 17 bp oligonucleotide complementary to the RNA at positions 577 to 592 (Fig. 3B). Protected fragments and Maxam and Gilbert sequencing reactions [19] were separated by electrophoresis on sequencing gels (6% w/v polyacrylamide 50% w/v urea) using the BRL Model S2 Sequencing Gel Electrophoresis System (Bethesda Research Laboratories, Gaithersburg, MD).

Results

Southern blots of total genomic DNA from *P. hollandica* digested with several restriction enzymes were probed with fragments from the *Anabaena* sp. strain PCC 7120 *psbB* [15], *Synechocystis psbH* [20], and maize *petB* and *petD* genes [3] (data not shown). These blots provided evidence that three of the four genes are

not closely linked in the prochlorophyte genome. The sizes of the *Eco* RI fragments that hybridized to *psbB* and *petBD* probes were approximately 6.0 kb and 6.6 kb, respectively (as in Fig. 1, panels A and B). Thus these three genes were likely to be represented in an existing library of *P. hollandica Eco* RI fragments cloned in λgt10 [22]. Positive clones were isolated from this library using the heterologous *Anabaena* (*psbB*) and maize (*petBD*) probes. The *Eco* RI fragment that hybridized with the *psbH* probe (Fig. 1, panel D) is too large to be represented in the λgt10 library, so we constructed a library of *Bam* HI fragments in λL47.1 [18]. The 3 kb *Bam* HI fragment that contains the *P. hollandica psbH* gene (Fig. 1, panel D) was isolated from this library by hybridization with the *Synechocystis* gene.

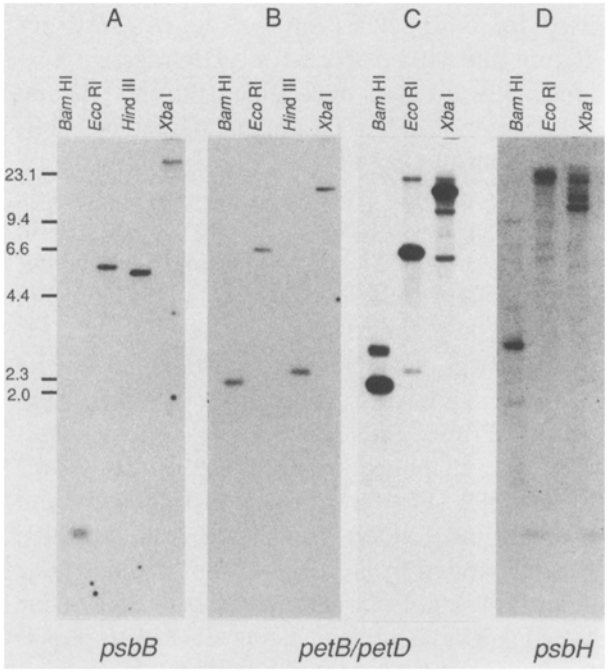


Fig. 1. Linkage comparisons of *P. hollandica psbB*, *petBD* and *psbH* genes by Southern analysis. *P. hollandica* genomic DNA was digested with the indicated restriction enzymes and prepared for Southern analysis. Panels A and B are from a single gel, as are panels C and D. Size markers, from λ DNA digested with *Hind* III, represent all panels. The following were used as probes: panel A, a 0.5 kb radioactive antisense RNA internal to the *P. hollandica psbB* gene [11]; panel B, a radiolabeled antisense RNA from a 2.2 kb fragment that includes all of the *P. hollandica petB* gene and 180 bp of *petD*; panel C, probe 1–8, a fragment that contains sequences upstream from *petB*; panel D, a 270 bp *Eco* RV-*Sal* I fragment from the *Synechocystis psbH* gene.

Heterologous probes or cloned *P. hollandica* gene probes for *psbB* and *psbH* (Fig. 1, panels A and D) hybridized to fragments that migrated differently from each other and from *petB* and *petD* fragments (Fig. 1, panel B). Heterologous internal probes from the maize *petB* and *petD* genes showed identical patterns (data not shown); the same pattern was generated by probing with a 2.2 kb *Bam* HI fragment (Fig. 2) carrying *petB* from *P. hollandica* (Fig. 1, panel B). This suggested that *petB* and *petD* are close together on the *P. hollandica* genome as is the case for those genes in *Nostoc* [14].

A probe that covered only the upstream flanking region of the *P. hollandica petB* gene (Fig. 2, probe 1–8) hybridized to numerous bands in each lane when the Southern blot was washed at moderate stringency (data not shown), suggesting a repeated sequence upstream of *petB*. After a high-stringency wash, probe 1–8 hybridized strongly to the major band for *petB* and *petD* and to one or two additional bands (Fig. 1, panel C). One of these bands comigrated with fragments recognized by the *psbH* probe from *Synechocystis* (Fig. 1, panel D). The heterologous *psbH* probe is not likely to contain the same upstream region, and it did not recognize the *petBD* band. A probe made from the *P. hollandica psbH* gene and its upstream region recognized the same bands as probe 1–8 (data not shown).

Figure 2 diagrams the cloned regions of the *P. hollandica* genome and the strategies used for sequencing *petBD* and *psbH*. The *P. hollandica psbB* gene was also isolated, mapped and sequenced [11]. *petB* and *petD* were originally subcloned on a 2.2 kb *Bam* HI fragment, but an overlapping 2.4 kb *Hind* III fragment that contains the entire coding sequence of both genes is also shown.

The three genes, *petB*, *petD* and *psbH*, are conserved with respect to those of chloroplasts and cyanobacteria (Fig. 3). At the amino acid level, *petB* is 92% and 90% similar, and 85% and 82% identical to the *Nostoc* [14] and maize [26] se-

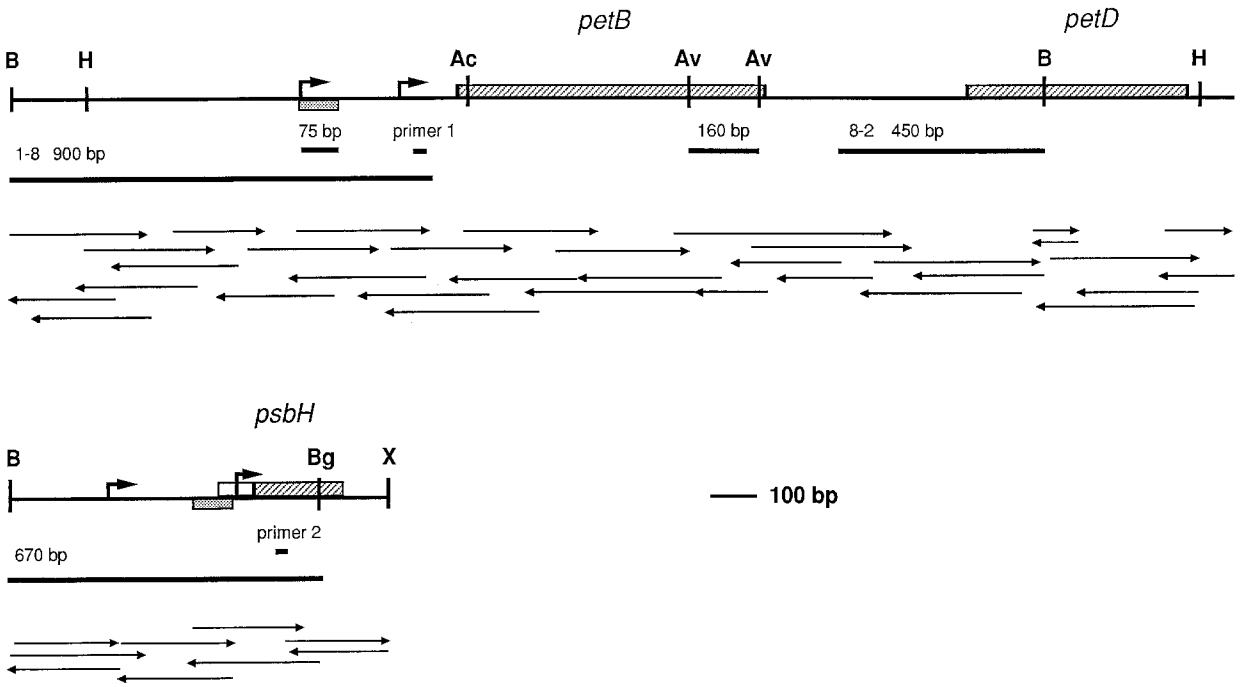
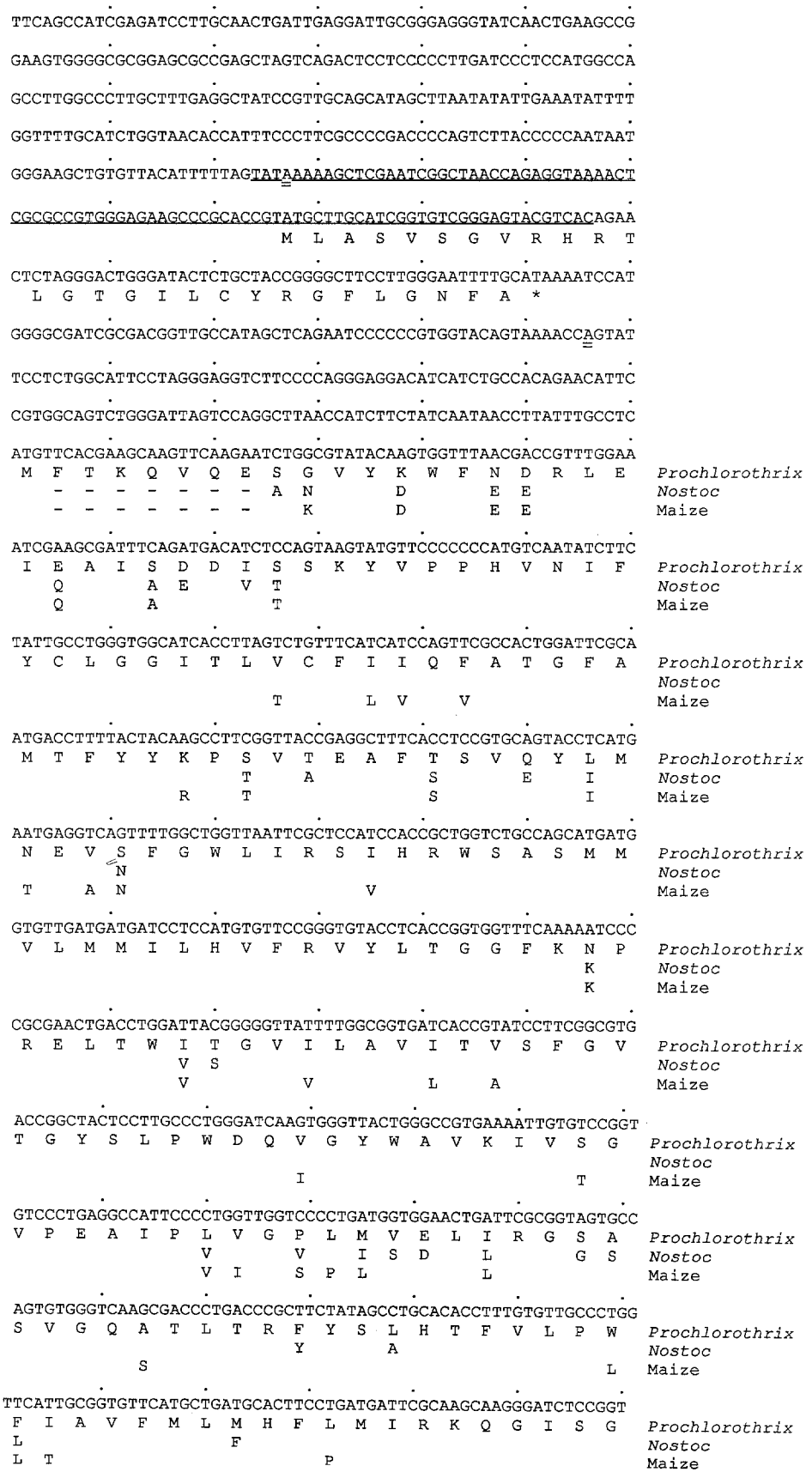


Fig. 2. Physical maps of *petB*, *petD* and *psbH*. Hatched boxes indicate potential open reading frames (ORFs) for *petB*, *petD* and *psbH*. The open box in front of the ORF for *psbH* indicates the alternative translational start site. The stippled boxes upstream from *petB* and *psbH* indicate the positions of the 93 bp repeated elements. Bent arrows above the restriction maps indicate transcript 5′ ends and the direction of transcription. Bars beneath the maps indicate relevant probes. These are identified by size if they are from naturally occurring restriction fragments or by a number and size if they were derived from exonuclease III digestion. Also indicated by small bars is the position of primers used for primer extension analysis of transcripts. Sequencing strategy for the three genes is displayed by arrows beneath the restriction maps. Abbreviations for restriction enzyme recognition sites are as follows: Ac, *Acc* I; Av, *Ava* II, B, *Bam* HI; Bg, *Bgl* II, H, *Hin*d III; X, *Xba* I.

quences, respectively. Likewise, *petD* is 85 and 86% similar and 76% identical to the *Nostoc* [14] and spinach [12] genes. The deduced amino acid sequence of the *psbH* open reading frame is slightly less well-conserved, at 94% and 66% similarity and 64% and 47% identity with *Synechocystis* [20] and spinach [32] sequences, respectively. If the *psbH* open reading frame begins at the first possible ATG, the product would be longer than other *psbH* genes (Fig. 2, *psbH*, open box). Translation from the second ATG would yield a product that is more similar to the *Synechocystis psbH*. Both potential start codons have possible ribosome binding sites: GGAGAAG 10 bp before the first ATG and AGG 6 bp before the second ATG (Fig. 3B).

The *P. hollandica petB* and *petD* genes are separated by 427 bp. This spacer includes a small open reading frame (Fig. 3A, nucleotides 1406–1528) which potentially encodes a 40 amino acid polypeptide that showed no strong similarity to any sequence in the GenBank or EMBL databases. As is true for the *Nostoc* genes [14] and those of the green alga, *Chlorella protothecoides* [25], there is no evidence for introns in the *P. hollandica petB* or *petD* genes. The amino terminal region of the *petB*-coding region is, however, different from all other sequences, and conservation begins at the base where splicing of the two exons occurs in the chloroplast *petB* gene (Fig. 3A, nucleotide 625).

Nucleotide sequence analysis showed an identical 93 bp sequence upstream of the *psbH* and *petB* genes, accounting for the comigration of bands that hybridized to *petB* and *psbH* probes (Fig. 2, stippled box; Fig. 3A and 3B, single un-


1 TTCAGCCATCGAGATCCTTGCAACTGATTGAGGATTGCGGGAGGGTATCAACTGAAGCCG
61 GAAGTGGGGCGCGGAGCGCCGAGCTAGTCAGACTCCTCCCCCTTGATCCCTCCATGGCCA
121 GCCTTGGCCCTTGCTTTGAGGCTATCCGTTGCAGCATAGCTTAATATATTGAAATATTTT
181 GGTTTTGCATCTGGTAACACCATTTCCCTTCGCCCCGACCCCAGTCTTACCCCCAATAAT
241 GGGAAGCTGTGTTACATTTTTAGTATAAAAAGCTCGAATCGGCTAACCAGAGGTAAAACT
301 CGCGCCGTGGGAGAAGCCCGCACCGTATGCTTGCATCGGTGTCGGGAGTACGTCACAGAA
M L A S V S G V R H R T
361 CTCTAGGGACTGGGATACTCTGCTACCGGGGCTTCCTTGGGAATTTTGCATAAAATCCAT
L G T G I L C Y R G F L G N F A *
421 GGGGCGATCGCGACGGTTGCCATAGCTCAGAATCCCCCCGTGGTACAGTAAAACCAGTAT
481 TCCTCTGGCATTCCTAGGGAGGTCTTCCCCAGGGAGGACATCATCTGCCACAGAACATTC
541 CGTGGCAGTCTGGGATTAGTCCAGGCTTAACCATCTTCTATCAATAACCTTATTTGCCTC
601 ATGTTCACGAAGCAAGTTCAAGAATCTGGCGTATACAAGTGGTTTAACGACCGTTTGGAA
M F T K Q V Q E S G V Y K W F N D R L E Prochlorothrix
- - - - - - - A N D E E Nostoc
- - - - - - - K D E E Maize
661 ATCGAAGCGATTTCAGATGACATCTCCAGTAAGTATGTTCCCCCCCATGTCAATATCTTC
I E A I S D D I S S K Y V P P H V N I F Prochlorothrix
Q A E V T Nostoc
Q A T Maize
721 TATTGCCTGGGTGGCATCACCTTAGTCTGTTTCATCATCCAGTTCGCCACTGGATTCGCA
Y C L G G I T L V C F I I Q F A T G F A Prochlorothrix
Nostoc
T L V V Maize
801 ATGACCTTTTACTACAAGCCTTCGGTTACCGAGGCTTTCACCTCCGTGCAGTACCTCATG
M T F Y Y K P S V T E A F T S V Q Y L M Prochlorothrix
T A S E I Nostoc
R T S I Maize
861 AATGAGGTCAGTTTTGGCTGGTTAATTCGCTCCATCCACCGCTGGTCTGCCAGCATGATG
N E V S F G W L I R S I H R W S A S M M Prochlorothrix
N Nostoc
T A N V Maize
921 GTGTTGATGATGATCCTCCATGTGTTCCGGGTGTACCTCACCGGTGGTTTCAAAAATCCC
V L M M I L H V F R V Y L T G G F K N P Prochlorothrix
K Nostoc
K Maize
981 CGCGAACTGACCTGGATTACGGGGGTTATTTTGGCGGTGATCACCGTATCCTTCGGCGTG
R E L T W I T G V I L A V I T V S F G V Prochlorothrix
V S Nostoc
V V L A Maize
1041 ACCGGCTACTCCTTGCCCTGGGATCAAGTGGGTTACTGGGCCGTGAAAATTGTGTCCGGT
T G Y S L P W D Q V G Y W A V K I V S G Prochlorothrix
Nostoc
I T Maize
1101 GTCCCTGAGGCCATTCCCCTGGTTGGTCCCCTGATGGTGGAACTGATTCGCGGTAGTGCC
V P E A I P L V G P L M V E L I R G S A Prochlorothrix
V V I S D L G S Nostoc
V I S P L L Maize
1161 AGTGTGGGTCAAGCGACCCTGACCCGCTTCTATAGCCTGCACACCTTTGTGTTGCCCTGG
S V G Q A T L T R F Y S L H T F V L P W Prochlorothrix
Y A Nostoc
S L Maize
1221 TTCATTGCGGTGTTCATGCTGATGCACTTCCTGATGATTCGCAAGCAAGGGATCTCCGGT
F I A V F M L M H F L M I R K Q G I S G Prochlorothrix
L F Nostoc
L T P Maize

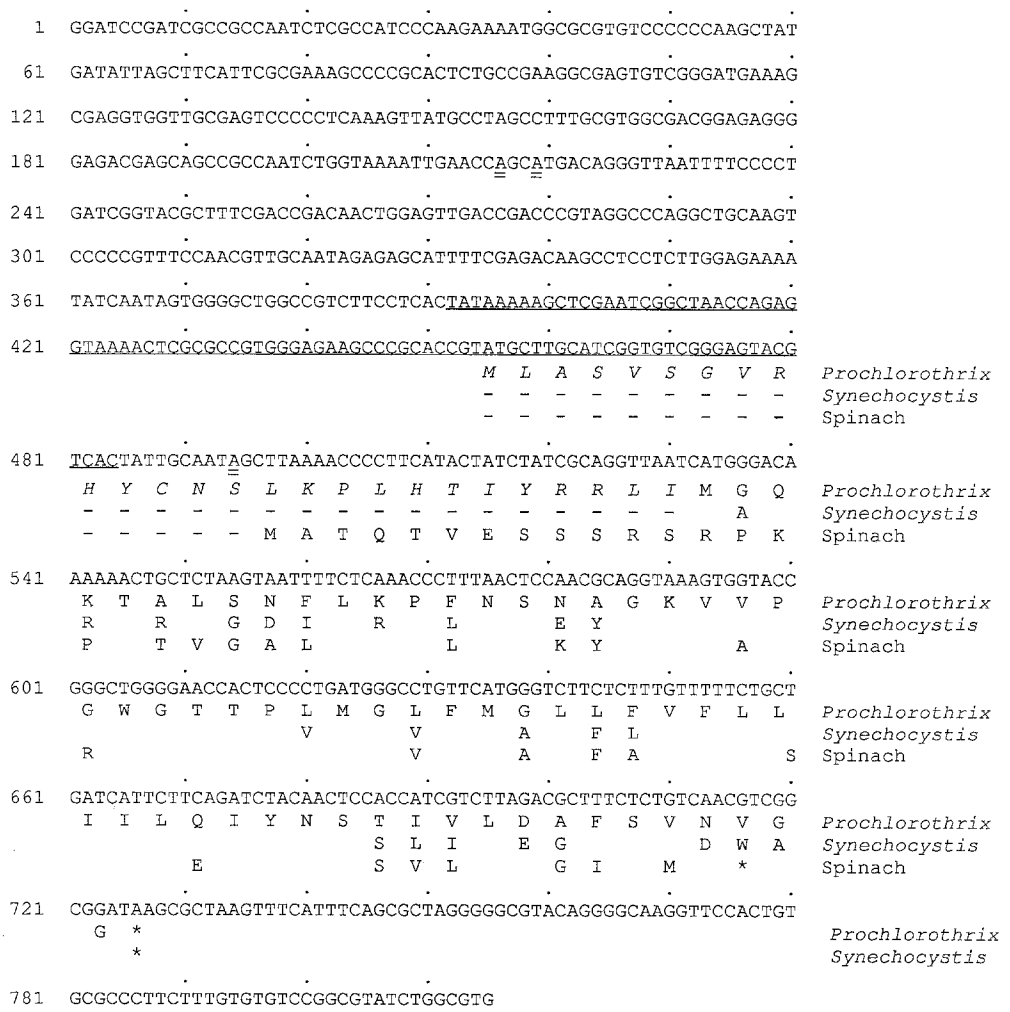


Fig. 3. Nucleotide and deduced amino acid sequences for *petBD* and *psbH*. The nucleotide sequence is shown for *petB* and *petD* in panel A (two previous pages) and for *psbH* in panel B (above). Deduced amino acid sequences are shown by the single letter code for these genes and for potential ORFs upstream from *petB* and between *petB* and *petD* (Fig. 3A, positions 327–400 and 1406–1528). The translation of the first ORF of the *psbH* sequence is shown in italics. Differences in amino acid sequences between the *P. hollandica* genes and genes from other prokaryotes and chloroplast genes are shown beneath the *P. hollandica* translations. Residues that are not present in the deduced amino acid sequence from *Synechocystis* and spinach are indicated by dashes. Double inderlines indicate transcript 5′ ends mapping by S1 nuclease-protection and primer extension assays. The single underline indicates the 93 bp repeated element upstream from both *petB* and *psbH*. Sequences shown in bold italics downstream from *petD* represent a potential stem loop that could form a *rho*-independent terminator. No sequences with such strong potential for secondary structure follow the *psbH* ORF.

derline). Northern analysis was performed to determine whether this element is cotranscribed with either of the known genes. Probes made from the 2.4 kb *Hin*d III fragment that contains the *petB* and *petD* coding regions (Fig. 4, panel A) as well as internal probes for each gene (data not shown) hybridized to at least two transcripts. The two major transcripts are approximately 2100 and 1900 nucleotides long. A probe that contains the majority of the *psbH* open reading frame and 500 bp of upstream sequence hybridized to three lower molecular weight transcripts, as well as to a transcript that comigrates with the larger of the two *petBD* transcripts (Fig. 4, panel C). The two

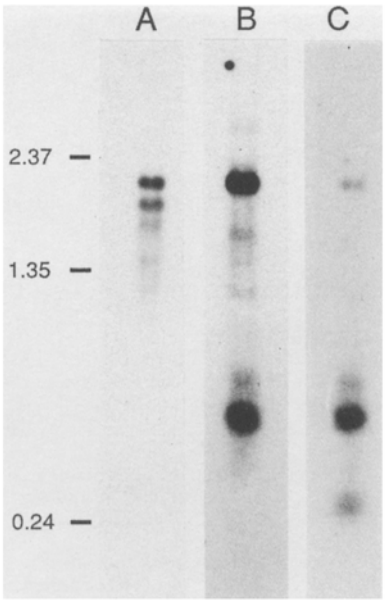


Fig. 4. RNA blot analysis of *pet* and *psbH* transcripts. Three identical sets of lanes from the same gel were probed with (A) a random primer-labeled fragment from the 3′ end of *petB* (Fig. 2, 160 bp probe), (B) a radiolabeled antisense RNA from a fragment internal to the 93 bp repeated element (Fig. 2, 75 bp probe) and (C) a radiolabeled antisense RNA from a fragment that contains most of the *psbH* open reading frame and upstream sequences (Fig. 2, 670 bp probe). Molecular sizes were estimated from a commercial 0.24–9.5 kb RNA ladder (BRL, Gaithersburg, MD).

shorter transcripts are approximately 680 and 450 nucleotides long. Each transcript identified by northern analysis is long enough to encode the gene(s) with which it hybridized. A northern blot probed with a fragment from within the repeated element (Fig. 2, 75 bp probe) specifically recognized a band that comigrated with the longer of the two transcripts attributed to the *petBD* and *psbH* genes (Fig. 4, panel B).

Transcripts were further analyzed by S1 nuclease-protection and primer-extension assays which revealed two transcript 5′ ends for both *psbH* and *petBD*. The presence of two transcript ends for both loci supports data from Northern analyses. Figure 5A shows the S1 nuclease-protection analysis of the two transcript 5′ ends, one for *petBD* and one for *psbH*, that are in close proximity to the repeated element. Figure 5B shows the S1 nuclease protection experiment that mapped the other transcripts for both genes. All four transcript ends were corroborated to within one or two bases by primer extension analyses (data not shown) using the 17 bp oligonucleotide primers indicated in Fig. 2 (primer 1 for *petBD* and primer 2 for *psbH*).

The two 5′ ends for the *petBD* cluster map 125 bp and 333 bp upstream of the start codon of the *petB* open reading frame (Fig. 3A). The longer transcript begins four bases into the repeated element. An open reading frame with a coding potential of 28 amino acids begins within the repeated element and continues past the end of the element into the upstream leader region of *petB*. This small unidentified open reading frame is preceded by a potential ribosome-binding site (GGAGAAG 10 bp upstream of the putative start codon), but we do not know if it is translated. The start of this open reading frame also corresponds to the first potential open reading frame for *psbH* (Fig. 3B). However, the longer open reading frame is not present on the shorter transcript identified for *psbH*. The *psbH* transcript ends map to 40 and 318 bases upstream from the second ATG of the open reading frame (Fig. 3B). The 5′ end of the shorter *psbH* transcript is 10 bases downstream from the right end of the repeated element.

Discussion

Four genes encoding elements of the photosynthetic apparatus, *psbB*, *psbH*, *petB* and *petD* have an arrangement in *P. hollandica* that is characteristic of cyanobacteria. This arrangement differs from that found in the chloroplast of most eukaryotic photoautotrophs in that only *petB* and *petD* are cotranscribed, and *psbB* and *psbH* are not located immediately upstream [14, 15]. Additionally, the higher-plant *petB* and *petD* genes have Group II introns [26, 30, 33], but there are no introns in the prochlorophyte *pet* genes.

All three genes show a great deal of similarity across evolutionary lines with a few interesting differences. One difference is that, even though there is no apparent intron in the *petB* gene from

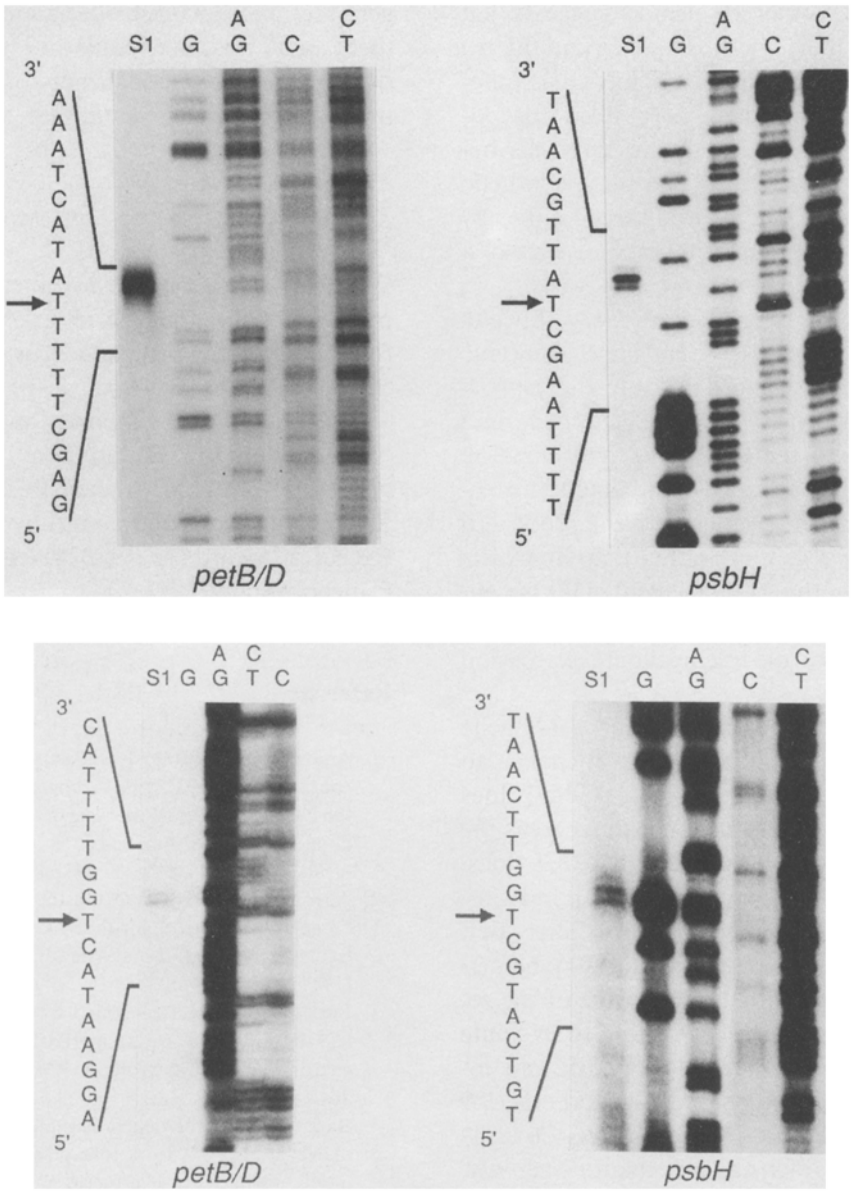


Fig. 5. S1 nuclease protection mapping of transcript 5′ ends. Nuclease protection reactions were performed as described in Materials and methods and bands were separated by electrophoresis on sequencing gels. The positions of 5′ ends were assigned by comparison of the migration of protected bands (lanes S1) with positions in DNA sequencing ladders generated from the same labeled DNA fragments (lanes G, AG, C, and CT). Arrows indicate probable transcript ends. Panel A (top): mapping of the *petBD* and *psbH* transcript ends that are proximal to the repeated element (the longer *petBD* and shorter *psbH* species); panel B (bottom): mapping of the element-distal transcript ends for those genes.

P. hollandica, the amino terminus encoded by the *petB* ORF is different from that of all other *petB* genes. The region of high similarity begins at a conserved valine residue that would correspond to the intron 3′ splice site for the chloroplast genes. The most striking unique feature is the presence of an identical repeated element upstream from *psbH* and *petB* in *P. hollandica*. The

presence and location of the element suggests that the chloroplast linkage between *psbH* and the *pet* genes is not merely fortuitous, but has some functional significance. If there is a need for co-regulation of genes, or coordinate synthesis between different complexes of the photosynthetic membrane, this purpose may be served either by juxtaposition of the genes or by the presence of a specific regulatory element.

When an internal fragment of the 93 bp element was used to probe *P. hollandica* genomic DNA at moderate stringency, only the *petBD*- and *psbH*-bearing fragments were identified (data not shown). Thus, the additional hybridization signals detected by probe 1–8 indicate the presence of other repeated elements on the probe and in the *P. hollandica* genome rather than many degenerate copies of the 93 bp element. This is consistent with the idea that this element may serve a specific regulatory role to coordinate expression of *petB* and *psbH*.

In greening wheat plants, *psbB* and *psbH* gene products accumulate differently in response to light [13]. The *psbB* gene product (CP47) does not accumulate in the dark in wheat, whereas the 10 kDa phosphoprotein encoded by *psbH* does. *psbH* and the *pet* genes appear to respond similarly to light in chloroplasts, and it is likely that they are co-regulated. Co-regulation may be mediated in *P. hollandica* by the presence of the repeated element. It will be interesting to evaluate whether the small ORF that begins in and includes the repeated element in front of the *petBD* cluster is expressed and whether that region or its potential gene product has a regulatory activity.

The presence of the repeated element suggests a common thread in the expression of genes for elements of the photosynthetic apparatus. It is likely that oxygenic photosynthetic prokaryotes have evolved different mechanisms to accomplish the goals they share with the chloroplast: assembly and maintenance of a functional photosynthetic apparatus. The constraints of coordinate expression between compartmentalized genomes has probably had a great influence on the arrangement and expression of chloroplast photosynthesis genes. The prochlorophyte and cyanobacterial genomes may have achieved similar gene arrangement, and possible regulatory strategies, either due to the shared constraints of prokaryotic genomes or through common lineage.

Acknowledgements

We thank Alice Barkan for plasmids containing portions of the maize *petB* and *petD* genes and Steve Mayes for a plasmid carrying part of the *Synechocystis psbH* gene.

This project was supported by grants from the National Science Foundation (DMB 8958089) and the Texas Advanced Research Program. S.S.G. was supported in part by a Junior Faculty Research Award (JFRA 224) from the American Cancer Society.

References

1. Anderson SL, McIntosh L: Partial conservation of the 5′ *ndhE-psaC-ndhD* 3′ gene arrangement of chloroplasts in the cyanobacterium *Synechocystis* sp PCC 6803: implications for NDH-D function in cyanobacteria and chloroplasts. Plant Mol Biol 16: 487–499 (1991).
2. Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K: Current protocols in molecular biology. Greene/Wiley-Interscience, New York (1987).
3. Barkan A: Proteins encoded by a complex chloroplast transcription unit are each translated from both monocistronic and polycistronic mRNAs. EMBO J 7: 2637–2644 (1988).
4. Barkan A: Tissue-dependent plastid RNA splicing in maize: transcripts from four plastid genes are predominantly unspliced in leaf meristems and roots. Plant Cell 1: 437–445 (1989).
5. Bullerjahn GS, Matthijs HCP, Mur LR, Sherman LA: Chlorophyll-protein composition of the thylakoid membrane from *Prochlorothrix hollandica*, a prokaryote containing chlorophyll *b*. Eur J Biochem 168: 295–300 (1987).
6. Burger-Wiersma T, Post AF: Functional analysis of the photosynthetic apparatus of *Prochlorothrix hollandica* (Prochlorales), a Chlorophyll *b* containing procaryote. Plant Physiol 91: 770–774 (1989).
7. Burger-Wiersma T, Veenhuis M, Korthals HJ, Van de Wiel CCM, Mur LR: A new prokaryote containing chlorophylls *a* and *b*. Nature 320: 262–264 (1986).
8. Bustos SA, Golden SS: Expression of the *psbDII* gene in *Synechococcus* sp. strain PCC 7942 requires sequences

downstream of the transcription start site. J Bact 173: 7525–7533 (1991).
9. Devereux J, Haeberli P, Smithies O: A comprehensive set of sequence analysis programs for the VAX. Nucl Acids Res 12: 387–395 (1984).
10. Golden SS, Brusslan J, Haselkorn R: Expression of a family of *psbA* genes encoding a photosystem II polypeptide in the cyanobacterium *Anacystis nidulans* R2. EMBO J 5: 2789–2798 (1986).
11. Greer KL, Golden SS: Nucleotide sequence of *psbB* from *Prochlorothrix hollandica*. Plant Mol Biol 17: 915–917 (1991).
12. Heinemeyer W, Alt J, Herrman RG: Nucleotide sequence of the clustered genes for apocytochrome b6 and subunit 4 of the cytochrome b/f complex in the spinach plastid chromosome. Curr Genet 8: 543–549 (1984).
13. Hird SM, Webber AN, Wilson RJ, Dyer TA, Gray JC: Differential expression of the *psbB* and *psbH* genes encoding the 47 kDa chlorophyll *a*-protein and the 10 kDa phosphoprotein of photosystem II during chloroplast development in wheat. Curr Genet 19: 199–206 (1991).
14. Kallas T, Spiller S, Malkin R: Characterization of two operons encoding the cytochrome b_6-*f* complex of the cyanobacterium *Nostoc* PCC 7906. J Biol Chem 263: 14334–14342 (1988).
15. Lang JD, Haselkorn R: Isolation, sequence and transcription of the gene encoding the photosystem II chlorophyll-binding protein, CP-47, in the cyanobacterium *Anabaena* 7120. Plant Mol Biol 13: 441–456 (1989).
16. Lewin RA: Extraordinary pigment composition of a prokaryotic alga. Nature 256: 735–737 (1975).
17. Lewin RA, Cheng L: Introduction. In: Lewin RA, Cheng L (eds) Prochloron: A Microbial Enigma, pp. 1–7. Chapman and Hall, New York (1989).
18. Maniatis T, Fritsch EF, Sambrook J: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1982).
19. Maxam AM, Gilbert W: Sequencing end-labeled DNA with base-specific chemical cleavages. Meth Enzymol 65: 499–560 (1980).
20. Mayes SR, Barber J: Nucleotide sequence of the *psbH* gene of the cyanobacterium *Synechocystis* 6803. Nucl Acids Res 18: 194 (1990).
21. Medveczky P, Chang C-W, Oste C, Mulder C: Rapid vacuum driven transfer of DNA and RNA from gels to solid supports. Biotechniques 5: 242–246 (1987).
22. Morden CW, Golden SS: *psbA* genes indicate common ancestry of prochlorophytes and chloroplasts. Nature 337: 382–385 (1989).
23. Morden CW, Golden SS: *psbA* genes indicate common ancestry of prochlorophytes and chloroplasts (Corrigendum). Nature 339: 400 (1989).
24. Morden CW, Golden SS: Sequence analysis and phylogenetic reconstruction of the genes encoding the large and small subunits of ribulose-1,5-bisphosphate carboxylase/oxygenase from the chlorophyll *b*-containing prokaryote *Prochorothrix hollandica*. J Mol Evol 32: 379–395 (1991).
25. Reimann A, Kück U: Nucleotide sequence of the plastid genes for apocytochrome b_6 (*petB*) and subunit IV of the cytochrome b_6-*f* complex (*petD*) from the green alga *Chlorella protothecoides*: lack of introns. Plant Mol Biol 13: 255–256 (1989).
26. Rock CD, Barkan A, Taylor WC: The maize plastid *psbB-psbF-petB-petD* gene cluster: spliced and unspliced *petB* and *petD* RNAs encode alternative products. Curr Genet 12: 69–77 (1987).
27. Schaefer MR, Golden SS: Differential expression of members of a cyanobacterial *psbA* gene family in response to light. J Bact 171: 3973–3981 (1989).
28. Shinozaki K, Ohme M, Tanaka M, Wakasugi 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: The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J 5: 2043–2049 (1986).
29. Spratt BG, Hedge PJ, te Heesen S, Edelman A, Broome-Smith JK: Kanamycin-resistant vectors that are analogues of plasmids pUC8, pUC9, pEMBL8 and pEMBL9. Gene 41: 337–342 (1986).
30. Tanaka M, Obokata J, Chunwongse J, Shinozaki K, Sugiura M: Rapid splicing and stepwise processing of a transcript from the *psbB* operon in tobacco chloroplasts: Determination of the intron sites in *petB* and *petD*. Mol Gen Genet 209: 427–431 (1987).
31. Tumer NE, Robinson SJ, Haselkorn R: Different promoters for the *Anabaena* glutamine synthetase gene during growth using molecular or fixed nitrogen. Nature 306: 337–342 (1983).
32. Westhoff P, Farchaus JW, Herrmann RG: The gene for the M_r 10000 phosphoprotein associated with photosystem II is part of the *psbB* operon of the spinach plastid chromosome. Curr Genet 11: 165–169 (1986).
33. Westhoff P, Herrmann RG: Complex RNA maturation in chloroplasts: the *psbB* operon from spinach. Eur J Biochem 171: 551–564 (1988).