J Mol Evol (2012) 75:184–197
DOI 10.1007/s00239-012-9526-z

Extreme Conservation of the *psaA/psaB* Intercistronic Spacer Reveals a Translational Motif Coincident with the Evolution of Land Plants

Elena L. Peredo · Donald H. Les · Ursula M. King · Lori K. Benoit

Received: 18 May 2012 / Accepted: 3 October 2012 / Published online: 29 November 2012

Abstract Although chloroplast transcriptional and translational mechanisms were derived originally from prokaryote endosymbionts, chloroplasts retain comparatively few genes as a consequence of the overall transfer to the nucleus of functions associated formerly with prokaryotic genomes. Various modifications reflect other evolutionary shifts toward eukaryotic regulation such as posttranscriptional transcript cleavage with individually processed cistrons in operons and gene expression regulated by nuclear-encoded sigma factors. We report a notable exception for the *psaA-psaB-rps14* operon of land plant (embryophyte) chloroplasts, where the first two cistrons are separated by a spacer region to which no significant role had been attributed. We infer an important function of this region, as indicated by the conservation of identical, structurally significant sequences across embryophytes and their ancestral protist lineages, which diverged some 0.5 billion years ago. The *psaA/psaB* spacers of embryophytes and their progenitors exhibit few sequence and length variants, with most modeled transcripts resolving the same secondary structure: a loop with projecting Shine-Dalgarno site and well-defined stem that interacts with adjacent coding regions to sequester the *psaB* start codon. Although many functions of the original endosymbiont have been usurped by nuclear genes or interactions, conserved functional elements of embryophyte *psaA/psaB* spacers provide compelling evidence that translation of *psaB* is regulated here by a *cis*-acting mechanism comparable to those common in prokaryotes. Modeled transcripts also indicate that spacer variants in some plants (e.g., aquatic genus *Najas*) potentially reflect ecological adaptations to facilitate temperature-regulated translation of *psaB*.

Keywords Charophytes · cpDNA · Embryophytes · Gene expression · Operon · Photosystem I · Polycistronic · Secondary structure · Shine-Dalgarno site

Introduction

By virtue of its endosymbiotic origin (Gray 1993), the higher plant chloroplast (cp) possesses genetic transcriptional and translational mechanisms derived originally from cyanobacteria (Peled-Zehavi and Danon 2007; Bock 2007). In addition to synthesizing 70S ribosomes, the 60–120 genes coded by cpDNA are often arranged in polycistronic operons (Peled-Zehavi and Danon 2007; Drechsel and Bock 2011), which reminisce their prokaryotic ancestry. However, chloroplasts contain far fewer than the 3000+ genes found in some prokaryotic genomes as many of the functions provided by the original endosymbiont have since been transferred to, or are now regulated by, the nucleus (Bock 2007).

Numerous modifications of chloroplast genes have evolved, which reflect a shift toward typical eukaryotic posttranscriptional-based regulation (Peled-Zehavi and Danon 2007). Examples include cleavage of the mRNA transcript and individual cistron processing in many of the operons (Marín-Navarro et al. 2007) in addition to the regulation of gene expression by nuclear-encoded sigma

Electronic supplementary material The online version of this article (doi:10.1007/s00239-012-9526-z) contains supplementary material, which is available to authorized users.

E. L. Peredo (✉) · D. H. Les · U. M. King · L. K. Benoit
Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT 06269-3043, USA
e-mail: elperedo@outlook.com

D. H. Les
e-mail: les@uconn.edu

factors (Allison 2000; Favory et al. 2005). A dual prokaryotic/eukaryotic nature of chloroplast gene expression is evident by the acquired ability to partially or completely utilize nuclear-encoded RNA polymerase (NEP) in place of the prokaryotic plastid encoded analog (PEP). Although transcription of chloroplast genes can be mediated by PEP or NEP (Gruissem et al. 1983; Hajdukiewicz et al. 1997), the key genes for photosystem I (PS I) (e.g., *psaA-psaB-rps14* operon) and photosystem II (PS II) rely exclusively on PEP; however, many others (e.g., *atpB*, *atpI*, *clpP*) have multiple promoters, which enable them to use either PEP or NEP, while a few others (*accD*, *ycf2*) use NEP exclusively (Hajdukiewicz et al. 1997).

In particular, variation in the initiation of translation associated with binding of the mRNA transcript to the ribosome has been presented as a "key difference" between prokaryotes and eukaryotic cpDNA genomes (Peled-Zehavi and Danon 2007). Prokaryotic translation involves a highly conserved "Shine-Dalgarno" (SD) sequence in the 5′ untranslated region (5′-UTR), which provides an mRNA binding site that is complimentary to the anti-SD sequence positioned at the 3′ end of the 16S rRNA (Peled-Zehavi and Danon 2007). The initiation of prokaryotic translation relies on a correct positional relationship between the SD site and AUG start codon of the gene (Ma et al. 2002). However, SD sequences are not conserved positionally in many higher plant chloroplast genomes (Hirose and Sugiura 2004) and some genes are translated in the absence of functional SD sequences (Fargo et al. 1998; Sugiura et al. 1998). In tobacco, SD mutants for *rps2* exhibited an increase rather than decrease in translation, indicating the interaction of trans-acting factors in translation of some chloroplast genes (Plader and Sugiura 2003). Because of such inconsistencies, the role of SD sequences as rRNA pairing elements in chloroplast genomes has remained controversial (Marín-Navarro et al. 2007).

However, functional prokaryotic SD translational systems also are known to operate in chloroplast genomes, which retain properly spaced SD sequences in the 5′-UTR of some genes and anti-SD sequences at the 3′ end of the 16S rRNA that are highly homologous with those of prokaryotes (Hirose and Sugiura 2004; Marín-Navarro et al. 2007; Peled-Zehavi and Danon 2007; Suzuki et al. 2011). In both dicots and monocots a distinct SD element in the 5′-UTR is spaced at the appropriate aligned distance from the initiation codon of the chloroplast *psaA-psaB-rps14* operon (Chen et al. 1992; Meng et al. 1988), which codes for critical subunits of PS I. In *Chlamydomonas*, site-directed mutagenesis of mRNA in the 5′-UTR of *psbA* (a vital component of PS II) established not only that the SD sequence was necessary for expression of this gene but also that changes in the secondary structure of that region influenced the level of translation substantially (Mayfield et al. 1994). Similarly, Zou et al. (2003) demonstrated that SD sequence recognition was involved in the translation of *psbA* in tobacco.

Drechsel and Bock (2011) concluded that SD sequences of the polycistronic operons of cpDNA generally are recognized most efficiently by the initial 5′ cistron but are recognized less-effectively by successive cistrons in an incremental fashion. Inefficacy of the internal SD sites of cpDNA operons has been postulated as an explanation for why most polycistronic transcripts are cleaved post-transcriptionally into monocistronic mRNAs (Drechsel and Bock 2011). Two notable exceptions occur where the polycistronic transcript is not cleaved in cpDNA: the 1.1 kb mRNA of the tetracistronic *psbE* operon, which codes for four small subunits (*psbE*, *psbF*, *psbJ*, and *psbL*) of PS II (Willey and Gray 1989), and the 5.2 kb mRNA of the tricistronic *psaA-psaB-rps14* operon (Meng et al. 1988), which codes for the two principal subunits of PS I (*psaA*, *psaB*) along with a ribosomal protein (*rps14*).

Acknowledging that the process of downstream SD site recognition in polycistronic cpDNA operons remains unclear, Drechsel and Bock (2011) suggested that in the case of *psaA*/*psaB*, the downstream SD recognition for *psaB* could be mediated by a regulator protein such as *TAB 2*. This conclusion is consistent with Marín-Navarro et al. (2007), who listed both *TAB 1* and *TAB 2* as trans-acting factors for *psaB*. Yet the influence of both factors in the regulation of *psaB* has been demonstrated only in *Chlamydomonas* (Stampacchia et al. 1997; Dauvillée et al. 2003; Rochaix et al. 2004), where *psaB* occurs not in an operon with *psaA*, but as a free-standing gene. Consequently, the target of *TAB 2* in *Chlamydomonas* actually occurs in the 5′-UTR of *psaB* (Barneche et al. 2006). Furthermore, an analogous protein in *Arabidopsis* (*ATAB 2*) interacts with the 5′-UTR of *psaA* (Barneche et al. 2006) rather than *psaB*. Reports of essential internal SD sites in other chloroplast operons (Suzuki et al. 2011) provide additional rationale to reconsider the importance of the *psaA*/*psaB* spacer region in translation. In particular, additional details regarding how internal SD sites influence translation of *psaB* in the operon remain to be elucidated.

Our interest in the structure of the *psaA-psaB-rps14* operon developed while assembling contigs derived from the cpDNA genome of *Najas flexilis*, a submersed aquatic angiosperm. A comparison to several other angiosperm chloroplast genomes revealed that the intercistronic spacer between *psaA* and *psaB* in *Najas* contained a three nucleotide (nt) insert relative to the other genera. Initial alignments also indicated that otherwise, the angiosperm spacer region appeared to be strongly conserved not only in length but also at the DNA sequence level, which included a distinct 6 nt SD site. Unlike previous appraisals, these observations suggested that the *psaA*/*psaB* spacer may be functional and encouraged a further survey of angiosperms

and other photosynthetic autotrophs to better characterize the extent of conservation associated with the region. The anomalous spacer of *Najas* also raised additional questions. Since *Najas* is a submersed aquatic plant, it is reasonable to presume that it might manifest unique photosynthetic modifications relating to its specialized ecological niche. Consequently, we considered whether the unusual spacer of *Najas* might be modified as an outcome of ecological adaptation.

The primary objective of our study was to conduct a comprehensive survey of *psaA*/*psaB* spacer variation not only among angiosperms, but across the evolutionary spectrum of photosynthetic autotrophs using the observed degree of conservation to assess the potential functionality of this region. We further evaluated functionality using mRNA in silico models to compare predicted structural features between common and rarer spacer variants. By summarizing this information, we hoped to better characterize the role of the *psaA*/*psaB* spacer region in the expression of vital photosynthetic genes, thus to further evaluate its importance in the evolution of land plants.

Methods

Sampling

A representative phylogenetic sample of *psaA*/*psaB* intergenic sequences was evaluated by comparing a total of 301 accessions representing major groups of photosynthetic eukaryotes. This survey included species from 10 protist groups (number of species indicated): Cercozoa (1), Alveolata (4), Euglenozoa (2), Glucophyta (1), Cryptophyta (1), Haptophyta (2), Rhodophyta (5), Heterokontophyta (10), Chlorophyta (21), and Charophyta (6) (Supplementary Table S1) and nine land plant divisions (Supplementary Table S2): Anthocerotophyta (1), Bryophyta (2) Marchantiophyta (3), Pteridophyta (8), Lycopodiophyta (4), Gnetophyta (3), Cycadophyta (1), Pinophyta (25), and Magnoliophyta (201). Initially, 244 *psaA/psaB* intergenic sequences were extracted from complete chloroplast genomes deposited in GenBank. Where no chloroplast genome was available, the relevant data were obtained by de novo sequencing of 57 additional taxonomic groups of interest. Each observed spacer sequence variant was assigned to a different type, which were then mapped onto two phylogenetic trees depicting: (1) relationships among major embryophyte groups and relevant ancestral lineages (charophytes) and (2) relationships among angiosperm orders. Both phylogenetic trees were drawn to depict consensus relationships summarized by Qiu et al. (2007) and Soltis et al. (2011). A phylogenetic tree that depicted *psaA*/*psaB* spacer length variation across a wide survey of embryophytes and various autotrophic protist lineages was also constructed following relationships depicted by Lewis and McCourt (2004) and ToL (Maddison and Schulz 2007).

De Novo Sequencing (*psaA*/*psaB* Spacer)

Material for de novo sequencing targeted those angiosperm orders lacking complete chloroplast genomes in GenBank and was acquired from plants available in the University of Connecticut EEB Plant Growth Facilities (Storrs, CT) or using DNA retained from previous systematic studies. Total genomic DNA was extracted from fresh leaf tissue or material preserved in saturated NaCl/CTAB solution (Rogstad 1992) using a standard protocol (Doyle and Doyle 1987) with minor modifications as described in Les et al. (2008).

Specific primers complimentary to *psaA* (NF1F 5′-GCGTCTCAGGTCATTCAGTC-3′) and *psaB* (NF3R 5′-ATACGACGAGTAGTGGGGTC-3′) were designed using NetPrimer (Biosoft International, Palo Alto, CA, USA) to target the intervening spacer. For each sample, the total 12.5 µl PCR reaction volume included 20 ng of total DNA, 0.15 mM of each dNTP, 0.2 µM of each primer (Promega, Madison, WI, USA), 1× Reaction Buffer (include $MgCl_2$ 1.5 mM), and 0.065 µl of Titanium Taq polymerase (Clontech, Mountain View, CA, USA). Thermal cycling involved 2 min initial denaturation at 94 °C, then 29 cycles of 30 s at 94 °C, 54 °C for 30 s, and 72 °C for 30 s; final extension was completed at 72 °C for 10 min. Amplicons were cleaned using an equal volume of PCR product and diluted (1:4) ExoSAp-IT (USB, Cleveland, OH, USA). The sequencing reaction was carried out in a final volume of 10 µl which included 1 µl BigDye® Terminator v1.1 Ready Reaction Mix (Applied Biosystems, Foster City, CA, USA), 2 µl of 5× Sequencing Buffer, and 1.25 mM primer. Amplified products were cleaned by gel filtration in lab prepared columns, using 600 µl of a 6.5 g Sephadex™ G-50 mix (GE Healthcare Bio-Sciences AB, Uppsala, Sweden) in 100 ml of double distilled water. Sequencing was conducted on an ABI PRISM 3130xl Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). Forward and reverse sequences were assembled using the program CodonCode Aligner (CodonCode Corporation, MA, USA). FASTA sequence files were imported to Geneious Pro 5.4 (Drummond et al. 2011) and aligned together with the extracted GenBank sequences using the MUSCLE plug-in.

Relative Divergence Levels (*psaA*, *psaB*)

In order to facilitate discussion, we designated each different *psaA/psaB* spacer length and/or sequence variant as a discrete type. The spacer most common in angiosperms

was denoted arbitrarily as "type I," and that most prevalent among embryophytes as "type II." The remaining variants (III–XVIII) essentially reflect the order they were detected. The relative extent of nt divergence was quantified for the *psaA* and *psaB* cistrons, which flanked the type I and type II spacers. Divergence analyses excluded other spacer types (represented by only one or few accessions), and newly sequenced taxa, which did not include complete *psaA* or *psaB* sequences.

Full *psaA* and *psaB* sequences were extracted from 114 (type I) or 38 (type II) complete chloroplast genomes available in Genbank using the start codons reported previously (142 sequences) or as annotated de novo (*Beta vulgaris*, *Brassica rapa*, *Limnanthes floccosa*; *Medicago truncatula*, *Pinus longaeva*). De novo annotations were made using ORF Finder and the Protein Code Prediction (EMBOSS, tcode tool) plug-in as implemented in Geneious 5.4 (Drummond et al. 2011). Five previously reported sequences (*Microlaena stipoides*, *Potamophila parviflora*, *Oryza australiensis*, *O. merdionalis*, *O. japonica*) identified different start codons for both genes, which were located 45 nt (*psaA*) or 12 nt (*psaB*) upstream from the most commonly designated start codon. Those reading frames were modified to match the predicted open reading frame (ORF) by removing the excess nt from the alignment. Sequences were aligned using the MUSCLE plug-in as implemented in Geneious Pro 5.4 (Drummond et al. 2011) and manually inspected for misalignments.

Uncorrected p-distances were calculated for each gene and accession using MEGA 5 (Tamura et al. 2011) and patristic distances were obtained with respect to maximum likelihood (ML) trees. *Amborella trichopoda* (GenBank: AJ506156) was selected as the reference to compute the uncorrected p-distance calculations and to root the angiosperm ML trees. A search for optimal ML trees was performed for each dataset (type I and type II spacers) and gene (*psaA* and *psaB*), using PHYML vr. 3.0 (Guindon et al. 2010) and raxml GUI vr. 1.1 (Stamatakis 2006; Silvestro and Michalak 2011), with the GTR + G + I model of sequence evolution determined under the AIC criterion using jModeltest (Posada 2008). The optimal ML trees were then imported into Geneious Pro 5.4 (Drummond et al. 2011) for calculation of the patristic distances.

In Silico mRNA Models

In order to evaluate potential functionality among *psaA*/*psaB* spacer variants, all *psaA*/*psaB* spacer sequences were converted to RNA using Geneious Pro 5.4 (Drummond et al. 2011) and treated as RNA transcripts for the secondary structure modeling. The predicted secondary structures of the mRNA transcripts were compared by modeling a region comprising the last 10 nt of the *psaA* coding region, the entire *psaA*/*psaB* spacer region, and first 12 nt of the *psaB* coding region.

We used the program Quikfold on the DINAMelt Server (Markham and Zuker 2005) to generate minimum ΔG predictions for the spacer. Two-dimensional secondary structures were modeled assuming linear sequence parameters and RNA version 3.0 energy rules at 37 °C, which approximates the average leaf surface temperatures reported for various flowering plants (Kuraishi and Nito 1980). In the case of the aquatic plant *Najas flexilis*, additional analyses under lower temperatures (which required version 2.3 energy rules) were performed. Results for each variant were corroborated using the additional modeling programs RNAfold (Hofacker et al. 2012) and RNAstructure version 5.3 (Reuter and Matthews 2010), which produced comparable results.

All modeled transcripts were evaluated for the presence of two putatively functional motifs: (1) a loop with projecting Shine-Dalgarno site and (2) a well-defined stem that interacted with adjacent coding regions to sequester the *psaB* start codon. Spacer transcript variants possessing both elements were categorized as having a conserved 2° structure; whereas, those lacking one or both element(s) were categorized as having an altered 2° structure. All spacer variants were color-coded accordingly when mapped onto the phylogenetic trees.

Results

De Novo Sequencing (*psaA*/*psaB* Spacer)

Using the primer set described above, we successfully amplified the target fragment containing the *psaA*/*psaB* spacer in a wide range of species ranging from ferns (*Dryopteris affinis*) to flowering plants. The resulting fragment varied in length from 379 nt (*Alvaradoa amorphoides*) to 414 nt (*Viscum minima*). By supplementing available Genbank data with 57 additional sequences, we were able to evaluate species representing 92 % of the 62 angiosperm orders represented in the ToL project (Maddison and Schulz 2007).

Analysis of the *psaA*/*psaB* Spacer Region (Protists)

A spacer length of 20–35 nt distinguished most photosynthetic protists surveyed except for diatoms and chlorophyte green algae, where the region length was conspicuously expanded (Fig. 1). Chlorophytes displayed a phylogenetic pattern, having short spacers (23 nt) in the basal prasinophyte orders (Mamiellales, Pyramimonadales) and an incremental length increase in more derived groups (Pseudoscourfieldiales, Ulvophyceae, Trebouxiophyceae). That trend culminated with Chlorophyceae, which possessed

the longest spacers (Fig. 1). Here, *psaA* and *psaB* are so widely spaced that they exist as free-standing genes and no longer comprise an operon.

Although nt sequences of the *psaA*/*psaB* spacer varied considerably among most of the photosynthetic protists, we surveyed two charophyte groups (Coleochaetophyceae, Zygnemophyceae) that possessed type II spacers, which were identical to those found throughout embryophyte lineages extending phylogenetically from bryophytes to angiosperms (Fig. 2).

Analysis of the *psaA*/*psaB* Spacer Region (Non-Angiosperm Embryophytes)

The length of the *psaA*/*psaB* spacer varied minimally among the non-angiosperm embryophytes surveyed. Like their charophyte ancestors, a 25 nt length was conserved in some conifers, cycads, ferns, marratioid ferns, horsetails, whisk ferns, lycophytes, mosses, and hornworts (Fig. 3; Supplementary Table S2). Spacer length varied from 17 to 26 nt in liverworts, gnetophytes, and some conifers (Fig. 3; Supplementary Table S2).

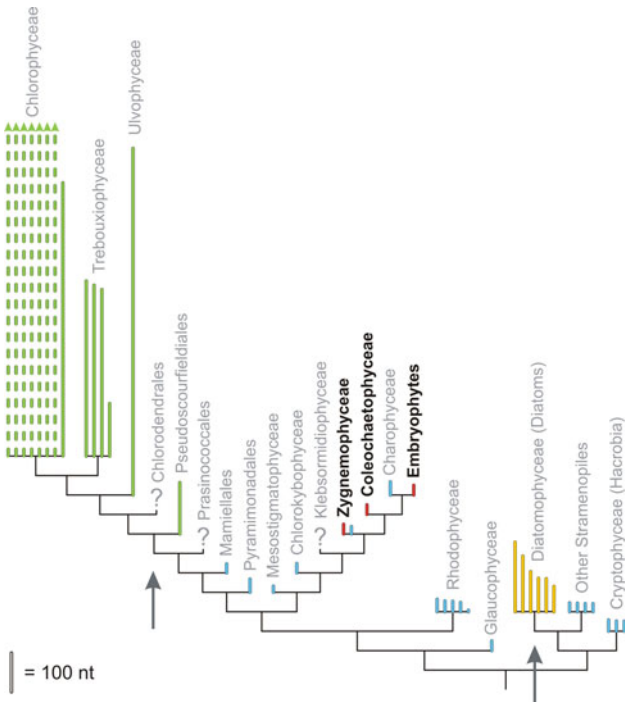

Fig. 1 Evolution of *psaA*/*psaB* spacer length in photosynthetic eukaryotes. The tree of relationships is synthesized from Lewis and McCourt (2004) and ToL (Maddison and Schulz 2007). Taxonomic names represent classes, or orders in the case of "Prasinophytes." *Solid colored vertical bars* represent lengths of the spacer region (nt) relative to scale shown. *Broken lines* represent taxa where *psaA* and *psaB* occur as widely separated, free-standing genes. *Blue* spacers <30 nt; *red* identical 25 nt spacer sequence ("type II") that is conserved among embryophytes and ancestral charophytes (groups in *bold*). Length amplifications (*arrows*) in green algae (*green*) and diatom (*gold*) clades indicate the ancestral condition of short spacers. *Question marks* indicate groups for which no data were available

Seventeen sequence variants were unique to non-angiospermous embryophytes, with each variant restricted to a single taxonomic group (Fig. 2). Most were confined to just three groups: conifers (G4–G9 variants), gnetophytes (G1–G3 variants), and lycophytes (P5–P7 variants). Five variants (II; G4–G7) were detected among the 21 *Pinus* accessions examined. The shorter variants (24 nt) were confined to the subgenus *Pinus*. Type G4 was detected in section *Pinus* and in the basal clade of section *Trifoliae* (taxonomy following Gernandt et al. 2005), while types G5–G6 (characterized by additional mutations), occurred only within the *Australis*/*Ponderosae* clade. The three types exclusive to Gnetophytes (G1–G3) were characterized by a reduced length (17–21 nt) and corrupted Shine-Dalgarno motif (Fig. 3). No size variation was detected among the ferns analyzed. One species (*Angiopteris evecta*; Marattiales) contained the type II spacer, whereas several leptosporangiate fern orders shared a variant (P1) that differed by just two mutations. Mosses possessed type I and II spacers; however, the type I spacer of *Physcomitrella* is regarded as convergent as it was not found in any other non-angiosperm. The B1 spacer of liverworts (Marchantiophyta) differed from type II by 1 nt insert (Fig. 3; Supplementary Table S2).

Analysis of the *psaA*/*psaB* Spacer Region (Angiosperms)

With few exceptions the length of the *psaA*/*psaB* spacer was conserved at 25 nt in most of the flowering plants we evaluated. Out of 166 angiosperm genera surveyed (201 species), only six genera contained species where the spacer length varied (Fig. 4; Supplementary Table S2). Two of the exceptional genera (*Cuscuta*, *Viscum*) comprise parasitic plants. Otherwise, the spacer length deviated only in *Alvaradoa*, *Carica*, *Dioscorea*, and in all examined species of the aquatic genus *Najas*.

Moreover, only 18 distinct sequence variants for the entire spacer region were found across a comprehensive phylogenetic range of angiosperms (Figs. 4, 5). The nt sequence of the spacer was identical for the majority of angiosperm genera surveyed (143/201; 71 %), with the most common type (type I) present in 124 different genera spanning 48 orders across the flowering plant phylogeny (Figs. 4, 5). The second most common sequence (type II) characterized 22 genera in ten orders (Figs. 4, 5) and deviated from type I by a single G/A transition (Fig. 4). Only 16 other sequence variants were found across the remaining angiosperms surveyed (Figs. 4, 5; Supplementary Table S2) and these occurred mainly in parasitic, aquatic, and cultivated plants.

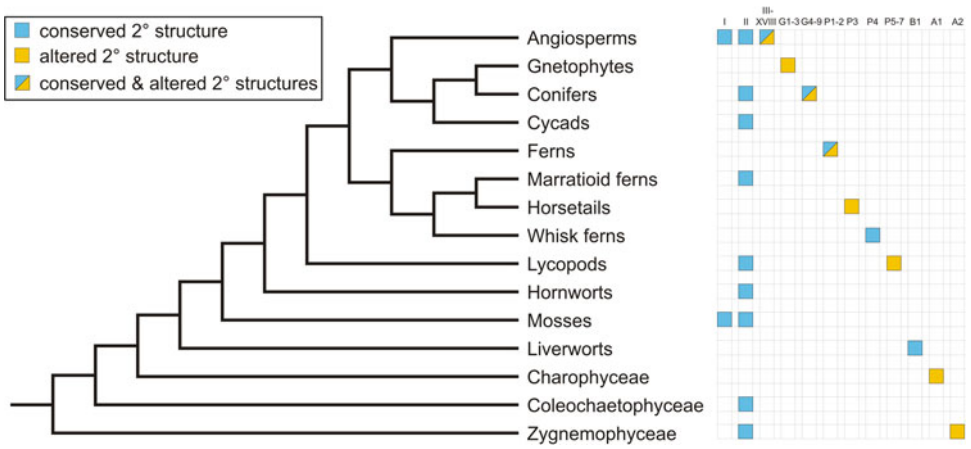

Fig. 2 Phylogenetic distribution of spacer types among embryophytes and charophytes. The tree of relationships is adapted from Qiu et al. (2007). Conserved *psaA/psaB* spacers exist in angiosperms (I–XVIII), gymnosperms (G), pteridophytes (P), bryophytes (B), and algae (A). Type II spacers extend from angiosperms to charophyte algae. Different spacer types do not always lose structural features (see Fig. 7); e.g., an extra base in liverworts (B1) or conifer variants (G8–G9) preserves essential type I and II features. *Blue* all structural elements retained; *gold* loss of at least one feature. A large number of types characterize aquatic protists ("algae"); however, two proposed sister groups of land plants (Coleochaetales and Zygnematales), include species with type II spacers

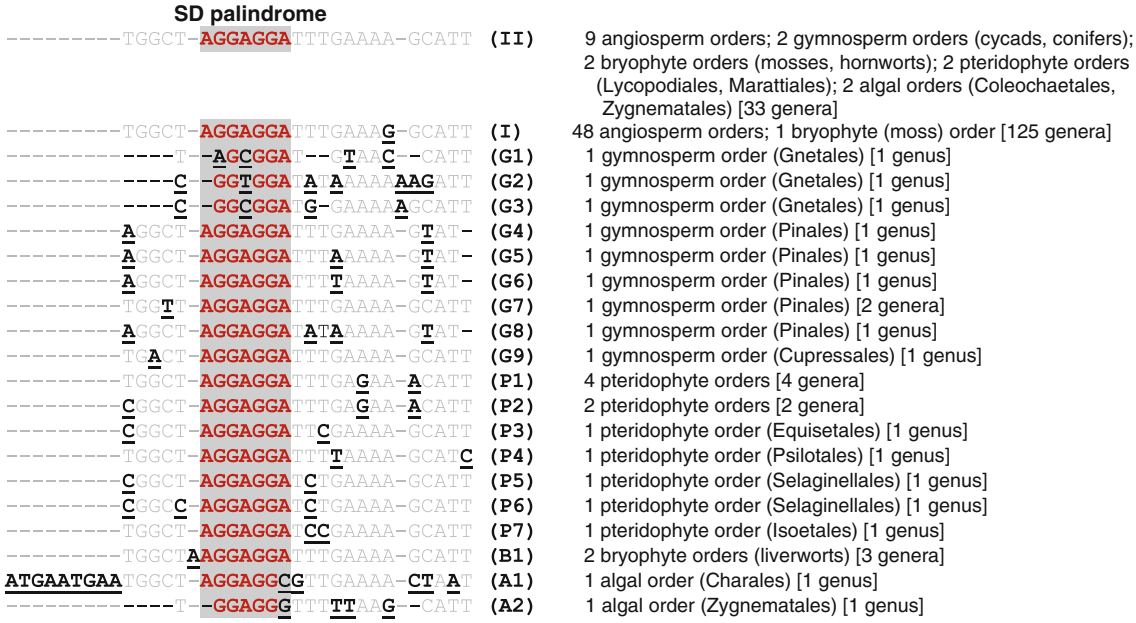

Fig. 3 *psaA/psaB* spacer types in embryophytes and charophyte algae. The common angiosperm types (I, II) and all observed gymnosperm (G), pteridophyte (P), bryophyte (B), and charophyte algae (A) types are shown. Nucleotide substitutions relative to the most widespread type (II) are *bold* and *underlined*; identical nucleotides are shown in *gray*; the Shine-Dalgarno palindrome is delimited by the *gray shaded box* (conserved sites in *red*)

Four variants occurred in parasitic genera (*Cuscuta*, *Viscum*) and five others were found in aquatic plants (*Astonia* and *Sagittaria*, *Ceratophyllum*, *Syringodium*, *Najas*, and *Thalassia*). The aquatic order Alismatales contained the largest total number of types (six). Cultivated species such as *Carica* (papaya), *Dioscorea*, *Erodium*, *Fagopyrum* (buckwheat), *Geranium*, *Glycine* (soybean), *Phaseolus* (bean), and *Vigna* (mung bean) contained most of the remaining (seven) sequence variants.

Relative Divergence Between the *psaA/psaB* Spacer vs. Flanking Cistrons

The *psaA* sequence alignment across 114 angiosperms with type I spacers was 2257 nt long and included five gaps. Three 1–2 nt gaps modified the reading frame, but were compensated by subsequent inserts (e.g., *Festuca arundinacea*) or posterior deletions (e.g., *Cucumis sativus*). A 6 nt deletion occurred in the parasitic *Cuscuta reflexa*. The

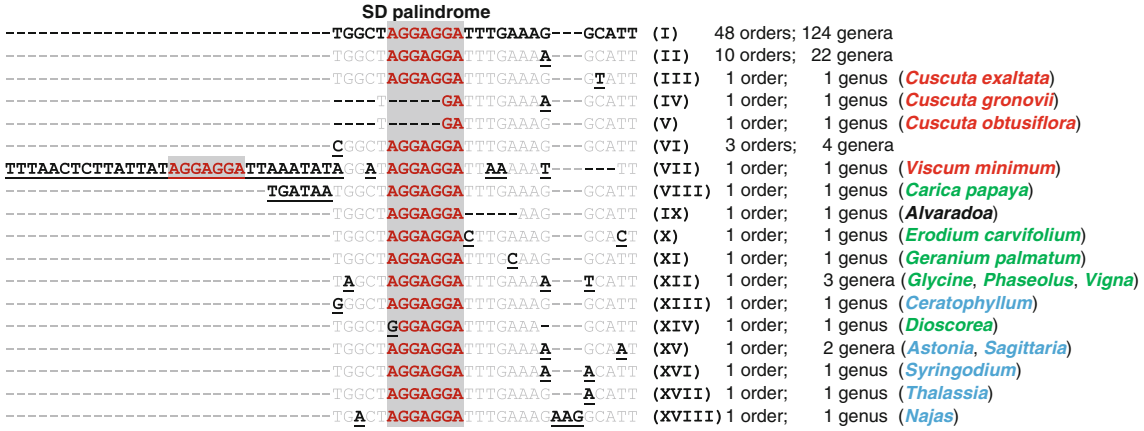

Fig. 4 *psaA*/*psaB* spacer types in angiosperms. Nucleotide substitutions relative to the most common type (I) are *bold* and *underlined*; identical nucleotides are shown in *gray*; the Shine-Dalgarno palindrome is delimited by the *gray shaded box* (conserved sites in *red*); taxonomic names represent parasitic (*red*), cultivated (*green*), or aquatic (*blue*) plants

alignment length of the *psaB* sequences was 2221 nt long and it included seven gaps, most of them in *F. arundinacea*. Only two other sequences (*Zea mays* and *Oncidium gower*) contained indels. The percentage of identical sites across alignments was 61.5 % (1387 nt) for *psaA* and 62.5 % (1386 nt) for *psaB*.

All ML trees obtained for type I angiosperms using *psaA* or *psaB* sequences were topologically equivalent to the one reported by ToL (Maddison and Schulz 2007). The uncorrected p-distances for *psaA* ranged from 0.041 (*Illicium oligandrum*) to 0.100 (*F. arundinacea*) while the patristic distances (PHYML/raxml) ranged from 0.115/0.121 to 0.387/0.414. The accession of *Amborella* also associated most closely with *I. oligandrum* for *psaB* (0.048/0.138/0.144; p-distance/PHYML/raxml); however, the parasitic *C. reflexa* had the most divergent sequence (0.116/0.405/0.426) for this gene. Genetic distances calculated among accessions sharing the same (type I) *psaA*/*psaB* spacer are summarized graphically in Fig. 6. We observed a similar trend among accessions sharing type II spacers, which also showed a relative increase in genetic divergence as taxonomic distance increased (data not shown).

In Silico mRNA Models

The resulting mRNA models obtained for most analyzed spacer variants shared several distinctive motifs: (1) a simple stem/loop structure with a projecting 7 nt SD palindrome; (2) complementary 5 nt regions (3′-UAAUA/5′-AUUAU); and (3) a conserved adenine (A) base. The latter immediately precedes the stop codon of *psaA* and forms an AUG motif across the base of the stem by combining with (and thereby sequestering) the UG portion of the actual *psaB* start codon (a feature we refer to as a "pseudo" start codon) (Fig. 7a). All three features characterize types I and II (Fig. 7a, b), the most common spacer variants. Conspicuous deviations occur in the secondary structure of some parasitic plants such as type V, in which all elements are retained except for a loss of the SD sequence in the loop (Fig. 7c); and in some aquatic plants like *Najas* (type XVIII), in which the SD site is embedded partly in the stem region and the pseudo start codon does not form (Fig. 8a). Over 85 % of the embryophyte spacers evaluated retained all three motifs described above (Supplementary Table S2). Modeling the variant *Najas* spacer transcript at lower temperatures (8–34 °C), to better approximate conditions of the waters in which these plants grow, generated in silico models with comparably conserved elements (Fig. 8b). The conformational change is facilitated by a region containing two 7 nt repeats, which flank the U of the start codon of *psaB* (Fig. 8b).

Discussion

The chloroplast *psaA-psaB-rps14* operon codes for the two largest subunits (A, B) of the PS I supercomplex (Amunts et al. 2010), which generates NADPH required by the Calvin cycle. This operon has been a fruitful region for evolutionary study where the conservation of various structural elements has disclosed a number of functional roles. Originating through gene duplication, the *psa* subunits are highly conserved, with some cyanobacteria and chloroplast sequences exhibiting as much as 80 % amino acid similarity (van der Staay et al. 2000). Strongly conserved regions include a group of four cysteines (two each provided by *psaA* and *psaB*) that comprise the Fe_4-S_4 Fx center, which is essential for functional electron transfer (Heathcote et al. 2003). In one case, the conservation of

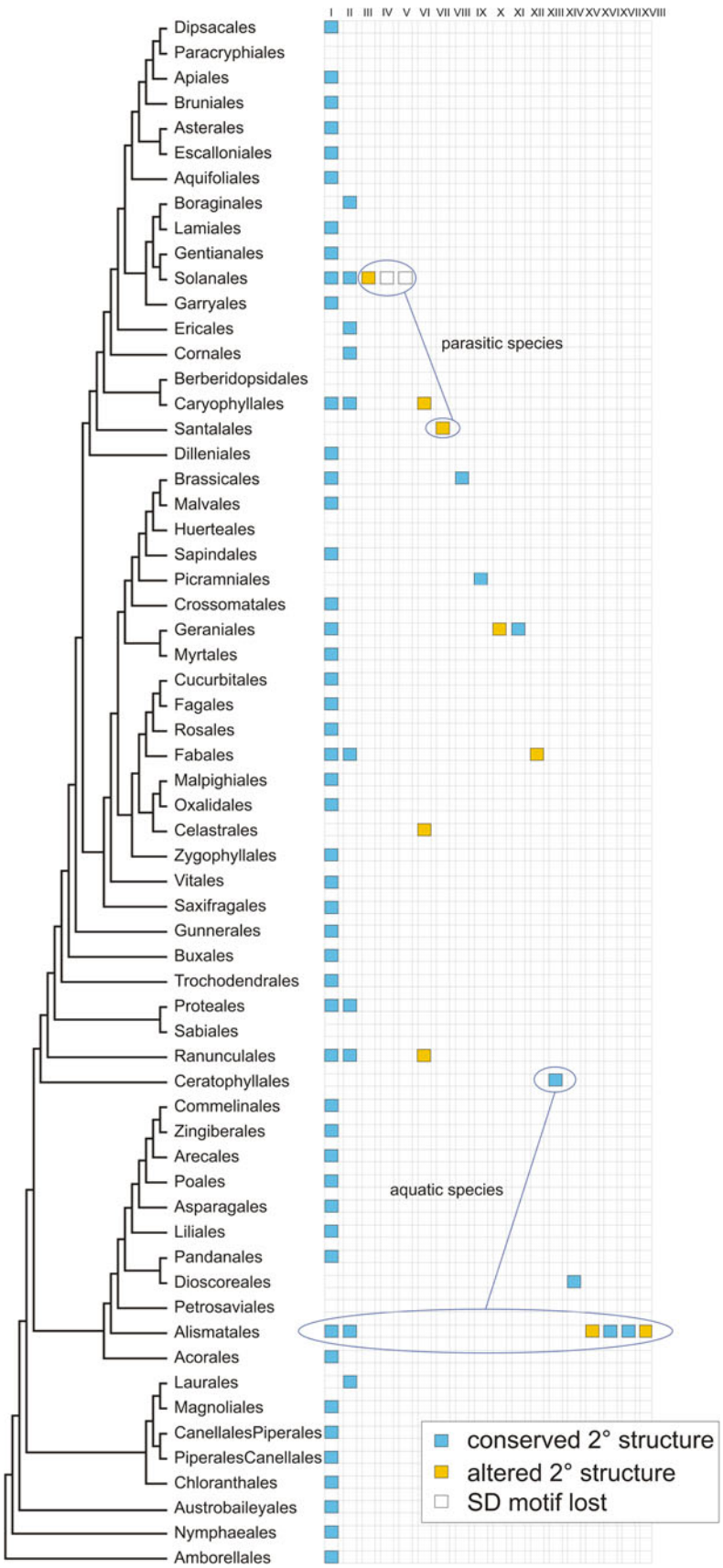

Fig. 5 Phylogenetic distribution of *psaA*/*psaB* intercistronic spacer types found among a survey of 166 genera representing 57 angiosperm orders. DNA sequences within each type (I–XVIII) are identical. A functional role is indicated for the spacer by the conservation of identical sequences (e.g., type I) in flowering plants since their origin some 125+ mya (tree adapted from Soltis et al. 2011). *Blue* spacer variants retaining all proposed structural motifs; *gold* and *white* variants lacking at least one structural element; *white* variants specifically lacking the SD sequence

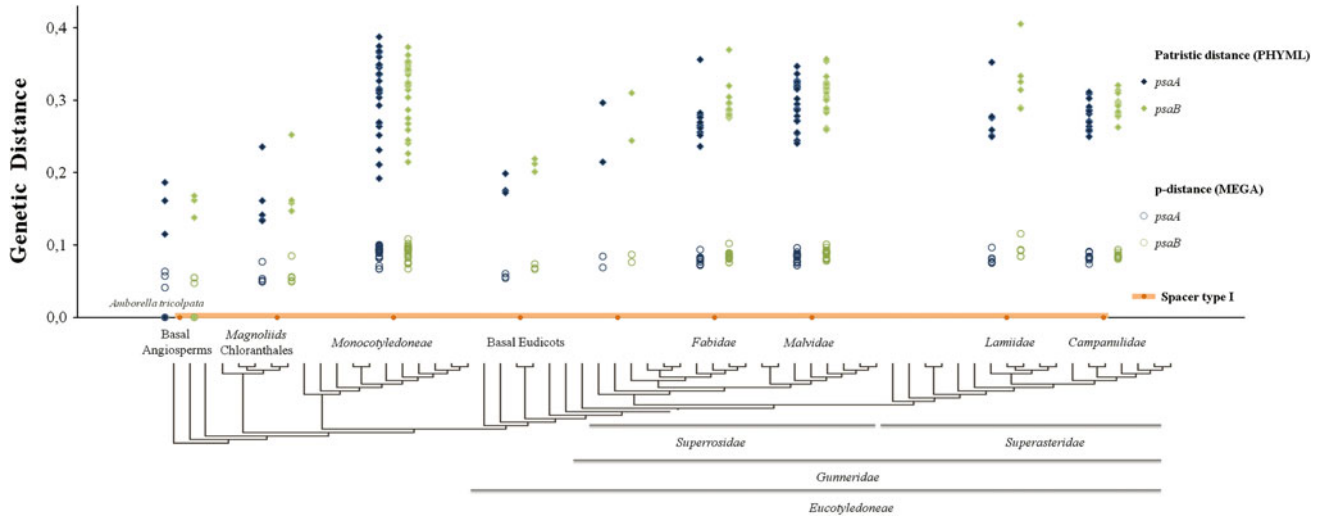

Fig. 6 Phylogenetic divergence of two highly conserved chloroplast genes (*psaA*, *psaB*) relative to the invariant, intervening spacer region. The 114 accessions evaluated (which share the type I spacer) were grouped taxonomically according to Soltis et al. (2011). The phylogenetic tree adapted from Soltis et al. (2011) is shown to illustrate the relationships among those groups. Genetic distances were calculated for *psaA* (*blue*) and *psaB* (*green*) as uncorrected p-distances (*open circles*) and patristic distances (*solid diamonds*) based on the optimal ML tree (see "Methods"). *Orange dots* and *line* represent the *psaA*/*psaB*, which was invariant across the same selection of taxa. Extreme conservation of the non-coding *psaA*/*psaB* spacer, even with respect to highly conserved chloroplast genes, argues strongly for its functional role

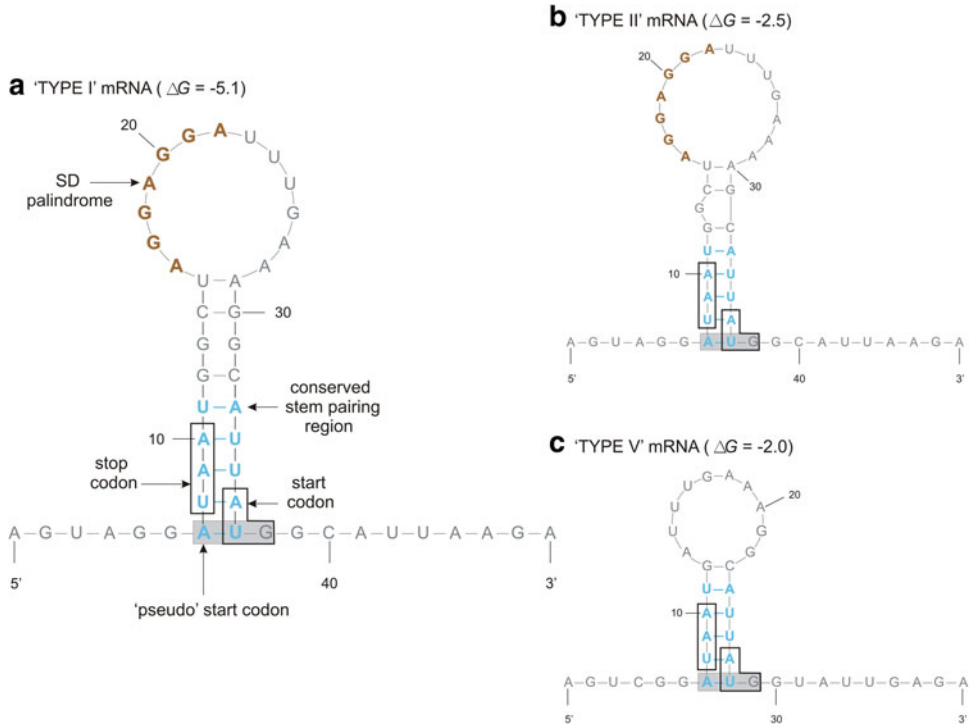

Fig. 7 In silico mRNA secondary models of the *psaA/psaB* spacer. **a** Type I stem-loop structure, the most prevalent in angiosperms. *Bold* nucleotides represent a 7 nt SD sequence in the loop. The "pseudo" start codon occurs at the stem base where the A of the GGA-codon and U of the *psaB* start codon interact. A–T bonds within a conserved pairing region stabilize the stem to sequester the start codon. **b** Type II (the most prevalent among embryophytes) differs from type I by one G/A transition but otherwise shares all structural elements. **c** Type V (in the parasitic *Cuscuta obtusiflora* and *C. gronovii*) lacks the SD site and 4 nt of the stem. *Brown* SD site; *blue* conserved stem pairing region; *boxes* start/stop codons; *gray shaded box* "pseudo" start codon

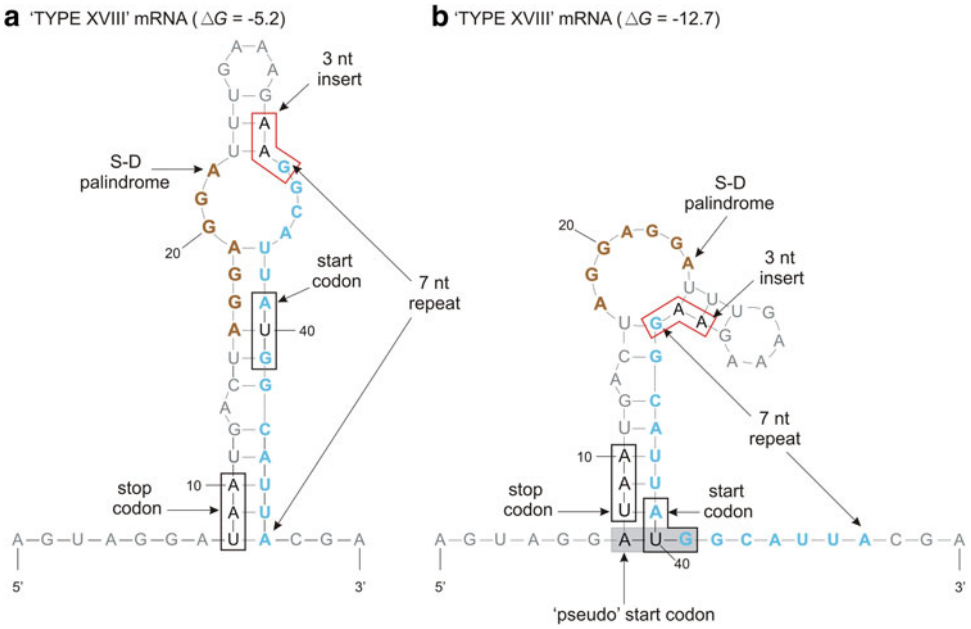

Fig. 8 Modified secondary structure in *Najas*, an aquatic plant. **a** When modeled at 37 °C, the type XVIII *psaA*/*psaB* spacer lacks all structural features because of a 3 nt insert, which causes the SD sequence to join with the stem. **b** When modeled at 20 °C, the spacer undergoes a conformational change in the secondary structure by exploiting a conserved 7 nt repeat (5′-GGCAUUA-3′), which reconstitutes the structural features. This conformation was the most probable in silico model rendered in the 8–24 °C range and coexisted with the modified structure, a, up to 34 °C. *Brown* SD site; *blue* conserved stem pairing region; *gray* "pseudo" start codon; *black boxes* start/stop codons; *red box* exclusive *Najas* insert

histidines in transmembrane segments of *psaA* and *psaB* revealed their role as axial ligands to P_{700} chlorophyll molecules (Redding et al. 1998). Since these genes code for critical PS I subunits, their conservation throughout the evolutionary history of photosynthetic autotrophs is entirely expected.

In contrast, no major function has been associated previously with the untranslated *psaA*/*psaB* spacer, which intervenes the *psaA* and *psaB* cistrons. Although the region is known to contain SD-like elements, this impression probably arose from earlier observations, which emphasized that the 16 nt distance between the SD sites and start codon of *psaB* exceeded the 5–13 nt range necessary to effectively initiate translation in prokaryotes (Ma et al. 2002; Hirose and Sugiura 2004). Consequently, several authors (Marín-Navarro et al. 2007; Drechsel and Bock 2011) presumed the interaction of nuclear regulatory factors in recognizing internal SD sites in the *psaA-psaB-rps14* operon despite the fact that the evidence of trans-factor regulation was derived from studies of *Chlamydomonas* (Stampacchia et al. 1997; Dauvillée et al. 2003; Rochaix et al. 2004), where *psaB* occurs as a free-standing gene, not in an operon. Our perspective of the *psaA*/*psaB* spacer is quite different, based on several lines of evidence that point to its functionality, at least in autotrophic embryophytes and their protist ancestors.

Many photosynthetic protists have relatively short spacers (<35 nt), with length amplifications (Fig. 1) restricted primarily to just two groups: diatoms (Diatomophyceae) and green algae (Chlorophyta). This pattern indicates that short spacers characterized ancestral eukaryotic autotrophs. All principal lineages leading to land plants are characterized by short spacers (19–34 nt). Although *Chara* itself presents a 34 nt spacer, two other charophyte classes (Coleochaetophyceae and Zygnemophyceae) included lengths (25 nt) identical to those conserved throughout the majority (184/192; 96 %) of embryophyte genera surveyed (Figs. 1, 3; Supplemental Table S2). This association may be significant because charophytes are implicated as embryophyte ancestors (Turmel et al. 2007). The few angiosperm taxa with atypical spacer lengths include parasitic plants (*Cuscuta* and *Viscum*), where divergence of photosynthetic-related regions is understandable (de Pamphilis and Palmer 1990), the cultivated *Carica* and *Dioscorea* (perhaps altered through artificial selection), and the aquatic genus *Najas*, which is discussed below. Otherwise, only *Alvaradoa* (Picramniaceae) exhibited an anomalous spacer length among the angiosperms compared (Fig. 4). In particular, the divergent spacer lengths observed in the two parasitic genera reciprocally indicate functionality in the autotrophic angiosperms, where the 25 nt length is highly conserved (160/164 genera; 98 %; Fig. 4; Supplemental Table S2).

Even more compelling than conserved *psaA*/*psaB* spacer length is the extreme level of nt sequence conservation present across embryophytes. Arguably, the type II spacer

is the ancestral embryophyte configuration, with various charophytes, bryophytes, lycophytes, ferns, gymnosperms, and angiosperms all possessing identical 25 nt sequences for the region (Figs. 2, 3). Here, 100 % sequence conservation across the roughly 0.5 billion years of divergence exemplified by these phylogenetically disparate groups (Becker and Marin 2009) is difficult to explain without implicating intense selection to retain functionality. Even the flanking *psaA* and *psaB* cistrons, which are regarded as highly conserved genes, are characterized by conspicuous divergence across the same phylogenetic spectrum. Similarly, these conserved genes exhibit moderate divergence with respect to the associated spacer type (I) just among angiosperms, which retain identical sequences throughout all major clades (Fig. 6). Given such extreme conservation of the *psaA/psaB* spacer, it is a nearly inescapable conclusion that some functional role must be involved.

Since the *psaA-psaB-rps14* operon of higher plants possesses a number of characteristics similar to prokaryotic translational mechanisms (i.e., PEP polymerase, presence of SD sites, and an unprocessed mRNA transcript), we reasoned that the low divergence of the spacer region could reflect the preservation of *cis*-acting features similar to those involved in prokaryotic gene expression (Kozak 2005). Several factors specifically indicate function as a *cis*-motif to initiate translation, which is the rate-limiting and mostly highly regulated phase of prokaryotic protein biosynthesis (Laursen et al. 2005).

One feature convincingly indicating such a regulatory function is the consistent presence of an SD sequence, which characterizes nearly all *psaA/psaB* spacer sequences examined (Figs. 3, 4). We found a distinct palindromic SD sequence (5′-AGGAGGA-3′) to be conserved within the *psaA/psaB* spacer of most embryophytes surveyed except for gnetophytes and parasitic plants (Figs. 3, 4), which are fully complimentary to the anti-SD site in the 16S rRNAs (5′-CCUCCU-3′). Although the 16 nt distance between most embryophyte SD sites and the start codon of *psaB* exceeds the 5–13 nt range that effectively initiates translation in prokaryotes (Hirose and Sugiura 2004; Ma et al. 2002), the strong 6 nt SD sequence in the *psaA/psaB* spacer presumably helps to overcome that distance. It is known that longer SD recognition sequences can ameliorate suboptimal spacing to maintain effective translation (Bonham-Smith and Bourque 1989) and even some shorter SD sequences (3–5 nt) can function within a broad (6–18 nt) distance range (Hirose and Sugiura 2004). Other pairing possibilities also exist due to extension of the SD sequence by an additional 3′ adenine, which occurs in most spacer sequences examined (Figs. 3, 4). The resulting palindrome provides two overlapping recognition regions complimentary to a shorter 4 nt motif (5′-UCCU-3′) within the ribosomal anti-SD site. Additionally, our in silico *psaA/psaB* spacer models (see below) consistently depict a loop that projects the SD sequence (Fig. 7), which also should further facilitate pairing with the 16 s rRNA. Selection to expose the SD site in the loop region is assumed, given that translation in other genes is reduced severely in mutated forms that position the SD site in configurations having higher secondary structure (Betts and Spremulli 1994).

Our in silico models of the embryophyte *psaA/psaB* spacer disclosed other conserved structural elements. The "pseudo" start codon (Figs. 7, 8) is a feature conserved at the base of our *psaA/psaB* spacer model structure. This motif is not unique to this region as we also observed a comparable feature in the secondary structures reported for the 5′-UTR of *atpH* in *Euglena* (Betts and Spremulli 1994). For *psaA/psaB*, our spacer models consistently paired the adenine of the extremely conserved GGA-glycine with the U of the *psaB* start codon. Although AT richness can bias third codon position substitutions in cpDNA (Shimada and Sugiura 1991), it is difficult to attribute the consistency of this site merely to chance as it is conserved universally across all 248 embryophyte *psaA* sequences evaluated. In contrast, the third position of the adjacent codon exhibited 15 substitution events. The adenine conservation likely results from its interaction with two essential features that delimit the spacer (*psaA* stop codon and *psaB* start codon) to maintain a consistently paired stem that resolves in the complementary 3′-UAAUA/5′-AUUAU regions (Fig. 7). Sequestering the start codon within the stem would prevent translation by precluding any binding of the 30S ribosomal subunit until the structure opened. The latter step would proceed as movement of the 70S ribosomal complex on the upstream cistron unfolded the mRNA. In essence, this mechanism emulates the prokaryotic system of "conformational masking" (Kozak 2005), whereby the level of expression is regulated by availability of the start codon.

Taken together, the conservation of length, sequence, and the secondary structure comprise persuasive evidence for the *cis*-acting role of the *psaA/psaB* spacer in translation of *psaB* in embryophytes. Weakly paired stems, like the A–U bonded regions of our model, are known to enable efficient translation from a structured ribosomal binding site (RBS) once the bonds dissociate (de Smit and van Duin 2003). Kinetics of rapid mRNA folding can be overcome by attachment of the 30S ribosome subunit in a flanking position prior to the opening of any secondary structure involving the RBS (de Smit and van Duin 2003). Similarly, we suggest that the *psaA/psaB* spacer structure situates the initiation complex using the projecting SD site to attract the 16S ribosomal subunit immediately before the onset of translation. Consequently, mutations or deletions disrupting the secondary structure should affect translation rate significantly by reducing exposure and availability of the SD site. Results consistent with this interpretation have been reported for

Nicotiana where the −25 to −5 sequence of *atpE* is essential for translation of *atpB-atpE*; mutations in that region reduced or arrested the expression of *atpE* (Suzuki et al. 2011). Our modeling of the same upstream region (but not of the mutant variants) depicted a structure equivalent to that of the *psaA/psaB* spacer. Therefore, it is likely that altered expression of *atpE* in *Nicotiana* mutants (Suzuki et al. 2011) occurs because of disrupted pairing in the stem-loop structure.

Intensively constrained divergence of the *psaA/psaB* spacer region throughout embryophyte evolution not only reflects strong selection for functionality but also urges the study of variants lacking the conserved features observed. In our survey, deviations in secondary structure ranged from loss of the SD site in the parasitic *Cuscuta* (Fig. 7c) to the apparent lack of functional features in the aquatic *Najas* (Fig. 8a). Although divergence of photosynthesis-related regions is not unusual in parasitic plants (de Pamphilis and Palmer 1990), it is interesting that most other spacer variants occurred either in cultivated species or aquatic angiosperms. It is possible that cultivation could result in the inadvertent selection of various spacer types, which might perform adequately under managed growing conditions, but perhaps would not fare as well in natural populations.

Interestingly, the lack of a conserved *psaA/psaB* spacer in most algal protists (except charophytes) raises questions of that region's functionality in these principally aquatic organisms (Fig. 1). However, we found that all major embryophyte taxa retain the type II spacer (Fig. 2), which first appears phylogenetically in charophytes, the presumed ancestors of embryophytes (Turmel et al. 2007). This observation indicates that the region was important to the evolutionary success of land plants as they colonized terrestrial habitats. Even in angiosperms, the most highly specialized embryophyte lineage, the prevalent spacer sequence (type I) differs from that assumed ancestral type by only a single transition, and retains all functional structural elements. However, for those angiosperm species that have since returned secondarily to a life in water, the spacer region is considerably divergent. Aquatic plants, which possess 7/18 (39 %) of the spacer types detected in angiosperms (Fig. 5), suggest an intriguing ecological explanation.

Unlike terrestrial plants, submersed species like *Najas* can grow in quite disparate conditions ranging from shallow, warm, sunny waters to cooler and darker waters, which can approach 4 °C and total light extinction (Hutchinson 1975). It is unclear how aquatic plants survive in such broad temperature and light regimes, especially as the latter conditions are known to degrade PS I in terrestrial plants (Tjus et al. 1999). Although in silico models indicate the loss of functional *psaA/psaB* spacer structure in *Najas* when modeled at temperatures typical of terrestrial plant photosynthesis (Fig. 8a), a structure possessing all conserved elements can be obtained for the same sequence when modeled at lower temperatures (Fig. 8b), which better approximate water conditions where these plants occur. Since heterogeneous pools of both functional and non-functional configurations would exist across a temperature gradient, these results indicate that *psaB* translation could be regulated accordingly. Presumably, lower temperatures would up regulate *psaB* translation overall, whereas higher temperatures would down regulate, thereby maintaining photosynthetic activity across the range of conditions. The conformational change in *Najas* is facilitated by a region containing two 7 nt repeats, which flank the U of the start codon of *psaB* (Fig. 8b). Since a conserved region of this magnitude would indicate a functional role, and because this repeat also occurs in most other spacer variants, its influence on the secondary structure should be evaluated further.

Although these hypotheses regarding the function of *psaA/psaB* spacer variants will require further testing, it is apparent by the extremely conserved nature of this region that some regulatory role is maintained. In particular, the possible modification of this region through selection to maximize photosynthetic rates under different environmental conditions certainly is an avenue deserving of more intensive investigation. Such endeavors may shed additional light on photosynthetic adaptations associated with the radiation of embryophyte lineages and potentially could elucidate novel factors involved in chloroplast gene regulation.

Acknowledgments We thank the University of Connecticut Center for Applied Genetics and Technology for their use of facilities and to C. Obergfell, R. O'Neill, and L. Strausbaugh for their technical assistance and support. We thank Prof. Jeffrey Palmer (University of Indiana) for critical reading of a previous version of this manuscript. Portions of this research were funded by grants from the Fulbright Foundation and Spanish MEC, NSF DEB–0841658, Irish Research Council for Science, Engineering and Technology (IRCSET) and University of Connecticut CLAS. Data reported in this paper are tabulated in the Supplementary Tables S1, S2, and archived in GenBank (KC129658–KC129716).

Conflict of interest The authors declare that they have no conflict of interest.

References

Allison LA (2000) The role of sigma factors in plastid transcription. Biochimie 82:537–548

Amunts A, Toporik H, Borovikova A, Nelson N (2010) Structure determination and improved model of plant photosystem I. J Biol Chem 285:3478–3486

Barneche F, Winter V, Crèvecoeur M, Rochaix J-D (2006) ATAB 2 is a novel factor in the signalling pathway of light-controlled synthesis of photosystem proteins. EMBO J 25:5907–5918

Becker B, Marin B (2009) Streptophyte algae and the origin of embryophytes. Ann Bot 103:999–1004

Betts L, Spremulli LL (1994) Analysis of the role of the Shine-Dalgarno sequence and mRNA secondary structure on the efficiency of translational initiation in the *Euglena gracilis* chloroplast *atpH* mRNA. J Biol Chem 269:26456–26463

Bock R (2007) Structure function and inheritance of plastid genomes. In: Bock R (ed) Cell and molecular biology of plastids. Springer, Heidelberg, pp 29–63

Bonham-Smith PC, Bourque DP (1989) Translation of chloroplast-encoded mRNA: potential initiation and termination signals. Nucleic Acids Res 17:2057–2080

Chen S-CG, Cheng M-C, Chung K-R, Yu N-J, Chen M-C (1992) Expression of the rice chloroplast *psaA-psaB-rps14* gene cluster. Plant Sci 81:93–102

Dauvillée D, Stampacchia O, Girard-Bascou J, Rochaix J-D (2003) Tab2 is a novel conserved RNA binding protein required for translation of the chloroplast *psaB* mRNA. EMBO J 22:6378–6388

de Pamphilis C, Palmer JD (1990) Loss of photosynthetic and chlororespiratory genes from the plastid genome of a parasitic flowering plant. Nature 348:337–339

de Smit MH, van Duin J (2003) Translational standby sites: how ribosomes may deal with the rapid folding kinetics of mRNA. J Mol Biol 331:737–743

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bul 19:11–15

Drechsel O, Bock R (2011) Selection of Shine-Dalgarno sequences in plastids. Nucleic Acids Res 39:1427–1438

Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J, Kearse M, Markowitz S, Moir R, Stones-Havas S, Sturrock S, Thierer T, Wilson A (2011) Geneious v5.4. http://www.geneious.com/

Fargo DC, Zhang M, Gillham NW, Boynton JE (1998) Shine-Dalgarno-like sequences are not required for translation of chloroplast mRNAs in *Chlamydomonas reinhardtii* chloroplasts or in *Escherichia coli*. Mol Gen Genet 257:271–282

Favory J-J, Kobayshi M, Tanaka K, Peltier G, Kreis M, Valay J-G, Lerbs-Mache S (2005) Specific function of a plastid sigma factor for *ndhF* gene transcription. Nucleic Acids Res 33:5991–5999

Gernandt DS, López GG, García SO, Liston A (2005) Phylogeny and classification of *Pinus*. Taxon 54:29–42

Gray MW (1993) Origin and evolution of organelle genomes. Curr Opin Genet Dev 3:884–890

Gruissem W, Greenberg BM, Zurawski G, Prescott DM, Hallick RB (1983) Biosynthesis of chloroplast transfer RNA in a spinach chloroplast transcription system. Cell 35:815–828

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59:307–321

Hajdukiewicz PT, Allison LA, Maliga P (1997) The two RNA polymerases encoded by the nuclear and the plastid compartments transcribe distinct groups of genes in tobacco plastids. EMBO J 16:4041–4048

Heathcote P, Jones MR, Fyfe PK (2003) Type I Photosynthetic reaction centres: Structure and function. Philos Trans R Soc Lond B 358:231–243

Hirose T, Sugiura M (2004) Functional Shine-Dalgarno-like sequences for translational initiation of chloroplast mRNAs. Plant Cell Physiol 45:114–117

Hofacker IL, Fontana W, Bonhoeffer S, Stadler PF, Lorenz R (2012) RNAfold 2.0.6. http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi

Hutchinson GE (1975) A treatise on limnology. Volume 3: Limnological botany. Wiley, New York

Kozak M (2005) Regulation of translation via mRNA structure in prokaryotes and eukaryotes. Gene 361:13–37

Kuraishi S, Nito N (1980) The maximum leaf surface temperatures of the higher plants observed in the Inland Sea area. Bot Mag (Tokyo) 93:209–220

Laursen BS, Sørensen HP, Mortensen KK, Sperling-Petersen HU (2005) Initiation of protein synthesis in bacteria. Mol Biol Rev 69:101–123

Les DH, Jacobs SWL, Tippery NP, Chen L, Moody ML, Wilstermann-Hildebrand M (2008) Systematics of *Vallisneria* (Hydrocharitaceae). Syst Bot 33:49–65

Lewis LA, McCourt RM (2004) Green algae and the origin of land plants. Am J Bot 91:1535–1556

Ma J, Campbell A, Karlin S (2002) Correlations between Shine-Dalgarno sequences and gene features and operon structures. J Bacteriol 184:5733–5745

Maddison DR, Schulz KS (eds) (2007) The tree of life web project. http://tolweb.org

Marín-Navarro J, Manuell AL, Wu J, Mayfield SP (2007) Chloroplast translation regulation. Photosynth Res 94:359–374

Markham NR, Zuker M (2005) DINAMelt web server for nucleic acid melting prediction. Nucleic Acids Res 33:577–581

Mayfield SP, Cohen A, Danon A, Yohn CB (1994) Translation of the *psbA* mRNA of *Chlamydomonas reinhardtii* requires a structured RNA element contained within the 5′ untranslated region. J Cell Biol 127:1537–1545

Meng BY, Tanaka M, Wakasugi T, Ohme M, Shinozaki K, Sugiura M (1988) Cotranscription of the genes encoding two P700 chlorophyll a apoproteins with the gene for ribosomal protein CS14: determination of the transcriptional initiation site by in vitro capping. Curr Genet 14:395–400

Peled-Zehavi H, Danon A (2007) Translation and translational regulation in chloroplasts. In: Bock R (ed) Cell and molecular biology of plastids. Springer, Heidelberg, pp 249–281

Plader W, Sugiura M (2003) The Shine-Dalgarno-like sequence is a negative regulatory element for translation of tobacco chloroplast *rps2* mRNA: an additional mechanism for translational control in chloroplasts. Plant J 34:377–382

Posada D (2008) jModelTest: phylogenetic model averaging. Mol Biol Evol 25:1253–1256

Qiu YL, Li L, Wang B, Chen Z, Dombrovska O, Lee J, Kent L, Li R, Jobson RW, Hendry TA, Taylor DW, Testa CM, Ambros M (2007) A non-flowering land plant phylogeny inferred from nucleotide sequences of seven chloroplast, mitochondrial and nuclear genes. Int J Plant Sci 168:691–708

Redding K, MacMillan F, Leibl W, Brettel K, Hanley J, Rutherford AW, Breton J, Rochaix JD (1998) A systematic survey of conserved histidines in the core subunits of Photosystem I by site-directed mutagenesis reveals the likely axial ligands of P_{700}. EMBO J 17:50–60

Reuter J, Matthews D (2010) RNAstructure: software for RNA secondary structure prediction and analysis. BMC Bioinformatics 11:129

Rochaix J-D, Perron K, Dauvillée D, Laroche F, Takahashi Y, Goldschmidt-Clermont M (2004) Post-transcriptional steps involved in the assembly of photosystem I in *Chlamydomonas*. Biochem Soc Trans 32:567–570

Rogstad SH (1992) Saturated NaCl-CTAB solution as a means of field preservation of leaves for DNA analyses. Taxon 41:701–708

Shimada H, Sugiura M (1991) Fine structural features of the chloroplast genome: comparison of the sequenced chloroplast genomes. Nucleic Acids Res 19:983–995

Silvestro D, Michalak I (2011) raxmlGUI: A graphical front-end for RAxML. Org Divers Evol. doi:10.1007/s13127-011-0056-0

Soltis DE, Smith SA, Cellinese N, Wurdack KJ, Tank DC, Brockington SF, Refulio-Rodriguez NF, Walker JB, Moore MJ, Carlsward BS, Bell CD, Latvis M, Crawley S, Black C, Diouf D, Xi Z, Gitzendanner MA, Sytsma KJ, Qiu YL, Hilu KW, Davis CC, Sanderson MJ, Olmstead RG, Judd WS, Donoghue MJ, Soltis PS (2011) Angiosperm phylogeny: 17 genes, 640 taxa. Am J Bot 98: 704–730

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690

Stampacchia O, Girard-Bascou J, Zanasco JL, Zerges W, Bennoun P, Rochaix JD (1997) A nuclear-encoded function essential for translation of the chloroplast *psaB* mRNA in *Chlamydomonas*. Plant Cell 9:773–782

Sugiura M, Hirose T, Sugita M (1998) Evolution and mechanism of translation in chloroplasts. Annu Rev Genet 32:437–459

Suzuki H, Kuroda H, Yukawa Y, Sugiura M (2011) The downstream atpE cistron is efficiently translated via its own cis-element in partially overlapping atpB-atpE dicistronic mRNAs in chloroplasts. Nucleic Acids Res 39(21):9405–9412. doi:10.1093/nar/gkr644

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739

Tjus SE, Møller BL, Scheller HV (1999) Photoinhibition of photosystem I damages both reaction centre proteins PSI-A and PSI-B and acceptor-side located small photosystem I polypeptides. Photosynth Res 60:75–86

Turmel M, Pombert J-F, Charlebois P, Otis C, Lemieux C (2007) The green algal ancestry of land plants as revealed by the chloroplast genome. Int J Plant Sci 168:679–689

van der Staay GW, Moon-van der Staay SY, Garczarek L, Partensky E (2000) Rapid evolutionary divergence of Photosystem I core subunits PsaA and PsaB in the marine prokaryote *Prochlorococcus*. Photosynth Res 65:131–139

Willey DL, Gray JC (1989) Two small open reading frames are co-transcribed with the pea chloroplast genes for the polypeptides of cytochrome b-559. Curr Genet 15:213–220

Zou Z, Eibl C, Koop H-U (2003) The stem-loop region of the tobacco psbA 5′UTR is an important determinant of mRNA stability and translation efficiency. Mol Genet Genom 269: 340–349