ISSN 1021-4437, Russian Journal of Plant Physiology, 2009, Vol. 56, No. 6, pp. 838–845. © Pleiades Publishing, Ltd., 2009.
Original Russian Text © A.V. Seliverstov, E.A. Lysenko, V.A. Lyubetsky, 2009, published in Fiziologiya Rastenii, 2009, Vol. 56, No. 6, pp. 926–934.

RESEARCH PAPERS

Rapid Evolution of Promoters for the Plastome Gene *ndhF* in Flowering Plants

A. V. Seliverstov[a,1], E. A. Lysenko[a,b,1], and V. A. Lyubetsky[b]

[a] *Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia*
[b] *Timiryazev Institute of Plant Physiology, Russian Academy of Sciences, Botanicheskaya ul. 35, Moscow, 127276 Russia; fax:7 (495) 977-8018; e-mail: genlysenko@mail.ru*

Received May 13, 2008

Abstract—Plastome is thought to be a very conservative part of plant genome but little is known about the evolution of plastome promoters. It was previously shown that one light-regulated promoter (LRP-*psbD*) is highly conserved in different flowering plant species and in black pine. We have undertaken search and demonstrated that gene *ndhF* is located in a plastome region that rarely underwent substantial rearrangements in terrestrial plants. However, alignment of sequences upstream *ndhF* suggests that promoters of this gene underwent comparatively rapid evolution in flowering plants. Probably, the ancestor of two basal Magnoliophyta branches (magnoliids and eudicotyledons) had the promoter P_A-*ndhF*, which was substituted with other promoters—P_B-*ndhF* and P_C-*ndhF*—in some phylogenetic lineages of dicots. We failed to reveal conservative sequences with potential promoters of –10/–35 type upstream *ndhF* genes of monocotyledonous plants, including nine representatives of the grass family (Poaceae). Multiple alignments of sequences from related taxa showed that the predicted *ndhF* promoters (A–C) underwent frequent mutations and these mutations are not only nucleotide substitutions but also small insertions and deletions. Thus, we can assume that at least some plastome promoters evolve rapidly.

Key words: flowering plants - plastome - promoter - evolution - ndhF

DOI: 10.1134/S1021443709060144

INTRODUCTION

Plastome is a part of plant genome located inside plastids and inherited probably from cyanobacteria. The set of plastome genes, their arrangement on the plastid DNA, and even intron positions are very similar in different phylogenetic lineages of terrestrial plants [1, 2]. The rate of synonymous and nonsynonymous nucleotide substitutions in plastome genes is lower than that in nuclear genes [3, 4]. Some noncoding plastome sequences are also highly conserved in algae and higher plants [5]. Therefore, plastome is thought to be a very conservative part of plant genome. However, in some cases plastome is able to evolve very rapidly. For example, the plastomes in various representatives of *Cuscuta* genus—green photosynthetic species and achlorophyllous ones—differ greatly in their size and gene number [6]. Thus we are sure that plastome genes are mainly conservative, but little is known about the evolution of their promoters.

Promoters regulate the first step of gene expression, i.e., the transcription initiation. Plastids of higher plants have two different transcription complexes: multisubunit RNA polymerase of cyanobacterial origin (plastid-encoded RNA polymerase, PEP) and monosubunit RNA polymerase homologous to the respective enzymes of bacteriophages and mitochondria (nucleus-encoded RNA polymerase, NEP) [7, 8]. NEP-dependent promoters were comprehensively investigated in [7], whereas PEP-dependent promoters have not been analyzed systematically in the last 20 years. As it was demonstrated in the early compilation [9], plastid promoters are very similar to bacterial promoters of σ^{70}-type: the consensus sequences of –10 (TAtaaT) and –35 (TTGaca) plastid boxes are the same as in bacteria. Evolutionary analysis was performed only for one PEP-dependent promoter and showed that light-regulated promoter of *psbD* gene (LRP-*psbD*) is highly conserved in different flowering plant species and in black pine [10]. Thus, according to limited available data, PEP promoters are evolutionary conservative, which enables us to extrapolate the results obtained with model species onto other plants. However, in the last few years it was demonstrated that transcription of the

[1] These authors contributed equally to this work.

Abbreviations: Indel—insertion or deletion; IR—inverted repeat; LRP-*psbD*—light-regulated promoter of *psbD* gene; NEP—nucleus-encoded plastid RNA polymerase; PEP—plastid-encoded RNA polymerase; ptDNA—plastid DNA; SSC—small single-copy region (in plastome).

same plastome genes is initiated at different sites in various plant species, which implies the existence of different promoters [11, 12]. Therefore, comprehensive evolutionary analysis is essential for understanding the degree of similarity in regulation of plastome gene transcription in plants.

NADPH plastoquinone oxidoreductase is a large protein complex of chloroplast thylakoid membranes; the complex participates in chlororespiration [13] and inhibits the generation of reactive oxygen species, thereby protecting the photosynthetic machinery [14]. Genes *ndhA-K* encode subunits of plastid NADPH plastoquinone oxidoreductase and usually are plastome genes; however in some plant species they disappeared from plastome [15]. We searched for these genes in annotated complete plastomes and revealed that they are present nearly in all higher plants and green algae belonging to Streptophyta phylum; plastomes lost *ndh* genes in few phylogenetic lineages only. The gene *ndhF* encodes the largest subunit of the complex and is a preferred object in molecular systematics. The nucleotide substitution rate in *ndhF* is twice higher than in plastome gene *rbcL* coding for the large subunit of Rubisco [16]. Hence, *ndhF* is one of actively evolving plastome genes. In *Arabidopsis thaliana, ndhF* is transcribed from the promoter of –10/–35 type, and initiation of transcription from this promoter is provided by a transcription factor Sig4 [17]. Analysis of the genomic environment revealed that *ndhF* lies in the conserved region that did not undergo substantial rearrangements in higher plants and characean algae belonging to Streptophytina group. All these data indicate that *ndhF* gene is an appropriate candidate for investigation of the plastome promoter evolution.

To retrace the evolution of *ndhF* promoters, we analyzed regions upstream this gene. In dicotyledonous plants and species from magnoliid group, we detected three potential promoters that are conservative in representatives of more than one family. We failed to find conservative promoter of –10/–35 type in many plant species, including all monocots studied. Alignments of sequences from related taxa showed that the predicted *ndhF* promoters (A–C) underwent mutations frequently and these mutations are not only the nucleotide substitutions but also small insertions and deletions. Thus, the data obtained suggest that *ndhF* promoters evolved rapidly, at least, in flowering plants.

MATERIALS AND METHODS

Complete plastome sequences presented at NCBI site (http://www.ncbi.nlm.nih.gov/genomes/ORGANELLES/plastids.html) were used for the analysis. All representatives of Streptophytina group having *ndhF* gene (84 species at the date of work completion) were searched for potential promoters. The search was limited to the region between *ndhF* and the nearest upstream gene: *rpl32* in flowering plants and *rpl21* in representatives of nonseed plants and characean algae (Fig. 1). The length of this region varied from 300 to 1000 bp and even more in different species.

Promoter prediction was based on the next procedure. First, we searched for sequences with maximal similarity to the minimal promoter of σ^{70}-type: TTGaca-17–18 bp–TAtaaT (more conservative nucleotides are given in capital letters; less conservative, in lower case). A penalty was imposed for deviations from consensus sequences: a "big" one was for deviations in more conservative positions and for shortening of spacer between –10 and –35 boxes to 16 bp; a "small" one was imposed for deviations in less conservative positions and for enlargement of the spacer up to 19 bp. The small penalty was one third of the big one. In addition, a "bonus" was set for the AT-reach spacer. The threshold level for sums of penalties and bonuses was chosen so that the number of promoter-like sequences selected at the first stage was minimal in each region upstream *ndhF*; in the results given below the number was no more than 5. For *A. thaliana*, sequences of –10 and –35 boxes just upstream the experimentally determined site of transcription initiation were used [17]. After that, multiple alignment of regions upstream *ndhF* was performed. The sequences were aligned by the program MultAl (A.A. Mironov, unpublished) and our procedure [5]; the alignments were corrected manually if needed. First, the sequences of representatives of a small taxon were aligned and then a group was enlarged as much as possible. Sequences being conservative in species of more than one family were considered as potential PEP-dependent promoters. Potential promoters from one group were usually located at similar distances from the coding *ndhF* region (Figs. 2–4); variability of nucleotides in a zone of potential promoter was lower than in adjoining regions. This was also considered to be a point in favor of prediction of a promoter. To verify the developed procedure, we used regions upstream *psbA, psbB, psbE, rbcL*, and *psaA* genes. The verification demonstrated that our procedure enables us to predict promoters coinciding with experimentally determined ones and does not lead to overprediction of promoters.

There is an explanation to be made why the conservativeness in representatives of more than one family was used as the criterion. Similarity of sequences inside one family may result from high conservativeness of plastid DNA and may be unrelated to functional significance of sequences. For example, in all 12 representatives of Brassicaceae family, we found a sequence that was very similar to minimal promoter of –10/–35 type: TTGACA–17 bp–TtTAAT (a nucleotide different from the consensus is lowercased). In *A. thaliana* this sequence is 198 bp upstream the coding region of *ndhF*; however, the transcription initiation in this region was not detected [17]. In a representative of a closely related family (Brassicales/Caricaceae), *Carica papaya*, the homologous region remarkably differs from the consensus: TTactA–17 bp–TtTAAc (–177 bp). It seems likely

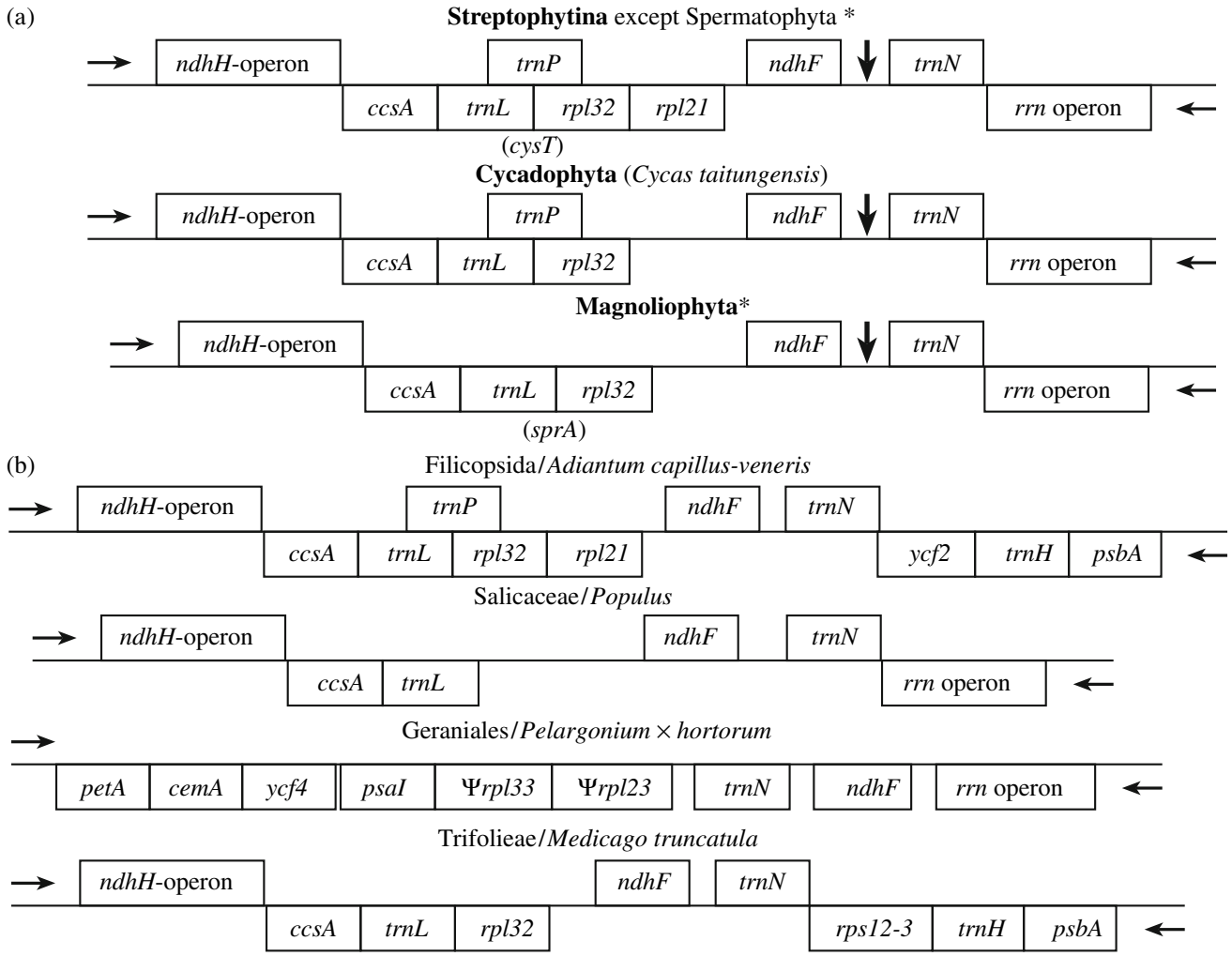

Fig. 1. Scheme of genomic surroundings of *ndhF* gene in representatives of Streptophytina group.

Horizontal arrows show the direction of transcription for genes located on different strands of ptDNA. Vertical arrows show the border between small single copy region, SSC (on the left) and inverted repeat IR_B (on the right). In some species, with expansion of inverted repeats, copies of genes (from one to three) from opposite side of SSC arise in this place (see text for explanation). The scheme shows an order of gene placement but does not reflect relative sizes of genes and intergene regions. Gene *trnP* does not overlap with *trnL* and *rpl32* genes.

(a) Scheme of the order of gene placement in groups. In annotated complete plastomes of Coniferophyta (*Cryptomeria, Pinus*), this region was subject to different rearrangements; therefore, no scheme was given for this group. (b) Exceptions. The largest/smallest taxa whose representatives may have a rearrangement are shown. For example, the recombination is observed in both representatives of genus *Populus*; therefore, it might be found in all representatives of the genus. The recombination is not observed in other representative of Malpighiales group (*M. esculenta*) that includes the family Salicaceae and genus *Populus*; consequently, the largest group that may have this rearrangement is Salicaceae.

Note:* In some representatives of hornworts (Anthocerotophyta) and liverworts (Marchantiophyta), there is an extra gene *cysT* between *rpl32* and *trnP* genes. In some members of Solanaceae family there is an extra gene *sprA* between *rpl32* and *trnL*.

that this region does not function as a promoter despite of similarity to the consensus and high degree of conservativeness inside Brassicaceae family. Conversely, the promoter before the transcription initiation site in *A. thaliana* is less similar to the consensus sequence [17] but is conservative in Brassicaceae species and *C. papaya* (Fig. 2). Therefore, the promoter prediction was made only for sequences whose conservativeness was revealed in representatives of more than one family.

We used systematics of plant and algae presented at NCBI site (www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html).

RESULTS

Genomic Analysis of ndh Genes

Before starting the analysis of sequences upstream *ndhF*, we investigated genomic organization of this

Fig. 2. Alignment of *ndhF* upstream regions in members of Brassicaceae family and other representatives of the taxon rosids. Potential promoter C (P_C-*ndhF*).

Potential –35 and –10 boxes are shown in bold type and underlined. Conservative nucleotides are given against gray background. The first nucleotide transcribed in *A. thaliana* is designated with white letter on a black background. At—*Arabidopsis thaliana,* Ah—*Arabis hirsute*, Ae c—*Aethionema cordifolium*, Ae g—*Aethionema grandiflorum*, Bv—*Barbarea verna,* Cb-p—*Capsella bursa-pastoris*, Cw—*Crucihimalaya wallichii*, Dn—*Draba nemorosa*, Lv—*Lepidium virginicum*, Lm—*Lobularia maritime*, No—*Nasturtium officinale*, Op—*Olimarabidopsis pumila.* Other species are given in the order of decreasing of their degree of relationship with Brassicacea family. Cp—*Carica papaya* (Brassicales), Cs—*Citrus sinensis* (eurosids II), Gos—*Gossypium* sp. (*G. barbadense, G. hirsutum*) (eurosids II), Eg—*Eucalyptus globulus* (rosids), Vv—*Vitis vinifera* (core eudicotyledons), Po—*Platanus occidentalis* (eudicotyledons). K means G (in *G. hirsutum*) and T (in *G. barbadense*).

Note:* There is an insertion of 20 bp in *C. wallichii*, which is not shown for simplicity; the insertion is identical to the fragment underlined and located right before it; thus, they are a tandem repeat without interval.

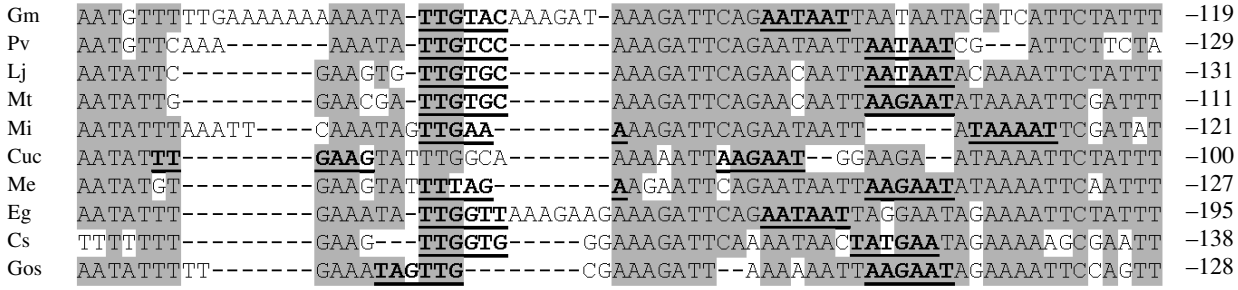

Fig. 3. Alignment of *ndhF* upstream regions in representatives of the taxon rosids. Potential promoter B (P_B-ndhF).

Potential –35 and –10 boxes are shown in bold type and underlined. Conservative nucleotides are given against gray background. Fabaceae/Papilionoideae: Gm—*Glycine max*, Pv—*Phaseolus vulgaris,* Lj—*Lotus japonicus*, Mt—*Medicago truncatula*; members of other groups from the taxon eurosids I: Mi—*Morus indica*, Cuc—*Cucumis sativus*, Me—*Manihot esculenta*; members of other groups from the taxon rosids: Cs—*Citrus sinensis* (eurosids II), Gos—*Gossypium* sp. (eurosids II), Eg—*Eucalyptus globulus* (Myrtales).

gene and all *ndh* genes in general. The performed analysis showed that *ndh* genes are absent from plastomes of living organisms in taxa Alveolata, Cercozoa, Cryptophyta, Euglenozoa, Glaucocystophyceae, Haptophyceae, Rhodophyta, and stramenopiles. Altogether 18 complete plastomes were studied including 5 plastomes of red algae. In Chlorophyta phylum comprising the majority of green algae, *ndh* genes were absent in

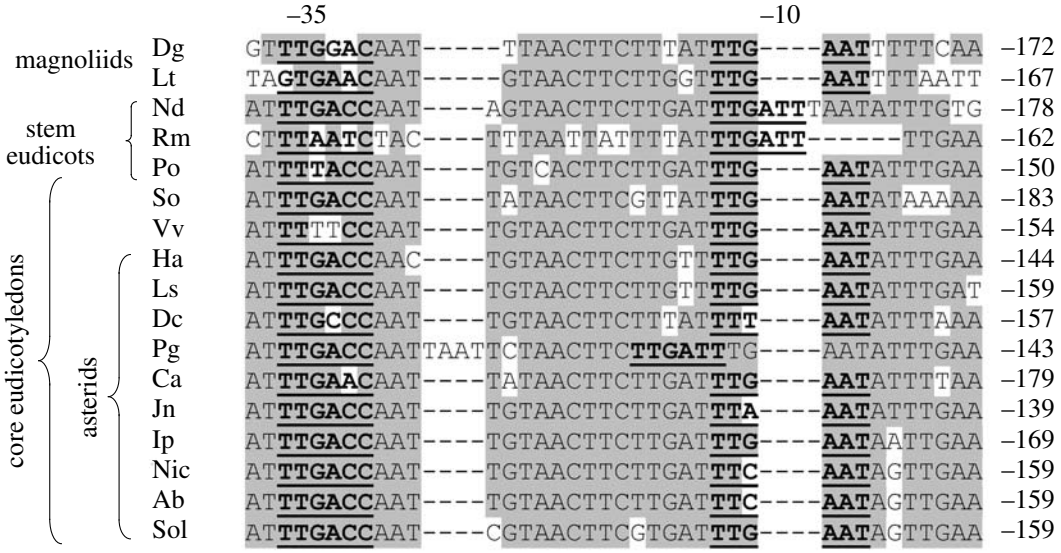

Fig. 4. Alignment of *ndhF* upstream regions in representatives of eudicotyledons and magnoliids. Potential promoter A (P_A-*ndhF*). Potential –35 and –10 boxes are shown in bold type and underlined. Conservative nucleotides are given against gray background. Dg—*Drimys granadensis*, Lt—*Liriodendron tulipifera*, Nd—*Nandina domestica*, Rm—*Ranunculus macranthus*, Po—*Platanus occidentalis*, So—*Spinacia oleracea*, Vv—*Vitis vinifera*, Ha—*Helianthus annuus*, Ls—*Lactuca sativa*, Dc—*Daucus carota*, Pg—*Panax ginseng*, Ca—*Coffea arabica*, Jn—*Jasminum nudiflorum*, Ip—*Ipomoea purpurea*, Ab—*Atropa belladonna*, Nic—*Nicotiana* sp. (*N. tabacum, N. tomentosiformis, N. sylvestris*), Sol—*Solanum* sp. (*S. bulbocastanum, S. lycopersicum, S. tuberosum*).

plastomes of eight out of nine species, whereas the plastome of *Nephroselmis olivacea* contained nearly all *ndh* genes except *ndhJ*. On the contrary, in Streptophyta phylum that includes land plants and a small number of green algae species, *ndh* genes occurred in nearly all representatives. Out of 98 species whose complete plastomes could be found in the corresponding part of NCBI base, *ndh* genes were absent (or presented by pseudogenes) in nonphotosynthesizing parasitic plants from *Epifagus* and *Cuscuta* genera. As to photosynthesizing organisms, these genes were absent from plastomes of a liverwort *Aneura mirabilis* (Marchantiophyta), in representatives of a conifer family Pinaceae, (*Pinus koraiensis* and *P. thunbergii*), and in a representative of monocotyledonous plant order Asparagales (*Phalaenopsis aphrodite*). At the same time, in liverwort *Marchantia polymorpha* from Marchantiophyta group, in a species *Cryptomeria japonica*, AP009377 from the family Cupressaceae closely related to Pinacea, and in members of other orders of monocotyledonous plants, the plastomes contained all 11 *ndh* genes.

The gene *ndhF* is located in a small single copy (SSC) region next to inverted repeat B (IR_B). In some species the inverted repeats may expand. If IR_B overlaps with *ndhF*, its complete or partial copy arises in IR_A next to the border with SSC region (e.g., in *Oenothera* species). If IR_A expands towards SSC, the gene copies emerge on the opposite side of SSC between *ndhF* and *trnN* (*ycf1, rps15, ndhH*, or *ndhA*; in characean algae, *chlL, chlN* or *ycf20*) (Fig. 1). Expansion of IR does not affect the *ndhF* promoter region. In many species, gene copies that arose as a result of IR expansion are pseudogenes (e.g. a new *ndhF* copy in *Oenothera*).

Genomic environment of *ndhF* vary in different phylogenetic branches of Streptophyta phylum. However, in Streptophytina group, which includes characean algae and land plants, this region is quite conservative and probably rarely underwent substantial rearrangements (Fig. 1). One of such rearrangements was the loss of *rpl21* gene, which most likely occurred in the last common ancestor of seed plants, because this gene is absent in representatives of angiosperms (Magnoliophyta) and two groups of gymnosperms: Cycadophyta (*Cycas taitungensis*) and Coniferophyta (*C. japonica, P. koraiensis*, and *P. thunbergii*). In some species, extra rearrangements of *ndhF* surroundings took place, but these recombinations are apparently characteristic of small taxa. For example, reorganization in *Medicago truncatula* may be typical of its tribe Trifolieae only, because, in members of other tribes belonging to the same subfamily Papilionoideae (*Glycine max, Lotus japonicus*, and *Phaseolus vulgaris*), the organization of this region is similar to that in the majority of flowering plants. Of all these rearrangements (Fig. 1), only the loss of *rpl32* gene in representatives of *Populus* genus (*P. alba* and *P. trichocarpa*) and the inversion of *ndhF* gene in *Pelargonium* × *hortorum* affect the region upstream *ndhF* and could influence the promoter zone.

Thus, the gene *ndhF* is located in quite a conserved plastome region. In the course of evolution, substantial rearrangements that may have changed the promoter zone occurred probably in the last common ancestor of

Lm	TAACGCAA**TTGGAA**CTTTTTGAATTATTTAA**AATAAT**GGATGC	–158
De	TATTAGAT**TTGAAG**CCAATTAATAAAATTC**CATAAT**GAGTCA	–385
Aco	TAATAAAT**TTGATA**ATAACTTAACTCCAAA**TAAACT**AACTAG	–266
Pop*	CGAGC**TTGAAG**GAGTTAATTCAATGTAA**TATTTT**TTTAG	–493/–552
Gos	TAATAA**TTGATA**TTTTCAGAAAAATATTAC**TATAAT**GAAAAAAAGAATT	–246
At**	AATCTT**TTGACA**GTAACTTAGTAATATTT**TTTAAT**CATTTCTAATTTT	–185

Fig. 5. Sequences of –10/–35 type found in some plants.

Potential –35 and –10 boxes are shown in bold type and underlined. Lm—*Lemna minor,* De—*Dioscorea elephantipes*, Aco—*Acorus* sp. (*A. americanus, A. calamus*), Pop—*Populus* sp. (*P. alba, P. trichocarpa*), Gos—*Gossypium* sp., At—*A. thaliana.*

Notes: * Distances from coding region: –493 (*P. trichocarpa*)/ –552 (*P. alba*).

** The sequence is conservative in all Brassicaceae species (see Materials and Methods for explanations).

seed plants (loss of *rpl21*) as well as in ancestors of small phylogenetic groups of dicotyledonous plants: genus *Populus* (loss of *rpl32*) and *P. hortorum* (inversion of *ndhF*) (Fig. 1).

Search for Potential Promoters

In *A. thaliana* the transcription initiation site of *ndhF* gene was experimentally determined [17]. The analysis of sequences revealed that promoter of –10/–35 type found in *A. thaliana* is conservative in members of Brassicaceae family, in a species from closely related family Caricaceae (*C. papaya*), in representatives of other families from the taxon eurosids II (*Citrus sinensis, Gossipium barbadense,* and *G. hirsutum*), and also in certain species from more diverged groups (*Eucalyptus globulus, Vitis vinifera*, and *Platanus occidentalis*) (Fig. 2). We denoted this group of potential promoters with letter C (P_C-*ndhF*). The alignment of sequences presented in Fig. 2 shows that, in the course of evolution, this promoter underwent remarkable changes even inside Brassicaceae family. Not only mononucleotide substitutions but also small deletions and insertions probably occurred in the promoter zone, and they changed the distance between promoter boxes and should have shifted one of the boxes. Alignment of sequences from the most diverged species (*E.globulus, V. vinifera*, and *P. occidentalis*) led us to suggest that they represent a more ancient variant of this promoter and that the promoter revealed in *A. thaliana* arose after the insertion of about 11 bp, which occurred in the last common ancestor of order Brassicales or the group eurosids II. We failed to find promoters homologous to P_C-*ndhF* in other species; therefore, we continued searching for other potential promoters of *ndhF* gene.

Another potential promoter was revealed in representatives of the taxon eurosids I and some other members of subclass rosids (Fig. 3); this group of promoters was designated with letter B (P_B-*ndhF*). This predicted promoter is less conservative than P_C-*ndhF*, even inside one subfamily (Papilionoideae) of legume plants. Insertion and deletions, which should cause the shift of boxes, were also found in the zone of P_B-*ndhF* promoters. Variability of 3'-end of –35 box (less conservative in the consensus promoter as well) and replacement of the first nucleotide in –10 box with A (instead of T in the consensus) are general characteristics of promoters in this group. Besides, shortening of spacer between boxes to 16 bp was observed in three species (Fig. 2). In *Gossipium* species only two nucleotides corresponded to the consensus motif. We failed to find promoters homologous to P_B-*ndhF* outside the subclass of rosids.

The potential promoter A (P_A-*ndhF*) was revealed in members of different branches of the core eudicotyledons taxon (asterids, Caryophyllales, and Vitales) as well as in representatives of stem eudicotyledons and even in some species belonging to a separate branch of flowering plants, magnoliids (*Drimys granadensis, Liriodendron tulipifera*) (Fig. 4). The zones of promoters P_A-*ndhF* are much more conservative. However, substitutions of nucleotides in promoter boxes as well as the insertion that should induce the shift of one of the boxes (in *Panax ginseng*) were also observed in this group.

We failed to find the conservative promoter of σ^{70}-type among monocotyledonous plants. We obtained a good alignment of sequences for the majority of cereals, both within this taxonomic group and with sequences of dicots having P_A-*ndhF* promoter (data not shown). This alignment demonstrates that, in plants from grass family, the potential promoter homologous to P_A-*ndhF* is absent and, generally, there is no conservative consensus promoter of –10/–35 type. Sequences of nine species belonging to eight genera representing both main clades (BEP and PACCAD) of Poaceae family were used for the analysis. In species from other orders of monocotyledonous plants—Acorales (*Acorus americanus, A. calamus*), Alismatales (*Lemna minor*), and Dioscoreales (*Dioscorea elephantipes*)—sequences similar to the consensus promoter of σ^{70}-type were detected (Fig. 5), but these sequences bear no homology to each other nor to conservative promoter-like sequences from dicotyledonous plants.

We found no conservative promoter of the consensus type in representatives of other branches of angiosperms—basal Magnoliophyta (*Amborella trichopoda, Illicum oligandrum, Chlorantus spicatus,*

Nuphar advena, Nymphaea alba) and Ceratophyllales (*Ceratophyllum demersum*)—as well as in a gymnosperm *C. taitungensis*, in seven species of nonseed plants, and in two species of characean algae.

DISCUSSION

We investigated sequences upstream *ndhF* plastome genes and revealed that flowering plants have no common promoter. In members of two out of five basal branches of angiosperms (magnoliids and eudicotyledons), a conservative potential promoter P_A-*ndhF* was found (Fig. 4); in one of dicot subclasses, rosids, this promoter was not detected but two other conservative potential promoters were revealed: P_B-*ndhF* (Fig. 3) and P_C-*ndhF* (Fig. 2). We failed to find conservative promoters similar to the consensus promoter of –10/–35 type outside magnoliids and eudicotyledons. The absence of a common conservative promoter cannot result from frequent genomic reorganizations because *ndhF* is located in quite a stable region of plastome (Fig. 1).

Conservative promoter-like sequences were revealed in the majority of representatives of dicots (eudicotyledons). In this group, conservative promoters were not detected in *Buxus microphylla, P. hortorum*, and two *Populus* and five *Oenothera* species. In *P. hortorum* and *Populus* substantial plastome rearrangements, which could affect the promoter zone of *ndhF*, were observed (Fig. 1); *Oenothera* genus is noted for high variability of ptDNA [18]. Probably, P_A-*ndhF* is the most ancient promoter of dicotyledonous or even flowering plants, because it was found not only in representatives of eudicotyledons but also in magnoliids (Fig. 4). An evolutionary more ancient promoter was possibly disrupted as a result of *rpl21* deletion that probably had occurred in a common ancestor of seed plants (Fig. 1). This may be the reason why we failed to find sequences homologous to P_A-*ndhF* in nonseed plants. Promoters P_B-*ndhF* and P_C-*ndhF* must have arisen later; however, the promoter P_C-*ndhF* was revealed not only in representatives of the subclass rosids but also in other groups including the stem eudicotyledons (*P. occidentalis*) (Fig. 2). In some species we detected two promoters: P_A-*ndhF* and P_C-*ndhF* (*P. occidentalis, V. vinifera*), P_B-*ndhF* and P_C-*ndhF* (*E. globulus, C. sinensis, Gossipium* sp.) (Figs. 2–4). It is conceivable that new promoters did not replace the elder ones but coexisted with them, and later in the evolution the majority of species lost one of them. The fact that "new" promoters (P_B-*ndhF* and P_C-*ndhF*) were detected in most members of the subclass rosids must be due to the loss of the "old" promoter (P_A-*ndhF*) by a common ancestor of this group.

The analysis of multiple alignments of sequences from related species led us to conclude that all predicted promoters (P_A-, P_B-, and P_C-*ndhF*) underwent changes in the course of evolution. Mutations that had to modify promoter features—nucleotide substitutions in –10 and –35 boxes, insertions and deletions in spacer between boxes—were observed in each group of promoters (Figs. 2–4). These observations are in good agreement with the information that small insertions or deletions (indels) are the major force driving the evolution of plastome noncoding sequences and that mononucleotide indels represent about 40% of total frequency; 2-, 3-, 4-, or 5-nucleotide indels constitute about 10% each; and, as the size of indels grows further, their frequencies decline gradually [19]. However, these calculations were based on the analysis of all noncoding sequences. Our data suggests that at least some regulatory *cis*-elements in plastome are targets for small indels, which probably do not destroy them but change their properties.

We succeeded in revealing conservative potential promoters in taxonomic groups where complete plastome sequences are annotated for many representatives. The fact that we failed to find conservative promoters for many groups of organisms (monocots, gymnosperms, ferns and horsetails, mosses, lichens, and characean algae) may be due to a small number of sequences presented in the analysis for these groups. It is possible that each of these groups contain their own conservative PEP-dependent promoters. However, the negative result obtained after analysis of seven sequences of nonseed plants and two characean algae makes us to conclude that nonseed representatives of Streptophytina group have no common conservative promoter for *ndhF* gene.

Investigation of sequences from nine Poaceae species led us to suggest that conservative PEP-dependent promoter is absent inside this family. As it was shown, *ndhF* is transcribed in cereals and its transcription is activated under heat shock [20, 21]. Apparently, in members of grass family, the gene is transcribed either by PEP from the nonconsensus type promoter, as it was shown for LRP-*psbD*, [10, 22] or from the NEP-dependent promoter.

Thus, the analysis of promoter zone of *ndhF* genes led us to suggest that plastid promoters evolve actively. The study of the promoter zone of *psbD* genes [10] favors the opposite conclusion. There are possibly genes with highly conserved promoters and genes, whose promoters evolve rapidly in the plastome. The data obtained indicate that the a priori assumption of high conservativeness of plastid promoters is not quite justified and that we are still unaware to what extent the plastid system of transcription regulation is conservative (or variable). Can we hope that molecular mechanisms of plastid transcription regulation revealed in a model plant (e.g. *A. thaliana*) are proper to other organisms? The results of our investigation demonstrate that the single Sig4-dependent promoter of *ndhF* found in *A. thaliana* [17] is typical for members of order Brassicales only, and in representatives of other taxa it differs substantially or is absent (Fig. 2). However, other mechanisms may be more conservative; therefore, a

comprehensive evolutionary analysis of regulatory cis-elements in plastid DNA is quite an important task.

ACKNOWLEDGMENTS

We thank M.S. Gelfand and A.G. Vitreshchak for helpful discussion.

This study was supported by the grant from International Science and Technology Center (no. 3807) and by the grant from the President of Russian Federation for Leading Scientific Schools (NSh-915.2008.4).

REFERENCES

1. Allen, J.F., The Function of Genomes in Bioenergetic Organelles, *Phil. Trans. R. Soc. Lond., B. Biol. Sci.*, 2003, vol. 358, pp. 19–37.
2. Turmel, M., Otis, C., and Lemieux, C., The Chloroplast and Mitochondrial Genome Sequences of the Charophyte *Chaetosphaeridium globosum*: Insights into the Timing of the Events That Restructured Organelle DNAs within the Green Algal Lineage That Led to Land Plants, *Proc. Natl. Acad. Sci. USA*, 2002, vol. 99, pp. 11275–11280.
3. Palmer, J.D., Comparison of Chloroplast and Mitochondrial Genome Evolution in Plants, *Cell Organelles*, Herrmann, R.G., Ed., Wien: Springer-Verlag, 1992, pp. 99–136.
4. Sorhannus, U. and Fox, M., Synonymous and Nonsynonymous Substitution Rates in Diatoms: A Comparison between Chloroplast and Nuclear Genes, *J. Mol. Evol.*, 1999, vol. 48, pp. 209–212.
5. Seliverstov, A. and Lyubetsky, V., Translation Regulation of Intron Containing Genes in Chloroplasts, *J. Bioinform. Comput. Biol.*, 2006, vol. 4, pp. 783–790.
6. Berg, S., Krupinska, K., and Krause, K., Plastids of Three *Cuscuta* Species Differing in Plastid Coding Capacity Have a Common Parasite-Specific RNA Composition, *Planta*, 2003, vol. 218, pp. 135–142.
7. Hess, W.R. and Borner, T., Organellar RNA Polymerases of Higher Plants, *Int. Rev. Cytol.*, 1999, vol. 190, pp. 1–59.
8. Lysenko, E.A. and Kusnetsov, V.V., Plastid RNA Polymerases, *Mol. Biol.* (Moscow), 2005, vol. 39, pp. 661–674.
9. Kung, S.D. and Lin, C.M., Chloroplast Promoters from Higher Plants, *Nucleic Acids Res.*, 1985, vol. 13, pp. 7543–7549.
10. Hoffer, P.H. and Christopher, D.A., Structure and Blue-Light-Responsive Transcription of a Chloroplast psbD Promoter from *Arabidopsis thaliana, Plant Physiol.*, 1997, vol. 115, pp. 213–222.
11. Lerbs-Mache, S., Regulation of rDNA Transcription in Plastids of Higher Plants, *Biochimie*, 2000, vol. 82, pp. 525–535.
12. Swiatecka-Hagenbruch, M., Liere, K., and Börner, T., High Diversity of Plastidial Promoters in *Arabidopsis thaliana, Mol. Genet. Genom.*, 2007, vol. 277, pp. 725–734.
13. Peltier, G. and Cournac, L., Chlororespiration, *Annu. Rev. Plant Biol.*, 2002, vol. 53, pp. 523–550.
14. Wang, P., Duan, W., Takabayashi, A., Endo, T., Shikanai, T., Ye, J.Y., and Mi, H., Chloroplastic NAD(P)H Dehydrogenase in Tobacco Leaves Functions in Alleviation of Oxidative Damage Caused by Temperature Stress, *Plant Physiol.*, 2006, vol. 141, pp. 465–474.
15. Wakasugi, T., Tsudzuki, J., Ito, S., Nakashima, K., Tsudzuki, T., and Sugiura, M., Loss of All *ndh* Genes as Determined by Sequencing the Entire Chloroplast Genome of the Black Pine *Pinus thunbergii, Proc. Natl. Acad. Sci. USA*, 1994, vol. 91, pp. 9794–9798.
16. Kim, K.-J. and Jansen, R.K., NdhF Sequence Evolution and the Major Classes in the Sunflower Family, *Proc. Natl. Acad. Sci. USA*, 1995, vol. 92, pp. 10379–10383.
17. Favory, J.-J., Kobayshi, M., Tanaka, K., Peltier, G., Kreis, M., Valay, J.-G., and Lerbs-Mache, S., Specific Function of a Plastid Sigma Factor for *ndhF* Gene Transcription, *Nucleic Acids Res.*, 2005, vol. 33, pp. 5991–5999.
18. Greiner, S., Wang, X., Rauwolf, U., Silber, M.V., Mayer, K., Meurer, J., Haberer, G., and Herrmann, R.G., The Complete Nucleotide Sequences of the Five Genetically Distinct Plastid Genomes of Oenothera, Subsection Oenotera: I. Sequence Evaluation and Plastome Evolution, *Nucleic Acids Res.*, 2008, vol. 36, pp. 2366–2378.
19. Yamane, K., Yano, K., and Kawahara, T., Pattern and Rate of Indel Evolution Inferred from Whole Chloroplast Intergenic Regions in Sugarcane, Maize and Rice, *DNA Res.*, 2006, vol. 13, pp. 197–204.
20. Kanno, A. and Hirai, A., A Transcription Map of the Chloroplast Genome from Rice (*Oryza sativa*), *Curr. Genet.*, 1993, vol. 23, pp. 166–174.
21. Zubo, Ya.O., Lysenko, E.A., Aleinikova, A.Yu., Kusnetsov, V.V., and Pshibytko, N.L., Changes in the Transcriptional Activity of Barley Plastome Genes under Heat Shock, *Russ. J. Plant Physiol.,* 2008, vol. 55, pp. 293–300.
22. Kim, M., Thum, K.E., Morishige, D.T., and Mullet, J.E., Detailed Architecture of the Barley Chloroplast psbD-psbC Blue Light-Responsive Promoter, *J. Biochem.*, 1999, vol. 274, pp. 4684–4692.