

## БЫСТРАЯ ЭВОЛЮЦИЯ ПРОМОТОРОВ ПЛАСТОМНЫХ ГЕНОВ *ndhF* У ЦВЕТКОВЫХ РАСТЕНИЙ

© 2009 г. А. В. Селиверстов\*<sup>1</sup>, Е. А. Лысенко\*<sup>1,2</sup>, В. А. Любецкий\*

\* Институт проблем передачи информации Российской академии наук, Москва

\*\* Институт физиологии растений им. К.А. Тимирязева Российской академии наук, Москва

Поступила в редакцию 13.05.2008 г.

Пластом считается очень консервативной частью генома растений. Однако мало что известно об эволюции пластомных промоторов. Ранее было показано, что один промотор (СРП-*psbD*) очень консервативен у различных цветковых растений и у черной сосны. Мы провели поиск, показавший, что ген *ndhF* находится в регионе, который у наземных растений редко подвергался крупным перестройкам. Однако результаты анализа последовательностей, предшествующих кодирующей области гена *ndhF*, указывают на относительно быструю эволюцию промоторов этого пластомного гена у цветковых растений. Вероятно, предок двух основных групп Magnoliophyta (magnoliids и eudicotyledons) имел промотор P<sub>A</sub>-*ndhF*, но в некоторых филогенетических ветвях двудольных растений он был замещен другими промоторами – P<sub>B</sub>-*ndhF* и P<sub>C</sub>-*ndhF*. Мы не смогли обнаружить консервативные последовательности, содержащие потенциальные промоторы –10/–35 типа у представителей однодольных растений, в том числе у 9 представителей семейства злаков (Poaceae). Множественные выравнивания последовательностей родственных организмов, позволили выявить, что предсказанные промоторы *ndhF* (A–C) часто подвергаются мутациям, причем не только заменам нуклеотидов, но и небольшим инсерциям и делециям. На основании полученных результатов можно высказать предположение, что, по меньшей мере, некоторые пластомные промоторы быстро эволюционируют.

Ключевые слова: цветковые растения – пластом – промоторы – эволюция – *ndhF*

### ВВЕДЕНИЕ

Пластом – это часть генома растений, которая находится в пластидах, и, по-видимому, унаследована от цианобактерий. Набор пластомных генов, их расположение вдоль пластидной ДНК и даже расположение интронов очень консервативно в различных филогенетических группах наземных растений [1, 2]. Темп синонимичных и несинонимичных замен нуклеотидов в пластидных генах ниже, чем в ядерных генах [3, 4]. Некоторые не кодирующие последовательности также очень консервативны в пластомах водорослей и высших растений [5]. Поэтому пластом считается очень консервативной частью генома растений. Однако в некоторых случаях пластом может изменяться

очень быстро. Например, у различных представителей рода *Cuscuta*, включающего как фотосинтезирующие, так и практически бесхлорофильные виды, пластомы очень сильно отличаются по размеру и количеству генов [6]. Таким образом, можно полагать, что пластомные гены преимущественно консервативны, но мало что известно об эволюции их промоторов.

Промоторы регулируют первый этап экспрессии генов – инициацию транскрипции. Пластиды высших растений обладают двумя различными транскрипционными комплексами: мультисубъединичной РНК-полимеразой цианобактериального происхождения (PEP от *plastid encoded RNA polymerase*) и моносубъединичной РНК-полимеразой, гомологичной соответствующим ферментам бактериофагов и митохондрий (NEP от *nucleus encoded RNA polymerase*) [7, 8]. Сравнительный анализ промоторов NEP был представлен в работе [7], тогда как системное исследование промоторов PEP было произведено более 20 лет назад. В ранней компиляции было продемонстрировано, что пластидные промоторы очень сходны с бактериальными: консенсусные последовательности пластидных –10 (TAtaaT) и –35 (TTGaca) боксов соответствуют промоторам  $\sigma^{70}$ -типа [9]. Эволюционный анализ был проведен только для одного

<sup>1</sup> Авторы внесли равный вклад в работу.

Сокращения: СРП-*psbD* – светорегулируемый промотор гена *psbD*; индель – инсерция или делеция; ИП – инвертированный повтор; МУО – малая уникальная область (пластома); птДНК – пластидная ДНК; NEP – пластидная РНК-полимераза ядерного кодирования (от *nucleus-encoded polymerase*); PEP – РНК-полимераза пластидного кодирования (от *plastid-encoded polymerase*).

Адрес для корреспонденции: Лысенко Евгений Анатольевич. 127276 Москва, Ботаническая ул., 35. Институт физиологии растений РАН. Факс: 007 (495) 977-80-18; электронная почта: genlysenko@mail.ru

РЕР-зависимого промотора, и этот анализ показал, что светорегулируемый промотор гена *psbD* (СРП-*psbD*) очень консервативен у различных видов цветковых растений и у черной сосны [10]. Таким образом, имеющиеся немногочисленные данные указывают на эволюционную консервативность промоторов РЕР, что позволяет экстраполировать результаты, полученные на модельных объектах, на другие виды растений. Однако в последние годы появляются экспериментальные данные, демонстрирующие несовпадение сайтов инициации транскрипции (а, следовательно, промоторов) одних и тех же пластомных генов у разных видов растений [11, 12]. Поэтому для того чтобы понять, насколько сходна регуляция транскрипции генов пластома у различных растений, требуется проведение обширного эволюционного анализа.

НАДФ · Н пластохиноноксидоредуктаза является крупным белковым комплексом тилакоидных мембран хлоропластов, который, по-видимому, участвует в хлоропластном дыхании [13], подавляет генерацию АФК и, тем самым, защищает фотосинтетический аппарат [14]. Гены *ndhA-K*, кодирующие субъединицы этого комплекса, у многих растений находятся в пластидах; однако известно, что в пластомах некоторых видов эти гены отсутствуют [15]. Проведенный нами анализ пластома, для которых определена полная нуклеотидная последовательность, показал, что эти гены характерны для подавляющего большинства видов группы Streptophyta, объединяющей высшие растения и небольшую группу зеленых водорослей, и лишь в отдельных филогенетических группах пластома утратили гены *ndh*. Ген *ndhF* кодирует самую большую субъединицу комплекса и является излюбленным объектом в молекулярной систематике. Темп нуклеотидных замен в *ndhF* в два раза выше, чем в пластомном гене *rbcL*, кодирующем большую субъединицу РБФК/О [16]. Следовательно, *ndhF* является одним из активно эволюционирующих генов пластома. У *Arabidopsis thaliana* транскрипция *ndhF* регулируется промотором –10/–35 типа, и инициация транскрипции с этого промотора обеспечивается транскрипционным фактором Sig4 [17]. Анализ геномного окружения показал, что *ndhF* находится в консервативной области, которая у представителей группы Streptophytina (объединяющей высшие растения и харовые водоросли) почти не подвергалась крупным перестройкам. Все эти данные указывают на то, что ген *ndhF* является подходящим кандидатом для изучения эволюции его промоторов.

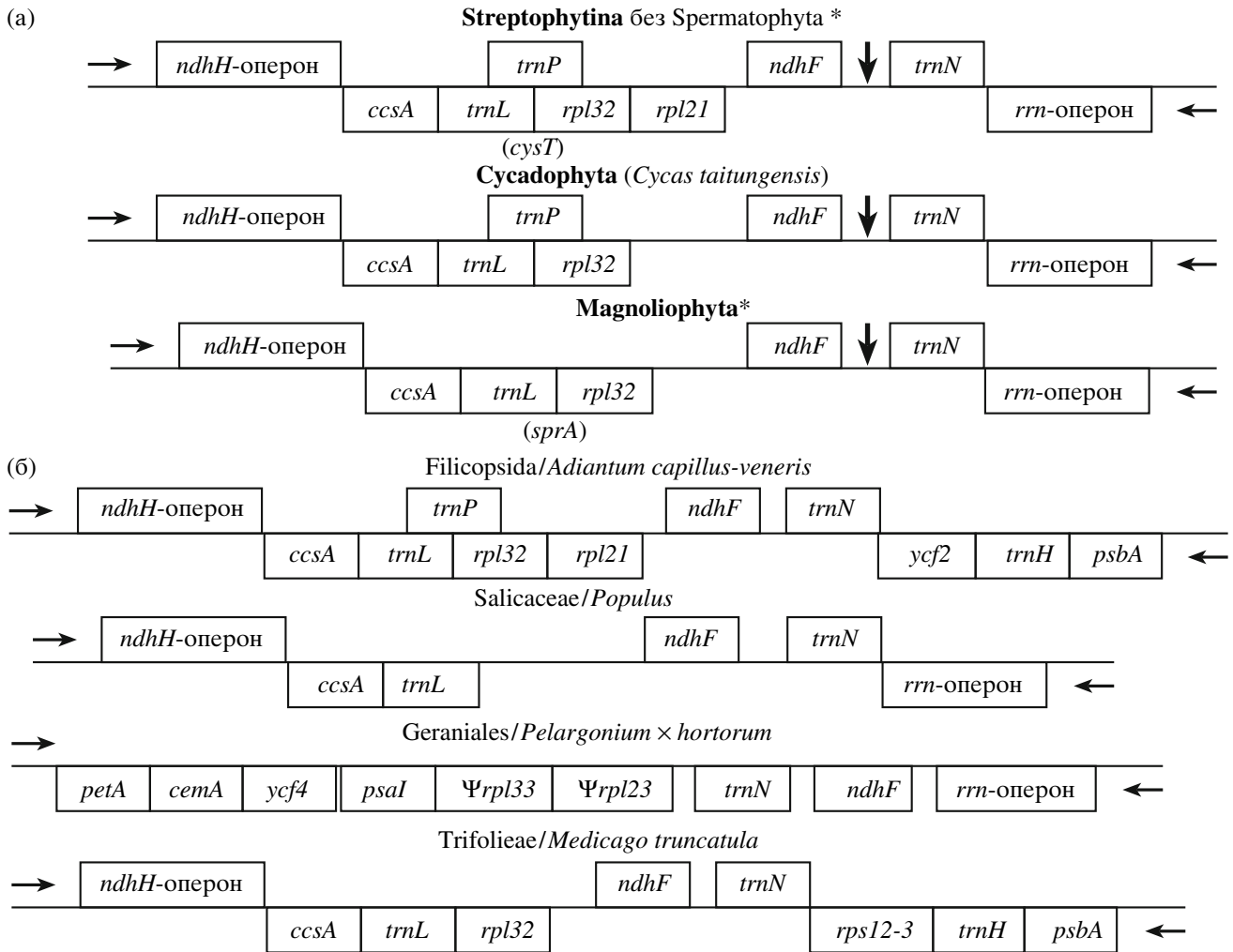
Чтобы проследить эволюцию промоторов *ndhF*, мы исследовали последовательности, расположенные с 5'-конца от кодирующей части гена. У представителей различных групп двудольных растений и magnoliids мы обнаружили три по-

тенциальных промотора, консервативных у представителей более чем одного семейства. У многих исследованных видов, в том числе у всех представителей однодольных растений, мы не смогли обнаружить существование консервативных промоторов –10/–35 типа. Выравнивания последовательностей родственных организмов позволило выявить, что предсказанные промоторы *ndhF* (А–С) часто подвергаются мутациям, причем не только заменам нуклеотидов, но и небольшим инсерциям и делециям. Таким образом, результаты проведенного анализа указывают на то, что промоторы гена *ndhF* активно эволюционировали, по крайней мере, у цветковых растений.

## МЕТОДИКА

Для анализа использовали последовательности полностью секвенированных пластома, представленные на сайте NCBI (<http://www.ncbi.nlm.nih.gov/genomes/ORGANELLES/plastids.html>). Поиск потенциальных промоторов производили для всех представителей группы Streptophytina, у которых имеется ген *ndhF* (84 вида на момент завершения работы). Поиск был ограничен зоной между кодирующими областями гена *ndhF* и ближайшего гена, расположенного выше (upstream): *rpl32* у цветковых растений и *rpl21* у представителей несеменистых растений и харовых водорослей (рис. 1). Размер этой области варьировал у разных видов от 300 до 1000 п. н. и более.

Для предсказания промоторов использовали следующую процедуру. Сначала проводили поиск последовательностей, имеющих наибольшее сходство с минимальным промотором  $\sigma^{70}$ -типа: TTGaca–17–18 п.н. – TAtaaT (более консервативные нуклеотиды представлены прописными буквами, менее консервативные – строчными). За отклонение от консенсусной последовательности начислялся штраф: “большой” – в случае отклонения по более консервативным позициям и в случае сокращения спейсера (т.е. интервала между –10 и –35 боксами промотора) до 16 п. н.; “малый” – за отклонения по менее консервативным позициям и за увеличение спейсера до 19 п.н. “Малый” штраф был в три раза меньше “большого”. Кроме того, назначался “приз” за АТ-богатый спейсер. Верхняя граница суммы штрафов и призов выбиралась так, чтобы в каждой лидерной области число кандидатов, отбираемых на первом этапе, было небольшим; в приводимых далее результатах это число не превышало пяти. Для *A. thaliana* использовали последовательность –10 и –35 боксов перед экспериментально установленным сайтом инициации транскрипции [17]. После этого производили множественное выравнивание лидерных областей. Выравнивание осуществляли при помощи программы MultAl (А.А. Миронов, не опубликовано) и разработан-



**Рис. 1.** Схема геномного окружения гена *ndhF* у представителей группы Streptophytina.

Горизонтальные стрелки указывают направление транскрипции генов, расположенных на разных нитях птДНК. Вертикальная стрелка показывает границу МУО (слева) и ИП<sub>Б</sub> (справа). У некоторых видов при расширении инвертированных повторов в это место попадают гены (от одного до трех) с противоположного конца МУО (пояснения в тексте). Схема отражает порядок расположения генов и не отражает относительные размеры генов и межгенных участков. Ген *trnP* не перекрывается с генами *trnL* и *rpl32*. а – схема порядка генов в группах. В известных полных пластомах Coniferophyta (*Cryptomeria*, *Pinus*) наблюдаются различные перестройки этого региона, поэтому для этой группы схема не приведена; б – исключения; показаны наибольший/наименьший таксон, у представителей которого можно ожидать наличие перестройки. Например, перестройка встречается у обоих представителей рода *Populus*, поэтому, вероятно, она встречается у всех представителей этого рода; перестройка не встречается у представителя группы Malpighiales (*M. esculenta*), к которой относится сем. Salicaceae и род *Populus*; поэтому наибольший таксон, для которого можно предполагать наличие такой перестройки, это Salicaceae. \* У представителей лишайников (Anthocerotophyta и Marchantiophyta) между генами *rpl32* и *trnP* присутствует ген *cysT*. У некоторых представителей семейства пасленовых (Solanaceae) между генами *rpl32* и *trnL* присутствует ген *sprA*.

ного нами метода [5] и корректировали “вручную” при необходимости. Сначала выравнивали участки, относящиеся к небольшой таксономической группе, которая затем расширялась насколько возможно. В том случае, если последовательность оказывалась консервативной у представителей более чем одного семейства, такая последовательность рассматривалась как потенциальный промотор РЕР. Обычно потенциальные промоторы одной группы находились у раз-

ных видов на сходном расстоянии от кодирующей области *ndhF* (рис. 2–4), и консервативность нуклеотидов в зоне потенциального промотора была выше, чем в соседних участках. Это также рассматривали как довод в пользу предсказания. Для проверки разработанной процедуры использовали последовательности, предшествующие генам *psbA*, *psbB*, *psbE*, *rbcL*, *psaA*. Проверка показала, что избранная процедура позволяет предсказывать промоторы, совпадающие с эксперимен-

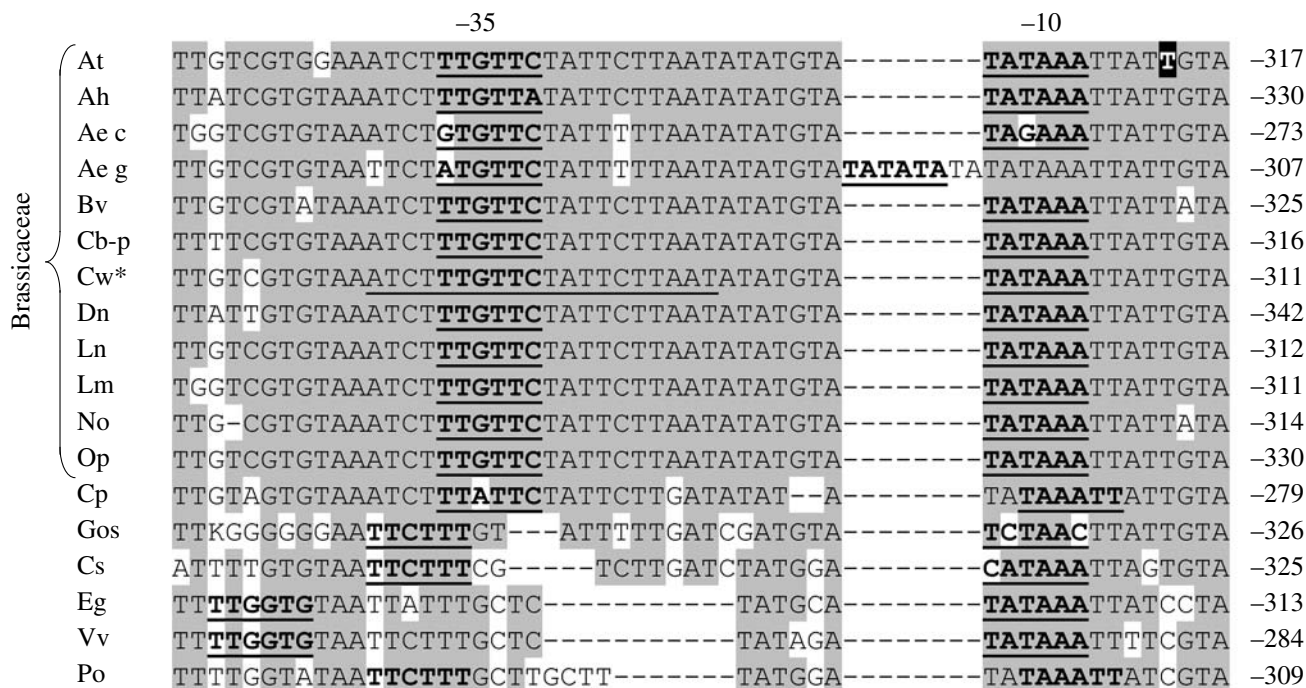


Рис. 2. Выравнивание последовательностей, предшествующих кодирующим областям гена *ndhF*, у видов сем. Brassicaceae и представителей родственных им групп. Потенциальный промотор С ( $P_C$ -*ndhF*).

Потенциальные -35 и -10 боксы даны жирным шрифтом и подчеркнуты. Консервативные нуклеотиды выделены серым фоном. Сайт инициации транскрипции у *A. thaliana* [17] показан белой буквой на черном фоне. 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*. Остальные виды приведены в порядке убывания их близости с крестоцветными: Cp – *Carica papaya* (Brassicales), Cs – *Citrus sinensis* (eurosids II), Gos – *Gossypium* sp. (*G. barbadense* и *G. hirsutum*) (eurosids II), Eg – *Eucalyptus globulus* (rosids), Vv – *Vitis vinifera* (core eudicotyledons), Po – *Platanus occidentalis* (eudicotyledons). К означает G (у *G. hirsutum*) или T (у *G. barbadense*). \* У *C. wallichii* имеется вставка 20 п.н., которая опущена для упрощения рисунка; вставка идентична подчеркнутому фрагменту и размещается непосредственно перед ним, так что они представляют тандемный повтор без интервала.

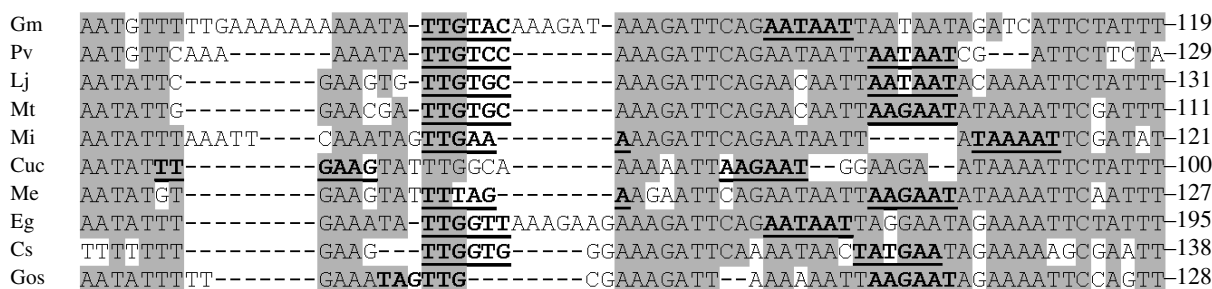


Рис. 3. Выравнивание последовательностей, предшествующих кодирующим областям гена *ndhF*, у представителей таксона rosids. Потенциальный промотор В ( $P_B$ -*ndhF*).

Потенциальные -35 и -10 боксы даны жирным шрифтом и подчеркнуты. Консервативные нуклеотиды выделены серым фоном. Fabaceae/Papilionoideae: Mt – *Medicago truncatula*, Gm – *Glycine max*, Lj – *Lotus japonicus*, Pv – *Phaseolus vulgaris*; представители других групп таксона eurosids I: Mi – *Morus indica*, Cuc – *Cucumis sativus*, Me – *Manihot esculenta*; представители других групп таксона rosids: Cs – *Citrus sinensis* (eurosids II), Gos – *Gossypium* sp. (eurosids II), Eg – *Eucalyptus globulus* (Myrtales).

		-35											-10														
magnoliids	Dg	GT	<b>TTGACA</b>	CAAT	----	T	T	A	A	C	T	T	T	A	<b>TTG</b>	----	<b>AAT</b>	T	T	T	T	C	A	-172			
	Lt	TA	<b>GTGAAC</b>	CAAT	----	G	T	A	A	C	T	T	C	T	<b>TTG</b>	----	<b>AAT</b>	T	T	T	T	A	A	T	-167		
stem eudicots	Nd	AT	<b>TTGACCA</b>	CAAT	----	A	G	T	A	A	C	T	T	C	<b>TTGATT</b>	----	T	A	A	T	A	T	T	T	G	T	-178
	Rm	CT	<b>TTAATC</b>	TAC	----	T	T	A	A	T	T	A	T	T	T	<b>TTGATT</b>	----	-----	T	T	G	A	A	-----	-162		
	Po	AT	<b>TTTACCA</b>	CAAT	----	T	G	T	C	A	C	T	T	C	<b>TTG</b>	----	<b>AAT</b>	A	T	T	T	G	A	A	-150		
core eudicotyledons	So	AT	<b>TTGACCA</b>	CAAT	----	T	A	T	A	A	C	T	T	C	<b>TTG</b>	----	<b>AAT</b>	A	T	A	A	A	A	A	-183		
	Vv	AT	<b>TTTCC</b>	CAAT	----	T	G	T	A	A	C	T	T	C	<b>TTG</b>	----	<b>AAT</b>	A	T	T	T	G	A	A	-154		
	Ha	AT	<b>TTGACCA</b>	CAAC	----	T	G	T	A	A	C	T	T	C	<b>TTG</b>	----	<b>AAT</b>	A	T	T	T	G	A	A	-144		
	Ls	AT	<b>TTGACCA</b>	CAAT	----	T	G	T	A	A	C	T	T	C	<b>TTG</b>	----	<b>AAT</b>	A	T	T	T	G	A	T	-159		
	Dc	AT	<b>TTGCCCA</b>	CAAT	----	T	G	T	A	A	C	T	T	C	<b>TTT</b>	----	<b>AAT</b>	A	T	T	T	A	A	A	-157		
	Pg	AT	<b>TTGACCA</b>	CAAT	T	A	A	T	C	T	A	A	C	T	<b>TTGATT</b>	T	G	-----	A	A	T	T	T	G	A	-143	
	Ca	AT	<b>TTGAACA</b>	CAAT	----	T	A	T	A	A	C	T	T	C	<b>TTG</b>	----	<b>AAT</b>	A	T	T	T	T	A	A	-179		
	Jn	AT	<b>TTGACCA</b>	CAAT	----	T	G	T	A	A	C	T	T	C	<b>TTG</b>	----	<b>AAT</b>	A	T	T	T	G	A	A	-139		
	Ip	AT	<b>TTGACCA</b>	CAAT	----	T	G	T	A	A	C	T	T	C	<b>TTG</b>	----	<b>AAT</b>	A	A	A	T	T	G	A	A	-169	
	Nic	AT	<b>TTGACCA</b>	CAAT	----	T	G	T	A	A	C	T	T	C	<b>TTT</b>	----	<b>AAT</b>	A	G	T	T	G	A	A	-159		
	Ab	AT	<b>TTGACCA</b>	CAAT	----	T	G	T	A	A	C	T	T	C	<b>TTT</b>	----	<b>AAT</b>	A	G	T	T	G	A	A	-159		
	Sol	AT	<b>TTGACCA</b>	CAAT	----	C	G	T	A	A	C	T	T	C	<b>TTG</b>	----	<b>AAT</b>	A	G	T	T	G	A	A	-159		

**Рис. 4.** Выравнивание последовательностей, предшествующих кодирующим областям гена *ndhF*, у представителей eu-dicotyledons и magnoliids. Потенциальный промотор А ( $P_A$ -*ndhF*).

Потенциальные -35 и -10 боксы даны жирным шрифтом и подчеркнуты. Консервативные нуклеотиды выделены серым фоном. 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*).

тально найденными, и не приводит к “перепредсказанию” промоторов.

Следует пояснить, почему в качестве критерия была избрана консервативность у представителей более, чем одного семейства. Сходство последовательностей в пределах одного семейства может быть результатом высокой консервативности пластидной ДНК и не отражать функциональной значимости последовательности. Например, у всех 12 представителей сем. Brassicaceae мы обнаружили последовательность, сходную с последовательностью минимального промотора -10/-35 типа: TTGACA-17 п. н. -TtTAAT (строчной буквой показан нуклеотид, отличающийся от консенсуса). У *A. thaliana* эта последовательность находится на расстоянии 198 п. н. от кодирующей области *ndhF*; однако инициация транскрипции в этом районе не была обнаружена [17]. У *Carica papaya*, который является представителем одного из ближайших семейств (Brassicales/Caricaceae), гомологичный участок существенно отличается от консенсуса: TТactA-17 п. н. -TtTAAc (-177 п. н.). Более вероятно, что данный участок не является промотором, несмотря на соответствие последовательности нуклеотидов консенсусу и их высокую консервативность внутри семейства крестоцветных, тогда как промотор перед сайтом инициации транскрипции у *A. thaliana* менее соответствует консенсусной последовательности [17], но консервативен у крестоцветных и у *C. papaya* (рис. 2). Поэтому

в качестве потенциальных промоторов были предсказаны только последовательности, для которых удалось обнаружить консервативность у представителей более, чем одного семейства.

В работе использована систематика растений и водорослей, представленная на сайте NCBI ([www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html](http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html)).

## РЕЗУЛЬТАТЫ

### Геномный анализ генов *ndh*

Аналізу лидерных последовательностей области гена *ndhF* предшествовал анализ геномной организации этого гена и *ndh* генов вообще. Проведенный анализ показал, что в пластомах организмов, представляющих таксоны Alveolata, Cercozoa, Cryptophyta, Euglenozoa, Glaucocystophyceae, Haptophyceae, Rhodophyta и stramenopiles гены *ndh* отсутствуют. Всего было изучено 18 пластов, для которых определена полная нуклеотидная последовательность, в том числе 5 пластов красных водорослей. В группе Chlorophyta, объединяющей большинство зеленых водорослей, в пластомах 8 из 9 видов гены *ndh* отсутствуют, но у *Nephroselmis olivacea* присутствуют почти все гены *ndh*, за исключением гена *ndhJ*. Напротив, в группе Streptophyta, объединяющей небольшое число видов зеленых водорослей и наземные растения, гены *ndh* присутствуют у большинства представителей. Из 98 видов, пластома которых

представлены в соответствующем разделе базы NCBI, гены *ndh* отсутствуют (или представлены псевдогенами) у нефотосинтезирующих паразитических растений родов *Epifagus* и *Cuscuta*. Среди фотосинтезирующих организмов эти гены отсутствуют в пластомах лишайника (Marchantiophyta) *Aneura mirabilis*, у представителей одного из семейств хвойных (Pinaceae) – *Pinus koraiensis* и *P. thunbergii*, и у представителя одного из порядков однодольных растений (Asparagales) – *Phalaenopsis aphrodite*. При этом у представителя Marchantiophyta (*Marchantia polymorpha*), у представителя родственного Pinaceae сем. Cupressaceae (*Cryptomeria japonica*, AP009377), а также у представителей других порядков однодольных растений в пластомах присутствуют все 11 генов *ndh*.

Ген *ndhF* располагается в малой уникальной области (МУО) пластома рядом с инвертированным повтором Б (ИП<sub>Б</sub>). У некоторых видов может происходить расширение инвертированных повторов. Если ИП<sub>Б</sub> захватывает ген *ndhF*, то возникает его полная или частичная копия в ИП<sub>А</sub> на границе с МУО (например, у видов рода *Oenothera*). Если в сторону МУО расширяется ИП<sub>А</sub>, то между *ndhF* и *trnN* (рис. 1) появляются копии генов с противоположной стороны МУО: *ycf1*, *rps15*, *ndhH* или *ndhA* (*chlL*, *chlN* или *ycf20* у харовых водорослей). Расширение ИП не оказывает влияния на промоторную область *ndhF*. У многих видов копии генов, возникшие в результате расширения ИП, являются псевдогенами (как новая копия *ndhF* у *Oenothera*).

Геномное окружение гена *ndhF* отличается у представителей различных филогенетических ветвей группы Streptophyta. Однако внутри группы Streptophytina, объединяющей харовые водоросли и наземные растения, этот регион довольно консервативен и, по-видимому, редко подвергался крупным перестройкам (рис. 1). Одной из таких перестроек была утрата гена *rpl21*, которая, скорее всего, произошла у общего предка семенных растений, так как этот ген отсутствует у представителей цветковых (Magnoliophyta) и двух групп голосеменных растений: Сусадophyta (*Cycas taitungensis*) и Coniferophyta (*C. japonica*, *P. koraiensis* и *P. thunbergii*). У некоторых видов наблюдаются дополнительные перестройки вблизи гена *ndhF* (рис. 1), однако эти перестройки, по-видимому, характерны для небольших таксономических групп. Так, перестройка у *Medicago truncatula* может быть характерна только для его трибы Trifolieae, так как у представителей других триб того же подсемейства Papilionoideae (*Glycine max*, *Lotus japonicus*, *Phaseolus vulgaris*) организация этого участка такая же, как у большинства цветковых растений. Из этих перестроек (рис. 1) только утрата гена *rpl32* у представителей рода *Populus* (*P. alba* и *P. trichocarpa*) и инверсия гена *ndhF* у *Pelargonium* × *hortorum* затрагивают область вы-

ше (upstream) гена *ndhF*, и могли повлиять на промоторную зону.

Таким образом, ген *ndhF* находится в довольно консервативном участке пластома. В ходе эволюции крупные перестройки, которые могли затронуть промоторную область, вероятно, происходили у общего предка семенных растений (утрата гена *rpl21*), а также у предков небольших групп двудольных растений: рода *Populus* (утрата гена *rpl32*) и *P. hortorum* (инверсия гена *ndhF*) (рис. 1).

#### Поиск потенциальных промоторов

У *A. thaliana* сайт инициации транскрипции гена *ndhF* установлен экспериментальным путем [17]. Анализ последовательностей показал, что промотор –10/–35 типа, характерный для гена *ndhF* *A. thaliana*, консервативен у представителей сем. Brassicaceae, у представителя родственного сем. Caricaceae (*C. papaya*), у представителей других семейств таксона eurosids II (*Citrus sinensis*, *Gossypium barbadense*, *G. hirsutum*), а также у отдельных представителей более отдаленных групп (*Eucalyptus globulus*, *Vitis vinifera*, *Platanus occidentalis*) (рис. 2). Мы обозначили эту группу потенциальных промоторов буквой С (P<sub>С</sub>-*ndhF*). Выравнивание последовательностей, представленное на рис. 2, свидетельствует о том, что в ходе эволюции этот промотор подвергался существенным изменениям, в том числе и у представителей семейства крестоцветных. Кроме мононуклеотидных мутаций, вероятно, в области промотора происходили небольшие делеции и инсерции, которые изменяли расстояние между боксами и должны были приводить к “смещению” одного из боксов. Выравнивание последовательностей наименее родственных видов (*E. globulus*, *V. vinifera*, *P. occidentalis*) позволяет предполагать, что они представляют более древний вариант промотора, и что промотор, выявленный у *A. thaliana*, возник в результате вставки примерно 11 п. н. у общего предка порядка Brassicales или группы eurosids II. У остальных видов нам не удалось обнаружить промоторы, гомологичные P<sub>С</sub>-*ndhF*; поэтому мы продолжили поиск других потенциальных промоторов гена *ndhF*.

Еще один потенциальный промотор был выявлен у представителей таксона eurosids I и у некоторых других представителей подкласса rosids (рис. 3); эта группа промоторов была обозначена буквой В (P<sub>В</sub>-*ndhF*). Этот предполагаемый промотор является еще менее консервативным, чем P<sub>С</sub>-*ndhF*, даже у представителей одного подсемейства (Papilionoideae) бобовых растений. В зоне промоторов P<sub>В</sub>-*ndhF* также наблюдаются вставки и делеции коротких фрагментов, которые должны приводить к “смещению” боксов. Общей особенностью промоторов этой группы является неконсервативность 3'-конца –35-боксов (менее консер-

Lm	ТААСГСАА <b>TTGGA</b> СТТТТТГААТТАТТТТАА <b>ААТААТ</b> GGATGC –158
De	ТАТТАГАТ <b>TTGAAG</b> ССААТТААТААААТТ <b>С</b> АТААТGAGTCA –385
Aco	ТААТАААТ <b>TTGATA</b> АТААСТТААСТССААА <b>ТАААСТ</b> ААСТAG –266
Pop*	СGAGС <b>TTGAAG</b> GAGТТААТТ <b>С</b> ААТГТАА <b>ТАТТТТ</b> ТТТАG –493/–552
Gos	ТААТАА <b>TTGATA</b> ТТТТ <b>С</b> GАААААТТАТТ <b>А</b> С <b>ТАТААТ</b> GАААААААGААТТ –246
At**	ААТСТТ <b>TTGACA</b> ГТААСТТАGТААТТАТТТ <b>ТТТААТ</b> САТТТТСТААТТТТ –185

Рис. 5. Последовательности –10/–35 типа, обнаруженные у отдельных растений.

Потенциальные –35 и –10 боксы даны жирным шрифтом и подчеркнуты. 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*.

\* Расстояние от кодирующей зоны: –493(*P. trichocarpa*)/–552 (*P. alba*). \*\* Последовательность консервативна у всех крестоцветных (пояснение в разделе МЕТОДИКА).

вативного и в консенсусном промоторе) и замена первого нуклеотида –10 бокса на А (вместо консенсусного Т). Кроме того, у трех видов наблюдается сокращение расстояния между боксами до 16 п. н. (рис. 2). У видов рода *Gossypium* только два нуклеотида соответствуют консенсусу. Вне подкласса rosids не удалось обнаружить последовательности гомологичные  $P_B\text{-}ndhF$ .

Потенциальный промотор А ( $P_A\text{-}ndhF$ ) был выявлен у представителей различных групп таксона core eudicotyledons (asterids, Caryophyllales, Vitales), а также у представителей stem eudicotyledons и даже у некоторых представителей отдельной ветви цветковых растений – magnoliids (*Drimys granadensis*, *Liriodendron tulipifera*) (рис. 4). Участки промоторов  $P_A\text{-}ndhF$  значительно более консервативны; однако и в этой группе наблюдаются замены нуклеотидов в боксах промотора, а также инсерция, которая должна приводить к “смещению” одного из боксов (у *Panax ginseng*).

У однодольных растений нам не удалось выявить консервативный промотор  $\sigma^{70}$ -типа. Удастся построить хорошее выравнивание большинства последовательностей злаков, как между собой, так и с последовательностями двудольных растений, у которых выявляется промотор  $P_A\text{-}ndhF$  (данные не приведены). На этом выравнивании видно, что у злаков нет потенциального промотора, гомологичного  $P_A\text{-}ndhF$ , и вообще отсутствует консервативный консенсусный –10/–35 промотор. Для анализа использовали последовательности 9 видов, принадлежащих к 8 родам, представляющим обе основные ветви сем. Poaceae (ВЕР и РАССАД). У представителей других порядков однодольных растений – Acorales (*Acorus americanus*, *A. calamus*), Alismatales (*Lemna minor*) и Dioscoreales (*Dioscorea elephantipes*) удастся обнаружить последовательности, обладающие хорошим сходством с консенсусным промотором  $\sigma^{70}$ -типа (рис. 5), но эти последовательности не являются гомологичными друг другу или консервативным промотор-подобным последовательностям двудольных растений.

У представителей остальных ветвей цветковых растений – basal Magnoliophyta (*Amborella trichopoda*, *Illicium oligandrum*, *Chloranthus spicatus*, *Nuphar advena*, *Nymphaea alba*) и Ceratophyllales (*Ceratophyllum demersum*), а также у голосеменного растения *C. taitungensis*, у семи видов несеменистых растений и двух харовых водорослей – нам также не удалось выявить консервативный промотор консенсусного типа.

## ОБСУЖДЕНИЕ

Мы провели анализ последовательностей, расположенных перед пластомными генами *ndhF* и обнаружили, что у цветковых растений нет промотора, который был бы общим для всех групп. У представителей двух из пяти основных ветвей цветковых растений (magnoliids и eudicotyledons) обнаружен консервативный потенциальный промотор  $P_A\text{-}ndhF$  (рис. 4); в одном из подклассов двудольных растений – rosids – этот промотор не обнаружен, но выявлены два других консервативных потенциальных промотора:  $P_B\text{-}ndhF$  (рис. 3) и  $P_C\text{-}ndhF$  (рис. 2). За пределами magnoliids и eudicotyledons нам не удалось найти консервативные последовательности, имеющие сходство с консенсусом промотора –10/–35 типа. Отсутствие единого для всех консервативного промотора не может быть объяснено частыми геномными перестройками, поскольку ген *ndhF* находится в довольно консервативном участке пластома (рис. 1).

Выявленные консервативные промотор-подобные последовательности обнаружены у большинства представителей двудольных растений (eudicotyledons). Среди представителей этой группы консервативные промоторы не обнаружены у *Buxus microphylla*, *P. hortorum*, двух видов рода *Populus* и пяти видов рода *Oenothera*. У *P. hortorum* и *Populus* наблюдаются крупные перестройки пластома, которые могли затронуть промоторную область гена *ndhF* (рис. 1), а род *Oenothera* известен высокой изменчивостью птДНК [18]. Вероятно, промотор  $P_A\text{-}ndhF$  является наиболее древним промотором двудольных или даже цвет-



ковых растений, так как он обнаружен не только у представителей eudicotyledons, но и у magnoliids (рис. 4). Предшествующий ему более древний промотор мог быть разрушен в результате делеции гена *rpl21*, которая, вероятно, произошла у общего предка семенных растений (рис. 1). И это может быть причиной того, что мы не смогли выявить последовательности, гомологичные  $P_A\text{-}ndhF$ , у несеманных растений. Промоторы  $P_B\text{-}ndhF$  и  $P_C\text{-}ndhF$ , по-видимому, возникли позже; однако промотор  $P_C\text{-}ndhF$  выявляется не только у представителей подкласса rosids, но и у представителей других групп, в том числе у представителя stem eudicotyledons – *P. occidentalis* (рис. 2). У некоторых видов обнаружены два промотора:  $P_A\text{-}ndhF$  и  $P_C\text{-}ndhF$  (*P. occidentalis*, *V. vinifera*),  $P_B\text{-}ndhF$  и  $P_C\text{-}ndhF$  (*E. globulus*, *C. sinensis*, *Gossypium* sp.) (рис. 2–4). Вероятно, новые промоторы возникали не вместо, а в дополнение к предшествующему, а потом один из них у большинства видов исчезал. То, что “новые” промоторы ( $P_B\text{-}ndhF$  и  $P_C\text{-}ndhF$ ) обнаружены у большинства представителей подкласса rosids, по-видимому, обусловлено тем, что у общего предка этой группы исчез “старый” промотор ( $P_A\text{-}ndhF$ ).

Анализ множественных выравниваний последовательностей из родственных видов позволяет сделать вывод, что все предсказанные промоторы ( $P_A$ -,  $P_B$ -,  $P_C\text{-}ndhF$ ) подвергаются изменению в процессе эволюции. В каждой группе промоторов наблюдаются мутации – замены нуклеотидов в зоне –10 и –35 боксов, инсерции и делеции в спейсере между боксами, делеции последовательностей одного из боксов (рис. 2–4), которые должны приводить к изменению свойств промоторов. Эти наблюдения согласуются с информацией о том, что небольшие инсерции или делеции (индели) являются основной причиной эволюции некодирующих последовательностей пластома, причем около 40% приходится на долю мононуклеотидных инделей, частота инделей 2, 3, 4 или 5 нуклеотидов составляет около 10% (для каждого), после 5 п. н. частота инделей постепенно снижается [19]. Однако эти расчеты основаны на анализе всех некодирующих последовательностей. Наши данные позволяют предполагать, что, по крайней мере, некоторые регуляторные *цис*-элементы пластома являются мишенями небольших инделей, которые, вероятно, не разрушают их, но изменяют свойства.

Нам удалось выявить консервативные потенциальные промоторы для таксономических групп, у многих представителей которых определена полная нуклеотидная последовательность пластома. То, что нам не удалось найти консервативные промоторы для многих групп организмов (однодольных, голосеменных, папоротникообразных, мхов, лишайников, харовых водорослей) может быть связано с тем, что в анализе эти группы были представ-

лены небольшим числом последовательностей. Возможно, внутри каждой из этих групп существуют свои консервативные промоторы PER. Однако отрицательный результат, полученный при анализе 7 последовательностей несеманных растений и двух последовательностей харовых водорослей заставляет сделать вывод, что у несеманных представителей группы Streptophytina также нет общего для всех консервативного промотора гена *ndhF*.

Анализ последовательностей 9 видов злаков, также позволяет предполагать, что внутри этого семейства нет консервативного консенсусного промотора PER. Известно, что у злаков *ndhF* транскрибируется и его транскрипция активируется при тепловом шоке [20, 21]. По-видимому, у злаков этот ген либо транскрибируется PER с промотора неконсенсусного типа, как в случае СРП-*psbD* [10, 22], либо он может транскрибироваться с промотора NEP.

Таким образом, результаты анализа промоторной зоны генов *ndhF* позволяют предполагать, что пластомные промоторы активно эволюционируют. Результат анализа промоторной зоны генов *psbD* [10] позволяет сделать противоположный вывод. Возможно, в пластомах есть гены с очень консервативными промоторами и гены, промоторы которых активно эволюционируют. Полученные нами результаты указывают на то, что *a priori* предположение о консервативности пластомных промоторов не вполне справедливо и что мы пока еще не представляем, насколько консервативна (или вариабельна) система регуляции транскрипции в пластидах. Можем ли мы ожидать, что молекулярные механизмы регуляции пластидной транскрипции, выявленные на модельном объекте (например, *A. thaliana*), будут характерны для других организмов? Например, результаты данного исследования указывают на то, что единственный обнаруженный у *A. thaliana* Sig4-зависимый промотор *ndhF* [17] характерен только для представителей порядка Brassicales, а у представителей других групп он существенно отличается или вообще отсутствует (рис. 2). Однако другие механизмы могут быть более консервативными, поэтому обширный эволюционный анализ регуляторных *цис*-элементов пластидной ДНК является весьма актуальной задачей.

*Авторы выражают благодарность М.С. Гельфанду и А.Г. Витрецаку за обсуждение работы.*

Исследование было поддержано грантом Международного научно-технического центра (№ 3807) и грантом Президента Российской Федерации для поддержки ведущих научных школ (НШ-915.2008.4).

#### СПИСОК ЛИТЕРАТУРЫ

1. Allen J.F. The Function of Genomes in Bioenergetic Organelles // Phil. Trans. R. Soc. Lond. B. Biol. Sci. 2003. V. 358. P. 19–37.



2. Turmel M., Otis C., 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. V. 99. P. 11 275–11 280.
3. Palmer J.D. Comparison of Chloroplast and Mitochondrial Genome Evolution in Plants // Cell Organelles / Ed. Herrmann R.G. Wien: Springer-Verlag, 1992. P. 99–136.
4. Sorhannus U., Fox M. Synonymous and Nonsynonymous Substitution Rates in Diatoms: A Comparison between Chloroplast and Nuclear Genes // J. Mol. Evol. 1999. V. 48. P. 209–212.
5. Seliverstov A., Lyubetsky V. Translation Regulation of Intron Containing Genes in Chloroplasts // J. Bioinform. Comput. Biol. 2006. V. 4. P. 783–790.
6. Berg S., Krupinska K., Krause K. Plastids of Three *Cuscuta* Species Differing in Plastid Coding Capacity Have a Common Parasite-Specific RNA Composition // Planta. 2003. V. 218. P. 135–142.
7. Hess W.R., Börner T. Organellar RNA Polymerases of Higher Plants // Int. Rev. Cytol. 1999. V. 190. P. 1–59.
8. Лысенко Е.А., Кузнецов В.В. РНК-полимеразы пластид // Мол. биол. 2005. Т. 39. С. 762–775.
9. Kung S.D., Lin C.M. Chloroplast Promoters from Higher Plants // Nucleic Acids Res. 1985. V. 13. P. 7543–7549.
10. Hoffer P.H., Christopher D.A. Structure and Blue-Light-Responsive Transcription of a Chloroplast psbD Promoter from *Arabidopsis thaliana* // Plant Physiol. 1997. V. 115. P. 213–222.
11. Lerbs-Mache S. Regulation of rDNA Transcription in Plastids of Higher Plants // Biochimie. 2000. V. 82. P. 525–535.
12. Swiatecka-Hagenbruch M., Liere K., Börner T. High Diversity of Plastidial Promoters in *Arabidopsis thaliana* // Mol. Genet. Genom. 2007. V. 277. P. 725–734.
13. Peltier G., Cournac L. Chlororespiration // Annu. Rev. Plant Biol. 2002. V. 53. P. 523–550.
14. Wang P., Duan W., Takabayashi A., Endo T., Shikanai T., Ye J.Y., Mi H. Chloroplastic NAD(P)H Dehydrogenase in Tobacco Leaves Functions in Alleviation of Oxidative Damage Caused by Temperature Stress // Plant Physiol. 2006. V. 141. P. 465–474.
15. Wakasugi T., Tsudzuki J., Ito S., Nakashima K., Tsudzuki T., 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. U.S.A. 1994. V. 91. P. 9794–9798.
16. Kim K.-J., Jansen R.K. *ndhF* Sequence Evolution and the Major Classes in the Sunflower Family // Proc. Natl. Acad. Sci. USA. 1995. V. 92. P. 10 379–10 383.
17. Favory J.-J., Kobayashi M., Tanaka K., Peltier G., Kreis M., Valay J.-G., Lerbs-Mache S. Specific Function of a Plastid Sigma Factor for *ndhF* Gene Transcription // Nucleic Acids Res. 2005. V. 33. P. 5991–5999.
18. Greiner S., Wang X., Rauwolf U., Silber M.V., Mayer K., Meurer J., Haberer G., Herrmann R.G. The Complete Nucleotide Sequences of the Five Genetically Distinct Plastid Genomes of *Oenothera*, Subsection *Oenothera*: I. Sequence Evaluation and Plastome Evolution // Nucleic Acids Res. 2008. V. 36. P. 2366–2378.
19. Yamane K., Yano K., Kawahara T. Pattern and Rate of Indel Evolution Inferred from Whole Chloroplast Intergenic Regions in Sugarcane, Maize and Rice // DNA Res. 2006. V. 13. P. 197–204.
20. Kanno A., Hirai A. A Transcription Map of the Chloroplast Genome from Rice (*Oryza sativa*) // Curr. Genet. 1993. V. 23. P. 166–174.
21. Зубо Я.О., Лысенко Е.А., Алейникова А.Ю., Кузнецов В.В., Пишбытко Н.Л. Изменение транскрипционной активности генов пластома ячменя в условиях теплового шока // Физиология растений. 2008. Т. 55. С. 323–331.
22. Kim M., Thum K.E., Morishige D.T., Mullet J.E. Detailed Architecture of the Barley Chloroplast psbD-psbC Blue Light-Responsive Promoter // J. Biochem. 1999. V. 274. P. 4684–4692.