

УДК 575.174.015.3

ВАРИАБЕЛЬНОСТЬ ИНТРОНА ПЛАСТИДНОГО ГЕНА *trnL* У ВИДОВ FABOIDEAE (СЕМ. FABACEAE)

© 2014 г. Е.А. Дьяченко^{1,2}, М.А. Филюшин¹, Е.П. Пронин², Е.З. Кочиева^{1,3}

¹ Центр «Биоинженерия» РАН, Москва, Россия,
e-mail: dyachenko-el@yandex.ru;

² Всероссийский научно-исследовательский институт селекции и семеноводства
овощных культур, поселок ВНИИССОК, Одинцовский район, Московская область, Россия;

³ Московский государственный университет им. М.В. Ломоносова,
кафедра биотехнологии, Москва, Россия

Поступила в редакцию 16 сентября 2014 г. Принята к публикации 21 октября 2014 г.

Инtron, расположенный между первым и вторым нуклеотидами антикодона лейциновой транспортной РНК, является единственным представителем интронов группы I у высших растений. В данной работе впервые охарактеризована последовательность интрона пластидного гена *trnL* у 16 видов бобовых и построена вероятная вторичная структура всего интрона и отдельных функциональных участков. Выявлен высокий уровень полиморфизма как между родами семейства Fabaceae, так и между видами одного рода. В последовательностях каталитического центра, которые считаются крайне консервативными, были детектированы единичные нуклеотидные замены.

Ключевые слова: растения, инtron группы I, хлоропластный геном, нуклеотидный полиморфизм, вторичная структура интрона.

ВВЕДЕНИЕ

Интроны группы I представляют собой последовательности, способные катализировать собственный сплайсинг из последовательности пре-РНК. Механизм сплайсирования включает две связанные реакции трансэтерификации в присутствии гуанозинового кофактора на первом этапе с последующими конформационными изменениями последовательности РНК интрона. Границы между экзонами и интроном обычно четко определены на уровне вторичной структуры пре-РНК. Механизм сплайсирования интронов группы I зависит от стабильной вторичной структуры пре-РНК, состоящей из 10 элементов (Nielsen *et al.*, 2009).

Вторичные структуры, образованные интронами группы I, представляют собой центральный (коровый) элемент и периферические элементы, функцией которых считается стабилизация центрального элемента. Центральный элемент ответственен за каталитическую актив-

ность и представляет собой три спиральные петли/структуры, которые формируют активный центр рибозима. Периферические элементы могут также включать различные последовательности (кодирующие, повторяющиеся и др.) (Nielsen *et al.*, 2009).

Единственным интроном группы I у фотосинтезирующих растений является инtron гена лейциновой тРНК *trnL* пластидного генома, который встроен между U и A в антикодонном триplete UAA тРНК^{Leu}. Считается, что инtron тРНК^{Leu} возник 2,7–3,5 млрд лет назад у предков цианобактерий и порядка 1 млрд лет назад был перенесен в составе пластидного генома в эукариотическую клетку растений (Paquin *et al.*, 1997, Summons *et al.*, 1999). Однако не ясно, сохранена ли рибозимная функция этого интрона, показанная для цианобактерий, в пластидах растений.

В настоящее время известно около 16 000 последовательностей интрона *trnL*, представленных в базе GISSD (Zhou *et al.*, 2008). По-

следовательности, включающие район гена *trnL*, широко используются для реконструкций филогении, в популяционной генетике и для ДНК-штрихкодирования на различных таксономических уровнях (Borsch *et al.*, 2003; Taberlet *et al.*, 2007; Yulita, 2013). Однако данные о вторичных структурах и изменениях во вторичных структурах интрона тРНК^{Leu} крайне ограничены. Поэтому представлялось интересным определить вторичные структуры интрона *trnL* у образцов различного таксономического уровня; определить вариабельность последовательностей, места локализации нуклеотидных замен и возможность изменения вторичной структуры интрона группы I гена *trnL*.

Семейство Fabaceae насчитывает более чем 40 культурных видов и включает 3 подсемейства и более 30 триб, в том числе трибы Viciae, Cicereae, Trifolieae и Phaseoleae подсемейства Faboideae, которое включает большинство агрономически ценных видов, таких как горох, бобы, фасоль, чечевица и др. (Weeden, 2007).

Таким образом, целью работы стало определение первичной последовательности интрона гена *trnL* у семейства Fabaceae, включая представителей всех пяти родов видов трибы Viciae, а также представителей триб Trifolieae, Cicereae и Phaseoleae; исследование нуклеотидного полиморфизма последовательности, в том числе в функционально значимых участках, а также построение вероятной вторичной структуры всего интрона и его отдельных частей.

МАТЕРИАЛЫ И МЕТОДЫ

Для проведения исследования из коллекции ГНУ ВНИИР им. Н.И. Вавилова РАСХН (Санкт-Петербург) были отобраны образцы 16 видов 8 родов, относящихся к трибам Viciae (Fabaceae), Cicereae, Trifolieae, и Phaseoleae подсем. Faboideae (табл. 1).

Выделение ДНК проводили по стандартной методике (Слугина и др., 2013). Амплификацию гена *trnL* проводили с праймерами, соответствующими концевым 5'- и 3'-последовательностям экзона гена *trnL*. Пара праймеров trnLF (CGAAATCGGTAGACGCTACG) и trnLR (GGGGATAGAGGGACTTGAAC) была

разработана на основании известных последовательностей интрона у бобовых, представленных в базе данных NCBI. Анализ полученного продукта проводили методом электрофореза в 1 %-м агарозном геле. Первичные последовательности амплифицированных фрагментов были определены методом прямого секвенирования с использованием праймеров. Выравнивание последовательностей производили с помощью программы MEGA 5.0. (Tamura *et al.*, 2011). Определение позиций единичных нуклеотидных замен и делеций и дупликаций в интроне проводили по последовательности *Lens culinaris*. Для построения вторичных структур пре-мРНК интрона была использована программа MFOLD (Zuker, 2003).

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Вариабельность последовательности интрона *trnL*

Последовательность гена *trnL* у 16 образцов Fabaceae была амплифицирована и секвенирована. Полученные первичные последовательности *trnL* были выровнены, определены границы экзонов и интрона. Как и ожидалось, длины последовательностей экзона I и II гена *trnL* у всех анализируемых образцов составляли 35 п.н. и 50 п.н. соответственно и были инвариантны (табл. 1).

Как и у всех растений, последовательность интрона *trnL* была А/Т богатой. GC состав интрона практически не различался у представителей анализируемых родов Fabaceae (31,4–33,3%). Так, было показано, что для представителей *Gnetum* GC состав варьировал в пределах 34–39%, у *Welwitschia* он составил 30% (Wop, Renner, 2005). GC состав у голосеменных рода *Taxus* составил 32%, у родственного *Cephalotaxus* – 35% (Gielly *et al.*, 1996).

В отличие от экзонов, последовательность интрона *trnL* анализируемых образцов характеризовалась высоким уровнем полиморфизма. Длина последовательности интрона *trnL* варьировала в пределах от 430 п.н. у образца вида *L. culinaris* (триба Viciae) до 555 п.н. у вида *S. arietinum* (триба Cicereae) (табл. 1). При этом следует отметить, что вариабельность по длине выявлялась не только между представителями

Таблица 1

Характеристика интрона гена *trnL* у представителей подсемейства Faboideae

№ п/п	Вид	Каталожный номер/происхождение	Длина интрона <i>trnL</i> , п.н.	Длина Р6 петли, п.н.	Длина Р8 петли, п.н.	GC состав, %
Триба Viciae						
1	<i>Lathyrus sativus</i>	701	442	57	205	32,1
2	<i>Lathyrus aphaca</i>	1293	442	61	201	32,4
3	<i>Lathyrus gmelinii</i>	453	441	61	200	33,3
4	<i>Lathyrus sphaericus</i>	1329	437	61	195	32,5
5	<i>Lathyrus ochrus</i>	380	436	61	195	33,3
6	<i>Lathyrus clymenum</i>	1370	441	61	200	32,9
7	<i>Pisum fulvum</i>	422009	437	57	200	32,7
8	<i>Pisum sativum</i>	3980	432	57	195	32,9
9	<i>Vavilovia formosa</i>	4/Армения	432	57	195	33,3
10	<i>Vicia unijuga</i>	k-35828	452	61	211	32,7
11	<i>Vicia villosa</i>	277	421	30	211	31,4
12	<i>Vicia faba</i>	1506	451	50	221	31,5
13	<i>Lens culinaris</i>	224	430	57	193	32,1
Триба Trifolieae						
14	<i>Trifolium repens</i>		554	58	316	31,6
Триба Ciceraceae						
15	<i>Cicer arietinum</i>		555	60	315	30,1
Триба Phaseoleae						
16	<i>Phaseolus vulgaris</i>		523	56	287	31,5

различных триб и родов, но и между видами одного рода (табл. 1). И если различия в длине интрона у анализируемых представителей родов *Pisum* и *Lathyrus* не превышали 6 нуклеотидов, то у представителей *Vicia* внутривидовая вариабельность по длине последовательности интрона составила 31 нуклеотид.

Выровненная длина интрона *trnL* для всех анализируемых образцов составила 593 п.н. (табл. 1). Последовательность интрона была крайне вариабельной: у анализируемых образцов 182 сайта были полиморфны, уровень полиморфизма составил 30,7 %. Число информативных в отношении парсимонии сайтов – 38 (6,4 %), что несколько выше, чем показано в среднем для покрытосеменных растений (Borsch *et al.*, 2003).

В последовательности интрона были обнаружены как многочисленные единичные нуклеотидные замены, так и инсерции и делеции. Для некоторых образцов были выявлены

видоспецифичные делеции и дупликации, остальные инсерции встречались у групп видов (табл. 2, рис. 1).

Вторичная структура интрона *trnL*

Вторичная структура интрона *trnL* была определена у достаточно ограниченного ряда растений, таких как нимфея (Borsch *et al.*, 2003), горчавка (Gielly *et al.*, 1996), и у представителей семейства Dipterocarpaceae (Yulita, 2013).

В качестве модельной считается вторичная структура интрона *trnL* у нимфеи, которая была подробно описана Borsch с соавт. (2003). В структуре интрона выделяют 10 петель и 2 функциональных домена: каталитический и субстратный. Наиболее вариабельными участками интрона считаются последовательности петель Р6 и Р8 (Taberlet *et al.*, 2007).

На основании полученных данных о первичной последовательности нами была построена

Таблица 2

Положение инсерций/делеций в последовательности P8 петли
у представителей подсемейства Faboideae

Положение в последовательности интрона (по <i>Lens culinaris</i>)	Инсерция/ делеция	Длина	Вид/род
199	Инсерция	114	<i>Trifolium repens</i>
		109	<i>Cicer arietinum</i>
		79	<i>Phaseolus vulgaris</i>
199	Тандемный повтор	5	<i>Pisum fulvum</i>
216	Делеция	15	<i>Phaseolus vulgaris</i>
229	Делеция	4	<i>Lathyrus sphaericus</i>
232	Делеция	4	<i>Lens culinaris</i>
		12	<i>Vicia unijuga</i> <i>Vicia villosa</i>
		17	<i>Vicia faba</i>
240	Инсерция	4	<i>Phaseolus vulgaris</i>
250	Инсерция	5	<i>Vicia faba</i>
261	Делеция	5	<i>Lathyrus ochrus</i>
			<i>Pisum fulvum</i>
			<i>Pisum sativum</i>
			<i>Vavilovia formosa</i>
278	Инсерция	3	<i>Cicer arietinum</i>
289	Делеция	5	<i>Cicer arietinum</i>
			1
	Инсерция	1	<i>Lathyrus sativus</i> <i>Pisum fulvum</i> <i>Pisum sativum</i> <i>Vavilovia formosa</i> <i>Lathyrus aphaca</i> <i>Lathyrus sphaericus</i> <i>Lathyrus gmelinii</i> <i>Lathyrus ochrus</i> <i>Lathyrus clymenum</i>
299	Тандемный повтор	5	<i>Lathyrus sativus</i>

вероятная вторичная структура этого интрона для анализируемых образцов Fabaceae. При построении вторичной структуры у видов Fabaceae были определены все 10 петель и последовательности обоих функциональных доменов. При этом было показано, что выявленные точковые замены локализовались даже в функциональных доменах и мотивах, которые считаются наиболее консервативными участками интрона (рис. 2). Так, в последовательности каталитического центра были выявлены четыре

замены: две, локализованные в одном положении в последовательности R2 центра, A/G (*Tr. repens*) и A/U (*Ph. vulgaris*), и две замены в последовательности S центра у видов *V. villosa* U/G и *V. faba* A/G (рис. 2).

Полиморфизм петли P6

Известно, что петля P6 интронов группы I может несколько различаться как по длине, так и по нуклеотидным последовательностям

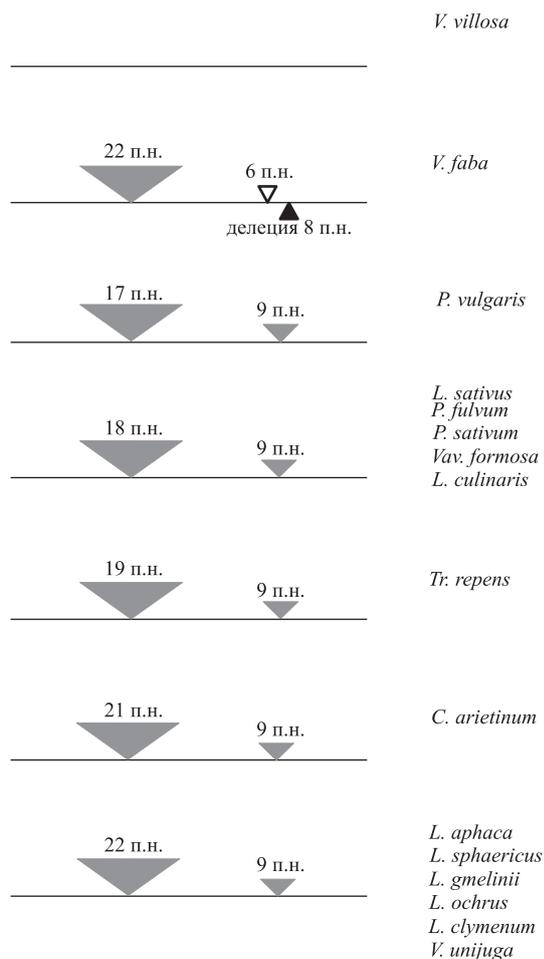


Рис. 1. Схема расположения инсерций/делеций в петле P6 у разных видов подсемейства Faboideae.

(Taberlet *et al.*, 2007). Анализ последовательностей петли P6 у Fabaceae позволил выявить значительные различия, при этом интересно то, что длина последовательностей, формирующих P6 петлю, различалась не только у представителей различных родов, но и у видов одного рода. Так, разница в размере P6 у образцов с максимальной длиной последовательности петли (*V. unijuga* и некоторые виды *Lathyrus*) и минимальной (*V. villosa*) составила 31 нуклеотид (табл. 1, рис. 1). P. Taberlet с соавт. (2007) в своих исследованиях показали невысокий уровень различия в длине петли P6 и в ее нуклеотидной последовательности внутри родов и близкородственных групп видов. Так, например, последовательность P6 петли у трех исследованных видов рода *Solanum* различалась лишь на один нуклеотид при одинаковой длине. Аналогично последовательность петли у 8 ро-

дов злаковых различалась по длине не более чем на 5 нуклеотидов, при этом последовательности петли у образцов родов *Triticum* и *Secale* были полностью идентичны, а у последовательности петли образца *Zea mays* по отношению к роду *Panicum* была детектирована только однонуклеотидная делеция (Taberlet *et al.*, 2007).

Для всех анализируемых образцов Fabaceae была построена вторичная структура P6 петли. Было получено пять разных типов вторичной структуры петли (рис. 3). Наибольшее число образцов относится к I и II типу. Отметим то, что некоторые образцы одного рода имели разный тип строения петли. Так, образцы родов *Pisum*, *Vavilovia*, *Lens* и *Cicer* формировали P6 петли I типа. Почти все образцы рода *Lathyrus* образовывали петли II типа. Исключение составил образец *L. sativus*, у которого был I тип строения петли. Все три образца рода *Vicia*, взятые в анализ, различались по типу строения P6 петли. Так, образец вида *V. unijuga* имел тип строения петли II, *V. faba* – тип IV, *V. villosa* – тип V. Такие различия в строении петли P6 можно объяснить тем, что данные виды относятся к разным секциям рода *Vicia*: *Vicilla*, *Vicia* и *Cracca* соответственно (Schaefer *et al.*, 2012).

Стоит отметить, что образцы родов *Pisum* и *Vavilovia* имели одинаковую длину и различались лишь тремя нуклеотидами, что подтверждает точку зрения об их близком родстве (Oskoueiyani *et al.*, 2010). Полностью идентичные последовательности P6 петли были детектированы у образцов *P. sativum* и *P. fulvum*; *L. ochrus* и *L. clymenum*; *L. sphaericus* и *L. aphaca* соответственно.

Полиморфизм петли P8

Последовательность петли P8 у Fabaceae была крайне вариативной и содержала в себе многочисленные нуклеотидные замены, инсерции и делеции. Ее протяженность варьировала от 193 п.н. у рода *Lens* до 316 п.н. у образцов родов *Trifolium* и *Cicer*. Среди единичных замен и делеций и дупликаций были детектированы как видоспецифичные, так и характерные для групп видов (табл. 2).

Наиболее протяженную инсерцию содержали в себе образцы *Tr. repens* (114 п.н.), *C. arieti-*

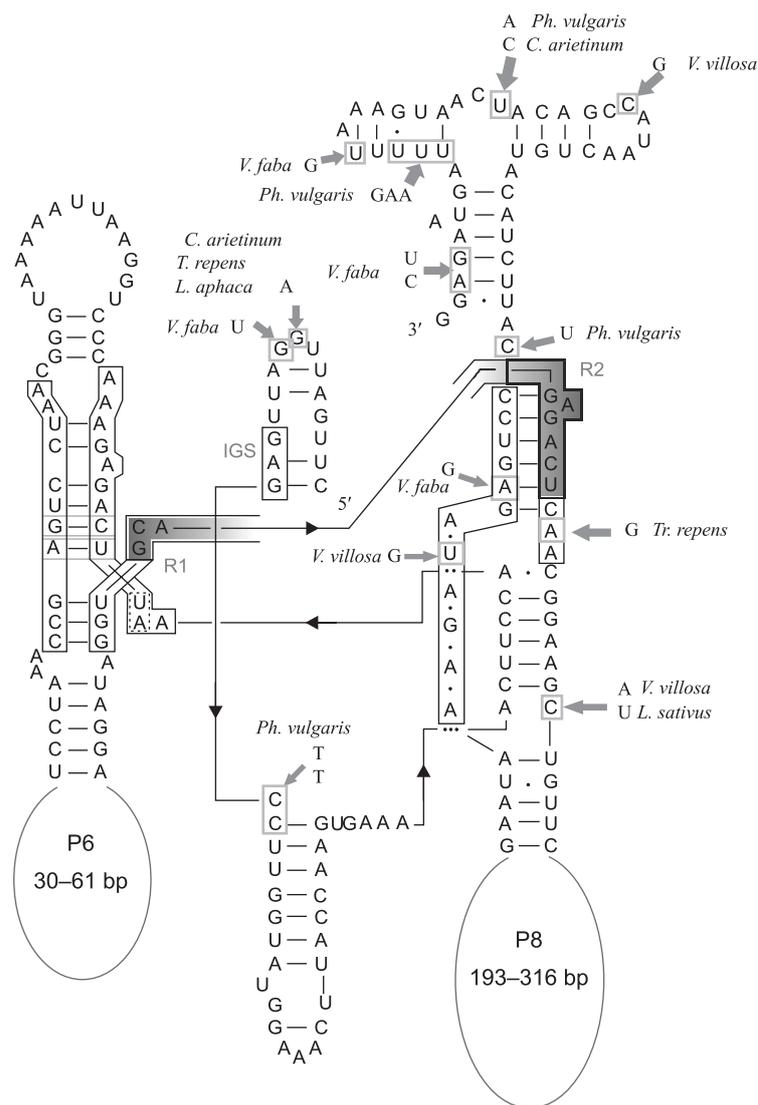


Рис. 2. Вторичная структура интрона гена *trnL*.

num (109 п.н.) и *Ph. vulgaris* (79 п.н.) (табл. 2). Образцы *L. ochrus* и *L. clymenum*, относящиеся к секции *Clymenum* (по классификации Kurіcha, 1976), различались наличием единственной инсерции/делеции в положении 261 п.н. (табл. 2), а также содержали единственную замену из всей последовательности интрона.

При исследовании последовательности участка пластидного генома *trnT-trnF* выделяют 8 «горячих точек/районов» с наивысшей вероятностью инсерций, из них две расположены в I и II спиральных элементах петли P8 (Borsch *et al.*, 2003). В нашем исследовании наиболее протяженные вставки находились в элементе I.

Последовательности петли P8 у образцов рода *Pisum* (*P. sativum* и *P. fulvum*) различались лишь наличием однокопийного тандемного пятинуклеотидного повтора у *P. fulvum* в районе 199 п.н. от первого нуклеотида интрона. Образец *V. formosa* в своей последовательности имел идентичные делеции и дупликации по отношению к образцу вида *P. sativum*, однако содержал 7 нуклеотидных замен.

Таким образом, в данной работе был проведен анализ полиморфизма интрона гена *trnL* и предложена его вторичная структура. Последовательность интрона была достаточно полиморфной и содержала замены даже в функциональных центрах. Были определены

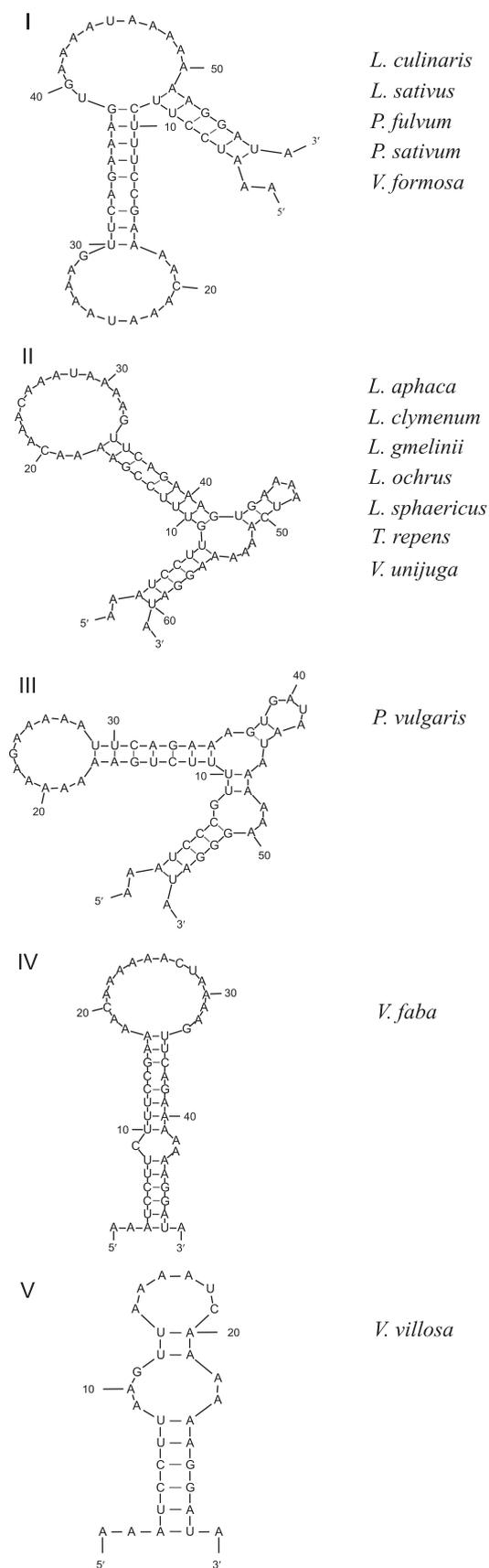


Рис. 3. Типы вторичной структуры петли Р6.

все функциональные участки и петли интрона, относящегося к интронам группы I. Было показано наличие пяти типов строения вторичной структуры Р6 петли. При этом виды одного рода имели различный тип строения петли Р6.

Работа выполнена при финансовой поддержке программ Президиума Российской академии наук «Динамика и сохранение генофондов» и «Молекулярная и клеточная биология».

ЛИТЕРАТУРА

- Слугина М.А., Снигирь Е.А., Рыжова Н.Н., Кочиева Е.З. Структура и полиморфизм фрагмента локуса *Pain-I*, кодирующего вакуолярную инвертазу *Solanum* // Молекуляр. биология. 2013. Т. 47. С. 243–250.
- Borsch T., Hilu K.W., Quandt D., Wilde V., Neinhuis C., Barthlott W. Noncoding plastid trnT-trnF sequences reveal a well resolved phylogeny of basal angiosperms // J. Evol. Biol. 2003. V. 16. P. 558–576.
- Gielly L., Yuan Y.-M., Kupfer P., Taberlet P. Phylogenetic use of noncoding regions in the genus *Gentiana* L.: chloroplast *trnL* (UAA) intron versus nuclear ribosomal internal transcribed spacer sequences // Mol. Phylogenet. Evol. 1996. V. 5. No. 3. P. 460–466.
- Kupicha F.K. The infrageneric structure of *Vicia* // Notes from the Royal Botanic Garden. Edinburg. 1976. V. 34. No. 3. P. 287–326.
- Nielsen H., Johansen S.D. Group I introns // RNA Biology. 2009. V. 6. No. 4. P. 375–383
- Oskoueian R., Osaloo S.K., Maassoumi A.A., Nejadstatti T., Mozaffarian V. Phylogenetic status of *Vavilovia formosa* (Fabaceae-Fabeae) based on nrDNA ITS and cpDNA sequences // Biochem. Syst. Ecol. 2010. V. 38. P. 313–319.
- Paquin B., Kathe S.D., Nierzwicki-Bauer S.A., Shub D.A. Origin and evolution of group – I introns in cyanobacterial tRNA genes // J. Bacteriol. 1997. V. 179. P. 6798–6806.
- Schaefer H., Hechenleitner P., Santos-Guerra A., Menezes de Sequeira M., Pennington R., Kenicer G., Carine M.A. Systematics, biogeography, and character evolution of the legume tribe Fabeae with special focus on the middle-Atlantic island lineages // BMC Evol. Biol. 2012. V. 12. P. 250.
- Summons R.E., Jahnke L.L., Logan G.A., Hope J.M. 2-Methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis // Nature. 1999. V. 398. P. 554–557.
- Taberlet P., Coissac E., Pompanon F., Gielly L., Miquel C., Valentini A., Vermet T., Corthier G., Brochmann C., Willerslev E. Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding // Nucl. Acids Res. 2007. V. 35. No. 3. e14.
- Tamura K., Peterson D., Peterson N., Stecher G., Nei M., Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods // Mol. Biol. Evol. 2011. V. 28. P. 2731–2739.
- Weeden N.F. Genetic changes accompanying the domestica-

- tion of *Pisum sativum*: is there a common genetic basis to the 'domestication syndrome' for legumes // *Ann. Bot.* 2007. V. 100. P. 1017–1025.
- Won H., Renner S.S. The chloroplast trnT–trnF region in the seed plant lineage Gnetales // *J. Mol. Evol.* 2005. V. 61. P. 425–436.
- Yulita K.S. Secondary structures of chloroplast *trnL* intron in dipterocarpaceae and its implication for the phylogenetic reconstruction // *Hayati J. Biosci.* 2013. V. 20. No. 1. P. 31–39.
- Zhou Y., Lu C., Wu Q.-J., Wang Y., Sun Z.-T., Deng J.-C., Zhang Y. GISSD: Group I intron sequence and structure database // *Nucl. Acids Res.* 2008. V. 36. D31–D37.
- Zuker M. Mfold web server for nucleic acid folding and hybridization prediction // *Nucl. Acids Res.* 2003. V. 31. P. 3406–3415.

PLASTID *TRNL* INTRON VARIABILITY IN FABOIDEAE SPECIES (FABACEAE)

E.A. D'yachenko^{1,2}, M.A. Filyushin¹, E.P. Pronin², E.Z. Kochieva^{1,3}

¹ Bioengineering Center, Russian Academy of Sciences, Moscow, Russia,
e-mail: dyachenko-el@yandex.ru;

² All-Russia Research Institute for Vegetable Breeding and Seed Production,
VNISSOK Village, Odintsovo region, Moscow oblast, Russia;

³ Chair of Biotechnology, Moscow State University, Moscow, Russia

Summary

The intron located between the first and second nucleotide of the leucine tRNA anticodon is the only representative of group I introns in higher plants. In this paper, for the first time the intron sequence of the plastid *trnL* gene is described in 16 legume species, and putative secondary structures of the entire intron and some of its functional domains are reconstructed. It has been found that genera of the Fabaceae family, as well as species within a single genus, are highly diverse in this sequence. Single nucleotide polymorphisms have been found in sequences of the catalytic center, believed to be highly conserved.

Key words: plants, group I intron, chloroplast genome, nucleotide polymorphism, intron secondary structure.