

Сравнительный анализ полных последовательностей пластидных геномов чеснока *Allium sativum* и лука репчатого *Allium cepa*


М.А. Филюшин , А.М. Мазур, А.В. Щенникова, Е.З. Кочиева

Федеральный исследовательский центр «Фундаментальные основы биотехнологии» Российской академии наук, Институт биоинженерии, Москва, Россия

Секвенирование и сравнительная характеристика пластидных геномов растений – актуальный инструмент для современных филогенетических и таксономических исследований. Род *Allium* L. (сем. Amaryllidaceae) объединяет более 900 видов растений, включая такие экономически значимые овощные культуры, как чеснок *Allium sativum*, лук репчатый *Allium cepa*, лук-порей *A. porrum* и др. В настоящей работе впервые определена полная нуклеотидная последовательность пластидного генома чеснока *A. sativum*. Пластом *A. sativum* имел размер 153172 п. н. и состоял из большой уникальной (LSC, 82035 п. н.) и малой уникальной (SSC, 18015 п. н.) копий, разделенных инвертированными повторами (IRa и IRb) размером 26561 п. н. каждый. В пластидном геноме чеснока аннотировано 134 гена, из них 82 белок-кодирующих, 38 тРНК, 8 рРНК и 6 псевдогенов. Сравнительный анализ пластидных геномов *A. sativum* и *A. cepa* выявил различия в размерах структурных элементов и спейсеров на границах инвертированных повторов. Общее число генов в пластомах *A. sativum* и *A. cepa* было одинаковым, однако генный состав различался: ген *rpl22* был функциональным у *A. sativum* и псевдогеном у *A. cepa*, а ген *rps16* – наоборот, псевдогеном у *A. sativum* и белок-кодирующим у *A. cepa*. В последовательностях пластидных геномов чеснока *A. sativum* и лука репчатого *A. cepa* были идентифицированы 32 микросателлитные последовательности, больше половины которых – динуклеотидные, остальные – тетра-, пента- и гексануклеотидные, а тринуклеотидные отсутствовали. Сравнимые пластидные геномы отличались числом повторов ряда микросателлитов, а некоторые микросателлиты встречались только у одного из видов.

Ключевые слова: пластидный геном; чеснок; *Allium sativum*; *Allium cepa*.

Comparative analysis of the complete plastomes of garlic *Allium sativum* and bulb onion *Allium cepa*

М.А. Filyushin , А.М. Mazur, А.В. Shchennikova, Е.З. Kochieva

Federal Research Centre "Fundamentals of Biotechnology", RAS, Institute of Bioengineering, Moscow, Russia

Sequencing and comparative characterization of plant plastid genomes, or plastomes, is an important tool for modern phylogenetic and taxonomic studies, as well as for understanding the plastome evolution. The genus *Allium* L. (family Amaryllidaceae) incorporates more than 900 species, includes economically significant vegetable crops such as garlic *A. sativum*, onion *A. cepa*, leek *A. porrum*, etc. In this work, the plastome of garlic *A. sativum* has been completely sequenced. The *A. sativum* plastome is 153172 bp in size. It consists of a large unique (LSC, 82035 bp) and small unique (SSC, 18015 bp) copies, separated by inverted repeats (IRa and IRb) of 26561 bp each. In the garlic plastome, 134 genes have been annotated: 82 protein-coding genes, 38 tRNA genes, 8 rRNA genes, and 6 pseudogenes. Comparative analysis of *A. sativum* and *A. cepa* plastomes reveals differences in the sizes of structural elements and spacers at the inverted repeat boundaries. The total numbers of genes in *A. sativum* and *A. cepa* are the same, but the gene composition is different: the *rpl22* gene is functional in *A. sativum*, being a pseudogene in *A. cepa*; conversely, the *rps16* gene is a pseudogene in *A. sativum* and a protein-coding gene in *A. cepa*. In the *A. sativum* and *A. cepa* plastomes, 32 SSR sequences have been identified. More than half of them are dinucleotides, and the remaining are tetra-, penta-, and hexanucleotides at the same time, trinucleotides were absent. The compared plastomes differ in the numbers of certain SSRs, and some are present in only one of the species.

Key word: plastid genome; garlic; *Allium sativum*; *Allium cepa*.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Филюшин М.А., Мазур А.М., Щенникова А.В., Кочиева Е.З. Сравнительный анализ полных последовательностей пластидных геномов чеснока *Allium sativum* и лука репчатого *Allium cepa*. Вавиловский журнал генетики и селекции. 2018;22(5):524-530. DOI 10.18699/VJ18.390

HOW TO CITE THIS ARTICLE:

Filyushin M.A., Mazur A.M., Shchennikova A.V., Kochieva E.Z. Comparative analysis of the complete plastomes of garlic *Allium sativum* and bulb onion *Allium cepa*. Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding. 2018;22(5):524-530. DOI 10.18699/VJ18.390 (in Russian)

Крупнейший род однодольных растений *Allium* L. (сем. Amaryllidaceae, Asparagales) объединяет более 900 видов (Seregin et al., 2015), произрастающих преимущественно в Северном полушарии. К этому роду относятся овощные культуры, например чеснок, лук репчатый, лук-порей и другие, а ряд видов используется в качестве декоративных. Многие виды луков являются редкими и эндемичными (Friesen et al., 2006). Чеснок *Allium sativum* и лук репчатый *A. cepa* – наиболее экономически значимые виды рода, общемировой объем производства которых, по данным Продовольственной и сельскохозяйственной организации ООН (FAO) за 2016 г., составил 26.5 и 57.2 млн т соответственно.

Пищевая ценность как чеснока, так и лука репчатого определяется, прежде всего, углеводным составом листьев и луковиц, а также наличием специфических вторичных метаболитов (Jones et al., 2004). Образование первичных углеводов происходит в хлоропластах в процессе фотосинтеза. Хотя фотосинтез признается ключевой функцией пластид, они играют также важную роль в других аспектах физиологии и развития растений, включая синтез аминокислот, нуклеотидов, жирных кислот, фитогормонов, витаминов и множества метаболитов (Daniell et al., 2016).

Пластидный геном растений, за исключением нефотосинтезирующих видов (Bellot, Renner, 2016; Ravin et al., 2016) и видов семейства Fabaceae (Sveinsson, Cronk, 2014), имеет консервативную четырехчастную структуру – большая уникальная (large single copy, LSC) и малая уникальная (small single copy, SSC) копии, разделенные инвертированными повторами (inverted repeat, IR), а также сходный порядок генов (Wicke et al., 2011). По данным NCBI, в настоящее время секвенированы полные нуклеотидные последовательности пластидных геномов более чем у 2400 видов высших растений.

Для определения последовательностей пластов существует три стратегии: 1) секвенирование тотальной ДНК и сборка пластома; 2) выделение и секвенирование хлоропластной ДНК; 3) амплификация с использованием в качестве матрицы тотальной ДНК перекрывающихся фрагментов пластома и их дальнейшее секвенирование и сборка (Dong et al., 2013). Геном представителей рода *Allium* имеет большой размер (Ricroch et al., 2005), что делает проблематичной сборку пластидного генома при секвенировании тотальной ДНК. Второй метод весьма трудозатратный и требует больших количеств растительного материала. С помощью этого метода была получена полная последовательность пластидного генома только у одного представителя рода *Allium* – лука репчатого *A. cepa* (von Kohn et al., 2013). Третьим методом определены последовательности пластома многих видов растений, например кувшинки *Nymphaea alba* (Gogemykin et al., 2004), огурца *Cucumis sativus* (Chung et al., 2007), ряски *Lemna minor* (Mardanov et al., 2008) и др. Так как к настоящему времени секвенирована полная последовательность пластидного генома нескольких генотипов *Allium cepa*, то представлялось возможным разработать универсальную систему праймеров для амплификации перекрывающихся участков пластома у других видов *Allium*. Поэтому цель нашей работы – разработка системы праймеров и

определение полной нуклеотидной последовательности пластидного генома чеснока *A. sativum* и ее последующее сравнение с пластомом *A. cepa*.

Материалы и методы

Для секвенирования был взят образец *A. sativum* из Узбекистана (каталожный номер k31) из коллекции Всероссийского института генетических ресурсов растений им. Н.И. Вавилова. Тотальную ДНК выделяли из листьев методом СТАВ (Doyle J.J., Doyle J.L., 1987) с дополнительной депротенизацией фенол-хлороформом (1:1). На основе ранее определенных последовательностей пластома нескольких генотипов лука репчатого *A. cepa* (von Kohn et al., 2013; Kim et al., 2015) было разработано 11 пар праймеров (табл. 1) для амплификации пластидного генома у родственного ему чеснока. Размеры участков варьировали от 6 до 21 т. п. н., их концы перекрывались на 500–1000 п. н. Амплификацию фрагментов проводили с использованием набора LongAmp® Hot Start Taq DNA Polymerase (New England Biolabs, США). Эквимольная смесь полученных амплификатов, содержащая по 100 нг каждого фрагмента, после прободопготовки была секвенирована с помощью Illumina HiSeq 1500 Sequencing System (ЗАО Геноаналитика, Москва, Россия). Сборку полученного пула чтений в контиги проводили с помощью Spades v.3.8 (Bankevich et al., 2012). Получено 11 контигов, которые затем были объединены в единую последовательность, в качестве референса использовали пластом *A. cepa* (KF728079). Пробелы были закрыты в программе Bandage (Wick et al., 2015), после этого чтения были сопоставлены с полученным одиночным контигом для обеспечения правильности готовой сборки и определения величины покрытия. Покрытие участков составило от 900x до 8000x, длина контигов варьировала от 6 до 21 т. п. н. Полученный пластидный геном был аннотирован с помощью программ DOGMA (<http://dogma.cccb.utexas.edu>) и tRNAscan-SE (Lowe, Eddy, 1997). Координаты всех генов проверены вручную. Физическая карта пластидного генома *A. sativum* была построена с помощью программы OGDRAW (рис. 1) (Lohse et al., 2013). Границы инвертированных повторов были перепроверены секвенированием с праймерами, специфичными к фланкирующим участкам: *ycf1-ndhF* (5'-ATTGGTATTATTCTGGATACGGC-3'; 5'-GATCTGTTTATTCATCTCTACAG-3'), *rpl22-rpl2* (5'-GATGCTCGATAAGGCATAAG-3'; 5'-GTAAGCGTCCTGTAGTAAG-3'), *rpl2-psbA* (5'-CTGTAGTAAGAGGAGTAGT-3'; 5'-TTGATAGTCAAGGTCGTGT-3').

Для проведения сравнительного анализа пластидных геномов использована последовательность стерильного генотипа *A. cepa* (KF728079). Последовательность *A. cepa* KF728079 была аннотирована нами *de novo*, так как в исходной аннотации (von Kohn et al., 2013) ряд генов, в частности tPHK, имеют неверные координаты, а, например, в гене *rps16* не определены границы экзонов, в результате чего он был признан псевдогеном, хотя таковым не является. Выравнивание последовательностей, определение GC-состава проводили в программе MEGA 7.0 (Kumar et al., 2016). Поиск микросателлитных последовательностей выполняли в программе SSRIT (Temnykh et al., 2001) со следующими параметрами: для ди- и тринуклеотидов – не

Таблица 1. Последовательности праймеров, разработанных для амплификации протяженных участков пластидного генома *A. sativum*

Участок пластидного генома	Последовательность праймеров	Температура отжига, °C	Размер ампликона, т. п. н.
<i>rpl2-atpA</i>	5'-GGCTATAATTGGAGATACCATT-3' 5'-GAAATAGTAATATTATCCGYGAA-3'	54	12
<i>atpA-rpoC1</i>	5'-TAATCGTTGACCTCTTGCCAATT-3' 5'-TCCCATCGAAGTTCACATGAAT-3'	57	10
<i>rpoC1-psaB</i>	5'-TGTAARGCTTCTTCKATTCTCG-3' 5'-ATTTCCAAGGTTTAGCCAAG-3'	53	20
<i>psaB-rbcL</i>	5'-CATAAAGATTCCAYTGACC-3' 5'-GTTGTCCAKGTACCAGTAGAAG-3'	53	18
<i>rbcL-psaJ</i>	5'-ATGTCACCMCAAACAGARACT-3' 5'-TACTGGGTCATTCATAGCATTAC-3'	57	10
<i>psaJ-psbB</i>	5'-ACTAAATTCATCRAGTTGTTCC-3' 5'-GGAATGCTCCAAATTCASACTT-3'	54	10
<i>psbB-ycf2</i>	5'-GAATTAGAYCGKGTACTTT-3' 5'-TCTCTGGATTGATCAGARGAT-3'	55	15
<i>ycf2-ndhB</i>	5'-TTAGACAATGTGTGGTTGKTA-3' 5'-ACGAAACCAAGAAATAACCC-3'	53	8
<i>ndhB-rnn23</i>	5'-AGGGTATCCTGAGCAATTKCAA-3' 5'-CAAACCTCCTGGCTGTCTCT-3'	58	11
<i>rnn23-ndhF</i>	5'-TGATCTATCCATGACCAGG-3' 5'-GATGAAATCTTAATGATAGTTGG-3'	54	6
<i>ndhF-trnN</i>	5'-GTAATAGATMCGAAACATATAA-3' 5'-CTCCCAAGTAGGATTCGA-3'	52	21

менее пяти повторов, для тетра-, пента- и гексануклеотидов – не менее трех повторов.

Результаты и обсуждение

Длина полной нуклеотидной последовательности пластидного генома чеснока *A. sativum* составила 153 172 п. н. (см. рис. 1). Последовательность пластома *A. sativum* была депонирована в базу данных NCBI (KX683282). Краткая аннотация пластидного генома чеснока *A. sativum* опубликована нами ранее (Filyushin et al., 2016). Пластидный геном *A. sativum* состоит из большой уникальной копии (82035 п. н.) и малой уникальной копии (18015 п. н.), разделенных инвертированными повторами (IRa и IRb) размером 26561 п. н. каждый. В пластидном геноме аннотировано 134 гена, из них 82 белок-кодирующих, 38 тРНК, 8 рРНК и 6 псевдогенов (Filyushin et al., 2016).

Белок-кодирующие гены, аннотированные в пластидном геноме *A. sativum*, включали гены большой субъединицы рибосомы (гены *rpl2*, *rpl14*, *rpl16*, *rpl20*, *rpl22*, *rpl23*, *rpl32*, *rpl33* и *rpl36*), малой субъединицы рибосомы (гены *rps3*, *rps4*, *rps7*, *rps8*, *rps11*, *rps12*, *rps14*, *rps15*, *rps18* и *rps19*), фотосистемы I (*psaA*, *psaB*, *psaC*, *psaI* и *psaJ*), фотосистемы II (*psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbT* и *psbZ*), субъединиц АТФ-синтазы (*atpA*, *atpB*, *atpE*, *atpF*, *atpH* и *atpI*) и других семейств. Большинство генов находится в пластидном геноме *A. sativum* в одной копии, только 18 генов представлены двумя копиями за счет нахождения в инвертированных повторах: шесть белок-кодирующих генов (*rps19*, *rpl2*, *rpl23*, *ycf2*, *ndhB* и *rps7*), восемь тРНК

(*trnR-ACG*, *trnL-CAA*, *trnV-GAC*, *trnH-GUG*, *trnI-CAU*, *trnI-GAU*, *trnA-UGC* и *trnN-GUU*) и все четыре гена рРНК (*rnn4.5*, *rnn5*, *rnn16* и *rnn23*). Шесть генов были признаны псевдогенами, так как либо содержали внутренний стоп-кодон (*rps2*, *rps16*, *infA*, два *ycf15* в IRs), либо были образованы в результате неполной дупликации (*ycf1*). Семнадцать генов содержали интроны, из них пятнадцать – один интрон (*atpF*, *rpoC1*, *trnK-UUU*, *trnG-GCC*, *trnL-UAA*, *trnV-UAC*, *ndhA*; четыре гена в IRs: *rpl2*, *ndhB*, *trnI-GAU* и *trnA-UGC*), а два (*clpP* и *ycf3*) – два интрона (Filyushin et al., 2016).

Сравнение пластидных геномов чеснока *A. sativum* и лука репчатого *A. cepa* (KF728079) выявило различия как в размерах структурных элементов, так и в генном составе (табл. 2). В результате анализа нуклеотидных последовательностей сравниваемых пластома идентифицировано 2063 вариабельных сайта (1.33 % от выровненной длины). Пластома *A. sativum* и *A. cepa* имели одинаковое количество белок-кодирующих генов, однако состав функциональных и псевдогенов был различен. Так, *rpl22* у *A. cepa* – псевдоген, а у *A. sativum* в этом гене не выявлено внутренних стоп-кодонов. В то же время ген *rps16* у *A. cepa* – белок-кодирующий, а у *A. sativum* этот ген считается псевдогеном, так как содержит стоп-кодон во втором экзоне.

Последовательности *rps2*, *infA*, *ycf15* являются псевдогенами и у *A. sativum* и у *A. cepa* (табл. 2). Ген *rps2*, как было показано в результате массивного параллельного секвенирования тотальной ДНК 48 представителей *Asparagus* (Steele et al., 2012), содержит преждевременный

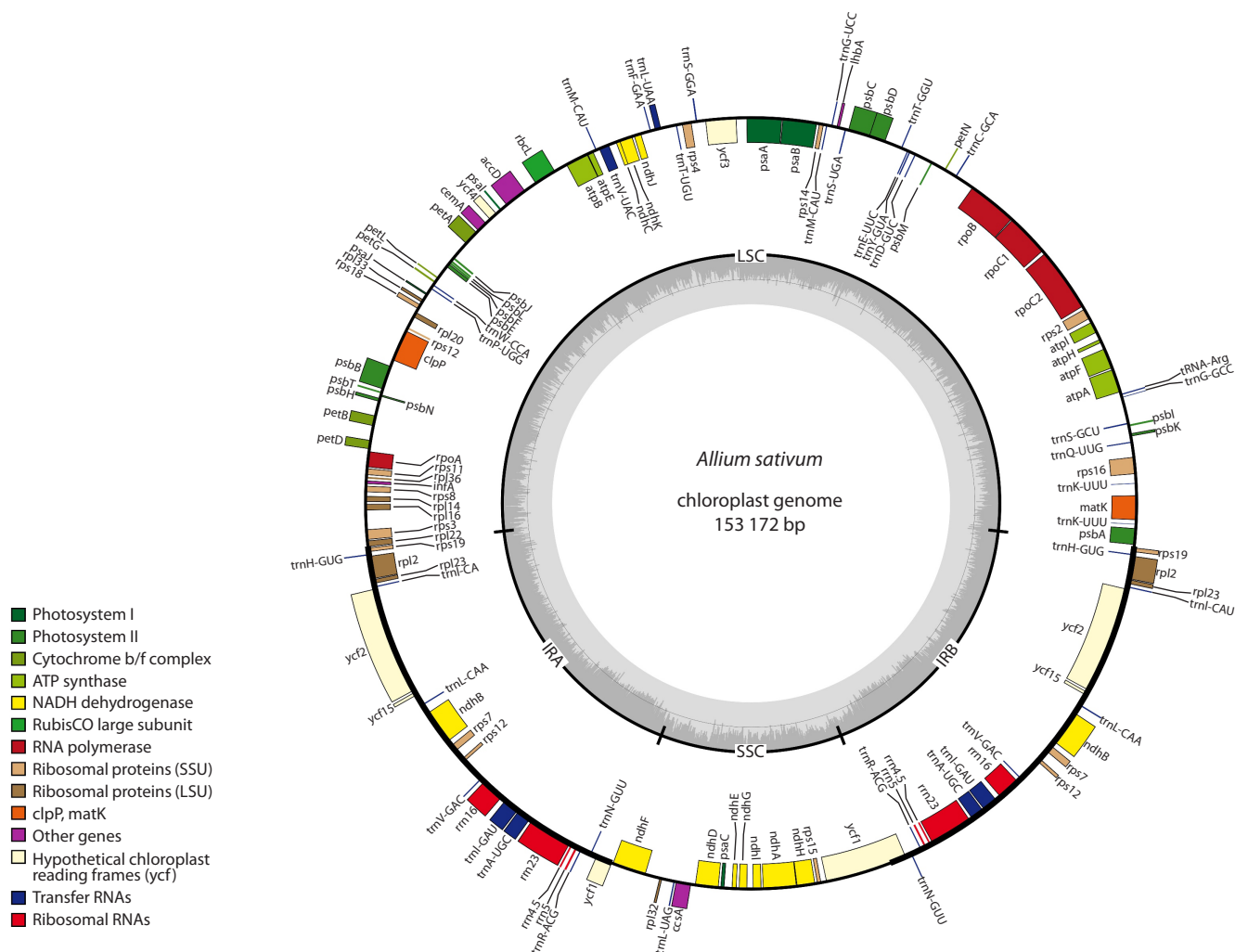


Рис. 1. Физическая карта пластидного генома *A. sativum* (KX683282).

Гены, находящиеся внутри круга, транскрибируются по часовой стрелке, а за пределами круга – против часовой стрелки. Темно-серый цвет во внутреннем круге соответствует содержанию GC.

стоп-кодон лишь у некоторых представителей этого порядка, включая виды *Allium*. Анализ доступных в NCBI последовательностей пластомеров показал, что *ycf15* – псевдоген не только у видов *Allium*, но и у видов других родов порядка Asparagales, в то время как *infA* (псевдоген у *A. sativum* и *A. cepa*) у видов *A. altaicum* (MH159130; собственные неопубликованные данные) и *A. obliquum* (NC_037199; Filyushin et al., 2018) – белок-кодирующий.

Для функционирования пластид необходимо более 2000 белков, при этом в самом пластидном геноме кодируются не более 90 белков (Синявская и др., 2015). Перенос генетического материала из цитоплазматических геномов в ядерный геном считается характерной чертой эволюции современных растений (Brandvain, Wade, 2009; Sloan, 2015). Для ряда видов растений ранее был показан перенос гена *rps16* в ядерный геном, при этом в пластидном геноме он либо сохраняется, либо отсутствует (полностью, частично или содержит стоп-кодон), а рибосомальный белок S16 транспортируется в пластиды из ядра (Keller et al., 2017). Наличие/отсутствие гена *rps16* у видов одного рода было описано ранее, например, для видов *Arabidopsis*

Таблица 2. Сравнительная характеристика пластидных геномов *A. sativum* и *A. cepa*

Показатель	<i>A. sativum</i> (KX683282)	<i>A. cepa</i> (KF728079)
Общая длина, п. н.	153172	153355
LSC	82035	82467
SSC	18015	17918
IRs	26561	26485
GC состав, %	36.7	36.8
Всего генов	134	134
Число		
белок-кодирующих генов	82	82
генов тРНК	38	38
генов рРНК	8	8
Число и состав псевдогенов	6 (<i>rps2</i> , <i>rps16</i> , <i>infA</i> , <i>ycf15</i> (IRs), <i>ycf1</i> (IRa)*)	6 (<i>rps2</i> , <i>rpl22</i> , <i>infA</i> , <i>ycf15</i> (IRs), <i>ycf1</i> (IRa)*)

* *ycf1* в IRa считается псевдогеном из-за неполной дупликации.

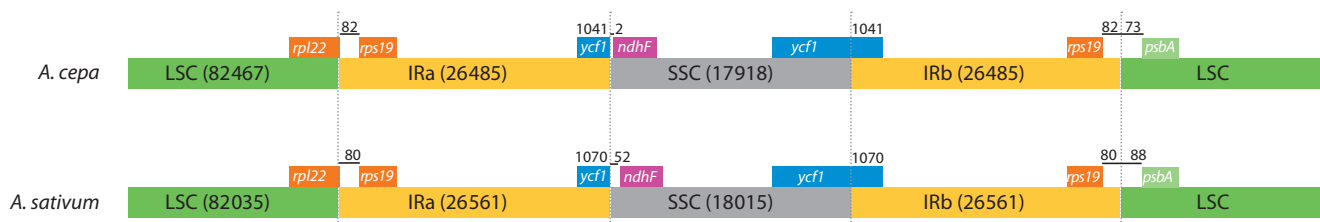


Рис. 2. Сравнение границ инвертированных повторов в пластидных геномах *A. cepa* (KF728079) и *A. sativum* (KX683282).

Таблица 3. Характеристика микросателлитных последовательностей в пластидных геномах *A. sativum* (KX683282) и *A. cepa* (KF728079)

Локализация	Область	Мотив	Число повторов		Локализация	Область	Мотив	Число повторов	
			<i>A. sativum</i>	<i>A. cepa</i>				<i>A. sativum</i>	<i>A. cepa</i>
<i>atpH-atpI</i>	Межгенная	AT	5	4	<i>psaC-ndhE</i>	Межгенная	TA	6	–
<i>rpoC2</i>	Ген	AT	5	5	<i>trnK</i>	Интрон	AATT	3	–
<i>trnT-psbD</i>	Межгенная	AT	6	6	<i>trnK-rps16</i>	Межгенная	ATAA	3	3
<i>trnS-psbZ</i>	»	TA	6	6	<i>ycf3</i>	Интрон	TTTC	3	3
<i>rps4-trnT</i>	»	TA	5	5	<i>ndhF-rpl32</i>	Межгенная	ATTG	–	3
<i>rps4-trnT</i>	»	TA	6	5	<i>cemA</i>	Ген	AATG	3	3
<i>trnT-trnL</i>	»	AT	8	5	<i>petA-psbJ</i>	Межгенная	TAAA	3	3
<i>ycf4-cemA</i>	»	AT	5	5	<i>clpP</i>	Интрон	TAAA	3	3
<i>cemA</i>	Ген	TC	5	5	<i>rpoC2-rpoC1</i>	Межгенная	ATTT	3	3
<i>psaJ-rpl33</i>	Межгенная	AT	7	8	<i>atpF</i>	Интрон	AATA	3	2
<i>rps18-rpl20</i>	»	TA	8	5	<i>petB-petD</i>	Межгенная	GGAT	3	3
<i>ycf2</i>	Ген	GA	5	5	<i>trnT-trnL</i>	»	TTTTA	3	–
<i>ndhF-rpl32</i>	Межгенная	AT	9	–	<i>accD-psaI</i>	»	TATAA	2	3
<i>rpl32-trnL</i>	»	TA	5	–	<i>trnK</i>	Интрон	TAAAA	–	3
<i>rpl32-trnL</i>	»	TA	5	–	<i>petN-psbM</i>	Межгенная	ATTGA	3	3
<i>psaC-ndhE</i>	»	TA	7	5	<i>rpl16-rps3</i>	»	TTTAT	–	3

(Roy et al., 2010). Можно предположить, что гены рибосомальных белков *rps16* (у *A. sativum*), *rpl22* (у *A. cepa*), *rps2* и ген *infA*, кодирующий фактор инициации трансляции, были перенесены в ядерный геном, так как без этих белков был бы невозможен процесс трансляции в пластидах. Что касается гена *ycf15*, то, во-первых, достоверно не известно о его функциях в пластидах, во-вторых, не известно его происхождение, так как ортологи этого гена не идентифицированы у эубактерий (Shi et al., 2013).

В пластидном геноме *A. sativum* в сравнении с пластоном *A. cepa* было обнаружено семь протяженных делеций, локализованных в межгенных спейсерах *trnC-petN* (221 п. н.), *rps18-rpl20* (62 п. н.), *rpoB-trnC* (56 п. н.), *cemA-petA* (53 п. н.) и *trnT-trnL* (три делеции 38, 23 и 18 п. н.).

Анализ нуклеотидных последовательностей границ инвертированных повторов в пластидных геномах *A. sativum* и *A. cepa* выявил различия в размерах фланкирующих спейсеров (рис. 2). Так, размер спейсера между *ycf1* и *ndhF* у *A. sativum* составляет 52 п. н., а у *A. cepa* – всего 2 п. н. У *A. sativum* и *A. cepa* различаются также длины спейсера LSC-IRb. Следует отметить, что в пластомах *A. sativum* и *A. cepa* гены *rps19* и *trnH* так же, как и у другого представителя порядка Asparagales, *Asparagus officinalis* (Steele

et al., 2012), полностью расположены в инвертированных повторах. В то же время у представителей других родов, например *Solanum*, в IRs находится только фрагмент гена *rps19*. Однако присутствие генов *rps19* и *trnH* в IRs не специфично для видов Asparagales, а также встречается у других порядков однодольных, например у Poales (Redwan et al., 2015).

Поиск микросателлитных последовательностей в пластомах *A. sativum* и *A. cepa* выявил всего 32 SSRs, из которых 17 были динуклеотидными (табл. 3). Интересно отметить отсутствие тринуклеотидных микросателлитов. Сравнимые пластоны отличались числом повторов некоторых микросателлитов, при этом для девяти микросателлитов было показано их отсутствие у одного из исследуемых видов. Большая часть SSRs была локализована в межгенных спейсерах, в то время как в генах и интронах обнаружены только четыре и пять микросателлитов соответственно. У *A. sativum* в межгенных спейсерах *ndhF-rpl32* и *rpl32-trnL*, локализованных в SSC, было идентифицировано три микросателлита ((AT)₉ и два повтора (TA)₅), которые полностью отсутствуют у *A. cepa*. Нужно отметить, что спейсер *rpl32-trnL* высоко вариабельный у разных групп растений (Shaw et al., 2007), в связи с этим

он наиболее часто используется для филогенетических исследований, в том числе и для рода *Allium* (Wheeler et al., 2013; Herden et al., 2016; Sinitsyna et al., 2016).

Таким образом, впервые определена и охарактеризована полная нуклеотидная последовательность пластидного генома *A. sativum* и проведено ее сравнение с пластомом родственного вида *A. cepa*. Полученную последовательность пластома *A. sativum* можно использовать в дальнейшем для филогенетических исследований рода *Allium*.

Благодарности

Работа выполнена при финансовой поддержке Федерального агентства научных организаций (ФАНО России).

Конфликт интересов

Авторы заявляют об отсутствии конфликта интересов.

Список литературы / References

- Синявская М.Г., Даниленко Н.Г., Луханина Н.В., Шимкевич А.М., Давыденко О.Г. Экспрессия хлоропластного генома: современные представления и экспериментальные пути изучения. Вавиловский журнал генетики и селекции. 2015;19(5):511-528. DOI 10.18699/VJ15.068. [Siniauskaya M.G., Danilenko N.G., Lukhanina N.V., Shymkevich A.M., Davydenko O.G. Expression of the chloroplast genome: modern concepts and experimental approaches. Russ. J. Genet.: Appl. Res. 2016;6(5):491-509. DOI 10.1134/S2079059716050117.
- Bankevich A., Nurk S., Antipov D., Gurevich A.A., Dvorkin M., Kulikov A.S., Lesin V.M., Nikolenko S.I., Pham S., Pribelski A.D., Pyshkin A.V., Sirotkin A.V., Vyahhi N., Tesler G., Alekseyev M.A., Pevzner P.A. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 2012;19(5):455-477. DOI 10.1089/cmb.2012.0021.
- Ballot S., Renner S.S. The plastomes of two species in the endoparasite genus *Ptilostyles* (Apodanthaceae) each retain just five or six possibly functional genes. Genome Biol. Evol. 2016;8(1):189-201. DOI 10.1093/gbe/evv251.
- Brandvain Y., Wade M.J. The functional transfer of genes from the mitochondria to the nucleus: the effects of selection, mutation, population size and rate of self-fertilization. Genetics. 2009;182(4):1129-1139. DOI 10.1534/genetics.108.100024.
- Chung S.M., Gordon V.S., Staub J.E. Sequencing cucumber (*Cucumis sativus* L.) chloroplast genomes identifies differences between chilling-tolerant and -susceptible cucumber lines. Genome. 2007;50:215-225.
- Daniell H., Lin C.-S., Yu M., Chang W.-J. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. 2016;17(1):134. DOI 10.1186/s13059-016-1004-2.
- Dong W., Xu C., Cheng T., Lin K., Zhou S. Sequencing angiosperm plastid genomes made easy: a complete set of universal primers and a case study on the phylogeny of saxifragales. Genome Biol. Evol. 2013;5(5):989-997. DOI 10.1093/gbe/evt063.
- Doyle J.J., Doyle J.L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. 1987;19:11-15.
- Filyushin M.A., Beletsky A.V., Mazur A.M., Kochieva E.Z. The complete plastid genome sequence of garlic *Allium sativum* L. Mitochondrial DNA Part B: Resources. 2016;1(1):831-832. DOI 10.1080/23802359.2016.1247669.
- Filyushin M.A., Beletsky A.V., Mazur A.M., Kochieva E.Z. Characterization of the complete plastid genome of lop-sided onion *Allium obliquum* L. (Amaryllidaceae). Mitochondrial DNA Part B: Resources. 2018;3(1):393-394. DOI 10.1080/23802359.2018.1456369.
- Friesen N., Fritsch R.M., Blattner F.R. Phylogeny and new intrageneric classification of *Allium* (Alliaceae) based on nuclear ribosomal DNA ITS sequences. Aliso. 2006;22:372-395.
- Goremykin V.V., Hirsch-Ernst K.I., Wolf S., Hellwig F.H. The chloroplast genome of *Nymphaea alba*: whole-genome analyses and the problem of identifying the most basal angiosperm. Mol. Biol. Evol. 2004;21:1445-1454.
- Herden T., Hanelt P., Friesen N. Phylogeny of *Allium* L. subgenus *Anguinum* (G. Don. ex W.D.J. Koch) N. Friesen (Amaryllidaceae). Mol. Phylogenet. Evol. 2016;95:79-93. DOI 10.1016/j.ympev.2015.11.004.
- Jones M.G., Hughes J., Tregova A., Milne J., Tomsett A.B., Collin H.A. Biosynthesis of the flavour precursors of onion and garlic. J. Exp. Bot. 2004;55(404):1903-1918.
- Keller J., Rousseau-Gueutin M., Martin G.E., Morice J., Boutte J., Coissac E., Ourari M., Ainouche M., Salmon A., Cabello-Hurtado F., Ainouche A. The evolutionary fate of the chloroplast and nuclear rps16 genes as revealed through the sequencing and comparative analyses of four novel legume chloroplast genomes from *Lupinus*. DNA Res. 2017;24(4):343-358. DOI 10.1093/dnares/dsx006.
- Kim S., Park J.Y., Yang T. Comparative analysis of the complete chloroplast genome sequences of a normal male-fertile cytoplasm and two different cytoplasmic conferring cytoplasmic male sterility in onion (*Allium cepa* L.). J. Hort. Sci. Biotechnol. 2015;90(4):459-468. DOI 10.1080/14620316.2015.11513210.
- Kumar S., Stecher G., Tamura K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 2016;33(7):1870-1874. DOI 10.1093/molbev/msw054.
- Lohse M., Drechsel O., Kahlau S., Bock R. OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic Acids Res. 2013;41:575-581. DOI 10.1093/nar/gkt289.
- Lowe T.M., Eddy S.R. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25:955-964.
- Mardanov A.V., Ravin N.V., Kuznetsov B.B., Samigullin T.H., Antonov A.S., Kolganova T.V., Skyabin K.G. Complete sequence of the duckweed (*Lemna minor*) chloroplast genome: structural organization and phylogenetic relationships to other angiosperms. J. Mol. Evol. 2008;66(6):555-564. DOI 10.1007/s00239-008-9091-7.
- Ravin N.V., Gruzdev E.V., Beletsky A.V., Mazur A.M., Prokhorchouk E.B., Filyushin M.A., Kochieva E.Z., Kadnikov V.V., Mardanov A.V., Skryabin K.G. The loss of photosynthetic pathways in the plastid and nuclear genomes of the non-photosynthetic myco-heterotrophic eudicot *Monotropa hypopitys*. BMC Plant Biol. 2016;16(Suppl. 3):153-161. DOI 10.1186/s12870-016-0929-7.
- Redwan R.M., Saidin A., Kumar S.V. Complete chloroplast genome sequence of MD-2 pineapple and its comparative analysis among nine other plants from the subclass Commelinidae. BMC Plant Biol. 2015;15:196. DOI 10.1186/s12870-015-0587-1.
- Ricroch A., Yockteng R., Brown S.C., Nadot S. Evolution of genome size across some cultivated *Allium* species. Genome. 2005;48(3):511-520.
- Roy S., Ueda M., Kadowaki K., Tsutsumi N. Different status of the gene for ribosomal protein S16 in the chloroplast genome during evolution of the genus *Arabidopsis* and closely related species. Genes Genet. Syst. 2010;85(5):319-326.
- Seregin A.P., Anačkov G., Friesen N. Molecular and morphological revision of the *Allium saxatile* group (Amaryllidaceae): geographical isolation as the driving force of underestimated speciation. Bot. J. Linn. Soc. 2015;178(1):67-101. DOI 10.1111/boj.12269.
- Shaw J., Lickey E.B., Schilling E.E., Small R.L. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. Am. J. Bot. 2007;94(3):275-288. DOI 10.3732/ajb.94.3.275.
- Shi C., Liu Y., Huang H., Xia E.-H., Zhang H.-B., Gao L.-Z. Contradiction between plastid gene transcription and function due to complex posttranscriptional splicing: an exemplary study of *yef15* function and evolution in angiosperms. PLoS ONE. 2013;8(3):e59620. DOI 10.1371/journal.pone.0059620.
- Sinitsyna T.A., Herden T., Friese N. Dated phylogeny and biogeography of the Eurasian *Allium* section *Rhizirideum* (Amaryllidaceae).

- Plant Syst. Evol. 2016;302:1311-1328. DOI 10.1007/s00606-016-1333-3.
- Sloan D.B. Using plants to elucidate the mechanisms of cytonuclear co-evolution. *New Phytol.* 2015;205(3):1040-1046. DOI 10.1111/nph.12835.
- Steele P.R., Hertweck K.L., Mayfield D., McKain M.R., Leebens-Mack J., Pires J.C. Quality and quantity of data recovered from massively parallel sequencing: Examples in Asparagales and Poaceae. *Am. J. Bot.* 2012;99(2):330-348. DOI 10.3732/ajb.1100491.
- Sveinsson S., Cronk Q. Evolutionary origin of highly repetitive plastid genomes within the clover genus (*Trifolium*). *BMC Evol. Biol.* 2014;14:228. DOI 10.1186/s12862-014-0228-6.
- Temnykh S., DeClerck G., Lukashova A., Lipovich L., Cartinour S., McCouch S. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. *Genome Res.* 2001; 11(8):1441-1452.
- von Kohn C.M., Kielkowska A., Havey M.J. Sequencing and annotation of the chloroplast DNAs of normal (N) male-fertile and male-sterile (S) cytoplasms of onion and single nucleotide polymorphisms distinguishing these cytoplasms. *Genome.* 2013;56(12):737-742. DOI 10.1139/gen-2013-0182.
- Wheeler E.J., Mashayekhi S., McNeal D.W., Columbus J.T., Pires J.C. Molecular systematics of *Allium* subgenus *Amerallium* (Amaryllidaceae) in North America. *Am. J. Bot.* 2013;100(4):701-711. DOI 10.3732/ajb.1200641.
- Wick R.R., Schultz M.B., Zobel J., Holt K.E. Bandage: interactive visualisation of *de novo* genome assemblies. *Bioinformatics.* 2015; 31(20):3350-3352. DOI 10.1093/bioinformatics/btv383.
- Wicke S., Schneeweiss G.M., dePamphilis C.W., Müller K.F., Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol. Biol.* 2011; 76(3-5):273-297. DOI 10.1007/s11103-011-9762-4.