

УДК 57.088+577.29+582.572.4

## Сравнительный анализ последовательностей митохондриальных геномов видов луков *Allium cepa* и *Allium schoenoprasum*

### Comparative analysis of the mitochondrial genomes for *Allium cepa* and *Allium schoenoprasum*

Беленикин М. С.<sup>1</sup>, Криницына А. А.<sup>2</sup>, Купцов С. В.<sup>2</sup>, Логачева М. Д.<sup>2</sup>, Сперанская А. С.<sup>2</sup>

Belenikin M. S.<sup>1</sup>, Krinitsina A. A.<sup>2</sup>, Kuptsov S. V.<sup>2</sup>, Logacheva M. D.<sup>2</sup>, Speranskaya A. S.<sup>2</sup>

<sup>1</sup> Московский физико-технический институт, Институтский переулок, 9, Долгопрудный, 141701, Россия  
E-mail: molecular.modeler@gmail.com

<sup>1</sup> Moscow Institute of Physics and Technology, Institutskiy per., 9, Dolgoprudny, 141701, Russia

<sup>2</sup> Московский государственный университет им. М.В. Ломоносова, Ленинские горы, 1, Москва, 119234  
E-mails: krinitsina@mail.ru, het\_mastin@rambler.ru, maria.log@gmail.com, hanna.s.939@gmail.com

<sup>2</sup> Lomonosov Moscow State University, Leninskiye gory, 1, Moscow, 119234, Russia

**Реферат.** Изучение последовательностей митохондриальных (мт) геномов растений наряду с хлоропластными геномами расширяет возможности филогенетического анализа и позволяет изучать эволюционные события, в том числе эволюционные процессы миграции генов. В данной работе проведено выделение мтДНК, высокопроизводительное секвенирование, *de novo* сборка чернового варианта мт-генома *Allium schoenoprasum*. Проведенный сравнительный анализ белок-кодирующих генов *A. cepa* и *A. schoenoprasum* продемонстрировал высокое сходство генов и позволил выделить лишь небольшое число нуклеотидных замен, приведших к изменению белковых последовательностей. Наиболее существенное из отличий в белок-кодирующих областях найдены в гене *ccmF<sub>NI</sub>*, продукты экспрессии которого имеют отличия 6 подряд идущих аминокислотных остатков.

**Summary.** Studying of mitochondrial genomes together with chloroplast genomes expands the possibilities of phylogenetic analysis allowing studying evolutionary events, including the evolutionary processes of gene migration. Here, in this study the mitochondrial DNA (mtDNA) extraction, next-generation sequencing and draft mitochondrial genome assembly have been made for *Allium schoenoprasum*. The comparative analysis of the protein-coding genes of *A. cepa* and *A. schoenoprasum* has shown high similarity of the gene sequences, giving only a few nucleotide substitutions resulted to a change in the protein sequence. The most significant difference in protein-coding regions is found at the *ccmF<sub>NI</sub>* gene, the expression products of which have 6 consecutive amino acid residues.

#### Введение

Род Лук – *Allium* L. (Alliaceae) насчитывает около 780 видов и является одним из крупнейших родов мировой флоры (Friesen et al., 2006). Среди видов *Allium* много экономически важных растений (ежегодное общемировое производство только репчатого лука и чеснока оценивается в десятки миллионов тонн), что объясняет практическую важность их исследований.

*Allium schoenoprasum* L. – широко известное культивируемое пищевое и декоративное растение, которое в диком виде встречается практически по всей северной Евразии от Британских о-вов до Камчатки, а также в Северной Америке на лугах, в долинах рек, реже на каменистых склонах (Введенский, 1935). Благодаря способности выдерживать низкие температуры растет в Арктических регионах до 70° N. Также этот вид растет в горных районах, например, в Северной Индии, в субальпийском и альпийском поясе (Fritsch, Friesen, 2002; Terpin, Dakskobler, 2012), при этом встречается на больших высотах (более 3300 м. над ур. м.) (Tuncer et al., 2016).

Изучение последовательностей митохондриальных геномов растений наряду с хлоропластными геномами расширяет возможности филогенетического анализа и обеспечивает более детальное изучение эволюционных событий, в том числе процессы миграции генов. Процессы миграции генов обычно протекают из хлоропластного генома в митохондриальный или ядерный, реже – в обратном направлении. В настоящее время информация о строении митохондриального генома растений рода

*Allium* ограничена единственным видом *Allium cepa* L. (KU318712) (Kim et al., 2016), что объясняется сложностью получения препаратов, обогащенных митохондриальной ДНК (секвенирование тотальной ДНК слишком дорого из-за большого размера генома луков), а также сборки (невозможна референсная сборка). В настоящей работе были проведены *de novo* сборка черного варианта митохондриального генома *A. schoenoprasum* и сравнительный анализ белок-кодирующих генов митохондриальных геномов *A. cepa* и *A. schoenoprasum*.

### Материалы и методы

Для исследования митохондриального генома *A. schoenoprasum* нами были использованы растения из коллекции Ботанического сада МГУ им. М. В. Ломоносова. Библиотеки для высокопроизводительного секвенирования были приготовлены из препаратов нуклеиновых кислот, обогащенных низкомолекулярной фракцией ДНК. Протокол выделения ДНК был рассчитан, в первую очередь, на выделение хлоропластной ДНК и описан нами ранее (Беленикин и др., 2016).

Библиотеки для высокопроизводительного секвенирования были приготовлены по стандартному протоколу Nextera (Illumina, США). Нами было проведено секвенирование двух независимых библиотек, различавшихся длиной вставки (~350 п.о. и ~450 п.о.). Амплификация библиотек проводилась с использованием высокоточной полимеразы KAPA HiFi DNA Polymerase (Kapa Biosystems). Секвенирование проводилось на приборе MiSeq (Illumina) на базе лаборатории эволюционной геномики факультета Биоинформатики и Биоинженерии МГУ им. М. В. Ломоносова.

Полученные в результате секвенирования данные после первичной обработки и фильтрации по качеству подвергались процедуре *de novo* сборки, аналогично проведенной для хлоропластных геномов (Беленикин и др., 2016). На следующем этапе проводилось сравнение собранных контигов с белок-кодирующими генами *A. cepa* с использованием BLAST (Altschul et al., 1990).

### Результаты и обсуждение

В результате проведенной *de novo* сборки митохондриального генома *A. schoenoprasum* был получен набор контигов, размер наибольшего из них составил более 110 тыс. п.о. Размер митохондриального генома *A. cepa* составляет 316,3 тысяч п.о. (Kim et al., 2016). По приблизительным оценкам получена информация об около 90 % митохондриального генома *A. schoenoprasum*. На первом этапе исследования был произведен поиск наиболее существенных различий в генах, согласно аннотации *A. cepa*. Результаты сравнения приведены в таблице. Всего 17 из 51 генов митохондриального генома *A. schoenoprasum* несут замены по сравнению с *A. cepa*.

Среди выявленных отличий следует отметить транс-сплайсинг гена *cox2*, который ранее был выявлен также в митохондриальном геноме *A. cepa* (Kim et al., 2016). Другой особенностью митохондриального генома *A. schoenoprasum* является разделение гена, кодирующего ген  $ccmF_N$  на два гена  $ccmF_{N1}$  и  $ccmF_{N2}$ , которые разделены вставкой. Ген  $ccmF_N$  кодирует белок, участвующий в процессе созревания цитохрома C (Thony-Meyer, 2002). Ген  $ccmF_N$  входит в состав митохондриальных геномов всех исследованных высших растений, однако вышеописанное эволюционное событие (расщепление гена  $ccmF_N$  на два) ранее было установлено только у видов, относящихся к семейству крестоцветных, и *A. cepa* (Rayapuram et al., 2008; Kim et al., 2016). Последовательности гена  $ccmF_{N1}$  *A. schoenoprasum* отличаются от *A. cepa*. В частности, в *A. schoenoprasum* имеется несколько несинонимичных однонуклеотидных замен, приводящих к изменению шести последовательно расположенных аминокислотных остатков в белке, кодируемом этим геном. Функциональная значимость этого участка остается невыясненной.

Наиболее существенным отличием некодирующих областей следует назвать наличие в одном из интронов гена *nad4* шести повторов TATTAG у *A. schoenoprasum* по сравнению с двумя такими повторами у *A. cepa*.

Известно, что митохондриальные геномы отличаются своей вариабельностью; однако количественная оценка вариабельности может быть произведена только при наличии хотя бы нескольких близкородственных геномов. В настоящее время для рода *Allium* митохондриальный геном собран лишь для одного представителя, *A. cepa*, подвид *Cepa* (Mill.) Radic, секция *Cepa* (Mill.) Prokh. Наши усилия направлены на завершение сборки полной последовательности митохондриального генома

близкородственного *A. schoenoprasum*, который относится к тому же подроду, но к другой секции *Schoenoprasum* Dumort. При этом полученные в настоящее время данные черновой сборки митохондриального генома *A. schoenoprasum* позволяют предполагать наличие крупных перестроек в митохондриальных геномах этих видов. Так, сравнение собранного нами фрагмента протяженностью около 110 тыс. п.н.о. с последовательностью *A. sepa* [KU318712] показало, что самый протяженный совпадающий участок имеет размер только около 18500 п.о. Участки, расположенные последовательно друг за другом в данном фрагменте *A. schoenoprasum*, соотносятся с участками *A. sepa*, распределенными по всему митохондриальному геному. Также необходимо отметить наличие протяженных несопадающих участков (до 7 тыс. п.о.).

Таблица

Сравнение последовательностей генов митохондриальных геномов луков *A. sepa* и *A. schoenoprasum*

Начало гена*	Конец гена*	Обозначение*	Общее число вариантов/число вариантов в белок-кодирующих областях
1082	3376	<i>nad5</i>	полное совпадение
12513	12900	<i>nad1</i>	полное совпадение
18617	18685	<i>trnM-CAU</i>	1
18764	19120	<i>nad3</i>	полное совпадение
19166	19543	<i>rps12</i>	полное совпадение
21417	28275	<i>nad7</i>	3/0
31139	34730	<i>nad2</i>	2/0
35286	35368	<i>trnY-GUA</i>	полное совпадение
35789	35852	<i>trnR-GCG</i>	полное совпадение
45960	47705	<i>nad1</i>	1/0
51272	52795	<i>atp1</i>	1/0
58436	58490	<i>trnV-GAC</i>	полное совпадение
63268	63525	<i>nad1</i>	полное совпадение
64084	66084	<i>matR</i>	2/2
66290	68717	<i>ccmFc</i>	3/3
75064	75135	<i>trnE-UUC</i>	полное совпадение
75188	77365	<i>orf725</i>	3/3
82415	83221	<i>atp6</i>	полное совпадение
85797	85859	<i>nad1</i>	полное совпадение
91121	93159	<i>nad2</i>	5/1
103573	103842	<i>atp9</i>	4/4
112812	115888	<i>A7Q16_gr03</i>	полное совпадение
126268	126732	<i>atp8</i>	полное совпадение
133535	133650	<i>A7Q16_gr02</i>	полное совпадение
134707	135456	<i>mttB</i>	полное совпадение
140525	140596	<i>trnQ-UUG</i>	полное совпадение
148328	149374	<i>ccmF<sub>N1</sub></i>	2/2
154350	154369	<i>nad5</i>	полное совпадение
155529	155602	<i>trnM-CAU</i>	полное совпадение
159945	168885	<i>nad4</i>	9/1
180834	181802	<i>nad6</i>	1/1
191422	191830	<i>cox2</i>	полное совпадение
201922	201977	<i>trnC-ACA</i>	полное совпадение
203377	203450	<i>trnM-CAU</i>	полное совпадение
205033	205530	<i>nad9</i>	полное совпадение
205540	205600	<i>trnY-AUA</i>	полное совпадение
210250	212160	<i>A7Q16_gr01</i>	3
213897	214712	<i>ccmC</i>	полное совпадение
238167	238976	<i>cox3</i>	полное совпадение
239506	239900	<i>cox2</i>	полное совпадение

Окончание таблицы

246284	246586	<i>nad4L</i>	полное совпадение
246733	247296	<i>atp4</i>	2/2
256454	256524	<i>trnM-CAU</i>	1
258012	258063	<i>trnH-AUG</i>	полное совпадение
261178	261250	<i>trnK-UUU</i>	полное совпадение
265211	265267	<i>trnL-UAA</i>	1
276019	276089	<i>trnE-CUC</i>	полное совпадение
278555	279175	<i>ccmB</i>	полное совпадение
287429	288037	<i>ccmF<sub>N2</sub></i>	полное совпадение
290152	291836	<i>nad5</i>	полное совпадение
301112	302302	<i>Cob</i>	полное совпадение

Примеч.: \* Обозначения генов и координаты даны в соответствии с аннотацией митохондриального генома *A. cepa* (NC\_030100). Гены отсортированы в порядке возрастания координаты начала участка. Для мультиэкзонных генов с подряд следующими экзонами приведены начало и окончание всего гена; для мультиэкзонных генов с экзонами, разделенными другими генами, координаты экзонов приведены отдельно для каждого участка. «Полное совпадение» означает совпадение нуклеотидных последовательностей *A. cepa* и *A. schoenoprasum*, при наличии различий они показаны цифрами, разделенными косой чертой – цифра до черты обозначает общее число вариантов, цифра после черты – число вариантов в белок-кодирующей части гена.

**Благодарности.** Работа выполнена при поддержке грантов РФФИ № 14-04-01852а (за исключением работ по секвенированию) и РФФИ № 14-50-00029 (проведение высокопроизводительного секвенирования).

#### ЛИТЕРАТУРА

- Беленикин М. С., Криницына А. А., Логачева М. Д., Купцов С. В., Сперанская А. С. Секвенирование *de novo* и сравнительный анализ хлоропластных геномов четырех видов рода *Allium*, произрастающих в условиях высокогорий или на равнинах // Проблемы ботаники Южной Сибири и Монголии: сборник науч. ст. по материалам XV междунар. науч.-практ. конф. (23–26 мая 2016 г., Барнаул). – Барнаул: Концепт, 2016. – 524 с.
- Введенский А. И. Род лук – *Allium* // Флора СССР / гл. ред. Комаров В. Л. – М.: Изд-во Академии наук СССР, 1935. – Т. 4. – 760 с.
- Altschul S. F., Gish W., Miller W., Myers E. W. & Lipman D. J. Basic local alignment search tool. // J. Mol. Biol., 1990. – Vol. 215 – P. 403–410.
- 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. – Vol. 22. – P. 372–395.
- Fritsch R. M., Friesen N. Evolution, Domestication and Taxonomy // *Allium* Crop Science: Recent Advances. Eds. H. D. Rabinowitch, L. Currah. – CAB International, Wallingford, U. K., 2002. – P. 5–30.
- Kim B., Kim K., Yang T., Kim S. Completion of the mitochondrial genome sequence of onion (*Allium cepa* L.) containing the CMS-S male-sterile cytoplasm and identification of an independent event of the *ccmF<sub>N</sub>* gene split // Curr Genet., 2016. – Vol. 62, № 4. – P. 873–885.
- Rayapuram N., Hagenmuller J., Grienenberger J. M., Bonnard G., Giege P. J. The three mitochondrial encoded CcmF proteins form a complex that interacts with CCMH and c-type apocytochromes in *Arabidopsis* // Biol. Chem., 2008. – Vol. 283, № 37. – P. 25200–25208.
- Terpin R., Dakskobler I. A new locality of *Allium schoenoprasum* subsp. *alpinum* in the Idrija hills, the first in Slovenia outside the Julian Alps // Folia Biologica et Geologica, 2012. – Vol. 53 – P. 1–2.
- Thony-Meyer L. Cytochrome c maturation: a complex pathway for a simple task? // Biochem. Soc. Trans., 2002. – Vol. 30. – P. 633–638.
- Tuncer B., Firat M., Yarali F., Sarikamis G. Morphology and utilization of *Allium* L. species used as herbs in cheese around Van province in Turkey // Acta Hort., 2016. – Vol. 1143. – P. 171–178.