

УДК 57.088+581.522.4

## Секвенирование *de novo* и сравнительный анализ хлоропластных геномов четырех видов рода *Allium*, произрастающих в условиях высокогорий или на равнинах

### *De novo* sequencing and comparative analysis of the chloroplast genome of four species of the genus *Allium*, growing in various altitude

М. С. Беленикин<sup>1</sup>, А. А. Криницына<sup>1</sup>, М. Д. Логачева<sup>2</sup>, С. В. Купцов<sup>3</sup>, А. С. Сперанская<sup>1</sup>

M. S. Belenikin, A. A. Krinitsina, M. D. Logacheva, C. V. Kuptsov, A. S. Speranskaya

<sup>1</sup>Московский государственный университет им. М.В. Ломоносова, Биологический факультет, 119234, Москва, Ленинские горы, д. 1, стр. 12, e-mail: molecularmodeler@gmail.com, krinitsina@mail.ru, hanna.s.939@gmail.com

<sup>2</sup>Московский государственный университет им. М.В. Ломоносова, факультет Биоинженерии и Биоинформатики, 119234, Москва, Ленинские горы, д. 1, стр. 73, e-mail: maria.log@gmail.com

<sup>3</sup>Московский государственный университет им. М.В. Ломоносова, Ботанический сад, 119899, Москва, Ленинские горы, ул. Менделеева, e-mail: het\_mastin@rambler.ru

**Реферат.** Исследование последовательностей хлоропластных (хп) геномов позволяет получить большое количество разнообразной информации об эволюции и физиологии растений. В данной работе проведено выделение хпДНК, высокопроизводительное секвенирование и сборка хп-геномов для видов рода *Allium*: *A. elatum* Regel., *A. obliquum* L., *A. paradoxum* (M. Bieb.) G. Don. Сравнительный анализ хп-геномов позволил установить различия, отражающие дивергенцию этих видов и *A. cepa* L. Также были найдены особенности, потенциально связанные с эволюционной адаптацией видов к различным условиям произрастания. В двух видах (*A. elatum* и *A. paradoxum*) были найдены инсерции в интроне I гена *ycf3*, в хп-геноме *A. elatum* установлена делеция гена *infA*.

**Summary.** Chloroplast genomes provide various genetic information for evolutionary and functional studies in plants. In this study, the chloroplast DNA (cpDNA) extraction, next-generation sequencing and chloroplast genome assembling were made for *A. elatum* Regel., *A. obliquum* L. and *A. paradoxum* (M. Bieb.) G. Don. Comparative analysis of these and *A. cepa* L. sequences provides insights into the evolutionary events in the *Allium* cpDNA. Comparison of protein-coding regions reveals insertions in the *ycf3* gene intron I for two species (*A. elatum* and *A. paradoxum*) and *infA* gene loss in the cpDNA of *A. elatum*. The potentially ecophysiological role of these genetic traits during the adaptation of *Allium* species to different habitats are discussed.

#### Введение

Род *Allium* L. (Alliaceae) насчитывает около 780 видов и является одним из крупнейших родов мировой флоры (Серегин, 2007). Представители данного рода произрастают в Северном полушарии, среди них довольно большое количество редких и эндемичных видов. Наличие в роде *Allium* большого числа экономически важных растений объясняет практическую важность исследований их диких родичей, в частности, приспособившихся в процессе эволюции к выживанию в неблагоприятных для сельскохозяйственной деятельности условиях, например, подверженных воздействию повышенной дозировки ультрафиолетового излучения и недостаточного увлажнения.

У видов, адаптировавшихся к произрастанию в подобных условиях, в частности, на различных высотных поясах гор, могут возникать адаптационно значимые перестройки хлоропластного генома. На сегодняшний день информации о возникновении нарушений в геномах пластома, в частности, в хлоропластных геномах, под действием условий высокогорья (повышенного уровня ультрафиолета и недостатка влаги) достаточно мало и связана она, в основном, с изучением системы репарации. Сравнительный анализ полных последовательностей хлоропластных геномов (хпДНК), полученных при помощи высокопроизводительного секвенирования, позволит выявить подобные перестройки, в том числе и функционально значимые нуклеотидные замены, которые могут быть использованы как для разработки системы генетических маркеров, например, детекторов состояния окружающей среды, связанных с влиянием на растительный организм ульт-

трафиолетового излучения, так и для отбора перспективных дикорастущих форм для использования в селекции.

До настоящего времени исследование полных последовательностей хпДНК растений рода *Allium* было ограничено видом *A. cepa* L. (von Kohn et al., 2013). В базе данных RefSeq представлено шесть полных последовательностей хпДНК для этого вида, но о последовательностях хпДНК из иных видов этого рода информация ограничена отдельными филогенетически информативными генами или межгенными участками. В настоящей работе было проведен сравнительный анализ последовательностей хпДНК *A. cepa* с хпДНК трех видов луков, отличающихся в природе ареалом и экологическими условиями обитания: *A. elatum* Regel. (syn. *A. macleanii* Baker.), *A. obliquum* L. и *A. paradoxum* (M. Bieb.) G. Don.

*A. elatum* – высокогорный вид, обитающий в безлесных долинах с выраженным сухим летним периодом. Этот вид произрастает в северном Пакистане, Афганистане, юго-восточном Казахстане, горном Туркменистане, южной Киргизии, западном Непале и северо-западной Индии (Кашмир, Уттар-Прадеш), южном Таджикистане (Wendelbo, 1971).

*A. obliquum* обитает на относительно влажных лугово-степных участках и в редколесьях лесостепного типа. Произрастает на западе Монголии, северо-западе Синцзян-Уйгурского АР (Китай), в Восточном Казахстане, Кыргызстане, Средней Азии, Монголии, Закарпатской Украине и Венгрии. Также встречается в южной части Западной Сибири от верхнего течения Иртыша до Енисея, в Новосибирской, Томской, Кемеровской областях, Алтайском крае, республике Алтай, Красноярском крае, в Хакасии, на Урале, юго-западном Предуралье и на юге европейской части России (Jiemei, Kamelin, 2000; Фризен, 1988; Fritsch, Friesen, 2002).

*A. paradoxum* обитает в нижнем ярусе сезонно-влажных листопадных лесов нижнего и среднего горного пояса, в особенности на влажной почве вдоль временных водотоков, избегает открытых мест. Широко распространен в Западной Европе и части Азии (Ricoch et al., 2005).

### Материалы и методы

Для исследования хлоропластных геномов растений рода *Allium* в качестве материала были использованы растения, в настоящее время произрастающие в коллекции Ботанического сада МГУ им. М.В. Ломоносова: *A. elatum* Regel. (syn. *A. macleanii* Baker.), *A. obliquum* L. и *A. paradoxum* (M. Bieb.) G. Don. Последовательности хпДНК *A. cepa* были получены из GenBank (NCBI).

Библиотеки для высокопроизводительного секвенирования были приготовлены из препаратов нуклеиновых кислот, обогащенных фракцией хпДНК. Для этого из свежего листового материала массой 2–20 г было проведено выделение хпДНК с помощью протокола, разработанного нами на основе методов, описанных в статьях Shi et al. (2012) и Do Nascimento Vieira et al. (2014). Для выделения хпДНК: (1) 2–20 г листового материала, предварительно выдержанного при +4 °С в течение 1 недели, гомогенизировали с помощью бытового погружного блендера в охлажденном буфере HSPS-A (pH 3,0–4,0), затем профильтровали через нетканую салфетку и центрифугировали при +4 °С в течение 15 мин, при ускорении 200 g. (2) Полученный осадок ресуспендировали в буфере HSPS-B, объемом 50 мл и повторно центрифугировали при +4 °С в течение 20 мин, при ускорении 3000 g. (3) Полученный осадок мягкой кисточкой ресуспендировали в буфере HSPS-wash объемом 2 мл и затем центрифугировали в градиенте сахарозы в пробирках объемом 15 мл при +4 °С в течение 1 часа, при ускорении 3500 g. (4) Интерфазу, полученную в результате центрифугирования, собирали с помощью пипетки и выделяли из нее ДНК согласно методу (Doyle J.J., Doyle J.L., 1987) с использованием лизирующего ЦТАБ-буфера, с последующей очисткой при помощи хлороформа и преципитацией изопропанолом.

Библиотеки для высокопроизводительного секвенирования были приготовлены по стандартному протоколу Nextera (Illumina, США). Для упрощения процедуры биоинформатической обработки данных для образцов *A. paradoxum* и *A. obliquum* проводилось секвенирование двух независимых библиотек, различавшихся длиной вставки (со средней длиной 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 имени М.В. Ломоносова.

Была проведена сборка последовательностей хлоропластных геномов. В качестве референсной последовательности использовали полные последовательности участков хпДНК *A. cepa* [KM088015.1, KM088014.1, KM088013.1, NC\_024813.1, KF728080.1, KF728079.1]. Референсные последовательности

были использованы в качестве первоначального фильтра для отбора прочтений, относящихся к хлоропластному геному. Далее прочтения, отобранные с помощью фильтра, собирали *de novo* в контиги с помощью ряда сборщиков (Velvet, MIRA). Оставшиеся после фильтрации прочтения также подвергли сборке *de novo* с последующим выявлением фрагментов хлоропластных геномов. Объединение контигов проводили с помощью Newbler, валидировали и проверяли в ручном режиме с помощью картирования собранных контигов на референсный геном *A. sepa*, а также на собранные *de novo* геномы. Сравнительный анализ осуществляли по результатам выравнивания геномов с помощью ClustalX.

### Результаты и обсуждение

Как показали результаты анализа (рис.) препараты, полученные с помощью приведенного протокола, оказались значительно обогащены фракцией хпДНК. Результаты высокопроизводительного секвенирования показали, что 50–60 % прочтений, полученных при секвенировании библиотек, приготовленных из обогащенных препаратов, относились к хпДНК. Для сравнения: секвенирование библиотек, приготовленных из геномной ДНК, выделенной по стандартному протоколу, основанному на применении ЦТАБ-буфера с последующей очисткой хлороформом, показало, что хлоропластным последовательностям относятся лишь 5–10 % прочтений. Использование описанного подхода значительно удешевляет проведение исследования хпДНК методом высокопроизводительного секвенирования.

В результате проведенной биоинформатической обработки данных высокопроизводительного секвенирования была получена информация о полной последовательности хлоропластного генома *A. obliquum*, ее длина составила 153049 п.о. Также получена информация о 90 % последовательностей хлоропластных геномов видов *A. elatum* и *A. paradoxum*.

По сравнению с *A. sepa*, хлоропластный геном *A. elatum* содержит значительное количество инсерций-делетий, которые обнаруживаются, в основном, в межгенных спейсерах, например, *rps16-trnQ* (инс. 602 п.о.), *trnS-trnG* (инс. 283 п.о.), *trnF-ndhJ* (инс. 107 п.о.), *trnE-trnT* (дел. 107 п.о.), *rpl36-infA* (дел. 305 п.о.).

В хп-геноме *A. obliquum* также имеется ряд заметных делеций в межгенных спейсерах, например, *ycf4-cemA* (дел. 148 п.о.), *petB-petD* (дел. 83 п.о.). Однако локализация инсерций-делетий в хп-геномах *A. elatum* и *A. obliquum*, как правило, не совпадает. Эти результаты совпадают с данными систематики, т.к. перечисленные виды относятся к различным под родам: *A. elatum* – подрод *Melanocrommyum* (Webb et Berth.) Rouy, *A. obliquum* – подрод *Polyprason* Radic по классификации N. Friesen с соавт. (2006, цит. по Серегин, 2007).

Кроме различий в межгенных спейсерных участках также обнаружен ряд изменений в последовательностях генов хп-геномов, которые кодируют ряд белков, участвующих в различных процессах, связанных с фотосинтезом. У двух видов, *A. elatum* и *A. paradoxum*, были найдены инсерции в последовательностях интрона I генов *ycf3* (по сравнению с *A. sepa*): инсерция 5 п.о. для *A. elatum* и инсерция 11 п.о. для *A. paradoxum*. Ген *ycf3*, найденный в геномах хлоропластов зеленых водорослей и сосудистых растений, кодирует белок, участвующий в сборке белкового ансамбля фотосистемы I. Этот ген состоит из трех экзонов, разделенных двумя интронами. Первый интрон гена *ycf3* участвует в прохождении сплайсинга и формировании зрелой мРНК. Наличие двух точечных мутаций в первом интроне этого гена (Т→С в позиции 528 и мононуклеотидная инсерция между нуклеотидами в положении 150 и 151) у ячменя приводит к формированию мутантного фенотипа (Landau et al., 2009). При этом фенотипические проявления – формирование светло-зеленой окраски листовой пластинки, наблюдаются только при выращивании в условиях повышенной температуры. Это связано с нарушением образования транскрипта гена

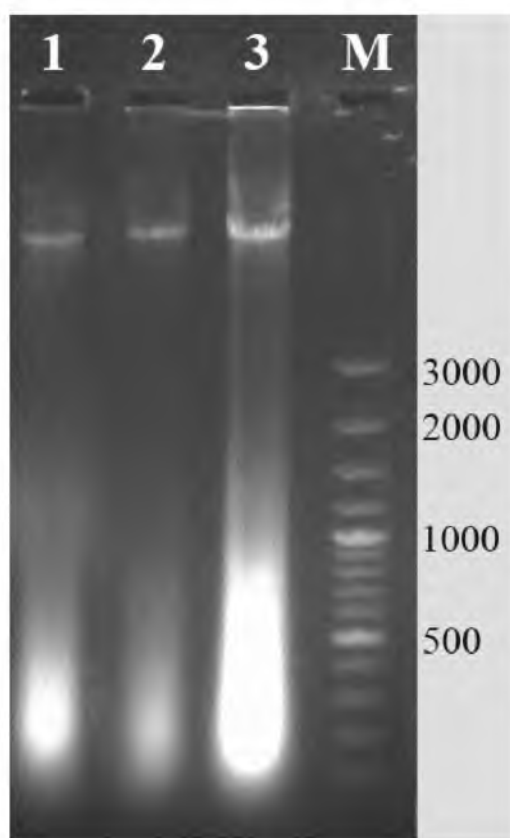


Рис. Электрофореграмма препаратов ДНК, обогащенных фракцией хпДНК. 1 – *A. elatum*, 2 – *A. obliquum*, 3 – *A. paradoxum*, М – маркер молекулярных масс.

*ycf3*. Похожий фенотип также имеют мутанты табака, у которых делеция первого интрона гена *ycf3* в условиях недостаточной освещенности приводит к снижению на 30 % количества хлорофилла в листовой пластинке, в результате чего листья становятся светло-зелеными (Petersen, 2011).

Также в хп-геноме *A. elatum* было обнаружено отсутствие гена *infA*, который кодирует фактор инициации трансляции 1. Большинство изученных видов покрытосеменных содержит в хп-геноме функциональный ген *infA*, хотя в 24 эволюционных линиях он утратил функциональную активность. При этом, как было показано на примере нескольких видов, отсутствие функционирующего гена *infA* в хп-геноме может сопровождаться наличием его экспрессирующейся копии в ядерном геноме (Millen et al., 2001). Авторы ряда работ, выполненных на ячмене, предполагают, что наличие мутации в этом гене может быть ассоциировано с формированием мутантного (хлорофилл-недостаточного) фенотипа (Landau et al., 2007; Landau et al., 2011).

Кроме вышеперечисленного, в хп-геноме *A. paradoxum* был найден ряд существенных структурных отличий, которые локализованы преимущественно в области, размером около 10 000 п.н. Одним из таких отличий является наличие делеций суммарным размером около 3 500 п.н. Однако дальнейшее уточнение положения данных делеций требует дополнительного изучения.

Можно предположить, что наличие некоторых вышеперечисленных делеций может быть связано с особенностями функционирования фотосистем и играть важную роль при адаптации растений к различным экологическим условиям таким, как уровень освещенности, годовые температурные максимумы.

### Благодарности

Работа выполнена при грантовой поддержке РФФИ №14-04-01852а (пробоподготовка и биоинформатический анализ полученных результатов) и РФФИ №14-50-00029 (высоко производительное секвенирование).

### ЛИТЕРАТУРА

- Введенский А. И.** Род лук – *Allium* // Флора СССР / гл. ред. Комаров В. Л. – М.: Изд-во Академии наук СССР, 1935. – Т. 4. – 760 с.
- Сергеев А. П.** Род *Allium* L. (Alliaceae) во флоре восточной Европы: Автореф. дис... канд. биол. наук. – М., 2007. – 27 с.
- Фризен Н. В.** Луковые Сибири: систематика, кариология, хорология. – Новосибирск: Наука, 1988. – 184 с.
- Do Nascimento Vieira L., Faoro H., Fraga H. P., Rogalski M., de Souza E. M., de Oliveira Pedrosa F, Nodari R. O., Guerra M. P** An improved protocol for intact chloroplasts and cpDNA isolation in conifers // PLoS One, 2014. – Vol. 9, No 1. – e84792.
- Doyle J. J., Doyle J. L.** A rapid DNA isolation procedure for small quantities of fresh leaf tissue // Phyt. Bull., 1987. – Vol. 19. – P. 11–15.
- 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 // In: Rabinowitch, H. D. and Currah, L., Eds., *Allium* Crop Science: Recent Advances. – CABI Publishing, Wallingford, U. K., 2002, P. 5–30.
- Jiemei X., Kamelin R. V.** *Allium* Linnaeus // Flora of China. – 2000. – Vol. 24. – P. 165–202. URL.: <http://www.efloras.org>
- Landau A, Pacheco M. G., Prina A. R.** A second *infA* plastid gene point mutation shows a compensatory effect on the expression of the cytoplasmic line 2 (CL2) syndrome in barley // Journal of Heredity, 2011. – Vol. 102, No. 5. – P. 633–639.
- Landau A, Paleo A. D., Civitillo R., Jauregui-alzo M., Prina A. R.** Two *infA* gene mutations independently originated from a mutator genotype in barley // Journal of Heredity, 2007. – Vol. 98, No. 3. – P. 272–276.
- Landau A. M., Lokstein H., Scheller H. V., Lainez V., Maldonado S., Prina A. R.** A cytoplasmically inherited barley mutant is defective in photosystem I assembly due to a temperature-sensitive defect in *ycf3* splicing // Plant Physiol., 2009. – Vol. 151, No. 4. – P. 1802–1811.
- Millen R. S., Olmstead R. G., Adams K. L., Palmer J. D., Lao N. T., Heggie L., Kavanagh T. A., Hibberd J. M., Gray J. C., Morden C. W., Calie P J., Jermiin L. S., Wolfed K. H.** Many Parallel Losses of *infA* from Chloroplast DNA during Angiosperm Evolution with Multiple Independent Transfers to the Nucleus // The Plant Cell, 2001. – Vol. 13. – P. 645–658.
- Petersen K., Schottler M. A., Karcher D., Thiele W., Bock R.** Elimination of a group II intron from a plastid gene causes a mutant phenotype // Nucleic Acids Res., 2011. – Vol. 39, No. 12. – P. 5181–5192.
- Ricroch A., Yockteng R., Brown S. C., Nadot S.** Evolution of genome size across some cultivated *Allium* species // Genome, 2005. – Vol. 48. – P. 511–520.

**Shi C., Hu N., Huang H., Gao J., Zhao Y. J., Gao L. Z.** An improved chloroplast DNA extraction procedure for whole plastid genome sequencing // PLoS One, 2012. – Vol. 7, No. 2. – e31468.

**von Kohn C., Kietkowska A., Havey M. J.** Sequencing and annotation of the chloroplast DNAs and identification of polymorphisms distinguishing normal male-fertile and male-sterile cytoplasms of onion // Genome, 2013. – Vol. 56. – P. 737–742.

**Wendelbo P.** Alliaceae // Flora Iranica / Rechinger K. H. (ed.). Vol. 76. – Austria: Akademische Druck und Verlagsanstalt, Graz, 1971. – 100 s.