

МОЛЕКУЛЯРНО-ГЕНЕТИЧЕСКИЕ МЕТОДЫ ИССЛЕДОВАНИЯ РАСТЕНИЙ И ХЕМОСИСТЕМАТИКА

УДК 582.572.225:577.212.3

Полиморфизм и возможная эволюция индели в интроне I хлоропластного гена *ycf3* в роде *Allium* L.

Polymorphism and possible evolution of an indel in the introne I of the chloroplast gene *ycf3* in genus *Allium* L.

Антипин М. И.¹, Криницына А. А.¹, Беленикин М. С.², Омельченко Д. О.³,
Купцов С. В.¹, Логачева М. Д.¹, Сперанская А. С.^{1,3}

Antipin M. I.¹, Krinitsina A. A.¹, Belenikin M. S.², Omelchenko D. O.³,
Kuptsov S. V.¹, Logacheva M. D.¹, Speranskaya A. S.^{1,3}

¹Московский государственный университет им. М. В. Ломоносова, Москва

¹Lomonosov Moscow State University, Moscow

²Московский физико-технический институт, Долгопрудный, Московская область

²Moscow Institute for Physics and Technology, Dolgoprudny

³ФБУН ЦНИИ Эпидемиологии Роспотребнадзора, Москва

³Central Research Institute of Epidemiology of The Federal Service on Customers' Rights Protection
and Human Well-being Surveillance, Moscow

E-mails: sagefool@yandex.ru, ankrina@gmail.com, molecular.modeler@gmail.com, omdeno@gmail.com,
het_mastin@rambler.ru, maria.log@gmail.com, hanna.s.939@gmail.com

Реферат. К роду *Allium* L. относится несколько древнейших видов культурных растений. Идентификация генетических особенностей, связанных с формированием важных признаков у культурных видов и их дикорастущих родственников важно как для фундаментальной науки, так и для экономики. Мы обнаружили полиморфизм в интроне I хлоропластного гена *ycf3*, кодирующего белок сборки фотосистемы I у видов *Allium*, относящихся к различным эволюционным линиям рода. Определение последовательностей гомологичных участков этого интрона у других шести родов Amaryllidaceae позволило нам выдвинуть гипотезу о возможных путях эволюции этого участка хлоропластного генома.

Summary. Genus *Allium* L. has several important crop plants among its species, some of which were domesticated in prehistoric times. Identification of genetic features, determining important traits of crop plants and their closest wild relatives can have a huge impact both on fundamental studies and economy. We have found a polymorphism in intron I of chloroplast gene *ycf3*, encoding the Photosystem I complex assembling protein, in species of *Allium* belonging to different evolutionary lines. An evolution history hypothesis of this site was suggested, based on new reads of homologous sequences of this intron in 6 other genera of Amaryllidaceae.

Геном хлоропластов кодирует многие ключевые белки растений, участвующие в процессе фотосинтеза и других важных метаболических процессах.

С появлением методов высокопроизводительного секвенирования в генетике и геномике хлоропластов произошел количественный и качественный прорыв. В настоящее время в базе данных Национального центра биотехнологической информации (NCBI) выложено в открытый доступ более 800 полных последовательностей хлоропластных геномов, в том числе более 300 последовательностей хозяйственно ценных растений (Daniell et al., 2016). Эти исследования имеют ключевое значение для понимания биологии и физиологии растений, неопределимы для уточнения и пересмотра представлений о филогенетических связях таксонов больших и малых рангов. Информация о взаимодействиях меж-

ду генотипом и средой, выявляемая благодаря геномике хлоропластов, особенно важна в случае культурных растений. Ближайшие родственники культурных видов составляют генофонд, пригодный для получения новых, высокоэффективных сортов методами селекции, гибридизации, геной инженерии (Wroczynska et al., 2016). Выявление родственных связей внутри таких таксономических групп позволяет пролить свет на происхождение культурных растений и сложные взаимосвязи морфологических, физиологических и генетических изменений, происходивших в процессе их доместификации.

Настоящая работа представляет собой лишь небольшую часть исследований, проводимых нашим коллективом авторов по изучению взаимодействий хлоропластного генома рода *Allium* и экологических параметров среды, имеющих прямое воздействие на физиологию растений (свет, влажность, температура). Виды этого рода, многие из которых имеют огромное экономическое значение, обладают рядом признаков, указывающих на их происхождение из ксерофильных сообществ, адаптированных к высоким температурам, недостатку влаги и значительной доле ультрафиолета в солнечном спектре.

Нами были получены полные последовательности хлоропластных геномов восьми видов рода *Allium* (Беленикин и др., 2016; Сперанская и др., 2017). При построении выравниваний обнаружилась значительная полиморфность многих участков на уровне видов, секций, подродов и так называемых эволюционных линий внутри рода. Особый интерес вызвал участок в первом интроне гена *ycf3*, продукты которого участвуют в сборке белкового ансамбля фотосистемы I. Известно, что первый интрон гена *ycf3* участвует в прохождении сплайсинга и формировании зрелой мРНК; точечные мутации этого участка у ячменя приводят к проявлению мутантного фенотипа в стрессовых (высокая температура) условиях (Landau et al., 2009).

Чтобы иметь возможность делать хотя бы самые осторожные выводы об эволюции этого вариабельного участка в роде *Allium* и предлагать внутренне непротиворечивые гипотезы о возникновении этой индели, нам были необходимы аутгруппы разной степени филогенетической близости. На момент написания данной работы в базе NCBI отсутствовали данные по последовательности гена для каких-либо других представителей Amaryllidaceae, кроме рода *Allium*. Для восполнения этого пробела мы дополнительно предприняли секвенирование начального участка гена *ycf3* у представителей следующих таксономических групп: рода *Tristagma*, по принятой в последнее время систематике входящего в подсемейство луковые внутри семейства амариллисовые, и пяти родов из «сердцевинных» (“core”) амариллисовых, или подсемейства *Amaryllidoideae* (цит. по: APG III – Angiosperm Phylogeny ..., 2009). На основании анализа ряда ядерных, рибосомальных и пластидных маркеров среди амариллисовых в узком смысле принято выделять, среди прочих, американскую ветвь (в нашем исследовании представленную родом *Rhodophiala*), евроазиатскую ветвь (в этой работе род *Pancratium*) и значительно дивергировавшую южноафриканскую ветвь (типовой род *Amaryllis*, а также роды *Gethyllis* и *Boophone*).

Материалы и методы

Образцы листьев шести видов амариллисовых, по одному виду на каждый род, площадью около 1-2 см² были получены из живой коллекции филиала ботанического сада МГУ «Аптекарский огород» (*Amaryllis belladonna* L., *Boophone disticha* (L. f.) Herb., *Gethyllis* aff. *namaquensis* (Schönland) Oberm., *Pancratium maritimum* L., *Rhodophiala ananuca* (Phil.) Traub, *Tristagma bivalve* (Hook. ex Lindl.) Traub). Образцы высушивались в зиплок-пакетах с силикагелем, после полного высыхания растирались индивидуально в ступках до порошкообразного состояния. Выделение ДНК проводилось с использованием СТАВ, по методу, описанному J. J. Doyle и J. L. Doyle (1987) с небольшими изменениями. Для амплификации 5'-региона интрона I гена *ycf3* были использованы праймеры *ycf3*-seq-F (5'-GAGAGTTTCCTACCTCATACGG) и *ycf3*-seq-R (5'-CCACTAAAGTATTGAGCAGCG). Амплифицированные фрагменты ожидаемого размера (~350 п.о.) очищались от реакционной смеси с помощью AMPure Beads (Beckman Coulter). Секвенирование по Сэнгеру проводилось в ЦКП «Геном» (ИМБ РАН).

Также была получена последовательность интрона I гена *ycf3* для *A. semenovii* Regel. из материала, собранного в Киргизии, в Ак-Суйском районе Иссык-Кульской области. Для этого было проведено высокопроизводительное секвенирование геномной библиотеки на MiSeq (Illumina). Подготовка материала к секвенированию проводилась, как описано нами ранее (Беленикин и др., 2016). Последо-

вательность гена *ucf3* была получена с помощью референсной сборки полученных прочтений с помощью программы CLC Genomics Workbench. В качестве референса были использованы последовательности гена *ucf3* *A. obliquum* L.

Помимо этого, в работе были использованы последовательности первого интрона гена *ucf3* для видов *Allium* (*A. obliquum* L., *A. macleanii* Baker. (syn. *A. elatum* Regel), *A. paradoxum* (M. Bieb.) G. Don.), полученные нами ранее в результате секвенирования и сборки полных последовательностей хлоропластных геномов (Беленикин и др., 2016; Сперанская и др., 2017). Также были привлечены данные, доступные из базы данных NCBI, о последовательностях генов *ucf3* для видов *A. sativum*, *A. cepa* и *A. fistulosum*, а также, в качестве аутгруппы, для других семейств порядка Asparagales и более удаленных групп однодольных. Выравнивание последовательностей проводилось алгоритмом Muscle, построение филогенетических деревьев – методом максимального сходства с помощью программы MEGA 7.

Результаты и обсуждение

В полученных нами и взятых из ГенБанка последовательностях на этом участке первого интрона протяженностью одиннадцать пар нуклеотидов у рода *Allium* наблюдаются индели, причем замечательно их распределение у изученных нами представителей трех эволюционных линий рода. У видов первой эволюционной линии участок представлен одиннадцатью нуклеотидами; второй вариант индели, в котором шесть из одиннадцати нуклеотидов отсутствуют, встречается как во второй, так и в третьей линии родства; и, наконец, третий вариант, в котором отсутствует весь одиннадцатинуклеотидный участок, обнаружен только у видов, принадлежащих к третьей эволюционной линии (см. рисунок 1).

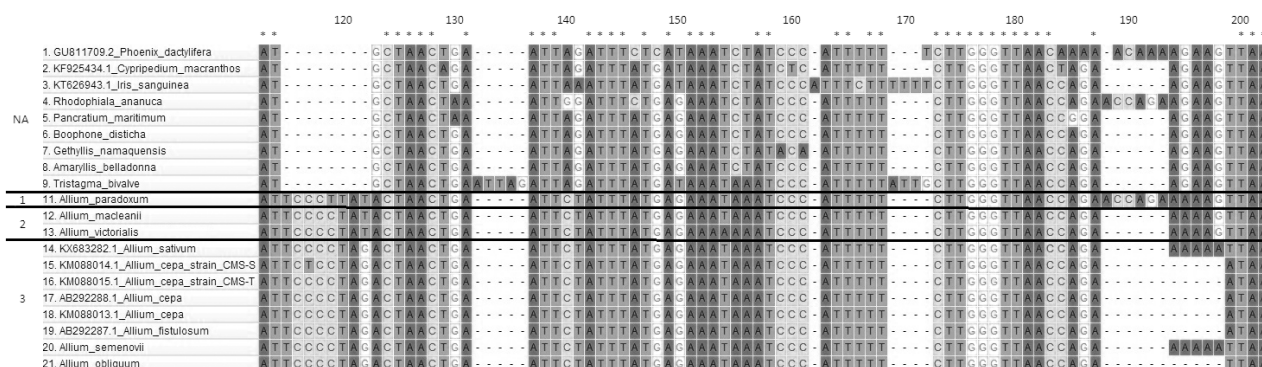


Рис. 1. Выравнивание последовательностей первого интрона гена *ucf3* для родов *Allium*, *Amaryllis*, *Boophone*, *Gethyllis*, *Pancratium*, *Rhodophiala*, *Tristagma*. Аутгруппа – *Phoenix dactylifera* (Arecaceae, Arecales), *Cyripedium macranthos* (Orchidaceae, Asparagales), *Iris sanguinea* (Iridaceae, Asparagales).

Сравнение последовательностей видов рода Лук и других шести родов амариллисовых показало, что на участке, гомологичном участку индели у *Allium*, у пяти из шести дополнительно изученных родов расположены пять нуклеотидов, как во втором варианте индели у луков. Однако если у рода *Allium* это пятикратно повторенный аденин, то у остальных амариллисовых это последовательность AGAAG. Присутствует она и у шестого рода, *Rhodophiala*, однако в этом случае имеется ещё одна вставка: шесть нуклеотидов, из которых пять соответствует нуклеотидам первого варианта индели у луков первой эволюционной линии.

Среди прочих идентифицированных особенностей следует отметить, что (1) только для рода *Allium* характерна вставка в восемь нуклеотидов в начале интрона. (2) У неафриканских амариллисовых имеется вставка из восьми нуклеотидов (позиции 59–66). (3) Делеция из 12 нуклеотидов (сайты 169–180) имеется у всех *Allioideae* (*Allium* + *Tristagma*).

Филогенетическое дерево, построенное методом максимального правдоподобия, в целом хорошо согласуется с филогенетическими деревьями, полученными для семейства амариллисовых в широком и узком смысле (Meerow et al., 1999; Meerow, Snijman, 2001). Виды рода *Allium* оказываются в отдельной кладе: южноафриканская ветвь амариллисовых в узком смысле является сестринской



Рис. 2. Филогенетическое дерево, построенное методом максимального сходства (maximum likelihood) для начальной последовательности гена *5S* с использованием данных из базы NCBI.

по отношению к американскому (*Rhodophiala*) и евроазиатскому (*Pancratium*) родам: подсемейства *Amaryllidoideae* и *Allioideae* – сестринские клады. Если привлечь для построения выравнивания более удаленные таксоны однодольных, становится очевидно, что вариант индели в пять нуклеотидов AGAAG, характерный для нелуковых амариллисовых, очень консервативен и встречается у всех представителей порядка Asparagales. Более того, эти же пять нуклеотидов в этом же положении мы видим и у коммелинид, и у лилейных. Можно выдвинуть предположение, что наличие этой индели в пять нуклеотидов является базальным для однодольных. При этом вариант, характерный для луков несет замену одного или обоих гуанинов на аденин.

Из этого предположения следует, что первый вариант индели у луков по отношению ко второму является инсерцией пяти нуклеотидов, а третий – делецией пяти, возникшей только в определенной части третьей группы родства в роде *Allium*.

Вне рода *Allium* вставка из шести нуклеотидов, отличающаяся от инсерции в первой группе луков только одним дополнительным нуклеотидом, встречается в роде *Rhodophiala*. Поскольку остальные пять нуклеотидов из одиннадцати у этого рода идентичны типу, гипотетически принятому нами за исходный, можно предположить, что эта инсерция возникла в этой ветви амариллисовых в узком смысле независимо от рода *Allium*.

Среди более отдаленных таксонов вставка в этом же положении, также из шести нуклеотидов, характерна для пальм (Arecaceae, Arecales, commelinids). Значительная филогенетическая удаленность этого таксона и совершенно иная последовательность шести нуклеотидов вставки у пальм позволяют сделать осторожные выводы о ее конвергентном, независимом происхождении.

Можно с осторожностью предположить, что задействованный в сплайсинге первый интрон гена *ycf3* у рода *Allium* находится под действием отбора, что объясняет его большую изменчивость внутри этого таксона по сравнению с большинством других родов порядка Asparagales, где этот участок заметно более консервативен.

Благодарности. Настоящая работа была выполнена в рамках Госзадания Московского государственного университета им. М. В. Ломоносова (темы № ААА-А16-116021660105-3 и ААА-А-А16-116021660049-0).

ЛИТЕРАТУРА

- Angiosperm Phylogeny Group III An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III // Botanical Journal of the Linnean Society. 2009. – Vol. 161, № 2. – P. 105–121.
- Brozynska M., Furtado A., Henry R. J. Genomics of crop wild relatives: expanding the gene pool for crop improvement // Plant Biotechnol J. 2016. – Vol. 14 – P.1070–85. DOI: 10.1111/pbi.12454.
- Daniell H., Lin Ch.-S., Yu M., Chang W.-J. Chloroplast genomes: diversity, evolution, and applications in genetic engineering // Genome Biology, 2016. – P.134. DOI: 10.1186/s13059-016-1004-2.
- Doyle J. J., Doyle J. L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue // Phytochemical Bulletin, 1987. – Vol. 19. – P. 11–15.
- 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 Physiology, 2009. – Vol. 151. – P. 1802–1811.
- Meerow A. W., Fay M. F., Charles L., Guy Ch. L., Li Q.-B., Zaman F. Q., Chase M. Systematics of Amaryllidaceae based on cladistic analysis of plastid sequence data // Am. J. Bot., 1999. – Vol. 86. – P. 1325–1345.
- Meerow, A. W., Snijman, D. A. Phylogeny of Amaryllidaceae tribe Amaryllideae based on nrDNA ITS sequences and morphology // Am. J. Bot. December, 2001. – Vol. 88. – P. 2321–2330.
- 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: Материалы XV междунар. науч.-практ. конф. – Барнаул: Изд-во АлтГУ, 2016. – С. 107–111.
- Сперанская А. С., Криницына А. А., Логачева М. Д., Беленикин М. С. Высокопроизводительное секвенирование и сравнительный анализ полных последовательностей хлоропластных геномов нескольких видов рода *Allium* (Alliaceae) // Систематика и эволюционная морфология растений: Материалы конференции, посвященной 85-летию со дня рождения В. Н. Тихомирова / Под ред. К. В. Авилова, А. В. Троицкий. – М.: Изд-во МАКС Пресс, 2017. – С. 369–370.