

Вставки прямых повторов в микроэволюции пластид и митохондрий семенных растений¹

О. А. Зверков*, Л. Ю. Русин^{*,**}, А. В. Селиверстов*, В. А. Любецкий*

*Институт проблем передачи информации им. А. А. Харкевича, Российская академия наук, Москва, Россия

**Московский государственный университет им. М. В. Ломоносова, Москва, Россия

Поступила в редколлегию 23.07.2012

Аннотация—Исследованы вставки точных прямых повторов и распределение длин повторяемых участков на ДНК в пластидах и митохондриях. По очереди рассмотрено несколько семейств семенных растений. Уточнены ранее опубликованные сведения об эволюции некодирующих областей ДНК. Предположено, что в ходе эволюции некодирующих участков ДНК прямые повторы часто возникают одновременно при удвоении участка ДНК. Длина повторяемого участка обычно равна одному или пяти. Распределения по длинам участков близки у пластид и митохондрий.

КЛЮЧЕВЫЕ СЛОВА: прямой повтор, пластида, митохондрия, семенные растения, микроэволюция.

1. ВВЕДЕНИЕ И ПОСТАНОВКА ЗАДАЧИ

Широко рассматривается задача поиска в одном геноме *неточных* повторов некоторого участка ДНК (слова). Часто поиск выполняется с помощью программы REPuter [1]. В [2] эта программа применена к пластоми *Phoenix dactylifera* и в нём найдены 11 прямых и обратных неточных повторов с длиной повторяемого слова 30 п. н. В [3] с помощью REPuter найдены прямые неточные повторы с длиной повторяемого слова от 10 до 100 п. н. в пластомах пяти видов рода *Oenothera*. В [2] рассматриваются неточные повторы в двух видах рода *Silene*. При секвенировании новых пластоми среди первичной информации о них обычно ищутся и неточные повторы слов [4–6]. Можно привести много примеров такого типа результатов, в которых рассматриваются единичные геномы, в частности, пластоми, и повторяемые слова не короткие (иначе возникает огромное перепредсказание), а повторы неточные (часто они имеют 90% точность).

В [8] на основе большого числа недавно секвенированных пластоми рассматривалась эволюция большого инвертированного повтора, что не связано с изучением прямых точных повторов слов, которые, как показывают, в частности, наши результаты, имеют сравнительно небольшую длину. Работа [9] посвящена полиморфизмам в почти совпадающих геномах митохондрий некоторых близких растений родов *Oryza* и *Brassica*; в ней отмечено, что ранее рассматривался и вид *Beta vulgaris*. В нашей заметке сопоставлены более далёкие друг от друга митохондрии видов из больших таксономических групп. Существенное обстоятельство, побуждающее к исследованию *произвольной длины точных повторов* некоторого слова, состоит в том, что модель независимых однонуклеотидных замен и вставок, обсуждаемая, например, в [7], не позволяет объяснить их в случае слова с длиной четыре и более п. н. в некодирующих областях генома. Хотя такие вставки часто встречаются, например, в пластидах. В [7, 10]

¹ Работа выполнена при частичной финансовой поддержке госконтрактами 14.740.11.0624, 14.740.11.1053, НК-421П, 14.740.12.0830 Министерства образования и науки РФ.

приведён перечень основных эволюционных событий в некодирующих областях пластид. И отмечена высокая частота вставок (точных) прямых повторов слов и особенно — многократной вставки одного нуклеотида. Однако в [7, 10] рассматривались только короткие специальные участки геномов, например, интроны первой группы гена *trnL* у asterids. Такие вставки использовались для построения и классификации деревьев видов [11].

Мы рассмотрим задачу поиска и подсчёта числа вставок (точных) прямых повторов любой длины в некодирующих областях, включая интроны, в пластидах и митохондриях видов семенных растений. Подчеркнём, что эта задача не сводится к поиску повторов в отдельных последовательностях и требует существенно *множественного выравнивания последовательностей*, например, ради возможности отличить вставку от серии независимых нуклеотидных замен. Чтобы пояснить разницу между повторами и вставками повторов, мы дадим следующее определение. Вставка прямого повтора наблюдается как множественное выравнивание нуклеотидных последовательностей, в котором имеются две строки: одна (образец) содержит точный прямой повтор вида $\varphi\varphi$ некоторого слова φ , а в другой строке это $\varphi\varphi$ выравнивается на $\varphi-$ или $-\varphi$, где знак “-” означает делецию всех букв в φ ; при этом для образца и каждой последовательности выполняется: $\varphi\varphi$ выравнивается на один из четырёх типов участков $\varphi\varphi$, $\varphi-$, $-\varphi$ или $--$. На рис. 1 приведено иллюстративное выравнивание с повтором слова длины 4 п. н., оно является частью выравнивания на рис. 2. На рис. 2 приведён пример неточного повтора, который произошёл у *Keteleeria davidiana* в результате, вероятно, сначала вставки точного повтора слова длины 11 п. н., а затем замены в одной из его позиций.

```

====ATTCA TTCATCTCT
=====ATTCA TCCT
=====ATTCA TCCT
====ATTCA TTCATCTCT
AAAGATTCA TTCATCTCT

```

Рис. 1. Множественное выравнивание участков 5'-лидерных областей гена *psbM*. Повторяемое слово φ выделено полужирным.

2. МАТЕРИАЛЫ И МЕТОДЫ

В части пластид нами проведён широкомасштабный поиск вставок прямых повторов в следующих семействах, которые охватывают почти все семейства семенных растений, включающих хотя бы два вида с полностью секвенированным пластомом. После названия вида указан номер пластома в базе GenBank.

Acoraceae: *Acorus americanus* NC 010093, *A. calamus* NC 007407.

Asteraceae: *Guizotia abyssinica* NC 010601, *Helianthus annuus* NC 007977, *Lactuca sativa* NC 007578, *Parthenium argentatum* NC 013553.

Brassicaceae: *Aethionema cordifolium* NC 009265, *A. grandiflorum* NC 009266, *Arabidopsis thaliana* NC 000932, *Arabis hirsuta* NC 009268, *Barbarea verna* NC 009269, *Capsella bursa-pastoris* NC 009270, *Crucihimalaya wallichii* NC 009271, *Draba nemorosa* NC 009272, *Lepidium virginicum* NC 009273, *Lobularia maritima* NC 009274, *Nasturtium officinale* NC 009275, *Olimarabidopsis pumila* NC 009267.

Convolvulaceae: *Cuscuta exaltata* NC 009963, *C. gronovii* NC 009765, *C. obtusiflora* NC 009949, *C. reflexa* NC 009766, *Ipomoea purpurea* NC 009808.

Fabaceae: *Cicer arietinum* NC 011163, *Glycine max* NC 007942, *Lathyrus sativus* NC 014063, *Lotus japonicus* NC 002694, *Medicago truncatula* NC 003119, *Phaseolus vulgaris* NC 009259, *Pisum sativum* NC 014057, *Trifolium subterraneum* NC 011828, *Vigna radiata* NC 013843.

Geraniaceae: *Erodium texanum* NC 014569, *Geranium palmatum* NC 014573, *Monsonia speciosa* NC 014582, *Pelargonium x hortorum* NC 008454.

Malvaceae: *Gossypium hirsutum* NC 007944, *G. barbadense* NC 008641.

Myrtaceae: *Eucalyptus globulus* NC 008115, *E. grandis* NC 014570.

Nymphaeaceae: *Nuphar advena* NC 008788, *Nymphaea alba* NC 006050.

Oleaceae: *Jasminum nudiflorum* NC 008407, *Olea europaea* NC 013707.

Onagraceae: *Oenothera elata* NC 002693, *O. argillicola* NC 010358, *O. glazioviana* NC 010360, *O. biennis* NC 010361, *O. parviflora* NC 010362.

Pinaceae: *Cathaya argyrophylla* NC 014589, *Cedrus deodara* NC 014575, *Keteleeria davidiana* NC 011930, *Picea sitchensis* NC 011152, *Pinus thunbergii* NC 001631, *P. koraiensis* NC 004677, *P. contorta* NC 011153, *P. gerardiana* NC 011154, *P. krempfii* NC 011155.

Poaceae (БЕР): *Agrostis stolonifera* NC 008591, *Bambusa oldhamii* NC 012927, *Brachypodium distachyon* NC 011032, *Dendrocalamus latiflorus* NC 013088, *Festuca arundinacea* NC 011713, *Hordeum vulgare* NC 008590, *Lolium perenne* NC 009950, *Oryza nivara* NC 005973, *O. sativa Indica Group* NC 008155, *O. sativa Japonica Group* NC 001320, *Triticum aestivum* NC 002762.

Poaceae (ПАССАД): *Coix lacryma-jobi* NC 013273, *Saccharum hybrid* SP80-3280 NC 005878, *S. officinarum* NC 006084, *Sorghum bicolor* NC 008602, *Zea mays* NC 001666.

Ranunculaceae: *Megaleranthis saniculifolia* NC 012615, *Ranunculus macranthus* NC 008796.

Saliceae: *Populus alba* NC 008235, *P. trichocarpa* NC 009143.

Solanaceae: *Atropa belladonna* NC 004561, *Nicotiana tabacum* NC 001879, *N. sylvestris* NC 007500, *N. tomentosiformis* NC 007602, *Solanum lycopersicum* NC 007898, *S. bulbocastanum* NC 007943, *S. tuberosum* NC 008096.

Также рассмотрен пластом *Cryptomeria japonica* NC 010548 из семейства **Cupressaceae**, который относительно плохо выравнивается с пластами других видов и на рис. 2 играет роль внешней группы.

В части митохондрий вместо семейства Brassicaceae рассмотрены порядки

Brassicales: *Arabidopsis thaliana* NC 001284, *Brassica napus* NC 008285, *Carica papaya* NC 012116;

Caryophyllales: *Beta vulgaris subsp. vulgaris* NC 002511, *B. vulgaris subsp. maritima* NC 015099.

А также рассмотрены митохондрии видов семейства **Poaceae:** *Triticum aestivum* NC 007579, *Sorghum bicolor* NC 008360, *Oryza sativa Indica Group* NC 007886, *O. sativa Japonica Group* NC 011033, *O. rufipogon* NC 013816, *Zea mays subsp. mays* NC 007982, *Z. perennis* NC 008331, *Z. mays subsp. parviglumis* NC 008332, *Z. luxurians* NC 008333, *Tripsacum dactyloides* NC 008362.

Метод исследования состоял в следующем. Для каждой пары видов искались пары позиционно сцепленных генов или экзонов из близких видов, при этом очень короткие неконсервативные гены игнорировались. Затем собранные вместе некодирующие области между такими генами или экзонами из всех видов одного семейства выравнивались, на выравнивании определялись вставки прямых повторов. Множественные выравнивания проводились программой MUSCLE [12]. Также использовались программы, описанные на официальном сайте лаборатории 6 ИППИ РАН <http://lab6.iitp.ru/ru/repeats/>.

3. РЕЗУЛЬТАТЫ

Получены следующие распределения числа вставок прямых повторов в зависимости от длины слова φ в семействах цветковых; после двоеточия используется запись “длина слова φ : число найденных вставок с этой длиной”. В семействе *Ascoraceae*: 1: 9. В семействе *Asteraceae*: 1: 153, 2: 10, 3: 5, 4: 9, 5: 27, 6: 31, 7: 9, 8: 4, 9: 1, 10: 5, 11: 1, 12: 2, 13: 1, 17: 2, 18: 2, 21: 2, 22: 3, 23: 1. В семействе *Brassicaceae*: 1: 1373, 2: 82, 3: 26, 4: 31, 5: 66, 6: 57, 7: 32, 8: 16, 9: 5, 10: 3, 11: 3, 12: 1, 13: 6, 15: 1, 16: 1, 17: 2, 20: 3, 22: 1, 23: 1. В семействе *Fabaceae*: 1: 783, 2: 84, 3: 41, 4: 91, 5: 71, 6: 25, 7: 10, 8: 3, 10: 1, 12: 2, 13: 1, 14: 1, 16: 2. В семействе *Geraniaceae*: 1: 186, 2: 14, 3: 6, 4: 25, 5: 34, 6: 21, 7: 9, 8: 2, 9: 1. В семействе *Convolvulaceae*: 1: 168, 2: 8, 3: 4, 4: 6, 5: 4, 6: 7, 7: 3, 8: 2, 15: 1, 18: 1, 19: 1. В семействе *Malvaceae*: 1: 45, 2: 1, 4: 2, 5: 8, 6: 4, 7: 2, 8: 2, 51: 1. В семействе *Myrtaceae*: 1: 34, 6: 1, 18: 1, 19: 1. В семействе *Nymphaeaceae*: 1: 58, 2: 8, 3: 3, 4: 22, 5: 32, 6: 8, 7: 5, 9: 2, 11: 1, 14: 2, 24: 1. В семействе *Oleaceae*: 1: 46, 2: 4, 3: 3, 4: 5, 5: 6, 6: 3, 7: 1, 8: 2, 11: 1, 12: 1, 14: 1, 17: 1. В семействе *Onagraceae*: 1: 114, 2: 6, 3: 1, 4: 5, 5: 14, 6: 10, 7: 4, 9: 1, 11: 1, 12: 3, 13: 2, 15: 2, 16: 2, 17: 1, 19: 1, 20: 1, 21: 2, 22: 2, 23: 2, 24: 2, 26: 1, 29: 1, 35: 1, 50: 1, 78: 1. В семействе *Pinaceae*: 1: 378, 2: 41, 3: 38, 4: 104, 5: 99, 6: 23, 7: 6, 8: 4, 9: 4, 10: 2, 11: 3, 12: 3, 13: 2, 16: 1. В кладе ВЕР семейства *Poaceae*: 1: 695, 2: 22, 3: 9, 4: 47, 5: 108, 6: 52, 7: 15, 8: 6, 9: 6, 10: 1, 11: 1, 13: 2, 14: 2, 15: 1, 16: 1, 18: 1, 19: 1, 21: 2, 24: 1, 27: 1. В кладе РАССАД семейства *Poaceae*: 1: 202, 2: 7, 3: 4, 4: 11, 5: 30, 6: 7, 7: 5, 8: 2, 9: 2, 14: 3, 16: 3, 17: 3, 18: 3, 19: 1, 20: 1, 26: 1. В семействе *Ranunculaceae*: 1: 111, 2: 13, 3: 3, 4: 12, 5: 22, 6: 7, 7: 5. В семействе *Salicaceae*: 1: 49, 2: 2, 4: 1, 6: 4, 7: 6, 8: 1, 9: 1, 10: 1, 11: 4, 13: 2, 14: 4, 16: 1, 17: 1, 18: 1, 20: 1. В семействе *Solanaceae*: 1: 238, 2: 18, 3: 2, 4: 19, 5: 15, 6: 13, 7: 14, 8: 10, 9: 7, 10: 3, 11: 3, 12: 1, 13: 5, 14: 4, 15: 1, 16: 4, 17: 3, 18: 3, 19: 1, 20: 1, 24: 1, 30: 1.

Прямые повторы слов с наибольшей длиной найдены в семействе *Onagraceae*: 50 п. н. между генами *petA* и *psbJ* и 78 п. н. между генами *atpH* и *atpI*. В семействе *Malvaceae* найден повтор слова длиной 51 п. н. между генами *petN* и *psbM*. Преобладают вставки слова из одного нуклеотида, их найдено 4642. Среди вставок прямых повторов слов большей длины наиболее часто встречаются слова длины 5 п. н., что видно из рис. 3. Вставки повторов слов длиной более 24 п. н. наблюдаются очень редко. Это — две вставки при длине слова 26 и по одной вставке при длинах слов 27, 29, 30, 35, 50, 51, 78. При поиске вставок отсутствовало какое-либо ограничение на длину повтора.

Рассмотрим пример вставок прямых повторов в 5'-лидерных областях гена *psbM* в семействе *Pinaceae*, рис. 2. Здесь имеются троекратные повторы слова длиной 11 п. н. у *K. davidiana* и *P. thunbergii*, которые расположены в разных местах относительно иницирующего кодона у *psbM* и различны по нуклеотидному составу, что указывает на их независимое возникновение в эволюции. У *C. argyrophylla*, *C. deodara*, *P. sitchensis*, *P. gerardiana*, *P. krempfii*, *P. koraiensis* перед этим геном отсутствуют повторы слов длины 11 п. н., у *P. contorta* и *P. thunbergii* имеется повтор слова АТГАГАТСААА, который отсутствует у остальных видов. У *K. davidiana* имеется повтор слова ТТСТАТТСАТТ с одним отличием в третьей копии, который отсутствует у двух видов сосен и расположен дальше от начала гена. Также в этой области наблюдаются вставки прямых повторов слова АААГ у *P. contorta* и *P. thunbergii* и слова АТТС у *C. argyrophylla*, *P. sitchensis* и *Pinus* spp. Можно думать, что 5'-лидерная область гена *psbM* удобна для возникновения повторов участков независимо от их нуклеотидного состава. У близкого вида *C. japonica*, не входящего в семейство *Pinaceae*, повторы перед *psbM* отсутствуют, что подтверждает возникновение, а не потерю повторов в этом семействе. Отметим, что области, расположенные по соседству с этими повторами, также обладают заметной консервативностью.

Ca TTAGTTAGATTTTGTATTTCGAACCAATTTTATAAATT-----ACATCTCTTTCA
 Cd TTAGTTAGATTCGGTTACTCGAACCGATTCTATTAATT-----CAATCACATCTATTTCA
 Kd TTAGTTATATCCATTACTCGAACCGATTTCTATTCATTTTCTATTCATTTTCTATTAATTCAATCATATCTATTTCA
 Ps TTAGTTAGATTCGGTTACTCGAACCGATTCTATCAATT-----ACATCTATTTCA
 Pc TTAGTTAGATTCGGTTACTCGAACCAATTCTATCAATG-----ACATAGATCTATTTCA
 Pg TTAGTTAGATTCGGTTACTCGAACCGATTCCATAGATA-----AAATACATCTATTTCA
 Pkr TTAGTTAGATTCGGTTACTCGAACCGATTCCATAGATA-----AAATACATCTATTTCA
 Pk TTAGTTAGATTCGGTTACTCGAACCGATTCCATAGATA-----AAATACATCTATTTCA
 Pt TTAGTTAGATTCGGTTACTCGAACCAATTCTATCAATG-----ACATACATCTATTTCA
 Cj TCAGATTGATCCTATTGATGGAATT-ACTCCATGGATT-----CAATCCATTTT

Ca TGAATCTTTCAAG----ATTCATTCATCTCT-----ATGAGATCAAATTCGAGCTATTTT
 Cd TGAATCTTGAAAG-----ATTCATCTCT-----ATGAGATCAAATTCGAGCTATTT
 Kd TGGATCTTGAAAG-----ATTCATCTCT-----ATGAGATAAAATTCGAGCTATTT
 Ps TGGATCTTGAAAG----ATTCATTCATCTCT-----ATGAAATTAATTCGAGCTATTT
 Pc TGGATCTTGAAAGAAAGATTCAATTCATCTCTATGAGATCAAATGAGATCAAATGAGATCAAATTCTCGAGCTATTT
 Pg TGGATCTTGAAAG----ATTCATTCATCTCT-----ATGAGATAAAATTCGAGCTATTT
 Pkr TGGATCTTGAAAG----ATTCATTCATCTCT-----ATGAGATAAAATTCGAGCTATTT
 Pk TGGATCTTGAAAG----ATTCATTCATCTCT-----ATGAGATAAAATTCGAGCTATTT
 Pt TGGATCTTGAAAGAAAGATTCAATTCATCTCTATGAGATCAAATGAGATCAAATGAGATCAAATTCTCGAGCTATTT
 Cj TTTGAC-----AAAAAATGAGATACTCT-----ATGAGATCAAATTCGAGTATTG

Ca T-GAACGAAGTAAAAATAAGGAGATC
 Cd TAGAACGAAGTAAAAATAAGGAGATC
 Kd T-AAACGAAGTAAAAATCAGGGGATC
 Ps TTTAACGAAGTAAAGATCAGGAGATC
 Pc TTGAACAAAGTAAAAATAAGGAGATC
 Pg TGGAACAAAGTGAAAATCAGGAGATC
 Pkr TGGAACAAAGTGAAAATCAGGAGATC
 Pk TGGAACAAAGTGAAAATCAGGAGATC
 Pt TTGAACAAAGTAAAAATAAGGAGATC
 Cj T----AAAACGAAGGGAAAATCAATC

Рис. 2. Множественное выравнивание участков 5'-лидерных областей гена *psbM*, в которых возникли троекратные прямые повторы слов длиной 11 п. н. и двукратные повторы слов длиной 4 п. н. Подчёркнуты повторы, а в случае *K. davidiana* подчёркнуто и продолжение повтора, которое отличается от повторяемого слова в одной позиции — выделена серым цветом. Полужирным выделены повторяемые слова φ . Обозначения видов: Ca — *C. argyrophylla*, Cd — *C. deodara*, Kd — *K. davidiana*, Ps — *P. sitchensis*, Pc — *P. contorta*, Pg — *P. gerardiana*, Pkr — *P. krempfii*, Pk — *P. koraiensis*, Pt — *P. thunbergii*, Cj — *C. japonica*.

При исследовании митохондрий получены следующие распределения числа вставок прямых повторов в зависимости от длины слова φ . В порядке Brassicales: 1: 159, 2: 20, 3: 7, 4: 12, 5: 16, 6: 5; в семействе Poaceae: 1: 164, 2: 12, 3: 6, 4: 71, 5: 181, 6: 31, 7: 3, 9: 1, 12: 1.

Так же как и в пластидах абсолютный максимум числа вставок прямых повторов приходится на длину 1 повторяемого слова, а локальный максимум — на длину 5.

4. ОБСУЖДЕНИЕ И ВЫВОДЫ

Нами анализировались пластиды и митохондрии близких видов с целью расширить и уточнить представления об эволюции некодирующих участков ДНК. Полученные результаты показывают, что вставки (точных) прямых повторов являются частыми элементарными событиями в микроэволюции относительно коротких некодирующих участков ДНК пластид и митохондрий. Повторяемое слово обычно имеет длину пять. Распределение длин этих слов в пластомах и митохондриях сходно, рис. 3. Неточный повтор у *Keteleeria davidiana*, по-видимому, возник в результате вставки (точного) прямого повтора слова длины 11 с последующей однонуклеотидной мутацией, рис. 2. Предполагается, что мгновенное появление вставок (точных) прямых повторов является результатом ошибок репликации, ведущих к дупликации некодирующих участков ДНК.

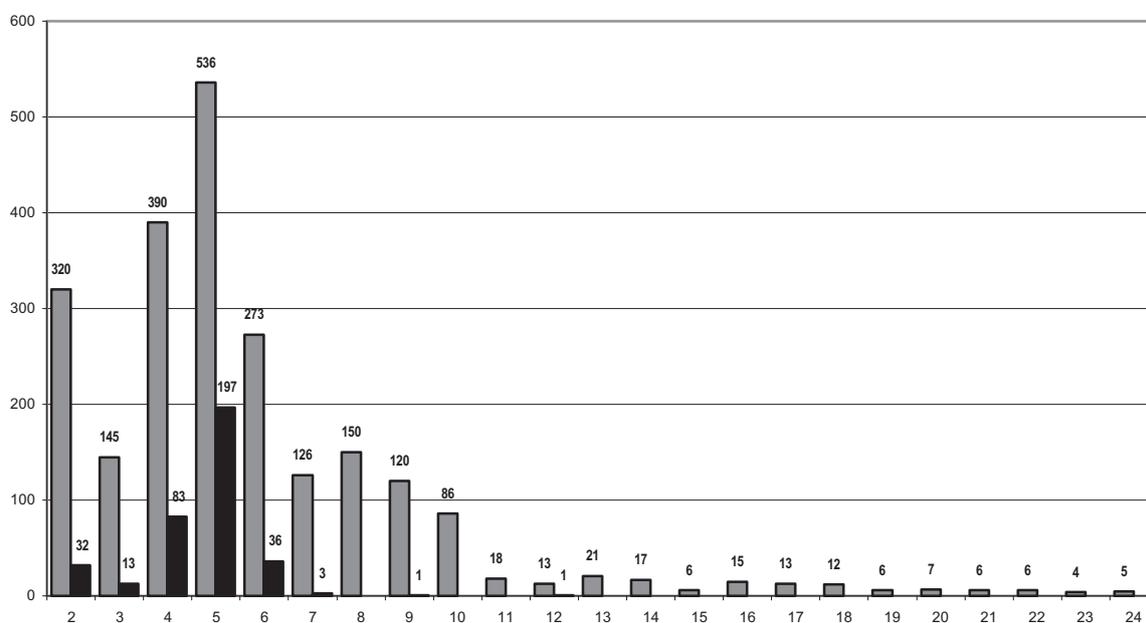


Рис. 3. Число всех найденных вставок прямых повторов в зависимости от длины повторяемого слова, результаты показаны для длин от 2 до 24 п. н. Чёрные столбики показывают числа вставок прямых повторов в пластидах, серые в митохондриях

СПИСОК ЛИТЕРАТУРЫ

1. Kurtz S., Choudhuri J. V., Ohlebusch E., Schleiermacher C., Stoye J., Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Research*, 2001, vol. 29, pp. 4633–4642.
2. Yang M., Xiaowei Zhang X., et al. The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.). *PLoS ONE*, 2010, vol. 5, no. 9, e12762.
3. Greiner S., Wang X., Rauwolf U., Silber M. V., Mayer K., Meurer J., Haberer G., Herrmann R. G. The complete nucleotide sequences of the five genetically distinct plastid genomes of *Oenothera*, subsection

- Oenothera*: I. Sequence evaluation and plastome evolution. *Nucleic Acids Research*, 2008, vol. 36, no. 7, pp. 2366–2378.
4. Ogiwara Y., Terachi T., Sasakuma T. Intramolecular recombination of chloroplast genome mediated by short direct-repeat sequences in wheat species. *Proc. Natl. Acad. Sci. USA*, 1988, vol. 85, no. 22, pp. 8573–8577.
 5. Cai Z., Guisinger M., Kim H. G., Ruck E., Blazier J. C., McMurtry V., Kuehl J. V., Boore J., Jansen R. K. Extensive reorganization of the plastid genome of *Trifolium subterraneum* (Fabaceae) is associated with numerous repeated sequences and novel DNA insertions. *Journal of Molecular Evolution*, 2008, vol. 67, no. 6, pp. 696–704.
 6. Timme R. E., Kuehl J. V., Boore J. L., Jansen R. K. A comparative analysis of the *Lactuca* and *Helianthus* (Asteraceae) plastid genomes: identification of divergent regions and categorization of shared repeats. *American Journal of Botany*, 2007, vol. 94, no. 3, pp. 302–312.
 7. Borsch T., Quandt D. Mutational dynamics and phylogenetic utility of noncoding chloroplast DNA. *Plant Systematics and Evolution*, 2009, vol. 282, pp. 169–199.
 8. Moore M. J., Hassan N., Gitzendanner M. A., Bruenn R. A., Croley M., Vandeventer A., Horn J. W., Dhingra A., Brockington S. F., Latvis M., Ramdial J., Alexandre R., Piedrahita A., Xi Z., Davis C. C., Soltis P. S., Soltis D. E. Phylogenetic analysis of the plastid inverted repeat for 244 species: insights into deeper-level angiosperm relationships from a long, slowly evolving sequence region. *International Journal of Plant Sciences*, 2011, vol. 172, no. 4, pp. 541–558.
 9. Honma Y., Yoshida Y., Terachi T., Toriyama K., Mikami T., Kubo T. Polymorphic minisatellites in the mitochondrial DNAs of *Oryza* and Brassica. *Current genetics*, 2011, vol. 57, no. 4, pp. 261–270.
 10. Kelchner S. A. The evolution of non-coding chloroplast DNA and its application in plant systematics. *Annals of the Missouri Botanical Garden*, 2000, vol. 87, pp. 482–498.
 11. Ingvarsson P. K., Ribstein S., Taylor D. R. Molecular evolution of insertions and deletion in the chloroplast genome of *Silene*. *Molecular Biology and Evolution*, 2003, vol. 20, no. 11, pp. 1737–1740.
 12. Edgar R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 2004, vol. 32, no. 5, pp. 1792–1797.

Insertions of direct repeats in plastid and mitochondrion microevolution of Spermatophyta

O. A. Zverkov, L. Y. Rusin, A. V. Seliverstov, V. A. Lyubetsky

The study focuses on insertions of perfect direct repeats of words of arbitrary length in DNA of plastids and mitochondria. The approach is exemplified on seed plants. Some families one after another were analyzed to further develop and refine published evidence on the evolution of non-coding DNA. The results suggest that insertions are common elementary events in microevolution of short non-coding DNA regions. The repeated word length is usually five, the word length distribution is similar between plastids and mitochondria.

KEYWORDS: direct repeat, plastid, mitochondrion, Spermatophyta, microevolution.