УДК 575.2:582.542.1

ПОВТОРЯЮЩИЕСЯ ПОСЛЕДОВАТЕЛЬНОСТИ ДНК В ГЕНОМАХ РАСТЕНИЙ

А.Б. Щербань

Федеральное государственное бюджетное учреждение науки Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия,
e-mail: atos@bionet.nsc.ru

Поступила в редакцию 25 сентября 2014 г. Принята к публикации 16 октября 2014 г.

Обзор посвящен характеристике основных классов повторяющихся последовательностей ДНК, включая кодирующие (гены рРНК) и некодирующие последовательности (тандемные и диспергированные повторы). Обсуждается особая роль этих компонентов в формировании структурно-функциональной организации генома высших растений, а также в обеспечении повышенной по сравнению с геномом животных генетической изменчивости на уровне как отдельных последовательностей, так и генома.

Ключевые слова: высшие растения, повторяющиеся последовательности ДНК, гены рРНК, тандемный повтор, мобильный элемент.

ДНК ядерного генома растений содержит уникальные или низкокопийные кодирующие последовательности ДНК, непосредственно связанные с ними регуляторные элементы, а также различные классы повторяющихся последовательностей (ПП) ДНК (Schmidt, Heslop-Harrison, 1998). Часть ПП ДНК может иметь кодирующую функцию как, например, гены рибосомальной РНК (рРНК). Другая часть этих последовательностей не несет функцию кодирования белков, необходимых для функционирования и воспроизводства организма. Однако они могут кодировать белки, важные для их собственного размножения в геноме (мобильные генетические элементы).

ПП ДНК являются основным компонентом генома растений. По разным данным, у злаков к ним может относиться до 90–95 % ядерной ДНК (Flavell *et al*., 1974; Paux *et al*., 2006). ПП представляют собой высокогетерогенную группу, представленную тысячами или даже десятками тысяч семейств, отличающихся по длине мотива, уровню копийности и организации в геноме (Flavell, 1986; Kubis *et al*., 1998; Heslop-Harrison, 2000). Причины поддержания столь значительного количества ПП ДНК, а также высокого разнообразия их состава до конца не изучены. Предполагают, что они играют важную роль в стабилизации и поддержании структуры хромосом, участвуют в «узнавании» и правильном расхождении хромосом во время митоза и мейоза (Vershinin *et al*., 1995; Kubis *et al*., 1998). Локусы некоторых семейств сателлитных ДНК выступают как точки рекомбинации хромосом в мейозе (Vershinin *et al*., 1995; Linares *et al*., 1998). Теломеро-ассоциированные повторы выполняют функцию защиты теломер и участвуют в регуляции генов, локализованных в субтеломерных районах (Sykorova *et al*., 2003). Показано, что видообразование у растений часто связано с быстрыми изменениями фракции ПП ДНК (Flavell, 1982; Dvorak *et al*., 1998; Cuadrado, Jouve, 2002).

ГЕНЫ рРНК

Наряду с первичной перетяжкой в области центромеры у метафазной хромосомы существует вторичная перетяжка в районе организации ядрышка (NOR – nucleolar organizer region). Этот район соответствует сайтам локализации генов, кодирующих рРНК, которая считывается

в виде единого транскрипта с коэффициентом седиментации 45S. Данная РНК затем подвергается процессингу с образованием 18S, 5,8S и 26S субъединиц рРНК. Тандемно организованные повторяющиеся единицы рДНК, наряду с цистроном 18S–5,8S–26S рРНК, содержат межгенный спейсерный район, в котором локализуются сайт инициации транскрипции для РНК-полимеразы I, сайты инициации и терминации транскрипции рРНК и другие регуляторные элементы (рис. 1.) (Gerlach, Bedbrook, 1979; Reeder, 1984; Vincentz, Flavell, 1989). Нетранскрибируемый спейсер содержит также ряд повторяющихся нуклеотидных последовательностей, так называемых субповторов A, B, C и D, протяженность и число которых варьируют у разных видов растений; при этом внутри отдельных кластеров длина повторяющейся единицы, как правило, одинаковая (Appels, Dvořák, 1982a, b). У злаков (Triticeae) имеется несколько дискретных локусов рДНК, расположенных на коротких плечах хромосом 1, 5 и 6, каждый из которых содержит ~ 1200 повторяющихся единиц длиной ~ 9 тыс. пар нуклеотидов (т.п.н.) (Flavell, O'Dell, 1976; Gerlach, Bedbrook, 1979). У арабидопсиса локусы рДНК расположены на двух хромосомах с числом копий ~360 единиц/локус, что в сумме составляет около 5 % всего генома (Copenhaver, Pikaard, 1996).

Гены, кодирующие 5S субъединицу рРНК, локализованы отдельно от 45S рРНК генов, но имеют тот же принцип организации. У злаков локусы 5S рРНК расположены на хромосомах 1 и 5 и содержат до нескольких тысяч единиц, каждая из которых состоит из консервативной кодирующей части длиной 120 пар нуклеотидов (п.н.) и вариабельного нетранскрибируемого спейсера (рис. 1) (Appels *et al*., 1980). В геноме Triticeae присутствуют два подсемейства: 5SDna-1 и 5SDna2 с длиной спейсеров 200–349 и 350–380 п.н. соответственно (Gerlach, Dyer, 1980). Различия в размере единиц 5S рДНК обусловлены инсерциями/ делециями в средней полиморфной части спейсера, тогда как 5′- и 3′-фланкирующие районы генов являются консервативными и содержат сайт связывания РНК-полимеразы III и сайт терминации соответственно (Appels *et al*., 1992). У некоторых видов *Brachypodium* были обнаружены очень короткие спейсеры длиной 30 п.н., поэтому можно предполагать, что у растений часть спейсерных последовательностей лишена какой-либо функции (Cox *et al*., 1992).

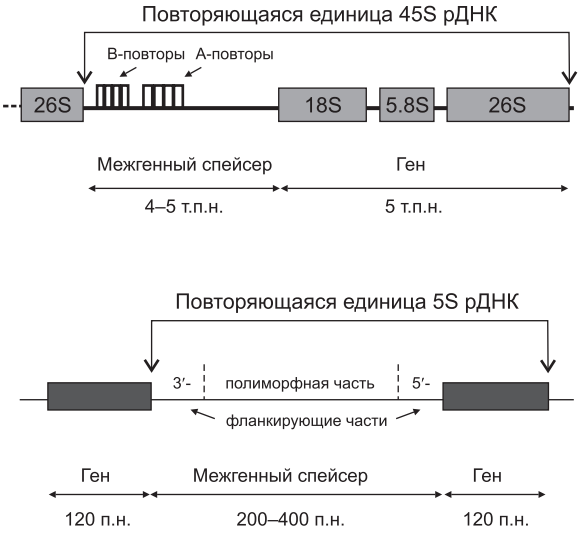

Рис. 1. Организация рибосомальных генов, кодирующих 45S и 5S рРНК.

Как 45S рРНК, так и 5S рРНК локусы часто подвергаются реорганизации в ходе эволюции. Так, у двух экотипов *A.thaliana* (Landsberg и Columbia) локусы 5S рРНК расположены на разных хромосомах (Murata *et al*., 1997). У злаков как размер локусов рРНК, так и их расположение на хромосомах сильно отличаются даже между близкородственными видами (Castilho, Heslop-Harrison, 1995).

ТАНДЕМНО ОРГАНИЗОВАННЫЕ ПОСЛЕДОВАТЕЛЬНОСТИ ДНК

Первые представления о тандемно организованных ПП были получены при изучении сателлитной ДНК. Сначала этим термином обозначали ту часть генома, которая отделялась при градиентном ультрацентрифугировании (Kit, 1961) и, следовательно, по плотности и содержанию AT/GC пар должна была отличаться от основной массы ДНК. В настоящее время под этим термином подразумевается характерный компонент эукариотического генома, состоящий из тандемно организованных повторов. Сателлитная ДНК не кодирует белки и локализована в конститутивном гетерохроматине хромосом (см. обзор: Хемлебенб и др., 2003). Сателлитная ДНК состоит из рядов идентичных повторяющихся единиц (мономеров), число

которых может варьировать от нескольких сот до более миллиона. В зависимости от размера повторяющихся единиц различают микро- и минисателлитную ДНК. Микросателлиты (SSR – simple sequence repeats) имеют длину мономера не более 10 п.н., тогда как у минисателлитов она колеблется в пределах 10–100 п.н. (Sharma, Raina, 2005). Остальные повторы с длиной мономера выше 100 п.н. относятся к сателлитной ДНК. В геномах пшеницы и ячменя микросателлитная ДНК составляет 1,2 % и 3,8 % соответственно и в основном представлена короткими высокоповторяющимися последовательностями $(GAA)_n(GAG)_m$ (Mitra, Bhatia, 1986). Были выделены и охарактеризованы динуклеотидные микросателлитные последовательности *Ae. tauschii*, донора генома D мягкой пшеницы (Pestsova *et al*., 2000a). Микросателлитные повторы имеют, главным образом, интерстициальное расположение на хромосомах и широко используются в качестве маркеров для построения молекулярно-генетических карт, в частности мягкой пшеницы (Pestsova *et al*., 2000a; Paillard *et al*., 2003; Somers *et al*., 2004). Для минисателлитной и сателлитной ДНК характерно более локализованное распределение на концах хромосом или в центромерных районах. Так, например, протяженные тандемные повторяющиеся единицы входят в состав теломерного гетерохроматина ржи *Secale cereale*, занимая 8–12 % генома (Bedbrook *et al*., 1980). Были детально исследованы четыре семейства теломерных повторов ржи, имеющих сложную иерархическую структуру. Основным является семейство 120 п.н. (p*Sc*119.2), первоначально клонированное из *Secale cereale* и впоследствии обнаруженное у многих видов трибы Triticeae и сестринской трибы Avenae (Bedbrook *et al*., 1980; McIntyre *et al*., 1990; Vershinin *et al*., 1995; Salina *et al*., 1998). Повторяющиеся единицы остальных семейств содержат простые субповторы длиной 120 п.н. в сочетании с другими последовательностями (Bedbrook *et al*., 1980).

Из *Aegilops speltoides* было выделено видоспецифичное теломерное семейство повторов Spelt1, составляющее около 2 % ядерной ДНК данного вида (Salina *et al*., 1998). Указанные повторы длиной 178 п.н. входят в состав теломерных районов хромосом наряду с более общим для злаков семейством Spelt52, при этом Spelt1 повторы расположены дистально (Salina *et al*., 2006). Семейство Spelt52 содержит мономеры двух типов: Spelt52.1 и Spelt52.2, которые имеют одинаковую консервативную часть длиной 283 п.н. и негомологичные участки длиной 92 и 106 п.н. соответственно (Salina *et al*., 2004).

Помимо уже указанных семейств тандемных повторов, у представителей трибы Triticeae был найден ряд других семейств, имеющих теломерную (центромерную) локализацию:

а) *Afa* (p*As*1)-семейство повторов длиной 340 п.н., которые локализованы в теломерных и интерстициальных хромосомных районах *Ae. tauschii*, *T. aestivum*, видов *Hordeum*, *Elymus trachycaulus* (Rayburn, Gill, 1986; Nagaki *et al*., 1995, 1998). К этой же группе можно отнести семейство тандемно организованных последовательностей p*Aes*KB52, расположенных в теломерных районах хромосом у некоторых видов Triticeae, относящихся к S-геному (*Ae. speltoides*, *Ae. sharonensis* и *Ae. longissima*) (Anamthawat-Josson, Heslop-Harrison, 1993; Zhang *et al*., 2002; Salina *et al*., 2004);

б) семейство *TaiI*, состоящее из тандемных повторов длиной 570 п.н., обнаруженных в теломерных районах большинства видов Triticeae, за исключением видов *Triticum* и *Aegilops*, у которых выявлено центромерное расположение указанных повторов (Kishii, Tsujimoto, 2002);

в) специфичный для центромер тандемный повтор *CentC* кукурузы *Zea mays* длиной 156 п.н. (Ananiev *et al*., 1998);

г) *Hv*RT-теломерные повторы *Hordeum vulgare* с длиной мономера 118 п.н., имеющие сложную иерархическую организацию (Kilian, Kleinhofs, 1992).

Тандемные повторы могут распространяться по геному посредством неравного кроссинговера между парами сестринских хроматид или гомологичных хромосом (Charlesworth *et al*., 1994), а также с помощью репликации экстрахромосомной ДНК по модели «катящегося колеса» (Navratilova *et al*., 2008).

ДИСПЕРГИРОВАННЫЕ ПОСЛЕДОВАТЕЛЬНОСТИ ДНК

Известно, что основная масса диспергированных ПП ДНК возникла в результате активности мобильных элементов. Мобильные

элементы (МЭ) генома подразделяют на два класса: элементы класса I (ретротранспозоны), которые перемещаются с помощью механизма «копирования–встраивания» с использованием РНК-посредников; и элементы класса II (ДНК-транспозоны), которые используют механизм «вырезания–встраивания», с образованием либо одно-, либо двухцепочечных разрывов ДНК (Wicker *et al*., 2007) (рис. 2).

Класс I, ретротранспозоны

Согласно общепринятой классификации, по наличию–отсутствию длинных концевых повторов (LTR) ретротранспозоны делятся на LTR-содержащие (LTR ретротранспозоны) и несодержащие ретротранспозоны. Последние в свою очередь делятся на длинные диспергированные (*LINE*) и короткие диспергированные (*SINE*) элементы (рис. 2). Кроме этих основных групп, выделяют дополнительные группы: *DIRS*- и *Penelope*-подобные элементы, а также неавтономные варианты LTR-ретротранспозонов: *TRIM*- (Terminal-Repeat Retrotransposons In Miniature) и *LARD*-элементы (Large Retrotransposon Derivatives) (см. обзор Сергеева, Салина, 2011).

Выделяют два основных суперсемейства LTR-ретротранспозонов: *Gypsy* и *Copia*, различающихся порядком расположения генов обратной транскриптазы и интегразы друг относительно друга, но имеющих одинаковый механизм транспозиции (Wicker *et al*., 2007). У растений LTR-ретротранспозоны являются преобладающей группой МЭ. Они составляют от 15 % генома у *Arabidopsis thaliana* и до 90 % у некоторых представителей *Liliaceae* (Flavell *et al*., 1992; Voytas *et al*., 1992; Bennetzen, 1996; SanMiguel *et al*., 1996; Suoniemi *et al*., 1998; Kumar, Bennetzen, 1999; Vicient *et al*., 2001; Vitte, Panaud, 2005; Sabot, Schulman, 2006; Vitte, Bennetzen, 2006). Длина этих элементов варьирует от нескольких т.п.н. до 25 т.п.н., а размер LTR – от нескольких сотен п.н. до нескольких т.п.н. LTR ограничены инвертированными динуклеотидами TG/CA и содержат промотор с энхансером, а также сайт полиаденилирования. При интеграции LTR ретротранспозоны образуют дупликацию сайта-мишени (TSD) длиной 4–6 п.н. Внутренняя область LTR ретротранспозонов содержит 2 открытые рамки считывания (ORF): GAG для структурного белка вирусоподобных частиц и POL, кодирующую аспарагиновую протеиназу (PR), обратную транскриптазу (RT), РНКазу H (RH) и интегразу (INT) (рис. 2). LTR ретротранспозоны также содержат специальные сигналы для упаковки, димеризации, обратной транскрипции и интеграции. Растения с большими геномами, такие как кукуруза, пшеница, ячмень, могут содержать

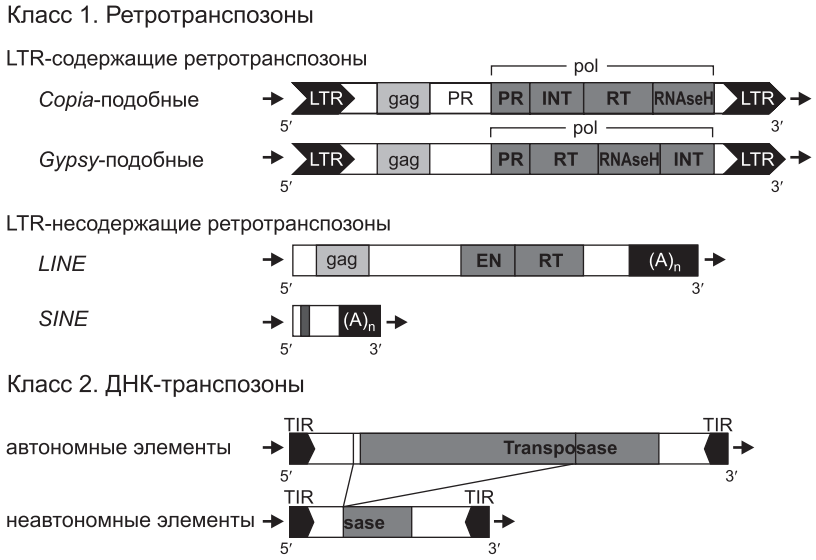

Рис. 2. Основные группы мобильных элементов (по: Wicker *et al*., 2007).

тысячи семейств LTR-ретротранспозонов. Однако в каждом геноме основную массу диспергированных повторов ДНК составляют, как правило, несколько или даже одно семейство ретротранспозонов, например *BARE1* у ячменя (Vicient *et al*., 1999), *Opie* у кукурузы (SanMiguel *et al*., 1998).

В настоящий момент нет точных данных о содержании *LINE* элементов в геномах растений из-за сложности их идентификации, однако, по-видимому, они составляют гораздо меньшую часть генома в сравнении с LTR ретротранспозонами (Сергеева, Салина, 2011). Автономные *LINE* элементы кодируют по меньшей мере RT и эндонуклеазу (EN) в одной рамке считывания (POL), необходимые для транспозиции (Ostertag, Kazazian, 2005) (рис. 2). У некоторых представителей *LINE* выявлена GAG-подобная ORF с 5′-конца от POL, однако ее функция еще не ясна. Хотя *LINE* и образуют TSD в результате транспозиции, редуцированные 5′-концы делают их сложными для определения. Редукция, возможно, является результатом преждевременной терминации обратной транскрипции (Petrov, Hartl, 1998). На 3′-конце *LINE* могут содержать поли(A)-хвост, тандемный повтор или A-богатый район. Таким образом, *LINE* элементы более разнообразны по своей структуре, чем LTR ретротранспозоны. К наиболее известным растительным *LINE* элементам относятся представители суперсемейств *L1* и *RTE* (Zupunski *et al*., 2001).

SINE элементы формально отнесены к классу 1, хотя имеют другое происхождение. Это неавтономные элементы, но они не являются делеционными производными других ретротранспозонов. В отличие от ретропроцессированных псевдогенов, они содержат внутренний промотор. *SINE* элементы перемещаются пассивно, под действием белковых факторов, кодируемых *LINE* элементами (RT) (Kajikawa, Okada, 2002; Kramerov, Vassetzky, 2005). Длина *SINE* элементов составляет 80–500 п.н. Они образуют TSD (5–15 п.н.). «Голова» или передняя часть *SINE* содержит промотор для РНК полимеразы III и определяет суперсемейство *SINE* элемента в зависимости от его происхождения: тРНК, 7SL РНК и 5S РНК. Внутренние районы *SINE* элементов (50–200 п.н.) специфичны для семейства и имеют различное происхождение, иногда могут быть результатом димеризации или тримеризации *SINE* элементов. Источник 3′-района обычно неясен (иногда – это укороченный *LINE* элемент). Он может быть AT-богатым, содержать 3–5 п.н. тандемные повторы или поли(T)-хвост, сигнал терминации (Kramerov, Vassetzky, 2005).

Класс II, ДНК транспозоны

ДНК транспозоны, как и ретротранспозоны, обнаружены почти у всех эукариот. Они обычно представлены в небольшом количестве, тем не менее некоторые, как, например, *Pogo-Fot1* (Daboussi, Capy, 2003) у грибов или *CACTA* у пшеницы (Wicker *et al*., 2003), распространились более успешно. Класс II содержит два подкласса: первый объединяет элементы, которые перемещаются с помощью классического механизма «вырезания–встраивания» с образованием двухцепочечного разрыва ДНК, а второй подкласс использует для перемещения механизм «катящегося кольца» (Сергеева, Салина, 2011).

Подкласс 1 содержит классические мобильные элементы, характеризующиеся наличием концевых инвертированных повторов (TIR) вариабельной длины (рис. 2). Транспозиция осуществляется с помощью белка транспозазы, который узнает концевые инвертированные повторы и разрезает обе цепи на обоих концах элемента. TIR-содержащие ДНК транспозоны представлены надсемействами *hAT*, *Tc-Mariner*, *Mutator* (*MULU*), *P*, *PIF-Harbinger* и *CACTA* (Wicker *et al*., 2007). Наиболее подробно изучены *hAT*-семейства: *Ac-Ds*-элементы кукурузы и *Tam3* львиного зева (Rubin *et al*., 2001). Суперсемейство *PIF-Harbinger* предпочтительно встраивается рядом с TAA-мотивом. Эти мобильные элементы содержат две ORF: одна кодирует ДНК-связывающий белок, другая – транспозазу (Jurka, Kapitonov, 2001).

CACTA элементы имеют характерную структурную черту: на внешних концах TIR присутствует консервативный мотив CACTA. Наиболее хорошо изученным является *CACTA*-транспозон семейства *Spm* (*Suppressor-Mutator*) кукурузы (Frey *et al*., 1990). Члены суперсемейства *CACTA* кодируют транспозазу и ORF с неизвестной функцией. У растений короткие

TIR ограничены высококонсервативными CACTA (иногда CACTG) мотивами и фланкированы 3 п.н. TSD (табл.). TIR обычно фланкируют сложные массивы субтерминальных повторов (Wicker *et al*., 2003).

К подклассу 1 принадлежит многочисленная и гетерогенная группа неавтономных элементов *MITE* (Miniature Inverted-Repeat Transposable Element) (Сергеева, Салина, 2011). Эти элементы имеют размер от нескольких десятков до нескольких сотен п.н.. Размер TIR весьма вариабелен и составляет от нескольких десятков до нескольких сотен п.н. *MITE* элементы образуют TSD длиной 9–11 п.н. Основываясь на нуклеотидной последовательности TIR, *MITE* разделяют на два больших надсемейства: *Stowaway*-подобные и *Tourist*-подобные элементы. В геномах растений эти элементы часто расположены рядом с генами (Bureau, Wessler, 1992, 1994).

Подкласс 2 содержит мобильные элементы, процесс транспозиции которых влечет за собой репликацию без двунитевых разрывов, чем резко отличается от подкласса 1. Эти мобильные элементы перемещаются путем репликации, приводящей к переносу только одной цепи. С подклассом 1 их связывает только признак отсутствия РНК-посредников, но не наличие общих предков (Wicker *et al*., 2007). Подкласс 2 в геномах растений представлен элементами суперсемейства *Helitron*, которые хорошо описаны в геноме кукурузы (Kapitonov, Jurka, 2001). Концы этих элементов не содержат TIR, но имеют TC или CTRR мотивы (где R – пурин). Автономные элементы *Helitron* кодируют тирозиновую рекомбиназу Y2-типа с геликазным доменом и способностью к инициации репликации. Также они могут кодировать белки, связывающие одноцепочечную ДНК, или другие белки (Kapitonov, Jurka, 2001). Любопытно, что многие *Helitron*-элементы несут случайно захваченные фрагменты генов хозяина. Хотя эти мобильные элементы были описаны в основном для растений, они также найдены у животных и грибов (Poulter, Goodwin, 2005).

Некоторые группы ДНК-транспозонов остаются неклассифицированными, поскольку для них известны последовательности только неавтономных вариантов. Такую группу малоизученных МЭ представляют *Foldback*-(FB) элементы, сходные с одноименным элементом *Drosophila melanogaster*. Эти элементы представлены в широком ряду организмов, в том числе и растений: у риса *Oryza sativa*, *Arabidopsis thaliana*, ржи *Secale cereale*, пасленовых Solanaceae (Feschotte, Pritham, 2007).

ИЗМЕНЧИВОСТЬ РАЗМЕРА ЯДЕРНОГО ГЕНОМА У ВЫСШИХ РАСТЕНИЙ

Размер ядерного генома у высших (покрытосеменных) растений варьирует в широких пределах, хотя для каждого вида он обычно имеет более или менее постоянное значение. Размер генома, измеряемый в количестве ДНК (в млн пар нуклеотидов, м.п.н.) на гаплоидный геном (1C), варьирует от 63 м.п.н. у плотоядного растения *Genlisea aurea* ($2n = 52$) до 149 000 м.п.н. у вида лилий *Paris japonica* ($2n = 40$), т. е. различия могут достигать 2 300 раз (Bennett, Leitch, 2011). Растения с маленьким геномом (< 200 м.п.н.) включают одно семейство однодольных и 13 семейств двудольных. Растения с огромным геномом часто встречаются среди видов однодольных, особенно относящихся к порядку Liliales, тогда как только 9 семейств двудольных имеют виды с геномом более 15 000 м.п.н. (Leitch *et al*., 2007; Bennett, Leitch, 2011). Средний размер генома покрытосеменных растений составляет 5 800 м.п.н. (у однодольных – 10 200 м.п.н.; у двудольных – 2 800 м.п.н.) (рис. 3).

Средняя длина гена (за исключением интронов) у эукариот равна 1 346 п.н. (с небольшой вариацией между таксонами) (Xu *et al*., 2006), тогда как число генов у диплоидных высших растений – около 30 000 (Ming *et al*., 2008), что составляет в сумме 40 м.п.н.. Если сюда добавить структурные районы хромосом (центромеры и теломеры), рДНК, регуляторные последовательности и интроны, то получим оценку ~ 60 м.п.н., сопоставимую с минимальным размером генома. Большинство известных растительных видов имеют размер генома, значительно превышающий минимальное значение. Увеличение размера ядерного генома может достигаться за счет мультипликации либо части генома, либо генома в целом (полиплоидизация). Полиплоидия происходила многократно в ходе эволюции растительных видов и была

Таблица

Классификация мобильных элементов и их распространенность у разных групп организмов (по: Wicker *et al*., 2007)

Классификация		Структура	TSD	Распространенность
Порядок	Суперсемейство			
Класс 1 – ретротранспозоны				
LTR	Copia	GAG AP INT RT RH	4–6	Р, Г, М, О
	Gypsy	GAG AP RT RH INT	4–6	Р, Г, М, О
	Bel-Pao	GAG AP RT RH INT	4–6	М
	Retrovirus	GAG AP RT RH INT ENV	4–6	М
	ERV	GAG AP RT RH INT ENV	4–6	М
DIRS	DIRS	GAG AP RT RH YR	0	Р, Г, М, О
	Ngaro	GAG AP RT RH YR	0	М, Г
	VIPER	GAG AP RT RH YR	0	О
PLE	Penelope	RT EN	вар.	Р, Г, М, О
LINE	R2	RT EN	вар.	М
	RTE	APE RT	вар.	М
	Jockey	ORF1 APE RT	вар.	М
	L1	ORF1 APE RT	вар.	Р, Г, М, О
	I	ORF1 APE RT RH	вар.	Р, Г, М
SINE	tRNA		вар.	Р, Г, М
	7SL		вар.	Р, Г, М
	5S		вар.	М, О
Класс 2 – ДНК-транспозоны. Подкласс 1				
TIR	Tc1-Mariner	Tase′	TA	Р, Г, М, О
	hAT	Tase′	8	Р, Г, М, О
	Mutator	Tase′	9–11	Р, Г, М, О
	Merlin	Tase′	8–9	М, О
	Transib	Tase′	5	М, Г
	P	Tase	8	Р, М
	PiggyBac	Tase	TTA	М, О
	PIF-Harbinder	Tase′ ORF2	3	Р, Г, М, О
	CACTA	Tase ORF2	2–3	Р, Г, М
Crypton	Crypton	YR	0	Г
Класс 2 – ДНК-транспозоны. Подкласс 2				
Helitron	Helitron	RPA // Y2 HEL	0	Р, Г, М
Maverick	Maverick	C-INT ATP // CYP POL B	6	Г, М, О

Примечания. ⟶ – LTR, длинный концевой повтор; ▭ – кодирующая область; ── – некодирующая область; ⟩── ──⟨ – TIR, концевые инвертированные повторы; ▬ – диагностические особенности в некодирующей области; –//– – район, содержащий одну или более дополнительных OFR.

AP – аспарагиновая протеаза; ENV – белок оболочки; POL B – ДНК-полимераза B; Tase – транспозаза (′ – с DDE-мотивом); APE – апуриновая эндонуклеаза; GAG – капсидный белок; RH – РНКаза H; ATP – упаковочная АТФаза; HEL – геликаза; RPA – репликативный белок A (только у растений); YR – тирозиновая рекомбиназа; C-INT – C-интеграза; CYP – цистеиновая протеаза; PFR – открытая рамка считывания с неизвестной функцией; RT – обратная транскриптаза; Y2 – YR с YY-мотивом; EN – эндонуклеаза; Р – растения; Г – грибы; М – Metazoa; О – остальные организмы.

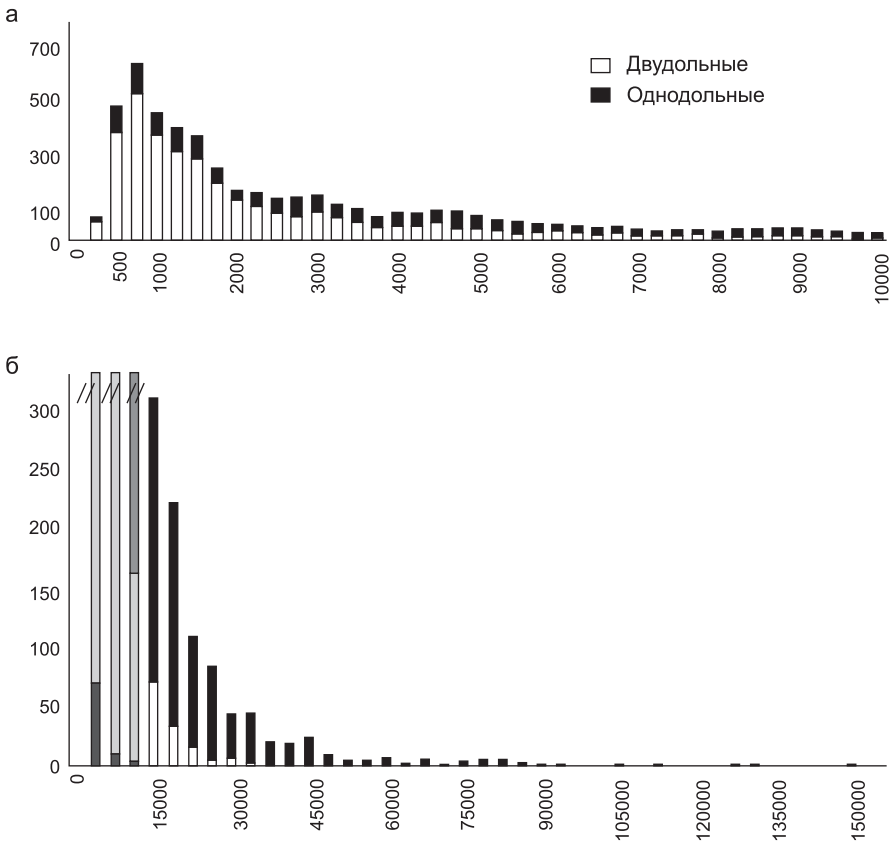

Рис. 3. Размер генома у покрытосеменных растений.

а – до 10 000 м.п.н.; б – от 15 000 до 150 000 м.п.н. По оси Х – размер генома (1С в м.п.н.); по оси У – частота. Данные взяты из: http://data.kew.org/cvalues/cvalOrigReference.html (по: Bennett, Leitch, 2011).

установлена в недавнем эволюционном прошлом у примерно половины изученных видов растений (Schranz *et al*., 2012). Второй источник вариации размера генома – амплификация последовательностей геномной ДНК. Помимо амплификации, в геноме могут происходить делеции как генов (особенно дуплицированных), так и повторяющихся некодирующих последовательностей, что может отражаться на размере генома. Например, геном *Arabidopsis thaliana* на примерно 50 % меньше генома близкородственного вида *A. lyrata*, хотя эти виды разошлись от общего предка сравнительно недавно – около 10 млн лет назад (Hu *et al*., 2011). В этом случае различие в размере генома обусловлено множественными делециями в некодирующих межгенных районах, насыщенных мобильными элементами. Межвидовая гибридизация довольно часто сопровождается делециями ДНК (Ma, Gustafson, 2008). Изучая эволюцию размера генома в семействе Brassicaceae, ученые обнаружили 16-кратные различия среди 185 видов (Lysak *et al*., 2009). Половина этих видов имели уменьшенный размер генома по сравнению с общим предком, несмотря на динамичные процессы, направленные на увеличение размера генома (транспозиции мобильных элементов и полиплоидизация). До сих пор не установлены общие закономерности и селекционные механизмы, которые лежат в основе вариации размера генома в ту или другую сторону в ходе эволюции растительных видов.

Большинство межвидовых различий в размере генома связаны с содержанием ПП ДНК, которое подвержено быстрым изменениям. У кукурузы существуют большие различия по размеру терминального гетерохроматина хромосом, состоящего из тандемных повторов (Laurie, Bennett, 1985). Большая вариация по содержанию гетерохроматина, отражающаяся на размере генома, наблюдается также у ржи *Secale cereale* ($2n = 14$), причем полиморфизм

затрагивает даже гомологичные хромосомы (Alkhimova *et al*., 2004). При некоторых условиях повторяющиеся последовательности ДНК терминальных районов хромосом теряются в ходе митоза. Была показана ограниченная вариабельность размера генома у диплоидных и тетраплоидных видов пшеницы, обусловленная активностью ретроэлементов (Ozkan *et al*., 2010). У льна обнаружены изменения числа копий генов рРНК в результате различных стрессовых воздействий (Cullis, 2005). У кукурузы количественные изменения затрагивают все концевые районы хромосом, в которых происходят интенсивные реорганизации (инсерции, дупликации и делеции) отдельных сегментов размером более 1 т.п.н. (Belo *et al*., 2010). Такой полиморфизм, по видимому, имеет важное значение для популяций видов растений, для их адаптации, устойчивости к стрессу и гетерозиса (Biemont, 2008; Belo *et al*., 2010).

ЗАКЛЮЧЕНИЕ

Структурное разнообразие повторяющихся элементов растительного генома отражает его ускоренную эволюцию, в ходе которой происходили многократные раунды полиплоидизации, а также процессы массовой реорганизации геномной ДНК, включающие амплификацию и замещение мобильных элементов, рекомбинацию тандемных повторов теломерного гетерохроматина, быстрые количественные изменения генов рРНК и др. И хотя адаптивная значимость большинства механизмов генетической реорганизации, связанных с ПП ДНК, остается невыясненной, тем не менее ряд данных указывает на селекционное преимущество генотипов, генерирующих большее разнообразие в составе различных фракций ПП ДНК.

Работа выполнена в рамках проекта по фундаментальным научным исследованиям (тема № VI.53.1.5.) при поддержке гранта РФФИ (№ 14-04-00297).

ЛИТЕРАТУРА

Сергеева Е.М., Салина Е.А. Мобильные элементы и эволюция генома растений // Вавилов. журн. генет. и селекции. 2011. Т. 15. № 2. С. 382–397.

Хемлебенб В., Беридзе Т.Г., Бахман Л., Коварик Я., Торрес Р. Сателлитные ДНК // Усп. биол. химии. 2003. Т. 43. С. 267–306.

Alkhimova O.G., Mazurok N.A., Potapova T.A., Zakian S.M. *et al.* Diverse patterns of the tandem repeats organization in rye chromosomes // Chromosoma. 2004. V. 113. P. 42–52.

Anamthawat-Josson K., Heslop-Harrison J.S. Isolation and characterization of genome-specific DNA sequences in Triticeae species // Mol. Gen. Genet. 1993. V. 240. P. 151–158.

Ananiev E.V., Phillips R.L., Rines H.W. Chromosome-specific molecular organization of maize (*Zea mays* L.) centromeric regions // Proc. Natl Acad. Sci. USA. 1998. V. 95. P. 13073–13078.

Appels R., Baum B.R., Clarke B.C. The 5S DNA units of bread wheat (*Triticum aestivum* L.) // Plant Syst. Evol. 1992. V. 183. P. 183–194.

Appels R., Dvořák J. Relative rates of divergence of spacer and gene sequences within the rDNA region of species in the Triticeae: Implications for the maintenance of homogeneity of a repeated gene family // Theor. Appl. Genet. 1982a. V. 63. P. 361–365.

Appels R., Dvořák J. The wheat ribosomal DNA spacer region: Its structure and variation in populations and among species // Theor. Appl. Genet. 1982b. V. 63. P. 337–348.

Appels R., Gerlach W.L., Dennis E.S. *et al*. Molecular and chromosomal organization of DNA sequences coding for the ribosomal RNAs in cereals // Chromosoma. 1980. V. 78. P. 293–311.

Bedbrook J.H., Jones J., O'Del M. A molecular distribution of telomeric heterochromatin in *Secale* species // Cell. 1980. V. 19. P. 545–560.

Belo A., Beatty M., Hondred D., Fengler K. *et al.* Allelic genome structural variations in maize detected by array comparative genome hybridization // Theor. Appl. Genet. 2010. V. 120. P. 355–367.

Bennett M., Leitch I.J. Nuclear DNA amounts in angiosperms: targets, trends and tomorrow // Ann. Bot. 2011. V. 107. P. 467–590.

Bennetzen J.L. The contribution of retroelements to plant genome organization, function and evolution // Trends Microbiol. 1996. V. 4. P. 347–353.

Biemont C. Genome size evolution: within-species variation in genome size // Heredity. 2008. V. 101. P. 297–298.

Bureau T.E., Wessler S.R. *Tourist*: a large family of inverted-repeat elements frequently associated with maize genes // Plant Cell. 1992. V. 4. P. 1283–1294.

Bureau T.E., Wessler S.R. *Stowaway*: a new family of inverted-repeat elements associated with genes of both monocotyledonous and dicotyledonous plants // Plant Cell. 1994. V. 6. P. 907–916.

Castilho A., Heslop-Harrison J.S. Physical mapping of 5S and 18S-25S rDNA and repetitive DNA sequences in *Aegilops umbellulata* // Genome. 1995. V. 38. P. 91–96.

Charlesworth B., Sniegowski P., Stephan W. The evolutionary dynamics of repetitive DNA in eukaryotes // Nature. 1994. V. 371. P. 215–220.

Copenhaver G.P., Pikaard C.S. RFLP and physical mapping with an rDNA-specific endonuclease reveals that nucleolus organizer regions of *Arabidopsis thaliana* adjoin the telomeres on chromosomes 2 and 4 // Plant J. 1996. V. 9. P. 259–272.

Cox A.V., Bennett M.D., Dyer T.A. Specific 5S ribosomal RNA primers for plant species identification in admixtures // Theor. Appl. Genet. 1992. V. 83. P. 684.

Cuadrado A., Jouve N. Evolutionary trends of different repetitive DNA sequences during speciation in the genus *Secale* // J. Hered. 2002. V. 93. P. 339–345.

Cullis C.A. Mechanisms and control of rapid genomic changes in flax // Ann. Bot. 2005. V. 95. P. 201–206.

Daboussi M., Capy P. Transposable elements in filamentous fungi // Annu. Rev. Microbiol. 2003. V. 57. P. 275–299.

Dvorák J., Luo M.C., Yang Z.L. Restriction fragment length polymorphism and divergence in the genomic regions of high and low recombination in self-fertilizing and cross-fertilizing *Aegilops* species // Genetics. 1998. V. 148. P. 423–434.

Feschotte C., Pritham E.J. DNA transposons and the evolution of eukaryotic genomes // Annu. Rev. Genet. 2007. V. 41. P. 331–368.

Flavell R.B. Amplification, deletion and rearrangement: Major sources of variation during species divergence // Genome Evolution / Eds G.A. Dover, R.B. Flavell. London: Acad. Press, 1982.

Flavell R.B. Repetitive DNA and chromosome evolution in plants // Philos. T. R. Soc. Lon. B. 1986. V. 312. P. 227–242.

Flavell R.B., Bennett M.D., Smith J.B., Smith D.B. Genome size and the proportion of repeated nucleotide sequence DNA in plants // Biochem. Genet. 1974. V. 12. P. 257–269.

Flavell R.B., O'Dell M. Ribosomal RNA genes on homologous chromosomes of groups 5 and 6 in hexaploid wheat // Heredity. 1976. V. 37. P. 372–385.

Flavell A.J., Dunbar E., Anderson R. *et al*. *Ty*1-copia group retrotransposons are ubiquitous and heterogeneous in higher plants // Nucl. Acids Res. 1992. V. 20. P. 3639–3644.

Frey M., Reinecke J., Grant S. *et al*. Excision of the *En/Spm* transposable element of *Zea mays* requires two element-encoded proteins // EMBO J. 1990. V. 9. P. 4037–4044.

Gerlach W.L., Bedbrook J.R. Cloning and characterization of ribosomal RNA genes from wheat and barley // Nucl. Acid Res. 1979. V. 7. P. 1869–1885.

Gerlach W.L., Dyer T.A. Sequence organization of the repeating units in the nucleus of wheat which contain 5S rRNA genes // Nucl. Acids Res. 1980. V. 8. P. 4851–4865.

Heslop-Harrison J.S. Comparative genome organization in plants: from sequence and markers to chromatin and chromosomes // Plant Cell. 2000. V. 12. P. 617–636.

Hu T.T., Pattyn P., Bakker E.G., Cao J. *et al.* The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change // Nat. Genet. 2011. V. 43. P. 476–481.

Jurka J., Kapitonov V.V. PIFs meet *Tourists* and *Harbingers*: a superfamily reunion // Proc. Natl Acad. Sci. USA. 2001. V. 98. P. 12315–12316.

Kajikawa M., Okada N. *LINE*s mobilize *SINE*s in the eel through a shared 3′ sequence // Cell. 2002. V. 111. P. 433–444.

Kapitonov V., Jurka J. Rolling-circle transposons in eukaryotes // Proc. Natl Acad. Sci. USA. 2001. V. 98. P. 8714–8719.

Kilian A., Kleinhofs A. Cloning and mapping of telomere-associated sequences from *Hordeum vulgare* L. // Mol. Gen. Genet. 1992. V. 235. P. 153–156.

Kishii M., Tsujimoto H. Genus-specific localization of the *TaiI* family of tandem-repetitive sequences in either the centromeric or subtelomeric regions in Triticeae species (Poaceae) and its evolution in wheat // Genome. 2002. V. 45. P. 946–955.

Kit S. Equilibrium sedimentation in density gradients of DNA preparations from animal tissues // J. Mol. Biol. 1961. V. 3. P. 711–716.

Kramerov D., Vassetzky N. Short retroposons in eukaryotic genomes // Int. Rev. Cytol. 2005. V. 247. P. 165–221.

Kubis S.E., Schmidt, T., Heslop-Harrison J.S. Repetitive DNA elements as a major component of plant genomes // Ann. Bot. 1998. V. 82. P. P. 45–55.

Kumar A., Bennetzen J. Plant retrotransposons // Annu. Rev. Genet. 1999. V. 33. P. 479–532.

Laurie D.A., Bennett M.D. Nuclear DNA content in the genera *Zea* and *Sorghum*. Intergeneric, interspecific and intraspecific variation // Heredity. 1985. V. 55. P. 307–313.

Leitch I.J., Beaulieu J.M., Cheung K., Hanson L. *et al.* Punctuated genome size evolution in Liliaceae // J. Evol. Biol. 2007. V. 20. P. 2296–2308.

Linares C., Ferrer E., Fominaya A. Discrimination of the closely related A and D genomes of the hexaploid oat *Avena sativa* L. // Proc. Natl Acad. Sci. USA. 1998. V. 95. P. 12450–12455.

Lysak M.A., Koch M.A., Beaulieu J.M., Meister A., Leitch I.J. The dynamic ups and downs of genome size evolution in Brassicaceae // Mol. Biol. Evol. 2009. V. 26. P. 85–98.

Ma X.F., Gustafson J.P. Allopolyploidization-accommodated genomic sequence changes in Triticale // Ann. Bot. 2008. V. 101. P. 825–832.

McIntyre C.L., Pereira S., Moran L.B., Appels R. New *Secale cereale* (rye) DNA derivatives for the detection of rye chromosome segments in wheat // Genome. 1990. V. 33. P. 635–640.

Ming R., Hou S., Feng Y. *et al.* The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* L.) // Nature. 2008. V. 452. P. 991–996.

Mitra R., Bhatia C.R. Repeated DNA sequences and polyploidy in cereal crops // DNA Systematics. V. II. Plants / Ed. S.K. Dutta. Boca Raton, Florida: CRC Press, 1986. P. 21–43.

Murata M., Heslop-Harrison J.S., Motoyoshi F. Physical mapping of the 5S ribosomal RNA genes in *Arabidopsis thaliana* by multi-color fluorescence *in situ* hybridization with cosmid clones // Plant J. 1997. V. 12. P. 31–37.

Nagaki K., Tsujimoto H., Isono K., Sasakuma T. Molecular characterization of a tandem repeat, *Afa* family, and its distribution among Triticeae // Genome. 1995. V. 38. P. 479–486.

Nagaki K., Tsujimoto H., Sasakuma T.H. Genome specific repetitive sequence, *pEt2*, of *Elimus trachycaulus* in part of *Afa* family of Triticeae // Genome. 1998. V. 41. P. 134–136.

Navratilova A., Koblizkova A., Macas J. Survey of extrachromosomal circular DNA derived from plant satellite repeats // BMC Plant Biol. 2008. V. 8. P. 90.

Ozkan H., Tuna M., Kilian B., Mori N., Ohta S. Genome size variation in diploid and tetraploid wild wheats // AoB Plants. 2010. doi: 10.1093/aobpla/plq015.

Paillard S., Schnurbusch T., Winzeler M., Messmer M. *et al.* An integrative genetic linkage map of winter wheat (*Triticum aestivum* L.) // Theor. Appl. Genet. 2003. V. 107. P. 1235–1242.

Paux E., Roger D., Badaeva E., Gay G. *et al.* Characterizing the composition and evolution of homoeologous genomes in hexaploid wheat through BAC-end sequencing on chromosome 3B // Plant J. 2006. V. 48. P. 463–474.

Pestsova E., Ganal M.W., Rцder M.S. Isolation and mapping of microsatellite markers specific for the D genome of bread wheat // Genome. 2000a. V. 43. P. 689–697.

Poulter R., Goodwin T. *DIRS 1* and the other tyrosine recombinase retrotransposons // Cytogenet. Genome Res. 2005. V. 110. P. 575–588.

Rayburn A.L., Gill B.S. Isolation of a D-genome specific repeated DNA sequence from *Aegilops squarosa* // Plant Mol. Biol. 1986. V. 4. P. 102–109.

Reeder R.H. Enhancers and ribosomal gene spacers // Cell. 1984. V. 38. P. 349–351.

Rubin E., Lithwick G., Levy A.A. Structure and evolution of the *hAT* transposon superfamily // Genetics. 2001. V. 158. P. 949–957.

Sabot F., Schulman A.H. Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome // Heredity. 2006. V. 97. P. 381–388.

Salina E.A., Adonina I.G., Vatolina T.Yu., Kurata N.A. Comparative analysis of the composition and organization of two subtelomeric repeat families in *Aegilops speltoides* Tausch. and related species // Genetica. 2004. V. 122. P. 227–237.

Salina E.A., Lim Y.K., Badaeva E.D. *et al*. Philogenetic reconstruction of *Aegilops* section Sitopsis and the evolution of tandem repeats in the diploids and derived wheat poliploids // Genome. 2006. V. 49. P. 1023–1035.

Salina E.A., Pestsova E.G., Adonina I.G., Vershinin A.V. Identification of a new family of tandem repeats in Triticeae genomes // Euphytica. 1998. V. 100. P. 231–237.

SanMiguel P., Gaut B.S., Tikhonov A., Nakajima Y., Bennetzen J.L. The paleontology of intergene retrotransposons in maize // Nature Genet. 1998. V. 20. P. 43–45.

SanMiguel P., Tikhonov A., Jin Y.K. *et al.* Nested retrotransposons in the intergenic regions of the maize genome // Science. 1996. V. 274. P. 765–768.

Schmidt T., Heslop-Harrison J.S. Genomes, genes and junk: the large-scale organization of plant chromosomes // Trends Plant Sci. 1998. V. 3. P. 195–199.

Schranz M.E., Mohammadin S., Edger P.P. Ancient whole genome duplications, novelty and diversification: the WGD Radiation Lag-Time Model // Curr. Opin. Plant Biol. 2012. V. 15. P. 147–153.

Sharma S., Raina S.N. Organization and evolution of highly repeated satellite DNA sequences in plant chromosomes // Cytogenet. Genome Res. 2005. V. 109. P. 15–26.

Somers D.J., Isaac P., Edwards K. *et al.* A high-density wheat microsatellite consensus map for bread wheat (*Triticum aestivum* L.) // Theor. Appl. Genet. 2004. V. 109. P. 1105–1114.

Suoniemi A., Tanskanen J., Schulman A.H. *Gypsy*-like retrotransposons are widespread in the plant kingdom // Plant J. 1998. V. 13. P. 699–705.

Sýkorová E., Lim K.Y., Kunická Z., Chase M.W. *et al*. Telomere variability in the monocotyledonous plant order Asparagales // Proc. Biol. Sci. 2003. V. 270. P. 1893–1904.

Vershinin A.V., Schwarzacher T., Heslop-Harrison J.S. The large-scale genomic organization of repetitive DNA families at the telomeres of rye chromosomes // Plant Cell. 1995. V. 7. P. 1823–1833.

Vicient C.M., Kalendar R., Anamthawat-Jonsson K., Schulman A.H. Structure, functionality, and evolution of the *BARE 1* retrotransposon of barley // Genetica. 1999. V. 107. P. 53–63.

Vicient C.M., Kalendar R., Schulman A.H. Envelope-class retrovirus-like elements are widespread, transcribed and spliced, and insertionally polymorphic in plants // Genome Res. 2001. V. 11. P. 2041–2049.

Vincentz M., Flavell R. Mapping of ribosomal RNA transcripts in wheat // Plant Cell. 1989. V. 1. P. 579–589.

Vitte C., Bennetzen J.L. Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution // Proc. Natl Acad. Sci. USA. 2006. V. 103. P. 17638–17643.

Vitte C., Panaud O. LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model // Cytogenet. Genome Res. 2005. V. 110. P. 91–107.

Voytas D.F., Cummings M.P., Konieczny A.K. *et al. Copia*-like retrotransposons are ubiquitous among plant // Proc. Natl Acad. Sci. USA. 1992. V. 89. P. 7124–7128.

Wicker T., Guyot R., Yahiaoui N., Keller B. *CACTA* transposons in Triticeae. A diverse family of high-copy repetitive elements // Plant Physiol. 2003. V. 132. P. 52–63.

Wicker T., Sabot F., Hua-Van A., Bennetzen J.L. *et al.* A unified classification system for eukaryotic transposable elements // Nat. Rev. Genet. 2007. V. 8. P. 973–982.

Xu L., Chen H., Hu X., Zhang R. *et al.* Average gene length is highly conserved in prokaryotes and eukaryotes and diverges only between the two kingdoms // Mol. Biol. Evol. 2006. V. 23. P. 1107–1108.

Zhang P., Friebe B., Gill B.S. Variation in the distribution of a genome-specific DNA sequences on chromosomes reveals evolutionary relations in the *Triticum* and *Aegilops* complex // Plant Syst. Evol. 2002. V. 235. P. 169–179.

Zupunski V., Gubensek F., Kordis D. Evolutionary dynamics and evolutionary history in the *RTE* clade of non-LTR retrotransposons // Mol. Biol. Evol. 2001. V. 18. P. 1849–1863.

REPETITIVE DNA SEQUENCES IN THE GENOMES OF PLANTS

A.B. Shcherban

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia,
e-mail: atos@bionet.nsc.ru

Summary

The main classes of repetitive DNA sequences, including coding (rRNA genes) and non-coding sequences (tandem and dispersed repeats) are characterized. Emphasis is placed on their special role in the formation of the structural and functional organization of the genomes of higher plants and in the support of their higher genetic variation, as compared to animal genomes, at the levels of individual sequences and of the whole genome.

Key words: higher plants, repetitive DNA sequences, rRNA genes, tandem repeat, mobile element.