Вавиловский журнал генетики и селекции. 2017;21(1):128-134
DOI 10.18699/VJ17.231

Plant diversity and transcriptional variability assessed by retrotransposon-based molecular markers

R.N. Kalendar ✉, K.S. Aizharkyn, O.N. Khapilina, A.A. Amenov, D.S. Tagimanova

Republican State Enterprise "National Center for Biotechnology" under the Science Committee of Ministry of Education and Science of the Republic of Kazakhstan, Astana, Kazakhstan

Molecular markers have become crucial part of genetics due to their use in various branches of it, such as positional cloning, which includes identification of genes responsible for desired traits and management of backcrossing programs, as well as in modern plant breeding, and human forensics. Retrotransposons are a major component of all eukaryotic genomes, which makes them suited as molecular markers. The retrotransposons comprise most of large genomes among plants; differences in their prevalence explain most of the variation in genome size. These ubiquitous transposable elements are scattered in all of genome and their replicative transposition allows insert itself into a genome without deletion of the original elements. Retrotransposon activity can occur during development, cell differentiation and stress, and a source of chromatin instability and genomic rearrangements. Both the overall structure of retrotransposons and the domains responsible for the various phases of their replication are highly conserved in all eukaryotes. A high proportion of the retroelements have lost their autonomous transposition ability, either by point mutations and/or deletions, many of them seem to embody defective elements with deletions. Various molecular marker systems have been developed that exploit the ubiquitous nature of these genetic elements and their property of stable integration into dispersed chromosomal loci that are polymorphic within species. The utility of LTR-retrotransposon-based markers, not only for genetic analysis and map construction, in addition also for the isolation and characterization of LTR retrotransposons, such as the long terminal repeats or the internal genes they contain. This review encompasses description of the range of retrotransposon-based marker systems established for plants and evaluation of the role of retrotransposon markers in genetic diversity analysis of plant species.

Key words: plants; polymorphism; molecular marker; repeats; retrotransposon; transposable element.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:
Календарь Р.Н., Айжаркын К.С., Хапилина О.Н., Аменов А.А., Тагиманова Д.С. Оценка разнообразия растений и изменчивости транскрипционной активности с использованием молекулярных маркеров на основе транспозонов. Вавиловский журнал генетики и селекции. 2017;21(1):128-134. DOI 10.18699/VJ17.231

HOW TO CITE THIS ARTICLE:
Kalendar R.N., Aizharkyn K.S., Khapilina O.N., Amenov A.A., Tagimanova D.S. Plant diversity and transcriptional variability assessed by retrotransposon-based molecular markers. Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding. 2017;21(1):128-134. DOI 10.18699/VJ17.231

Received 27.10.2016
Accepted for publication 15.12.2016

✉ e-mail: ruslan.kalendar@mail.ru

Оценка разнообразия растений и изменчивости транскрипционной активности с использованием молекулярных маркеров на основе ретротранспозонов

Р.Н. Календарь ✉, К.С. Айжаркын, О.Н. Хапилина, А.А. Аменов, Д.С. Тагиманова

Республиканское государственное предприятие «Национальный центр биотехнологии» Комитета науки Министерства образования и науки Республики Казахстан, Астана, Казахстан

Молекулярные маркеры играют важную роль в генетике, поскольку используются в исследованиях различных уровней: при позиционном клонировании, которое включает определение генов, контролирующих желаемые признаки, при беккроссировании, а также в современной селекции и судебной медицине. Ретротранспозоны являются основными компонентами эукариотических геномов, что делает их удобными для использования в качестве молекулярных маркеров. Они составляют основную часть хромосом крупных геномов растений; различия в размерах геномов объясняются разным количеством ретротранспозонов. Распространенность этих мобильных элементов по всему геному объясняется их способностью к репликативной транспозиции, т. е. возможностью встраивания в геном без удаления исходного элемента. Структура ретротранспозонов в целом и их доменов, отвечающих за различные фазы их репликации, являются высококонсервативными для эукариот. Значительная часть ретротранспозонов утратила способность перемещаться самостоятельно по причине накопления множественных точечных мутаций и/или делеций. Активность ретротранспозонов может проявляться в процессе развития, на этапе дифференциации клеток, при воздействии стресса, а также может быть источником нестабильности хроматина и геномных перестроек. Для детекции генетического полиморфизма были разработаны различные системы молекулярных маркеров, основанных на распространенности ретротранспозонов по всему геному и их способности к стабильной интеграции в любые локусы хромосом. Маркеры на основе ретротранспозонов целесообразно использовать не только для проведения генетического анализа или картирования, но и для выделения, клонирования и характеристики ретротранспозонов или генов, содержащихся в них. В настоящем обзоре описаны маркерные системы, созданные на основе ретротранспозонов для исследования растений, и оценена их роль в генетическом анализе разнообразия видов растений.

Ключевые слова: растения; полиморфизм; молекулярный маркер; повторы; ретротранспозон; мобильный элемент.

DNA markers have become a key part of genetic variability and diversity research. Scope of their usage includes creating linkage maps, diagnosis of individuals or lines carrying certain linked genes. The DNA marker system is tightly linked to developments in molecular biology and biochemistry (Lewontin, Hubby, 1966). Due to shortcomings of biochemically based markers, development of markers based on DNA polymorphisms have emerged (Kan, Dozy, 1978). This DNA marker system utilizes "fingerprints", distinctive patterns of DNA fragments resolved by electrophoresis and detected by staining or labelling. Molecular markers work by finding nucleotide sequence variation at a particular location in the genome and when this nucleotide sequence is different between the parents of the chosen cross to be distinguishable between plant accessions and to finally study its pattern of inheritance. Molecular marker technologies experienced a huge progress when polymerase chain reaction (PCR) was introduced, enabling execution of many fingerprinting methods. Two categories can be distinguished depending on number of loci detections: single or multiplex.

Large portion of many eukaryotic organisms' genome consists of interspersed repetitive sequences, transposable elements (TEs) in particular. Interspersed repeats in most of studies species weren't distributed uniformly, but rather unevenly, with some of them being clustered around telomeres or centromeres. Variation in copy number of repeat elements and internal rearrangements on both homologous chromosomes occur after the induction of recombinational processes during the meiotic prophase. The resulting heterogeneity in the arrangement of distinguishable repeats is used for certain molecular markers techniques by targeting mentioned repeat elements.

In eukaryotic genome retrotransposons are two major transposable elements, which are defined according to their mode of propagation. They fall under class I TEs and transpose via RNA intermediate, in contrast to other transposons of class II that don't involve RNA intermediate (Finnegan, 1990). Depending on their structure and transposition cycle, retrotransposons can be classified into two main subclasses: LTR retrotransposons and the non-LTR retrotransposons: long interspersed repetitive elements (LINE) and short interspersed nuclear elements (SINE), determined by the presence or absence of long terminal repeats (LTRs) at their ends. All groups are accompanied by their respective non-autonomous forms that lack one or more of the genes essential for transposition: MITEs (miniature inverted-repeat tandem elements) for class II, SINEs for non-LTR retrotransposons, and TRIMs (terminal-repeat retrotransposons in miniature) and LARDs (large retrotransposon derivatives) for LTR retrotransposons (Kalendar et al., 2004, 2008).

Retrotransposons and retroviruses share common similarities, such as overall structural features and basic stages of the life cycle (Frankel, Young, 1998; Vicient et al., 2001; Mita, Boeke, 2016). However, unlike retroviruses, retrotransposons don't leave genome in order to infect new individuals, but insert the new copies only into their host genomes. If integration appears within a cell lineage from which pollen or egg cells are ultimately derived, new polymorphism is formed. These newly integrated copies are useful for discriminating breeding lines, varieties, or populations of plants from each other.

Retrotransposon-based marker systems

Retrotransposons are one the most fluid genomic components, fluctuating immensely in copy number over relatively short evolutionary timescale and represent a major component of the structural evolution of plant genomes (Flavell et al., 1992; Voytas et al., 1992; Macas et al., 2015). In plants, LTR retrotransposons tend to be more abundant than non-LTR (Macas et al., 2011). In many crop plants between 40–70 % of the total DNA is comprised of LTR retrotransposons (Pearce et al., 1996; Goke, Ng, 2016). Most of retrotransposons are nested, mixed, inverted or truncated in chromosomal sequences. Fragments of LTR with retrotransposons internal part are located near other retrotransposons, which allows the use LTR sequences for PCR amplification. Sites of genome with high density of retrotransposons can be used to detect their chance association with other retrotransposons. Event in which new genome integrations result from retrotransposon activity or recombination can be used for distinguishing reproductively isolation plant line. In this case, amplified bands derived from new insert or recombination will be polymorphic, appearing only in plant lines in which the insertions or recombination have taken place.

Different ways of using transposable elements as molecular markers have been designed. Their qualities such as abundance, general dispersion, and activity allow perfect conditions for developing molecular markers. By using retrotransposon sequences as molecular markers, many methods were developed as primers in the polymerase chain reaction (Kalendar, 2011; Kalendar et al., 2011). The inter-repeat amplification polymorphism techniques such as inter-retrotransposon amplified polymorphism (IRAP), retrotransposon microsatellite amplification polymorphisms (REMAP) or inter-MITE amplification have used abundant dispersed repeats such as the LTRs of retrotransposons and SINE-like sequences (inter-SINE amplified polymorphism – ISAP) (Bureau, Wessler, 1992; Kalendar, Schulman, 2006; Wenke et al., 2011; Seibt et al., 2012), also called Alu-PCR or SINE-PCR (Charlieu et al., 1992).

Positive correlation was detected between the genome size of studied organisms and the efficiency of repeat-based amplification techniques. The larger the genome, the easier it is to develop good primers for revealing multiple bands for polymorphism detection (main cereals); organisms with small genome, such as *Brachypodium dystachyon* or *Vitis vinifera*, are the hardest examples for PCR marker development (Kalendar, Schulman, 2014).

It has been proven that TE families evolve with different profiles, so TE marker systems based on different TEs show different levels of resolution and can be chosen to fit with the required analysis (Leigh et al., 2003; Kalendar, Schulman, 2006; Smykal et al., 2009; Hosid et al., 2012). Retrotransposons insertions behave as Mendelian loci (Manninen et al., 2006; Tanhuanpaa et al., 2008). Thus, retrotransposon-based markers would be expected to be co-dominant and involve a different level of genetic variability, i. e. transposition events, then arbitrary markers systems such as RAPD or AFLPs, which detect polymorphism from simple nucleotide changes to genomic rearrangements. Depending on method and primer combinations polymorphism detection tools can further be expanded knowing nearby TEs that are found in

different orientations in the genome (head-to-head, tail-to-tail, or head-to-tail).

PCR primers from one species can be used on others because related species have phylogenetically related TE sequences (retroelements or transposons), in which scenario primers designed to conservative TE sequences are advantageous. Being scattered at whole chromosome, TEs often are mixed with other elements and repeats; thus, PCR fingerprints can be improved if combination of PCR primers is used.

Following retrieval of LTR sequences of a selected family of retrotransposon, their alignment can be made to find out the most conserved region in them. The related plant species have conservative regions in LTR for identical retroelement; therefore, conservative regions can be identified through alignment of a few LTR sequences from one species or mixture with sequences from the related species (Kalendar et al., 2004; Yin et al., 2013; Moisy et al., 2014). The conservative parts of retrotransposon regions are used for the design of inverted primers for long distance PCR, for cloning of whole element and also for IRAP techniques.

Most of the retrotransposon techniques are anonymous, producing fingerprints from multiple sites of retrotransposon insertion in the genome. All of them use the combination of a known retrotransposon sequence and a variety of adjacent sequences. Target for primers are usually designed for LTRs near to the joint, in domains that are conserved within families but that differ between families. Despite regions internal to the LTR containing conserved segments could be used for this purpose, commonly to minimize size of the target to be amplified choice falls on LTR. Primer facing outward from the left or 5′ LTR will necessarily face inward from the right, or 3′ LTR because of LTRs being direct repeats. Depending on the nature of the second primer, the inward facing primer will either not amplify a product, produce a monomorphic band, or will detect polymorphism resulting from a nested insertion pattern. By using infrequent cutting enzyme, removal of internal amplicon can be done. To simplify the process, transposon specific primer can be obtained from an internal sequence present only once per element for retrotransposons with relatively short LTRs. Also, simplified digestion and amplification protocols can be used for S-SAP (sequence-specific amplified polymorphism) with low copy number elements. S-SAP is a modified AFLP method based on BARE-1 retroelement. The core of this method is shredding genomic DNA by using two different enzymes that produce a template for the specific primer PCR: amplification between retrotransposon and adaptors ligated at restriction sites (usually *MseI* and *PstI* or any other restriction enzyme) using selective bases in the adaptor primer. Normally LTR regions is the site where primers are produced; however, in some cases it can take place at internal part of the element, such as polypurine tract (PPT) that is located internal to the 3′-LTR in retrotransposons.

Generally, compared to AFLP, S-SAP displays more polymorphism, more chromosomal distribution, and more codominance, but in order to provide sites for adapter ligation as in AFLP method, for S-SAP method restriction digestion of genomic DNA is required. False genotyping results could be caused by sensitivity of commonly used restriction enzymes to DNA methylation. When the same technique used for retrotransposons is applied to DNA transposons, it is named transposon display (TD) (Van den Broeck et al., 1998). In the *Oryza* genus, Rim2/Hipa-TD produced highly polymorphic profiles with ample reproducibility within a species as well as between species (Kwon et al., 2005).

IRAP and REMAP methods have been used in gene mapping in cereals (Manninen et al., 2000; Boyko et al., 2002; Tanhuanpaa et al., 2007), in studies of genome evolution in the Solanaceous crop species (Tek et al., 2005; Tam et al., 2009), in a variety of applications, including measurement of genetic diversity and population structure, chromatin modification and epigenetic reprogramming, similarity and cladistic relationships, determination of essential derivation, marker-assisted selection, detection of somaclonal variation (Ishizaki, Kato, 2005; Lightbourn et al., 2007; Tam et al., 2009; Belyayev et al., 2010; Smykal et al., 2011; Wenke et al., 2011; Hosid et al., 2012; Yin et al., 2013; Moisy et al., 2014; Sharma, Nandineni, 2014; Tang et al., 2014; Paz et al., 2015).

Generation of virtually unlimited number of unique markers are gained through combination of different LTR primers or with combinations with microsatellite primers (REMAP). Same primers depending if used alone or in combination produce completely different banding patterns, demonstrating most of IRAP/REMAP bands were derived from sequences bordered by other LTR or a microsatellite on one side, and by an LTR on the other. In general, more variable pattern was shown in REMAP than in ISSR; also, frequently, but not always, depending on LTR sequence, single priming PCR show less variability than IRAP pattern with primer combinations.

LTR amplification technique was derived to reach quick, robust and economic marker system for genotyping in plant breeding and marker-assisted selection (Tam et al., 2009). Genetically inherited retrotransposon families can serve as markers that can ultimately protect the rights of breeders. The pattern obtained will be related to the TE copy number, insertion pattern and size of the TE family. Amplification of a series of bands (DNA fingerprints) using primers homologous to these high copy number repeats is achievable because of association of these sequences with each other and produced markers are very informative genetic markers (Yin et al., 2013).

Transposon display has been used also as a sensitive method for detecting genomic copies of retrotransposons amidst retrotransposon cDNAs (Jaaskelainen et al., 1999), in detecting cDNA polymorphism and clonal differences resulting from retrotransposon activities or retrotransposon recombination after crossing-over (Kalendar et al., 2010; Monden et al., 2014). LTR amplification technique displays is also efficient in examine genome structure and evolution in Solanaceous crop species, and in chromosome structure and transmission (Manetti et al., 2007, 2009; Novakova et al., 2009; Park et al., 2012; Michael, 2014; Na et al., 2014; Tang et al., 2014).

Insertion polymorphism of active retrotransposon families (Rtsp-1 and Lib) was used for DNA fingerprinting in sweet potato (*Ipomoea batatas*). Constructed phylogenetic tree using these insertion sites showed strong correspondence with pedigree information, proving this method could be utilized for genetic diversity studies. Thus, without a need for whole genome sequence information genome-wide comparative analysis of active retrotransposon insertion sites is effective approach for DNA fingerprinting. This method could

facilitate development of cultivar diagnostic system based on PCR and determination of genetic relationships (Monden et al., 2014).

Due to abundance of SINE repetitive sequences in almost all plant genomes, it can be effectively used for genotyping (Wenke et al., 2011; Seibt et al., 2012). Potato served as a sample plant to develop ISAP method, and it is also possible to apply this method on another species (Seibt et al., 2012; Wenke et al., 2015). Two selected SINE families, SolS-IIIa and SolS-IV, were shown to be highly but differently amplified in Solanaceae, *Solaneae* tribe, including wild and cultivated potatoes, tomato, and eggplant. Genome-wide distribution of SolS-IIIa and SolS-IV along potato chromosomes, which is the basis for genotype discrimination and differentiation of somaclonal variants by ISAP markers, was shown through fluorescent *in situ* hybridization (Seibt et al., 2012). Study of activity of retrotransposons in inter- and intraspecific hybrids between *Solanum kurtzianum* and *S. microdontum* observed that at morphological level intraspecific hybrids' genotypes remained same as their parents', while genotypes of interspecific hybrids have been altered. Analysis of genotypes showed mobility of both retrotransposons (Tto1 and Tnt1) used, ranging from 0 to 7.8 %. In comparison to their parental genotype, hybrids were epigenetically changed by demethylation in the vicinity of Tnt1 and Tto1, which correlates with the activity of retrotransposons. Those results indicate that retrotransposon activation can lead to genetic variability in tuber-bearing species of *Solanum* via hybridization (Paz et al., 2015).

Retrotransposons and transcriptional variability of genome of *Solanaceous* crop species

Two most important Solanaceae species from *Solaneae* tribe, *Solanum tuberosum* and *S. lycopersicum*, are almost fully sequenced with approximately 85 % of *S. tuberosum* genome and 95 % of *S. lycopersicum* genome being known to this day. Out of total genome size of 810.6 MB, 15.78 % (127,958,425 bp) of *S. tuberosum* has yet to be sequenced, and as for *S. lycopersicum*, 5 % (44,030,063 bp) out of 781.6 MB isn't still sequenced (Mehra et al., 2015). Many plant species contain repetitive elements in their genome, sometimes reaching 80 %, such as wheat. As for *S. tuberosum* and *S. lycopersicum*, ~49 % and ~60 % of their genome is comprised of repetitive elements, respectively. Among other genetic elements, there are 629,713 complex repetitive elements in *S. tuberosum* and 589,561 in *S. lycopersicum*, with both having chromosome 12 as most repeat rich. Among identified repeat families, DNA transposons and retrotransposons are included (Tang et al., 2014). There are increasing number of reports suggesting repetitive elements carry important functions in the genome, for instance, it was detected that they're abundant in gene-coding region. Many repetitive elements have been detected upstream of protein coding genes – regulatory regions. Some are found in introns, where they become exonized or domesticated. Most prevalent repetitive super family in both species is LTR/Gypsy, with 334,474 and 306,511 repetitive elements in *S. tuberosum* and *S. lycopersicum*, respectively. Following LTR Gypsy, other super families, such as LTR Copia and LINE elements L1, occupy much of both species' genome. Thus, LTR is the most abundant complex repetitive element in *S. tuberosum* and *S. lycopersicum*. Twelve chromosomes of both species are highly syntenic to each other, having high similarity for genes and repeats distribution (Mehra et al., 2015).

The study of genic regions has shown that, 99.29 % (38,740 genes out of 39,021 genes) of *S. tuberosum* genes had repeats overlapping either with their coding sequence or with the 5 kb upstream region, whereas in *S. lycopersicum* 98.92 % (34,303 genes out of 34,675 genes) genes had repeats overlapping with their genic regions and/or with the 5 kb upstream region. These results suggest that in both species repetitive sequences have impact on majority of protein coding genes. Analysis of various repeat families (LINE elements RTE-BovB, SINE elements and DNA transposon Stowaway) in *S. tuberosum* indicated big portion of repetitive elements are located within genic regions, as well as in *S. lycopersicum*, where the DNA transposons hAT-Tag1, hAT-Tip100, PIF-Harbinger, RC Helitron and SINEs were found in abundance in genic region, while genic region showed notable preference for LTR/ERV1 repeat family (Mehra et al., 2015).

Introns with repetitive elements have shown to impact the spatio-temporal expression of genes, creation of cryptic splices sites and other effects, while insertion of repetitive elements is thought to be more destructive, and associated with many disease conditions. Thus, detailed study of insertion of repetitive elements is crucial to further understand mechanism that lets repetitive elements influence genes and their products. In *S. tuberosum*, insertion preference of repeat families in either exonic or intronic regions wasn't noticed, whereas in *S. lycopersicum* DNA transposon MULE/MuDR and LTR/ERV1 prevailed in exonic region and DNA transposons, TcMar-Stowaway, LINE elements RTE-BovB and SINE elements accumulated in intronic regions.

Epigenetic control and retrotransposon activity

Methylation status of TEs in plants was correlated with lower transcription of genes with TE insertions. Also, more systematic knowledge about the influence of stress or environmental cues on epigenetic control of retrotransposons as well as impact of TEs on phenotypic plasticity is still unclear (Hollister, Gaut, 2009). The stochastic and sometime incomplete nature of epigenetic silencing of retrotransposons may help explain stress surviving, heterosis and the genome dominance phenomenon for intraspecific cross hybrids. Repetitive element mobilization represents a destabilizing process for the host cell. Several mechanisms such as DNA and histone methylation and RNAi, actively suppress retrotransposon expression (Vetukuri et al., 2011). The epigenetic mechanisms controlling retroelements may well follow retrotransposons during their movement 'around' the genome and thereby modify the epigenetic control of retrotransposition targeted loci.

In the plant genome, insertional inactivation and other genome rearrangements lead to a wide spectrum of recombination and chromosomal instability (Belyayev et al., 2010). Retroelement-induced genetic rearrangements can lead to non-allelic homologous recombination or insertional mutagenesis due to the 'hopping' of retrotransposons within gene coding sequences; it causes diverse effects on target gene expression depending on intragenic location, orientation, length of the inserted sequence and other factors, or activation mobilization of small RNAs.

Studies have suggested repetitive elements cause speciation through regulatory variability. It was found that transcription factors (TFs) that partake in main metabolic pathways and defence response are associated with repetitive elements. In *S. tuberosum*, out of total binding sites of I-box, member of Myb-group of transcription factors, gained/lost in the 2 kb upstream region of the genes, around 36 % were found to be overlapping with repetitive elements. I-box promoter motif is considered to be involved in response mechanism to light and another TF SORLIP2 (sequences over-represented in light-induced promoters – SORLIPs) has been linked to light-induced genes in some plants, which was found to have significant gain of their TF binding sites in the orthologous genes of *S. tuberosum* with ~23 % of them occurring in repetitive region. Similarly, other TFs, such as G-Box and MADS, were associated with repetitive regions. Functions of these TFs include response to stress, light, abscisic acid, and other metabolites as well as taking part in development processes (flower development and gametophyte, embryo and seed development). Above examples indicate integration of repetitive sequences in plants is beneficial for their survival purposes, as opposed to initial speculations of them being "junks" (Mehra et al., 2015).

In addition, repetitive elements have been found to be associated with miRNAs, small non-coding RNAs responsible for regulation of 60–70 % genes in an organism. Examining multiple loci of miRNA in both species showed that most of them were overlapping with different repetitive elements, 242 and 77 miRNA in *S. tuberosum* and *S. lycopersicum*, respectively. Most prevalent repeat families were found to be LTR/Gypsy in *S. tuberosum*, and DNA transposons in *S. lycopersicum*. By performing binomial test, probability of miRNA enrichment was calculated, where *p*-values of *S. tuberosum* ($4.136e^{-10}$) and *S. lycopersicum* ($1.819e^{-12}$) were obtained. Calculated numbers indicate miRNAs were enriched around repetitive elements.

Another small non-coding RNAs repetitive elements have been associated with are siRNA, which can silence repetitive elements through post transcriptional gene silencing mechanisms (PTGS) by creating feed-back-loop. Transcriptional activity of repetitive elements besides controlling repetitive elements, can as well provide tissue specific expression of certain genes (Mehra et al., 2015). A fair number of expressive repetitive elements have been linked to small RNAs/siRNA biogenesis, some of which are involved in gene regulation in either *cis* or *trans* manner. While some sRNAs partake in post transcriptional gene silencing, other such RNAs are involved in *de novo* DNA methylation in plant genome. With increasing number of reports, sRNAs are now thought to be core members of post transcriptional as well as RdDM based transcriptional gene regulatory processes. Involvement of repetitive elements in biogenesis of sRNAs indicate their importance in gene regulatory system of *Solanum* species.

Perspectives and implications

Many features of retrotransposons, such as ubiquity and dispersion in eukaryotic genome, make them appealing as the basis of molecular marker systems. Because of their repetitive nature, retrotransposons are a source of chromatin instability and genomic rearrangements with deleterious consequences (Belyayev et al., 2010). Newly inserted retrotransposons created instability and influence gene expression of flanking regions by modifying their methylation status. Retrotransposons can also impact gene regulation simply by inserting their own internal regulatory sequences (promoters, enhancers) in new genomic loci upon retrotransposition. A high proportion of the retroelements have lost their autonomous transposition ability, either by point mutations and/or deletions, many of them seem to embody defective elements with deletions.

Genome diversification results from their past activity and by recombination events, which provides a means of its detection. Their integration can be detected by conserved sequences. Retrotransposons are long and produce a large genetic change at the point of insertion, thereby providing conserved sequences that can be used to detect their own integration. This event isn't related to deletion of the transposable element from another locus, as it is for DNA transposons. Even the loss of the core domain of a retrotransposon by LTR-LTR recombination is invisible to the marker methods using outward-facing LTR primers. The ancestral state of a retrotransposon insertion is obvious – it is the empty site, which is very useful in pedigree and phylogenetic analyses. Original empty sites are unlikely to be regenerated by later recombination processes at a full site. Retroelements were used to clarify the relationships between related species.

Previously mentioned DNA markers based on LTR retrotransposons are usually referred to as "transposon display". The applications range from investigations of retrotransposon activation and mobility to studies of biodiversity, genome evolution, chromatin modification, epigenetic reprogramming, mapping of genes and the estimation of genetic distance, to assessment of essential derivation of varieties, detection of somaclonal variation and cDNA fingerprinting. Only those retrotransposon insertions are useful, which are passed into the egg cells and pollen. Thus, they could possibly be considered as sexually transmitted diseases, but that moves by a cellular, rather than extracellular, pathway into the new host.

The utility of LTR-retrotransposon-based markers, not only for genetic analysis and map construction, in addition also for the isolation and characterization of LTR retrotransposons, such as the long terminal repeats or the internal genes they contain.

In plants, analogous approaches have been adopted to the non-LTR retrotransposons, specifically to SINE elements. The insertion pattern of the human Alu, a SINE and the most prevalent transposable element in the human genome, were not only used for research on human population structure, but as well as in studies of heritable diseases. In essence, effective application of retrotransposon- or endogenous retrovirus-based molecular markers could be established for use on animals, including mammals and birds.

Platforms for commercial next-generation DNA sequencing techniques (NGS) of have been developed and found wide range of use for major crops, domestic animals, and humans. Abundance of sequence data is crucial for the development of new molecular markers. While genetic analysis by shotgun sequencing appears to be a promising method, cost is still the limit; therefore, cheap, generic, easily applied retrotransposon marker systems will stay as most applicable method for the foreseeable future.

Conflict of interest

The authors declare no conflict of interest.

References

Belyayev A., Kalendar R., Brodsky L., Nevo E., Schulman A.H., Raskina O. Transposable elements in a marginal plant population: temporal fluctuations provide new insights into genome evolution of wild diploid wheat. Mobile DNA. 2010;1(6):1-16. DOI 10.1186/1759-8753-1-6.

Boyko E., Kalendar R., Korzun V., Fellers J., Korol A., Schulman A.H., Gill B.S. A high-density cytogenetic map of the *Aegilops tauschii* genome incorporating retrotransposons and defense-related genes: insights into cereal chromosome structure and function. Plant Mol. Biol. 2002;48(5):767-790. DOI 10.1023/A:1014831511810.

Bureau T.E., Wessler S.R. Tourist: a large family of small inverted repeat elements frequently associated with maize genes. Plant Cell. 1992;4(10):1283-1294. DOI 10.1105/tpc.4.10.1283.

Charlieu J.P., Laurent A.M., Carter D.A., Bellis M., Roizes G. 3′ Alu PCR: a simple and rapid method to isolate human polymorphic markers. Nucl. Acids Res. 1992;20(6):1333-1337. DOI 10.1093/nar/20.6.1333.

Finnegan D.J. Transposable elements and DNA transposition in eukaryotes. Curr. Opin. Cell Biol. 1990;2(3):471-477. DOI 10.1016/0955-0674(90)90130-7.

Flavell A.J., Dunbar E., Anderson R., Pearce S.R., Hartley R., Kumar A. Ty1-copia group retrotransposons are ubiquitous and heterogeneous in higher plants. Nucl. Acids Res. 1992;20(14):3639-3644. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/1379359.

Frankel A.D., Young J.A. HIV-1: fifteen proteins and an RNA. Annu. Rev. Biochem. 1998;67:1-25. DOI 10.1146/annurev.biochem.67.1.1.

Goke J., Ng H.H. CTRL+INSERT: retrotransposons and their contribution to regulation and innovation of the transcriptome. EMBO Rep. 2016;17(8):1131-1144. DOI 10.15252/embr.201642743.

Hollister J.D., Gaut B.S. Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome Res. 2009;19(8):1419-1428. DOI 10.1101/gr.091678.109.

Hosid E., Brodsky L., Kalendar R., Raskina O., Belyayev A. Diversity of long terminal repeat retrotransposon genome distribution in natural populations of the wild diploid wheat *Aegilops speltoides*. Genetics. 2012;190(1):263-274. DOI 10.1534/genetics.111.134643.

Ishizaki T., Kato A. Introduction of the tobacco retrotransposon Tto1 into diploid potato. Plant Cell Rep. 2005;24(1):52-58. DOI 10.1007/s00299-005-0919-y.

Jaaskelainen M., Mykkanen A.H., Arna T., Vicient C.M., Suoniemi A., Kalendar R., Savilahti H., Schulman A.H. Retrotransposon BARE-1: expression of encoded proteins and formation of virus-like particles in barley cells. Plant J. 1999;20(4):413-422. DOI 10.1046/j.1365-313x.1999.00616.x.

Kalendar R. The use of retrotransposon-based molecular markers to analyze genetic diversity. Ratarstvo i Povrtarstvo. 2011;48(2):261-274. DOI 10.5937/ratpov1102261K.

Kalendar R., Antonius K., Smýkal P., Schulman A.H. iPBS: a universal method for DNA fingerprinting and retrotransposon isolation. Theor. Appl. Genet. 2010;121(8):1419-1430. DOI 10.1007/s00122-010-1398-2.

Kalendar R., Flavell A.J., Ellis T.H., Sjakste T., Moisy C., Schulman A.H. Analysis of plant diversity with retrotransposon-based molecular markers. Heredity. 2011;106(4):520-530. DOI 10.1038/hdy.2010.93.

Kalendar R., Schulman A. IRAP and REMAP for retrotransposon-based genotyping and fingerprinting. Nature Protocols. 2006;1(5):2478-2484. DOI 10.1038/nprot.2006.377.

Kalendar R., Schulman A.H. Transposon-based tagging: IRAP, REMAP, and iPBS. Methods Mol. Biology. 2014;1115:233-255. DOI 10.1007/978-1-62703-767-9_12.

Kalendar R., Tanskanen J., Chang W., Antonius K., Sela H., Peleg O., Schulman A.H. *Cassandra* retrotransposons carry independently transcribed 5S RNA. Proc. Natl. Acad. Sci. USA. 2008;105(15):5833-5838. DOI 10.1073/pnas.0709698105.

Kalendar R., Vicient C.M., Peleg O., Anamthawat-Jonsson K., Bolshoy A., Schulman A.H. Large retrotransposon derivatives: abundant, conserved but nonautonomous retroelements of barley and related genomes. Genetics. 2004;166(3):1437-1450. DOI 10.1534/genetics.166.3.1437.

Kan Y.W., Dozy A.M. Polymorphism of DNA sequence adjacent to human beta-globin structural gene: relationship to sickle mutation. Proc. Natl. Acad. Sci. USA. 1978;75(11):5631-5635. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC393021/.

Kwon S.J., Park K.C., Kim J.H., Lee J.K., Kim N.S. *Rim 2*/*Hipa* CACTA transposon display: a new genetic marker technique in *Oryza* species. BMC Genet. 2005;6:15. DOI 10.1186/1471-2156-6-15.

Leigh F., Kalendar R., Lea V., Lee D., Donini P., Schulman A.H. Comparison of the utility of barley retrotransposon families for genetic analysis by molecular marker techniques. Mol. Genet. Genomics. 2003;269(4):464-474. DOI 10.1007/s00438-003-0850-2.

Lewontin R.C., Hubby J.L. A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. Genetics. 1966;54(2):595-609. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/5968643.

Lightbourn G.J., Jelesko J.G., Veilleux R.E. Retrotransposon-based markers from potato monoploids used in somatic hybridization. Genome. 2007;50(5):492-501. DOI 10.1139/g07-026.

Macas J., Kejnovsky E., Neumann P., Novak P., Koblizkova A., Vyskot B. Next generation sequencing-based analysis of repetitive DNA in the model dioecious [corrected] plant *Silene latifolia*. PLoS ONE. 2011;6(11):e27335. DOI 10.1371/journal.pone.0027335.

Macas J., Novak P., Pellicer J., Cizkova J., Koblizkova A., Neumann P., Fukova I., Dolezel J., Kelly L.J., Leitch I.J. In depth characterization of repetitive DNA in 23 plant genomes reveals sources of genome size variation in the legume tribe *Fabeae*. PLoS ONE. 2015;10(11):e0143424. DOI 10.1371/journal.pone.0143424.

Manetti M.E., Rossi M., Costa A.P., Clausen A.M., Van Sluys M.A. Radiation of the Tnt1 retrotransposon superfamily in three Solanaceae genera. BMC Evol. Biol. 2007;7:34. DOI 10.1186/1471-2148-7-34.

Manetti M.E., Rossi M., Nakabashi M., Grandbastien M.A., Van Sluys M.A. The Tnt1 family member Retrosol copy number and structure disclose retrotransposon diversification in different *Solanum* species. Mol. Genet. Genomics. 2009;281(3):261-271. DOI 10.1007/s00438-008-0408-4.

Manninen O., Jalli M., Kalendar R., Schulman A., Afanasenko O., Robinson J. Mapping of major spot-type and net-type net-blotch resistance genes in the Ethiopian barley line CI 9819. Genome. 2006;49(12):1564-1571. DOI 10.1139/g06-119.

Manninen O., Kalendar R., Robinson J., Schulman A. Application of BARE-1 retrotransposon markers to the mapping of a major resistance gene for net blotch in barley. Mol. Gener. Genetics. 2000;264(3):325-334. DOI 10.1007/s004380000326.

Mehra M., Gangwar I., Shankar R. A deluge of complex repeats: the *Solanum* genome. PLoS ONE. 2015;10(8):e0133962. DOI 10.1371/journal.pone.0133962.

Michael T.P. Plant genome size variation: bloating and purging DNA. Brief Funct. Genomics. 2014;13(4):308-317. DOI 10.1093/bfgp/elu005.

Mita P., Boeke J.D. How retrotransposons shape genome regulation. Curr. Opin. Genet. Dev. 2016;37:90-100. DOI 10.1016/j.gde.2016.01.001.

Moisy C., Schulman A.H., Kalendar R., Buchmann J.P., Pelsy F. The Tvv1 retrotransposon family is conserved between plant genomes separated by over 100 million years. Theor. Appl. Genet. 2014;127(5):1223-1235. DOI 10.1007/s00122-014-2293-z.

Monden Y., Yamamoto A., Shindo A., Tahara M. Efficient DNA fingerprinting based on the targeted sequencing of active retrotransposon insertion sites using a bench-top high-throughput sequencing platform. DNA Res. 2014;21(5):491-498. DOI 10.1093/dnares/dsu015.

Na J.K., Wang J., Ming R. Accumulation of interspersed and sex-specific repeats in the non-recombining region of papaya sex chromosomes. BMC Genomics. 2014;15:335. DOI 10.1186/1471-2164-15-335.

Novakova A., Simackova K., Barta J., Curn V. Potato variety identification by molecular markers based on retrotransposon analyses. Czech J. Genet. Plant Breeding. 2009;45(1):1-10. Retrieved from <Go to ISI>://WOS:000264852600001.

Park M., Park J., Kim S., Kwon J.K., Park H.M., Bae I.H., Yang T.J., Lee Y.H., Kang B.C., Choi D. Evolution of the large genome in *Capsicum annuum* occurred through accumulation of single-type long terminal repeat retrotransposons and their derivatives. Plant. J. 2012;69(6):1018-1029. DOI 10.1111/j.1365-313X.2011.04851.x.

Paz R.C., Rendina González A.P., Ferrer M.S., Masuelli R.W. Short-term hybridisation activates Tnt1 and Tto1 Copia retrotransposons in wild tuber-bearing *Solanum* species. Plant Biol. 2015;17(4):860-869. DOI 10.1111/plb.12301.

Pearce S.R., Harrison G., Li D., Heslop-Harrison J., Kumar A., Flavell A.J. The *Ty1-copia* group retrotransposons in *Vicia* species: copy number, sequence heterogeneity and chromosomal localisation. Mol. Gener. Genetics. 1996;250(3):305-315. DOI 10.1007/BF02174388.

Seibt K.M., Wenke T., Wollrab C., Junghans H., Muders K., Dehmer K.J., Diekmann K., Schmidt T. Development and application of SINE-based markers for genotyping of potato varieties. Theor. Appl. Genet. 2012;125(1):185-196. DOI 10.1007/s00122-012-1825-7.

Sharma V., Nandineni M.R. Assessment of genetic diversity among Indian potato (*Solanum tuberosum* L.) collection using microsatellite and retrotransposon based marker systems. Mol. Phylogenet. Evol. 2014;73:10-17. DOI 10.1016/j.ympev.2014.01.003.

Smykal P., Bacova-Kerteszova N., Kalendar R., Corander J., Schulman A.H., Pavelek M. Genetic diversity of cultivated flax (*Linum usitatissimum* L.) germplasm assessed by retrotransposon-based markers. Theor. Appl. Genet. 2011;122(7):1385-1397. DOI 10.1007/s00122-011-1539-2.

Smykal P., Kalendar R., Ford R., Macas J., Griga M. Evolutionary conserved lineage of *Angela*-family retrotransposons as a genome-wide microsatellite repeat dispersal agent. Heredity. 2009;103(2):157-167. DOI 10.1038/hdy.2009.45.

Tam S.M., Lefebvre V., Palloix A., Sage-Palloix A.M., Mhiri C., Grandbastien M.A. LTR-retrotransposons Tnt1 and T135 markers reveal genetic diversity and evolutionary relationships of domesticated peppers. Theor. Appl. Genet. 2009;119(6):973-989. DOI 10.1007/s00122-009-1102-6.

Tang X., Datema E., Guzman M.O., de Boer J.M., van Eck H.J., Bachem C.W., Visser R.G., de Jong H. Chromosomal organizations of major repeat families on potato (*Solanum tuberosum*) and further exploring in its sequenced genome. Mol. Genet. Genomics. 2014;289(6):1307-1319. DOI 10.1007/s00438-014-0891-8.

Tanhuanpaa P., Kalendar R., Schulman A.H., Kiviharju E. A major gene for grain cadmium accumulation in oat (*Avena sativa* L.). Genome. 2007;50(6):588-594. DOI 10.1139/g07-036.

Tanhuanpaa P., Kalendar R., Schulman A.H., Kiviharju E. The first doubled haploid linkage map for cultivated oat. Genome. 2008;51(8):560-569. DOI 10.1139/G08-040.

Tek A.L., Song J., Macas J., Jiang J. Sobo, a recently amplified satellite repeat of potato, and its implications for the origin of tandemly repeated sequences. Genetics. 2005;170(3):1231-1238. DOI 10.1534/genetics.105.041087.

Van den Broeck D., Maes T., Sauer M., Zethof J., De Keukeleire P., D'Hauw M., Van Montagu M., Gerats T. *Transposon Display* identifies individual transposable elements in high copy number lines. Plant J. 1998;13(1):121-129. DOI 10.1046/j.1365-313X.1998.00004.x.

Vetukuri R.R., Tian Z., Avrova A.O., Savenkov E.I., Dixelius C., Whisson S.C. Silencing of the *PiAvr3a* effector-encoding gene from *Phytophthora infestans* by transcriptional fusion to a short interspersed element. Fungal Biol. 2011;115(12):1225-1233. DOI 10.1016/j.funbio.2011.08.007.

Vicient C.M., Jaaskelainen M.J., Kalendar R., Schulman A.H. Active retrotransposons are a common feature of grass genomes. Plant Physiol. 2001;125(3):1283-1292. DOI 10.1104/pp.125.3.1283.

Voytas D.F., Cummings M.P., Koniczny A., Ausubel F.M., Rodermel S.R. *copia*-like retrotransposons are ubiquitous among plants. Proc. Natl. Acad. Sci. USA. 1992;89(15):7124-7128. DOI 10.1073/pnas.89.15.7124.

Wenke T., Dobel T., Sorensen T.R., Junghans H., Weisshaar B., Schmidt T. Targeted identification of short interspersed nuclear element families shows their widespread existence and extreme heterogeneity in plant genomes. Plant Cell. 2011;23(9):3117-3128. DOI 10.1105/tpc.111.088682.

Wenke T., Seibt K.M., Dobel T., Muders K., Schmidt T. Inter-SINE Amplified Polymorphism (ISAP) for rapid and robust plant genotyping. Methods Mol. Biol. 2015;1245:183-192. DOI 10.1007/978-1-4939-1966-6_14.

Yin H., Liu J., Xu Y., Liu X., Zhang S., Ma J., Du J. *TARE1*, a mutated *Copia*-like LTR retrotransposon followed by recent massive amplification in tomato. PLoS ONE. 2013;8(7):e68587. DOI 10.1371/journal.pone.0068587.