

ОГЛЯДИ

UDC 577.2.08

R. N. KALENDAR

PhD, Head of the laboratory, RSE «National Center for Biotechnology»,
Kazakhstan

e-mail: ruslan.kalendar@biocenter.kz

RETROTRANSPOSON-BASED MOLECULAR MARKERS

Molecular markers play an essential role in all aspects of genetics, modern plant breeding, in human forensics, for map-based cloning of genes, ranging from the identification of genes responsible for desired traits to the management of backcrossing programs. Retrotransposons are ideally suited as molecular markers. As dispersed and ubiquitous transposable elements, their «copy and paste» life cycle of replicative transposition leads to new genome insertions without excision of the original element. Both the overall structure of retrotransposons and the domains responsible for the various steps of their replication are highly conserved in all eukaryotes. Following the demonstration that retrotransposons are ubiquitous, active, and abundant in plant genomes, various marker systems were developed to exploit polymorphisms in retrotransposon insertion patterns. This review provides an insight into the spectrum of retrotransposon-based marker systems developed for plant species, evaluates the contributions of retrotransposon markers to the analysis of population diversity in plants and the way for the rapid isolation of retrotransposon termini.

Key words: molecular marker, phylogeny, polymorphism, repeats, retrotransposon, transposable element.

Introduction. Markers play an essential role in the study of genetically variability and diversity, in the construction of linkage maps, and in the diagnosis of individuals or lines carrying certain linked genes. The emergence of marker systems has, for the last 40 years [1], closely tracked developments in biochemistry and molecular biology. The shortcomings of biochemically-based markers, such as isozymes, drove the development of markers based on DNA polymorphisms [2]. These marker types generate «fingerprints,» distinctive patterns of DNA fragments resolved by electrophoresis and detected by staining or labelling. A molecular marker in essence detects nucleotide sequence variation at a particular location in the genome. This nucleotide sequence must be different between the parents of the chosen cross to be distinguishable between plant accessions and to finally study its pattern of inheritance. The advent of the polymerase chain reaction (PCR) was a break-

through for molecular marker technologies, and made possible many fingerprinting methods. These fall into two broad categories, namely methods that detect single loci and multiplex methods that detect multiple loci simultaneously.

Interspersed repetitive sequences comprise a large fraction of the genome of many eukaryotic organisms and they are predominantly comprised of transposable elements (TEs). In most species that have been studied interspersed repeats are distributed unevenly across the nuclear genome and some repeats have a tendency to cluster around the centromeres or telomeres. Following the induction of recombinational processes during the meiotic prophase variation in the copy number of repeat elements and internal rearrangements on both homologous chromosomes can ensue. The resulting heterogeneity in the arrangement of distinguishable repeats has been exploited for specific molecular markers technique targeted this repeat element.

It is therefore not surprising that many DNA marker techniques that are based on these repeats have been devised.

Numerous methods have been developed that exploit repeated sequences as molecular markers. In an early example, Restriction Fragment Length Polymorphism (RFLP), probes derived from repetitive sequences were hybridized to Southern blots of restriction-digested genomic DNA to produce a highly variable pattern [3]. The RFLP technique was used extensively in the past, but has been replaced by PCR-based methods due to the slowness of Southern blotting.

Alternatively, repeats can be used as single primers in the polymerase chain reaction. The first multiplex methods to be developed were named Randomly Amplified Polymorphic DNA (RAPD) [4, 5] and DNA Amplification Fingerprinting (DAF)[6] respectively, and involve amplification of random repetitive sites in the genome using short primers, typically 8–12 nt in length. The approaches involve quick and easy reaction set-up and no genome sequence information is needed to design the primers. However, problems in reproducibility due to the presence of huge numbers of potential priming sites in the genome and the low annealing temperatures in the reactions, derived from the nature of the primers themselves, have led to this method largely disappearing from the molecular marker toolkit today.

Nucleotide sequences matching repetitive sequences showing polymorphism in RFLP analyses have also been used as PCR primers for the inter-repeat amplification polymorphism marker method [7–9]. Such repetitive sequences include microsatellites, such as $(CA/GT)_n$ or $(CAC/GTG)_n$ which are distributed throughout the genome. A derived approach was developed to generate PCR markers based on amplification of microsatellites near the 3' end of the *Alu* (SINE) transposable elements (TEs), called *Alu*-PCR or SINE-PCR [10]. The successful application of microsatellite-specific oligonucleotides as PCR primers was first described by [7, 8, 11, 12] who amplified DNA from different sources, for example, with primers $(GATA)_n$, $(GACA)_n$.

Retrotransposons. Retrotransposons present one of the most fluid genomic components, varying greatly in copy number over relatively short evolutionary timescale and represent a major component of the structural evolution of plant genomes [13, 14, 15].

Retrotransposons are one of the two major groups of transposable elements in eukaryotic genomes and are defined according to their mode of propagation. Retrotransposons belongs to class I TEs and transpose via an RNA intermediate in contrast to other transposons (class II) that do not have an RNA intermediate (Fig. 1). Retrotransposons are separated in two major subclasses that differ in their structure and transposition cycle. These are the LTR retrotransposons and the non-LTR retrotransposons (long interspersed repetitive elements (LINE) and short interspersed nuclear elements (SINE), are distinguished by the respective presence or absence of long terminal repeats (LTRs) at their ends. All groups are complemented by their respective non-autonomous forms which 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 [16].

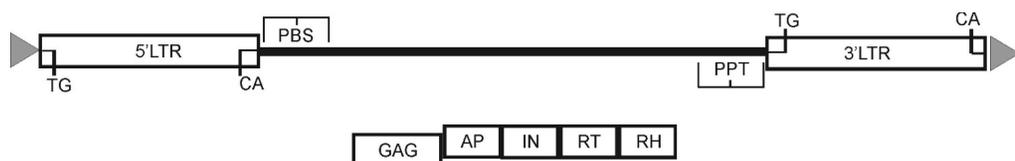


Fig. 1. Organization of an LTR retrotransposon. The retrotransposon is bounded by long terminal repeats (LTRs) which contain the transcriptional promoter and terminator. The LTRs contain short inverted repeats at either end, shown as filled triangles. Reverse transcription is primed at the PBS and PPT domains, respectively for the (–)– and (+)–strands of the cDNA. The internal region of the retrotransposon codes for the proteins necessary for the retrotransposon life cycle: the capsid protein (GAG), which packages the transcript into a virus-like particle; aspartic proteinase (AP), which cleaves the polyprotein (AP); integrase (IN), which inserts the cDNA copy into the genome; reverse transcriptase (RT) and RNaseH (RH), which together copy the transcript into cDNA. The GAG is often in a different reading frame, shown as a shifted box. The order of coding domains shown here is typical for *copia* – like elements; in *gypsy* – like elements, the IN domain is found at the 3' end of the internal domain. The internal region contains evolutionarily conserved domains, necessary for function that can be used to isolate retrotransposons from previously unstudied plant species. These are noted below the element as black boxes. The LTRs are generally well-conserved within families, and can serve for the design of primers to generate DNA footprints (Fig. 2). The areas of conservation suitable for primer design are noted as hatched boxes below the element diagram. Direct repeats in the flanking genomic DNA are generated upon retrotransposon integration; these are depicted as large arrow heads. The flanking genomic DNA is shown as a wavy line. The apposition of a long element bearing conserved sequences, with genomic DNA of random sequence is the basis for retrotransposon marker methods

LTR retrotransposons are transcribed from one LTR of an integrated element to produce a nearly full length RNA copy containing a single copy of the LTR split between its two ends (the LTR provides both the start site and polyadenylation signal for the element; Fig. 1, 2). This RNA is then reverse-transcribed into an extrachromosomal cDNA, reconstituting the full length element that is ultimately integrated back into the genome. Immediately internal to the LTRs are the priming sites for reverse transcription. The large central part of the retrotransposon encodes the structural components of a virus-like particle into which the RNA is inserted, together with reverse transcriptase and integrase enzymes.

Both the overall structural features as well as the basic stages of the life cycle are shared by the retrotransposons and the retroviruses [17, 18].

However, rather than escaping the genome to infect new individuals as do retroviruses, retrotransposons insert the new copies only into their host genomes. If the integration takes place within a cell lineage from which pollen or egg cells are ultimately derived, then a new polymorphism is contributed to the gene pool. These new copies are useful for distinguishing breeding lines, varieties, or populations of plants from each other.

In plants, the LTR retrotransposons are typically more plentiful and active than their non-LTR relatives. In many crop plants between 40–90 % of the total DNA is comprised of LTR retrotransposons [19–21]. Although most prevalent retrotransposons are dispersed throughout the genome, at least in the cereals and citrus they are often locally nested one into another and in extensive domains that have been referred to as «retrotransposon seas» surrounding gene islands [19–21]. Their abundance, general dispersion, and activity make them ideal sources for the development of molecular markers.

Retrotransposons as molecular markers. The emergence of retrotransposon-based methods followed the basic research that demonstrated their ubiquity and activity in the plants [14, 15, 22]. The most recent marker methods based on retrotransposons rely on PCR. Transposable elements have been exploited as molecular markers in various ways. For example, mammalian SINE like Alu-repeats are dispersed throughout their genomes. Primer sequences complementary to any of these repeats may produce many non-specific bands from single-primer amplification and be used as markers for detecting Alu-repeat polymorphisms [23, 24].

It has been proved 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 [25–32]. Retrotransposon insertions behave as Mendelian loci [33]. Hence, retrotransposon-based markers would be expected to be co-dominant and involve a different level of genetic variability, i.e. transposition events, than arbitrary markers systems such as RAPD or AFLPs, which detect polymorphism from simple nucleotide changes to genomic rearrangements. Nearby TEs may be found in different orientations in the genome (head-to-head, tail-to-tail, or head-to-tail) in-

creasing the range of tools to available to detect polymorphism depending on the method and primer combinations.

Most of the retrotransposon techniques are anonymous, producing fingerprints from multiple sites of retrotransposon insertion in the genome. They all exploit the combination of a known retrotransposon sequence and a variety of adjacent sequences. Primers are generally designed to the LTRs near to the joint, in domains that are conserved within families but that differ between families (Fig. 2). Although regions internal to the LTR that also contain conserved segments can be used for this purpose, generally the LTRs are chosen to minimize the size of the target to be amplified. Because the LTRs are direct repeats, a primer facing outward from the left or 5' LTR will necessarily face inward from the right, or 3' LTR.

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. The internal amplicon can also be removed by judicious use of an infrequent cutting enzyme. For retrotransposons with relatively short LTRs the transposon specific primer can derive from an internal sequence present only once per element, simplifying this process. For S-SAP with low copy number elements it is also possible to use simplified digestion and amplification protocols.

The various retrotransposon marker systems differ in the nature of the second primer used in the amplification reactions (Fig. 2). The second primer can be any feature in the genome that is dispersed and conserved.

S-SAP/TD. The Amplified Fragment Length Polymorphism method (AFLP), introduced in the mid 1990s, is an anonymous marker method. It detects restriction sites by amplifying a subset of all the sites for a given enzyme pair in the genome by PCR between ligated adapters. S-SAP (Sequence-Specific Amplified Polymorphism), a modified AFLP method based on BARE-1, was described by Waugh and co-authors in 1997 [34]. This method is based on the sharing of genomic DNA with two different enzymes to generate a template for the specific primer PCR: amplification between retrotransposon and adapters ligated at restriction sites (usually *MseI* and *PstI* or any other restriction enzyme) using selective bases in the adaptor primer. The S-SAP method can be considered as a modification of AFLP but it usually displays a higher level of polymorphism than AFLPs. Primers are usually designed in the LTR region, but could also correspond to internal part of the element, like to the polypurine tract (PPT) which is found internal to the 3'-LTR in retrotransposons.

Nonselective primers could be used when enzymes used for digestion have a larger recognition sequence, or when the copy number of the TE is lower. For high-copy-number families, the number of selective bases may be increased. The use of two enzymes in S-SAP corresponds to a reduction in genomic complexity as does the use of selective bases on the primers associated with the adapters. Low copy number TEs are not well suited to methods

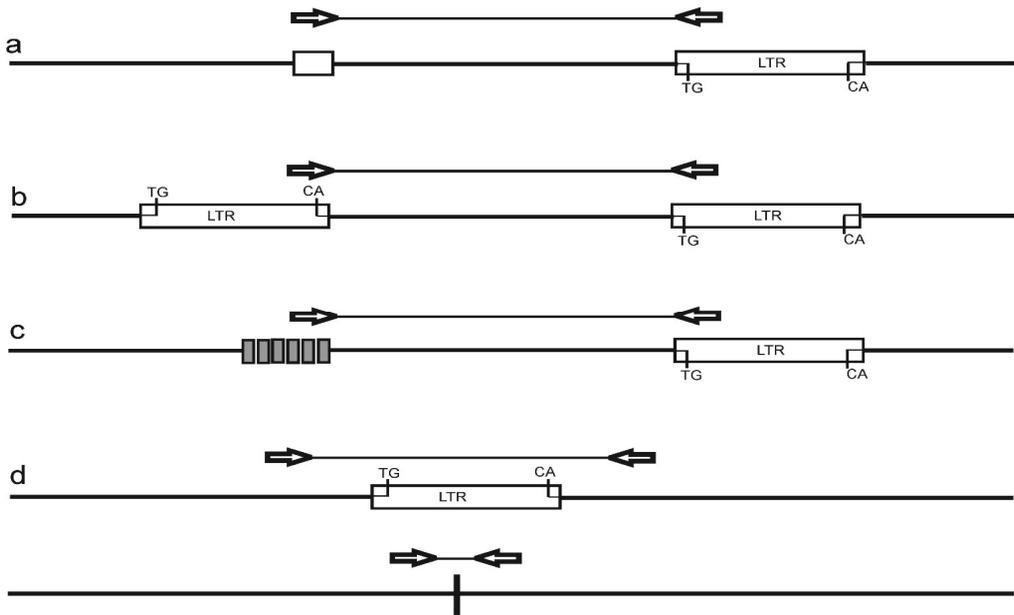


Fig. 2. Retrotransposon-based molecular marker methods. Multiplex products of various lengths from different loci are indicated by the bars beneath the diagrams of each reaction. **(a)** The SSAP method. The geometry of amplification from a DNA fragment cut with two restriction enzymes (E1, E2), containing a retrotransposon LTR (shaded, labeled) and ligated to an adapter (stippled) is shown. Primers used for amplification match the adapter (P_A) and retrotransposon (P_T). **(b)** The IRAP method. Amplification takes place between retrotransposons (internal regions hatched) near each other in the genome (open bar), using retrotransposon primers (P_T). The elements are shown oriented head-to-head, using a single primer. **(c)** The REMAP method. Amplification takes place between a microsatellite domain (vertical bars) and a retrotransposon, using a primer anchored to the proximal side of the microsatellite (P_M) and a retrotransposon primer (P_T). **(d)** RBIP. Full sites, depicted left, are scored by amplification between a primer in the flanking genomic DNA (here, primer P_{FL} matches the left flank) and a retrotransposon primer (P_T). The single product is shown as one bar beneath the diagram. The alternative reaction between the primers for the left and right flanks, P_{FL} and P_{FR} , respectively, is inhibited in the full site by the length of the retrotransposon. The product that is not amplified is indicated by the grey bar beneath the diagram. The flanking primers are able to amplify the empty site, right, depicted as a bar beneath the diagram

that involve such reduction in genomic complexity, the use of single enzyme digests with selective bases (or infrequent cutting enzymes) allows the survey of all insertion sites for a given TE, and can be considered as a variant of anchored PCR.

A S-SAP marker system based on three long terminal repeat (LTR) sequences of Ty1-copia retrotransposons displayed a higher level of polymorphism than AFLPs in cucumber (*Cucumis sativus* L.) [35]. S-SAP insertion patterns of 8 retrotransposon families on 10 *Vitis* accessions showed that these retrotransposon families are present across the *Vitis* genus and only a

few insertion sites are fixed in all accessions, which should have been maintained during speciation. Most of the scored bands are polymorphic, indicating that these families have been active after speciation across the genus [36]. S-SAP can be used mainly to measure the distribution and structure of specific retroelement populations in an organism. It has been used to evaluate the distribution and structure of specific retrotransposons populations in oat [37, 38], *Medicago sativa* L. [39], barley [25, 34], wheat *Triticum aestivum* and wild relatives [38, 40], sweet potato [41], pea [42], tomato [43], *Aegilops* species [44], apple [45], artichoke, lettuce [46, 47] and *Medicago sativa* [39]. S-SAP was also used to show evolutionary history in *Zea* [48], wheat [38] and in tobacco [49]. S-SAP has been optimised for multiple plant species and protocols for rapidly obtaining retrotransposon sequence information for S-SAP primer design have been developed [47].

The same technique was named Transposon Display (TD) when applied to DNA transposons rather than retrotransposons [50]. Rim2/Hipa-TD produced highly polymorphic profiles with ample reproducibility within a species as well as between species in the *Oryza* genus.

Usually, S-SAP shows more polymorphism, more co-dominance and more chromosomal distribution than AFLP. However, S-SAP also requires restriction digestion of genomic DNA to provide sites for adapter ligation as in AFLP method. Sensitivity of commonly used restriction enzymes to DNA methylation could provide false genotyping results.

IRAP/REMAP. In plants, the inter-repeat amplification polymorphism techniques such as inter-retrotransposon amplified polymorphism (IRAP), retrotransposon microsatellite amplification polymorphisms (REMAP) or inter-MITE amplification [32, 51- 53] have exploited the highly abundant dispersed repeats such as the LTRs of retrotransposons and SINE-like sequences. The association of these sequences with each other makes possible to amplify a series of bands (DNA fingerprints) using primers homologous to these high copy number repeats. The markers generated are very informative genetic markers. IRAP detects retrotransposon insertional polymorphisms by amplifying the portion of DNA between two retroelements [52]. One or two primers are used pointing outwards from an LTR, and therefore amplifies the tract of DNA between two nearby retrotransposons. IRAP can be carried out with a single primer matching either the 5' or 3' end of the LTR but oriented away from the LTR itself, or with two primers. The two primers may be from the same retrotransposon element family or may be from different families. The PCR products, and therefore the fingerprint patterns, result from amplification of hundreds to thousands of target sites in the genome (Fig.3). Retrotransposons generally tend to cluster together in «repeat seas» surrounding «genome islands», and may even nest within each other. Hence, the pattern obtained will be related to the TE copy number, insertion pattern and size of the TE family.

The REMAP (Retrotransposon-Microsatellite Amplified Polymorphism) method is similar to IRAP, but one of the two primers matches a microsatellite

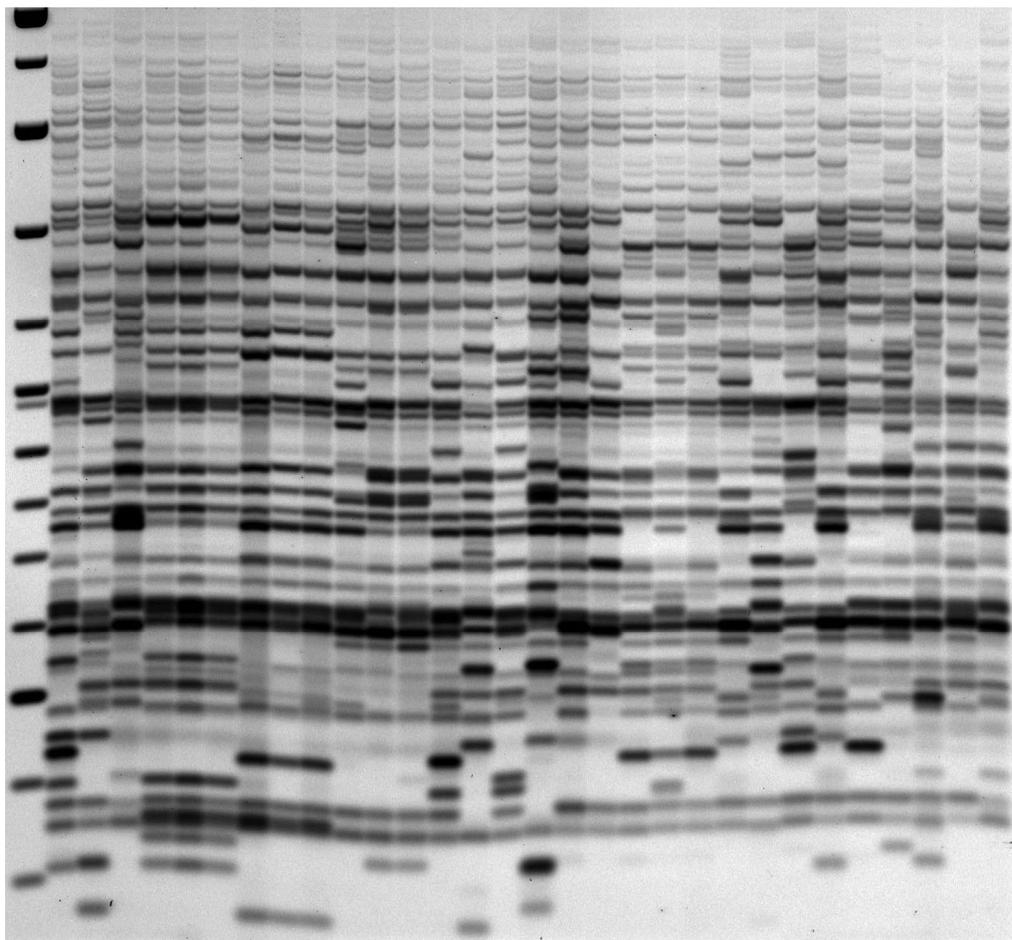


Fig. 3. Utility of IRAP for a diversity of plant species. The phenogram of 30 genotypes of populations of *H. spontaneum* based on IRAP analysis are shown as negative images of ethidium bromide — stained agarose gels following electrophoresis. Results for BARE-1 LTR primer 1369 (5'-TGCCTCTAGGGCATATTTCCAACAC) are shown. A 100 bp DNA ladder is present on the left

motif [52]. Founded throughout genomes, microsatellites appear to be associated with retrotransposons and have high mutation rates due to polymerase slippage. Therefore, they may show much variation at individual loci within a species. In REMAP, anchor nucleotides are used at the 3' end of the SSR primer to avoid slippage of the primer within the SSR. It also prevents the detection of variation in repeat numbers within the microsatellite.

IRAP and REMAP methods have been used in gene mapping in barley, wheat, oat, rice blast pathogen (*Magnaporthe grisea* SP.), in studies of genome evolution in the grasses, 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, in barley [54], banana [26], grapevine, *Pisum*, apple, in citrus, *Aegilops* and *Triticum* [55],

banana, rice, flax [56], sunflower [29] and medicinal plants — *Adonis vernalis*, *Paeonia anomala*, *Adenophora lilifolia*, *Digitalis grandiflora* [57]. REMAP has been used also as a sensitive method for detecting genomic copies of retrotransposons amidst retrotransposon cDNAs, to examine genome evolution in wild barley [58].

RBIP/ TAM. RBIP (Retrotransposons-based insertion polymorphism) was described as a simple PCR-based detection of retrotransposon insertions using PCR between primers flanking the insertion site and primers from the insertion itself. The basic RBIP method has been developed for high-throughput applications by replacing gel electrophoresis with hybridization to a filter. PCR reactions detecting the occupied sites and unoccupied sites are carried out together, the products spotted onto membranes, and probed with a locus-specific probe. By using sensitive, oligo-based hybridisation to spotted PCR products, TAM, has allowed the dot blot approach to be scaled down to microarrays with the attendant advantages in throughput, efficiency and data collection. Using three primers, RBIP can detect both the presence and absence of the TE insertion and generates single-locus codominant markers. In the case of a retrotransposon, a primer designed in the LTR is used together with a primer designed in the flanking region and allow the amplification of an insertion site, when primers specific of both 5' and 3' flanking regions are used to scored the corresponding empty site. TE insertions are usually more than thousands of bases long so the empty site primers do not generate an amplicon from the occupied site. Hence, RBIP detects both the presence and absence of the insertion but requires that the sequence of the 5' and 3' flanking sequences of the TE insertions are known. RBIP analysis was used to show evolutionary history in pea and rice.

TAM (Tagged Microarray Marker) is a microarray-based method developed from RBIP for scoring thousands of DNAs for a co-dominant molecular marker on a glass microarray slide. RBIP also works well with single nucleotide polymorphism (SNP) markers [59]. In this approach, biotin-terminated allele-specific PCR products are spotted unpurified onto streptavidin-coated glass slides and visualised by hybridisation of fluorescent detector oligonucleotides to tags attached to the allele-specific PCR primers. Two tagged primer oligonucleotides are used per locus and each tag is detected by hybridisation to form a concatameric DNA probe labelled with multiple copies of a fluorochrome.

Inter PBS amplification (iPBS), a universal method for isolating and displaying retrotransposon polymorphisms. A major disadvantage of all retrotransposon-based molecular markers techniques is the need for sequence information to design element-specific primers. Although rapid retrotransposon isolation methods based on PCR with conservative primer for TE have been designed, it maybe still necessary to clone and sequence hundreds of clones to obtain a few good primers sequences. The LTRs contain no conserved motifs, which would allow their direct amplification by PCR.

There are several restriction and adaptor-based methods for LTR cloning, based on conservancy of reverse transcriptase domain, especially for Ty1-copia. Major classes of retroelements include the Pseudoviridae (Ty1-copia), the Metaviridae (Ty3 -gypsy) and the Retroposineae *LINE* (non-LTR) groups. All reverse transcribing elements can be obtained by PCR with degenerate primers. For example, Ty1-copia two degenerate primers were designed for RT domain encoding TAF₁₀LHG and reverse site YVDDML also encoding QMDVKT and reverse YVDDML. For Ty3-gypsy element degenerate primers were designed for RT domain encoding RMCVDYR or LSGYHQI or YPLPRID and reverse encoding site YAKLSKC and LSGYHQI. The reverse transcriptase based isolation method is limited to the families of retrotransposons, which contain this sequence. Thus, for example TRIM or LARDs and unknown yet classes LTR-retrotransposons cannot be found using this approach [32].

LTR retrotransposons and all retroviruses contain tRNA conservative primer binding site for tRNA^{iMet}, tRNA^{Lys}, tRNA^{Pro}, tRNA^{Trp}, tRNA^{Asn}, tRNA^{Ser}, tRNA^{Arg}, tRNA^{Phe}, tRNA^{Leu} and tRNA^{Gln}. Elongation from the 3'-terminal nucleotides of the respective tRNA results in the conversion of the viral/retrotransposon RNA genome to double-stranded DNA prior to its integration into the host DNA. While the process of reverse transcription is conserved among all retroviruses, the specific tRNA capture varies for different retroviruses and retroelements. The primer binding sequences (PBS) is universally present in all LTR-retrotransposons sequences. Hence, an isolation method for retrotransposon LTRs, which is based on the PBS sequence, has potential for cloning all possible LTR-retrotransposons.

Kalendar et al. [60] describes the development of exceedingly universal and efficient method, which utilizes the conserved parts of PBS sequences, both for direct visualization of polymorphism between individuals, transcription profile polymorphism, as for fast cloning of LTR parts from genomic DNA, or in a form of database search. In this way, any eukaryotic organism possessing LTR type of retrotransposon can be investigated. Primers, which were designed to match the conserved regions of the primer binding sequences in LTR retrotransposons, proved to be very efficient in PCR amplification of eukaryotic genomic DNA. Single PBS primers can only amplify nested inverted retrotransposons or related elements' sequences dispersed through genome DNA. PCR amplification occurs between two nested PBS and contains two LTR sequences. The PBS sequences are nested near each other in all eukaryotes (Fig.4).

Most of retrotransposons are nested, mixed, inverted or truncated in chromosomal sequences, and can be easily amplified using conservative PBS primers in any plant species tested. Fragments of LTR with retrotransposons internal part are located near other retrotransposons. Therefore, PBS sequences are very often located near to each other. This situation allows the use PBS sequences for cloning LTR.

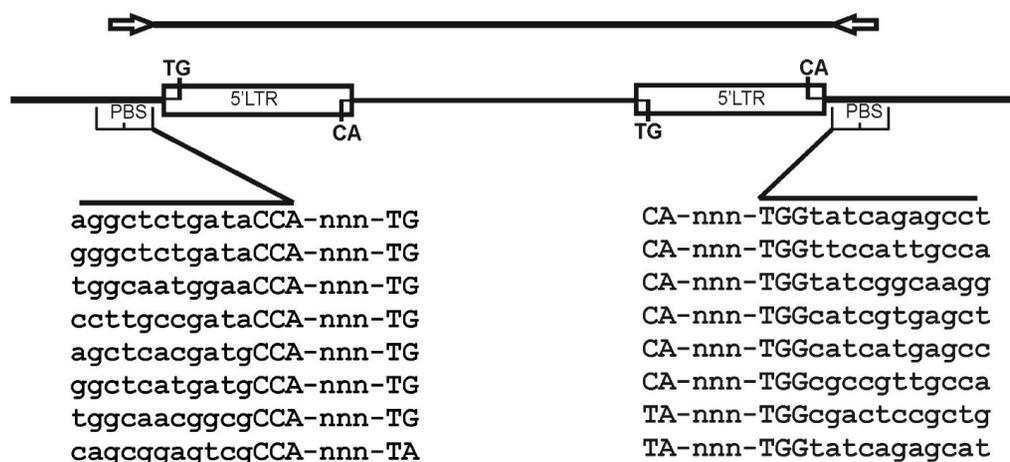


Fig. 4. The inter PBS amplification (iPBS) scheme and LTR retrotransposon structure. LTR and PBS sequence. Two nested LTR retrotransposons in inverted orientations amplified from single primer or two different primers from primer binding sites. PCR product contains both LTRs and PBS sequences as PCR primers in the termini. In figure schematically showing general structure for PBS and LTR sequences, between 5'LTR (5'-.CA) and PBS (5'-TGG..3') is spacer with several nucleotides (0–5 bases)

Where the retrotransposon density is high within genome, PBS sequences can be exploited for detection of their chance association with other retrotransposons. When retrotransposon activity or recombination has led to new genome integrations, this can be used to distinguish reproductively isolated plant line. In this case, amplified bands derived from newly inserted or recombination will be polymorphic, appearing only in plant lines in which the insertions or recombination have taken place.

After retrieving LTR sequences of a selected family of retrotransposon, and alignment is made of them to find out the most conserved region in them. The related plant species have conservative regions in LTR for identical retroelement, thus alignments of several LTR sequences from one species or mixing with sequences from the related species will identify conservative regions. Subsequently this conservative parts of LTR regions are used for inverted primers design for long distance PCR, for cloning of whole element and also for IRAP, REMAP or SSAP techniques.

The iPBS amplification technique shows about the same level of polymorphism in comparisons with IRAP and REMAP techniques and an efficient method for the detection of cDNA polymorphism and clonal differences resulting from retrotransposon activities or retrotransposon recombination after crossing-over [60].

Retrotransposon-based molecular markers to analyse genetics diversity. The analysis of genetic diversity and relatedness between or within different populations, species and individuals is a central task in genetics. The combination of different LTR primers or with combinations with microsatellite

primers (REMAP) allows for the generation of an almost unlimited number of unique markers.

Banding patterns were completely different if the same primers were used alone or in combinations, indicating that the majority 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. Usually, the REMAP pattern was considerably more variable than the corresponding ISSR pattern; and often (but not always, depending from LTR sequence) IRAP pattern with primer combinations shown more variability than a single priming PCR [25; 2].

Related species have phylogenetically related TE sequences (retroelements or transposons) meaning that PCR primers from one species can be used in another. In this case, primers designed to conservative TE sequences are advantageous. TEs are dispersed at whole chromosomes and most often mixed with other elements and repeats, that is the combinations of PCR primers from different repeats help to improving PCR fingerprint.

To study closely related varieties or breeding lines, one should develop a native retrotransposon system. This requires the cloning and sequencing of elements from the new species by using iPBS amplification method or technique based on conservancy of reverse transcriptase domain. This process begins with amplification and cloning of segments between retrotransposon domains that highly or universally conserved, development of new primers specific for the retrotransposon families found, and testing these for their efficacy as markers.

The genome size of studied organism is positively correlated with the efficiency of repeat-based amplification techniques; the larger genome the most easy developing good primers for revealing multiple bands for polymorphism detection (barley, wheat); small genome organism like *Brachypodium distachyon* or *Vitis vinifera* is most difficult to PCR marker development.

S-SAP is generally carried out on sequencing gels due to the large number of products generated, whereas IRAP and REMAP are used on agarose systems. However, IRAP and REMAP can be adapted to sequencers. These methods generate tens to hundreds of products in each amplification reaction, depending on the prevalence of the retrotransposon family, the selection of the second primer — the restriction site and number of selective bases in S-SAP -, and the organization of the genome of the plant.

A marker from any of the multilocus, anonymous systems (S-SAP, IRAP, and REMAP) can be converted into a corresponding RBIP marker and vice versa. Markers from the former methods are very easy to harvest and they can be quickly examined for their informativeness before taking on the investment of developing a corresponding RBIP marker. Electrophoretically resolved bands from S-SAP, IRAP, and REMAP are derived from one side of a retrotransposon insertion. Sequencing of the isolated, informative bands will enable the design of a PCR primer corresponding to the flanking genomic DNA at one side of the insertion, assuming that the sequence is not repetitive

and therefore unusable. However, the genomic sequence flanking the other side of the element needs to be found in order to score the empty site. This can be obtained by screening germplasm accessions that are polymorphic for the original band, then carrying out S-SAP reaction on these, where the LTR primer is replaced with a primer designed to the known flank that is facing toward the insertion site.

Conclusions. Many features of retrotransposons make them appealing as the basis of molecular marker systems. They are ubiquitous, abundant, dispersed components of eukaryotic genomes. Their activity simultaneously leads to genome diversification and provides a means of its detection. 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 is not linked to removal 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. This is very helpful in pedigree and phylogenetic analyses. Later recombination events at a full site are highly unlikely to regenerate the original empty site. In contrast, microsatellites, SNPs, and methods relying on gain or loss of restriction sites (in essence SNPs), suffer from a lack of temporal directionality in the changes they detect, resulting in the problem of homoplasy. For example, SINE elements have served to trace human roots to Africa, to determine the relationship of whales to even-toed ungulates. 1997), and to clarify the relationships between wild rice species.

Markers based on LTR retrotransposons, in one or other of the manifestations described above, generically referred to as «transposon display. The applications range from investigations of retrotransposon activation and mobility to studies of biodiversity, genome evolution and chromatin modification and epigenetic reprogramming to the mapping of genes and the estimation of genetic distance, to assessment of essential derivation of varieties, detection of somaclonal variation and cDNA fingerprinting. The retrotransposon insertions that provide useful polymorphisms are, of course, only those that are passed into the egg cells and pollen. One can thus think of the retrotransposons as sexually transmitted diseases, albeit ones that moves by a cellular, rather than extracellular, pathway into the new host.

Because LTR retrotransposons are ubiquitous, these methods are generic. Furthermore, similar approaches have been applied to the non-LTR retrotransposons in the plants, in particular to the SINE elements. The insertion pattern of the human Alu, a SINE and the most prevalent transposable element in the human genome, has not only served as a tool in many studies of human population structure, but also been linked to various heritable diseases. In principle, retrotransposon- or endogenous retrovirus-based molecular markers could prove highly useful in animals, including mammals and birds.

Commercial platforms for SNP detection (e.g., Illumina) have been developed and garnered much popularity for major crops, domestic animals, and humans. Development of SNPs depends on having abundant sequence data. The costs of acquiring this data, as well as of applying commercial assays, represent a barrier for research on underfunded tropical crops and wild species. Furthermore, evolutionary studies with SNPs are affected by the problems of homoplasy in SNP state, the lack of neutrality of genic markers, and the uneven chromosomal distribution of the highly expressed genes that are used to generate SNPs. While genetic analysis by shotgun sequencing remains a tantalizing possibility, the cost is still prohibitive. For these reasons, cheap, generic, easily applied retrotransposon marker systems will remain a viable choice for genetic markers for the foreseeable future.

REFERENCES

1. Lewontin R. C. 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* / R. C. Lewontin, J. L. Hubby // *Genetics*. — 1966. — Vol. 54, № 2. — P. 595–609.
2. Kan Y.W. Polymorphism of DNA sequence adjacent to human beta-globin structural gene: relationship to sickle mutation / Y.W. Kan, A. M. Dozy // *Proc. Natl. Acad. Sci. U. S. A.* — 1978. — Vol. 75, № 11. — P. 5631–5635.
3. Lee D. A copia-like element in *Pisum* demonstrates the uses of dispersed repeated sequences in genetic analysis / D. Lee, T. Ellis, L. Turner [et al.] // *Plant Mol. Biol.* — 1990. — Vol. 15. — P. 707–722.
4. Williams J. DNA polymorphisms amplified by arbitrary primers are useful as genetic-markers / J. Williams, A. Kubelik, K. Livak [et al.] // *Nucleic Acids Res.* — 1990. — Vol. 18, № 22. — P. 6531–6535.
5. Welsh J. Fingerprinting genomes using PCR with arbitrary primers / J. Welsh, M. McClelland // *Nucleic Acids Res.* — 1990. — Vol. 18, № 24. — P. 7213–7218.
6. Caetano-Anolles G. DNA amplification fingerprinting using very short arbitrary oligonucleotide primers / G. Caetano-Anolles, B. Bassam, P. Gresshoff // *Bio/Technology*. — 1991. — Vol. 9. — P. 553–557.
7. Meyer W. Hybridization probes for conventional DNA-fingerprinting used as single primers in the polymerase chain-reaction to distinguish strains of *Cryptococcus neoformans* / W. Meyer, T. Mitchell, E. Freedman [et al.] // *J. Clin. Microbiol.* — 1993. — Vol. 31, № 9. — P. 2274–2280.
8. Sivolap Y. The genetic polymorphism of cereals demonstrated by PCR with random primers / Y. Sivolap, R. Kalendar, S. Chebotar // *Cytology and Genetics*. — 1994. — Vol. 28, № 6. — P. 54–61.
9. Salimath S. S. Assessment of genome origins and genetic diversity in the genus *Eleusine* with DNA markers / S. Salimath, A. de Oliveira, I. Godwin [et al.] // *Genome*. — 1995. — Vol. 38, № 4. — P. 757–763.
10. Charlieu J. P. 3' Alu PCR: a simple and rapid method to isolate human polymorphic markers / J. Charlieu, A. Laurent, D. Carter [et al.] // *Nucleic Acids Res.* — 1992. — Vol. 20, № 6. — P. 1333–1337.

11. Tautz D. Hypervariability of simple sequences as a general source for polymorphic DNA markers / D. Tautz // *Nucleic Acids Res.* — 1989. — Vol. 17, № 16. — P. 6463–6471.
12. Zietkiewicz E. Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification / E. Zietkiewicz, A. Rafalski, D. Labuda // *Genomics.* — 1994. — Vol. 20, № 2. — P. 176–183.
13. Finnegan D. Eukaryotic transposable elements and genome evolution / D. Finnegan // *Trends in Genetics.* — 1989. — Vol. 5. — P. 103–107.
14. Flavell A. Ty1-copia group retrotransposons are ubiquitous and heterogeneous in higher plants / A. Flavell, E. Dunbar, R. Anderson [et al.] // *Nucleic Acids Res.* — 1992. — Vol. 20, № 14. — P. 3639–3644.
15. Voytas D. Copia-like retrotransposons are ubiquitous among plants / D. Voytas, M. Cummings, A. Konieczny [et al.] // *Proc. Natl. Acad. Sci. U. S. A.* — 1992. — Vol. 89. — P. 7124–7128.
16. Kalendar R. Large retrotransposon derivatives: abundant, conserved but nonautonomous retroelements of barley and related genomes / R. Kalendar, C. Vicent, O. Peleg et al. // *Genetics.* — 2004. — Vol. 166, № 3. — P. 1437–1450.
17. Frankel A. HIV-1: Fifteen proteins and an RNA / A. Frankel, J. Young // *Annu. Rev. Biochem.* — 1998. — Vol. 67. — P. 1–25.
18. Wicker T. A unified classification system for eukaryotic transposable elements / T. Wicker, F. Sabot, A. Hua-Van [et al.] // *Nature Rev. Genet.* — 2007. — Vol. 8, № 12. — P. 973–982.
19. Paterson A. The *Sorghum bicolor* genome and the diversification of grasses / A. Paterson, J. Bowers, R. Bruggmann et al. // *Nature.* — 2009. — Vol. 457, № 7229. — P. 551–556.
20. Consortium I. A physical, genetic and functional sequence assembly of the barley genome / I. Consortium, K. Mayer, R. Waugh [et al.] // *Nature.* — 2012. — Vol. 491, № 7426. — P. 711–716.
21. International Wheat Genome Sequencing C. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome // *Science.* — 2014. — Vol. 345, № 6194. — P. 1251788.
22. Grandbastien M. *Tnt1*, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics / M. Grandbastien, A. Spielmann, M. Caboche // *Nature.* — 1989. — Vol. 337. — P. 376–380.
23. Nelson D. Alu polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources / D. Nelson, S. Ledbetter, L. Corbo [et al.] // *Proc. Natl. Acad. Sci. U. S. A.* — 1989. — Vol. 86, № 17. — P. 6686–6690.
24. Sinnott D. Alu polymorphisms—human DNA polymorphisms detected by polymerase chain reaction using Alu-specific primers / D. Sinnott, M. Deragon, L. Simard [et al.] // *Genomics.* — 1990. — Vol. 7, № 3. — P. 331–334.
25. Leigh F. Comparison of the utility of barley retrotransposon families for genetic analysis by molecular marker techniques / F. Leigh, R. Kalendar, V. Lea [et al.] // *Mol. Genet. Genomics.* — 2003. — Vol. 269, № 4. — P. 464–474.
26. Teo C. Genome constitution and classification using retrotransposon-based markers in the orphan crop banana / C. Teo, S. Tan, C. Ho [et al.] // *J. Plant Biol.* — 2005. — Vol. 48, № 1. — P. 96–105.
27. Antonius-Klemola K. TRIM retrotransposons occur in apple and are polymorphic between varieties but not sports / K. Antonius-Klemola, R. Kalendar

- dar, A. Schulman // *Theor. Appl. Genet.* — 2006. — Vol. 112, № 6. — P. 999–1008.
28. Schulman A. A movable feast: diverse retrotransposons and their contribution to barley genome dynamics / A. Schulman, R. Kalendar // *Cytogen. Genome Res.* — 2005. — Vol. 110, № 1–4. — P. 598–605.
 29. Vukich M. Genetic variability in sunflower (*Helianthus annuus* L.) and in the *Helianthus* genus as assessed by retrotransposon-based molecular markers / M. Vukich, A. Schulman, T. Giordani [et al.] // *Theor. Appl. Genet.* — 2009. — Vol. 119, № 6. — P. 1027–1038.
 30. Kalendar R. The use of retrotransposon-based molecular markers to analyze genetic diversity / R. Kalendar // *Ratarstvo i povrtarstvo.* — 2011. — Vol. 48, № 2. — P. 261–274.
 31. Kalendar R. Analysis of plant diversity with retrotransposon-based molecular markers / R. Kalendar, A. Flavell, T. Ellis et al. // *Heredity.* — 2011. — Vol. 106, № 4. — P. 520–530.
 32. Kalendar R. Transposon-based tagging: IRAP, REMAP, and iPBS / R. Kalendar, A. Schulman // *Methods Mol Biol.* — 2014. — Vol. 1115. — P. 233–255.
 33. Manninen O. Application of BARE-1 retrotransposon markers to the mapping of a major resistance gene for net blotch in barley / O. Manninen, R. Kalendar, J. Robinson [et al.] // *Mol. General Genet.* — 2000. — Vol. 264, № 3. — P. 325–334.
 34. Waugh R. Genetic distribution of BARE-1-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). / R. Waugh, K. McLean, A. Flavell [et al.] // *Mol. General Genet.* — 1997. — T. 253. — P. 687–694.
 35. Lou Q. Ty1-copia retrotransposon-based SSAP marker development and its potential in the genetic study of cucurbits / Q. Lou, J. Chen // *Genome.* — 2007. — Vol. 50, № 9. — P. 802–810.
 36. Moisy C. The *Tw1* retrotransposon family is conserved between plant genomes separated by over 100 million years / C. Moisy, A. Schulman, R. Kalendar [et al.] // *Theor. Appl. Genet.* — 2014. — Vol. 127, № 5. — P. 1223–1235.
 37. Yu G. An anchored AFLP- and retrotransposon-based map of diploid *Avena* / G. Yu. R. Wise // *Genome.* — 2000. — Vol. 43. — P. 736–749.
 38. Queen R. Retrotransposon-based molecular markers for linkage and genetic diversity analysis in wheat / R. Queen, B. Gribbon, C. James [et al.] // *Mol. General Genomics.* — 2004. — Vol. 271. — P. 91–97.
 39. Porceddu A. Development of S-SAP markers based on an LTR-like sequence from *Medicago sativa* L. / A. Porceddu, E. Albertini, G. Baracaccia [et al.] // *Mol. Genet. Genomics.* — 2002. — Vol. 267. — P. 107–114.
 40. Gribbon B. Phylogeny and transpositional activity of *Ty1*-copia group retrotransposons in cereal genomes / B. Gribbon, S. Pearce, R. Kalendar [et al.] // *Mol. Gen. Genet.* — 1999. — Vol. 261, № 6. — P. 883–891.
 41. Tahara M. Isolation of an active element from a high-copy-number family of retrotransposons in the sweetpotato genome / M. Tahara, T. Aoki, S. Suzuka et al. // *Mol. General Genomics.* — 2004. — Vol. 272. — P. 116–127.
 42. Ellis T. Polymorphism of insertion sites of *Ty1*-copia class retrotransposons and its use for linkage and diversity analysis in pea / T. Ellis, S. Poyser, M. Knox et al. // *Mol. General Genet.* — 1998. — Vol. 260. — P. 9–19.

43. Tam S. Comparative analyses of genetic diversities within tomato and pepper collections detected by retrotransposon-based SSAP, AFLP and SSR / S. Tam, C. Mhiri, A. Vogelaar et al. // *Theor. Appl. Genet.* — 2005. — Vol. 110. — P. 819–831.
44. Nagy E. Characterization of chromosome-specific S-SAP markers and their use in studying genetic diversity in *Aegilops* species / E. Nagy, I. Molnar, A. Schneider et al. // *Genome.* — 2006. — Vol. 49. — P. 289–296.
45. Venturi S. Retrotransposon characterisation and fingerprinting of apple clones by S-SAP markers / S. Venturi, L. Dondini, P. Donini [et al.] // *Theor. Appl. Genet.* — 2006. — Vol. 112. — P. 440–444.
46. Syed N. Sequence-specific amplification polymorphisms (SSAPs): a multi-locus approach for analyzing transposon insertions / N. Syed, A. Flavell // *Nature Protocols.* — 2006. — Vol. 1, № 6. — P. 2746–2752.
47. Syed N. A detailed linkage map of lettuce based on SSAP, AFLP and NBS markers / N. Syed, A. Sørensen, R. Antonise et al. // *Theor. Appl. Genet.* — 2006. — Vol. 112, № 3. — P. 517–527.
48. García-Martínez J. Study on the evolution of the grande retrotransposon in the *Zea* genus / J. García-Martínez, J. Martínez-Izquierdo // *Mol. Biol. Evol.* — 2003. — Vol. 20, № 5. — P. 831–841.
49. Petit M. Differential impact of retrotransposon populations on the genome of allotetraploid tobacco (*Nicotiana tabacum*) / M. Petit, K. Lim, E. Julio [et al.] // *Mol. Genet. Genomics.* — 2007. — Vol. 278, № 1. — P. 1–15.
50. Van den Broeck D. Transposon Display identifies individual transposable elements in high copy number lines / D. van den Broeck, T. Maes, M. Sauer [et al.] // *Plant J.* — 1998. — Vol. 13. — P. 121–129.
51. Chang R. Inter-MITE polymorphisms (IMP): a high throughput transposon-based genome mapping and fingerprinting approach / R. Chang, L. O'Donoghue, T. Bureau // *Theor. Appl. Genet.* — 2001. — Vol. 102, № 5. — P. 773–781.
52. Kalendar R. IRAP and REMAP: Two new retrotransposon-based DNA fingerprinting techniques / R. Kalendar, T. Grob, M. Regina et al. // *Theor. Appl. Genet.* — 1999. — Vol. 98, № 5. — P. 704–711.
53. Kalendar R. IRAP and REMAP for retrotransposon-based genotyping and fingerprinting / R. Kalendar, A. Schulman // *Nature Protocols.* — 2006. — Vol. 1, № 5. — P. 2478–2484.
54. Brik A. F. IRAP-and REMAP-analyses of barley varieties of Odessa breeding / A. Brik, R. Kalendar, O. Stratula et al. // *Cytology and Genetics.* — 2006. — Vol. 40, № 3. — P. 24–33.
55. Boyko E. A high-density cytogenetic map of the *Aegilops tauschii* genome incorporating retrotransposons and defense-related genes: insights into cereal chromosome structure and function / E. Boyko, R. Kalendar, V. Korzun [et al.] // *Plant Mol. Biol.* — 2002. — Vol. 48, № 5. — P. 767–790.
56. Smykal P. Genetic diversity of cultivated flax (*Linum usitatissimum* L.) germplasm assessed by retrotransposon-based markers / P. Smykal, N. Bacova-Kerteszo-va, R. Kalendar et al. // *Theor. Appl. Genet.* — 2011. — Vol. 122, № 7. — P. 1385–1397.
57. Boronnikova S. Using IRAP markers for analysis of genetic variability in populations of resource and rare species of plants / S. Boronnikova, R. Kalendar // *Russian J. Genet.* — 2010. — Vol. 46, № 1. — P. 36–42.

58. Kalendar R. Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence / R. Kalendar, J. Tanskanen, S. Immonen [et al.] // Proc. Natl. Acad. Sci. U. S. A. — 2000. — Vol. 97, № 12. — P. 6603–6607.
59. Flavell A. A microarray-based high throughput molecular marker genotyping method: the tagged microarray marker (TAM) approach / A. Flavell, V. Bolshakov, A. Booth [et al.] // Nucleic Acids Res. — 2003. — Vol. 31. — P. e115.
60. Kalendar R. iPBS: a universal method for DNA fingerprinting and retrotransposon isolation / R. Kalendar, K. Antonius, P. Smýkal [et al.] // Theor. Appl. Genet. — 2010. — Vol. 121, № 8. — P. 1419–1430.

Receive 10.07.2015.

УДК 577.2.08

Календар Р. М.

МОЛЕКУЛЯРНІ МАРКЕРИ НА ОСНОВІ РЕТРОТРАНСПОЗОНІВ

Молекулярні маркери набули широкого застосування в сучасній генетиці, селекції рослин, генетичному картуванні та ідентифікації генів, у судово-медичній експертизі людини та ветеринарії. Ретротранспозони є ідеальними маркерами, адже вони локалізовані у всіх хромосомах еукаріот і мають потенційну можливість копіювати себе способом реплікативної транспозиції. Оскільки ретротранспозони складають основну частину геному еукаріот і потенційно активні, на їхній основі були розроблені різні системи ДНК-маркерів для виявлення генетичного поліморфізму. Огляд дає уявлення про основні напрями застосування послідовностей ретротранспозонів як молекулярних маркерів, розроблених для рослин, оцінює можливості їхнього застосування при аналізі генетичного різноманіття досліджуваних культур. Також розглянуто способи швидкого виділення ретротранспозонів для нових видів.

УДК 577.2.08

Календарь Р. Н.**МОЛЕКУЛЯРНЫЕ МАРКЕРЫ НА ОСНОВЕ РЕТРОТРАНСПОЗОНОВ**

Молекулярные маркеры широко используются в современной генетике, селекции растений, при генетическом картировании и идентификации генов, в судебно-медицинской экспертизе человека, ветеринарии. Ретротранспозоны являются идеальными молекулярными маркерами, так как локализованы во всех хромосомах эукариот и имеют потенциальную возможность копировать себя способом репликативной транспозиции. Поскольку ретротранспозоны составляют основную часть генома эукариот и потенциально активны, на их основе были разработаны различные системы ДНК-маркеров для выявления генетического полиморфизма. Обзор дает представление об основных подходах применения последовательностей ретротранспозонов в качестве молекулярных маркеров, разработанных для растений, оценивает возможности их применения при анализе генетического разнообразия исследуемых культур. Также рассмотрены способы быстрого выделения ретротранспозонов для новых видов.