CSIRO PUBLISHING

Functional Plant Biology, 2019, **46**, 15–29
https://doi.org/10.1071/FP18098

Review

Use of retrotransposon-derived genetic markers to analyse genomic variability in plants

Ruslan Kalendar[A,C], *Asset Amenov*[B] *and Asset Daniyarov*[B]

[A]Department of Agricultural Sciences, PO Box 27 (Latokartanonkaari 5), FI-00014 University of Helsinki, Helsinki, Finland.
[B]RSE 'National Center for Biotechnology', 13/5 Kurgalzhynskoye Road, Astana, 010000, Kazakhstan.
[C]Corresponding author. Email: ruslan.kalendar@helsinki.fi

Abstract. Transposable elements (TEs) are common mobile genetic elements comprising several classes and making up the majority of eukaryotic genomes. The movement and accumulation of TEs has been a major force shaping the genes and genomes of most organisms. Most eukaryotic genomes are dominated by retrotransposons and minimal DNA transposon accumulation. The 'copy and paste' lifecycle of replicative transposition produces new genome insertions without excising the original element. Horizontal TE transfer among lineages is rare. TEs represent a reservoir of potential genomic instability and RNA-level toxicity. Many TEs appear static and nonfunctional, but some are capable of replicating and mobilising to new positions, and somatic transposition events have been observed. The overall structure of retrotransposons and the domains responsible for the phases of their replication are highly conserved in all eukaryotes. TEs are important drivers of species diversity and exhibit great variety in their structure, size and transposition mechanisms, making them important putative actors in evolution. Because TEs are abundant in plant genomes, various applications have been developed to exploit polymorphisms in TE insertion patterns, including conventional or anchored PCR, and quantitative or digital PCR with primers for the 5′ or 3′ junction. Alternatively, the retrotransposon junction can be mapped using high-throughput next-generation sequencing and bioinformatics. With these applications, TE insertions can be rapidly, easily and accurately identified, or new TE insertions can be found. This review provides an overview of the TE-based applications developed for plant species and assesses the contributions of TEs to the analysis of plants' genetic diversity.

Additional keywords: genetic diversity, molecular marker, transposable element.

Received 17 April 2018, accepted 23 August 2018, published online 4 October 2018

Introduction

All eukaryotic genomes contain DNA sequences named "repetitive elements" that are present in multiple copies throughout the genome. These repetitive sequences can be arrayed in tandem (e.g. in telomeric DNA). Alternatively, repetitive elements, such as mobile elements and processed pseudogenes, can be interspersed throughout the genome.

Transposable elements (TEs) are highly abundant mobile genetic elements that have multiple classes and constitute a large fraction of most eukaryotic genomes. TEs can be subdivided on the basis of their size, with short interspersed elements being less than 1000 bp long and the rest considered to be long interspersed elements.

The class known as retrotransposons, for example, comprises ~10–90% of eukaryote genomes. Retrotransposons and related elements are highly abundant in eukaryotic genomes, where, for example, the copy number of a single short interspersed nuclear element (SINE) may exceed 10^6. TEs, particularly long-terminal repeat (LTR) retrotransposons, are also predominantly located in heterochromatic regions of the genome. In plants, LTR retrotransposons tend to be more abundant than non-LTR retrotransposons (Macas *et al*. 2011). In many crop plants, between 40% and 70% of the total DNA comprises LTR retrotransposons (Pearce *et al*. 1996; Goke and Ng 2016). Cereals and citrus fruits often have retrotransposons locally nested in one another and in extensive domains, referred to as "retrotransposon seas", that surround gene islands, despite the most prevalent retrotransposons being dispersed throughout the genome (Neumann *et al*. 2011). Their qualities, such as abundance, general dispersion and activity, provide perfect conditions for developing molecular markers (Kalendar 2011; Kalendar and Schulman 2014).

These elements use extensive cellular resources in their replication, expression and amplification, and, as a result of the negative effects of their transposition, contribute to genetic variation. Thus mobile elements are potentially intracellular agents that attack the host genome and exploit cellular resources, and also occasionally have a positive influence on genome evolution.

TEs are among the most fluid genomic components, fluctuating immensely in copy number over a relatively short evolutionary timescale, and represent a major component of the

structural evolution of plant genomes (Flavell *et al*. 1992; Voytas *et al*. 1992; Mascagni *et al*. 2017).

The movement and accumulation of TEs has been a major force in shaping the genes and genomes of almost all organisms (Kim *et al*. 2017). Retrotransposable elements (RTEs) and other TEs represent a massive reservoir of potential genomic instability and RNA-level toxicity. TEs mostly appear static and nonfunctional. However, some TEs are capable of replicating and mobilising to new positions in the genome, and even immobile TE copies can be expressed (Levin 1995). Moreover, endogenous transposition itself has been detected in the germline in which TEs have been most extensively investigated (Van Sluys *et al*. 1987; Schulman 2013). In addition, somatic transposition events have been observed in early embryonic development.

These observations do not, however, address the extent to which TEs are expressed or mobilised in plant development during germination, much less the possible functional consequences of such activation. Clarifying these issues may afford a new mechanistic view of plant adaptation and evolution (Mascagni *et al*. 2017).

Classification of TEs

TEs are classified into two main groups in eukaryotic genomes and defined according to their mechanism of transposition (Piegu *et al*. 2015). They are classified as Class I TEs transposing through an RNA intermediary, and other transposons (Class II), which do not have an RNA intermediary (Finnegan 1990) (Fig. 1). Two main subclasses of retrotransposons can be classified according to their structure and transposition cycle: LTR retrotransposons and non-LTR retrotransposons (long interspersed repetitive elements and short interspersed nuclear elements (SINE)), determined by the presence or absence of LTRs at their ends. All groups are complemented by degraded members of their nonautonomous forms, which lack

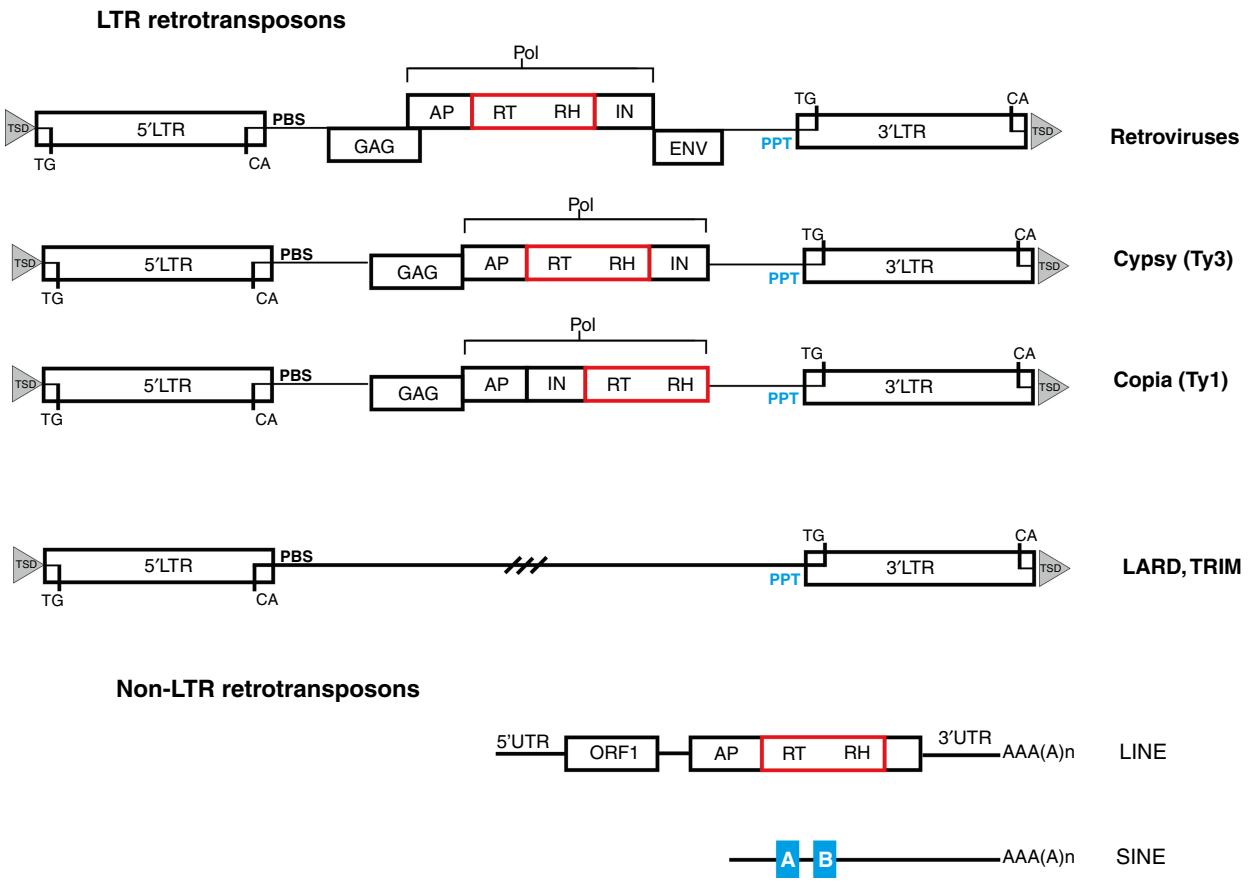

Fig. 1. Retrotransposon architecture. The main groups of autonomous and nonautonomous retrotransposons. (*a*) Retroviruses and autonomous long-terminal repeat (LTR) retrotransposons. Above: the basic structure of an LTR retrotransposon, comprising: target site duplication; LTRs; the primer-binding site (PBS), which is the (–)-strand priming site for reverse transcription; the polypurine tract (PPT), which is the (+)-strand priming site for reverse transcription. The PBS and PPT are part of the internal domain, which, in autonomous elements, includes the protein-coding open reading frame(s). The open reading frame(s) of the internal domain are: GAG, encoding the capsid protein Gag; AP, aspartic proteinase; RT-RH, reverse transcriptase – RNase H; INT, integrase; ENV, envelope protein. (*b*) Nonautonomous retrotransposons. Large retrotransposon derivative (LARD) elements have a long internal domain with a conserved structure but lack a coding capacity. Terminal-repeat retrotransposons in miniature (TRIM) elements have virtually no internal domain except for the PBS and PPT signals. (*c*) Autonomous and nonautonomous non-LTR retrotransposons. The autonomous order long interspersed repetitive elements (LINE) of the L1 superfamily and the nonautonomous order short interspersed nuclear elements (SINE) are shown.

the genes that are essential for transposition: miniature inverted-repeat tandem elements for Class II, SINEs for non-LTR retrotransposons, and terminal-repeat retrotransposons in miniature and large retrotransposon derivatives for LTR retrotransposons (Kalendar *et al*. 2004; Piegu *et al*. 2015). Given the complexity and diversity of TE origins, a universal TE classification could be composed of eight classes (Piegu *et al*. 2015). The 'class' level would be similar to virus classification, namely a grouping of entities with common biological characteristics but not necessarily requiring that they have to share a common origin. New independent classes correspond to transposons for the retroposon class (including various non-LTR retrotransposons: long interspersed repetitive elements, Penelope-like elements and Group II introns).

The LTR of an integrated element serves as a template for the transcription of LTR retrotransposons. In this process, a full-length RNA copy is produced that contains a single copy of the LTR divided between its two ends (the LTR provides both the start site and the polyadenylation signal for the element). With the reverse transcription of this RNA into extrachromosomal cDNA, a full-length element is eventually integrated back into the genome. The target sites for reverse transcription are located immediately internal to the LTRs. Reverse transcriptase and integrase enzymes together with RNA are integrated into the structural component of a virus-like particle, which is encoded from the large central part of the retrotransposon.

Three basic types of LTR retrotransposon structures are illustrated in Fig. 1, showing two LTRs. An LTR varying from a single length of a few kb to 100 bp generally starts, and its inverted repeat sequence 5′ to 5′-TG–CA-3′ is the end. They tend to form direct repeats of 4–6 bp (target site duplications) at both ends of the transposon upon insertion into the host genome. An LTR retrotransposon comprises a gene encoding a variety of proteins, including the *GAG* (encoding structural proteins forming the shell, the synthesis of reverse transcription) and the poly *POL* gene (encoding a series of reverse transcription enzymes). LTR retrotransposons further comprise transcription initiation and termination related to a tRNA binding site (a primer-binding site (PBS)) and a polypurine sequence (the polypurine tract). Based on the similarity of the order and sequence of the enzyme transposase genes, LTR retrotransposons can be subdivided into the *Tyl-copia* type and the *Ty3-gypsy* type.

LTR retrotransposons are autonomous elements in that although they are dependent on many cellular proteins for their amplification cycle, they do encode all the necessary proteins within the element (Frankel and Young 1998). LTR retrotransposons are similar in structure to retroviruses, with transcriptional regulatory sequences located in the flanking LTRs, an initiation site to allow priming of the reverse transcription located downstream of the first LTR, and several open reading frames encoding the proteins required for retrotranspositions. These proteins include domains for an endonuclease to cleave the genomic integration site and reverse transcriptase to copy the RNA to DNA. Unlike retroviruses, however, LTR retrotransposons lack envelope genes and genomic components required for creating a functional viral capsule. Nonautonomous, degenerated versions of LTR retrotransposons also exist, in which the LTR structure and PBS are maintained but the coding capacity is removed.

However, unlike retroviruses, instead of leaving the cell to infect new cells, retrotransposons have a shorter lifecycle, and they only move in the nucleus and insert the new copies into their host genomes. New polymorphisms are developed in the gene pool if integration occurs within a cell lineage from which pollen or egg cells are ultimately derived or in the somatic cells of a clonally propagated plant. These newly integrated copies are applicable for the genetic identification of lines, varieties or populations of plants.

Horizontal transfers and TE diversity

Horizontal transfers of TEs are very marginal and extraordinary events that can drive TE diversity between lineages. In recent years, several studies have reported cases of TE transfer (Gao *et al*. 2018), as shown by the establishment of the horizontally transferred TE database (http://lpa.saogabriel.unipampa.edu.br:8080/httdatabase/, accessed 6 September 2018) (Dotto *et al*. 2015).

The invasion of a TE from an unrelated species by bypassing species barriers and entering into a new genome is an extremely rare and special event. Special conditions are required to determine which horizontal transfers will take place. For example, for symbiotic species, the probability of horizontal transfers of TEs is much higher. We have shown that particular TEs are universally distributed among closely and distantly related species. There is no unique set of TEs for a particular species (Antonius-Klemola *et al*. 2006; Kalendar *et al*. 2008; Smykal *et al*. 2009; Hosid *et al*. 2012; Moisy *et al*. 2014; Masuta *et al*. 2018). Related species have phylogenetically related TE sequences (retroelements or transposons). Phylogenetic analysis of TEs has demonstrated that their patterns of conservation match the plant family from which the retrotransposon was isolated. Both the LTRs and the central part show conservation that is consonant with their parent plant families. Generally, retrotransposons have not been extensively explored as phylogenetic markers, except in a few papers that have discussed the phylogenetic relationships among concrete retrotransposon sequences (Kalendar *et al*. 2008; Moisy *et al*. 2014; Ivancevic *et al*. 2016). Many advantages can be gained by using high-copy retrotransposons for eukaryotic phylogenetic studies, because RTEs are widely distributed and diverse in eukaryotes. The main reasons for identifying false horizontal transfers of TEs are associated with the imperfect classification of TEs. For example, a particular TE has different names in separate species.

Retrotransposable element-based genetic marker applications

Retrotransposable elements, which are among the TEs that are abundantly present in the genome of plants, are also known to be excellent DNA markers. Retrotransposons are mobile elements that insert themselves into new genomic locations via a mechanism that involves the reverse transcription of an RNA intermediary. Retrotransposons may be grouped into at least three classes that are structurally distinct and retrotranspose

using radically different mechanisms. These three families are exogenous retroviruses, retrovirus-like LTR retrotransposons and non-LTR elements such as human Long interspersed nuclear elements-1 (LI) and SINE elements (Alu family).

In most studied species, interspersed repeats are not evenly (but rather unevenly) distributed around the nuclear genome and some tend to cluster around the centromeres or telomeres, and are often integrated into introns and promoter sites. Changes in the copy number of repeat elements and internal rearrangements on both homologous chromosomes occur after the induction of recombinational processes during the meiotic prophase (Sanchez *et al*. 2017; Klein and O'Neill 2018). The recombination takes place together with the formation of recombinant TEs. Each transposition burst generates a new progeny population of chromosomally integrated LTR retrotransposons consisting of pairwise recombination products produced in a process. This explains the high rates of sequence diversification in retrotransposons (Sanchez *et al*. 2017). The resulting heterogeneity in the arrangement of discernible repeats is utilised in certain molecular marker techniques targeting the mentioned repeat elements.

The insertion of LTR retrotransposons is random and it occurs during the transposition process in the continuous evolution of species. This can provide a wealth of information for the study of evolution and species, and differentiation of the genome. The transposition mechanism for the LTR–LTR retrotransposon sequence determines the ends after transposition and is completely consistent. Therefore, by comparing the sequence LTR ends of the complete transposon, the insertion time can be calculated from their mutation rates.

In plants, it has been demonstrated that the mobility of TEs is limited by DNA methylation and certain histone marks (Martinez and Slotkin 2012). The suppression of DNA methylation in genetic mutants can therefore result in the mobilisation of TEs. It has also been shown that abiotic stress, which reduces DNA methylation, can mobilise certain DNA TEs (Masuta *et al*. 2018). Furthermore, it has been reported that stresses imposed on plants that are defective in RNA-directed DNA methylation can activate TEs.

TEs are very rarely activated under normal growth conditions and few active TEs are currently known (Martinez and Slotkin 2012). However, the requirement for genetic mutants in components involved in the defence against TEs limits the possibility to activate TEs in nonmodel organisms or organisms that are difficult to transform. Therefore, the exploitation of endogenous TEs to obtain genetic and epigenetic diversity is currently very limited.

TEs have proven to be very useful genetic tools and have been broadly exploited for gene disruption and transgenesis in a wide variety of organisms. The emergence of retrotransposon-related applications has followed basic research demonstrating their ubiquity and activity in plants (Debladis *et al*. 2017). Most marker methods based on retrotransposons rely on DNA amplification and next-generation sequencing (NGS). Different ways of using TEs as molecular markers have been designed. For instance, in mammals, SINE-like Alu repeats are scattered all over their genome and any nonspecific bands can be produced by performing single-primer amplification. Thus Alu-repeat polymorphisms can be detected when a primer complementary to any Alu repeat is used (Nelson *et al*. 1989; Sinnett *et al*. 1990).

RTE-based DNA marker applications have become a key part of research into genetic variability and diversity (Wu *et al*. 2018). The scope of their usage includes creation of genetic maps and the identification of individuals or lines carrying certain genetic polymorphic variations. The DNA marker system is related to developments in molecular genetics and biochemistry (Lewontin and Hubby 1966). Markers based on DNA polymorphisms have been developed because of the shortcomings of biochemical markers (Kan and Dozy 1978). This DNA marker system utilises "fingerprints" (i.e. distinctive patterns of DNA fragments) resolved by gel electrophoresis and by NGS. Molecular markers work by finding polymorphisms in a nucleotide sequence at a particular location in the genome. When this nucleotide sequence varies between the parents of the chosen cross, it can be discernible between plant accessions and its pattern of inheritance can be investigated. Molecular marker technologies have progressed immensely since NGS was introduced, enabling the implementation of many DNA fingerprinting methods.

It has been proven that RTE families evolve with different PCR profiles, but because of RTE evolution, RTE marker systems based on different RTEs show different amplification profiles and can be chosen to fit the required analysis (Leigh *et al*. 2003; Kalendar and Schulman 2006; Smykal *et al*. 2009). Retrotransposon insertions behave as Mendelian loci (Manninen *et al*. 2006; Tanhuanpaa *et al*. 2008). Thus retrotransposon-based markers would be expected to be codominant and involve a different level of genetic variability. Depending on DNA amplification or the NGS application, polymorphism detection tools can further be expanded by knowing nearby RTEs that are found in different orientations in the genome (head-to-head, tail-to-tail or head-to-tail).

Sequence-specific amplified polymorphism

Most of the retrotransposon PCR techniques are anonymous (unpredictable results before analysis), producing fingerprints from multiple sites of retrotransposon insertion in the genome. However, when one analyses closely related species, it is possible to predict the part of the common PCR amplicons expected in the sample via a phylogenetic approach (Kalendar *et al*. 2017). All of the techniques use the combination of a known retrotransposon sequence and a variety of adjacent sequences. The targets for PCR primers are generally designed for LTRs close to the joint in domains that are conserved within families but vary between families (Fig. 2). Although the internal regions of RTEs containing conserved segments could also be applied for this purpose, to minimise the distance between the targets to be amplified, the LTR is commonly chosen. Primer design needs to be done in both directions: primers facing outward from the left or 5′ LTR will necessarily face inward from the right or 3′ LTR because LTRs are direct repeats. Depending on the location and direction of the second primer, the inward-facing primer will either not amplify a product and produce a monomorphic band, or will detect a polymorphism resulting from a nested insertion pattern. To simplify the process, a retrotransposon-specific primer can be designed from an internal sequence that is

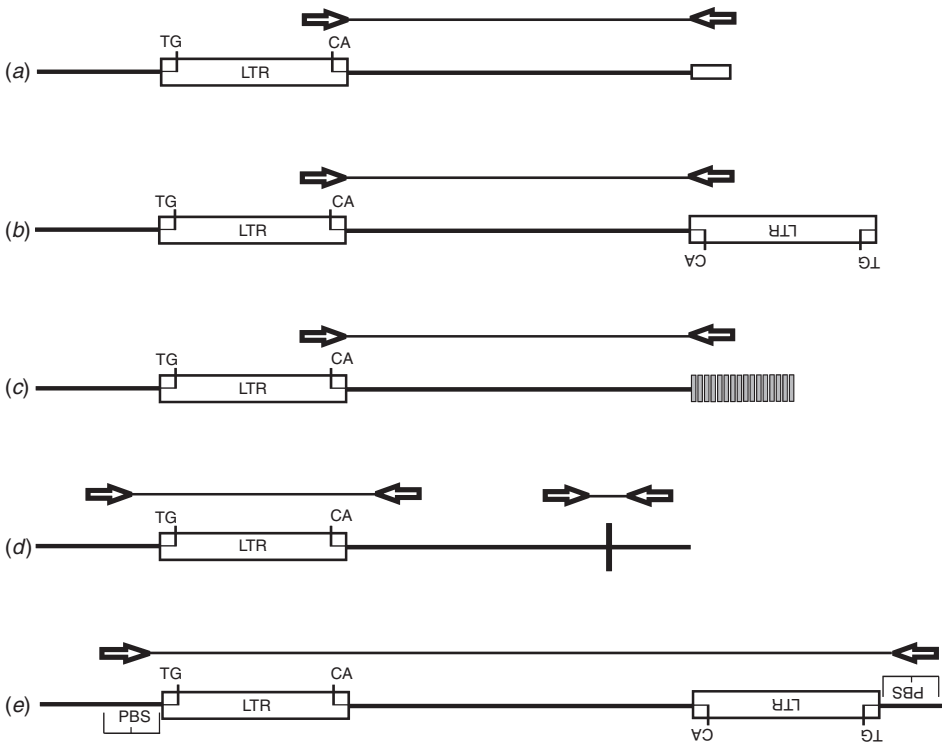

Fig. 2. Retrotransposon-based molecular marker methods. Multiplex products of various lengths from different loci are indicated by the bars above or beneath the diagrams for each reaction. Primers are indicated by arrows. (*a*) The sequence-specific amplified polymorphism method. The primers used for amplification match the adaptor (empty box) and retrotransposon (the long-terminal repeat (LTR) box). (*b*) The inter-retrotransposon amplified polymorphism method. Amplification takes place between retrotransposons (left and right LTR boxes) near each other in the genome (open bar), using retrotransposon primers. The elements are shown oriented head-to-head, using a single primer. (*c*) The retrotransposon microsatellite amplification polymorphisms method. Amplification takes place between a microsatellite domain (vertical bars) and a retrotransposon, using a primer anchored to the proximal side of the microsatellite and a retrotransposon primer. (*d*) Retrotransposon-based insertion polymorphism. Full sites, depicted on the left, are scored by amplification between a primer in the flanking genomic DNA and a retrotransposon primer. The single product is shown as one bar beneath the diagram. The alternative reaction between the primers for the left and right flanks is inhibited in the full site by the length of the retrotransposon. The product that is not amplified is indicated by a grey bar beneath the diagram. The flanking primers are able to amplify the empty site, on the right, depicted as a bar beneath the diagram. (*e*) The inter-primer-binding site amplification (iPBS) scheme and LTR retrotransposon structure. Two nested LTR retrotransposons in inverted orientations are amplified from a single primer or two different primers from primer binding sites. The PCR product contains both LTRs and PBS sequences as PCR primers in the termini. In the figure, the general structure for PBS and LTR sequences and the several-nucleotide-long spacer between the 5′ LTR and PBS are schematically shown.

present only once per element for retrotransposons with relatively short LTRs. Furthermore, simplified digestion and amplification protocols can be used for sequence-specific amplified polymorphism (S-SAP) for elements that have a low copy number (Waugh *et al*. 1997).

Retrotransposon marker systems differ according to the second primer used in the amplification reactions (Fig. 2). This primer can be any feature in the genome that is dispersed and conserved (Kalendar and Schulman 2014).

The amplified fragment length polymorphism (AFLP) method, proposed in the mid-1990s, is an anonymous marker method. Restriction sites in this method are detected by amplifying a subset of all the mobilisations for a given enzyme pair in the genome by PCR between ligated adapters (Vos *et al*. 1995). First reported by (Waugh *et al*. 1997), S-SAP is a modified AFLP method based on the *BARE-1* retroelement (Manninen and Schulman 1993). The foundation of this method is the cutting of genomic DNA using two different enzymes, which produces a template for the specific primer PCR: amplification between the retrotransposon and adapters ligated at restriction sites (generally *MseI* and *Pst*I or any restriction enzyme) using selective bases in the adaptor primer. S-SAP in general demonstrates a higher level of polymorphism than AFLPs, although it could be regarded as a modified version of AFLP. Usually, primers are designed in the LTR region but could coincide with the internal part of the

element as well. A nonselective primer could be employed when the copy number of the RTE is insufficient or when enzymes used for digestion have a larger recognition sequence. The number of discriminatory bases may be augmented for high-copy-number families. The usage of selective bases on the primers associated with the adapters or the use of two enzymes in S-SAP correlates with a reduction in genomic complexity. TEs with an insufficient copy number are not well suited for such a reduction in genomic complexity, but the use of single-enzyme digestion with discriminatory bases (or rare cutting enzymes) enables the surveying of all insertion sites for a given TE and can be regarded as a variant of anchored PCR.

The S-SAP marker system, based on LTR sequences of *Ty1-copia* retrotransposons, shows a greater level of polymorphism compared with AFLPs (Sorkheh *et al*. 2017). The S-SAP insertion patterns of the retrotransposon *Ty1-copia*-like element (*Tmc1*) in myrtle (*Myrtus communis* L.) were used to specifically characterise four myrtle accessions belonging to different areas in the province of Caserta in Italy. The high level of polymorphism detected in isolated LTRs makes *Tmc1* a good molecular marker for this species (D'Onofrio *et al*. 2010; Woodrow *et al*. 2010, 2012). Measuring the distribution and structure of a specific retroelement population in an organism is the main application of the S-SAP method. It has been used for evaluating the distribution and structure of specific retrotransposon populations in many plant species, including cereals, alfalfa (*Medicago sativa* L.), sweet potato (*Ipomoea batatas* (L.) Lam.), banana (*Musa acuminata* Colla), grapes (*Vitis vinifera* L.), pea (*Pisum sativum* L.) , *Iris* spp., cotton (*Gossypium hirsutum* L.), peanut (*Arachis hypogaea* L.), peppers (*Capsicum* spp.), tomato (*Solanum lycopersicum* L.), apple (*Malus* spp.), artichoke (*Cynara cardunculus* L.), lettuce (*Latuca sativa* L.) and flax (*Linum usitatissimum* L.) (Acquadro *et al*. 2006; Woodrow *et al*. 2010, 2012; Smykal *et al*. 2011; Galindo-Gonzalez *et al*. 2016; Sorkheh *et al*. 2017; Lee *et al*. 2018).

S-SAP generally displays more polymorphism dominance and a far greater chromosomal allocation compared with AFLP, but in order to ensure sites for adaptor ligation, as in the AFLP method, restriction digestion of genomic DNA is necessary for the S-SAP method. The sensitivity of the commonly used restriction enzymes to DNA methylation could generate false genotyping results. Retrotransposon-derived polymorphism can be used to differentiate among plant varieties, and the close association of numerous insertions with particular genes grants an advantageous source of potential mutations that could be related to phenotypic changes that result in diversifying processes.

When applied to DNA transposons, the same technique used for retrotransposons is termed transposon display (Van den Broeck *et al*. 1998). *Rim2/Hipa* transposon display yielded highly polymorphic profiles with ample reproducibility within a species as well as between species in the genus *Oryza* (Kwon *et al*. 2005).

Inter-repeat amplification polymorphism

Inter-repeat amplification polymorphism techniques such as inter-retrotransposon amplified polymorphism (IRAP), retrotransposon microsatellite amplification polymorphisms (REMAP) and inter-miniature inverted-repeat tandem element amplification have been used with abundant dispersed repeats such as the LTRs of retrotransposons and SINE-like sequences (inter-SINE amplified polymorphism) (Bureau and Wessler 1992; Kalendar and Schulman 2006, 2014). The amplification of a series of bands (DNA fingerprints) using primers homologous to these high-copy-number repeats is achievable because of the association of these sequences with each other, and the markers thus produced are very informative genetic markers. Retrotransposon insertional polymorphisms are detected by IRAP through amplification of the portion of DNA between two retroelements (Kalendar and Schulman 2006, 2014). Outwards from the LTR, one or two primers are used, and the tract of DNA between two nearby retrotransposons is thus amplified. In order to perform IRAP single primer matching, either the 5′ or 3′ end of the LTR could be used, oriented away from the LTR itself. Two primers could also be used when they are from the same retrotransposon element family or from different families. PCR products and consequently fingerprint patterns are an outcome of the amplification of hundreds to thousands of target sites in the genome. Retrotransposons generally tend to cluster together in 'repeat seas' surrounding 'genome islands' and may even nest within each other (Shirasu *et al*. 2000; Wicker and Keller 2007). Therefore, the pattern obtained will be related to the RTE copy number, the insertion pattern and the size of the RTE family.

The REMAP method is similar to IRAP, but one of the two primers is anchored to a microsatellite motif (Kalendar and Schulman 2006). Being spread throughout the genome, microsatellites appear to be associated with retrotransposons and have high mutation rates caused by polymerase slippage (Smykal *et al*. 2009). Therefore, they may show considerable variation at individual loci within a species. In REMAP, at the 3′ end of the microsatellite primer, anchor nucleotides are used to avoid slippage of the primer within the microsatellite site, which also prevents detection of the variation in repeat numbers within the microsatellite.

The IRAP and REMAP methods have been used in gene mapping in barley (*Hordeum vulgare* L.) (Manninen *et al*. 2000), wheat (*Triticum aestivum* L.) (Boyko *et al*. 2002; Vuorinen *et al*. 2018) and oats (*Avena sativa* L.) (Tanhuanpaa *et al*. 2007); in studies on genomic evolution in grasses (Vicient *et al*. 2001) and in a variety of applications, including measurement of genetic diversity and population structures, chromatin modification and epigenetic reprogramming, similarity and cladistic relationships, the determination of essential derivation, and marker-assisted selection (Kalendar *et al*. 2000; Belyayev *et al*. 2010; Smykal *et al*. 2011; Pakhrou *et al*. 2017; Paz *et al*. 2017; Sorkheh *et al*. 2017; Roy *et al*. 2018; Vuorinen *et al*. 2018).

Generally, IRAP and REMAP are carried out by using an agarose gel electrophoresis system; however, because of the large number of PCR products, S-SAP is used on sequencing gels (Kalendar and Schulman 2014). However, IRAP and REMAP can be used in NGS and yield 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 the number of

discriminatory bases in S-SAP, and the organisation of the plant genome.

Bands produced from IRAP techniques result in one side of a retrotransposon insertion. Sequencing of the isolated informative bands enables the design of a PCR primer corresponding to the flanking genomic DNA on one side of the insertion, assuming that the sequence is not repetitive and therefore unusable. However, in order to score the empty site, the genomic sequence flanking the other side of the element needs to be found, which can be performed by screening germplasm accessions that are polymorphic for the original band, followed by an S-SAP reaction on these, where the LTR primer is replaced with a primer designed for the known flank that faces towards the insertion site. Genetically inherited retrotransposon families can serve as markers that can ultimately protect the rights of breeders.

PCR primers from one species can be used on others because related species have phylogenetically related TE sequences. In this scenario, primers designed for conservative TE sequences are advantageous (Fig. 3). Being scattered over the whole chromosome, TEs are often mixed with other elements and repeats; thus PCR fingerprints can be improved if a combination of PCR primers is used.

A positive correlation has been 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 efficient PCR primers to reveal multiple bands for polymorphism detection (especially in the major cereals); organisms with a small genome, such as fungi, are the most difficult examples for RTE-based genetic marker development (Kalendar and Schulman 2014) (Fig. 4).

Generation of a virtually unlimited number of unique markers is possible through the combination of different LTR primers or by using combinations with microsatellite primers (REMAP) (Kalendar *et al*. 2017). The same primers produce completely different banding patterns depending on whether they are used alone or in combination, demonstrating that most IRAP and REMAP bands were derived from sequences bordered by an LTR or a microsatellite on one side and by another LTR on the other (Mandoulakani *et al*. 2015). In general, a more variable and stable pattern has been observed in IRAP than in inter simple sequence repeats (ISSR) or random amplification of polymorphic DNA (RAPD): frequently (but not always), depending on the LTR sequence, single priming PCR also shows less variability than the IRAP pattern with primer combinations (Sorkheh *et al*. 2017).

Inter-PBS amplification, a universal method for isolating and displaying retrotransposon polymorphisms

The main shortcoming of all RTE-based molecular marker techniques is the need for sequence information to design element-specific primers. Although rapid retrotransposon isolation methods have been designed based on PCR with a conservative primer for the TE, it might still be necessary to clone and sequence hundreds of clones or use NGS for the studied

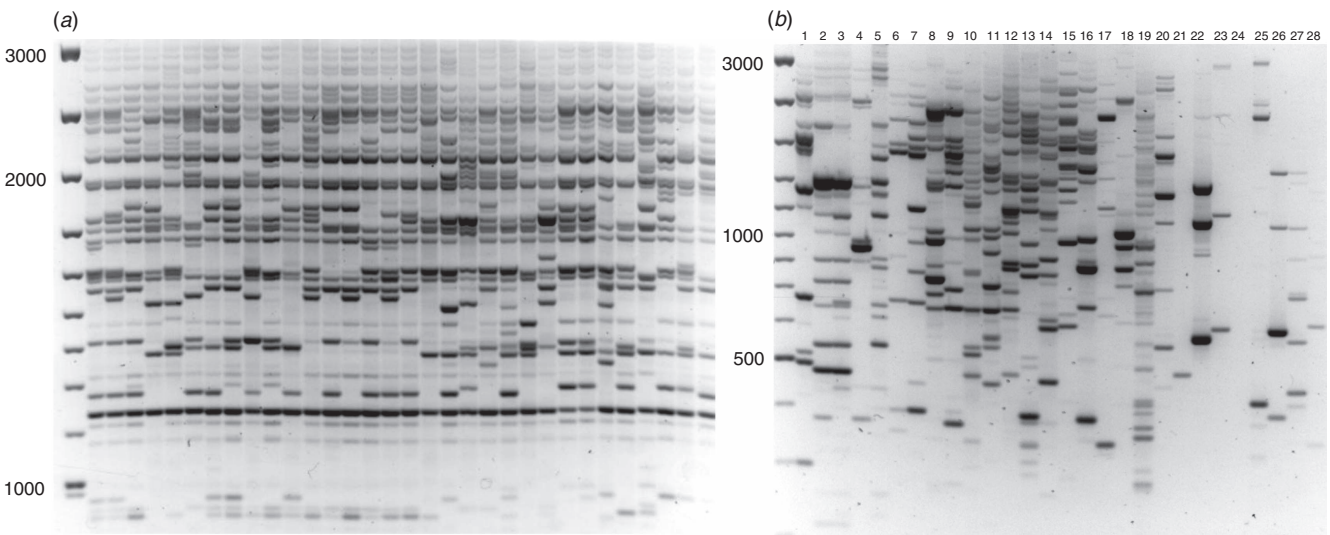

Fig. 3. (*a*) The use of inter-retrotransposon amplified polymorphism (IRAP) in the diversity analysis of plant species. A phenogram of 30 genotypes of populations of *Hordeum vulgare* based on IRAP analysis is shown as negative images of ethidium bromide-stained agarose gels following electrophoresis. Results for the LTR retrotransposon *Sukkula* (LTR primer 432: 5′-GATAGGGTCGCATCTTGGGCGTGAC-3′) are shown. A 100-bp DNA ladder is presented on the left. (*b*) IRAP fingerprints for *Triticeae* species with the same primer from a barley *Sukkula* LTR primer. 1, *Psathyrostachys fragilis* (Boiss.) Nevski; 2, *Triticum aestivum*; 3, *Triticum durum* Desf.; 4, *Aegilops tauschii* Coss.; 5, *Triticum dicoccoides* (Körn. ex Asch. & Graebn.) Schweinf.; 6, *Secale cereale* L.; 7, *Secale strictum* C.Presl., 8, *Hordeum erectifolium* Bothmer, N.Jacobsen & R.B.Jørg.; 9, *Hordeum pusillum*; 10, *Hordeum marinum* Huds.; 11, *Hordeum murinum* ssp. *glaucum* (Steud.) Tzvelev; 12, *Hordeum spontaneum* K.Koch; 13, *Hordeum patagonicum* (Hauman) Covas; 14, *Hordeum muticum* J.Presl; 15, *Hordeum roshevitzii* Bowden; 16, *Hordeum euclaston* Steud.; 17, *Hordeum brachyantherum* Nevski; 18, *Elymus repens* (L.) Gould; 19, *Eremopyrum distans* (K.Koch) Nevski; 20, *Eremopyrum triticeum* (Gaertn.) Nevski; 21, *Lophopyrum elongatum* (Host) Á.Löve; 22, *Taeniatherum caput-medusae* (L.) Nevski; 23, *Pseudoroegneria spicata* (Pursh) Á.Löve; 24, *Heteranthelium piliferum* (Sol.) Hochst. ex Jaub. & Spach; 25, *Amblyopyrum muticum* Eig; 26, *Comopyrum comosum* (Sm.) Á.Löve; 27, *Aegilops speltoides* Tausch; 28, *Dasypyrum vilosum* (L.) Borbás.

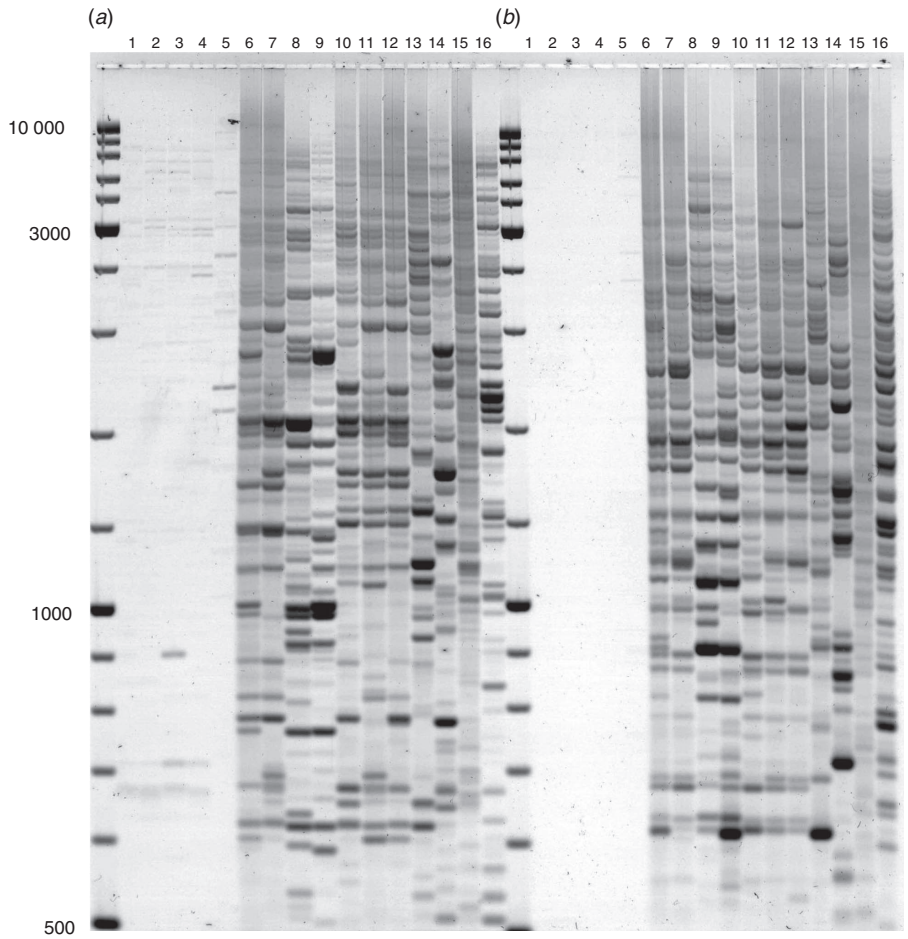

Fig. 4. The effectiveness of inter-retrotransposon amplified polymorphism (IRAP) amplification according to genome size. For the small genome of *Brachypodium distachyon* (L.) P.Beauv., there is no IRAP amplification, whereas for the large genomes of *Triticeae* species, multiple amplicons are observed. An IRAP gel produced with long-terminal repeat (LTR) primers: (*a*) the LTR retrotransposon *Sabrina* (primer 489: 5′-TCTCCCCTCCGGCAGGGTGC-3′) and (*b*) the LTR retrotransposon *Wham* (primer 515: 5′-ACACCCCCTATACTTGTGGGTCA-3′) are shown. A size marker is present on both sides, the GeneRuler DNA Ladder Mix (Thermo Fisher Scientific) (100–3000 bp), marked on the left in bp. DNA samples of *Triticeae* species with a small genome: 1–5, *Brachypodium distachyon* lines; species with a with a large genome: 6, *Triticum aestivum* (ABD); 7, *Triticum durum* (AB); 8–9 - *Aegilops tauschii* (D); 10–12, *Triticum dicoccoides* (AB); 13, *Aegilops peregrina* (Hack.) Maire & Weiller (S); 14, *Phleum pratense* L.; 15, *Avena sativa*; 16, *Secale strictum* (H4342).

genomes in order to obtain a few good primer sequences. Conserved motifs are not present in LTRs, which would allow their direct amplification by PCR. Based on the conservation of the reverse transcriptase domain, particularly for the *Ty1-copia* type, a few restrictions and adaptor-based methods for LTR cloning have been developed (Pearce *et al*. 1999). Major classes of retrotransposons include the Pseudoviridae (*Ty1-copia*), Metaviridae (*Ty3-gypsy*) and Retroposineae *LINE* (non-LTR) groups. PCR with degenerate primers can produce all the reverse transcribing elements. For instance, two degenerate *Ty1-copia* primers have been designed for the RT domain encoding *TAFLHG* and the reverse site *YVDDML*, also encoding *QMDVKT* and reverse *YVDDML* (Flavell *et al*. 1992; Hirochika and Hirochika 1993; Ellis *et al*. 1998). For the *Ty3-gypsy* element, degenerate primers have been designed for the RT domain encoding *RMCVDYR*, *LSGYHQI* or *YPLPRID*, and the reverse encoding sites *YAKLSKC* and *LSGYHQI*. The method based on reverse transcriptase can only be applied to the family of retrotransposons that contains this sequence. Therefore, for example, terminal-repeat retrotransposons in miniature or large retrotransposon derivatives and unknown classes of LTR retrotransposons cannot be found via this approach (Witte *et al*. 2001; Kalendar *et al*. 2004, 2008).

LTR retrotransposons and all retroviruses contain tRNA conservative PBS 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 leads to conversion of the viral or retrotransposon RNA genome to double-stranded DNA before its integration into the host DNA. The specific tRNA capture fluctuates or differs

among retroviruses and retroelements, but the process of reverse transcription is conserved among all retroviruses. All LTR retrotransposon sequences have primer binding sequences; therefore, there is the potential for an isolation method for retrotransposon LTRs to clone all possible LTR retrotransposons, since this method is based on the PBS sequence.

A generic and efficient method (inter-PBS (iPBS) amplification) has been developed that exploits the conserved parts of PBS sequences for direct visualisation of polymorphisms among individuals, the transcription profile of polymorphism and fast cloning of LTR parts from genomic DNA (Kalendar *et al*. 2010) (Fig. 2). This method permits the investigation of the LTR type of retrotransposons in any eukaryotic organism. It has been determined that primers designed to correspond with the conserved regions of the PBSs in LTR retrotransposons are very efficient in the PCR amplification of eukaryotic genomic DNA. Solitary PBS primers can only enhance nested reverse retrotransposons or sequences of related elements scattered through the genomic DNA. PCR amplification occurs between two nested PBS and consists of two LTR sequences. The PBS sequences are nested adjacent to each other in all eukaryotes. Most of the retrotransposons are blended, nested, reversed or edged in chromosomal sequences, and in all tested plant species, the amplification process has advanced readily with conservative PBS primers. Fragments of LTRs with the internal part of retrotransposons are in the neighbouring retrotransposons. Thus, PBS sequences are frequently located adjacent to each other, allowing the use of PBS sequences in cloning LTRs. The sites of the genome with a high density of retrotransposons can be applied to identify their chance association with other retrotransposons. New genome integrations result from an event, which means that retrotransposon activity or recombinations can be exploited to discern reproductively isolated plant lines (Qiu and Ungerer 2018). In this case, the amplified bands obtained from new inserts or recombinations will be polymorphic, appearing solely in plant lines where the insertions or recombination have occurred (Kalendar *et al*. 2010; Kalendar and Schulman 2014; Monden *et al*. 2014*a*, 2014*b*; Doungous *et al*. 2015; Coutinho *et al*. 2018).

Following the retrieval of the LTR sequences of a selected family of retrotransposons, they can be aligned to determine the most conserved region in them. Related plant species have conservative regions in LTRs for identical retroelements; therefore, conservative regions can be identified through the alignment of a few LTR sequences from one species or a mixture of sequences from related species. These conservative parts of LTR regions are used in the design of inverted primers for long-distance PCR, for cloning of the whole element and also for other inter-repeat amplification polymorphism techniques.

iPBS amplification is efficient in finding cDNA polymorphism and clonal differences resulting from retrotransposon activities or retrotransposon recombinations after crossing over and demonstrates roughly the same level of polymorphism as IRAP techniques (Kalendar *et al*. 2010). In order to obtain a vigorous, rapid and economical marker system for genotyping in plant breeding and marker-assisted selection, iPBS amplification was elaborated.

Further research on related varieties or breeding lines could be carried out through the development of a native RTE system, which requires the cloning and sequencing of elements from new a species by using iPBS amplification or a technique based on the conservation of the reverse transcriptase domain.

Next-generation sequencing allows small-scale, inexpensive genome sequencing with a turnaround time measured in days (Debladis *et al*. 2017; Qiu and Ungerer 2018). However, as NGS is generally performed and currently understood, all regions of the genome are sequenced with roughly equal probability, meaning that a large amount of a genomic sequence is collected and discarded to collect sequence information from the relatively low percentage of areas where the function is understood well enough to interpret potential mutations.

Nevertheless, the development of single-molecule sequencing technologies with long reads provides fresh perspectives on many aspects of genomics. The recent development of NGS enables the sequencing of a single molecule, and new possibilities for the detection of TE transposition events could arise for the generation of long reads. The new NGS platforms available from Pacific Biosciences and Oxford Nanopore Technologies enable the generation of reads that are kilobases long. This could improve the authenticity of disclosing novel TE insertions by ensuring enough sequence information to map new TE insertion sites accurately. The dependable genome-wide characterisation of structural variations either at a particular level (e.g. somatic variations) or within populations will aid in revealing novel functional aspects of genome dynamics in plants and animals (Debladis *et al*. 2017).

Use of RTEs to investigate genetic variability in plants

The study of genetic diversity and similarity between or within various populations, species and individuals is an essential objective in genetics. The application of various TEs enables the generation of a virtually unrestricted number of unique markers.

Completely different RTE amplification banding patterns are obtained if the same LTR primers are used alone or in combinations indicating that the majority of IRAP bands are derived from sequences bordered by one LTR or a microsatellite on one side, and by another LTR on the other side (Leigh *et al*. 2003; Kalendar and Schulman 2006; Boronnikova and Kalendar 2010; Hosid *et al*. 2012; Abdollahi Mandoulakani *et al*. 2015; Tanhuanpää *et al*. 2016).

Since related species have phylogenetically related TE sequences, PCR primers from one species can be used in another. In this case, primers designed for conservative TE sequences are advantageous. As TEs are dispersed throughout whole chromosomes and are very often mixed with other elements and repeats, combinations of primers from different repeats help to improve PCR fingerprinting.

To study genetic variation within varieties or breeding lines in a particular species, a native RTE system should be developed. This requires the cloning and sequencing of elements from a new species by using iPBS amplification, a technique based on the conservantion of the reverse transcriptase domain or genome sequencing with NGS. This process begins with the amplification and cloning of segments

between retrotransposon domains that are highly or universally conserved, the development of new primers specific for the retrotransposon families found and the testing of these for their efficacy as markers (Kalendar and Schulman 2014).

A marker from any of the anonymous multilocus RTE-based applications can be modified into an equivalent retrotransposon-based insertion polymorphism marker (Fig. 2) and vice versa (Jing *et al*. 2010, Jiang *et al*. 2015). Markers from the former methods are straightforward to harvest and can be rapidly analysed for their informativeness before investing in the advancement of a matching retrotransposon-based insertion polymorphism marker. One side of a retrotransposon insertion results in electrophoretically resolved bands from the inter-repeat amplification polymorphism techniques. Sequencing of the descriptive and sequestered bands will allow the design of a PCR primer matching the flanking genomic DNA on one side of the insertion, provided that the sequence is not monotonous and thus impractical. However, in order to score the empty site, the genomic sequence flanking the other side of the element is required. This can be achieved via the screening of polymorphic germplasm accessions for the initial band and then performing an S-SAP reaction on these, where the LTR primer is replaced by a primer composed for the known flank that faces towards the insertion site.

The application of TE-induced mutagenesis to link DNA sequences to functions has been demonstrated by the phenotype of knockout *Arabidopsis thaliana* (L.) Heynh. plants and, subsequently, by enzyme assays to encode the flavanol synthase gene (Wisman *et al*. 1998). These examples from flavonoid biosynthesis demonstrate the successful use of a TE-mutagenised population and PCR-based screens to assign gene functions unequivocally.

The plant genome contains families of all of the major TE classes, which are differently enriched in particular genomic regions. Whole genome sequencing with NGS and DNA methylation profiling of hundreds of natural accessions for several plant species (Chen *et al*. 2015; Underwood *et al*. 2017) have revealed that TEs exhibit significant intraspecific genetic and epigenetic variation, and that genetic variation often underlies epigenetic variation. Together, epigenetic modification and the forces of selection define the scope within which TEs can contribute to and control genome evolution.

Spontaneous interspecies crosses can induce TE activity, which may explain some of the new phenotypes observed (Vela *et al*. 2011; Guerreiro 2014; Debladis *et al*. 2017). TEs may also play a role in the diploidisation that follows polyploidisation events (Vicient and Casacuberta 2017). Investigating the multiple factors controlling TE dynamics and the nature of ancient and recent polyploid genomes may shed light on these processes.

Epigenetic control and retrotransposon activity

Retrotransposons can rapidly increase in copy number as a result of periodic bursts of transposition. Such bursts are mutagenic and thus potentially deleterious. The methylation status of TEs in plants has been correlated with lowered transcription of genes with TE insertions. Also, more systematic knowledge is needed about the influence of stress or environmental cues on the epigenetic control of retrotransposons, as well as the impact of TEs on phenotypic plasticity (Shang *et al*. 2017; Xia *et al*. 2017). The stochastic and sometimes incomplete nature of the epigenetic silencing of retrotransposons may help explain stress survival, heterosis and the genome dominance phenomenon for intraspecific hybrids (Guerreiro 2014; Fultz and Slotkin 2017; Gaubert *et al*. 2017; Zhou *et al*. 2018). Repetitive element mobilisation represents a destabilising process for the host cell. Several mechanisms such as DNA and histone methylation, and RNAi actively suppress retrotransposon expression (Vetukuri *et al*. 2011; Fultz and Slotkin 2017; Zakrzewski *et al*. 2017). 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 (Cho 2018).

In the plant genome, insertional inactivation and other genome rearrangements lead to a wide spectrum of recombination and chromosomal instability (Raskina *et al*. 2008; Belyayev *et al*. 2010; Brueckner *et al*. 2012; Hosid *et al*. 2012). RTE-induced genetic rearrangements can lead to nonallelic homologous recombination (Yu *et al*. 2012; Ben-David *et al*. 2013) or insertional mutagenesis caused by retrotransposons 'hopping' within gene coding sequences; it causes diverse effects on target gene expression, depending on the intragenic location, the orientation, the length of the inserted sequence and other factors, or the activation and mobilisation of small RNAs (Nuthikattu *et al*. 2013; Forestan *et al*. 2017; Masuta *et al*. 2017; Schorn *et al*. 2017). For example, the genes with nearby TE insertions are those most strongly affected by RNA polymerase IV-mediated gene silencing. The modulation of nearby gene expression by TEs is linked to alternative methylation profiles on gene flanking regions, and these profiles are strictly dependent on the specific characteristics of the TE member inserted (Forestan *et al*. 2017).

TEs have been found to be associated with microRNAs (miRNAs), small noncoding RNAs responsible for regulating the activities of 60–70% of genes in an organism. Other small noncoding RNAs that repetitive elements have been associated with include small interfering RNA (siRNA), which can silence repetitive elements through post-transcriptional gene silencing mechanisms by creating feedback loops. Besides controlling repetitive elements, the transcriptional activity of repetitive elements can also enable the tissue-specific expression of certain genes (Debladis *et al*. 2017). A fair number of expressive repetitive elements have been linked to the biogenesis of small RNAs or siRNA, some of which are involved in gene regulation in either a *cis* or *trans* manner. Although some sRNAs participate in post-transcriptional gene silencing, other RNAs are involved in *de novo* DNA methylation in the plant genome. Following an increasing number of reports, sRNAs are now thought to be core members of post-transcriptional as well as RNA-directed DNA methylation-based transcriptional gene regulatory processes (Nuthikattu *et al*. 2013; Forestan *et al*. 2017). The involvement of repetitive elements in the biogenesis of sRNAs indicates their importance in the gene regulatory system of plant species.

Long noncoding RNAs (lncRNAs) derived from TEs often appear under specific stress conditions and exhibit a

tissue-specific expression pattern (Paszkowski 2015; Wang *et al*. 2017). The TEs that are associated with the tissue-specificity of lncRNA expression can serve as one of the functional elements in lncRNAs (Chishima *et al*. 2018; Cho 2018).

Conclusions

Transposable elements are highly abundant mobile genetic elements that comprise multiple classes and constitute a large part of most eukaryotic genomes. Depending on the mechanism of transposition, they are mainly divided into transposons and retrotransposons. The movement and accumulation of TEs has been a major force in shaping the genes and genomes of almost all organisms. TEs are a source of chromatin instability and genomic rearrangements with deleterious consequences, and are important drivers of species diversity. They exhibit great variety in their structure, size and mechanisms of transposition, making them important putative actors in genome evolution. TEs can also impact gene regulation simply by inserting their own internal regulatory sequences (promoters, enhancers) in new genomic loci upon insertion. A high proportion of TEs have lost their autonomous transposition ability because of point mutations, deletions or both, and many of them appear to embody defective elements with deletions.

A large-scale analysis of genome sequencing data revealed that the TE landscape is very dynamic, and transcriptomic, epigenomic as well as phenotypic variations are attributed to TEs. TEs are a common component in many epigenetic mechanisms and represent a massive reservoir of potential genomic instability and RNA-level toxicity. Newly inserted TEs create instability and influence the gene expression of flanking regions by modifying their methylation status. Many TEs appear to be static and nonfunctional. However, some TEs are capable of replicating and mobilising to new positions in the genome, and even immobile TE copies can be expressed as somatic transposition events that have been observed in plant development. Only retrotransposon insertions that are passed into egg cells and pollen are inherited. Thus they could possibly be considered to be sexually transmitted diseases, but ones that move by cellular rather than extracellular pathways into the new host.

Many features of TEs, such as their ubiquity, abundance and dispersion in the eukaryotic genome, make them appealing as a basis for molecular marker systems. Genome diversification results from their activity, which provides a means for its detection. Their integration can be detected by conserved sequences. TEs are long and produce a sizable genetic change at the point of insertion, thereby providing conserved sequences that can be used to detect their own integration. Various applications have been developed to exploit polymorphisms in TE insertion patterns, including conventional or anchored PCR, and quantitative or digital PCR with primers designed for the 5′ or 3′ junction. The retrotransposon junction can be mapped by high-throughput NGS and bioinformatics. According to these ‘transposon display’ applications, the TE insertion can be rapidly, easily, conveniently and accurately identified, or a new TE insertion can be found. The applications range from investigations of retrotransposon activation and mobility to studies on biodiversity, genome evolution, chromatin modification, epigenetic reprogramming, the mapping of genes and the estimation of genetic distance, as well as assessment of the essential derivation of varieties, the detection of somaclonal variation and study of the tissue-specificity of noncoding RNA expression.

The development of single-molecule sequencing technologies with long reads presents novel opportunities for many aspects of genomics and could open new prospects for detecting unique TE transposition events, as well as the problematic investigation of TE movements during the lifecycle. The trustworthy genome-wide characterisation of structural variations, either at the individual level (single cell and single TE) or within populations, will aid in revealing novel functional aspects of genome dynamics in plants and animals.

Conflicts of interest

The authors declare they have no conflicts of interest.

Acknowledgements

This work was supported by the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan in the framework of the program funding for research (BR05236574 and AP05130266). The authors thank Dr Roy Siddall (University of Helsinki) for outstanding editing and proofreading of the manuscript.

References

Abdollahi Mandoulakani B, Yaniv E, Kalendar R, Raats D, Bariana HS, Bihamta MR, Schulman AH (2015) Development of IRAP- and REMAP-derived SCAR markers for marker-assisted selection of the stripe rust resistance gene *Yr15* derived from wild emmer wheat. *Theoretical and Applied Genetics* **128**, 211–219. doi:10.1007/s00122-014-2422-8

Acquadro A, Portis E, Moglia A, Magurno F, Lanteri S (2006) Retrotransposon-based S-SAP as a platform for the analysis of genetic variation and linkage in globe artichoke. *Genome* **49**, 1149–1159. doi:10.1139/g06-074

Antonius-Klemola K, Kalendar R, Schulman A (2006) TRIM retrotransposons occur in apple and are polymorphic between varieties but not sports. *Theoretical and Applied Genetics* **112**, 999–1008. doi:10.1007/s00122-005-0203-0

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

Ben-David S, Yaakov B, Kashkush K (2013) Genome-wide analysis of short interspersed nuclear elements SINES revealed high sequence conservation, gene association and retrotranspositional activity in wheat. *The Plant Journal* **76**, 201–210. doi:10.1111/tpj.12285

Boronnikova SV, Kalendar RN (2010) Using IRAP markers for analysis of genetic variability in populations of resource and rare species of plants. *Russian Journal of Genetics* **46**, 36–42. doi:10.1134/S1022795410010060

Boyko E, Kalendar R, Korzun V, Fellers J, Korol A, Schulman AH, Gill BS (2002) A high-density cytogenetic map of the *Aegilops tauschii* genome incorporating retrotransposons and defense-related genes: insights into cereal chromosome structure and function. *Plant Molecular Biology* **48**, 767–790. doi:10.1023/A:1014831511810

Brueckner LM, Sagulenko E, Hess EM, Zheglo D, Blumrich A, Schwab M, Savelyeva L (2012) Genomic rearrangements at the *FRA2H* common fragile site frequently involve non-homologous recombination events across LTR and L1(LINE) repeats. *Human Genetics* **131**, 1345–1359. doi:10.1007/s00439-012-1165-3

Bureau TE, Wessler SR (1992) Tourist: a large family of small inverted repeat elements frequently associated with maize genes. *The Plant Cell* **4**, 1283–1294. doi:10.1105/tpc.4.10.1283

Chen H, Yang P, Guo J, Kwoh CK, Przytycka TM, Zheng J (2015) ARG-walker: inference of individual specific strengths of meiotic recombination hotspots by population genomics analysis. *BMC Genomics* **16**(Suppl 12), S1. doi:10.1186/1471-2164-16-S12-S1

Chishima T, Iwakiri J, Hamada M (2018) Identification of transposable elements contributing to tissue-specific expression of long non-coding RNAs. *Genes* **9**(1), 23. doi:10.3390/genes9010023

Cho J (2018) Transposon-derived non-coding RNAs and their function in plants. *Frontiers in Plant Science* **9**, 600. doi:10.3389/fpls.2018.00600

Coutinho JP, Carvalho A, Martin A, Lima-Brito J (2018) Molecular characterization of Fagaceae species using inter-primer binding site (iPBS) markers. *Molecular Biology Reports* **45**, 133–142. doi:10.1007/s11033-018-4146-3

D'Onofrio C, De Lorenzis G, Giordani T, Natali L, Cavallini A, Scalabrelli G (2010) Retrotransposon-based molecular markers for grapevine species and cultivars identification. *Tree Genetics & Genomes* **6**, 451–466. doi:10.1007/s11295-009-0263-4

Debladis E, Llauro C, Carpentier MC, Mirouze M, Panaud O (2017) Detection of active transposable elements in *Arabidopsis thaliana* using Oxford Nanopore Sequencing technology. *BMC Genomics* **18**, 537. doi:10.1186/s12864-017-3753-z

Dotto BR, Carvalho EL, Silva AF, Duarte Silva LF, Pinto PM, Ortiz MF, Wallau GL (2015) HTT-DB: horizontally transferred transposable elements database. *Bioinformatics* **31**, 2915–2917. doi:10.1093/bioinformatics/btv281

Doungous O, Kalendar R, Adiobo A, Schulman AH (2015) Retrotransposon molecular markers resolve cocoyam (*Xanthosoma sagittifolium*) and taro (*Colocasia esculenta*) by type and variety. *Euphytica* **206**, 541–554. doi:10.1007/s10681-015-1537-6

Ellis THN, Poyser SJ, Knox MR, Vershinin AV, Ambrose MJ (1998) Polymorphism of insertion sites of *Ty1-copia* class retrotransposons and its use for linkage and diversity analysis in pea. *Molecular & General Genetics* **260**, 9–19. doi:10.1007/PL00008630

Finnegan DJ (1990) Transposable elements and DNA transposition in eukaryotes. *Current Opinion in Cell Biology* **2**, 471–477. doi:10.1016/0955-0674(90)90130-7

Flavell AJ, Dunbar E, Anderson R, Pearce SR, Hartley R, Kumar A (1992) *Ty1-copia* group retrotransposons are ubiquitous and heterogeneous in higher plants. *Nucleic Acids Research* **20**, 3639–3644. doi:10.1093/nar/20.14.3639

Forestan C, Farinati S, Aiese Cigliano R, Lunardon A, Sanseverino W, Varotto S (2017) Maize RNA PolIV affects the expression of genes with nearby TE insertions and has a genome-wide repressive impact on transcription. *BMC Plant Biology* **17**, 161. doi:10.1186/s12870-017-1108-1

Frankel AD, Young JA (1998) HIV-1: fifteen proteins and an RNA. *Annual Review of Biochemistry* **67**, 1–25. doi:10.1146/annurev.biochem.67.1.1

Fultz D, Slotkin RK (2017) Exogenous transposable elements circumvent identity-based silencing, permitting the dissection of expression-dependent silencing. *The Plant Cell* **29**, 360–376. doi:10.1105/tpc.16.00718

Galindo-Gonzalez L, Mhiri C, Grandbastien MA, Deyholos MK (2016) *Ty1-copia* elements reveal diverse insertion sites linked to polymorphisms among flax (*Linum usitatissimum* L.) accessions. *BMC Genomics* **17**, 1002. doi:10.1186/s12864-016-3337-3

Gao D, Chu Y, Xia H, Xu C, Heyduk K, Abernathy B, Ozias-Akins P, Leebens-Mack JH, Jackson SA (2018) Horizontal transfer of non-LTR retrotransposons from arthropods to flowering plants. *Molecular Biology and Evolution* **35**, 354–364. doi:10.1093/molbev/msx275

Gaubert H, Sanchez DH, Drost HG, Paszkowski J (2017) Developmental restriction of retrotransposition activated in *Arabidopsis* by environmental stress. *Genetics* **207**, 813–821. doi:10.1534/genetics.117.300103

Goke J, Ng HH (2016) CTRL+INSERT: retrotransposons and their contribution to regulation and innovation of the transcriptome. *EMBO Reports* **17**, 1131–1144. doi:10.15252/embr.201642743

Guerreiro MPG (2014) Interspecific hybridization as a genomic stressor inducing mobilization of transposable elements in *Drosophila*. *Mobile Genetic Elements* **4**. doi:10.4161/mge.34394

Hirochika H, Hirochika R (1993) *Ty1-copia* group retrotransposons as ubiquitous components of plant genomes. *Japanese Journal of Genetics* **68**, 35–46. doi:10.1266/jjg.68.35

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

Ivancevic AM, Kortschak RD, Bertozzi T, Adelson DL (2016) LINEs between species: evolutionary dynamics of LINE-1 retrotransposons across the eukaryotic tree of life. *Genome Biology and Evolution* **8**, 3301–3322. doi:10.1093/gbe/evw243

Jiang S, Zong Y, Yue X, Postman J, Teng Y, Cai D (2015) Prediction of retrotransposons and assessment of genetic variability based on developed retrotransposon-based insertion polymorphism (RBIP) markers in *Pyrus* L. *Molecular Genetics and Genomics* **290**, 225–237. doi:10.1007/s00438-014-0914-5

Jing R, Vershinin A, Grzebyta J, Shaw P, Smykal P, Marshall D, Ambrose MJ, Ellis TH, Flavell AJ (2010) The genetic diversity and evolution of field pea (*Pisum*) studied by high throughput retrotransposon based insertion polymorphism (RBIP) marker analysis. *BMC Evolutionary Biology* **10**, 44. doi:10.1186/1471-2148-10-44

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

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

Kalendar R, Schulman AH (2014) Transposon-based tagging: IRAP, REMAP, and iPBS. *Methods in Molecular Biology (Clifton, N.J.)* **1115**, 233–255. doi:10.1007/978-1-62703-767-9_12

Kalendar R, Tanskanen J, Immonen S, Nevo E, Schulman AH (2000) Genome evolution of wild barley (*Hordeum spontaneum*) by *BARE*-1 retrotransposon dynamics in response to sharp microclimatic divergence. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 6603–6607. doi:10.1073/pnas.110587497

Kalendar R, Vicient CM, Peleg O, Anamthawat-Jonsson K, Bolshoy A, Schulman AH (2004) Large retrotransposon derivatives: abundant, conserved but nonautonomous retroelements of barley and related genomes. *Genetics* **166**, 1437–1450. doi:10.1534/genetics.166.3.1437

Kalendar R, Tanskanen J, Chang W, Antonius K, Sela H, Peleg O, Schulman AH (2008) Cassandra retrotransposons carry independently transcribed 5S RNA. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 5833–5838. doi:10.1073/pnas.0709698105

Kalendar R, Antonius K, Smykal P, Schulman AH (2010) iPBS: a universal method for DNA fingerprinting and retrotransposon isolation. *Theoretical and Applied Genetics* **121**, 1419–1430. doi:10.1007/s00122-010-1398-2

Kalendar R, Khassenov B, Ramanculov E, Samuilova O, Ivanov KI (2017) FastPCR: an *in silico* tool for fast primer and probe design and advanced sequence analysis. *Genomics* **109**, 312–319. doi:10.1016/j.ygeno.2017.05.005

Kan YW, Dozy AM (1978) Polymorphism of DNA sequence adjacent to human beta-globin structural gene: relationship to sickle mutation. *Proceedings of the National Academy of Sciences of the United States of America* **75**, 5631–5635. doi:10.1073/pnas.75.11.5631

Kim S, Park J, Yeom SI, Kim YM, Seo E, Kim KT, Kim MS, Lee JM, Cheong K, Shin HS, Kim SB, Han K, Lee J, Park M, Lee HA, Lee HY, Lee Y, Oh S, Lee JH, Choi E, Choi E, Lee SE, Jeon J, Kim H, Choi G, Song H, Lee J, Lee SC, Kwon JK, Lee HY, Koo N, Hong Y, Kim RW, Kang WH, Huh JH, Kang BC, Yang TJ, Lee YH, Bennetzen JL, Choi D (2017) New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication. *Genome Biology* **18**, 210. doi:10.1186/s13059-017-1341-9

Klein SJ, O'Neill RJ (2018) Transposable elements: genome innovation, chromosome diversity, and centromere conflict. *Chromosome Research* **26**, 5–23. doi:10.1007/s10577-017-9569-5

Kwon SJ, Park KC, Kim JH, Lee JK, Kim NS (2005) *Rim 2/Hipa* CACTA transposon display: a new genetic marker technique in *Oryza* species. *BMC Genetics* **6**, 15. doi:10.1186/1471-2156-6-15

Lee SI, Gim JA, Lim MJ, Kim HS, Nam BH, Kim NS (2018) *Ty3/Gypsy* retrotransposons in the Pacific abalone *Haliotis discus hannai*: characterization and use for species identification in the genus *Haliotis*. *Genes & Genomics* **40**, 177–187. doi:10.1007/s13258-017-0619-3

Leigh F, Kalendar R, Lea V, Lee D, Donini P, Schulman AH (2003) Comparison of the utility of barley retrotransposon families for genetic analysis by molecular marker techniques. *Molecular Genetics and Genomics* **269**, 464–474. doi:10.1007/s00438-003-0850-2

Levin HL (1995) A novel mechanism of self-primed reverse transcription defines a new family of retroelements. *Molecular and Cellular Biology* **15**, 3310–3317. doi:10.1128/MCB.15.6.3310

Lewontin RC, Hubby JL (1966) 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* **54**, 595–609.

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

Mandoulakani BA, Yaniv E, Kalendar R, Raats D, Bariana HS, Bihamta MR, Schulman AH (2015) Development of IRAP- and REMAP-derived SCAR markers for marker-assisted selection of the stripe rust resistance gene *Yr15* derived from wild emmer wheat. *Theoretical and Applied Genetics* **128**, 211–219. doi:10.1007/s00122-014-2422-8

Manninen I, Schulman AH (1993) *BARE-1*, a *copia*-like retroelement in barley (*Hordeum vulgare* L.). *Plant Molecular Biology* **22**, 829–846. doi:10.1007/BF00027369

Manninen O, Kalendar R, Robinson J, Schulman A (2000) Application of *BARE-1* retrotransposon markers to the mapping of a major resistance gene for net blotch in barley. *Molecular & General Genetics* **264**, 325–334. doi:10.1007/s004380000326

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

Martinez G, Slotkin RK (2012) Developmental relaxation of transposable element silencing in plants: functional or byproduct? *Current Opinion in Plant Biology* **15**, 496–502. doi:10.1016/j.pbi.2012.09.001

Mascagni F, Giordani T, Ceccarelli M, Cavallini A, Natali L (2017) Genome-wide analysis of LTR-retrotransposon diversity and its impact on the evolution of the genus *Helianthus* (L.). *BMC Genomics* **18**, 634. doi:10.1186/s12864-017-4050-6

Masuta Y, Nozawa K, Takagi H, Yaegashi H, Tanaka K, Ito T, Saito H, Kobayashi H, Matsunaga W, Masuda S, Kato A, Ito H (2017) Inducible transposition of a heat-activated retrotransposon in tissue culture. *Plant & Cell Physiology* **58**, 375–384. doi:10.1093/pcp/pcw202

Masuta Y, Kawabe A, Nozawa K, Naito K, Kato A, Ito H (2018) Characterization of a heat-activated retrotransposon in *Vigna angularis*. *Breeding Science* **68**, 168–176. doi:10.1270/jsbbs.17085

Moisy C, Schulman AH, Kalendar R, Buchmann JP, Pelsy F (2014) The *Tvv1* retrotransposon family is conserved between plant genomes separated by over 100 million years. *Theoretical and Applied Genetics* **127**, 1223–1235. doi:10.1007/s00122-014-2293-z

Monden Y, Fujii N, Yamaguchi K, Ikeo K, Nakazawa Y, Waki T, Hirashima K, Uchimura Y, Tahara M (2014*a*) Efficient screening of long terminal repeat retrotransposons that show high insertion polymorphism via high-throughput sequencing of the primer binding site. *Genome* **57**, 245–252. doi:10.1139/gen-2014-0031

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

Nelson DL, Ledbetter SA, Corbo L, Victoria MF, Ramirez-Solis R, Webster TD, Ledbetter DH, Caskey CT (1989) Alu polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 6686–6690. doi:10.1073/pnas.86.17.6686

Neumann P, Navratilova A, Koblizkova A, Kejnovsky E, Hribova E, Hobza R, Widmer A, Dolezel J, Macas J (2011) Plant centromeric retrotransposons: a structural and cytogenetic perspective. *Mobile DNA* **2**, 4. doi:10.1186/1759-8753-2-4

Nuthikattu S, McCue AD, Panda K, Fultz D, DeFraia C, Thomas EN, Slotkin RK (2013) The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21–22 nucleotide small interfering RNAs. *Plant Physiology* **162**, 116–131. doi:10.1104/pp.113.216481

Pakhrou O, Medraoui L, Yatrib C, Alami M, Filali-Maltouf A, Belkadi B (2017) Assessment of genetic diversity and population structure of an endemic Moroccan tree (*Argania spinosa* L.) based in IRAP and ISSR markers and implications for conservation. *Physiology and Molecular Biology of Plants* **23**, 651–661. doi:10.1007/s12298-017-0446-7

Paszkowski J (2015) Controlled activation of retrotransposition for plant breeding. *Current Opinion in Biotechnology* **32**, 200–206. doi:10.1016/j.copbio.2015.01.003

Paz RC, Kozaczek ME, Rosli HG, Andino NP, Sanchez-Puerta MV (2017) Diversity, distribution and dynamics of full-length *Copia* and *Gypsy* LTR retroelements in *Solanum lycopersicum*. *Genetica* **145**, 417–430. doi:10.1007/s10709-017-9977-7

Pearce SR, Harrison G, Li D, Heslop-Harrison J, Kumar A, Flavell AJ (1996) The *Ty1-copia* group retrotransposons in *Vicia* species: copy number, sequence heterogeneity and chromosomal localisation. *Molecular and General Genetics MGG* **250**, 305–315. doi:10.1007/BF02174388

Pearce SR, Stuart-Rogers C, Knox MR, Kumar A, Ellis TH, Flavell AJ (1999) Rapid isolation of plant *Ty1-copia* group retrotransposon LTR sequences for molecular marker studies. *The Plant Journal* **19**, 711–717. doi:10.1046/j.1365-313x.1999.00556.x

Piegu B, Bire S, Arensburger P, Bigot Y (2015) A survey of transposable element classification systems – a call for a fundamental update to meet the challenge of their diversity and complexity. *Molecular Phylogenetics and Evolution* **86**, 90–109. doi:10.1016/j.ympev.2015.03.009

Qiu F, Ungerer MC (2018) Genomic abundance and transcriptional activity of diverse *gypsy* and *copia* long terminal repeat retrotransposons in three wild sunflower species. *BMC Plant Biology* **18**, 6. doi:10.1186/s12870-017-1223-z

Raskina O, Barber JC, Nevo E, Belyayev A (2008) Repetitive DNA and chromosomal rearrangements: speciation-related events in plant genomes. *Cytogenetic and Genome Research* **120**, 351–357. doi:10.1159/000121084

Roy NS, Lee SI, Nkongolo K, Kim NS (2018) Retrotransposons in *Betula nana*, and interspecific relationships in the Betuloideae, based on inter-retrotransposon amplified polymorphism (IRAP) markers. *Genes & Genomics* **40**, 511–519. doi:10.1007/s13258-018-0655-7

Sanchez DH, Gaubert H, Drost HG, Zabet NR, Paszkowski J (2017) High-frequency recombination between members of an LTR retrotransposon family during transposition bursts. *Nature Communications* **8**, 1283. doi:10.1038/s41467-017-01374-x

Schorn AJ, Gutbrod MJ, LeBlanc C, Martienssen R (2017) LTR-retrotransposon control by tRNA-derived small RNAs. *Cell* **170**, 61–71.e11. doi:10.1016/j.cell.2017.06.013

Schulman AH (2013) Retrotransposon replication in plants. *Current Opinion in Virology* **3**, 604–614. doi:10.1016/j.coviro.2013.08.009

Shang Y, Yang F, Schulman AH, Zhu J, Jia Y, Wang J, Zhang XQ, Jia Q, Hua W, Yang J, Li C (2017) Gene deletion in barley mediated by LTR-retrotransposon *BARE*. *Scientific Reports* **7**, 43766. doi:10.1038/srep43766

Shirasu K, Schulman AH, Lahaye T, Schulze-Lefert P (2000) A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion. *Genome Research* **10**, 908–915. doi:10.1101/gr.10.7.908

Sinnett D, Deragon JM, Simard LR, Labuda D (1990) Alumorphs–human DNA polymorphisms detected by polymerase chain reaction using *Alu*-specific primers. *Genomics* **7**, 331–334. doi:10.1016/0888-7543(90)90166-R

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

Smykal P, Bacova-Kerteszova N, Kalendar R, Corander J, Schulman AH, Pavelek M (2011) Genetic diversity of cultivated flax (*Linum usitatissimum* L.) germplasm assessed by retrotransposon-based markers. *Theoretical and Applied Genetics* **122**, 1385–1397. doi:10.1007/s00122-011-1539-2

Sorkheh K, Dehkordi MK, Ercisli S, Hegedus A, Halasz J (2017) Comparison of traditional and new generation DNA markers declares high genetic diversity and differentiated population structure of wild almond species. *Scientific Reports* **7**, 5966. doi:10.1038/s41598-017-06084-4

Tanhuanpaa P, Kalendar R, Schulman AH, Kiviharju E (2007) A major gene for grain cadmium accumulation in oat (*Avena sativa* L.). *Genome* **50**, 588–594. doi:10.1139/G07-036

Tanhuanpaa P, Kalendar R, Schulman AH, Kiviharju E (2008) The first doubled haploid linkage map for cultivated oat. *Genome* **51**, 560–569. doi:10.1139/G08-040

Tanhuanpää P, Erkkilä M, Kalendar R, Schulman AH, Manninen O (2016) Assessment of genetic diversity in Nordic timothy (*Phleum pratense* L.). *Hereditas* **153**, 5. doi:10.1186/s41065-016-0009-x

Underwood CJ, Henderson IR, Martienssen RA (2017) Genetic and epigenetic variation of transposable elements in *Arabidopsis*. *Current Opinion in Plant Biology* **36**, 135–141. doi:10.1016/j.pbi.2017.03.002

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

Van Sluys MA, Tempe J, Fedoroff N (1987) Studies on the introduction and mobility of the maize *Activator* element in *Arabidopsis thaliana* and *Daucus carota*. *The EMBO Journal* **6**, 3881–3889.

Vela D, Guerreiro MP, Fontdevila A (2011) Adaptation of the AFLP technique as a new tool to detect genetic instability and transposition in interspecific hybrids. *BioTechniques* **50**, 247–250. doi:10.2144/000113655

Vetukuri RR, Tian Z, Avrova AO, Savenkov EI, Dixelius C, Whisson SC (2011) Silencing of the *PiAvr3a* effector-encoding gene from *Phytophthora infestans* by transcriptional fusion to a short interspersed element. *Fungal Biology* **115**, 1225–1233. doi:10.1016/j.funbio.2011.08.007

Vicient CM, Casacuberta JM (2017) Impact of transposable elements on polyploid plant genomes. *Annals of Botany* **120**, 195–207. doi:10.1093/aob/mcx078

Vicient CM, Kalendar R, Schulman AH (2001) Envelope-class retrovirus-like elements are widespread, transcribed and spliced, and insertionally polymorphic in plants. *Genome Research* **11**, 2041–2049. doi:10.1101/gr.193301

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* **23**, 4407–4414. doi:10.1093/nar/23.21.4407

Voytas DF, Cummings MP, Koniczny A, Ausubel FM, Rodermel SR (1992) *Copia*-like retrotransposons are ubiquitous among plants. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 7124–7128. doi:10.1073/pnas.89.15.7124

Vuorinen A, Kalendar R, Fahima T, Korpelainen H, Nevo E, Schulman A (2018) Retrotransposon-based genetic diversity assessment in wild emmer wheat (*Triticum turgidum* ssp. *dicoccoides*). *Agronomy (Basel)* **8**, 107. doi:10.3390/agronomy8070107

Wang D, Qu Z, Yang L, Zhang Q, Liu ZH, Do T, Adelson DL, Wang ZY, Searle I, Zhu JK (2017) Transposable elements (TEs) contribute to stress-related long intergenic noncoding RNAs in plants. *The Plant Journal* **90**, 133–146. doi:10.1111/tpj.13481

Waugh R, McLean K, Flavell AJ, Pearce SR, Kumar A, Thomas BB, Powell W (1997) Genetic distribution of *Bare-1*-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). *Molecular & General Genetics* **253**, 687–694. doi:10.1007/s004380050372

Wicker T, Keller B (2007) Genome-wide comparative analysis of *copia* retrotransposons in Triticeae, rice, and *Arabidopsis* reveals conserved ancient evolutionary lineages and distinct dynamics of individual *copia* families. *Genome Research* **17**, 1072–1081. doi:10.1101/gr.6214107

Wisman E, Hartmann U, Sagasser M, Baumann E, Palme K, Hahlbrock K, Saedler H, Weisshaar B (1998) Knock-out mutants from an *En*-1 mutagenized *Arabidopsis thaliana* population generate phenylpropanoid biosynthesis phenotypes. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 12432–12437. doi:10.1073/pnas.95.21.12432

Witte CP, Le QH, Bureau T, Kumar A (2001) Terminal-repeat retrotransposons in miniature (TRIM) are involved in restructuring plant genomes. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 13778–13783. doi:10.1073/pnas.241341898

Woodrow P, Pontecorvo G, Fantaccione S, Fuggi A, Kafantaris I, Parisi D, Carillo P (2010) Polymorphism of a new *Ty1-copia* retrotransposon in durum wheat under salt and light stresses. *Theoretical and Applied Genetics* **121**, 311–322. doi:10.1007/s00122-010-1311-z

Woodrow P, Pontecorvo G, Ciarmiello LF (2012) Isolation of *Ty1-copia* retrotransposon in myrtle genome and development of S-SAP molecular marker. *Molecular Biology Reports* **39**, 3409–3418. doi:10.1007/s11033-011-1112-8

Wu L, Gingery M, Abebe M, Arambula D, Czornyj E, Handa S, Khan H, Liu M, Pohlschroder M, Shaw KL, Du A, Guo H, Ghosh P, Miller JF, Zimmerly S (2018) Diversity-generating retroelements: natural variation, classification and evolution inferred from a large-scale

genomic survey. *Nucleic Acids Research* **46**, 11–24. doi:10.1093/nar/gkx1150

Xia C, Zhang L, Zou C, Gu Y, Duan J, Zhao G, Wu J, Liu Y, Fang X, Gao L, Jiao Y, Sun J, Pan Y, Liu X, Jia J, Kong X (2017) A TRIM insertion in the promoter of *Ms2* causes male sterility in wheat. *Nature Communications* **8**, 15407. doi:10.1038/ncomms15407

Yu C, Han F, Zhang J, Birchler J, Peterson T (2012) A transgenic system for generation of transposon *Ac/Ds*-induced chromosome rearrangements in rice. *Theoretical and Applied Genetics* **125**, 1449–1462. doi:10.1007/s00122-012-1925-4

Zakrzewski F, Schmidt M, Van Lijsebettens M, Schmidt T (2017) DNA methylation of retrotransposons, DNA transposons and genes in sugar beet (*Beta vulgaris* L.). *The Plant Journal* **90**, 1156–1175. doi:10.1111/tpj.13526

Zhou M, Liang L, Hanninen H (2018) A transposition-active *Phyllostachys edulis* long terminal repeat (LTR) retrotransposon. *Journal of Plant Research* **131**, 203–210. doi:10.1007/s10265-017-0983-8