Transposon Activity in Plant Genomes

Chapter · August 2016

DOI: 10.1007/978-3-319-31703-8_4

4 authors:

Nermin Gozukirmizi
Istanbul University
127 PUBLICATIONS **1,090** CITATIONS
SEE PROFILE

Aslihan Temel
Istanbul University
25 PUBLICATIONS **115** CITATIONS
SEE PROFILE

Sevgi Marakli
Amasya University
17 PUBLICATIONS **31** CITATIONS
SEE PROFILE

Sibel Yilmaz
Yeni Yüzyil Üniversitesi
19 PUBLICATIONS **59** CITATIONS
SEE PROFILE

Transposon Activity in Plant Genomes

Nermin Gozukirmizi, Aslihan Temel, Sevgi Marakli, and Sibel Yilmaz

Contents

N. Gozukirmizi (✉) • A. Temel • S. Marakli • S. Yilmaz
Department of Molecular Biology and Genetics, Faculty of Science, Istanbul University, 34134, Vezneciler, Istanbul, Turkey
e-mail: nermin@istanbul.edu.tr

K.R. Hakeem et al. (eds.), *Plant Omics: Trends and Applications*,
DOI 10.1007/978-3-319-31703-8_4

Abstract Transposable elements (TEs) were first discovered in maize plants. However, they exist in all plant species investigated so far. Although plants with small genomes have smaller transposon percentages, plants with large genomes have high transposon percentages. For example, *Arabidopsis thaliana* has a genome size of 125 Mb, which comprises 14 % transposons, and the *Hordeum vulgare* genome (5300 Mb) has 80 %. TEs are classified into two major groups based on their transposition mechanism. Class I elements are characterized by DNA sequences with homology to reverse transcriptase, and they are often referred to as retroelements, retrotransposons, or retrovirus-like elements. Retrotransposons function by a copy-and-paste transposition mechanism. Class II TEs (DNA transposons) move by a cut-and-paste mechanism. TEs affect the genome dynamics of plants by regulation of gene expression and chromosomal mutations (such as duplications, insertions/deletions, and structural variations). Transposition rates among generations are about 10^{-3} to 10^{-4}, which is a higher rate than spontaneous mutations. All TEs in a cell are named as transposomes, and transposomics is a new area to work with transposomes. Although some bioinformatics software has recently been developed for the annotation of TEs in sequenced genomes, there are very few computational tools strictly dedicated to the identification of active TEs using genome-wide approaches. In this review article, after a brief introduction and review of the transposable elements, we discuss the effects of TEs in plant gene expression and evolution, and also present our recent research data on barley retrotransposons.

Keywords Mobile elements • Plant genome dynamics • Transposomic • Plant evolution

1 Introduction

Transposon, a segment of DNA that moves to a new location in a chromosome, or to another chromosome or cell, and alters the existing genetic structure, sometimes causes significant changes. Transposons were first described by Barbara McClintock, a maize cytogeneticist who was rewarded with the Nobel prize (1984) 30 years later than her exploration of the relationship between chromosome breaks and maize grain color alterations. Today, we are aware that gene and genome dynamics caused by transposons exist in somatic tissues not only of plants but also of almost all living organisms. Different terms, such as jumping genes, mobile genetics elements, controlling elements, and transposable elements, are used in synonymous ways. Today, we have gained incredible knowledge about the structure, types, and life cycles of these genomic sequences; however, their origin, functions, roles in gene expression, and evolutionary processes still need to be investigated. Developments in DNA sequencing, combined with advances in functional genomics and bioinformatics, have affected studies with transposons because they concern the genomes of living organisms, and the majority of these transposable elements (TEs) are either defective,

fossilized copies, or potentially active copies that are restrained by host silencing systems. However, active transposition evidenced by instances of mutagenic (yet potentially evolutionarily significant) insertions has been demonstrated. For example, TEs have been shown to silence or alter expression of genes adjacent to insertion sites, contribute to chromosomal rearrangements via recombination, epigenetically alter regional methylation patterns, and provide template sequences for RNA interference (Feschotte et al. 2002; Bennetzen 2005; Morgante et al. 2007; Weil and Martienssen 2008; Slotkin et al. 2012; Zhao et al. 2016). These diverse functional impacts of TEs, and their intrinsic contribution to genomic plasticity, suggest that these elements have a major function in molecular diversification and, ultimately, species divergence. In this chapter, we provide the reader with the fundamentals of TE biology, with an emphasis on plant elements. We begin with an overview of TE classification and transposition mechanisms, followed by an examination of the extensive variability in both inter- and intraspecific TE content across diverse plant taxa, and transposon markers, transposomics, and their effects on gene expression and evolutionary processes.

2 Structure of Transposons

Transposons are classified into two classes: class I and class II (retrotransposons and DNA transposons, respectively). Before we discuss the types and classification of transposons, a general explanation of their structures will be helpful to understand the following parts of this review. Transposons use many different enzymes for their transposition. Although some transposons can encode these enzymes (autonomous), others cannot (nonautonomous) and use enzymes of autonomous transposons.

Retrotransposons have a more complex structure than DNA transposons. Group-specific antigen (GAG) is the first protein-coding region of a retrotransposon (Fig. 1). It encodes proteins that pack retrotransposon mRNA in the cytoplasm, and this structure is named virus-like particle (VLP). Reverse transcription of retrotransposon mRNA occurs in VLPs (Jaaskelainen et al. 1999). Although it usually has the same open reading frame (ORF) as other domains, it is transcribed from a different ORF in some retrotransposons (Ohtsubo et al. 1999).

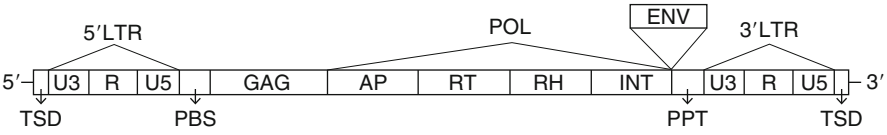

Fig. 1 Schematic demonstration of a retrotransposon having LTR regions. *LTR* long terminal repeat, *R* repeated region, *U3* unique for 3′-end of RNA, *U5* unique for 5′-end of RNA, *PBS* primer binding site, *GAG* group-specific antigen, *POL* polyprotein, *AP* aspartic peptidase, *RT* reverse transcriptase, *RH* ribonuclease H, *INT* integrase, *ENV* envelope, *PPT* polypurine tract, *TSD*, target site duplication

The other coding region of retrotransposons is the polyprotein (POL), which region encodes four enzymes: aspartic peptidase (AP), reverse transcriptase (RT), RNase H (RH), and integrase (INT). These enzymes can act either together as part of a protein or individually. AP is a proteolytic enzyme that is responsible for maturation of proteins required for the retrotransposon life cycle. RT is an RNA-dependent DNA polymerase that is necessary for reverse transcription of retrotransposon mRNA. RH is a hydrolytic enzyme that is responsible for the hydrolysis of the original RNA template that is part of the RNA/DNA hybrid generated after reverse transcription. INT catalyzes the insertion of new retrotransposon copies into the genome (Flavell 1995; Malik and Eickbush 2001). Some retrotransposons have a tyrosine recombinase (YR) domain instead of INT in the POL region. YR is typically involved in site-specific recombinations between similar or identical DNA sequences (Coates et al. 2005). Similar to YR, apurinic endonuclease (*APE*) is also encoded instead of INT in certain types of retrotransposons and provides site-specific insertion (Yadav et al. 2009).

In addition to these domains, some retrotransposons can have extra domains that give retrotransposons several properties. The envelope (ENV) domain is one of these domains and was first defined in viruses. This domain encodes a protein that is responsible for cell-to-cell transfer of virus. Most of the retrotransposons lost their ENV domains during the evolutionary process. However, the domain is still present in some retrotransposons (Wright and Voytas 1998).

Retrotransposons also have some sequences that do not encode a protein product but are essential for the life cycle of the retrotransposon. Long terminal repeats (LTRs) sequences are direct repeats that are present at two borders (the 5′- and 3′-ends) of some retrotransposons. They vary within the retrotransposon families. However, the LTR sequence of a retrotransposon is well conserved between plant species during the evolutionary process. LTR sequences do not encode a protein product but have important functions for transposition. An LTR sequence has three regions: Unique for the 3′-end of RNA, Repeated region, and Unique for the 5′-end of RNA (U3, R, and U5, respectively). In the U3 region, the promoter and some *cis*-acting elements are present (Pouteau et al. 1994; Takeda et al. 1998). The R region functions during reverse transcription of retrotransposon mRNA, and the U5 region constitutes the first portion of the retrotranscribed genome (SanMiguel et al. 1998; Jiang et al. 2002).

In addition to LTR sequences, retrotransposons have three noncoding regions: primer-binding site (PBS), polypurine tract (PPT), and target site duplication (TSD). PBS is a region about 18 nucleotides in length located between the 5′-LTR and GAG domains. During reverse transcription of retrotransposon mRNA, the 3′-end of a cellular tRNA binds to this region and acts as a primer (Wilhelm and Wilhelm 2001). PBS sequences are used for construction of a sequencing library to identify LTR sequences and their insertion sites (Monden et al. 2014). PPT is located between 3′-LTR and the internal domains: it contains about ten purine (A-G) bases and is involved in the synthesis of double-stranded cDNA from single-stranded (Mak and Kleiman 1997; Gabus et al. 1998). TSD is a short direct repeat typically four to eight nucleotides in length and that is present at the two borders. TSD is characteristic of both retrotransposons and DNA transposons.

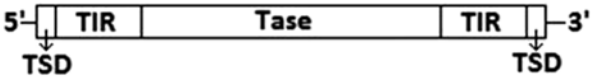

Fig. 2 Schematic demonstration of a DNA transposon. *TIR* terminal inverted repeat, *Tase* transposase, *TSD* target site duplication

Despite the complex structure of retrotransposons, DNA transposons have a more simple structure (Fig. 2). DNA transposons encode a transposase (Tase) enzyme. This enzyme cuts the DNA transposon and integrates it to a new location.

DNA transposons also contain noncoding repeat regions, such as LTR, at both borders, but these repeats are present in inverted orientation (Poulter and Goodwin 2005). Thus, these repeats are called terminal inverted repeats (TIRs). TIRs are essential for the transposition of most transposons. Generally, this is a characteristic feature of DNA transposons. However, some retrotransposon groups also have these repeat sequences. Generally, two functional regions are present in TIR sequences. Although the first region is involved in DNA cleavage and strand transfer reactions, the second one is required for specific recognition and binding (Szabo et al. 2010).

3 Types and Classification of Transposons

After the great discovery of transposons by Barbara McClintock, numerous studies in many areas have been performed regarding transposons. It was found that transposons are present as billions of copies in almost every organism investigated so far. In these studies, transposons with different features were also defined. Therefore, a classification system was needed, and the first classification was proposed by Finnegan (1989). Finnegan divided transposons into two classes, class I and class II, based on their intermediate of transposition. In the following years, this simple classification system has become inadequate. Some hierarchical classifications in detail were proposed by the International Committee on the Taxonomy of Viruses (ICTV) and Capy et al. (1998) for certain types of transposons. In this review, we describe a different type of transposons based on a unified classification system proposed by Wicker et al. (2007). In this hierarchical system, transposons are classified in five levels as class, subclass, order, superfamily, and family. In the first level, class, transposons are divided based on the presence or absence of an RNA transposition intermediate, as previously proposed by Finnegan (1989). The second level is the subclass, and transposons are separated according to their transposition mechanisms, either copy-paste or cut-paste. Because all class I transposons (retrotransposons) use copy-paste transposition, only class II transposons (DNA transposons) have different subclasses. Order, the third level, bases major differences in the insertion mechanisms and enzymology between transposons that belong to the same subclass. In the superfamily, the fourth level, transposons are grouped based on structural differences of their proteins or noncoding regions. The fifth level, family, is based on DNA sequence similarities.

3.1 Class I Transposons (Retrotransposons)

Retrotransposons are mobile genetic elements that transpose an RNA intermediate and have an important function in genome evolution. Because of their copy-paste transposition mechanism, they cause genome enlargement. They are present at high copy numbers in most plants, constituting more than 50 % of the nuclear genome in many cases. Retrotransposons resemble retroviruses because of both their structure and replication mechanism. Because all retrotransposons have copy-paste mechanisms, they do not have different subclasses. All belong to the same subclass, which is divided into five orders.

3.1.1 Order 1: LTR Retrotransposons

LTR retrotransposons are the largest group in all transposons. They are flanked by variable-size long terminal repeats (LTRs) in both borders. The LTR retrotransposon order is divided into five superfamilies: *Copia, Gypsy, Bel/Pao, Retrovirus*, and *Endogenous Retrovirus (ERV)*.

Copia is the first superfamily and shows identical function and similar genome organization with other superfamilies of LTR retrotransposons. However, it differs from others in the position of the INT domain that located between the AP and RT domains. It is widespread in almost all organisms: plants, animals, fungi, etc. In addition to the classification proposed by Wicker et al. (2007), the International Committee on the Taxonomy of Viruses (ICTV) divided it into three genera: *Pseudovirus, Hemivirus*, and *Sirevirus*. In these genera, only *Sirevirus* have ENV domain in this superfamily.

The second superfamily is *Gypsy* and, similar to *Copia*, it is one of the most abundant groups in living organisms. Its INT domain is located between the RH domain and the 3′-LTR. Some retrotransposons that belong in this superfamily also have an ENV domain and resemble retroviruses. Detailed classification of the *Gypsy* superfamily is complicated. Previously, it was divided into two genera, as *Metaviridae* and *Errantiviridae*, but lately is divided as *Chromavirus* and *non-Chromavirus*.

The last three superfamilies of LTR retrotransposons—*Bel/Pao, Retrovirus*, and *Endogenous Retrovirus (ERV)*—have the same organization with internal domains as in *Gypsy*. They exist only in Metazoan genomes (for more details of these groups, see Eickbush and Jamburuthugoda 2008; Llorens et al. 2009).

3.1.2 Order 2: DIRS

This order is characterized by a tyrosine recombinase (YR) that is typically involved in site-specific recombinations between similar or identical DNA sequences, instead of INT in the POL region. Therefore, this order is also called tyrosine recombinase

retroelements. The DIRS order is divided into three superfamilies: DIRS, Ngaro, and VIPER (Goodwin et al. 2004; Lorenzi et al. 2006). Analyses of the reverse transcriptase sequences of the first superfamily of DIRS suggested that it was distantly related to the *Gypsy* superfamily. However, DIRS elements differ from *Gypsy* by their terminal repeats in addition to the YR domain: DIRS elements are bordered by TIRs whereas *Gypsy* has LTRs (Poulter and Goodwin 2005). The DIRS superfamily contains internal ORFs and an internal complementary region (ICR) derived from the duplication of TIR sequences (Piednoel et al. 2011). Only the DIRS superfamily of the DIRS order is present in plants. The other two superfamilies of the DIRS order, Ngaro and VIPER, have not been found in plants to date. Although orientation of their internal domains is the same with DIRS, their terminal repeats are present as a direct position as in LTR retrotransposons.

3.1.3 Order 3: PENELOPE (PLE)

This order contains only one superfamily, named Penelope. Up to now this order has been described in many animals, plants, protists, and fungi. Retrotransposons that belong this group also contain LTRs in both borders. However, these LTR sequences might be either direct or inverted orientations. This order has a coding region that contains RT and endonuclease (EN) domains. The RT of PLE is related to the RT of telomerase, but EN is similar to an intron-encoded endonuclease (Pyatkov et al. 2004).

3.1.4 Order 4: LINE

This order is also known as non-LTR retrotransposons because elements of this order have neither LTR nor TIR sequences at borders. They are widely distributed in eukaryotic genomes. They usually have a poly-A tail at the 3′-end and some deletions at the 5′-end. This order is divided into five superfamilies: R2, RTE, Jockey, L1, and I (Wicker et al. 2007). The first three superfamilies, R2, RTE, and Jockey, are unique for metazoans whereas the last two, L1 and I, are found in various organisms including plants. Because of their capability to encode the enzymes required for transposition, they are also accepted as autonomous non-LTR retrotransposons (Eickbush and Jamburuthugoda 2008). Although R2 encode RT and EN enzymes, others encode APE and RT. Only the last superfamily, I, has an RH domain together with APE and RT.

3.1.5 Order 5: SINE

The last order of class I transposons, is also known as non-LTR retrotransposons, similar to LINE. This order lacks both repeat regions, TIR and LTR, at the borders and has a poly-A tail. In contrast to the LINE order, SINE does not have internal domains. Because SINE does not encode their own reverse transcriptase, it has been

proposed that SINE uses the enzymatic machinery of LINE for transposition (Wallace et al. 2008; Kroutter et al. 2009). Therefore, elements belonging to the SINE order are accepted as nonautonomous non-LTR retrotransposons. This order shows similarity with reverse-transcribed RNAs transcribed by RNA polymerase III into tRNA, rRNA, and other small nuclear RNAs (Malik and Eickbush 1998). The SINE order is divided into three superfamilies, tRNA, 7SL, and 5S. The first two superfamilies are found in plantsbut the last one is not.

3.2 Class II Transposons (DNA Transposons)

DNA transposons have a different transposition mechanism than retrotransposons. They move through a DNA intermediate, in either the cut-paste or copy-paste mechanism. For this reason, DNA transposons are divided into two subclasses. The first subclass uses the cut-paste mechanism and the second uses copy-paste. Further, both these subclasses have two orders.

3.2.1 Subclass I Order I: TIR

This order shows characteristic structures of DNA transposons, a Tase enzyme that is flanked by two TIR sequences. During transposition, both strands of these elements are cut and inserted into new locations so that their copy numbers remain constant. However, if they transpose during DNA replication (S-phase of the cell cycle), the copy number can increase. The TIR order is divided into nine superfamilies: Tc1/mariner, hAT, Mutator, P, PIF/Harbinger, CACTA, Merlin, Transib, and PiggyBac. The first six superfamilies are present in plants but the last three are not (Wicker et al. 2007).

3.2.2 Subclass I Order II: Crypton

The order Crypton also uses the cut-paste transposition mechanisms as does the TIR order, but in contrast the Crypton order has neither TIR sequences nor Tase domain. Instead of Tase, they have an YR domain. It only has one superfamily, Crypton, that has been found only in fungi so far.

3.2.3 Subclass II Order I: Helitron

Although helitrons are DNA transposons, they use the copy-paste transposition mechanism as do retrotransposons. However, their transpositions do not include an RNA intermediate. This DNA-intermediated transposition mechanism is achieved by cleavage of just one strand of a DNA transposon. Then, a complementary strand

of released single-strand DNA is synthesized and inserted to a new location. The Helitron order has just one superfamily, Helitron. This is the only replicative DNA transposon type defined in plants so far.

3.2.4 Subclass II Order II: Maverick

The last order of transposons is Maverick, which also uses the DNA-mediated replicative form of transposition. This order has only one superfamily, which is not present in plants. Differing from Helitron, DNA transposons belonging to the Maverick superfamily have TIRs at both borders. These transposons can have 11 domains that encode proteins needed for transposition, but the number and orientation of these domains vary from one transposon to another.

3.3 Transposons in Different Plant Species

After their discovery and characterization in maize, transposons have been found in all organisms investigated so far, with just one exception: in some species of *Plasmodium*, transposons were not found (Vitte and Bennetzen 2006; Huang et al. 2012). Percentages and types of transposons in the genomes can vary among species (Feschotte and Pritham 2007). Although transposons compose 1–3 % of prokaryotic genomes, in eukaryotic organisms, such as plants, their percentage may reach 85 % or more. In animals, transposons constitute 3–45 % of the genome. The human genome contains about 45 % transposons, and the most abundant type is non-LTR retrotransposons (7–33 %). However, LTR retrotransposons are the most abundant transposon type in the genome of many organisms but they constitute just 8.5 % of the human genome (Cordaux and Batzer 2009). Generally plants, especially cereals, have the highest percentage of transposons compared with others, such as prokaryotes, animals, and mammals.

4 Transposon Markers

A molecular marker is defined as a particular segment of DNA that is representative of the differences at the genome level. An ideal molecular marker should be polymorphic and evenly distributed throughout the genome, generate multiple, independent, and reliable markers, be simple, quick, and inexpensive, need only small amounts of DNA samples, etc. (Agarwal et al. 2008). Many features of LTR retrotransposons in plant genomes have made them excellent sources of molecular markers (Kalendar and Schulman 2006; Poczai et al. 2013; Yuzbasioglu et al. 2016a). A schematic demonstration of some LTR retrotransposon-based marker techniques is given in Fig. 3. Retrotransposon sequences alone or combined with

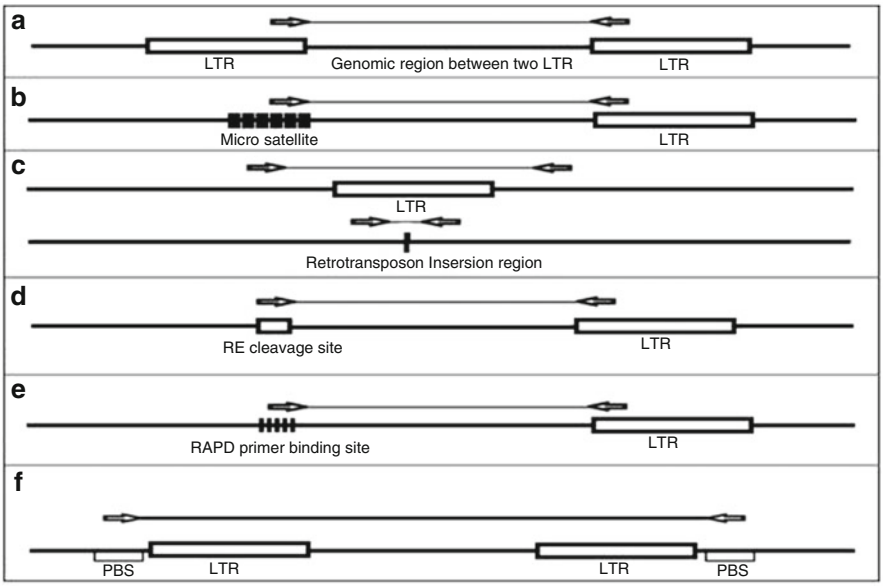

Fig. 3 LTR retrotransposon-based marker techniques: (**a**) IRAP, (**b**) REMAP, (**c**) RBIP, (**d**) SSAP, (**e**) RRAP, (**f**) IPBS (Gozukirmizi et al. 2015)

various sequences in the genome provide primer-binding sites. LTRs of a retrotransposon have conserved sequences between different organisms. Thus, a primer pair designed for a specific organism can be used for other organisms. However, different retrotransposons of an organism have different LTR sequences so a primer pair designed for a specific retrotransposon cannot bind to another LTR. In the following section, some of the transposon-based marker techniques are explained and compared, and their advantages and disadvantages are discussed.

4.1 Inter-Retrotransposon Amplified Polymorphism (IRAP)

IRAP is a multiplex and dominant technique. This technique was first developed using the *BARE-1* ('*BArley RetroElement 1*') sequence (Kalendar et al. 1999). The *BARE-1* family has approximately 1.5×10^4 full-length copies and 1.7×10^5 solo LTRs (Vicient et al. 1999). IRAP uses PCR primers designed in an outward direction from the conserved sequences of LTR. Internal regions between two LTRs or solo LTRs (without retrotransposon) are amplified. Band sizes of amplification products support the local clustering yet large-scale dispersion of *BARE-1* elements (Kalendar et al. 1999). Retrotransposons tend to be nested into another element in cereal genomes (Vicient et al. 2001). Targets (5′- or 3′-end of LTR) of primers produce different amplicons and alter banding patterns. The appearance of a band

reflects a new insertion that is in an amplifiable distance of another element. In this technique, DNA samples are amplified with a single or two primers. The major advantages of this technique are that restriction digestion or ligation steps do not exist and the polymerase chain reaction (PCR) products are separated on agarose gels. Because of the high band numbers, DNA amounts should be optimized. High template DNA concentrations may cause overamplification (Kalendar et al. 1999; Kalendar and Schulman 2006). Slow electrophoresis with high-quality agaroses improves the resolution of the bands. Impurities in DNA samples may interfere with PCR (Schulman et al. 2004).

4.2 Retrotransposon-Microsatellite-Amplified Polymorphism (REMAP)

This method is also a dominant and multiplex marker system. REMAP resembles IRAP but uses one LTR primer and a primer specific to a nearby microsatellite. Therefore, not only regions between two LTRs but also regions between an LTR and a microsatellite are amplified (Kalendar et al. 1999). Microsatellites, also known as simple sequence repeats (SSRs), are a type of repetitive sequences in plants and animals. These motifs can consist of a single base pair or a small number of bases (usually ranging from 1 to 6) that are repeated several times (Saha et al. 2003). SSRs are abundant, highly polymorphic, and dispersed throughout the genome (Li et al. 2002). SSRs also provide an independent molecular marker technique (Zietkiewicz et al. 1989). Moreover, SSRs are associated with retrotransposons in barley (Davila et al. 1999). Microsatellite primers in the REMAP method are different from original SSR primers that anneal to flanking regions. In other words, a LTR primer is combined with a primer consisting of SSR itself. PCR conditions and analysis of PCR products are similar to IRAP conditions (Kalendar et al. 1999; Kalendar and Schulman 2006). However, some bands may be derived solely from SSR primers (Leigh et al. 2003).

4.3 Retrotransposon-Based Insertional Polymorphism (RBIP)

RBIP was considered as an alternative to the multiplex S-SAP method. Individual transposon insertions are detected by RBIP using host-specific PCR primers and a transposon-specific primer. Two PCR primers are designed from the target (host) where the transposon is inserted and an additional transposon-specific primer was designed. DNA samples are amplified with either host primers or a host primer combined with a transposon primer. Insertion of a transposon alters the size of the locus. Polymorphism can be determined by agarose gel electrophoresis or dot-blot analysis. Hence, allelic states of a particular locus are determined. RBIP is locus

specific and codominant as are SSRs. The most important limitation to both RBIP and SSR markers is the need for sequence information from the chromosomal region surrounding each marker. Although SSR markers are obtained using by screening genomic libraries with simple sequence repeats, RBIP markers can be identified using the S-SAP approach (Flavell et al. 1998). DNA quality is not an issue (Schulman et al. 2004).

A major advantage of RBIP is that it can easily be automated, using gel-free procedures such as TaqMan or DNA chip technology to increase sample throughput.

4.4 Sequence-Specific Amplified Polymorphism (S-SAP)

S-SAP is a dominant and multiplex marker system. It is the first retrotransposon-based molecular marker and was modified from amplified fragment length polymorphism (AFLP) (Waugh et al. 1997). The AFLP technique is based on the selective PCR amplification of restriction fragments from a total digest of genomic DNA (Vos et al. 1993). In S-SAP, genomic DNA is first digested with two restriction endonucleases and ligated with adaptors specific to restriction endonucleases. The template is subjected to preamplification with primers from adaptors and then selectively amplified with two primers, one designed from a conserved region of a LTR and the other a selective adaptor primer. Selective adaptor primers have additional nucleotides (one to three) at the 3′-end. PCR products are denatured with gel loading buffer containing formamide and then resolved on polyacrylamide gel electrophoresis (PAGE). Although S-SAP generally yielded fewer PCR bands than AFLP, the polymorphism percentage was higher in S-SAP (Waugh et al. 1997). S-SAP is very similar to the transposon display (TD) method (van den Broeck et al. 1998).

4.5 RAPD Retrotransposon-Amplified Polymorphism (R-RAP)

R-RAP was proposed as a combination of RAPD (random amplified polymorphic DNA) and IRAP techniques, which may have some limitations. In the case of RAPD, low reproducibility may be a problem and in the case of IRAP, primer-binding sites may be too far apart to produce an amplification product. DNA samples are amplified using one random 10-mer primer and a LTR primer. Melting temperatures of RAPD and IRAP differ greatly; therefore, identification of optimal annealing temperature by gradient PCR is necessary. Amplification products are resolved on agarose gels. The number of R-RAP loci varies from 10 to 17 with high resolution and reproducibility (Aalami et al. 2012).

5 Sequencing and Transposomics

Transposon-insertion sequencing is a powerful technique for the rapid connection of genotype to phenotype, particularly in bacteria. A number of studies have improved our understanding of basic gene functions, establishing requirements for colonization and infection, mapping complex metabolic pathways, and exploring noncoding genomic regions (Barquist et al. 2013). Several methods were developed concurrently for high-throughput sequencing of transposon-insertion sites: transposon-directed insertion site sequencing (TraDIS) (Langridge et al. 2009), insertion sequencing (INSeq) (Goodman et al. 2009), high-throughput insertion tracking by deep sequencing (HITS) (Gawronski et al. 2009), and ransposon sequencing (Tn-seq) (van Opijnen et al. 2009), followed by Tn-seq circle (Gallagher et al. 2011). All these protocols have the same basic workflow with minor variations (Fig. 4) (Barquist et al. 2013).

The Tn-seq and INSeq methods are highly similar but INSeq includes a PAGE purification step following adaptor ligation and PCR whereas Tn-Seq includes an agarose gel purification at this point. Goodman et al. (2011) introduced additional steps to the original INSeq protocol: a linear PCR step using a biotinylated primer and subsequent purification of the product with magnetic streptavidin beads were added following adapter ligation. These steps reduce both the amount of sample and

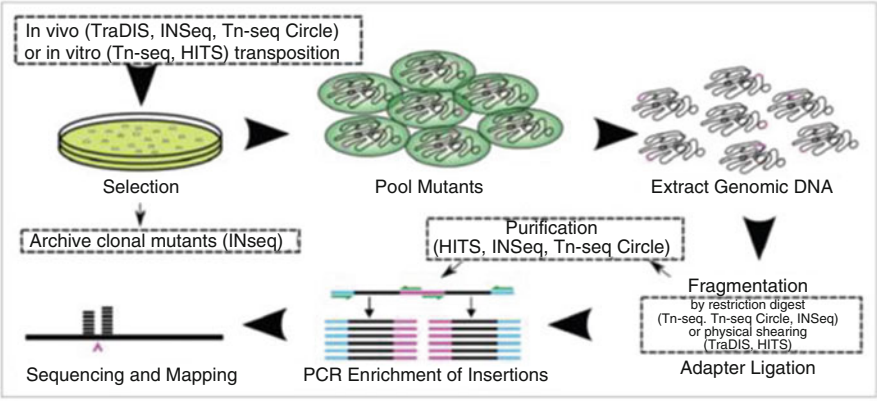

Fig. 4 An illustration of the typical workflow of transposon-insertion sequencing protocols. Transposons are represented by *pink lines*, sequencing adaptors by *blue*, genomic DNA by *black*, and PCR primers by *green*. Mutants are generated through either in vivo or in vitro transposition and subsequent selection for antibiotic resistance. These mutants are pooled, and genomic DNA is extracted and fragmented by restriction digest or physical shearing. Sequencing adaptors are ligated; some protocols then perform a step to purify fragments containing transposon insertions. Adaptor-specific primers are used to specifically enrich for transposon-containing fragments. The fragments are then sequenced and mapped back to a reference genome to uniquely identify insertion sites. *Dashed boxes* indicate steps that differ between protocols. (From Barquist et al. 2013)

the amount of enzymes needed. Although they make the protocol more laborious, the results suggest that these modifications increase the sensitivity of the technique. Moreover, HITS and TraDIS methods are more similar to each other than to Tn-seq and INSeq: after shearing of the DNA, the DNA ends are repaired, and a poly-A tail is added. However, the methods diverge after the PCR step; in HITS, the PCR products undergo size selection (on a gel) and affinity purification before sequencing, whereas in TraDIS, the PCR products are sequenced directly (van Opijnen and Camilli 2013).

Tn-seq was applied to the pathogen *Streptococcus pneumoniae* to identify genes that are essential for growth in rich medium, to determine the fitness effect (advantageous or disadvantageous) of each nonessential gene and to identify 97 genetic interactions between five query genes and the rest of the genome (van Opijnen et al. 2009). Moreover, Tn-seq has been used in *Pseudomonas aeruginosa* to identify antibiotic resistance genes (Gallagher et al. 2011) and *Mycobacterium tuberculosis* (Griffin et al. 2011) to identify essential genes and pathways that are involved in the utilization of cholesterol. A slight variation on the Tn-seq method, the so-called Tnseq circle, was used to track large numbers of transposon mutants of *Pseudomonas aeruginosa* to study its inherent resistance to the aminoglycoside antibiotics (Gallagher et al. 2011). The INSeq method was used to determine whether the human symbiont *Bacteroides thetaiotaomicron* harbors specific genes that are necessary for survival in the colon (Goodman et al. 2009). The HITS method was applied to a library of approximately 75,000 *Haemophilus influenzae* mutants (Gawronski et al. 2009). TraDIS was used to screen for essential genes in *Salmonella typhimurium*, the causative agent of typhoid fever (Langridge et al. 2009).

Transposon sequencing is not restricted to screening for the fitness effects of gene knockouts. In one study (Christen et al. 2011), this technology was used to construct a dense library of *Caulobacter crescentus* insertion mutants (with 8-bp resolution). In another study, by combining RNA-seq with Tn-seq, 56 new putative sRNAs in the noncoding regions of *S. pneumoniae* (Mann et al. 2012) were identified. Transposon sequencing has the potential to be used for other purposes. For example, probing for differences in genotype–phenotype interactions between strains or closely related species has the potential to reveal divergent evolutionary trajectories for shared genes or pathways, such as functional pathways that have undergone reconfiguration either in terms of their regulation or to accomplish new tasks (van Opijnen and Camilli 2013). Although it has been applied almost exclusively to bacterial species, transposon sequencing could be expanded to study other microorganisms. For example, it could be used to probe gene function and gene networks in viruses. All that is needed is a means of introducing transposon insertions into the viral genome; this could be accomplished by building an inducible transposition system within the host, such that insertions are introduced into the viral genome during viral replication (van Opijnen and Camilli 2013).

In vitro and in vivo transposition have been linked by the use of Transposomes™ (Goryshin et al. 2000). A transposome is the complex of transposase and transposon that is created in the early steps of the IVT (in vitro transposition) when the transposase attaches to the transposon for excision. At that point there is no recipient

DNA or cations to activate transposition. The transposome is introduced into a host by transformation and, because of the internal cation concentration, transposition into the host genome follows. This technology is dependent on a host transformation system and overcomes the host barrier for in vivo transposition and the need for homologous recombination (Hamer et al. 2001). In conclusion, the application of transposon sequencing give highly valuable information for discovering gene function, and further technological advances are likely to see it continue to be applied as a useful, high-throughput screening method for some time to come (van Opijnen and Camilli 2013).

6 Transposons and Gene Expression

Heterochromatin is commonly regarded as silent DNA. It consists of large regions of repetitive nucleotide sequences and transposons. Transposons, however, must be suppressed because they constitute two dangers for the genome: (1) their repeated units can cause spurious homologous recombination; and (2) their ability to transpose can lead to disruption or misregulation of important genes. Both these dangers are suppressed by heterochromatinization (Madlung and Comai 2004). Insertion of a transposable element into a gene causes mutation. A proper instance comes from human genetics: for example, retrotransposition of L1 element (LINE) into factor VIII gene causes hemophilia (Kazazian and Moran 1998). In the case of plants, transposon-like insertion into the starch-branching enzyme interrupts gene expression and causes wrinkled seeds in maize (Bhattacharyya et al. 1990). In addition, tases can fulfill important functions for the host as in the case of *Arabidopsis thaliana*. A tase, named DAYSLEEPER, binds to a specific motif in the upstream region of a repair gene. Plants lacking DAYSLEEPER exhibit developmental abnormalities (Bundock and Hooykaas 2005). Transposable elements produce small RNAs that regulate expression of specific genes in *Drosophila* and *Arabidopsis* (reviewed by McCue and Slotkin 2012). Insertion of a transposable element, *Mu*, generates a recessive mutation referred as *hcf106* that causes pale green seedlings and is lethal in maize. However, when the *Mu* element is inactivated by hypermethylation, plants exhibit a normal phenotype despite the continued presence of the transposon within the gene (Martienssen et al. 1990). Shortly after this publication, the same group reported that the *Mu* element lies within the 5′-untranslated leader of the *Hcf106* mRNA and, when active, this insertion interferes with the accumulation of *hcf106* mRNA (Barkan and Martienssen 1991). *Flowering locus C* (*FLC*) encodes a MADS domain protein and is a repressor of flowering in *Arabidopsis* (Michaels and Amasino 1999). The Landsberg *erecta* (Ler) ecotype of *Arabidopsis* has a Mutator-like (MULE) transposable element insertion in the first intron of FLC gene (Gazzani et al. 2003; Michaels et al. 2003), and this insertion causes FLC-Ler to be expressed at low levels (Michaels et al. 2003). It was proposed that microRNA (miRNA) genes have been evolved from miniature inverted-repeat transposable elements (MITEs). *hsa-mir-548*, a family of human miRNA genes, were found to be derived from

Made1 elements, which are members of MITEs (Piriyapongsa and Jordan 2007). miR441 and miR446, miRNA sequences of rice, were reported to have originated from *Stowaway1* (Li et al. 2011). TE-derived siRNA inhibits the formation of a host protein that represses TE activity (McCue et al. 2013).

7 Transposons and Plant Evolution

7.1 Genome Size

TEs have found to be important in genome structure and genetic diversity. For example, in some plant species such as maize, *Gossypium* spp., and *Oryza australiensis* (a wild relative of rice), TE amplifications have at least doubled the genome size within the last 5 million years (SanMiguel et al. 1998; Piegu et al. 2006; Hawkins et al. 2006). TE content strongly correlates with genome size variation. TE-derived DNA mostly makes up the 20–30 % of the genome even for plant species with small genomes such as *Brachypodium distachyon* and *Arabidopsis* spp. (The *Arabidopsis* Genome Initiative 2000; The International *Brachypodium* Initiative 2010). Species with larger genomes, such as maize and barley, have larger TE-derived DNA content, up to 85 % (Wicker et al. 2005; Schnable et al. 2009). Evidence suggests that repetitive DNA may be removed more slowly from species with larger genome sizes than smaller genomes. A study on *Copia* retrotransposons showed that they remained intact and active for much longer time periods in the larger barley and wheat genomes than the smaller *A. thaliana* and rice genomes (Wicker and Keller 2007). The TE proportion of the genomes not only differs between species but also within species. A comparison of a 1.2-Mb region of the *indica* and *japonica* subspecies of rice revealed that 13 % of the sequence was not shared and that the difference was caused by differential TE insertions (Ma and Bennetzen 2004). Again, the important differences in genome size between *A. thaliana* and *A. lyrata* have also shown to be caused by a reduction in TE activity and potentially more efficient TE elimination in *A. thaliana* (Hu et al. 2011). Moreover, two sequenced maize genomes revealed different genome sizes by 22 %, with 90 % of this difference caused by repetitive elements (Wang and Dooner 2006).

7.2 Genome Organization and Maintenance

Most eukaryotic organisms accommodate high numbers of retrotransposons in centromeres and telomeres. Plant centromeres contain infused TEs within short centromeric repeats (Ma et al. 2007). Pericentromeric regions, similarly, are composed mostly of silenced TEs and pseudogenes (Hall et al. 2006). A centromere-specific LTR-retrotransposon has been identified in rice (Cheng et al. 2002) and some

other grasses (Miller et al. 1998). Additionally, in maize, the centromere-specific retroelements have been demonstrated to have an interaction with the kitetochore protein CENH3 (Zhong et al. 2002). On the other hand, the exact role of these sequences in centromere or telomere function is still not clear (Joly-Lopez and Bureau 2014).

Allopolyploidization has been shown to be connected with rapid structural and functional alterations in genomes, especially in the repetitive content (Leitch and Leitch 2008). Broad epigenetic rearrangements are induced by TEs after polyploidy (Parisod et al. 2010). About one third of the identified imprinted genes in *A. thaliana* are located near transposable elements or repeated sequences that encounter demethylation. It could be implemented that transposon insertions near gene regulatory regions will recruit DNA methylation machinery that targets the invading foreign DNA and eventually cause silencing of the affected gene (Jians 2012). In plants, methylated TEs are found further away from genes than are unmethylated TEs (Hollister and Gaut 2009).

7.3 Generation of Variation and Evolutionary Innovation

Transposons have a significant role in constructing eukaryotic genomes by creating interspecies/intraspecies diversity and potential for adaptation to changing environments. Some TEs stimulate genome rearrangements, including inversion, duplication, or deletion of adjacent DNA, by chromosome breaking, aborted transposition, or ectopic recombination between homologous transposable elements at different chromosomal locations (Feschotte and Pritham 2007). The transposons not only contribute to the generation of allelic diversity in natural populations but also to shaping the genomic and epigenetic properties of their hosts and to the divergence of new genes. MULEs have been observed to capture genic sequences in both maize and *Arabidopsis* (Talbert and Chandler 1988; Yu et al. 2000). MULEs have also been shown to be a major feature of rice genome construction (Jiang et al. 2004).

TE insertions disrupt the coding sequence of a gene and inhibit the production of the gene product. However, insertions within promoters, introns, and untranslated regions can directly change the phenotype by genetic or epigenetic regulations (Feschotte and Pritham 2007). There are also transposase-induced rearrangements that have a strong potential for chromosome restructuring. Chromosomal inversions, duplications, and deletions of more than 100 kb can occur depending on the orientation of the TEs at the chromosomal location (Zhang and Peterson 2004). These events can lead to exon shuffling and create new functional genes. Studies on flowering plants demonstrate a high degree of gene creation by transposon capture and exon shuffling. Late-flowering *Arabidopsis* accessions possess an epigenetically silenced TE insertion at the first intron of the *FLC* (flowering locus C) gene (Liu et al. 2004). Wheat, barley, and maize flowering studies also highlighted the TEs on quantitative variation via modulation of gene expression (Yan et al. 2006;

Salvi et al. 2007). Additionally, as class II *Gyno-hAT* transposon causes methylation spreading to the promoter of the *CmWIP1* transcription factor gene, transition from male to female flowers occurs in melon (Martin et al. 2009).

The duplication of the host genes and exon shuffling by TEs may be important in the genome evolution. TEs use three modes: alteration of gene functions through insertion, induction of chromosomal rearrangements, and divergence of new genes and regulatory sequences (Feschotte and Pritham 2007). Transposons assist in the arrangement of the chromosomal domains through distinct epigenetic marks and transcriptional activity. These marks can be changed by genetic stress, such as DNA damage, interspecific hybridization or polyploidization, and environmental cues such as pathogen infection, and abiotic stresses help increase homologous recombination occurrence and chromosomal rearrangements. Such changes can release the transposons from epigenetic control, allowing stress-inducible TEs to propagate stress-inducible promoters to other genes through transposition (Ito et al. 2011) and can influence the genome organization both somatically and heritably (Boyko et al. 2010), offering an opportunity for natural selection to establish new chromosomal domains and regulatory circuits, eventually leading to speciation (Feschotte and Pritham 2007).

Plant resistance genes have been shown to be rapidly evolving. Race-specific resistance genes are often found in clusters forming large tandem repeats of polymorphic genes (Casacuberta and Santiago 2003). Likely, the rice *Xa21* gene family is demonstrated to incorporate a high number of TEs located within different genes (Richter and Ronald 2000). In plants, genes with similar or identical functions can be positioned in very different chromosomal locations in different species. For instance, the *adh1* gene is found in an entirely new chromosome in maize and sorghum when compared to more distant relatives such as rice or wheat, disrupting chromosomal synteny (Ilic et al. 2003).

The term TE exaptation can be referred to molecular domestication where donation of the protein-coding sequences contributes to the evolution of new genes with new functions and to the host phenotype. For instance, different genes captured by the same transposon are transcribed as a single compound mRNA, which can then evolve into new genes encoding novel proteins (Morgante et al. 2005). The host defends its genome against transposons via DNA methylation, small RNAs, cytosine deaminases, and DNA-repair factors (Levin and Moran 2011). In case TEs produce beneficial phenotypes, domesticated genes will be less silenced by the host silencing machineries. TE insertions can generate new promoters, enhancers, insulators, and regulatory regions. Analysis of TE-derived genes proposes that transposase-encoding DNA transposon sequences are the most frequently domesticated parts of TEs (Feschotte and Pritham 2007).

8 Our Work with Transposons

Our group used retrotransposon-based molecular markers mainly for analysis of somaclonal variation. The stability of aging barley calli and callus-regenerated shoots was investigated by IRAP using a primer derived from *BARE-1* (Evrensel et al. 2011;

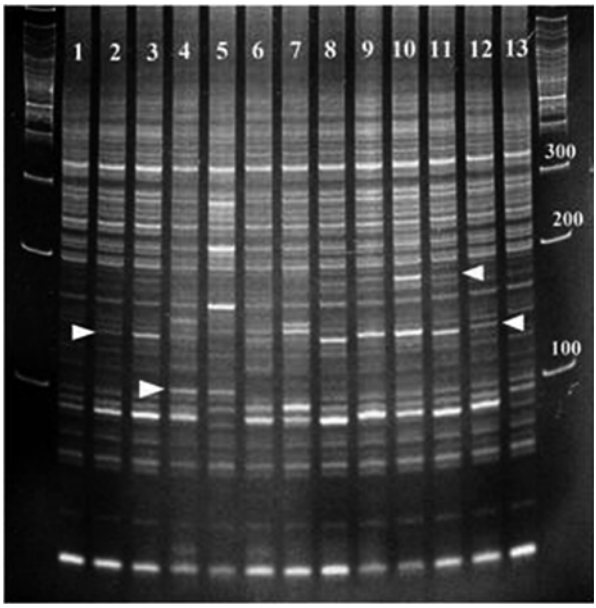

Fig. 5 Inter-retrotransposon amplified polymorphism (IRAP) profiles of calli and shoots for *BAGY2*. *1*, noncultured embryo; *2–13*, tissue culture materials (*2*, *6*, *10* 45-day-old calli; *3*, *7*, *11* shoots regenerated from 45-day-old calli; *4*, *8*, *12* 90-day-old calli; *5*, *9*, *13* shoots regenerated from 90-day-old calli. (From Yilmaz et al. 2014)

Yilmaz and Gozukirmizi 2013) and *Nikita* (Bayram et al. 2012) sequences. Callus culture conditions activate *BARE-1* and *Nikita* element. In the paper by Marakli et al. (2012), similarity of mature embryo, leaf, and root tissues grown from the same barley plant were investigated in terms of *BARE-1* and *BAGY2* movements. *BAGY2* was found to be more stable than *BARE-1*. Not all callus induction conditions increase retrotransposon activity (Temel and Gozukirmizi 2013). However, in addition to a *BAGY2* retrotransposon-specific IRAP polymorphism (Fig. 5), they also caused increase in copy numbers of internal domains of *BAGY2* (Yilmaz et al. 2014). Transformation of tobacco plant with the *dehE* gene, which degrades the herbicide Dalapon, was shown to cause activation of *Tto1* retrotransposon, one of the few active retrotransposons in tobacco (Kaya et al. 2013). Moreover, we investigated the nonautonomous retrotransposon *Sukkula* in barley (Kartal et al. 2014). Our group utilized the IRAP technique to assess the genotoxicity of some drugs, for example, epirubicine (Hamat-Mecbur et al. 2014) and amiprophos-methyl (Temel and Gozukirmizi 2014). Recently, we used IRAP markers for identification of variation in single seed derived leaves and roots in rice (Yuzbasioglu et al. 2016b).

9 Conclusions and Future Perspective

Transposable elements, particularly retroelements, are an important part of the genomes of complex organisms. Retroelements affect various aspects of human health (St. Laurent et al. 2010), but discussion of this impact is beyond the scope of

this chapter. Yet, review of recent developments in transposon research is essential. In this chapter, we tried to summarize the structure, types, and classification of transposons and present information about the most widely used transposon markers. We also mentioned the relationship of transposons with sequencing technologies, regulation of gene expression, and genome evolution. Data of our studies with transposons were briefly mentioned. We tried to cite as many papers as possible, but we apologize to those authors whose works went unmentioned in this chapter.

Acknowledgments We are grateful to the Research Fund of Istanbul University for financial support (Projects 20212 and 20316).

References

Aalami A, Safiyar S, Mandoulakani BA (2012) R-RAP: a retrotransposon-based DNA fingerprinting technique in plants. Plant Omics 5:359–364

Agarwal M, Shrivastava N, Padh H (2008) Advances in molecular marker techniques and their applications in plant sciences. Plant Cell Rep 27:617–631

Barkan A, Martienssen RA (1991) Inactivation of maize transposon *Mu* suppresses a mutant phenotype by activating an outward-reading promoter near the end of *Mu1*. Proc Natl Acad Sci USA 88:3502–3506

Barquist L, Boinett CJ, Cain AK (2013) Approaches to querying bacterial genomes with transposon-insertion sequencing. RNA Biol 10:1161–1169

Bayram E, Yilmaz S, Hamat-Mecbur H, Kartal-Alacam G (2012) *Nikita* retrotransposon movements in callus cultures of barley (*Hordeum vulgare* L.). Plant Omics 5:211–215

Bennetzen JL (2005) Transposable elements, gene creation and genome rearrangement in flowering plants. Curr Opin Genet Dev 15:621–627

Bhattacharyya MK, Smith AM, Noel Ellis TH, Hedley C, Martin C (1990) The wrinkled-seed character of pea described by Mendel is caused by a transposon-like insertion in a gene encoding starch-branching enzyme. Cell 60:115–122

Boyko A, Blevins T, Yao Y, Golubov A, Bilichak A, Ilnytskyy Y, Hollander J, Meins F Jr (2010) Transgenerational adaptation of *Arabidopsis* to stress requires DNA methylation and the function of dicer-like proteins. PLoS One 5:e9514

Bundock P, Hooykaas P (2005) An *Arabidopsis* hAT-like transposase is essential for plant development. Nature (Lond) 436:282–284

Capy P, Bazin C, Higuet D, Langin T (1998) Dynamics and evolution of transposable elements. Library of Congress, Austin, TX

Casacuberta JM, Santiago N (2003) Plant LTR-retrotransposons and MITEs: control of transposition and impact on the evolution of plant genes and genomes. Gene (Amst) 311:1–11

Cheng Z, Dong F, Langdon T, Ouyang S, Buell CR, Gu M, Blattner FR, Jiang J (2002) Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. Plant Cell 14:1691–1704

Christen B, Abeliuk E, Collier JM, Kalogeraki VS, Passarelli B, Coller JA, Fero MJ, McAdams HH, Shapiroa L (2011) The essential genome of a bacterium. Mol Syst Biol 7:1–7

Coates CJ, Kaminski JM, Summers JB, Segal DJ, Miller AD, Kolb AF (2005) Site-directed genome modification: derivatives of DNA-modifying enzymes as targeting tools. Trends Biotechnol 23:407–419

Cordaux R, Batzer MA (2009) The impact of retrotransposons on human genome evolution. Nat Rev Genet 10:691–703

Davila JA, Loarce Y, Ramsay L, Waugh R, Ferrer E (1999) Comparison of RAMP and SSR markers for the study of wild barley genetic diversity. Hereditas 131:5–13

Eickbush TH, Jamburuthugoda VK (2008) The diversity of retrotransposons and the properties of their reverse transcriptases. Virus Res 134:221–234

Evrensel C, Yilmaz S, Temel A, Gozukirmizi N (2011) Variations in *BARE-1* insertion patterns in barley callus cultures. Genet Mol Res 10:980–987

Feschotte C, Pritham EJ (2007) DNA transposons and the evolution of eukaryotic genomes. Annu Rev Genet 41:331–368

Feschotte C, Jiang N, Wessler SR (2002) Plant transposable elements: where genetics meets genomics. Nat Rev Genet 3:329–341

Finnegan DJ (1989) Eukaryotic transposable elements and genome evolution. Trends Genet 5:103–107

Flavell AJ (1995) Retroelements reverse transcriptase and evolution. Comp Biochem Physiol B Biochem Mol Biol 110:3–15

Flavell AJ, Knox MR, Pearce SR, Ellis TH (1998) Retrotransposon-based insertion polymorphisms (RBIP) for high throughput marker analysis. Plant J 16:643–650

Gabus C, Ficheux D, Rau M, Keith G, Sandmeyer S, Darlix JL (1998) The yeast Ty3 retrotransposon contains a 5′–3′ bipartite primer-binding site and encodes nucleocapsid protein NCp9 functionally homologous to HIV-1 NCp7. EMBO J 17:4873–4880

Gallagher LA, Shendure J, Manoil C (2011) Genome-scale identification of resistance functions in *Pseudomonas aeruginosa* using Tn-seq. MBio 2(1):e00315–e00310

Gawronski JD, Wong SM, Giannoukos G, Ward DV, Akerley BJ (2009) Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for *Haemophilus* genes required in the lung. Proc Natl Acad Sci USA 106:16422–16427

Gazzani S, Gendall AR, Lister C, Dean C (2003) Analysis of the molecular basis of flowering time variation in *Arabidopsis* accessions. Plant Physiol 132:1107–1114

Goodman AL, McNulty NP, Zhao Y, Leip D, Mitra RD, Lozupone CA, Knight R, Gordon JI (2009) Identifying genetic determinants needed to establish a human gut symbiont in its habitat. Cell Host Microbe 6:279–289

Goodman AL, Wu M, Gordon JI (2011) Identifying microbial fitness determinants by insertion sequencing using genome-wide transposon mutant libraries. Nat Protoc 6:1969–1980

Goodwin TJD, Poulter RTM, Lorenzen MD, Beeman RW (2004) DIRS retroelements in arthropods: identification of the recently active TcDirs1 element in the red flour beetle *Tribolium castaneum*. Mol Genet Genomics 272:47–56

Goryshin I, Jendrisak J, Hoffman L, Meis R, Reznikoff W (2000) Insertional transposon mutagenesis by electroporation of released Tn5 transposition complexes. Nat Biotechnol 18:97–100

Gozukirmizi N, Yilmaz S, Marakli S, Temel A (2015) Retrotransposon-based molecular markers: tools for variation analysis in plants. In: Tashki-Ajdukovic K (ed) Applications of molecular markers in plant genome analysis and breeding. Research Signpost/Transworld Research Network, Ontario, pp 19–45

Griffin J, Gawronski JD, DeJesus MA, Ioerger TR, Akerley BJ, Sassetti CM (2011) High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. PLoS Pathog 7:e1002251

Hall AE, Kettler GC, Preuss D (2006) Dynamic evolution at pericentromeres. Genome Res 16:355–364

Hamat-Mecbur H, Yilmaz S, Temel A, Sahin K, Gozukirmizi N (2014) Effects of epirubicin on barley seedlings. Toxicol Ind Health 30:52–59

Hamer L, DeZwaan TM, Montenegro-Chamorro MV, Frank SA, Hamer JE (2001) Recent advances in large-scale transposon mutagenesis. Curr Opin Chem Biol 5:67–73

Hawkins JS, Kim H, Nason JD, Wing RA, Wendel JF (2006) Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. Genome Res 16:1252–1261

Hollister JD, Gaut BS (2009) Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome Res 19:1419–1428

Hu TT, Pattyn P, Bakker EG, Cao J, Cheng JF, Clark RM, Fahlgren N, Fawcett JA, Grimwood J, Gundlach H et al (2011) The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. Nat Genet 43:476–481

Huang CRL, Burns KH, Boeke JD (2012) Active transposition in genomes. Annu Rev Genet 46:651–675

Ilic K, SanMiguel PJ, Bennetzen JL (2003) A complex history of rearrangement in an orthologous region of the maize, sorghum and rice genomes. Proc Natl Acad Sci USA 100:12265–12270

Ito H, Gaubert H, Bucher E, Mirouze M, Vaillant I, Paszkowski J (2011) An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. Nature (Lond) 472:115–119

Jaaskelainen M, Mykkanen AH, Arna T, Vicient CM, Suoniemi A, Kalendar R, Savilahti H, Schulman AH (1999) Retrotransposon BARE-1: expression of encoded proteins and formation of virus-like particles in barley cells. Plant J 20:413–422

Jiang N, Bao Z, Temnykh S, Cheng Z, Jiang J, Wing RA, McCouch SR, Wessler SR (2002) *Dasheng*: a recently amplified nonautonomous long terminal repeat element that is a major component of pericentromeric regions in rice. Genetics 161:1293–1305

Jiang N, Bao Z, Zhang X, Eddy SR, Wessler SR (2004) Pack-MULE transposable elements mediate gene evolution in plants. Nature (Lond) 431:569–573

Jians H (2012) Evolution, function, and regulation of genomic imprinting in plant seed development. J Exp Bot 63:4713–4722

Joly-Lopez Z, Bureau TE (2014) Diversity and evolution of transposable elements in *Arabidopsis*. Chromosome Res 22:203–216

Kalendar R, Schulman AH (2006) IRAP and REMAP for retrotransposon-based genotyping and fingerprinting. Nat Protoc 1:2478–2484

Kartal G, Yilmaz S, Marakli S, Gozukirmizi N (2014) *Sukkula* retrotransposon insertion polymorphism in barley. Russ J Plant Physiol 61:828–833

Kaya Y, Yilmaz S, Gozukirmizi N, Huyop F (2013) Evaluation of transgenic *Nicotiana tabacum* with *dehE* gene using transposon-based IRAP markers. Am J Plant Sci 4:41–44

Kazazian HH, Moran JV (1998) The impact of L1 retrotransposons on the human genome. Nat Genet 19:19–24

Kroutter EN, Belancio VP, Wagstaff BJ, Roy-Engel AM (2009) The RNA polymerase dictates ORF1 requirement and timing of LINE and SINE retrotransposition. PLoS Genet 5:e1000458

Langridge GC, Phan MD, Turner DJ, Perkins TT, Parts L, Haase J, Charles I, Maskell DJ, Peters SE, Dougan G et al (2009) Simultaneous assay of every *Salmonella typhi* gene using one million transposon mutants. Genome Res 19:2308–2316

Leigh F, Lea V, Law J, Wolters P, Powell W, Donini P (2003) Assessment of EST- and genomic microsatellite markers for variety discrimination and genetic diversity studies in wheat. Euphytica 133:359–366

Leitch AR, Leitch IJ (2008) Genomic plasticity and the diversity of polyploid plants. Science 320:481–483

Levin HL, Moran JV (2011) Dynamic interactions between transposable elements and their hosts. Nat Rev Genet 12:615

Li Y, Li C, Xia J, Jin Y (2011) Domestication of transposable elements into microRNA genes in plants. PLoS One 6:e19212

Liu J, He Y, Amasino R, Chen X (2004) siRNAs targeting an intronic transposon in the regulation of natural flowering behavior in *Arabidopsis*. Genes Dev 18:2873–2878

Llorens C, Munoz-Pomer A, Bernad L, Botella H (2009) Network dynamics of eukaryotic LTR retroelements beyond phylogenetic trees. Biol Direct 4:41

Lorenzi HA, Robledo G, Levin MJ (2006) The VIPER elements of trypanosomes constitute a novel group of tyrosine recombinase-enconding retrotransposons. Mol Biochem Parasitol 145:184–194

Ma JX, Bennetzen JL (2004) Rapid recent growth and divergence of rice nuclear genomes. Proc Natl Acad Sci USA 101:12404–12410

Ma J, Wing RA, Bennetzen JL, Jackson SA (2007) Plant centromere organization: a dynamic structure with conserved functions. Trends Genet 23:134–139

Madlung A, Comai L (2004) The effect of stress on genome regulation and structure. Ann Bot 94:481–495

Mak J, Kleiman L (1997) Primer tRNAs for reverse transcription. J Virol 71:8087–8095

Malik HS, Eickbush TH (1998) The RTE class of non-LTR retrotransposons is widely distributed in animals and is the origin of many SINEs. Mol Biol Evol 15:1123–1134

Malik HS, Eickbush TH (2001) Phylogenetic analysis of ribonuclease H domains suggests a late, chimeric origin of LTR-Retrotransposable elements and retroviruses. Genome Res 11: 1187–1197

Mann B, van Opijnen T, Wang J, Obert C, Wang YD, Carter R, McGoldrick DJ, Ridout G, Camilli A, Tuomanen EI, Rosch JW (2012) Control of virulence by small RNAs in *Streptococcus pneumoniae*. PLoS Pathog 8:e1002788

Marakli S, Yilmaz S, Gozukirmizi N (2012) *BARE1* and *BAGY2* retrotransposon movements and expression analyses in developing barley seedlings. Biotechnol Biotechnol Equip 26:3451–3456

Martienssen R, Barkan A, Taylor WC, Freeling M (1990) Somatically heritable switches in the DNA modification of Mu transposable elements monitored with a suppressible mutant in maize. Gene Dev 4:331–343

Martin A, Troadec C, Boualem A, Rajab M, Fernandez R, Morin H, Pitrat M, Dogimont C, Bendahmane A (2009) A transposon induced epigenetic change leads to sex determination in melon. Nature (Lond) 461:1135–1138

McClintock B (1984) The significance of responses of the genome to challenge. Science 226:792–801

McCue AD, Slotkin RK (2012) Transposable element small RNAs as regulators of gene expression. Trends Genet 28:616–623

McCue AD, Nuthikattu S, Slotkin RK (2013) Genome-wide identification of genes regulated in trans by transposable element small interfering RNAs. RNA Biol 10:1379–1395

Michaels SD, Amasino RM (1999) *FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering. Plant Cell 11:949–956

Michaels S, He Y, Scortecci KC, Amasino R (2003) Attenuation of *FLOWERING LOCUS C* activity as a mechanism for the evolution of summer-annual flowering behavior in *Arabidopsis*. Proc Natl Acad Sci USA 100:10102–10107

Miller JT, Dong F, Jackson SA, Song J, Jiang J (1998) Retrotransposon-related DNA sequences in the centromeres of grass chromosomes. Genetics 150:1615–1623

Monden Y, Yamaguchi K, Tahara M (2014) Application of iPBS in high-throughput sequencing for the development of retrotransposon-based molecular markers. Curr Plant Biol doi:10.1016/j.cpb.2014.09.001

Morgante M, Brunner S, Pea G, Fengler K, Zuccolo A, Rafalski A (2005) Gene duplication and exon shuffling by Helitron-like transposons generate intraspecies diversity in maize. Nat Genet 37:997–1002

Morgante M, De Paoli E, Radovic S (2007) Transposable elements and the plant pan-genomes. Curr Opin Plant Biol 10:149–155

Ohtsubo H, Kumekawa N, Ohtsubo E (1999) *RIRE2* a novel *gypsy*-type retrotransposon from rice. Genes Genet Syst 74:83–91

Parisod C, Alix K, Just J, Petit M, Sarilar V, Mhiri C, Ainouche M, Chalhoub B, Grandbastien MA (2010) Impact of transposable elements on the organization and function of allopolyploid genomes. New Phytol 186:37–45

Piednoel M, Goncalves IR, Higuet D, Bonnivard E (2011) Eukaryote DIRS1-like retrotransposons: an overview. BMC Genomics 12:621

Piegu B, Guyot R, Picault N, Roulin A, Sanyal A, Kim H, Collura K, Brar DS, Jackson S, Wing RA, Panaud O (2006) Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. Genome Res 16:1262–1269

Piriyapongsa J, Jordan IK (2007) A family of human microRNA genes from miniature inverted-repeat transposable elements. PLoS One 2:e203
Poczai P, Varga I, Laos M, Cseh A, Bell N, Valkonen JPT, Hyvonen J (2013) Advances in plant gene-targeted and functional markers: a review. Plant Methods 9:6
Poulter RT, Goodwin TJ (2005) DIRS-1 and the other tyrosine recombinase retrotransposons. Cytogenet Genome Res 110:575–588
Pouteau S, Grandbastien MA, Boccara M (1994) Microbial elicitors of plant defence responses activate transcription of a retrotransposon. Plant J 5:535–542
Pyatkov KI, Arkhipova IR, Malkova NV, Finnegan DJ, Evgen'ev MB (2004) Reverse transcriptase and endonuclease activities encoded by *Penelope*-like retroelements. Proc Natl Acad Sci USA 101:14719–14724
Richter TE, Ronald PC (2000) The evolution of disease resistance genes. Plant Mol Biol 42: 195–204
Saha S, Karaca M, Jenkins JN, Zipf AE, Reddy UK, Kantety RV (2003) Simple sequence repeats as useful resources to study transcribed genes of cotton. Euphytica 130:355–364
Salvi S, Sponza G, Morgante M, Tomes D, Niu X, Fengler KA, Meeley R, Ananiev EV, Svitashev S, Bruggemann E et al (2007) Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. Proc Natl Acad Sci USA 104:11376–11381
Sanmiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL (1998) The paleontology of intergene retrotransposons of maize. Nat Genet 20:43–45
Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115
Schulman AH, Flavell AJ, Ellis THN (2004) The application of LTR retrotransposons as molecular markers in plants. Methods Mol Biol 260:145–173
Slotkin RK, Nuthikattu S, Jiang N (2012) The impact of transposable elements on gene and genome evolution. Plant Genome Divers 1:35–58
St. Laurent G III, Hammell N, McCaffrey TA (2010) A LINE-1 component to human aging: do LINE elements exact a longevity cost for evolutionary advantage? Mech Ageing Dev 131:299–305
Szabo M, Kiss J, Olasz F (2010) Functional organization of the inverted repeats of IS30. J Bacteriol 192:3414–3423
Takeda S, Sugimoto K, Otsuki H, Hirochika H (1998) Transcriptional activation of the tobacco retrotransposon *Tto1* by wounding and methyl jasmonate. Plant Mol Biol 36:365–376
Talbert LE, Chandler VL (1988) Characterization of a highly conserved sequence related to Mutator transposable elements in maize. Mol Biol Evol 5:519–529
Temel A, Gozukirmizi N (2013) Analysis of retrotransposition and DNA methylation in barley callus culture. Acta Biol Hung 64:86–95
Temel A, Gozukirmizi N (2014) Genotoxicity of metaphase-arresting methods in barley. Turk J Biol. doi:10.3906/biy-1405-58
The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature (Lond) 408:796–815
The International Brachypodium Initiative (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature (Lond) 463:763–768
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. Plant J 13:121–129
van Opijnen T, Camilli A (2013) Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. Nat Rev Microbiol 11:435–442
van Opijnen T, Bodi KL, Camilli A (2009) Tn-seq: highthroughput parallel sequencing for fitness and genetic interaction studies in microorganisms. Nat Methods 6:767–772
Vicient CM, Suoniemi A, Anamthawat-Johnsson K, Tanskanen J, Beharav A, Nevo E, Schulman AH (1999) Retrotransposon *BARE-1* and its role in genome evolution in the genus *Hordeum*. Plant Cell 11:1769–1784

Vicient CM, Jaaskelainen M, Kalendar R, Schulman AH (2001) Active retrotransposons are a common feature of grass genomes. Plant Physiol 125:1283–1292

Vitte C, Bennetzen JL (2006) Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution. Proc Natl Acad Sci USA 103:17638–17643

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1993) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414

Wallace N, Wagstaff BJ, Deininger PL, Roy-Engel AM (2008) LINE-1 ORF1 protein enhances Alu SINE retrotransposition. Gene (Amst) 419:1–6

Wang Q, Dooner HK (2006) Remarkable variation in maize genome structure inferred from haplotype diversity at the *bz* locus. Proc Natl Acad Sci USA 103:17644–17649

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). Mol Gen Genet 253:687–694

Weil C, Martienssen R (2008) Epigenetic interactions between transposons and genes: lessons from plants. Curr Opin Genet Dev 18:188–192

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 Res 17:1072–1081

Wicker T, Zimmermann W, Perovic D, Paterson AH, Ganal M, Graner A, Stein N (2005) A detailed look at 7 million years of genome evolution in a 439-kb contiguous sequence at the barley *Hv-eIF4E* locus: recombination, rearrangements and repeats. Plant J 41:184–194

Wicker T, Sabot F, Hua-Van A, Benetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O et al (2007) A unified classified classification system for eukaryotic transposable elements. Nat Rev Genet 8:973–982

Wilhelm M, Wilhelm FX (2001) Reverse transcription of retroviruses and LTR-retrotransposons. Cell Mol Life Sci 58:1246–1262

Wright DA, Voytas DF (1998) Potential retroviruses in plants: Tat1 is related to a group of *Arabidopsis thaliana* Ty3/*gypsy* retrotransposons that encode envelope-like proteins. Genetics 149:703–715

Yadav VP, Mandal PK, Rao DN, Bhattacharya S (2009) Characterization of the restriction enzyme-like endonuclease encoded by the *Entamoeba histolytica* non-long terminal repeat retrotransposon EhLINE1. FEBS J 276:7070–7082

Yan L, Fu D, Li C, Blechl A, Tranquilli G, Bonafede M, Sanchez A, Valarik M, Yasuda S, Dubcovsky J (2006) The wheat and barley vernalization gene *VRN3* is an orthologue of FT. Proc Natl Acad Sci USA 103:19581–19586

Yilmaz S, Gozukirmizi N (2013) Variation of retrotransposon movement in callus culture and regenerated shoots of barley. Biotechnol Biotechnol Equip 27:4227–4230

Yilmaz S, Marakli S, Gozukirmizi N (2014) *BAGY2* retrotransposon analyses in barley calli cultures and regenerated plantlets. Biochem Genet 52:233–244

Yu Z, Wright SI, Bureau TE (2000) Mutator-like elements in *Arabidopsis thaliana*: structure, diversity and evolution. Genetics 156:2019–2031

Yuzbasioglu G, Yilmaz S, Gozukirmizi N (2016a) Houba retrotransposon-based molecular markers: a tool for variation analysis in rice. Turk J Agric For doi:10.3906/tar-1509-2

Yuzbasioglu G, Yilmaz S, Marakli S, Gozukirmizi N (2016b) Analysis of *Hopi/Osr27* and *Houba/Tos5/Osr13* retrotransposons in rice. Biotechnol Biotech Eq 30:213–218

Zhang J, Peterson T (2004) Transposition of reversed Ac element ends generates chromosome rearrangements in maize. Genetics 167:1929–1937

Zhao D, Ferguson AA, Jiang N (2016) What makes up plant genomes: The vanishing line between transposable elements and genes. Biochimica et Biophysica Acta 1859:366–380

Zhong CX, Marshall JB, Topp C, Mroczek R, Kato A, Nagaki K, Birchler JA, Jiang J, Dawe RK (2002) Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. Plant Cell 14:2825–2836

Zietkiewicz E, Rafalski A, Labuda D (1989) Genome singerprinting by simple sequence sepeat (SSR)-anchored polymerase chain reaction amplification. Genomics 20:176–183