



OPEN ACCESS

EDITED BY
Ruslan Kalendar,
University of Helsinki, Finland

REVIEWED BY
Tony Heitkam,
Technical University
Dresden, Germany
Kenji K. Kojima,
Genetic Information Research
Institute, United States
David Roquis,
Technical University of
Munich, Germany

*CORRESPONDENCE
Miroslav Baránek
miroslav.baranek@mendelu.cz

SPECIALTY SECTION
This article was submitted to
Technical Advances in Plant Science,
a section of the journal
Frontiers in Plant Science

RECEIVED 26 October 2022
ACCEPTED 17 November 2022
PUBLISHED 01 December 2022

CITATION
Bajus M, Macko-Podgórní A,
Grzebelus D and Baránek M (2022) A
review of strategies used to identify
transposition events in plant genomes.
Front. Plant Sci. 13:1080993.
doi: 10.3389/fpls.2022.1080993

COPYRIGHT
© 2022 Bajus, Macko-Podgórní,
Grzebelus and Baránek. This is an open-
access article distributed under the
terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use,
distribution or reproduction is
permitted which does not comply with
these terms.

A review of strategies used to identify transposition events in plant genomes

Marko Bajus¹, Alicja Macko-Podgórní², Dariusz Grzebelus²
and Miroslav Baránek^{1*}

¹Mendeleum—Institute of Genetics, Faculty of Horticulture, Mendel University in Brno, Lednice, Czechia, ²Department of Plant Biology and Biotechnology, Faculty of Biotechnology and Horticulture, University of Agriculture in Krakow, Kraków, Poland

Transposable elements (TEs) were initially considered redundant and dubbed 'junk DNA'. However, more recently they were recognized as an essential element of genome plasticity. In nature, they frequently become active upon exposition of the host to stress conditions. Even though most transposition events are neutral or even deleterious, occasionally they may happen to be beneficial, resulting in genetic novelty providing better fitness to the host. Hence, TE mobilization may promote adaptability and, in the long run, act as a significant evolutionary force. There are many examples of TE insertions resulting in increased tolerance to stresses or in novel features of crops which are appealing to the consumer. Possibly, TE-driven *de novo* variability could be utilized for crop improvement. However, in order to systematically study the mechanisms of TE/host interactions, it is necessary to have suitable tools to globally monitor any ongoing TE mobilization. With the development of novel potent technologies, new high-throughput strategies for studying TE dynamics are emerging. Here, we present currently available methods applied to monitor the activity of TEs in plants. We divide them on the basis of their operational principles, the position of target molecules in the process of transposition and their ability to capture real cases of actively transposing elements. Their possible theoretical and practical drawbacks are also discussed. Finally, conceivable strategies and combinations of methods resulting in an improved performance are proposed.

KEYWORDS

transposable elements, transposon mobilization, course of transposition, detection methods, eccDNA, bioinformatics tools

Introduction

Transposable elements (TEs) were found and described in the early 1950s by Barbara McClintock in maize, as entities causing chromosome breakage, with breaking points capable of changing their chromosomal positions (McClintock, 1950). The importance of her observation has eventually been recognized as fundamental and finally, more than

30 years after publishing her seminal paper, McClintock was awarded the Nobel prize (Ravindran, 2012).

TEs are abundant structural genome components inhabiting genomes throughout the course of life evolution (Chuong et al., 2017). Initially, TEs were considered unnecessary or even harmful components of the genome (Sotero-Caio et al., 2017). At present, it is commonly accepted that their interactions with the host genome are far more complex and still not fully understood. In plants, TEs are important drivers of genome evolution, propelling phenotypic variability in the course of crop domestication and improvement. Their representation in plant genomes varies, ranging from approximately 20% in small genomes, such as *Arabidopsis* to more than 80% in maize (Kim, 2017).

TEs are divided into two classes, according to their mechanism of transposition: Class I (retrotransposons) and Class II (DNA transposons). Retrotransposons use an RNA intermediate to be copied and subsequently inserted as a novel copy at a new position in the genome, which results in an increase of their copy numbers (Feschotte et al., 2002). Retrotransposons are further subdivided into those harboring long terminal repeats (long terminal repeat retrotransposons, LTR-RTs) and non-LTR retrotransposons, including Long Interspersed Nuclear Elements (LINEs) and Short Interspersed Nuclear Elements (SINEs). LTR-RTs are predominant in the TE landscape of plant genomes (Satheesh et al., 2021). In contrast, most DNA transposons physically excise and reinsert (a 'cut and paste' mechanism), while those classified as Helitrons utilize a 'rolling circle' mechanism for their transposition. Thus, transposition of Class II TEs does not involve any RNA intermediate. DNA transposons are widespread and active across many bacterial, archaeal and eukaryotic species, while their activity in mammals is low (Rodriguez-Terrones and Torres-Padilla, 2018). The distribution of TEs in plant genomes has been reviewed in more detail by Sahebi et al. (2018).

Most successful TE mobilization events are neutral or even deleterious to the host. They can cause changes in the pattern of gene expression and alter gene function by up- or down-regulating adjacent genes following insertion into promoter regions, introns, exons or downstream regions (Makarevitch et al., 2015; Deneweth et al., 2022). Also, they may become a source of small interfering RNAs (siRNAs) (Piriyapongsa and Jordan, 2008; Gill et al., 2021). In order to protect integrity of the host genome, TEs are silenced and the state is epigenetically heritable (Fultz et al., 2015). In general, *de novo* silencing of active TE involves DNA methylation and repressive modifications of histones. These epigenetic marks are maintained across subsequent mitotic divisions and transmitted from generation to generation. Importantly, precise mechanisms resulting in TE inactivation depend on the location of a TE copy in the genomic context (Sigman and Slotkin, 2016)

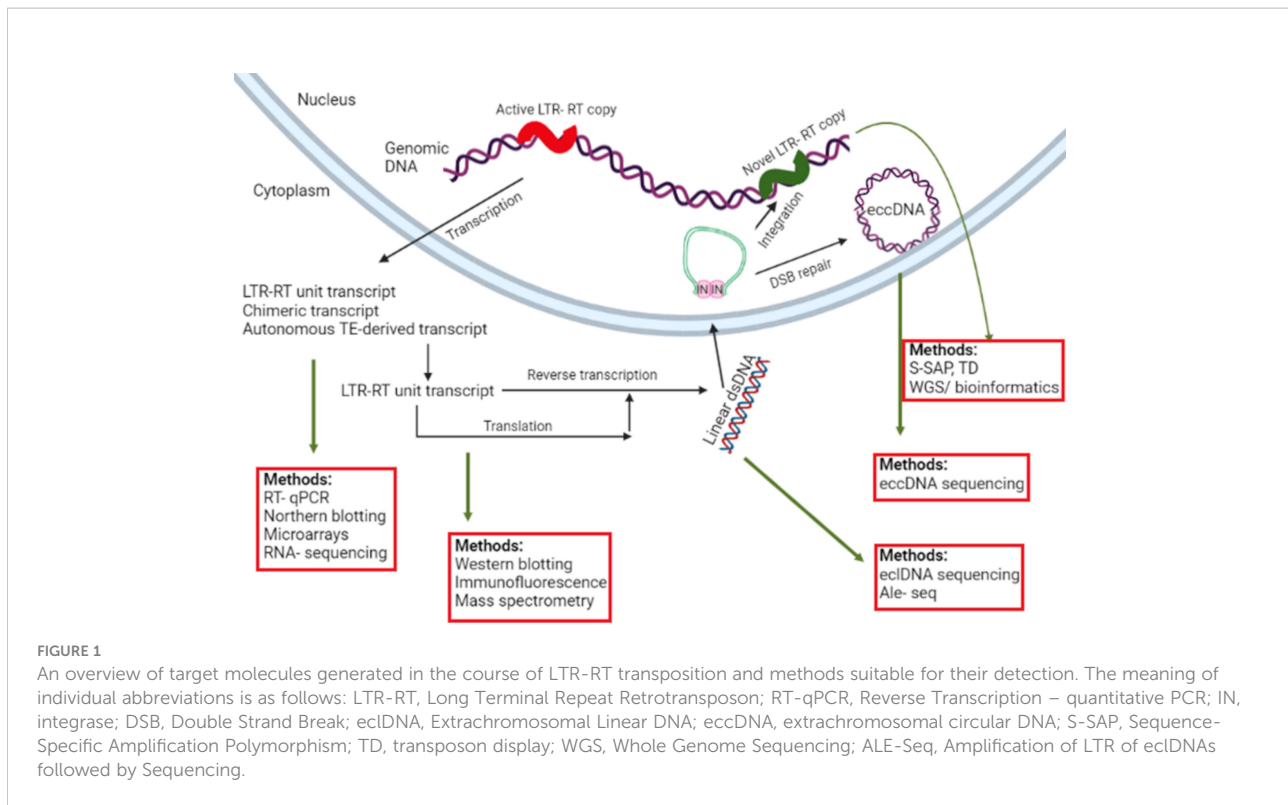
In order to recognize TEs showing ongoing activity, it is necessary to use tools targeting one of the molecules produced in the course of mobilization, i.e. RNA transcripts, extrachromosomal linear DNA (ecDNA), extrachromosomal circular DNA (eccDNA), small RNA or TE-encoded proteins (Figure 1). It is also important to monitor whether mobilized copies are competent to successfully reintegrate with the host genome to produce novel insertion sites.

The approach used by B. McClintock can be viewed as the first method of monitoring TE activity, as she observed that *Ac* as an activator autonomous TE mobilized non-autonomous *Ds* elements resulting in chromatid breakage. Fortunately, we have come a long way since then, and new possibilities and approaches are constantly emerging. The subject of the review is to summarize methods used for the analysis of TE activity and to discuss their advantages and specific applications. Special attention is paid to the LTR-RTs, which are considered the most abundant TEs in plant genomes (Deniz et al., 2019). The described methods are divided on the basis of their operational principles, the position of target molecules in the process of transposition and their ability to capture real cases of actively transposing elements. Their possible theoretical and practical drawbacks are also discussed. Finally, conceivable strategies and combinations of methods resulting in an improved performance are proposed.

Detection of TE-derived transcripts

As LTR-RTs require the formation of an RNA intermediate, it is the first target usable for the evaluation of their activity. Generally, LTR-RT-derived RNAs can be identified using tools similar to those used for monitoring gene expression, i.e. techniques based on nucleic acid hybridization (northern blotting, microarrays), PCR (RT-qPCR), or transcriptome sequencing (RNA-seq).

Historically, northern blotting was used as the first method of choice (Manninen and Schulman, 1993; Meyer et al., 1994; Pozueta-Romero et al., 1995). With the development of new technologies, its significance gradually declined due to the complexity of protocols and necessity to ensure high amounts of input RNA. Subsequently, methods based on RT-qPCR started to be utilized to monitor TE activity in plants (Marcon et al., 2015; Paz et al., 2015; Jiang et al., 2016; Voronova, 2019; Usai et al., 2020). An important limitation of RT-qPCR is that it targets individual copies or TE families grouping very similar copies and specificity is provided by primers used for qPCR. Hence, the assay requires prior knowledge about propensity of the studied copy to be mobilized. On the other hand, it may be problematic to design specific primers to investigate TEs from different families (Morillon et al., 2002). Another limitation is the fact that the target sequence may include nucleotide substitutions and/or indels in transcripts produced from



different copies. In such case, northern blotting seems to be a good complementary method, as it may reveal the size distribution of TE-derived transcripts, including full length TEs (Böhner et al., 2020).

A global analysis of TE-derived transcripts can be produced with microarrays (Picault et al., 2009; Rocheta et al., 2016). Comprehensive information about the whole spectrum of actively transcribed TEs can also be captured by RNA-seq based on massive parallel DNA sequencing technologies (Gürkök, 2017; Oberlin et al., 2017; Qiu and Ungerer, 2018; Vangelisti et al., 2019; Jiménez-Ruiz et al., 2020; Kirov et al., 2020). RNA-seq data have been utilized and interpreted differently in reports aiming at the description of global activity of TEs. While some reports simply presented a spectrum of TEs captured in RNA-seq reads (Gürkök, 2017; Jiménez-Ruiz et al., 2020), in other reports, especially those concerning plant species for which high quality reference genomes were available, TE-derived transcripts were mapped to the reference genome assembly (Li et al., 2010; Hollister et al., 2011; Valdebenito-Maturana and Riadi, 2018). However, owing to the fact that some TE families comprise numerous copies and the evolutionary relationships among TE families can be complex, interpretation of the RNA-seq data remains challenging. Different strategies have been implemented, solely or in combination, to confirm TE expression from RNA-seq data, i.e. mapping TE-derived reads to a reference genome, a TE pseudogenome and a model transcriptome (Lanciano and

Cristofari, 2020). Precision of the mapping process can be significantly improved by using longer reads provided by PacBio or Oxford Nanopore technologies (Sexton and Han, 2019). When using them it is much easier to predict if the sequenced TE-derived transcript has a potential to complete its full life cycle, or vice versa, whether it does not contain signs of inactive forms such as chimeric transcripts. Available bioinformatic tools and techniques for TE mapping to reference genomes were recently reviewed by O'Neill et al. (2020).

In general, with respect to all TE-derived transcript targeting techniques, it is necessary to be aware that there are issues that can impact clarity of results when the primary interest is to investigate only actively transposing TEs. It is because a significant share of TEs is transcribed by PolII and processed into 21~24 nt siRNA, involved in epigenetic silencing of TEs (Tang et al., 2022). Moreover, stress-dependent genome demethylation (Pandey et al., 2017; Liang et al., 2019) may result in increased expression of TEs. Also, transcripts containing sequences derived from TEs may also include chimeric transcripts containing both TE and genic fragments, e.g. those resulting from the initiation of transcription from a TE promoter or from exonization of intronic TE insertions. Such transcripts are obviously not an indication of ongoing transposition activity, but still they can be abundant in RNA samples. Besides, active post-transcriptional suppression mechanisms by TE-derived sequences was also described

(Fultz et al., 2015). The above-described drawbacks and the fact that transcription is only an initial step in the process of transposition suggest that monitoring TE-derived transcripts is not an optimal strategy aiming at the identification of TEs capable of completing new insertion. There is a serious risk of misinterpretations and incorrect conclusions deeply discussed also by Deininger et al. (2017). However, expression-based assays can be used to support results concerning TE mobility produced by using other approaches.

Detection of TE-encoded proteins

One of the possible manifestations of TE mobilization is translation of TE-encoded proteins constituting an essential transposition machinery. Thus, theoretically such proteins can also be used for monitoring an ongoing process of TE mobilization. It is necessary to emphasize that some types of TEs, e.g. SINEs or MITEs, referred to as non-autonomous, do not encode any proteins and utilize transposition machinery provided by their autonomous counterparts, LINEs and related DNA transposons, respectively. Historically, proteomic studies related to TE activity were based on western blotting. Western blot is an analytical technique used to detect a specific protein in a mixture of all proteins extracted from a tissue sample. Thus, TE mobilization-related experiments focus on a limited group of TE-derived proteins, such as transposases (Torres et al., 2013). The advantage of western blotting is that it can reveal events where internal mutations within coding regions of a TE prevent protein translation and subsequently hamper TE transposition. Such cases remained unrevealed by the analysis of TE-derived transcripts. Drawbacks of western blotting include limited availability and sensitivity of reagents, potential nonspecific activity of antibodies between related families of TEs, and necessity to produce large quantities of the starting material.

One of the most promising approaches for proteomic analysis is the application of methods based on mass spectrometry (MS) that may provide broad-spectrum results. Generally, MS is used to determine the mass of particles in order to determine the elemental composition and chemical structure of molecules, including complex substances, such as peptides. In the case of peptide analysis, combination of liquid chromatography (LC) with MS (LC-MS or LC-MS/MS), allowing for broad-spectrum analyses even down to the level of their amino acid sequences, are the most frequently used techniques. Obtained sequences can subsequently be evaluated with respect to the presence and the type of TE-derived proteins in analysed samples (Maringer et al., 2017). For example, Vuong et al. (2019) used MS to identify proteins of human TEs belonging to the L1 family of LINEs. In turn, Wang et al. (2008) used LC-MS/MS to study proteins activated by the moss *Physcomitrella patens* upon high salinity stress, revealing TE-derived proteins as being differentially expressed. Matrix

Assisted Laser Desorption Ionization - Time of Flight (MALDI-TOF-TOF) combined with MS was also used to reveal proteomic background of sporadic flowering in bamboo species, suggesting a direct relationship of TE activation and the induction of flowering (Louis et al., 2015).

With respect to the fact that proteins are synthesized in initial stages of the TE transposition process, it is necessary to realize that proteomics, while allowing for detection of actively transposing TE, also bears some limitations. Feschotte and Pritham (2007) reported that ancient TEs were less likely to be actively transposing, however they might still express proteins, especially when they originated from domesticated TEs, and at present those proteins fulfill essential host cell functions. Altogether, proteomic techniques may provide unique insights to investigations on the TE activity, e.g. involvement of TE-derived proteins in the assembly of protein complexes. However, the employment of complementary strategies is needed to obtain a comprehensive landscape of actively transposing TEs. Proteomics Informed by Transcriptomics (PIT) may be one such prospective strategy. In this method, proteomic MS/MS spectra are searched against open reading frames derived from assembled RNA-Seq transcripts. This approach can reveal previously unknown translated genomic elements or can also identify hotspots of incomplete genome annotation. PIT was initially generated in general principle, however, it can be easily tuned to investigate TE ongoing activity (Davidson et al., 2017; Maringer et al., 2017).

Detection of extrachromosomal linear DNA

The formation of extrachromosomal linear DNA (eclDNA) molecules is inherent to the process of LTR-RT mobilization. LTR-RTs contain two ORFs, Gag encoding a coat protein, and Pol encoding a polyprotein comprising four domains, i.e. reverse-transcriptase (RT), RNase H (RH), aspartic protease (AP) and integrase (INT). The life cycle of LTR-RTs begins with transcription of an active LTR-RT copy by a host-encoded RNA polymerase II, followed by synthesis of LTR-RT-encoded proteins, formation of virus-like particles (VLPs) encapsulating the RNA template, and its reverse transcription resulting in the formation of eclDNA. Subsequently, eclDNA enters the nucleus and integrates with the host genome (Havecker et al., 2004). Thus, the detection of eclDNAs seems to be an exquisite approach to mine for actively transposing LTR-RTs (Grandbastien, 2015), as they represent the final intermediates in LTR-RT retrotransposition (Figure 1). However, eclDNA can occur in cells also as a result of other events, such as cell lysis-originating eclDNA, as cells are constantly being lysed, or extrachromosomal linear microDNA interspersed with microRNAs (Sun et al., 2019). All these eclDNA sources may contain LTR-RT sequences, but only in the case of linear products resulting from the transposition process, the

identified fragment is expected to be terminated with LTR sequences, without additional fragments of genomic DNA sequence. Thus, a stage allowing selection of LTR-RTs should be included. A strategy based on PCR amplification utilizing a primer annealing to the tRNA primer binding site (PBS) could be used. It was originally applied to generate PCR-based iPBS molecular markers (Kalendar et al., 2010), while later it became the basis of SIRT (Sequence-Independent Retrotransposon Trapping) – the first method using LTR-RT-derived ecDNAs as targets (Griffiths et al., 2018). It took advantage from the fact that ecDNA ends are blunt-ended and competent for ligation of synthetic adaptors. Subsequently, using PCR primers complementary to the adaptor and to the PBS, a segment comprising the 5' LTR was amplified. When compiling complementary PBS primers, they used the fact that actively transposing LTR-RTs described in plants use predominantly as the initiator methionine tRNA (Met-iCAT) (Wicker et al., 2007; Kalendar et al., 2010). Thus, PBS sequences consist of 12 nucleotides complementary to the terminal nucleotides of the MET-iCAT tRNA. To ensure specific PCR amplification, the PBS-specific primers were therefore extended using the knowledge that two terminal nucleotides of 5' LTR mostly end in cytidine and adenosine (Griffiths et al., 2018). The disadvantage of the SIRT method is that it utilizes Sanger sequencing and that PBS-anchored primers are specific to particular LTR-RTs, which limits its usefulness for a global analysis of all LTR-RT families. It also turned out that the concept cannot be applied to large and TE-rich genomes,

To eliminate these disadvantages, the ALE-Seq (amplification of LTR of ecDNAs followed by sequencing) approach was developed (Cho et al., 2019). In comparison to SIRT, the ALE-Seq protocol utilizes more versatile primers complementary to PBS (or their combinations), high throughput sequencing, and is more elaborate as it includes adapter ligation, transcription and reverse transcription targeted to PBS domains. On the other hand, the ALE-Seq protocol is markedly more selective and efficient than SIRT, which relies on the single PCR amplification (Cho et al., 2019). The method is relatively recent, its applicability has been proved by the identification of actively transposing LTR-RTs in rice and tomato. On the basis of subsequent clustering of sequenced reads some retroelements were recognized as newly identified families for the respective genome. To summarize, ALE-Seq has potential for future use allowing reference-free annotation of new, active retroelements, what is especially important in plant species for which no reference genome assemblies are available (Satheesh et al., 2021).

Detection of extrachromosomal circular DNA

Some LTR-RT-derived ecDNA molecules were shown to be circularized. As integrase (IN) molecules are attached to LTRs of ecDNAs, their homodimerization causes the formation of a

pseudocircular but unclosed structures. Following their recognition as double strand breaks by DNA repair machineries in the nucleus, they are ligated resulting in closed extrachromosomal circular DNA (eccDNA) molecules (Figure 1). As such, they do not directly participate in the process of transposition and can be seen as mobilization by-products, however, their presence provides information about actively transposing LTR-RTs (Lanciano et al., 2017).

It should be stressed LTR-RT transposition is not a sole source of eccDNAs; they can also occur as a result of other cellular processes. They are common in eukaryotes and can be very heterogenic in number, length, origin, and role as reviewed by Cao et al. (2021).

The first methods of eccDNA detection, i.e. inverse PCR amplification of LTR-LTR junctions and electron microscopy, suggested that some circles originated from TEs, mostly LTR-RTs (Hirochika and Otsuki, 1995) and Mutator-like class II elements (Sundaresan and Freeling, 1987). Advances in sequencing techniques contributed to the development of efficient eccDNA detection methods along with the bioinformatics tools for analysis of such data.

The first high-throughput method of sequencing eccDNA, Circle-Seq, was developed for yeast and consisted of alkaline-based extraction of circular DNA, followed by digestion of linear DNA, eccDNA amplification using ϕ 29 DNA polymerase and sequencing on the Illumina platform using SE mode (Møller et al., 2015). Soon after, based on similar assumptions, a standardized Mobilome-seq protocol of extraction and Illumina SE sequencing of eccDNA from plant tissues was established (Lanciano et al., 2017). Another approach, CIDER-Seq (Circular DNA Enrichment sequencing) method, originally developed for analysis of plants infected with viruses, utilizes electrophoresis-based size-selection as the first step of sample preparation, followed by random amplification of circular DNA with ϕ 29 DNA polymerase, repair by DNA polymerase I and sequencing using Single Molecule Real Time sequencing (Pacific Biosciences) (Mehta et al., 2019).

The production of large amounts of sequencing data raises the need for simultaneous development of analytical tools. Circle-Map (Prada-Luengo et al., 2019) and Circle_finder (Kumar et al., 2020) were developed for identification of human tumor related eccDNA sequenced using short-reads technology. The downside to these tools is that they both require a reference genome as an input file and they were not tested on plant data. Short reads can be also analysed using ECCsplorer (Mann et al., 2022), a tool for mapping reads to the reference genome, identifying genomic origin of eccDNAs on the basis of read distribution, coverage, discordant mapping, and split reads, but also enabling reference-free clustering of reads. This helps to identify and annotate LTR-RTs enriched in eccDNA libraries. eccDNA analysis from long reads is possible using the CIDER-seq2 (Mehta et al., 2020). Although the method was developed for identification and characterization

of plant virus genomes, and includes the ‘annotate’ module that is restricted to viruses annotation, part of the pipeline that outputs eccDNA candidates and their genomic localization can be used for the identification of LTR-RTs. Other long-reads based tools, such as CReCIL (Wanchai et al., 2022) allow not only efficient identification of circular DNA but also annotation and Circos-based visualization of assembled circles, but its performance was tested only on long-reads from mammals eccDNA sequencing. Another tool, ecc_finder (Zhang et al., 2021) is based on a pipeline applied for the analysis of Mobilome-seq data originated from plant tissues (Lanciano et al., 2017). The pipeline allows analysis of both short and long reads and can be run in the reference genome and reference-free modes.

The eccDNA identification was reported to be useful for monitoring mobilization of previously known actively transposing TEs in Arabidopsis, rice and tomato (Lanciano et al., 2017; Benoit et al., 2019; Lanciano et al., 2021; Roquis et al., 2021; Wang et al., 2021; Zhang et al., 2021; Mann et al., 2022) and *de novo* identification of mobilized LTR-RTs, as shown for potato (Esposito et al., 2019), poplar (Sow et al., 2021) and carrot (Kwolek et al., 2022).

Mapping ecDNA or eccDNA sequencing reads to the reference genome may provide a clue as to what is the TE copy that has been undergoing mobilization. Ideally, a reference genome highly related to the individual used for ecDNA or eccDNA should be used. However, the typical properties of TEs, such as their highly repetitive character and the fact that TE families can be highly interrelated within a given species may complicate conclusions driven from such analyses.

Identification of novel insertion sites produced by actively transposing elements

The life cycle of a TE is completed upon its insertion into a new position in the host genome (Figure 1). Such *de novo* insertions are thus present in the progeny while they are absent in the ancestral plants. In earlier studies, these uncommon events were recognized only when they resulted in changed phenotypes. Obviously, these events represent a very small proportion of the total number of successful transpositions resulting in the integration occurring in genic regions.

Historically, the principles of positional (genetic map-based) cloning were used to identify insertional polymorphisms in the genome. However, mapping with high resolution requires numerous mapping populations and many genetic markers, thus it is costly and time consuming. It is therefore not suitable for mapping newly transposed TEs, although one can find some examples here as well (Bortiri et al., 2006). Identification of TE insertion sites and resulting transposon insertion polymorphisms (TIPs) can be also performed using marker systems derived from

conservative sequences specific to certain TEs (Kalendar and Schulman, 2006) or by a modification of the amplified fragment length polymorphism (AFLP) protocol (Vos et al., 1995). It is based on comparing the distribution of copies of a particular TE family in a collection of closely related accessions and works especially well for TE families with a number of copies highly uniform in their sequence, which is a proxy for recent or ongoing transposition. Two AFLP modifications aiming at the identification of TIPs have been developed, i.e. sequence-specific amplification polymorphism (S-SAP), used for the identification of LTR-RT insertions, where the final amplification is performed with a retrotransposon-specific and a *MseI*-adaptor-specific primer (Vaughn et al., 1997), and transposon display (TD) using two rounds of PCR with nested transposon-specific primers (Casa et al., 2000; Grzebelus et al., 2007) and applied mostly to identify TIPs produced by DNA transposons. Those methods have often been used to identify TIPs derived from few known TE families. One of the first attempts where the S-SAP method was successfully applied to identify a newly inserted LTR-RT was reported by Tahara et al. (2004). They identified Ty1-copia retrotransposons in sweet potato activated in the callus. Similar approach was used by Yamashita and Tahara (2006), where a polymorphic S-SAP product was identified as a LINE retroelement activated in meristem stem cells. There are examples of S-SAP being successfully used also to identify ongoing transpositions upon stress other than *in vitro* cultures. For example, Woodrow et al. (2010) identified Ty1-copia transposition in durum wheat under salt and light stress. The effect of interspecific hybridization and polyploidization on the actively transposing LTR-RT using S-SAP was evaluated by Gantuz et al. (2022). Another TIP identification system named palindromic sequence-targeted PCR (PST-PCR v.2) was proposed by Kalendar et al. (2021). It relies on the use of capturing primers targeting palindromic sequences arbitrarily present in natural DNA templates in combination with a sequence-specific primer. PST-PCR v.2 consists of two rounds of PCR. The first round utilizes a combination of one sequence-specific primer with one capturing (PST) primer. The second round uses a combination of a single (preferred) or two universal primers; one anneals to a 5' tail attached to the sequence-specific primer and the other anneals to a different 5' tail attached to the PST primer. The key advantage of PST-PCR v.2 is to quickly produce amplified PCR fragments containing a portion of the template flanked by the sequence-specific and capturing primers. The approach allowed characterization of Ac transposon integration sites (Kalendar et al., 2021). Lack of restriction digestion and adapter ligation, i.e. steps required in S-SAP or TD, reduces the cost and time of identifying new insertion sites.

All wet-lab methods are primarily useful for monitoring the mobilization of previously identified TEs, e.g. under stress conditions or in a range of genetically diverse accessions, since they require the use of primers with a sequence specific to the sequence of the investigated TE. Moreover, the specificity of the

amplification and the reliability of the new insertion sites should be confirmed by sequencing.

In 2004, the 454 technology became commercially available next generation sequencing (NGS) platform. Since then, NGS began to be widely applied to study plant TEs. In the early stages, they were usually combined with other techniques based on PCR amplification of regions specific to TEs. As an example, [Monden et al. \(2014\)](#), produced a LTR-RT libraries derived from eight strawberry cultivars, based on the primer binding site (PBS) adjacent to the conserved 5' LTR motif and sequenced them using Illumina HiSeq2000. It allowed detection of cultivar-specific LTR-RT insertion sites.

Another approach for genome-wide TIPs detection produced by a single TE family includes AFLP-based enrichment of DNA fragments in TE sequences followed by Illumina library preparation and sequencing. The recently published TEaseq pipeline ([Lyu et al., 2021](#)) developed for maize Ds transposons consists of samples barcoding, TE enrichment, library preparation and Illumina sequencing. The bioinformatics workflow for sequencing data analysis starts from de-barcoding, next reads containing the TE sequence are identified, the TE-portion of the read is trimmed and the remaining portion of the sequence is mapped against the reference genome to identify the insertion site. The method was successfully used for the identification of 35,696 putative germinal insertion sites in over 1,600 Ds insertional mutants. The major advantage of such approach is not only more detailed information about the number of TE insertions and the level of polymorphism among tested individuals but also the availability of sequences of regions flanking insertions, that is vital for verification of novel insertion sites and their downstream analyses.

With the advent of high throughput sequencing technologies, strategies have been developed to mine for TE insertion sites using raw reads and a suite of bioinformatics tools is currently available ([Serrato-Capuchina and Matute, 2018](#); [Vendrell-Mir et al., 2019](#); [Fan et al., 2022](#)). Depending on the purpose of the analysis and the type of investigated TEs, different tools and approaches are being developed. Some tools like the TRACKPOSON ([Carpentier et al., 2019](#)) can identify TIPs very quickly and efficiently using discordant reads identified in the process of reads mapping against a TE sequence for the identification of insertions based on their position in the reference genome. It shortens time of the analysis at the expense of the precise determination of the site of insertion. Nevertheless, the identification of 'insertion signatures', i.e. TE sequences in specified genomic windows rather than their precise locations, might be the first choice for large-scale analysis of LTR-RTs, including thousands of re-sequenced genomes, as shown for the analysis of 3,000 rice genomes ([Carpentier et al., 2019](#)). The method reports both reference and non-reference insertions and does not require any prior TE annotation in the reference genome.

Tools based on the usage of discordant reads and split-reads report precise localization of insertion sites. That group of tools

often requires high quality annotation of TEs in the reference genome, which in case of non-model organisms may limit their utility. In spite of higher computation demands, they can be efficiently used for large-scale population studies. Evaluation of this type of analysis make easier if another selective step is included in the experiment, such as the principle of TE sequence capture described firstly by [Baillie et al. \(2011\)](#) on the example of retrotranspositions registered in the human brain. Subsequently, this principle was used by [Quadrana et al. \(2016\)](#) in mining of transposition events within sequencing data for 211 *Arabidopsis thaliana* accessions. The SPLITREADER used here ([Quadrana et al., 2016](#)) was utilised for a global analysis of LTR-RTs in 602 tomato accessions and TIP-based GWAS (TE-GWAS; TIP-GWAS), that allowed identification of retrotransposon insertions associated with important phenotypic traits, such as flavor ([Domínguez et al., 2020](#)), while insertional polymorphism of class II MITEs in 3,000 rice genomes was analysed using PoPoolationTE2 ([Kofler et al., 2016](#)) and TIP-based GWAS showed association of particular MITE copies with MITE copy number, suggesting that MITE subfamilies originate from few "master" copies ([Castanera et al., 2021](#)). Another short read based method, RelocaTE2 ([Chen et al., 2017](#)) was used to analyse copy number and distribution of mPing, Ping and Pong class II elements actively transposing in rice in 3,000 rice genomes ([Chen et al., 2019](#)) and to detect *de novo* insertions of mPing in 272 rice recombinant inbred lines (RILs) developed from a cross between Nipponbare and HEG4 known to carry active mPing ([Chen et al., 2020](#)).

The obvious prerequisite for their utilization is availability of a high quality reference genome. The combination of high throughput sequencing and *in silico* discovery of new TE insertion events currently seems to be the most efficient strategy. Nevertheless, the risk that some new insertions are not being recorded still remains, but can be reduced by sufficient amount of reads i.e. it is necessary to achieve a high sequencing coverage.

It is also possible to utilize a pan-genome approach, i.e. to compare two or more genome assemblies representing the same or closely related species, with the intention of finding TIPs differentiating those genomes. However, availability of multiple genome assemblies limits the usage of such approach to the identification of TIPs and analyses of contribution of TEs to genome organization, as shown for four maize genotypes ([Anderson et al., 2019](#)), rather than for tracking or identification of active TEs.

A further significant improvement in the identification of TIPs may be achieved by the use of long-read NGS techniques, such as the Oxford Nanopore technology (ONT) ([Ellison and Cao, 2020](#); [Ewing et al., 2020](#)). While short reads technologies work well in identifying insertion sites of small TEs, such as MITEs, long reads significantly improve the efficiency of analysis of longer elements, especially LTR-RTs that are the most abundant TEs in plant genomes. For example, the utility ONT was shown for detection of novel insertions of actively

transposing LTR-RTs in Arabidopsis; EVD (Debladis et al., 2017) and ONSSEN (Kirov et al., 2021), as well as for the identification of TIPs in collections of insertional mutants of Medicago and soybean (Song et al., 2021). Along with the development of long read sequencing, tools dedicated to the identification of insertions in such data are becoming available. The first tool identifying TIPs in long read data was PALMER (Pre-mASKing Long reads for Mobile Element insertion), based on the alignment of reads to the genome and masking reference insertions of the investigated TE family in the reads sequence. Subsequently, the TE sequence is identified in the unmasked part of the read and, based on the presence of specific features, the software identifies ends of TE, and the remaining part of the read is used to detect non-reference insertion sites (Zhou et al., 2020). The method was successfully applied to the human genome and it was adjusted to the most common human TEs (L1, Alu, SVA). Hence it may not work for other types of TEs, e.g. those abundant in plant genomes. Another pipeline, also developed to screen actively transposing human TEs, utilizes a slightly different strategy, as in the reads the portion mapped to the genome is masked, while the remaining part is mapped to a TE library, TE sequences are reconstructed and the remaining part of the sequence is re-mapped to the reference genome to identify non-reference insertions. In addition to TIPs identification, this pipeline allows analysis of TEs methylation, that is called by the software dedicated for identification of CpG methylation in ONT reads (Ewing et al., 2020). The long read sequencing methods produce reads overlapping full TE sequences and their flanking regions, providing opportunity for comprehensive characterization of those sequences. They also allow identification of TEs insertions within repetitive regions.

However, for the identification of novel insertions of actively transposing elements, especially in plants, the Illumina platform is still a method of choice, as efficient bioinformatic tools have been available and the cost of sequencing is still much lower. The Cas9-targeted sequence capture to enrich library with TE sequences, in combination with long read sequencing, may be an alternative solution, that would reduce the cost of sequencing while still benefiting from the advantage provided by long reads (McDonald et al., 2021).

Long read sequencing also improves genome assemblies in TE-rich regions, TE detection, annotation and identification of TIPs (Shahid and Slotkin, 2020), opening new perspectives for better understanding of the TE biology and activity.

Based on the information provided, a screening was carried out to estimate the popularity of selected perspective approaches in the last period (see Figure 2). Here it is confirmed that the frequency of their use is generally increasing, especially in the last 2 years, while the use of Oxford Nanopore technology seems to be as most frequently used from compared approaches. Finally, the most important advantages and disadvantages of all discussed detection techniques were summarized (see Table 1).

Concluding remarks and future perspectives

Historically, the importance of TEs in plant genomes has been neglected. However, it turned out that their presence affects many areas important for the life and development of plants, as well as in terms of their possible use in the field of plant breeding. It puts pressure on the availability of suitable analytical methods to trace the

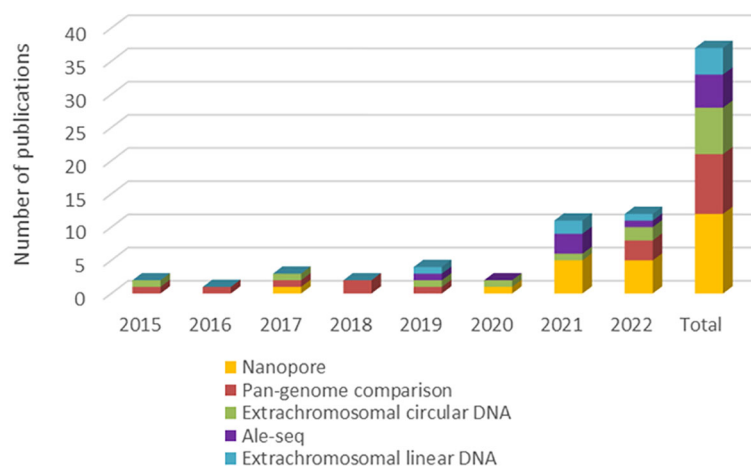


FIGURE 2

Popularity estimation of selected perspective approaches based on the frequency of their use in recent scientific articles. * The number of publications was generated by a search combining core keywords "plant + transposable + activ*" and keywords corresponding to individual perspective approaches.

TABLE 1 Summary of approaches used to identify actively transposing elements.

Strategy used to identify actively transposing TEs	Main drawbacks	Recommendations for efficient targeting actively transposing TEs
Targeting TE-derived transcripts	- existence of TE-derived transcripts not competent for transposition (chimeric transcripts; transcripts involved in epigenetic silencing of TEs; post-transcriptional suppression mechanisms by TE-derived sequences)	- combine with another technique targeting products from the final phases of the transposition process (e.g. eclDNA, eccDNA)
Targeting TE-derived proteins	- non-transposing TEs can still express proteins - requires equipment that is not so common in molecular genetics laboratories	- combine with another technique targeting products from the final phases of the transposition process (e.g. eclDNA, eccDNA) - PIT (Proteomics Informed by Transcriptomics)
Targeting eclDNA	- eclDNA can occur as a result of other cellular processes (e.g. cell lysis, existence of micro-eclDNA)	- include a selective step to enrich TE-derived eclDNAs (e.g. PBS complementary to MET-iCAT tRNA) - combine with high throughput sequencing (ALE-Seq)
Targeting eccDNA	- eccDNA does not directly participate in the process of transposition - eccDNA can occur as a result of other cellular processes	- combine with high throughput sequencing to identify novel insertion sites
Identification of novel insertion sites by using TE-based genotyping platforms	- laborious and time consuming and error-prone	- use PST-PCR v.2 as a less laborious method
High throughput sequencing	- availability of a high quality reference genome or a large set of resequenced genomes of related accessions - inaccuracies related to short reads provided by the Illumina technology (problems with longer TEs, such as LTR-RTs; insertions in repetitive regions)	- use technologies producing long reads, e.g. Oxford Nanopore

pathways of actively transposing TEs. However, the interpretation of results produced by the above-presented methods can be difficult owing to the inherent properties of TEs. This review seeks to present techniques that can be used to obtain information about mobilized TEs and some pitfalls associated with the interpretation of results. The methods were divided on the basis of the context of their use with respect to the process of transposition.

Apparently, the use of some of the older methods mentioned above can be expedient in some specific cases and can bring unique information at relatively low price and experimental demands. The most comprehensive results are seemingly achievable by the methods based on massive parallel sequencing, however, they have also their limits. One such limitation is the fact that the created evaluation tools detect only a limited part of TEs. Related to this is also the need for thorough genomic TE annotation as an important prerequisite for appropriate detection of new copies. Some of shortcomings in the accuracy in bioinformatics data interpretation can be significantly improved by NGS techniques producing long reads. Generally, the strengths of one method are usually offset by other shortcomings. To obtain a comprehensive picture, a combination of methods based on different principles, seems to be the most effective. One of such examples is a strategy combining RNA-seq and MS, for which the designation Proteomics Informed by Transcriptomics is used. From the principle of the matter, a combination of methods targeting molecules originating from the final stages of the transposition

process of actively transposing TEs seems to be the most suitable. Namely, it means to focus on methods aimed at detecting novel insertion sites, eclDNA and eccDNA. From this perspective, coupling WGS and analysis of the intermediates or by-signals of actively transposing TEs, such as eccDNA, ALE-Seq or multi-genomic comparisons, seems to be a promising approach to reveal complete information regarding TEs activity and their impact on host genome.

Author contributions

MBaj wrote first draft of the manuscript and perform graphical support. AP wrote sections of the manuscript that referred to current bioinformatics tools. DG contributed to conception, compiled and revised author contributions. MBar established conception and design, wrote some parts of manuscript and revised author contributions. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by Internal Grant of Mendel University (IGA-ZF/2021-SI1007) and by the project CZ.02.2.69/0.0/0.0/16_018/0002333 Research Infrastructure for

Young Scientists, this is co-financed from Operational Programme Research, Development and Education.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Anderson, S. N., Stitzer, M. C., Brohammer, A. B., Zhou, P., Noshay, J. M., O'Connor, C. H., et al. (2019). Transposable elements contribute to dynamic genome content in maize. *Plant J.* 100 (5), 1052–1065. doi: 10.1111/tj.14489
- Baillie, J. K., Barnett, M. W., Upton, K. R., Gerhardt, D. J., Richmond, T. A., De Sapiro, F., et al. (2011). Somatic retrotransposition alters the genetic landscape of the human brain. *Nature* 479 (7374), 534–537. doi: 10.1038/nature10531
- Benoit, M., Drost, H.-G., Catoni, M., Gouli, Q., Lopez-Gomollon, S., Baulcombe, D., et al. (2019). Environmental and epigenetic regulation of rider retrotransposons in tomato. *PLoS Genet.* 15 (9), e1008370. doi: 10.1371/journal.pgen.1008370
- Böhrer, M., Rymen, B., Himber, C., Gerbaud, A., Pflieger, D., Laudencia-Chingcuanco, D., et al. (2020). "Integrated genome-scale analysis and northern blot detection of retrotransposon siRNAs across plant species," in *RNA Tagging* (Humana, New York, NY, Springer), 387–411.
- Bortiri, E., Jackson, D., and Hake, S. (2006). Advances in maize genomics: the emergence of positional cloning. *Curr. Opin. Plant Biol.* 9 (2), 164–171. doi: 10.1016/j.pbi.2006.01.006
- Cao, X., Wang, S., Ge, L., Zhang, W., Huang, J., and Sun, W. (2021). Extrachromosomal circular DNA: Category, biogenesis, recognition, and functions. *Front. Vet. Sci.* 8. doi: 10.3389/fvets.2021.693641
- Carpentier, M.-C., Manfroi, E., Wei, F.-J., Wu, H.-P., Lasserre, E., Llauro, C., et al. (2019). Retrotranspositional landscape of Asian rice revealed by 3000 genomes. *Nat. Commun.* 10 (1), 1–12. doi: 10.1038/s41467-018-07974-5
- Casa, A. M., Brouwer, C., Nagel, A., Wang, L., Zhang, Q., Kresovich, S., et al. (2000). The MITE family heartbreaker (Hbr): molecular markers in maize. *Proc. Natl. Acad. Sci.* 97 (18), 10083–10089. doi: 10.1073/pnas.97.18.10083
- Castanera, R., Vendrell-Mir, P., Bardil, A., Carpentier, M. C., Panaud, O., and Casacuberta, J. M. (2021). Amplification dynamics of miniature inverted-repeat transposable elements and their impact on rice trait variability. *Plant J.* 107 (1), 118–135. doi: 10.1111/tj.15277
- Chen, J., Lu, L., Benjamin, J., Diaz, S., Hancock, C. N., Stajich, J. E., et al. (2019). Tracking the origin of two genetic components associated with transposable element bursts in domesticated rice. *Nat. Commun.* 10 (1), 1–10. doi: 10.1038/s41467-019-08451-3
- Chen, J., Lu, L., Robb, S. M., Collin, M., Okumoto, Y., Stajich, J. E., et al. (2020). Genomic diversity generated by a transposable element burst in a rice recombinant inbred population. *Proc. Natl. Acad. Sci.* 117 (42), 26288–26297. doi: 10.1073/pnas.2015736117
- Chen, J., Wrightsman, T. R., Wessler, S. R., and Stajich, J. E. (2017). RelocaTE2: a high resolution transposable element insertion site mapping tool for population resequencing. *PeerJ* 5, e2942. doi: 10.7717/peerj.2942
- Cho, J., Benoit, M., Catoni, M., Drost, H.-G., Brestovitsky, A., Oosterbeek, M., et al. (2019). Sensitive detection of pre-integration intermediates of long terminal repeat retrotransposons in crop plants. *Nat. Plants* 5 (1), 26–33. doi: 10.1038/s41477-018-0320-9
- Chuong, E. B., Elde, N. C., and Feschotte, C. (2017). Regulatory activities of transposable elements: from conflicts to benefits. *Nat. Rev. Genet.* 18 (2), 71–86. doi: 10.1038/nrg.2016.139
- Davidson, A. D., Matthews, D. A., and Maringer, K. (2017). Proteomics technique opens new frontiers in mobile research. *Mob. Genet. Elements* 7 (4), 1–9. doi: 10.1080/2159256X.2017.1362494
- Debladis, E., Llauro, C., Carpentier, M.-C., Mirouze, M., and Panaud, O. (2017). Detection of active transposable elements in arabidopsis thaliana using Oxford nanopore sequencing technology. *BMC Genomics* 18 (1), 1–8. doi: 10.1186/s12864-017-3753-z
- Deininger, P., Morales, M. E., White, T. B., Baddoo, M., Hedges, D. J., Servant, G., et al. (2017). A comprehensive approach to expression of L1 loci. *Nucleic Acids Res.* 45 (5), e31. doi: 10.1093/nar/gkw1067
- Deneweth, J., Van de Peer, Y., and Vermeirssen, V. (2022). Nearby transposable elements impact plant stress gene regulatory networks: A meta-analysis in a thaliana and s. lycopersicum. *BMC Genomics* 23 (1), 1–19. doi: 10.1186/s12864-021-08215-8
- Deniz, Ö., Frost, J. M., and Branco, M. R. (2019). Regulation of transposable elements by DNA modifications. *Nat. Rev. Genet.* 20 (7), 417–431. doi: 10.1038/s41576-019-0106-6
- Dominguez, M., Dugas, E., Benchouaia, M., Leduque, B., Jiménez-Gómez, J. M., Colot, V., et al. (2020). The impact of transposable elements on tomato diversity. *Nat. Commun.* 11 (1), 1–11. doi: 10.1038/s41467-020-17874-2
- Ellison, C. E., and Cao, W. (2020). Nanopore sequencing and Hi-c scaffolding provide insight into the evolutionary dynamics of transposable elements and piRNA production in wild strains of drosophila melanogaster. *Nucleic Acids Res.* 48 (1), 290–303. doi: 10.1093/nar/gkz1080
- Esposito, S., Barteri, F., Casacuberta, J., Mirouze, M., Carputo, D., and Aversano, R. (2019). LTR-TEs abundance, timing and mobility in solanum commersonii and s. tuberosum genomes following cold-stress conditions. *Planta* 250 (5), 1781–1787. doi: 10.1007/s00425-019-03283-3
- Ewing, A. D., Smits, N., Sanchez-Luque, F. J., Faivre, J., Brennan, P. M., Richardson, S. R., et al. (2020). Nanopore sequencing enables comprehensive transposable element epigenomic profiling. *Mol. Cell* 80 (5), 915–928.e915. doi: 10.1016/j.molcel.2020.10.024
- Fan, W., Wang, L., Chu, J., Li, H., Kim, E. Y., and Cho, J. (2022). Tracing mobile DNAs: From molecular to population scales. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.837378
- Feschotte, C., Jiang, N., and Wessler, S. R. (2002). Plant transposable elements: where genetics meets genomics. *Nat. Rev. Genet.* 3 (5), 329–341. doi: 10.1038/nrg793
- Feschotte, C., and Pritham, E. J. (2007). DNA Transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* 41, 331–368. doi: 10.1146/annurev.genet.40.110405.090448
- Fultz, D., Choudury, S. G., and Slotkin, R. K. (2015). Silencing of active transposable elements in plants. *Curr. Opin. Plant Biol.* 27, 67–76. doi: 10.1016/j.pbi.2015.05.027
- Gantuz, M., Morales, A., Bertoldi, M. V., Ibañez, V. N., Duarte, P. F., Marfil, C. F., et al. (2022). Hybridization and polyploidization effects on LTR-retrotransposon activation in potato genome. *J. Plant Res.* 135 (1), 81–92. doi: 10.1007/s10265-021-01354-9
- Gill, R. A., Scossa, F., King, G. J., Golic, A., Tong, C. B., Snowdon, R. J., et al. (2021). On the role of transposable elements in the regulation of gene expression and subgenomic interactions in crop genomes. *Crit. Rev. Plant Sci.* 40 (2), 157–189. doi: 10.1080/07352689.2021.1920731
- Grandbastien, M. A. (2015). LTR Retrotransposons, handy hitchhikers of plant regulation and stress response. *Biochim. Biophys. Acta* 1849 (4), 403–416. doi: 10.1016/j.bbagr.2014.07.017
- Griffiths, J., Catoni, M., Iwasaki, M., and Paszkowski, J. (2018). Sequence-independent identification of active LTR retrotransposons in arabidopsis. *Mol. Plant* 11 (3), 508–511. doi: 10.1016/j.molp.2017.10.012
- Grzebelus, D., Jagosz, B., and Simon, P. W. (2007). The DcMaster transposon display maps polymorphic insertion sites in the carrot (*Daucus carota* L.) genome. *Gene* 390 (1–2), 67–74. doi: 10.1016/j.gene.2006.07.041

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Gürkök, T. (2017). Transcriptome-wide identification and expression analysis of brachypodium distachyon transposons in response to viral infection. *Turkish J. Agriculture-Food Sci. Technol.* 5 (10), 1156–1160. doi: 10.24925/turjaf.v5i10.1156-1160.1260
- Havecker, E. R., Gao, X., and Voytas, D. F. (2004). The diversity of LTR retrotransposons. *Genome Biol.* 5 (6), 1–6. doi: 10.1186/gb-2004-5-6-225
- Hirochika, H., and Otsuki, H. (1995). Extrachromosomal circular forms of the tobacco retrotransposon ttol. *Gene* 165 (2), 229–232. doi: 10.1016/0378-1119(95)00581-P
- Hollister, J. D., Smith, L. M., Guo, Y.-L., Ott, F., Weigel, D., and Gaut, B. S. (2011). Transposable elements and small RNAs contribute to gene expression divergence between arabidopsis thaliana and arabidopsis lyrata. *Proc. Natl. Acad. Sci.* 108 (6), 2322–2327. doi: 10.1073/pnas.1018222108
- Jiang, S., Cai, D., Sun, Y., and Teng, Y. (2016). Isolation and characterization of putative functional long terminal repeat retrotransposons in the pyrus genome. *Mob DNA* 7 (1), 1. doi: 10.1186/s13100-016-0058-8
- Jiménez-Ruiz, J., Ramírez-Tejero, J. A., Fernández-Pozo, N., Leyva-Pérez, M. D. L. O., Yan, H., Rosa, R. D. L., et al. (2020). Transposon activation is a major driver in the genome evolution of cultivated olive trees (*Olea europaea* L.). *Plant Genome* 13 (1), e20010. doi: 10.1002/tpg2.20010
- Kalendar, R., Antonius, K., Smýkal, P., and Schulman, A. H. (2010). iPBS: a universal method for DNA fingerprinting and retrotransposon isolation. *Theor. Appl. Genet.* 121 (8), 1419–1430. doi: 10.1007/s00122-010-1398-2
- Kalendar, R., and Schulman, A. H. (2006). IRAP and REMAP for retrotransposon-based genotyping and fingerprinting. *Nat. Protoc.* 1 (5), 2478–2484. doi: 10.1038/nprot.2006.377
- Kalendar, R., Shustov, A. V., and Schulman, A. H. (2021). Palindromic sequence-targeted (PST) PCR, version 2: an advanced method for high-throughput targeted gene characterization and transposon display. *Front. Plant Sci.* 12, 691940. doi: 10.3389/fpls.2021.691940
- Kim, N. S. (2017). The genomes and transposable elements in plants: are they friends or foes? *Genes Genomics* 39 (4), 359–370. doi: 10.1007/s13258-017-0522-y
- Kirov, I., Merkulov, P., Dudnikov, M., Polkhovskaya, E., Komakhin, R. A., Konstantinov, Z., et al. (2021). Transposons hidden in arabidopsis thaliana genome assembly gaps and mobilization of non-autonomous LTR retrotransposons unravelled by nanotie pipeline. *Plants* 10 (12), 2681. doi: 10.3390/plants10122681
- Kirov, I., Omarov, M., Merkulov, P., Dudnikov, M., Gvaramiya, S., Kolganova, E., et al. (2020). Genomic and transcriptomic survey provides new insight into the organization and transposition activity of highly expressed LTR retrotransposons of sunflower (*Helianthus annuus* L.). *Int. J. Mol. Sci.* 21 (23), 9331. doi: 10.3390/ijms21239331
- Kofler, R., Gómez-Sánchez, D., and Schlötterer, C. (2016). PoPoolationTE2: comparative population genomics of transposable elements using pool-seq. *Mol. Biol. Evol.* 33 (10), 2759–2764. doi: 10.1093/molbev/msw137
- Kumar, P., Kiran, S., Saha, S., Su, Z., Paulsen, T., Chatrath, A., et al. (2020). ATAC-seq identifies thousands of extrachromosomal circular DNA in cancer and cell lines. *Sci. Adv.* 6 (20), eaba2489. doi: 10.1126/sciadv.aba24
- Kwolek, K., Kędzierska, P., Hankiewicz, M., Mirouze, M., Panaud, O., Grzebelus, D., et al. (2022). Diverse and mobile-eccDNA-based identification of carrot low-copy LTR retrotransposons active in callus cultures. *Plant J* 110, 1811–1828. doi: 10.1111/tpj.15773
- Lanciano, S., Carpentier, M. C., Llauro, C., Jobet, E., Robakowska-Hyzorek, D., Lasserre, E., et al. (2017). Sequencing the extrachromosomal circular mobilome reveals retrotransposon activity in plants. *PLoS Genet.* 13 (2), e1006630. doi: 10.1371/journal.pgen.1006630
- Lanciano, S., and Cristofari, G. (2020). Measuring and interpreting transposable element expression. *Nat. Rev. Genet.* 21 (12), 721–736. doi: 10.1038/s41576-020-0251-y
- Lanciano, S., Zhang, P., Llauro, C., and Mirouze, M. (2021). “Identification of extrachromosomal circular forms of active transposable elements using mobilome-seq” in *Plant transposable elements* (Humana, New York, NY, Springer), 87–93.
- Liang, X., Hou, X., Li, J., Han, Y., Zhang, Y., Feng, N., et al. (2019). High-resolution DNA methylome reveals that demethylation enhances adaptability to continuous cropping comprehensive stress in soybean. *BMC Plant Biol.* 19 (1), 1–17. doi: 10.1186/s12870-019-1670-9
- Li, B., Ruotti, V., Stewart, R. M., Thomson, J. A., and Dewey, C. N. (2010). RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics* 26 (4), 493–500. doi: 10.1093/bioinformatics/btp692
- Louis, B., Waikhom, S. D., Goyari, S., Jose, R. C., Roy, P., and Talukdar, N. C. (2015). First proteome study of sporadic flowering in bamboo species (*Bambusa vulgaris* and *dendrocalamus manipureanus*) reveal the boom is associated with stress and mobile genetic elements. *Gene* 574 (2), 255–264. doi: 10.1016/j.gene.2015.08.010
- Lyu, M., Liu, H., Waititu, J. K., Sun, Y., Wang, H., Fu, J., et al. (2021). TEaseq-based identification of 35,696 dissociation insertional mutations facilitates functional genomic studies in maize. *J. Genet. Genomics* 48 (11), 961–971. doi: 10.1016/j.jgg.2021.07.010
- Møller, H. D., Parsons, L., Jørgensen, T. S., Botstein, D., and Regenberg, B. (2015). Extrachromosomal circular DNA is common in yeast. *Proc. Natl. Acad. Sci.* 112 (24), E3114–E3122. doi: 10.1073/pnas.150882511
- Makarevitch, I., Waters, A. J., West, P. T., Stitzer, M., Hirsch, C. N., Ross-Ibarra, J., et al. (2015). Transposable elements contribute to activation of maize genes in response to abiotic stress. *PLoS Genet.* 11 (1), e1004915. doi: 10.1371/journal.pgen.1004915
- Manninen, I., and Schulman, A. H. (1993). BARE-1, a copia-like retroelement in barley (*Hordeum vulgare* L.). *Plant Mol. Biol.* 22 (5), 829–846. doi: 10.1007/BF00027369
- Mann, L., Seibt, K. M., Weber, B., and Heitkam, T. (2022). ECCsplorer: a pipeline to detect extrachromosomal circular DNA (eccDNA) from next-generation sequencing data. *BMC Bioinf.* 23 (1), 1–15. doi: 10.1186/s12859-021-04545-2
- Marcon, H. S., Domingues, D. S., Silva, J. C., Borges, R. J., Matioli, F. F., Fontes, M. R., et al. (2015). Transcriptionally active LTR retrotransposons in eucalyptus genus are differentially expressed and insertional polymorphic. *BMC Plant Biol.* 15 (1), 198. doi: 10.1186/s12870-015-0550-1
- Maringer, K., Yousuf, A., Heesom, K. J., Fan, J., Lee, D., Fernandez-Sesma, A., et al. (2017). Proteomics informed by transcriptomics for characterising active transposable elements and genome annotation in aedes aegypti. *BMC Genomics* 18 (1), 101. doi: 10.1186/s12864-016-3432-5
- McClintock, B. (1950). The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. U.S.A.* 36 (6), 344–355. doi: 10.1073/pnas.36.6.344
- McDonald, T. L., Zhou, W., Castro, C. P., Mumm, C., Switzenberg, J. A., Mills, R. E., et al. (2021). Cas9 targeted enrichment of mobile elements using nanopore sequencing. *Nat. Commun.* 12 (1), 1–13. doi: 10.1038/s41467-021-23918-y
- Mehta, D., Cornet, L., Hirsch-Hoffmann, M., Zaidi, S. S. A., and Vanderschuren, H. (2020). Full-length sequencing of circular DNA viruses and extra-chromosomal circular DNA using CIDER-seq. *Nat. Protoc.* 15, 1673–1689. doi: 10.1038/s41596-020-0301-0
- Mehta, D., Hirsch-Hoffmann, M., Were, M., Patrignani, A., Zaidi, S.S.E.A., and Were, H. (2019). A new full-length circular DNA sequencing method for viral-sized genomes reveals that RNAi transgenic plants provoke a shift in geminivirus populations in the field. *Nucleic Acids Res.* 47 (2), e9. doi: 10.1093/nar/gky914
- Meyer, C., Pouteau, S., Rouze, P., and Caboche, M. (1994). Isolation and molecular characterization of dTnp1, a mobile and defective transposable element of nicotiana plumbaginifolia. *Mol. Gen. Genet.* 242 (2), 194–200. doi: 10.1007/BF00391013
- Monden, Y., Fujii, N., Yamaguchi, K., Ikee, K., Nakazawa, Y., Waki, T., et al. (2014). Efficient screening of long terminal repeat retrotransposons that show high insertion polymorphism via high-throughput sequencing of the primer binding site. *Genome* 57 (5), 245–252. doi: 10.1139/gen-2014-0031
- Morillon, A., Benard, L., Springer, M., and Lesage, P. (2002). Differential effects of chromatin and Gcn4 on the 50-fold range of expression among individual yeast Ty1 retrotransposons. *Mol. Cell. Biol.* 22 (7), 2078–2088. doi: 10.1128/Mcb.22.7.2078-2088.2002
- O'Neill, K., Brocks, D., and Hammell, M. G. (2020). Mobile genomics: tools and techniques for tackling transposons. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 375 (1795), 20190345. doi: 10.1098/rstb.2019.0345
- Oberlin, S., Sarazin, A., Chevalier, C., Voinnet, O., and Mari-Ordóñez, A. (2017). A genome-wide transcriptome and translome analysis of arabidopsis transposons identifies a unique and conserved genome expression strategy for Ty1/Copia retroelements. *Genome Res.* 27 (9), 1549–1562. doi: 10.1101/gr.220723.117
- Pandey, G., Yadav, C. B., Sahu, P. P., Muthamilarasan, M., and Prasad, M. (2017). Salinity induced differential methylation patterns in contrasting cultivars of foxtail millet (*Setaria italica* L.). *Plant Cell Rep.* 36 (5), 759–772. doi: 10.1007/s00299-016-2093-9
- Paz, R. C., Rendina Gonzalez, A. P., Ferrer, M. S., and Masuelli, R. W. (2015). Short-term hybridisation activates Tnt1 and Tto1 copia retrotransposons in wild tuber-bearing solanum species. *Plant Biol. (Stuttg)* 17 (4), 860–869. doi: 10.1111/plb.12301
- Picault, N., Chaparro, C., Piegu, B., Stenger, W., Formey, D., Llauro, C., et al. (2009). Identification of an active LTR retrotransposon in rice. *Plant J.* 58 (5), 754–765. doi: 10.1111/j.1365-313X.2009.03813.x
- Piriyaopongsa, J., and Jordan, I. K. (2008). Dual coding of siRNAs and miRNAs by plant transposable elements. *RNA* 14 (5), 814–821. doi: 10.1261/rna.916708
- Pozueta-Romero, J., Klein, M., Houlne, G., Schantz, M. L., Meyer, B., and Schantz, R. (1995). Characterization of a family of genes encoding a fruit-specific wound-stimulated protein of bell pepper (*Capsicum annuum*): identification of a new family of transposable elements. *Plant Mol. Biol.* 28 (6), 1011–1025. doi: 10.1007/BF00032663

- Prada-Luengo, I., Krogh, A., Maretty, L., and Regenberg, B. (2019). Sensitive detection of circular DNAs at single-nucleotide resolution using guided realignment of partially aligned reads. *BMC Bioinf.* 20 (1), 1–9. doi: 10.1186/s12859-019-3160-3
- Qiu, F., and Ungerer, M. C. (2018). Genomic abundance and transcriptional activity of diverse gypsy and copia long terminal repeat retrotransposons in three wild sunflower species. *BMC Plant Biol.* 18 (1), 6. doi: 10.1186/s12870-017-1223-z
- Quadrana, L., Silveira, A. B., Mayhew, G. F., LeBlanc, C., Martienssen, R. A., Jeddeloh, J. A., et al. (2016). The arabidopsis thaliana mobilome and its impact at the species level. *eLife* 5, e15716. doi: 10.7554/eLife.15716.046
- Ravindran, S. (2012). Barbara McClintock and the discovery of jumping genes. *Proc. Natl. Acad. Sci. U.S.A.* 109 (50), 20198–20199. doi: 10.1073/pnas.1219372109
- Rocheta, M., Coito, J. L., Ramos, M. J. N., Carvalho, L., Becker, J. D., Carbonell-Bejerano, P., et al. (2016). Transcriptomic comparison between two vitis vinifera l. varieties (Trincadeira and touriga nacional) in abiotic stress conditions. *BMC Plant Biol.* 16 (1), 1–19. doi: 10.1186/s12870-016-0911-4
- Rodriguez-Terrones, D., and Torres-Padilla, M. E. (2018). Nimble and ready to mingle: Transposon outbursts of early development. *Trends Genet.* 34 (10), 806–820. doi: 10.1016/j.tig.2018.06.006
- Roquis, D., Robertson, M., Yu, L., Thieme, M., Julkowska, M., and Bucher, E. (2021). Genomic impact of stress-induced transposable element mobility in arabidopsis. *Nucleic Acids Res.* 49 (18), 10431–10447. doi: 10.1093/nar/gkab828
- Sahebi, M., Hanafi, M. M., van Wijnen, A. J., Rice, D., Rafii, M. Y., Azizi, P., et al. (2018). Contribution of transposable elements in the plant's genome. *Gene* 665, 155–166. doi: 10.1016/j.gene.2018.04.050
- Satheesh, V., Fan, W., Chu, J., and Cho, J. (2021). Recent advancement of NGS technologies to detect active transposable elements in plants. *Genes Genomics* 43 (3), 289–294. doi: 10.1007/s13258-021-01040-z
- Serrato-Capuchina, A., and Matute, D. R. (2018). The role of transposable elements in speciation. *Genes* 9 (5), 254. doi: 10.3390/genes9050254
- Sexton, C. E., and Han, M. V. (2019). Paired-end mappability of transposable elements in the human genome. *Mob DNA* 10 (1), 29. doi: 10.1186/s13100-019-0172-5
- Shahid, S., and Slotkin, R. K. (2020). The current revolution in transposable element biology enabled by long reads. *Curr. Opin. Plant Biol.* 54, 49–56. doi: 10.1016/j.pbi.2019.12.012
- Sigman, M. J., and Slotkin, R. K. (2016). The first rule of plant transposable element silencing: location, location, location. *Plant Cell* 28 (2), 304–313. doi: 10.1105/tpc.15.00869
- Song, R., Wang, Z., Wang, H., Zhang, H., Wang, X., Nguyen, H., et al. (2021). InMut-finder: a software tool for insertion identification in mutagenesis using nanopore long reads. *BMC Genomics* 22 (1), 1–7. doi: 10.1186/s12864-021-08206-9
- Sotero-Caio, C. G., Platt, R. N. 2nd, Suh, A., and Ray, D. A. (2017). Evolution and diversity of transposable elements in vertebrate genomes. *Genome Biol. Evol.* 9 (1), 161–177. doi: 10.1093/gbe/evw264
- Sow, M. D., Le Gac, A. L., Fichot, R., Lanciano, S., Delaunay, A., Le Jan, I., et al. (2021). RNAi suppression of DNA methylation affects the drought stress response and genome integrity in transgenic poplar. *New Phytol.* 232 (1), 80–97. doi: 10.1111/nph.17555
- Sundaresan, V., and Freeling, M. (1987). An extrachromosomal form of the mu transposons of maize. *Proc. Natl. Acad. Sci.* 84 (14), 4924–4928. doi: 10.1073/pnas.84.14.4924
- Sun, T., Wang, K., Liu, C., Wang, Y., Wang, J., and Li, P. (2019). Identification of extrachromosomal linear microDNAs interacted with microRNAs in the cell nuclei. *Cells* 8 (2), 111. doi: 10.3390/cells8020111
- Tahara, M., Aoki, T., Suzuka, S., Yamashita, H., Tanaka, M., Matsunaga, S., et al. (2004). Isolation of an active element from a high-copy-number family of retrotransposons in the sweetpotato genome. *Mol. Genet. Genomics* 272 (1), 116–127. doi: 10.1007/s00438-004-1044-2
- Tang, Y., Yan, X., Gu, C., and Yuan, X. (2022). Biogenesis, trafficking, and function of small RNAs in plants. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.825477
- Torres, A. R., Rodrigues, E. P., Batista, J. S., Gomes, D. F., and Hungria, M. (2013). Proteomic analysis of soybean [Glycine max (L.) Merrill] roots inoculated with bradyrhizobium japonicum strain CPAC 15. *Proteomics Insights* 6, 7–11. doi: 10.4137/PRI.S13288
- Usai, G., Mascagni, F., Vangelisti, A., Giordani, T., Ceccarelli, M., Cavallini, A., et al. (2020). Interspecific hybridisation and LTR-retrotransposon mobilisation-related structural variation in plants: A case study. *Genomics* 112 (2), 1611–1621. doi: 10.1016/j.ygeno.2019.09.010
- Valdebenito-Maturana, B., and Riadi, G. (2018). TEcandidates: Prediction of genomic origin of expressed transposable elements using RNA-seq data. *Bioinformatics* 34 (22), 3915–3916. doi: 10.1093/bioinformatics/bty423
- Vangelisti, A., Mascagni, F., Giordani, T., Sbrana, C., Turrini, A., Cavallini, A., et al. (2019). Arbuscular mycorrhizal fungi induce the expression of specific retrotransposons in roots of sunflower (Helianthus annuus L.). *PLoS One* 14 (2), e0212371. doi: 10.1371/journal.pone.0212371
- Vendrell-Mir, P., Barteri, F., Merenciano, M., González, J., Casacuberta, J. M., and Castanera, R. (2019). A benchmark of transposon insertion detection tools using real data. *Mobile DNA* 10 (1), 1–19. doi: 10.1186/s13100-019-0197-9
- Voronova, A. (2019). Retrotransposon expression in response to *in vitro* inoculation with two fungal pathogens of scots pine (Pinus sylvestris L.). *BMC Res. Notes* 12 (1), 243. doi: 10.1186/s13104-019-4275-3
- Vos, P., Hogers, R., Bleeker, M., Reijans, M., Lee, T., Hornes, M., et al. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* 23 (21), 4407–4414. doi: 10.1093/nar/23.21.4407
- Vuong, L. M., Pan, S., and Donovan, P. J. (2019). Proteome profile of endogenous retrotransposon-associated complexes in human embryonic stem cells. *Proteomics* 19 (15), e1900169. doi: 10.1002/pmic.201900169
- Wanchai, V., Jenjaroenpun, P., Leangapichart, T., Arrey, G., Burnham, C. M., Tümmeler, M. C., et al. (2022). CRE-SIL: Accurate identification of extrachromosomal circular DNA from long-read sequences. *Brief. Bioinformatics* 23 (6), bbac422. doi: 10.1093/bib/bbac422
- Wang, K., Tian, H., Wang, L., Wang, L., Tan, Y., Zhang, Z., et al. (2021). Deciphering extrachromosomal circular DNA in arabidopsis. *Comput. Struct. Biotechnol. J.* 19, 1176–1183. doi: 10.1016/j.csbj.2021.01.043
- Wang, X., Yang, P., Gao, Q., Liu, X., Kuang, T., Shen, S., et al. (2008). Proteomic analysis of the response to high-salinity stress in physcomitrella patens. *Planta* 228 (1), 167–177. doi: 10.1007/s00425-008-0727-z
- Waugh, R., McLean, K., Flavell, A., Pearce, S., Kumar, A., Thomas, B., et al. (1997). Genetic distribution of bare-1-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). *Mol. Gen. Genet.* MGG 253 (6), 687–694. doi: 10.1007/s004380050372
- Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., et al. (2007). A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 8 (12), 973–982. doi: 10.1038/nrg2165
- Woodrow, P., Pontecorvo, G., Fantaccione, S., Fuggi, A., Kafantaris, I., Parisi, D., et al. (2010). Polymorphism of a new Ty1-copia retrotransposon in durum wheat under salt and light stresses. *Theor. Appl. Genet.* 121 (2), 311–322. doi: 10.1007/s00122-010-1311-z
- Yamashita, H., and Tahara, M. (2006). A LINE-type retrotransposon active in meristem stem cells causes heritable transpositions in the sweet potato genome. *Plant Mol. Biol.* 61 (1), 79–84. doi: 10.1007/s11103-005-6002-9
- Zhang, P., Peng, H., Llauro, C., Bucher, E., and Mirouze, M. (2021). Ecc_finder: A robust and accurate tool for detecting extrachromosomal circular DNA from sequencing data. *Front. Plant Sci.* 12, 743742. doi: 10.3389/fpls.2021.743742
- Zhou, W., Emery, S. B., Flasch, D. A., Wang, Y., Kwan, K. Y., Kidd, J. M., et al. (2020). Identification and characterization of occult human-specific LINE-1 insertions using long-read sequencing technology. *Nucleic Acids Res.* 48 (3), 1146–1163. doi: 10.1093/nar/gkz1173