Chapter 23

Transposon-Based Tagging In Silico Using FastPCR Software

Ruslan Kalendar, Dana Kospanova, and Alan H. Schulman

Abstract

Retrotransposons are ubiquitous, generally dispersed components of eukaryotic genomes. These properties, together with their "copy and paste" lifecycle that generates insertional polymorphism without need for excision, makes them widely useful as a molecular-genetic tags. Various tagging systems have been developed that exploit the sequence conservation of retrotransposon components, such as those found in their long terminal repeats (LTRs). To detect polymorphisms for retrotransposon insertions, marker systems generally rely on PCR amplification between the termini and some component of flanking genomic DNA. As complements to various "wet lab" protocols for retrotransposon tagging, in silico bioinformatics approaches are useful for predicting likely outcomes from unsequenced accessions on the basis of reference genomes. In this chapter, we describe protocols for in silico retrotransposon-based fingerprinting techniques using the FastPCR software as an integrated tools environment for in silico PCR primer design and analysis.

Key words Retrotransposon, Transposon tagging, DNA fingerprinting, Bioinformatic genome analysis, PCR primer design, Degenerate PCR

1 Introduction

Interspersed repetitive sequences comprise a large fraction of the genome of most eukaryotic organisms; they are predominantly comprised of transposable elements (TEs) [1]. Among the TEs, in both plants and animals, the retrotransposons, also called Type I transposable elements, tend to predominate both by number and by the share of the genome they occupy. These elements replicate by a "copy and paste" lifecycle that leaves the mother element in place in the genome while new copies are generated and propagated. The remaining TEs are either DNA transposons, also called Type II transposable elements, which replicate by a "cut-and-paste" cycle of excision and reinsertion, or helitrons, which replicate by a rolling-circle mechanism [1].

Jungnam Cho (ed.), *Plant Transposable Elements: Methods and Protocols*, Methods in Molecular Biology, vol. 2250, https://doi.org/10.1007/978-1-0716-1134-0_23,

In plant genomes, the Long Terminal Repeat (LTR) retrotransposons (RLX, by the system of [1]) predominate by number and share of the genome [2]. The RLX elements replicate and propagate by a cycle of transcription, reverse transcription into cDNA, and integration of the cDNA copies back into the genome [3]. The lifecycle matches the intracellular replication phase of lentiviruses such as HIV. The key proteins needed for replication—reverse transcriptase, RNAse H, integrase—as well as the capsid protein (Gag) that forms the virus-like particles are encoded by the RLXs themselves. These, together with the motifs in the LTRs required for transcription and integration and those in the body of the element that are needed for reverse transcription provide conserved domains that are suitable for identification of RLX families and for PCR methods as described below. In addition to autonomous copies encoding proteins needed for replication, there are abundant groups of non-autonomous elements that retain only the signals for transcription, reverse transcription, and integration [4].

Because RLX propagation is not linked to excision, genomes diversify by the insertion of new copies, but old copies persist. Large plant genomes contain hundreds of thousands of these elements, *e.g.*, the barley [5] or gymnosperm [6] genomes together forming the vast majority of the total DNA. Their abundance in the genome is generally highly correlated with genome size. Elements can be lost through intra-molecular recombination, often between the LTRs, which can counter growth of genome size [7, 8]. However, intra-element LTR: LTR recombination leaves the joints between RLXs and the flanking genomic DNA intact, and thus does not affect their use in fingerprinting methods.

Human and other mammalian genomes also contain an abundance of retrotransposons. The majority of these, however, are not LTR retrotransposons but LINEs and SINEs, which replicate by a different copy-and-paste mechanism called through target site-primed reverse transcription [9]. The L1 family of LINE elements and the *Alu* family of SINE elements comprise together roughly 30% of human genomic DNA in nearly 2 million copies [10]. In addition, integrated retroviruses, which are remnants of ancient infections, are also abundant in mammalian genomes. These elements, called "endogenous retroviruses" (ERVs, HERVs in humans), are functionally equivalent to LTR retrotransposons. The features of integration activity, persistence, dispersion, conserved structure and sequence motifs, and high copy number together suggest that retrotransposons are well-suited genomic features on which to build molecular marker systems.

Various applications have been developed to exploit polymorphisms in TE insertion patterns. These applications are generally built on the principle of carrying out PCR using one primer matching a conserved TE motif and the other matching variously a TE motif, a conserved non-TE genomic feature, or an adapter such

as for a restriction site [11]. For retrotransposons, perhaps the earliest marker method (1990) was Alu-PCR or SINE-PCR, which amplified between the end of Alu (SINE) elements and some other feature [12, 13]. For plant RLXs, the first published marker method, Sequence-Specific Amplified Polymorphism (S-SAP or SSAP) amplified between the LTR of an RLX and a restriction site adapter [14]. That was soon followed (1999) by Inter-Retrotransposon Amplified Polymorphism (IRAP) and Retrotransposon-Microsatellite Amplified Polymorphism (REMAP), which produce PCR products, respectively, between two RLX elements and between an RLX and a microsatellite domain, such as $(CA/GT)_n$ or $(CAC/GTG)_n$ [15]. Methods for RLXs generally use a primer near the LTR terminus and facing towards the joint with genomic DNA, due to the size of the elements (upwards of 8–9 kb). IRAP succeeds where pairs of elements are sufficiently close to generate PCR products; such sites are fairly frequent at least in plant genomes. However, the method can be made more general by mixing primers for RLXs and SINEs or other elements in the same reaction.

A universal method requiring no species-specific sequence information is the Inter Primer Binding Site polymorphism (iPBS), uses the conserved PBS sites found adjacent to virtually all RLX elements, and is akin in concept to IRAP [16, 17]. This and all foregoing methods are anonymous, in the sense that the identity and map position of marker bands are unknown. The Retrotransposon-Based Insertional Polymorphism (RBIP) method, in contrast, displays the presence or absence of a specific TE insertion based on known flanking sequences [18]. While RBIP is low-throughput, the Tagged Array Marker (TAM) adaptation makes it possible to do multiple loci simultaneously [19]. Overall, the TE marker methods have been relatively well received. As of the end of 2019, over 3200 papers have been published on the subject.

The development of the DNA sequencing technology and in particular the emergence of high-throughput ("next-generation") sequencing has led to progressive accumulation of huge amounts of raw genome sequence data, as well as the many assembled and annotated prokaryotic and eukaryotic genomes. For this reason, there is a growing demand for in silico approaches to extract useful information from sequence data, process the information using virtual tools, and predict experimental results already at the planning stage of wet-lab experiments with unsequenced biological accessions. One such approach is in silico PCR. The usual goal of in silico PCR is to predict which PCR products are synthesized from one or multiple DNA templates using already designed oligonucleotide primers [20, 21].

This chapter describes the FastPCR software as a solution for in silico transposon-based genome fingerprinting. Specific examples for the use of FastPCR to achieve high throughput in silico

PCR assays on large amounts of data will be shown. We also provide a quick start guide for immediate application of FastPCR for in silico investigation of genomes during planning of experiments and for prediction of expected results.

2 Materials

2.1 Software, General Information

2.1.1 Supported Platforms and Dependencies

The online version of the FastPCR software (https://primerdigital.com/tools/pcr.html) is written in Java with NetBeans IDE (Apache) and requires the Java Runtime Environment with OpenWebStart software (https://openwebstart.com/download/) installed on the computer. The program can be used with any operating system (a 64-bit OS is preferred for large chromosome files). The stand-alone version of the FastPCR software (https://primerdigital.com/fastpcr.html) can be used with Microsoft Windows.

2.1.2 Downloading and Installing

The online version of the FastPCR software requires the Java Runtime Environment (https://www.oracle.com/technetwork/java/javase/downloads/). Users need to add the URL (https://primerdigital.com/) of this application to the Exception Site List (https://www.java.com/en/download/faq/exception_sitelist.xml), which is located under the Security tab of the Java Control Panel (http://www.java.com/en/download/help/appsecuritydialogs.xml). Adding this application URL to the list will allow it to run after presenting some security warnings. Existence of the application URL in the Exception list allows users to run Rich Internet Applications (RIAs) that would normally be blocked by security checks. The exception site list is managed in the Security tab of the Java Control Panel (*see* **Note 1**). The list is shown in the tab. To add, edit or remove a URL from the list, use the following:

- Click on the Edit Site List button.
- Click the Add in the Exception Site List window.
- Click in the empty field under Location field to enter the URL: https://primerdigital.com/.
- Click OK to save the URL that you entered. If you click Cancel, the URLs will not be saved.
- Click Continue in the Security Warning dialog.

Running and downloading online FastPCR software to a desktop computer using the OpenWebStart software (https://openwebstart.com/download/). Users can run the software directly from the WEB site: https://primerdigital.com/tools/pcr.html. The program manual and files for installation are available on

the internet at https://primerdigital.com/fastpcr/. The YouTube tutorial videos are located at https://www.youtube.com/user/primerdigital.

2.2 The Interface

2.2.1 Inputs to FastPCR

The software contains the menus, the toolbars, the ribbon, and three text editors. The ribbon is designed to help users quickly find the commands that are needed to complete a task. Commands are organized into logical groups, which are displayed together under tabs (Fig. 1). Each tab relates to a type of task, such as "PCR Primer Design", "in silico PCR", and "Primer Test".

Getting started with a basic project in the FastPCR software is as easy as opening a new or existing file as well as a copy-paste or starting to type. There are three independent text editors at different tabs: "General Sequence(s)", "Additional sequence(s) or pre-designed primers (probes) list", and "Result report". The two first text editors are necessary for loading sequences for analysis. The "General Sequence(s)" text editor is designed for working

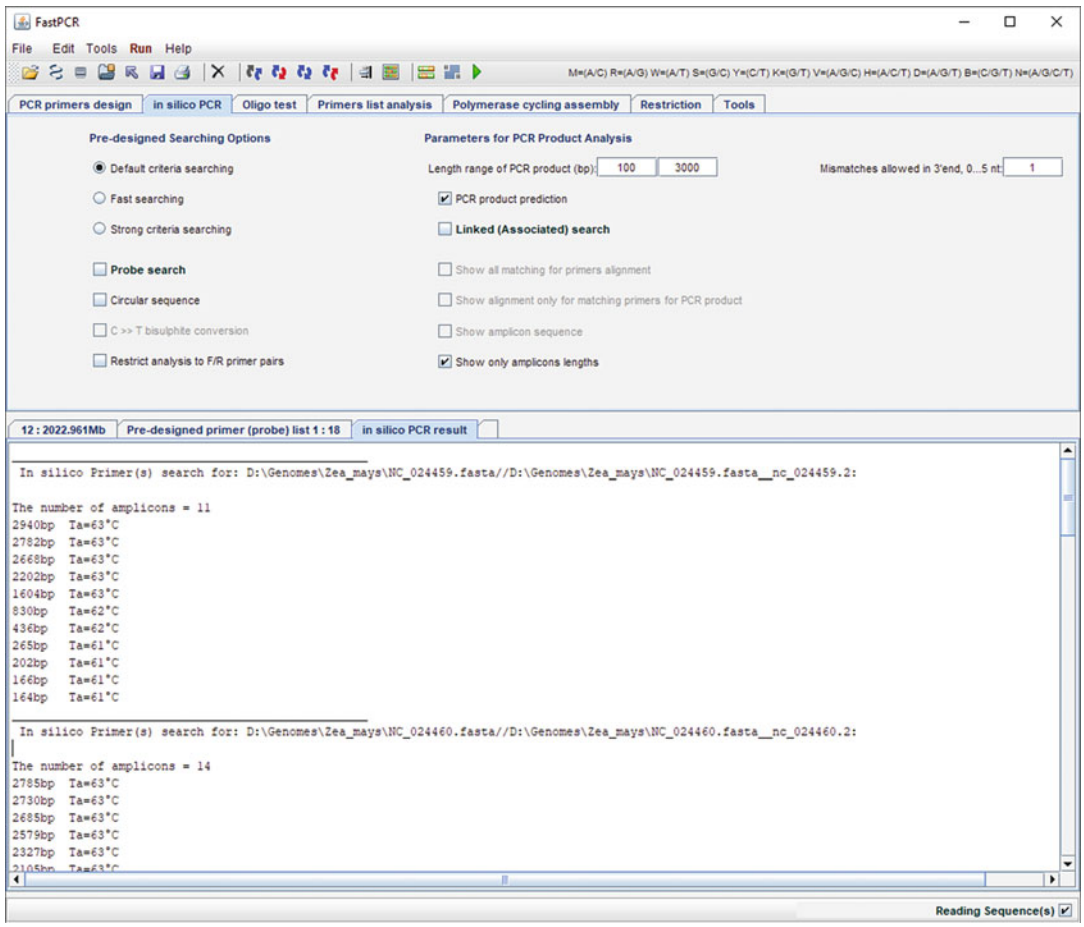

Fig. 1 Input for FastPCR, primer sequence editor, and user interface

with the project sequences; the "Additional sequence(s) or pre-designed primers (probes) list" text editor is appropriate for special and additional sequences, such as pre-designed primers, multiple query sequences, or for the numbers for input.

2.2.2 Sequence Entry

The program takes either a single sequence or accepts multiple separate DNA sequences in FASTA (http://blast.ncbi.nlm.nih.gov/blastcgihelp.shtml) or text file format (tab delimited). The FASTA format is preferred; it is simply the raw sequence proceeded by a definition line. The definition line begins with a ">" sign that can be optionally followed by a sequence name of any length and amount of words with no space in between. There can be many sequences listed in the same file. The format requires that a new sequence always starts with a new ">" symbol. Degenerate DNA sequences are accepted as IUPAC code that is an extended vocabulary of 11 letters that allows the description of ambiguous DNA code [22]. Each letter represents a combination of one or several nucleotides: M=(A/C), R=(A/G), W=(A/T), S=(G/C), Y=(C/T), K=(G/T), V=(A/G/C), H=(A/C/T), D=(A/G/T), B=(C/G/T), N=(A/G/C/T), U=T and I (Inosine) (*see* **Note 2**).

2.2.3 Program Output

The FastPCR software automatically generates results in the third text editor, which are named as "Result report". The results are produced in a tabulated format for transferring the output to a Microsoft Excel sheet from a clipboard using the copy-paste method or for saving them as .XLS or .RTF text files, compatible with both MS Excel and Open Office.

3 Methods

Stable hybridization of a primer to the template DNA is essential for primer extension by DNA polymerase. Mismatches affect the stability of the primer-template duplex and the efficiency with which the polymerase extends the primer. Although any mismatch affects PCR specificity, mismatches at the 3′-end of a primer have a profoundly detrimental effect on primer extension. A two-base mismatch at the 3′-end of the primer results in PCR failure. Therefore, our algorithm pays particular attention to the 3′-end of the primer and calculates its similarity to the template with a user-defined level of stringency.

The user can define the desired size of the PCR product. By default, the distance between the forward and reverse primers ranges from 50 to 5000 bases. PCR product size prediction is possible for linear or circular templates as modeled for standard, inverse, or multiplex PCR. The algorithm accepts single or multiple DNA sequences as primers or amplification targets.

Primer melting temperatures (T_m) for in silico PCR experiments with oligonucleotides that have mismatches to their targets are calculated using averaged nearest neighbor thermodynamic parameters [23–26]. The optimal PCR annealing temperature (T_a) is calculated as the value for the primer with the lowest T_m plus the natural logarithm of the PCR product length [21].

3.1 In Silico PCR Analysis Steps

Running in silico PCR analysis by FastPCR is a stepwise process. A user has the ability to set all the parameters required to run in silico PCR analysis interactively:

1. Select a task type on the ribbon—"in silico PCR".
2. Input user data
   (a) All target sequences to be used for the in silico PCR search should be pasted in the "sequences" TAB text area.
   (b) The entire the list of primer sequences should be pasted in the "additional sequence(s) or pre-designed primers (probes) list" TAB text area.
3. Choose an algorithm
   (a) By default—uses the standard algorithm to search for sites of stable primer binding within the target.
   (b) Probe search.
   (c) Linked search.
4. Set parameters
   (a) By default—"PCR product prediction" and "show all sites matching the primer."
   (b) For whole genome analysis—select "Show only amplicon lengths".
   (c) Change the "Length range of PCR product (bp)" to 100–3000.
5. Run.
6. Visualize results.

3.2 Additional Options Relating to Representation of the Analysis Results

The additional configuration settings allow optimization of the search for primer or probe binding sites and increases the representativeness of the results. This is determined mainly by the following parameters:

"Show all matching sites of primer binding"—checked by default, the software shows the result including all sites in which the primer would stably bind to the target. Although, given particular specified assay conditions, the primer combinations would not necessarily be able to produce a PCR product, the user can nevertheless examine the stability of primer binding sites, the

primer orientations, and the site coordinates in the target (*see* **Note 3**).

"Show alignment only for primers able to produce PCR product"—in the previous option, all primer binding sites were represented; here, the primer analysis will report back only those matching the target and able to produce a PCR product.

"Show only amplicon lengths"—checking this option allows the user to collect only amplicon lengths without analysis of the sites primers match the template. This option is recommended for use with in silico analysis of whole genomes or of full chromosomes, or analyses with highly abundant repeated sequences.

"Linked (associated) search"—programmable searching, when binding sites for primers or probes are searched within a determined distance.

3.3 FastPCR as a Robust Solution for In Silico PCR Assays

3.3.1 Example for Validation of the Software

We used the NCBI database of genome sequences (https://www.ncbi.nlm.nih.gov/genome/browse/) as a target set and retrotransposon primer and probe sequences as query sets. The use of subsets of retrotransposon sequences allowed us to assess how the software handles complex tasks such as in silico degenerate PCR and linked searching. The software is compatible with single-primer methods, such as SINE-PCR, IRAP, REMAP, ISSR, and RAPD, which are based on TEs and sequence repeats; it is also compatible with methods that utilize multiple primers, such as LAMP and multiplex PCR. It is likewise compatible with retrotransposon-based genotyping and DNA fingerprinting methods. Primer and probe lengths can be set to between 12 and 500 nucleotides. The maximum PCR product (amplicon) length is limited only by the polymerase processivity and not by the software. Analysis results include PCR product sequences and their lengths and T_a, and are presented separately for nuclear DNA, plastids, and mitochondrial DNA. Another application of the software is detection of plasmid sequences based on the ends of an integrated sequence. For this purpose, sequence fragments flanking an insertion should be used instead of primer sequences and the "probe search" option should be selected in the software.

3.3.2 In Silico PCR Application: An IRAP Example for the Maize Genome

To carry out an in silico IRAP analysis on the maize genome, the latest public chromosomal sequences (B73 RefGen_v4) were downloaded (https://www.ncbi.nlm.nih.gov/genome/12) to the same file folder. For in silico IRAP analysis, we used a single LTR primer corresponding to a sequence highly conserved in the RLXs examined with the default options. The length of potential PCR products varied from 100 to 3000 bp and allowed a single mismatch within the 3′-termini of the LTR primers.

The FastPCR software was used for in silico IRAP in the following way:

1. Launch the FastPCR software.
2. Open all files from folder from Menu-File: "working with all files in folder" (Ctrl-J). Automatically, the program will receive an indication of these files and determine the total number of files and the total size of all sequences.
3. The primer list of sequences should be pasted in the "additional sequence(s) or pre-designed primers (probes) list" in next TAB text area. When the target sequences or primers list is recognized by the program, the software will indicate features of these sequences, such as format, length, CG% content, T_m, and so on. Primer names can contain any character, but only one space. Furthermore, names of primers even can be identical.
4. Select the "in silico PCR" ribbon. Optionally, the user can specify search options, including stringency and PCR product detection options. Select "show only amplicon lengths" and change the "length range of PCR product (bp)" to 100–3000.
5. Execute the search task by clicking **F5** or **Run** in the drop-down menu.

3.3.3 Results Interpretation

Once the in silico PCR analysis is complete, the results will appear in the third Tab **Result** text editor as **In silico PCR Result**, when any target has been found from the designated genome or sequences. The results in the **In silico PCR Result** text editor reports the specificity of the primers (locations, including target position, similarity, and T_m) and a summary of primer pairs in relation to the PCR template, as well as detailed information on each primer pair, including its length and T_a. A description line of a primer begins with "In silico primer(s) search for": followed by target name and FASTA description of target genome sequence. The description line of a template begins with a ">" sign followed by its identification. In a query, the primer begins without a ">" sign, but includes its identification and original sequence. It found the actual targets of the primers will be listed along with detailed alignments between the primers and targets.

Features of the individual reports for the query primers include representations of their alignments with the target sequence. The actual target fragments will be listed along with detailed alignments with linked query sequences and detailed information on each query sequence including locations on target position and similarity to the primers. Alignment of query primers to their target template will be shown along with their starting and ending coordinates. Nucleotides on the template that perfectly match with the aligned query are embodied by a vertical bar and those mismatched

nucleotides are given as a colon (at least similarity 60%) or by a space. The products are grouped by the target template they are found in. One or multiple products will be displayed with product sizes and T_a, including both original primer sequences in FASTA format and the target position and orientation.

Depending on the task being run, the user can get full information about all potential primer binding sites within the DNA target and within the PCR template. For whole genome analysis, such full information can be redundant. In this case, the summary of each primer pair and the length of amplicons will be sufficient. Results can be saved using the "Save as" function in the menu **File** or can be copied and pasted into any text editor for further study.

4 Notes

1. In order to enhance security, the certificate revocation checking feature has been enabled by default (starting from Java 7). Before Java attempts to launch a signed application, the associated certificate will be validated to ensure that it has not been revoked by the issuing authority. This feature has been implemented using both Certificate Revocation Lists (CRLs) and Online Certificate Status Protocol (OCSP) mechanisms.

   Optionally, users can download self-signed certificates file (https://primerdigital.com/j/primerdigital.cer) and import it to "Signer CA" (Certificate Authority) from the Java Control Panel. Finally, users need to set "Security Level" to "High" under the Security tab of the Java Control Panel (as it is shown on the picture: https://primerdigital.com/image/primerdigital_certificate_big.png).

2. The user can type or import data from file(s) into the "General Sequence(s)", "Additional sequence(s), or pre-designed primers (probes) list" editors. In the FastPCR software, users have several options for how open a file while starting the program. The user can open the original file as read-only to work with text editors, or open a file to memory without opening to text editors, which allows the user to open a larger file(s).

   For genome analysis, users can open all files from a selected folder and the program will open each file while executing the task without opening it to a text editor. When a sequence file is open, the FastPCR software displays the information about the opened sequences and the sequence formats. The information status bar shows the number of sequences, total sequence length (in nucleotides), nucleotide composition, and the purine, pyrimidine, and % CG contents.

3. In some cases, users can use the options of "probe search" or "advanced (linked) search":

   "Probe search" helps the user execute a search of binding sites not only for primers but also for probes (e.g., TaqMan, Molecular Beacon, microarray, etc.). The default value of K-mers in this analysis is set to equal 9 with a maximum single mismatch within the K-mer. This option recommended for use in cases where primer binding sites are not found or for searches of probe binding sites for which only for part of sequence is expected to be complementary, e.g., for Molecular Beacons (neither terminus is complementary to the target).

Acknowledgments

This work was supported by the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan in the framework of program funding for research (Grant No. AP08855353), and by the Academy of Finland Decision 314961.

References

1. Wicker T, Sabot F, Hua-Van A, Bennetzen J, Capy P, Chalhoub B, Flavell AJ, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH (2007) A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8:973–982
2. Galindo-Gonzalez L, Mhiri C, Deyholos MK, Grandbastien MA (2017) LTR-retrotransposons in plants: engines of evolution. Gene 626:14–25
3. Schulman AH (2013) Retrotransposon replication in plants. Curr Opin Virol 3:604–614
4. Schulman AH (2012) Hitching a ride: nonautonomous retrotransposons and parasitism as a lifestyle. In: Grandbastien M-A, Casacuberta JM (eds) Plant transposable elements. Springer Verlag, Berlin, pp 71–88
5. Wicker T, Schulman AH, Spannagl M, Haberer G, Mayer KFX, Mascher M, Springer NM, Li Q, Langridge P, Braumann I, Waugh R, Li C, Zhang G, Stein N, Gundlach H (2017) The repetitive landscape of the barley genome. Mob DNA 8:22
6. Kovach A, Wegrzyn JL, Parra G, Holt C, Bruening GE, Loopstra CA, Hartigan J, Yandell M, Langley CH, Korf I, Neale DB (2010) The *Pinus taeda* genome is characterized by diverse and highly diverged repetitive sequences. BMC Genomics 11:420
7. Shang Y, Yang F, Schulman AH, Zhu J, Jia Y, Wang J, Zhang XQ, Jia Q, Hua W, Yang J, Li C (2017) Gene deletion in barley mediated by LTR-retrotransposon BARE. Sci Rep 7:43766
8. Schulman AH (2016) Genome size and the role of transposable elements. In: Vogel JP (ed) Genetics and genomics of brachypodium. Springer International Publishing, Cham, pp 81–106
9. Ostertag EM, Kazazian HH Jr (2001) Biology of mammalian L1 retrotransposons. Annu Rev Genet 35:501–538
10. Platt RN 2nd, Vandewege MW, Ray DA (2018) Mammalian transposable elements and their impacts on genome evolution. Chromosom Res 26:25–43
11. Schulman AH, Flavell AJ, Ellis THN, Paux E (2012) The application of LTR retrotransposons as molecular markers in plants. Methods Mol Biol 859:115–153
12. Breukel C, Wijnen J, Tops C, vd Klift H, Dauwerse H, Khan PM (1990) Vector-Alu PCR: a rapid step in mapping cosmids and YACs. Nucleic Acids Res 18:3097
13. Brooks-Wilson AR, Goodfellow PN, Povey S, Nevanlinna HA, de Jong PJ, Goodfellow PJ (1990) Rapid cloning and characterization of new chromosome 10 DNA markers by Alu element-mediated PCR. Genomics 7:614–620

14. Waugh R, McLean K, Flavell AJ, Pearce SR, Kumar A, Thomas WTB, 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
15. Kalendar R, Grob T, Regina M, Suoniemi A, Schulman AH (1999) IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. Theor Appl Genet 98:704–711
16. Kalendar R, Antonius K, Smykal P, Schulman AH (2010) iPBS: a universal method for DNA fingerprinting and retrotransposon isolation. Theor Appl Genet 121:1419–1430
17. Kalendar R, Muterko A, Boronnikova S (2021) Retrotransposable elements: DNA fingerprinting and the assessment of genetic diversity. Methods Mol Biol 2222:1–11. https://doi.org/10.1007/978-1-0716-0997-2_15
18. Flavell AJ, Knox MR, Pearce SR, Ellis THN (1998) Retrotransposon-based insertion polymorphisms (RBIP) for high throughput marker analysis. Plant J 16:643–650
19. Jing R, Vershinin A, Grzebyta J, Shaw P, Smýkal P, Marshall D, Ambrose MJ, Ellis TH, Flavell AJ (2010) The genetic diversity and evolution of field pea (Pisum) studied by high throughput retrotransposon based insertion polymorphism (RBIP) marker analysis. BMC Evol Biol 10:44
20. Kalendar R, Lee D, Schulman AH (2011) Java web tools for PCR, in silico PCR, and oligonucleotide assembly and analysis. Genomics 98(2):137–144. https://doi.org/10.1016/j.ygeno.2011.04.009
21. Kalendar R, Khassenov B, Ramanculov E, Samuilova O, Ivanov KI (2017) FastPCR: an in silico tool for fast primer and probe design and advanced sequence analysis. Genomics 109 (3-4):312–319. https://doi.org/10.1016/j.ygeno.2017.05.005
22. Cornish-Bowden A (1985) Nomenclature for incompletely specified bases in nucleic acid sequences: recommendations 1984. Nucleic Acids Res 13:3021–3030
23. Peyret N, Seneviratne PA, Allawi HT, SantaLucia J Jr (1999) Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A.A, C.C, G.G, and T.T mismatches. Biochemistry 38:3468–3477
24. SantaLucia J Jr, Allawi HT, Seneviratne PA (1996) Improved nearest-neighbor parameters for predicting DNA duplex stability. Biochemistry 35:3555–3562
25. Lane AN, Chaires JB, Gray RD, Trent JO (2008) Stability and kinetics of G-quadruplex structures. Nucleic Acids Res 36:5482–5515
26. Ho PS (1994) The non-B-DNA structure of d (CA/TG)n does not differ from that of Z-DNA. Proc Natl Acad Sci U S A 91:9549–9553