Chapter 12

Transposon-Based Tagging: IRAP, REMAP, and iPBS

Ruslan Kalendar and Alan H. Schulman

Abstract

Retrotransposons are a major component of virtually all eukaryotic genomes, which makes them useful as molecular markers. Various molecular marker systems have been developed that exploit the ubiquitous nature of these genetic elements and their property of stable integration into dispersed chromosomal loci that are polymorphic within species. To detect polymorphisms for retrotransposon insertions, marker systems generally rely on PCR amplification between the retrotransposon termini and some component of flanking genomic DNA. The main methods of IRAP, REMAP, RBIP, and SSAP all detect the polymorphic sites at which the retrotransposon DNA is integrated into the genome. Marker systems exploiting these methods can be easily developed and are inexpensively deployed in the absence of extensive genome sequence data. Here, we describe protocols for the IRAP, REMAP, and iPBS techniques, including methods for PCR amplification with a single primer or with two primers, and agarose gel electrophoresis of the product using optimal electrophoresis buffers; we also describe iPBS techniques for the rapid isolation of retrotransposon termini and full-length elements.

Key words Retrotransposon, Molecular marker, IRAP, REMAP, iPBS

1 Introduction

Interspersed repetitive sequences comprise a large fraction of the genome of most eukaryotic organisms, and they are predominantly composed of transposable elements (TEs) [1]. In most species that have been studied, interspersed repeats are distributed unevenly across the nuclear genome, with some repeats having a tendency to cluster around the centromeres or telomeres [2–4]. Following the induction of recombinational processes during meiotic prophase, variation in the copy number of repeat elements and internal rearrangements on both homologous chromosomes can ensue.

Nucleotide sequences matching repetitive sequences showing polymorphism in RFLP analyses have been used as polymerase chain reaction (PCR) primers for the inter-repeat amplification polymorphism marker method [5, 6]. Such repetitive sequences include microsatellites, such as $(CA/GT)_n$ or $(CAC/GTG)_n$, which are distributed throughout the genome. A related approach was

Pascale Besse (ed.), *Molecular Plant Taxonomy: Methods and Protocols*, Methods in Molecular Biology, vol. 1115,
DOI 10.1007/978-1-62703-767-9_12,

developed to generate PCR markers based on amplification of microsatellites near the 3′ end of the Alu (SINE) transposable elements (TEs), called Alu-PCR or SINE-PCR [7]. Successful applications of microsatellite-specific oligonucleotides as PCR primers were first described in the early 1990s [5, 6, 8].

1.1 LTR Retrotransposons

Long terminal repeat (LTR) retrotransposons, or type I transposable elements, replicate by a process of reverse transcription, as do the lentiviruses such as HIV [9]. The retrotransposons themselves encode the proteins needed for their replication and integration back into the genome [10]. Their "copy-and-paste" life cycle means that they are not excised in order to insert a copy elsewhere in the genome. Hence, genomes diversify by the insertion of new copies, but old copies persist. Their abundance in the genome is generally highly correlated with genome size. Large plant genomes contain hundreds of thousands of these elements, together forming the vast majority of the total DNA [11].

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 somewhat different copy-and-paste mechanism [12, 13]. The L1 family of LINE elements and the *Alu* family of SINE elements comprise together roughly 30 % of human genomic DNA and nearly two million copies [14]. Nevertheless, integrated retroviruses, which are remnants of ancient infections, are abundant in mammalian genomes [15]. 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 [16, 17].

1.2 Retrotransposons as DNA Markers

Retrotransposon-based systems (Fig. 1) detect the insertion of elements hundreds to thousands of nucleotides long, although generally only the insertion joint itself is monitored due to the impracticality of amplifying and resolving long fragments and discriminating their insertion sites. The LTRs that bound a complete retrotransposon contain ends that are highly conserved in a given family of elements. Newly inserted retrotransposons, therefore, form a joint between the conserved LTR ends and flanking, anonymous genomic DNA. Most retrotransposon-based marker systems use PCR to amplify a segment of genomic DNA at this joint. Generally, one primer is designed to match a segment of the LTR conserved with a given family of elements but different in other families. The primer is oriented towards the LTR end. The second primer is designed to match some other feature of the genome. The first retrotransposon method described was SSAP or S-SAP

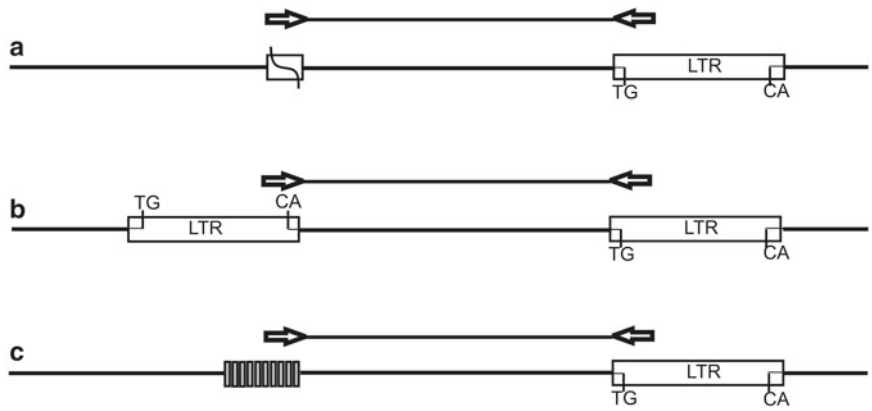

Fig. 1 Retrotransposon-based molecular marker methods. Multiplex products of various lengths from different loci are indicated by the *bars* above or beneath the diagrams of each reaction. Primers are indicated as *arrows.* (**a**) The SSAP method. Primers used for amplification match the adapter (restriction site shown as *empty box*) and retrotransposon (*LTR box*). (**b**) The IRAP method. Amplification takes place between retrotransposons (*left* and *right LTR boxes*) near each other in the genome (*open bar*), using retrotransposon primers. The elements are shown oriented head-to-head using a single primer. (**c**) The REMAP method. Amplification takes place between a microsatellite domain (*vertical bars*) and a retrotransposon, using a primer anchored to the proximal side of the microsatellite and a retrotransposon primer

(sequence-specific amplified polymorphism, *see* Fig. 1a), where one primer matched the end of the *BARE1* retrotransposon of barley and the other matched an AFLP-like restriction site adapter [18].

1.3 Sequence-Specific Amplified Polymorphism

Sequence-specific amplified polymorphism (SSAP) was described by Waugh and coworkers in 1997 [18], but has several origins and forms [19–22]. The SSAP method can be considered to be a modification of AFLP [23] or as a variant of anchored PCR [24]. The method described by Waugh and colleagues [18] has many similarities to AFLP, especially in that two different enzymes are used to generate the template for the specific primer PCR and that selective bases are used in the adapter primer.

In the SSAP procedure, it is important to maximize the sequence complexity of the template for the specific primer amplification, so a single enzyme digestion is used [25]. As with the method described for *BARE1* [18], the adapter primer is selective. This is a matter of convenience, and nonselective primers could be substituted where the enzyme used for digestion has a larger recognition sequence, or if the copy number were lower. In general, LTR ends are convenient for the design of SSAP primers [26]. However, for the *PDR1* retrotransposon in *Pisum*, the LTR is exceptionally short at 156 bp, so a GC-rich primer could be designed corresponding to the polypurine tract (PPT) which is found internal to the 3′ LTR in retrotransposons. For *BARE1* in barley and other high-copy-number families, the number of selective bases may be increased compared to the first version of the

protocol [18, 27]. Furthermore, *BARE1* and most other retrotransposons have long LTRs, necessitating an anchor primer in the LTR near to the external terminus. The main feature of the SSAP procedure that may be modified for various situations is the location of the sequence-specific primer [28]. The choice of this primer is critical and can be modified according to need. For example, internal primer sites have been exploited to describe structural variation within retrotransposons [29], and the primers can be applied to defined sequences other than the LTR or PPT.

1.4 Inter-retrotransposon Amplification Polymorphism (IRAP) and Retrotransposon-Microsatellite Amplification Polymorphism (REMAP)

The IRAP (Fig. 1b) and REMAP (Fig. 1c) methods represent a departure from SSAP, because no restriction enzyme digestion or ligation step is needed and because the products can be resolved by conventional agarose gel electrophoresis without resort to a sequencing apparatus. The IRAP method detects retrotransposon insertional polymorphisms by amplifying the portion of DNA between two retroelements. It uses one or two primers pointing outwards from an LTR and therefore amplifies the tract of DNA between two nearby retrotransposons. IRAP can be carried out with a single primer matching either the 5′ or 3′ end of the LTR but oriented away from the LTR itself or with two or more primers. The two primers may be from the same retrotransposon element family or may be from different families. The PCR products, and therefore the fingerprint patterns, result from amplification of hundreds to thousands of target sites in the genome.

The complexity of the pattern obtained will be influenced by the retrotransposon copy number, which mirrors genome size, as well as by their insertion pattern and by the size of the retrotransposon families chosen for analysis. Furthermore, thousands of products can neither be simultaneously amplified to detectable levels nor resolved on a gel system. Hence, the pattern obtained represents the result of competition between the targets and products in the reaction. As a result, the products obtained with two primers do not represent the simple sum of the products obtained with the primers individually.

If retrotransposons were fully dispersed within the genome, IRAP will either produce products too large to give good resolution on gels or target amplification sites too far apart to produce products with the available thermostable polymerases. This is because retrotransposons generally tend to cluster together in "repeat seas" surrounding "gene islands" and may even nest within each other. For example, the *BARE1* retrotransposon of barley, an abundant superfamily *Copia* element, is present as about 13,000 full-length copies of about 8.9 kb and 90,000 solo LTRs of 1.8 kb in the cultivar Bomi. Given a genome of roughly 5×10^9 bp, these elements comprise 5.6 % of the genome but would occur only about once every 46 kb if they were fully interspersed. Nevertheless, IRAP with *BARE1* primers displays a range of products from 100 bp upwards of 10 kb (Fig. 2).

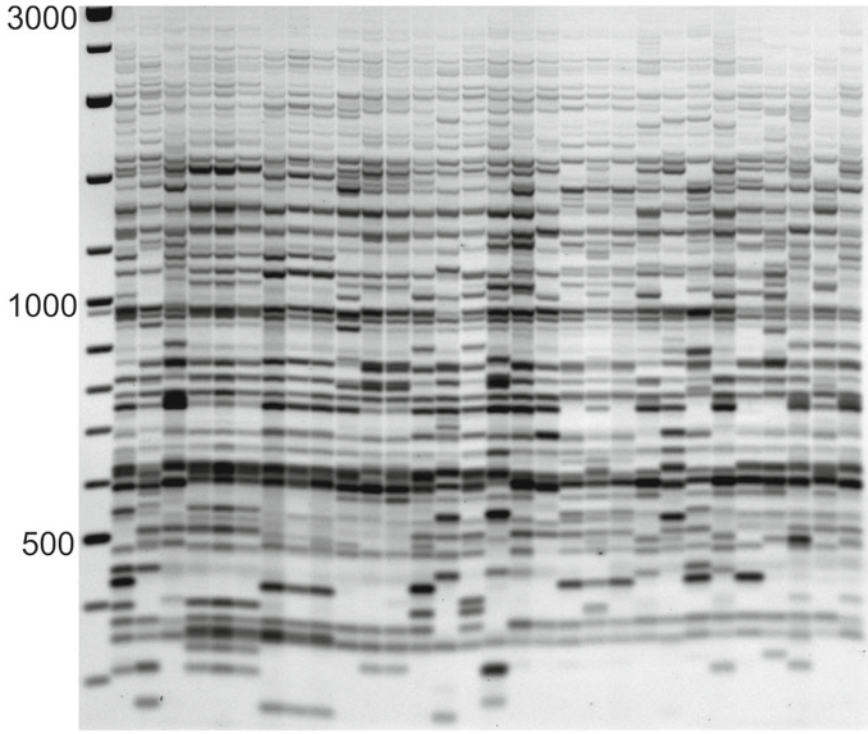

Fig. 2 Utility of IRAP for a diversity analysis of plant species. The phenogram of 30 genotypes of populations of *H. spontaneum* based on IRAP analysis is shown as negative images of ethidium bromide-stained agarose gels following electrophoresis. Results for *BARE1* LTR primer 1369 are shown. A 100 bp DNA ladder is present on the *left*

The REMAP method is similar to IRAP, except that one of the two primers matches an SSR motif with one or more non-SSR anchor nucleotides present at the 3′ end of the primer. Microsatellites of the form $(NN)_n$, $(NNN)_n$, or $(NNNN)_n$ are found throughout plant and animal genomes. In cereals, they furthermore appear to be associated with retrotransposons [30]. Differences in the number of SSR units in a microsatellite are generally detected using primers designed to unique sequences flanking microsatellites. Alternatively, the stretches of the genome present between two microsatellites may be amplified by ISSR [6, 8], in a way akin to IRAP. In REMAP, anchor nucleotides are used at the 3′ end of the SSR primer both to avoid slippage of the primer within the SSR, which would produce a "stutter" pattern in the fingerprint, and to avoid detection of variation in repeat numbers within the SSR. REMAP uses primer types that are shared by IRAP and ISSR. Although it would appear that the SSR primers in REMAP should also yield ISSR products and the LTR primers also IRAP products, in practice this is rarely the case. This is probably due to a combination of factors including both genome structure and competition within the PCRs.

1.5 Inter-primer Binding Site Polymorphism (iPBS)

A major disadvantage of all retrotransposon-based molecular marker techniques is the need for sequence information to design element-specific primers. The primary requirement is the sequence of an LTR end, harvested either from a database or produced by cloning and sequencing the genomic DNA that flanks conserved segments of retrotransposons. Although rapid retrotransposon isolation methods based on PCR with conserved primers for TE have been designed, it maybe still necessary to clone and sequence hundreds of clones to obtain a few good primer sequences. The LTRs contain no conserved motifs, which would allow their direct amplification by PCR.

Several restriction and adaptor-based methods for LTR cloning have been developed, which are based on the conservation of reverse transcriptase domain, especially for the superfamily *Copia* retrotransposons [28]. In general, however, all reverse-transcribing elements, including LTR retrotransposons of superfamily *Gypsy* as well as LINE retrotransposons, can be obtained by PCR with degenerate primers. For example, for *Copia*, two degenerate primers were designed for the RT motifs encoding TAFLHG and, for the reverse primer, the downstream YVDDML, as well as for QMDVKT and YVDDML respectively, in order to amplify the family-specific domain in between [31–33]. For *Gypsy* elements, degenerate primers were designed for the RT motifs encoding RMCVDYR, LSGYHQI, or YPLPRID and for the reverse primers for the domains YAKLSKC or LSGYHQI. The RT-based isolation method is limited, of course, to the families of retrotransposons that contain RT and the chosen domains. Thus, for example, non-autonomous groups such as TRIMs, LARDs, and SINEs cannot be found using this approach [34, 35].

The LTR retrotransposons and all retroviruses contain a conserved binding site for tRNA. Generally $tRNAi^{Met}$ is the most common, but also $tRNA^{Lys}$, $tRNA^{Pro}$, $tRNA^{Trp}$, $tRNA^{Asn}$, $tRNA^{Ser}$, $tRNA^{Arg}$, $tRNA^{Phe}$, $tRNA^{Leu}$, and $tRNA^{Gln}$ can be found. Elongation from the 3′-terminal nucleotides of the respective tRNA results in the conversion of the retroviral or retrotransposon RNA genome to double-stranded DNA prior to its integration into the host DNA. While the process of reverse transcription is conserved among virtually all retroelements, the specific tRNA capture varies for different retroviruses and retroelements. The primer binding sequences (PBS) are almost universally present in all LTR-retrotransposon sequences. Hence, an isolation method for retrotransposon LTRs, which is based on the PBS sequence, has the potential for cloning all possible LTR retrotransposons.

The inter-PBS amplification (iPBS) technique has led to the development of a virtually universal and exceedingly efficient method, which utilizes the conserved parts of PBS sequences, both for direct visualization of polymorphism between individuals, polymorphism in transcription profiles, and fast cloning of LTR segments from genomic DNA, as well as for database searches of LTR retrotransposons (Fig. 3). Many retrotransposons are nested, recombined, inverted, or truncated, yet can be easily amplified

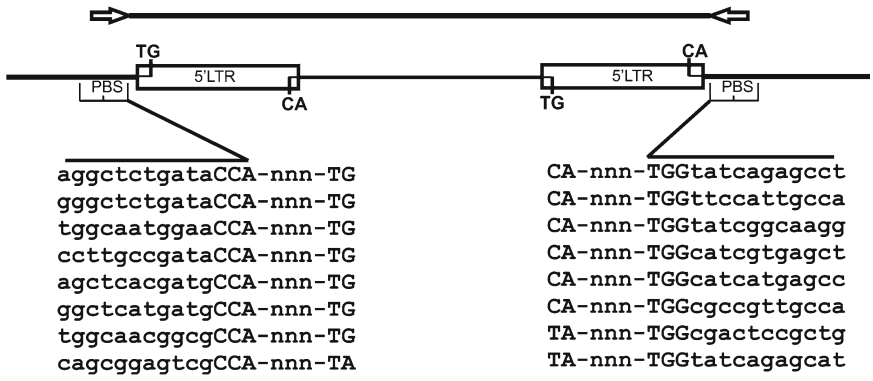

Fig. 3 The inter-PBS amplification (iPBS) scheme and LTR retrotransposon structure. Two nested LTR retrotransposons in inverted orientation amplified from single primer or two different primers from primer binding sites. PCR product contains both LTRs and PBS sequences as PCR primers in the termini. In the figure, general structure for PBS and LTR sequences and several nucleotides long spacer between 5′LTR (5′-CA) and PBS (5′-TGG3′) are schematically shown

using conserved PBS primers in all plant species tested. Fragments of retrotransposons containing a 5′ LTR and part of the internal domain are often located near other entire or similarly truncated retrotransposons. Therefore, PBS sequences are very often located sufficiently near to each other to allow amplification. This situation allows the use of PBS sequences for cloning LTRs. Where the retrotransposon density is high within a genome, PBS sequences can be exploited for detection of their chance association with other retrotransposons. When retrotransposon activity or recombination has led to new genome integration sites, the iPBS method can be used to distinguish reproductively isolated plant lines. In this case, amplified bands derived from a new insertion event or from recombination will be polymorphic, appearing only in plant lines in which the insertions or recombination have taken place.

The PBS primer(s) can amplify nested inverted retrotransposons or related elements' sequences dispersed throughout genomic DNA. The PCR amplification occurs in this case between two nested elements' PBS domains and produces fragments containing the insertion junction between the two nested LTRs. After retrieving LTR sequences of a selected family of retrotransposons, an alignment is made of them to find the most conserved region [36]. The related plant species have conserved regions in LTR for members of the same retrotransposon family. Thus, alignments of several LTR sequences from several species will identify these conserved regions. Subsequently, these conserved domains of LTRs can be used for inverted primers designed for long distance PCR, for cloning of whole retrotransposons, and also for the IRAP, REMAP, or SSAP marker techniques. The iPBS amplification technique shows about the same level of polymorphism in comparison with IRAP and REMAP, and it is an efficient method for the detection of cDNA polymorphism and clonal differences resulting from retrotransposon activities or recombination.

2 Materials

2.1 Reagents

Prepare all solutions using Milli-Q or equivalent ultrapure water and analytical grade reagents.

1. TE buffer (10×): 100 mM Tris–HCl (pH 8.0), 10 mM EDTA. DNA and primers should be stored in a 1× TE solution.
2. Electrophoresis buffer (10× THE): 200 mM Tris-HEPES (pH 8.06), 5 mM EDTA. Weigh 24.2 g Tris-base and 47.7 g HEPES (free acid), add 10 mL 0.5 M EDTA (pH 8.0), dissolve in water; bring final volume to 1 L. Store at +4 °C. While we get best results with 1× THE, standard 1× TBE (50 mM Tris-H_3BO_3, pH 8.8), 1× TAE (40 mM Tris-CH_3COOH, pH 8.0), or 1× TPE (40 mM Tris-H_3PO_4, pH 8.0) buffers may also be used.
3. Gel loading buffer (10×): 20 % (w/w) Polysucrose 400, 100 mM Tris–HCl (pH 8.0), 10 mM EDTA, ~0.01 % (w/w) Orange G, and ~0.01 % Xylene Cyanol FF. Dissolve 20 g Polysucrose 400 (Ficoll 400) in 80 mL 10× TE buffer. Add Orange G and Xylene Cyanol FF according on the desired color intensity. Store at +4 °C.
4. Thermostable polymerase: many types and sources of recombinant thermostable polymerases are effective. Most preferable for PCR use are the recombinant polymerases with 3′–5′ exonuclease proofreading activity that permit a "hot start," such as Phire® Hot Start II DNA Polymerase (Thermo Scientific) from *Pyrococcus furiosus*. Another excellent choice is *Thermus thermophilus* Biotools DNA polymerases (Biotools S.A., Madrid, Spain). Any *Thermus aquaticus* (*Taq*) DNA polymerase is applicable, however. We have tested several *Taq* DNA polymerases, including those of DreamTaq™ (Thermo Scientific), FIREPol® (Solis BioDyne), MasterAmp™ (Epicentre), and GoTaq® (Promega). Other thermostable polymerases, such as that from *Thermus brockianus* (DyNAzyme™ II, Thermo Scientific), was also tested to determine if the choice of polymerase enzyme had an effect on the products amplified. A polymerases mix consisting of 100 U of *Taq* DNA polymerase and 0.5 U *Pfu* DNA polymerase improves amplification of long bands and the accuracy of the PCR. Long distance PCR is performed with DyNAzyme™ EXT (Thermo Scientific) or Phusion® High-Fidelity (Thermo Scientific) or LongAmp™ *Taq* DNA polymerases (New England Biolabs).
5. PCR buffers (1×): several PCR buffers for *Taq* polymerase are suitable for PCR: Buffer 1: 20 mM Tris–HCl (pH 8.8), 2 mM $MgSO_4$, 10 mM KCl, 10 mM $(NH_4)_2SO_4$; Buffer 2: 10 mM Tris–HCl (pH 8.8), 2 mM $MgCl_2$, 50 mM KCl, 0.1 % Triton X-100; Buffer 3: 50 mM Tris–HCl (pH 9.0), 2 mM $MgCl_2$,

15 mM $(NH_4)_2SO_4$, 0.1 % Triton X-100. The PCR and its efficiency depend on which buffer and enzyme combination is used (*see* **Note 1**).

6. Ethidium bromide solution in water, 0.5 mg/mL. Store at room temperature.
7. SYBR Green I gel staining solution (50×) in 50 % dimethyl sulfoxide (DMSO) in Milli-Q water. The 10,000× concentrate is diluted to 50× with 50 % DMSO in Milli-Q water. Store at –20 °C.
8. DNA ladder for electrophoresis GeneRuler™ DNA Ladder Mix (Thermo Scientific), 100–10,000 base range, or similar. DNA ladder diluted with 1× gel loading buffer to final concentration 25 ng/μL.
9. Agaroses: RESolute Wide Range (BIOzym), D1 low EEO (Conda), Premium (Serva), MP (AppliChem). Agaroses of the LE type are not effective for fine resolution of fingerprinting bands. Cambrex NuSieve 3:1 and MetaPhor agaroses have low gel strength and are uncomfortable for gel manipulation; 1 % agaroses with gel strength >1,700 g/cm^2 can be used. Electrophoresis gels of enhanced selectivity can be produced by adding a performed polymer (additive) to a polymerization solution.
10. 10× FastDigest® buffer for restriction DNA with FastDigest® restriction enzymes: *MseI*, *PstI*, *TaqI*, and *TaiI*.
11. MinElute PCR Purification (Qiagen) or similar kit, for PCR products purified.
12. QIAEX II Gel Extraction Kit (Qiagen) or similar kit, for silica-membrane-based purification of DNA fragments from 40 bp to 50 kb from gel or enzymatic reactions.
13. PCR product TA cloning kits, TOPO® TA Cloning® Kit (pCR®2.1 plasmid vector) with TOP10 *E. coli* (Invitrogen) or alternative kits. PCR products should be amplified with Taq, Biotools, or DyNAzyme II DNA polymerases and have single 3′ adenine overhangs. Primers should use a hydroxyl group at 5′ termini. The alternative to TA cloning, BigEasy® Long PCR Cloning Kits (Lucigen) can be effectively used for GC cloning of difficult and long PCR fragments.
14. Competent *E. coli* cells (10^9 cfu/μg), One Shot® TOP10 (Invitrogen), JM109 (Promega).

2.2 Equipment

1. Thermal cycler for 0.2 mL tubes or plates (96 well), with a rapid heating and cooling capacity between 4 and 99 °C, so that the temperature can be changed by 3–5 °C/s, for example, the Mastercycler Gradient (Eppendorf AG) or the PTC-100 Programmable Thermal Controller (Bio-Rad Laboratories).
2. Power supply (minimum 300 V, 400 mA) for electrophoresis.

3. Horizontal electrophoresis apparatus without special cooling. Most commercially available medium- or large-scale horizontal DNA gel electrophoresis systems are suitable, for example, from such suppliers as GE Healthcare, Hoefer, or Bio-Rad. These include the GNA-200, Hoefer HE 99X Max Submarine, BioExpress (Wide Maxi Horizontal Gel System (E-4123-1)), and Sigma (Maxi-Plus). Small electrophoresis boxes and short gel trays are not suitable due to the large number of PCR products that need to be resolved. We routinely employ an apparatus with a run length of 20 cm.
4. Gel comb, a 36 or more well comb, 1 mm thickness, forming 3–4 mm wide wells, with a 1 mm well spacing. This comb is ideal for analysis of any PCR amplification product or DNA restriction enzyme digest. The small space between the slots is important for analysis of banding patterns and for comparing lanes across the gel. Also this thickness of comb improves band resolution.
5. UV transilluminator, for visualization of ethidium bromide-stained or SYBR green-stained nucleic acids, with a viewing area of 20×20 cm.
6. Dark Reader (Clare Chemical Research), for visualization and isolation of SYBR green-stained nucleic acids.
7. Imaging system. A digital gel electrophoresis scanner for detection of ethidium bromide-stained nucleic acids by fluorescence (532 nm green laser) or SYBR green-stained nucleic acids (473 nm blue laser), with a resolution of 50–100 μm. Examples include the FLA-5100 imaging system (Fuji Photo Film GmbH., Germany). Software such as the Aida Image Analyzer and Adobe Photoshop is required for image analysis and manipulation.

3 Methods

3.1 Primer Design

PCR primers are designed to match an LTR sequence near to either its 5′ or 3′ end, with the primer oriented so that the amplification direction is towards the nearest end of the LTR. Generally it is best to base the design on a sequence alignment for representative LTRs from a particular family of elements and to place the primer within the most conserved region for that family. For long LTRs, it is often useful to test primers at several locations within the LTR or internal part of the retrotransposon and in both orientations, particularly if there is evidence for nested insertions in the genome. Primers can be placed directly at the end of the LTR facing outwards, provided that they do not form dimers or loops. For primers placed at the edge of the LTR, one or more additional selective bases can be added at 3′ end in order to reduce the number of amplification targets (Fig. 4). This can be tried in a second round of primer design, if the initial primer yields amplification

TGGGAAGTAGTAAGTAGATGA**TGGGTTGCTCGAGTGACAGAAGC**TTAAACCCCAGTTTATGTGTTG**CTTCGTGAGGGGCTGATTTGGATC**CA
TGGGA-GTAGTAAGTAGATGA**TGGGTTGCTAGAGTGACAGAAGG**TTAAACCCCGGTCTATGCGTTG**CTTCGTAAGGGGCTGATTTGGATC**CA
TGGGAAGTAGTAAGTAGATGA**TGGGTTGCTAGAGTGACAGAAGC**TTAAACCCTAGTTTATGCGTTG**CTTCGTAAGGGGCTGATTTGGACC**CA
TGGGAAGTAGTAAGTAGATGA**TGGGTTGCTAGAGTGACAGAAGC**TTAAACCCTAGTTTATGCGTTG**CTTCGTAAGAGGCTGATTTGGATA**CA
TGGGAAGTACTAAGTAGATGA**TGGGTTGCTAGAGTAACAGAAGC**TTAAACCCTAGTTTATGCGTTG**CTTCGTAAGGGGCTGATTTGGATC**CA
TGGGAAGTCATAAGTAGATGA**TGGGTTGCTAGAGTGACAGAAGC**TTAAACCCTAGTTTATGCGTTG**CTTCGTAAGGGGCTGATTTGGATC**CA
TGGGAAGTAGTAAGTAGATGA**TGGGTTGCTAGAGTGACAGAAGC**TTAAACCCTAGTTTATGCGTTG**CTTCATAAGGGGCTGAATTGGATC**CA
TGGGAAGTAATAAGTAGATGA**TGGGTTGCTAGAGTGACAGGAGC**TTAAACCCTAGTTTATGCGTTT**CTTCGTAAGGGGCTGATTTGGATC**CA
TGGGAAGTAGTAAGTAGATGA**TTGGTTGCTAGAGTGACAGAAGC**TTAAACCCTAGTTTATGCGTTG**CTTCGTAAGGGGCTAATTTGGATC**CA
TAGGAAGTAATAAGTAGATTA**TGGGTTGCTAGAGTGACAGAAGC**TTAAACCCTAGTTTATGCGTTG**CTTCGTAAGGGGCTGATTTGGATC**CA
TGGGAAGTAATAAGTAGATGA**TGGGTTGCTAGAGTGACAGAAGC**ATAAACCCTAGTTTATGTGTTG**CTTCGTAAGGGGCTGATTTGGATC**CA
TGGGAAGTAATAAGTAGATGA**TGGGTTGCTAGAGTGACATAAGC**TTAAACCCTAGTTTATGCGTTC**CTTTGTTAGGGGCTGATTTGGATC**CA
TGGGAAGTAATAAGTAGATGA**TGGGTTGCTAGAGTGACACAAGC**TTAAACCCTAGTTTATGCGTTG**CTTCATAAGGGGCTGATTTGGATC**CA
TGGGAAGTAATAAGTAGATGA**TGGGTTGCTAGAGTGACACAAGC**TTAAACCCTAGTTTATGCGTTG**CTTCATAAGGGGCTGATTTGGATC**CA
TGGGAATTAATAAGTAGATGA**TGGGTTGCTAGAGTGACAGAAGC**TTAAACCCTAGTTTATGCGTTG**CTTCGTTAGGGGCTGATTTGGATC**CA
GGGGAAGTAATAAGTAGATGA**TGGGTTTCTAGAGTGACAAAAGC**TTAAACCCTAGTTTATGTGTTG**CTTCATTAGGGGCTGATATGGATC**CA

5′-**TGGGTTGCTAGAGTGACAGAAGC** 5'-**CTTCGTAAGGGGCTGATTTGGATC**

Fig. 4 Example of LTR primer design: conserved 3′ end of *Cereba* LTRs and matching primers

Table 1
Microsatellite primers

Microsatellites sequence	Primers sequences (5′–3′)	T_m (°C)[a]
Two-base-repeat microsatellites		
$(CT)_n$	(CT)10G; (CT)10T; (CT)10A	52.1; 51.5; 50.3
$(CA)_n$	(CA)10G; (CA)10T; (CA)10A	56.6; 56.2; 55.9
$(TC)_n$	(TC)10G; (TC)10C; (TC)10A	53.6; 53.5; 52.4
$(TG)_n$	(TG)10G; (TG)10C; (TG)10A	58.3; 58.8; 57.0
$(AG)_n$	(AG)10G; (AG)10C; (AG)10T	53.4; 54.0; 52.9
$(AC)_n$	(AC)10G; (AC)10C; (AC)10T	58.4; 58.3; 57.6
$(GA)_n$	(GA)10T; (GA)10C; (GA)10A	51.3; 52.0; 51.1
$(GT)_n$	(GT)10T; (GT)10C; (GT)10A	56.4; 56.6; 55.1
Three-base-repeat microsatellites		
$(CTC)_n$	(CTC)6G; (CTC)6T; (CTC)6A	58.0; 57.0; 56.7
$(GTG)_n$	(GTG)6C; (GTG)6T; (GTG)6A	61.7; 60.6; 59.7
$(CAC)_n$	(CAC)6G; (CAC)6T; (CAC)6A	61.2; 60.4; 60.0
$(ACC)_n$	(ACC)6G; (ACC)6T; (ACC)6C	63.0; 62.2; 63.0
$(TCG)_n$	(TCG)6G; (TCG)6C; (TCG)6A	61.3; 61.7; 59.9

[a]Oligonucleotide concentration is 200 nM

products containing too many weak individual bands for analysis by gel electrophoresis. If the LTRs are short (<300 bp), the primers may also be designed to match internal regions, but this will concomitantly increase the size of the amplified products.

Database searches can sometimes be used to find unannotated, native LTR sequences matching characterized retrotransposons from other species. However, care should be taken in defining the ends of the LTRs. Generally, mapping of the RT primer binding sites PBS and PPT is needed in order to define the LTR ends with confidence. Microsatellite primers for REMAP or ISSR should be

Table 2
Retrotransposon LTR primers

Name	Sequence	TE, source	T_m (°C)[a]	Optimal annealing T_a (°C)
560	TTGCCTCTAGGGCATATTTCCAACA	*Wis2*, LTR	58.2	58–65
554	CCAACTAGAGGCTTGCTAGGGAC		58.8	60–68
2105	ACTCCATAGATGGATCTTGGTGA		54.9	55–61
2106	TAATTTCTGCAACGTTCCCCAACA		57.3	58–65
2107	AGCATGATGCAAAATGGACGTATCA	*Wilma*, LTR	57.2	58–65
833	TGATCCCCTACACTTGTGGGTCA		59.5	60–68
2108	AGAGCCTTCTGCTCCTCGTTGGGT		64.2	64–72
516	TCCTCGTTGGGATCGACACTCC		60.5	60–68
2109	TACCCCTACTTTAGTACACCGACA	*Daniela*, LTR	56.3	57–62
2110	TCGCTGCGACTGCCCGTGCACA		67.8	68–72
2111	CAGGAGTAGGGTTTTACGCATCC		57.2	58–65
2112	TGCTGCGACTGCCCGTGCACA		66.5	66–72
2113	TACGCATCCGTGCGGCCCGAAC		66.5	66–72
2114	GGACACCCCCTAATCCAGGACTCC	*Fatima*, LTR	62.4	62–68
2115	CAAGCTTGCCTTCCACGCCAAG		61.6	62–67
2116	CGAACCTGGGTAAAACTTCGTGTC		57.9	58–64
2117	AGATCCGCCGGTTTTGACACCGACA		64.1	64–72
728	TGTCACGTCCAAGATGCGACTCTATC	*Sabrina*, LTR	59.8	60–66
2118	GTAGATAATATAGCATGGAGCAATC		50.8	55–61
2119	AGCCACTAGTGAAACCTATGG		54.4	55–63
2120	GTGACCTCGAAGGGATTGACAACC		59.5	60–65
2121	ACTGGATTGATACCTTGGTTCTCAA		55.8	55–62
2122	AGGGAAATACTTACGCTACTCTGC		56.3	57–64
432	GATAGGGTCGCATCTTGGGCGTGAC	*Sukkula*, LTR	63.0	63–68
480	GGAACGTCGGCATCGGGCTG		63.3	63–68
1319	TGTGACAGCCCGATGCCGACGTTCC		66.8	66–72
2123	GGAAAAGTAGATACGACGGAGACGT	*Wham*, LTR	57.8	58–63
483	TCTGCTGAAAACAACGTCAGTCC		57.8	58–63
1623	TGCGATCCCCTATACTTGTGGGT		60.1	60–65
552	CGATGTGTTACAGGCTGGATTCC	*Bagy1*, LTR	57.7	58–64
1369	TGCCTCTAGGGCATATTTCCAACAC	*BARE1*, LTR	59.0	60–65

[a]Oligonucleotide concentration is 200 nM

designed according to two principles: first, the primer length should be between 19 and 22 bases; second, the last base at 3′-end of the primer should be designed as a selective base, which is absent in repeat unit itself. We have provided examples of LTR conservation and consequent primer design for LTRs and microsatellites (Fig. 4 and Tables 1, 2, and 3).

We have designed primers using the FastPCR software [37] or Java Web tools [38]. Occasionally, not all primers (those derived from retrotransposons or SSR primers) will work in the PCR. The genome may contain too few retrotransposon or microsatellite target sites, or they may be too dispersed for the generation of PCR

Table 3
PBS 18-mer primers

Name	Sequence	T_m (°C)[a]	Optimal annealing T_a (°C)
2217	ACTTGGATGTCGATACCA	52.5	51.4
2218	CTCCAGCTCCGATTACCA	56.1	51.0
2219	GAACTTATGCCGATACCA	51.5	53.0
2220	ACCTGGCTCATGATGCCA	59.0	57.0
2221	ACCTAGCTCACGATGCCA	58.0	56.9
2222	ACTTGGATGCCGATACCA	55.7	53.0
2224	ATCCTGGCAATGGAACCA	56.6	55.4
2225	AGCATAGCTTTGATACCA	50.5	55.0
2226	CGGTGACCTTTGATACCA	54.2	53.1
2228	CATTGGCTCTTGATACCA	51.9	54.0
2229	CGACCTGTTCTGATACCA	53.5	52.5
2230	TCTAGGCGTCTGATACCA	54.0	52.9
2231	ACTTGGATGCTGATACCA	52.9	52.0
2232	AGAGAGGCTCGGATACCA	56.6	55.4
2237	CCCCTACCTGGCGTGCCA	65.0	55.0
2238	ACCTAGCTCATGATGCCA	55.5	56.0
2239	ACCTAGGCTCGGATGCCA	60.4	55.0
2240	AACCTGGCTCAGATGCCA	58.9	55.0
2241	ACCTAGCTCATCATGCCA	55.5	55.0
2242	GCCCCATGGTGGGCGCCA	69.2	57.0
2243	AGTCAGGCTCTGTTACCA	54.9	53.8
2244	GGAAGGCTCTGATTACCA	53.7	49.0
2245	GAGGTGGCTCTTATACCA	53.1	50.0
2246	ACTAGGCTCTGTATACCA	50.9	49.0
2249	AACCGACCTCTGATACCA	54.7	51.0
2251	GAACAGGCGATGATACCA	54.3	53.2
2252	TCATGGCTCATGATACCA	52.7	51.6
2253	TCGAGGCTCTAGATACCA	53.4	51.0
2255	GCGTGTGCTCTCATACCA	57.1	50.0
2256	GACCTAGCTCTAATACCA	49.6	51.0
2257	CTCTCAATGAAAGCACCA	52.4	50.0

(continued)

Table 3
(continued)

Name	Sequence	T_m (°C)[a]	Optimal annealing T_a (°C)
2295	AGAACGGCTCTGATACCA	55.0	60.0
2298	AGAAGAGCTCTGATACCA	51.6	60.0
2373	GAACTTGCTCCGATGCCA	57.9	51.0
2395	TCCCCAGCGGAGTCGCCA	66.0	52.8
2398	GAACCCTTGCCGATACCA	57.1	51.0
2399	AAACTGGCAACGGCGCCA	63.4	52.0
2400	CCCCTCCTTCTAGCGCCA	61.6	51.0
2401	AGTTAAGCTTTGATACCA	47.8	53.0
2402	TCTAAGCTCTTGATACCA	49.0	50.0
2415	CATCGTAGGTGGGCGCCA	62.5	61.0

[a]Oligonucleotide concentration is 1,000 nM

products. Alternatively, sequence divergence in ancient retrotransposon insertions or polymorphisms between heterologous primers and native elements may lead to poor amplification. Some primers generate smears under all PCR conditions. Many sources can contribute to this problem, ranging from primer structure to variability in the target site and competition from other target sites. Generally, it is more efficient to design another primer than to try to identify the source of the problem. Furthermore, primers which produce a single, very strong band are not suitable for fingerprinting.

3.2 Template DNA

The quality of template DNA plays an important role in the quality of the resulting fingerprint. Standard DNA extraction methods are sufficient to yield DNA of high quality from most samples. DNA should be free of polysaccharides, pigments, and secondary metabolites. Some tissue materials contain much polysaccharides, pigments, oils, or polyphenols, which can reduce the efficiency of PCR. Furthermore, contaminated DNAs will decline in PCR performance during prolonged (a month or more) periods of storage, due to chemical modification (*see* **Note 2**). Such DNAs (e.g., from *Brassica* spp.) should be extracted, for example, with methods involving guanidine thiocyanate at weakly acidic pH (below pH 6), followed by hot chloroform DNA extraction. DNA templates should be diluted with 1× TE solution for the appropriate working concentration (5 ng/μL) and stored at 4 °C. High-quality DNA can be stored at 4 °C for many years without showing any PCR inhibition or decrease in amplification efficiency for the longer bands.

3.3 Polymerase Chain Reaction

3.3.1 Protocol for IRAP, REMAP, and iPBS

The method described below is for reactions with standard *Taq* polymerase or with proofreading Phire® Hot Start II DNA Polymerase. PCR products can be separated by agarose gel electrophoresis or, if fluorescent-labeled primers are used after *Taq*I digestion of PCR fragments, by sequencing gel systems instead. For separation on sequencing systems, fluorescent, Cy5- or Cy3-labeled primers may be used; no special reaction conditions are needed.

1. The PCR can be set up at room temperature. Prepare a master mix for the appropriate number of samples. The DNA polymerase is the last component added to the PCR mixture. Mix well the master mix and centrifuge the tube. The reaction volume may vary from 10 to 25 μL; 10 μL is enough for running two gels. The final primer concentration(s) in the reaction can vary from 200 to 400 nM for primers in combination. For a single PCR primer, use 400 nM for IRAP and 1,000 nM for iPBS amplification. Although higher primer concentrations increase PCR efficiency and the rapidity of DNA amplification, they also produce over-amplified products.
2. Perform PCR with Phire® Hot Start II DNA Polymerase in a 25 μL reaction mixture containing 25 ng DNA (*see* **Note 3**), 1× Phire Reaction Buffer (containing 1.5 mM $MgCl_2$), 0.2–1 μM primer(s), 200 μM dNTP, 0.2 μL Phire® Hot Start II DNA Polymerase.
3. Alternative protocol for PCR with Taq polymerase: use a 25 μL reaction mixture containing 25 ng DNA, 1× Taq PCR Buffer (including 1.5 mM $MgCl_2$), 0.2–1 μM primer(s), 200 μM dNTP, 0.2 μL (1 U) Taq DNA polymerase (5 U/μL).
4. Centrifuge all tubes or the plate before starting amplification.
5. The PCR with Phire® Hot Start II DNA Polymerase (40 min) should consist of a 2 min initial denaturation step at 98 °C and 30–32 cycles of 5–10 s at 98 °C, 30 s at 55–72 °C (*see* **Note 4**), and 30 s at 72 °C; complete with a 2 min final extension at 72 °C. The denaturation step in PCR needs to be as short as possible. Usually 5 s at 98 °C is enough for most templates. For some templates requiring longer denaturation time, up to 10 s can be used, with the initial denaturation time extended up to 3 min.
6. The standard PCR with Taq polymerase (70 min total) should consist of: a 3 min initial denaturation step at 95 °C; 30–32 cycles of: 15 s at 95 °C, 30 s at 55–72 °C, and 60 s at 72 °C; a 5 min a final extension at 72 °C. PCR thermal conditions can be varied without large effects on the resulting band pattern.
7. The time of the annealing step can vary from 30 to 60 s, and the annealing temperature depends on the melting temperature of the primer; it should be between 55 and 68 °C (60 °C is optimal for almost all primers and their combinations in IRAP and REMAP; *see* **Note 5**).
8. PCRs can be stored at 4 °C overnight.

3.3.2 Long Distance Inverted PCR to Isolate Complete LTR Retrotransposons

Complete LTR retrotransposons can be identified and extracted using long distance PCR with inverted LTR primers and running low numbers of PCR cycles [10–15] to select for abundant elements. The iPBS amplification technique helps with cloning of LTR segments from genomic DNA or with database searches. After retrieving LTR sequences of a selected family of retrotransposons, align them to identify the most conserved regions. The conserved segments of the LTR are used for design of inverted primers for long distance PCR, such as for cloning of whole elements (*see* **Note 6**).

Several primer pairs, oriented away from each other as for inverse PCR, are designed for each identified element. Inverted primers of 25–30 nt with high T_m (>60 °C) need to be designed from the LTR. This allows annealing and polymerase extension in one step at 68–72 °C, thereby increasing the efficiency of the amplification of long fragments. The PCR product will be consist of complementary fragments of both LTRs and the central part of retroelement. To avoid formation of nonspecific PCR products and, assuming a high copy number for the retroelement of interest, the reaction is carried out with a low number of PCR cycles [10–15].

The PCRs can be set up at room temperature.

1. The 100 μL reaction volume contains 1× Phusion™ HF buffer, 100 ng DNA, 300 nM of each primer, 200 μM dNTP, 2 U Phusion™ High-Fidelity DNA Polymerase.
2. The reaction cycle consists of a 1 min initial denaturation step at 98 °C; 10–15 cycles of 10 s at 98 °C, 4 min at 72 °C; a final extension of 5 min at 72 °C.
3. PCRs can be stored at 4 °C overnight.

3.3.3 Digestion of PCR Products with Restriction Endonucleases Without Prior Purification

For separation on sequencing systems, fluorescent, Cy5- or Cy3-labeled primers need to be used. The PCR products are digested using restriction enzymes recognizing four nucleotides when they exceed the system resolution range (generally over 500 bp; Fig. 5). Such enzymes include *Alu*I, *Csp6*I, *Msp*I, *Taï*I, and *Taq*I.

1. Reactions are set up at room temperature. Prepare a 2× master mix for the appropriate number of samples to be amplified.
2. Perform PCR product digestion with *Taï*I restriction enzyme in a 20 μL reaction mixture.
3. To 10 μL PCR products (from IRAP, REMAP, or iPBS amplification), add 10 μL mix of 2× FastDigest® buffer with 1 μL of FastDigest® *Taï*I.
4. Centrifuge all tubes or the plate before starting reaction. Incubate at 65 °C for 30 min.

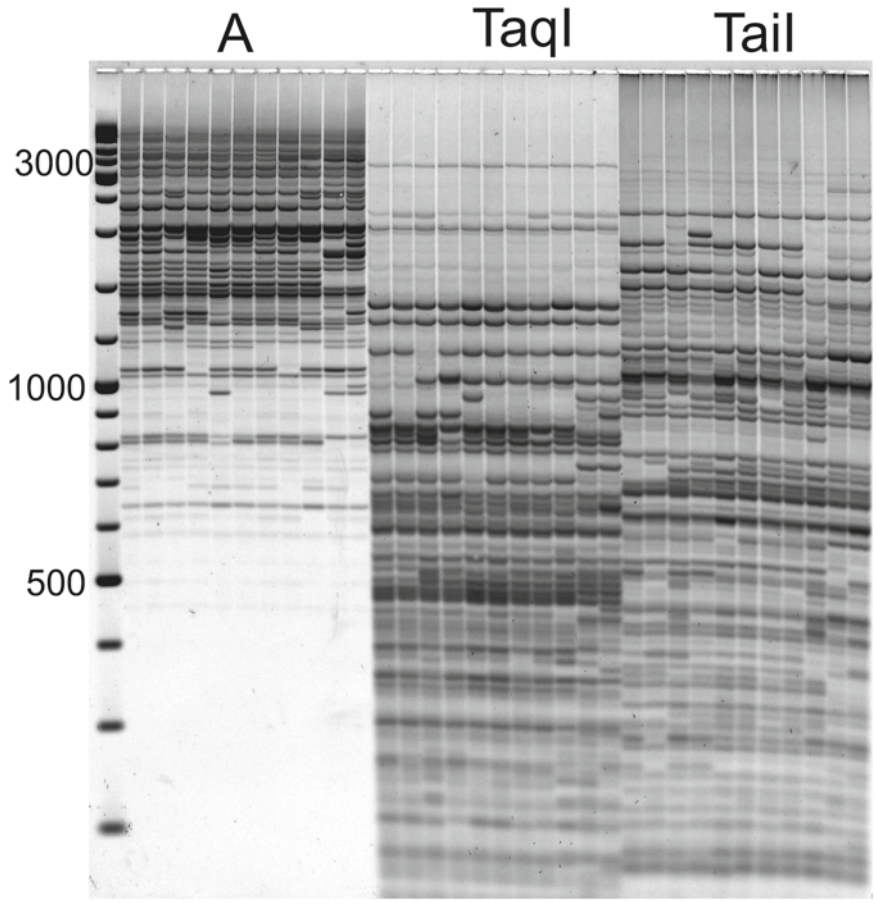

Fig. 5 Digestion of IRAP PCR products with restriction endonucleases. IRAP amplification with *Sukkula* LTR primer (432) is shown (*A*) prior to digestion; (*TaqI*) digestion with *TaqI* and (*TaiI*) with *TaiI* enzymes. A 100 bp DNA ladder is present on the *left*

3.4 Sample Preparation and Loading

Add an equal volume of 2× loading buffer to the completed PCRs in tubes or plates and mix well. Collect the mixture by a short centrifugation (by turning a benchtop microcentrifuge on and immediately off again). Load the gels with a sample volume of 8–10 μL. The DNA concentration plays an important role in gel resolution. Overloaded lanes will result in poor resolution.

3.5 Casting the Agarose Gel

1. Prepare 200 mL of 1.6 % (w/v) agarose containing 1× THE buffer in a 500 mL bottle. This volume is required for one gel with the dimensions 0.4 cm × 20 cm × 20 cm. Dissolve and melt the agarose in a microwave oven. The bottle should be closed, but the plastic cap must not be tightened! The agarose gel must be completely melted in the microwave and then allowed to slowly cool until its temperature drops to about 50–60 °C.

At that point, if desired, add the ethidium bromide solution at a rate of 50 μL per 100 mL, to bring the final concentration to 0.5 μg per mL (alternatively the gel can be stained at the end of the run). Take care not to boil over the agarose. Add ethidium bromide only after removing the agarose from the microwave oven to minimize risks from boilover. The agarose gel must melt and dissolve properly. Small undissolved inclusions will severely hamper the quality of the results. Do not allow the gel to cool unevenly before casting, for example, by leaving it stand on the benchtop or in cool water. The best way to cool the agarose is by shaking it at 37 °C for 15 min. Careful casting of gels is critical to success. Small, undissolved agarose inclusions in the gels will result in bands with spiked smears.

2. Pour the agarose into the gel tray (20×20 cm). Allow the agarose to solidify at room temperature for 1 h minimum. For optimal resolution, cast horizontal gels 3–4 mm thick. The volume of gel solution needed can be estimated by measuring the surface area of the casting chamber and then multiplying by gel thickness.
3. Fill the chamber with 1× THE running buffer until the buffer reaches about 3–5 mm over the surface of the gel.

3.6 Gel Electrophoresis

Select running conditions appropriate to the configuration of your electrophoresis box. For a standard 20×20 cm gel, carry out electrophoresis at a constant 80–100 V for 5–9 h (a total of 700–900 Vh). Electrophoresis may cause the gels to deteriorate after several hours; their temperature should not be allowed to exceed 30 °C, above which electrophoretic resolution will be impaired. Still better results are obtained with a slower run. We routinely use 90 V for 7 h or overnight at 50 V for 14 h (700 Vh). As the end of the run approaches, it is helpful to check the run with a UV transilluminator. For samples with many or large (>500 bp) bands, perform the gel electrophoresis at a constant voltage of 50 V overnight (17 h).

3.7 DNA Visualization

A high-quality gel scanner with good sensitivity and resolution is also very important. Older video systems, which may be suitable for checking the success of restriction digests, cloning reactions, or simple PCRs, are not suitable for analysis of complex banding patterns. DNA can be visualized directly by casting ethidium bromide into gel as described above or by incubating in an ethidium bromide solution of equivalent strength following electrophoresis. The gels are scanned on an FLA-5100 imaging system (Fuji Photo Film GmbH., Germany) or equivalent scanner with a resolution of 50–100 μm, or on a digital gel electrophoresis scanner for detection of ethidium bromide-stained nucleic acids by fluorescence using a second-harmonic-generation (SHG) green laser, 532 nm, or by SYBR Green (Molecular Probes), GelGreen (Biotium), GelStar (Lonza), using a SHG blue laser, 473 nm.

3.8 Cloning PCR Fragment

This protocol is for T/A-end cloning of blunt-ended DNA fragments that are amplified with a proofreading DNA polymerase (*Phusion*™, *Phire*® II, and *Pfu* DNA polymerases). When the PCR mixture contains more than one band of amplified DNA, purify the target fragment by electrophoresis in an agarose gel with SYBR Green I stained DNA. If not purified by gel electrophoresis, PCR-amplified DNA should be prepared for ligation by purification with a QIAEX II Gel Extraction Kit.

1. Non-templated 3′ adenosine is added to the blunt-end PCR fragments by adding 5 U of Taq polymerase per 100 μl reaction volume and 1 μM dATP (a final concentration) directly to PCR mix that is amplified with proofreading DNA polymerase and incubating for 30 min at 72 °C.
2. Mix the samples of DNA with 10× loading buffer and 50× SYBR Green I solution to 1× final concentrations; load them into the slots of the gel. Electrophoresis is performed with 1 % agarose gel in 1× THE buffer and at a constant voltage of 70 V, about 3 h or longer.
3. Following agarose gel electrophoresis, purify the band. Under a Dark Reader, use a sharp scalpel or razor blade to cut out a slice of agarose containing the band of interest and transfer it to a clean, disposable plastic tube.
4. After gel extraction with QIAEX II Gel Extraction Kit, PCR fragments are with TOPO® TA Cloning® Kit.

4 Notes

1. Most enzymes are supplied with their own recommended buffer; these buffers are often suitable for other thermostable polymerases as well. The concentration of $MgCl_2$ ($MgSO_4$) can be varied from 1.5 to 3 mM without influencing the fingerprinting results. A higher $MgCl_2$ concentration can increase the PCR efficiency and allow reduction in the number of PCR cycles from 30 to 28 and also help in PCRs containing somewhat impure DNA. Additional components such as (listed at their final concentration in the reaction buffer) 5 % acetamide, 0.5 M betaine (*N*,*N*,*N*-trimethylglycine), 3–5 % DMSO, 5–10 % glycerol, 5 % PEG 8000, and 5–20 mM TMA (Tetramethylammonium chloride) can increase the PCR efficiency for multiple templates and PCR products [39].
2. The DNA quality is very important, as it is for most PCR-based methods. DNA purification with a spin-column containing a silica-gel membrane is not a guarantee of high DNA quality for all plant samples or tissues. One sign of DNA contamination is that, after some period of time (a month or more) in storage, only short bands can be amplified.

3. The amount of DNA template should be about 1 ng DNA per 1 μL of PCR volume. Much higher DNA concentrations will produce smears between the bands, which is a sign of over-amplification.
4. The result from primer T_m calculation can vary significantly depending on the method used. We have provided a system for T_m calculation and corresponding instructions on the website http://primerdigital.com/tools/ to determinate the T_m values of primers and optional annealing temperature. If using a two-step PCR protocol, where both annealing and extension occur in a single step at 68–72 °C, the primers should be designed accordingly. If necessary, use a temperature gradient to find the optimal temperature for each template-primer pair combination.
5. The optimal annealing temperature (T_a) is the range of temperatures where efficiency of PCR amplification is maximal without nonspecific products. Primers with high T_ms (>60 °C) can be used in PCRs with a wide T_a range compared to primers with low T_ms (<50 °C). The optimal annealing temperature for PCR is calculated directly as the value for the primer with the lowest T_m (T_m^{min}) plus the natural logarithm of fragment size $T_a = T_m^{min} + \ln(L)$, where L is length of PCR fragment [38].
6. Development of a new marker system for an organism in which retrotransposons have not been previously described generally takes about 1 month. The availability of heterologous and conserved primers as well as experience in primer design, sequence analysis, and testing speeds up the development cycle. Routine analysis of samples with optimized primers and reactions may be carried out thereafter. Retrotransposons have several advantages as molecular markers. Their abundance and dispersion can yield many marker bands, the pattern possessing a high degree of polymorphism due to transpositional activity. The LTR termini are highly conserved even between families, yet longer primers can be tailored to specific families. Unlike DNA transposons, the new copies are inserted but not removed. Even intra-element recombination resulting in the conversion of a full-length element to a solo LTR does not affect its performance in IRAP or REMAP. Retrotransposon families may vary in their insertional activity, allowing the matching of the family used for marker generation to the phylogenetic depth required. The primers for different retrotransposons and SSRs can be combined in many ways to increase the number of polymorphic bands to be scored. Furthermore, the length and conservation of primers to the LTRs facilitate cloning of interesting marker bands and the development of new retrotransposons for markers. The IRAP and REMAP fingerprinting patterns can be used in a variety of applications, including measurement

of genetic diversity and population structure [17, 40–43], determination of essential derivation, marker-assisted selection, and recombinational mapping [3, 44]. In addition, the method can be used to fingerprint large genomic clones (e.g., BACs) for the purpose of assembly. The method can be extended, as well, to other prevalent repetitive genomic elements such as MITEs. Although SSAP is somewhat more general than IRAP or REMAP, requiring only a restriction site near the outer flank of a retroelement, its requirement for two additional enzymatic steps introduces the possibility of artifacts from DNA impurities, methylation, and incomplete digestion or ligation. Furthermore, SSAP generally requires selective nucleotides on the 3′ ends of the retrotransposon primers in order to reduce the number of amplification products and increase their yield and resolvability. As for IRAP and REMAP, the resulting subsets of amplifiable products are not additive [18]. The strength of all these methods is that the degree of phylogenetic resolution obtained depends on the history of activity of the particular retrotransposon family being used. Hence, it is possible to analyze both ancient evolutionary events such as speciation and the relationships and similarities of recently derived breeding lines. The IRAP and REMAP can be generalized, furthermore, to other transposable element systems, such as to MITEs, and to other organisms. For example, the SINE element *Alu* of humans has been used in a method called *Alu*-PCR in a way similar to IRAP and REMAP [7, 45, 46].

Acknowledgments

This work was supported by Academy of Finland grant 134079. Anne-Mari Narvanto is thanked for excellent technical assistance.

References

1. Wicker T, Keller B (2007) Genome-wide comparative analysis of *copia* retrotransposons in *Triticeae*, rice, and *Arabidopsis* reveals conserved ancient evolutionary lineages and distinct dynamics of individual *copia* families. Genome Res 17:1072–1081
2. Cheng ZJ, Murata M (2003) A centromeric tandem repeat family originating from a part of Ty3/gypsy-retroelement in wheat and its relatives. Genetics 164:665–672
3. Boyko E et al (2002) Combined mapping of *Aegilops tauschii* by retrotransposon, microsatellite, and gene markers. Plant Mol Biol 48:767–790
4. Lamb JC et al (2007) Plant chromosomes from end to end: telomeres, heterochromatin and centromeres. Curr Opin Plant Biol 10:116–122
5. Meyer W et al (1993) Hybridization probes for conventional DNA fingerprinting used as single primers in the polymerase chain reaction to distinguish strains of *Cryptococcus neoformans*. J Clin Microbiol 31:2274–2280
6. Sivolap IM, Kalendar RN, Chebotar SV (1994) The genetic polymorphism of cereals demonstrated by PCR with random primers. Tsitol Genet 28:54–61
7. Charlieu JP et al (1992) 3′ Alu PCR: a simple and rapid method to isolate human polymorphic markers. Nucleic Acids Res 20: 1333–1337
8. Zietkiewicz E, Rafalski A, Labuda D (1994) Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. Genomics 20:176–183

9. Feschotte C, Jiang N, Wessler S (2002) Plant transposable elements: where genetics meets genomics. Nat Rev Genet 3:329–341
10. Sabot F, Schulman AH (2006) Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome. Heredity 97: 381–388
11. Schulman AH, Kalendar R (2005) A movable feast: diverse retrotransposons and their contribution to barley genome dynamics. Cytogenet Genome Res 110:598–605
12. Hedges DJ, Batzer MA (2005) From the margins of the genome: mobile elements shape primate evolution. Bioessays 27:785–794
13. Ostertag EM, Kazazian HH (2005) Genetics: LINEs in mind. Nature 435:890–891
14. Jurka J (2004) Evolutionary impact of human Alu repetitive elements. Curr Opin Genet Dev 14:603–608
15. Bannert N, Kurth R (2004) Retroelements and the human genome: new perspectives on an old relation. Proc Natl Acad Sci U S A 101:14572–14579
16. Leigh F et al (2003) Comparison of the utility of barley retrotransposon families for genetic analysis by molecular marker techniques. Mol Genet Genomics 269:464–474
17. Kalendar R et al (2010) Analysis of plant diversity with retrotransposon-based molecular markers. Heredity 106:520–530
18. Waugh R 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 253:687–694
19. Vogel JM, Morgante M (1992) A microsatellite-based, multiplexed genome assay, In *Plant Genome III Conference*. San Diego, CA USA
20. Korswagen HC et al (1996) Transposon Tc1-derived, sequence-tagged sites in *Caenorhabditis elegans* as markers for gene mapping. Proc Natl Acad Sci U S A 93: 14680–14685
21. Van den Broeck D et al (1998) *Transposon Display* identifies individual transposable elements in high copy number lines. Plant J 13:121–129
22. Ellis THN et al (1998) Polymorphism of insertion sites of Ty1-copia class retrotransposons and its use for linkage and diversity analysis in pea. Mol Gen Genet 260:9–19
23. Vos P et al (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 21:4407–4414
24. Yu G-X, Wise RP (2000) An anchored AFLP- and retrotransposon-based map of diploid *Avena*. Genome 43:736–749
25. Porceddu A et al (2002) Development of S-SAP markers based on an LTR-like sequence from *Medicago sativa* L. Mol Genet Genomics 267:107–114
26. Lee D et al (1990) A *copia*-like element in *Pisum* demonstrates the uses of dispersed repeated sequences in genetic analysis. Plant Mol Biol 15:707–722
27. Syed NH, Flavell AJ (2006) Sequence-specific amplification polymorphisms (SSAPs): a multilocus approach for analyzing transposon insertions. Nat Protoc 1:2746–2752
28. Pearce SR et al (1999) Rapid isolation of plant Ty1-copia group retrotransposon LTR sequences for molecular marker studies. Plant J 19:711–717
29. Vershinin AV, Ellis TH (1999) Heterogeneity of the internal structure of PDR1, a family of Ty1/copia-like retrotransposons in pea. Mol Gen Genet 262:703–713
30. Ramsay L et al (1999) Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. Plant J 17:415–425
31. Hirochika H, Hirochika R (1993) Ty1-copia group retrotransposons as ubiquitous components of plant genomes. Jpn J Genet 68: 35–46
32. Flavell AJ et al (1992) *Ty1-copia* group retrotransposons are ubiquitous and heterogeneous in higher plants. Nucleic Acids Res 20:3639–3644
33. Ellis THN et al (1998) *Ty1-copia* class retrotransposon insertion site polymorphism for linkage and diversity analysis in pea. Mol Gen Genet 260:9–19
34. Kalendar R et al (2008) *Cassandra* retrotransposons carry independently transcribed 5S RNA. Proc Natl Acad Sci U S A 105:5833–5838
35. Witte CP et al (2001) Terminal-repeat retrotransposons in miniature (TRIM) are involved in restructuring plant genomes. Proc Natl Acad Sci U S A 98:13778–13783
36. Corpet F (1988) Multiple sequence alignment with hierarchical-clustering. Nucleic Acids Res 16:10881–10890
37. Kalendar R, Lee D, Schulman AH (2009) FastPCR software for PCR primer and probe design and repeat search. Genes, Genomes and Genomics 3:1–14
38. Kalendar R, Lee D, Schulman AH (2011) Java web tools for PCR, *in silico* PCR, and oligonucleotide assembly and analysis. Genomics 98:137–144
39. Kovarova M, Draber P (2000) New specificity and yield enhancer of polymerase chain reactions. Nucleic Acids Res 28:E70
40. Baumel A et al (2002) Inter-retrotransposon amplified polymorphism (IRAP), and retrotransposon-microsatellite amplified polymorphism (REMAP) in populations of the young

allopolyploid species *Spartina angelica* Hubbard (Poaceae). Mol Biol Evol 19:1218–1227
41. Boronnikova SV, Kalendar RN (2010) Using IRAP markers for analysis of genetic variability in populations of resource and rare species of plants. Russ J Genet 46:36–42
42. Belyayev A et al (2010) Transposable elements in a marginal plant population: temporal fluctuations provide new insights into genome evolution of wild diploid wheat. Mob DNA 1:6
43. Smýkal P et al (2011) Genetic diversity of cultivated flax (*Linum usitatissimum* L.) germplasm assessed by retrotransposon-based markers. Theor Appl Genet 122:1385–1397
44. Manninen OM et al (2006) Mapping of major spot-type and net-type net-blotch resistance genes in the Ethiopian barley line CI 9819. Genome 49:1564–1571
45. Tang JQ et al (1995) Alu-PCR combined with non-Alu primers reveals multiple polymorphic loci. Mamm Genome 6:345–349
46. Shedlock AM, Okada N (2000) SINE insertions: powerful tools for molecular systematics. Bioessays 22:148–160