# IRAP and REMAP for retrotransposon-based genotyping and fingerprinting

Ruslan Kalendar[1] & Alan H Schulman[1,2]

[1]MTT/BI Plant Genomics Laboratory, Institute of Biotechnology, Viikki Biocenter, University of Helsinki, P.O. Box 56, Viikinkaari 4, FIN-00014 Helsinki, Finland. [2]Plant Genomics, Biotechnology and Food Research, MTT Agrifood Research Finland, Myllytie 10, FIN-31600 Jokioinen, Finland. Correspondence should be addressed to A.H.S. (alan.schulman@helsinki.fi).

Published online 11 January 2007; doi:10.1038/nprot.2006.377



**Retrotransposons can be used as markers because their integration creates new joints between genomic DNA and their conserved ends. To detect polymorphisms for retrotransposon insertion, marker systems generally rely on PCR amplification between these ends and some component of flanking genomic DNA. We have developed two methods, retrotransposon-microsatellite amplified polymorphism (REMAP) analysis and inter-retrotransposon amplified polymorphism (IRAP) analysis, that require neither restriction enzyme digestion nor ligation to generate the marker bands. The IRAP products are generated from two nearby retrotransposons using outward-facing primers. In REMAP, amplification between retrotransposons proximal to simple sequence repeats (microsatellites) produces the marker bands. Here, we describe protocols for the IRAP and REMAP techniques, including methods for PCR amplification with a single primer or with two primers and for agarose gel electrophoresis of the product using optimal electrophoresis buffers and conditions. This protocol can be completed in 1–2 d.**

## INTRODUCTION

Long terminal repeat (LTR) retrotransposons, or Type I transposable elements, replicate by a process of reverse transcription resembling that of the lentiviruses (such as HIV)[1]. The retrotransposons themselves encode the proteins needed for their own replication and integration back into the genome[2]. Their 'copy-and-paste' life cycle means that they are excised in order to insert a copy elsewhere in the genome. Hence, genomes diversify through 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[3]. Human and other mammalian genomes also contain an abundance of retrotransposons. The majority of these, however, are not LTR retrotransposons but long interspersed elements (LINEs) and short interspersed elements (SINEs), which replicate by a somewhat different copy-and-paste mechanism[4,5]. The L1 family of LINE elements and the *Alu* family of SINE elements together comprise roughly 30% of human genomic DNA and nearly 2 million copies[6]. The features of integration activity, persistence, dispersion, conserved structure and sequence motifs, and high copy number together suggest that retrotransposons are highly appropriate genomic features on which to build molecular marker systems.

Marker systems based on transposable elements, in contrast to other methods, detect large changes in the genome. By comparison, restriction fragment length polymorphism (RFLP), single nucleotide polymorphism (SNP) and, to some extent, amplified fragment length polymorphism (AFLP) analyses detect single nucleotide changes that are bidirectional (have a fairly high reversion frequency), whereas simple sequence repeat (SSR) or microsatellite polymorphism analysis tracks the gain or loss of generally less than 20 nucleotides. Microsatellite alleles differ in the number of SSRs they contain and, like single nucleotide changes, also suffer from homoplasy because the number of SSRs can increase or decrease reversibly, making it impossible to distinguish ancestral and derived states.

Retrotransposon-based systems (**Fig. 1**) detect the insertion of elements hundreds to thousands of nucleotides long. 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

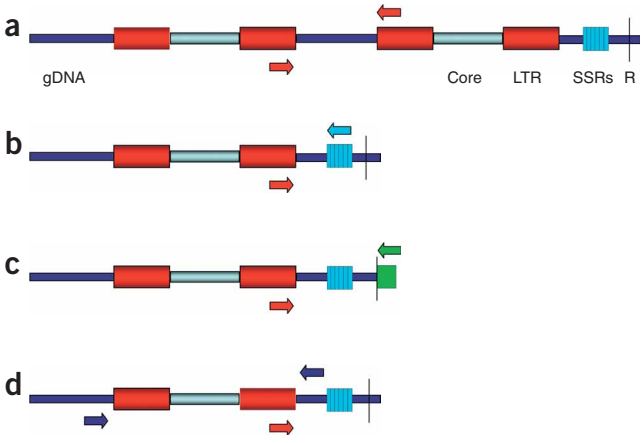


**Figure 1** | Retrotransposon-based marker methods. This shows the genomic features and the positions of the priming sites for the major methods described in the text. (**a**) IRAP. Amplification is carried out between the LTRs of two retrotransposons. Genomic DNA (gDNA) is shown as a solid blue line, primers as arrows above and below the genome segment, and the retrotransposon as comprised of LTRs and a core domain (core). Other features, such as microsatellites (SSRs) or restriction sites (R), may also be present in the genome, but the IRAP method does not take them into account. (**b**) REMAP. Amplification is carried out between primers matching an LTR and a microsatellite domain (SSRs). (**c**) SSAP. Amplification is carried out between primers matching an LTR and a restriction site adaptor ligated to genomic DNA digested with a restriction enzyme. (**d**) RBIP. Full sites (containing a retrotransposon) are generally scored by an amplification reaction with an LTR primer and a primer in flanking, single-copy DNA. Empty sites are scored by amplification between the left and right flanks of the presumptive integration site.

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 that is conserved with a given family of elements but different in other families. The primer is oriented toward the LTR end. The second primer is designed to match some other feature of the genome. The first retrotransposon method described was sequence-specific amplified polymorphism (S-SAP or SSAP) analysis (**Fig. 1c**), where one primer matched the end of the *BARE*-1 retrotransposon of barley and the other matched an AFLP-like restriction site adaptor[7]. The S-SAP method has since been applied using other retrotransposons in barley[8] as well as in other plants including wheat and its relatives[9,10], oat[11], apple[12], artichoke[13], lettuce[14], pea and other legumes[15–17], pepper and tomato[18] and sweet potato[19].

The inter-retrotransposon amplification polymorphism (IRAP; **Fig. 1a**) and retrotransposon-microsatellite amplification polymorphism (REMAP; **Fig. 1b**) methods represent a departure from S-SAP, 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 outward 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 primers. The two primers may be from the same retrotransposon element family or 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 the insertion pattern and by the size of the retrotransposon families chosen for analysis. Furthermore, thousands of products will neither be simultaneously amplifiable to detectable levels nor resolvable 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.

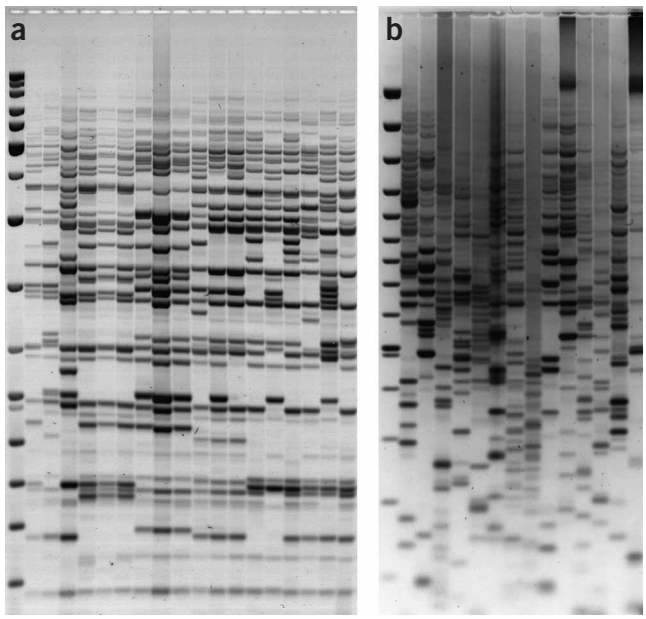


**Figure 2** | IRAP gel fingerprints. This illustrates the results achieved following agarose gel electrophoresis with correct and incorrect conditions. (**a**) Standard amplification, with almost all DNA samples at the same concentration. (**b**) PCR over-amplification, resulting from an overly high primer concentration, too many cycles, too much template or loaded sample, or a combination thereof.

If retrotransposons were fully dispersed within the genome, IRAP would 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. However, IRAP has been successful for all genomes tried to date (see below). This is because retrotransposons generally tend to cluster together in 'repeat seas' surrounding 'genome islands', and may even nest within each other. For example, the *BARE*-1 retrotransposon of barley, an abundant *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[20]. Given a genome of roughly $5 \times 10^9$ 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 *BARE*-1 primers displays a range of products from 100 bp to upward of 10 kb (for example, **Figs. 2** and **3**, below).

The REMAP method is similar to IRAP, but 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[21]. Owing to phenomena including polymerase slippage, microsatellites have high mutation rates and therefore may show much variation at individual loci within a species. Differences in the number of SSR units in a microsatellite are generally detected using primers designed to locate unique

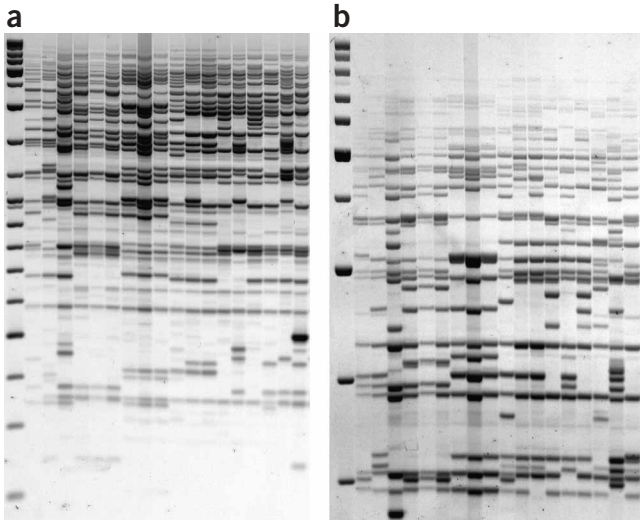


**Figure 3** | IRAP gel fingerprints. This illustrates how to increase the resolution and number of scorable bands by running agarose gel electrophoresis for both short and long periods of time. (**a**) 10 h electrophoresis at 80 V in a 1.7% agarose gel. (**b**) Samples and gel matrix as in **a**, but electrophoresed for 20 h at 70 V.

sequences flanking microsatellites. Alternatively, the stretches of the genome present between two microsatellites may be amplified by ISSR[22], 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 PCR reactions.

Still another retrotransposon-based marker method has been developed in addition to S-SAP, IRAP and REMAP. However, this method, retrotransposon-based insertional polymorphism (RBIP) analysis (**Fig. 1d**)[23,24], is conceptually more similar to the microsatellite method where SSR domain sizes are scored. Unlike the other methods displaying retrotransposon insertion sites, which fingerprint multiple loci simultaneously and anonymously, RBIP types a single locus. RBIP requires knowledge of unique sequences flanking a retrotransposon insertion so that a particular locus can be scored. Hence, development of a set of RBIP markers requires either extensive sequencing of insertion flanks or a fairly large genomic database for primer design. Therefore, it has not been extensively applied beyond *Pisum*, where it was initially described. Its advantage, however, is that it enables codominant scoring: it can detect both the full and empty allelic states for a retrotransposon insertion site. Codominant scoring is very powerful for pedigree reconstruction; tracking of SINE insertions has served to link the cetacean lineage to that of ungulates[25].

The REMAP and IRAP methods require comparatively little sequence information before they can be implemented in a new plant species. The primary requirement is the sequence of an LTR end, either harvested from a database or produced by cloning and sequencing the genomic DNA that flanks conserved segments of retrotransposons[26]. For REMAP, anchored SSR primers can be designed without reference to sequence data and then tested for usefulness. Following their initial description[27], IRAP and REMAP have been applied in species ranging from barley, wheat and their relatives[8,28,29] to oat[30], apple[31], banana[32], citrus[33], grapevine[34], pea[35] and sawgrass (*Spartina*)[36]. It has also been applied to genotyping fungi[37,38]. In addition to these organisms, we have developed and applied IRAP and REMAP in canola and turnip (*Brassica napus* and *Brassica rapa*) as well as in sunflower and *Brachypodium* (unpublished data). The wide applicability of these methods shows that the retrotransposons are organized in other genomes in a manner sufficiently similar to that of the cereals[39], where the methods were originally developed.

The retrotransposon methods described above provide consistent data[40]. Although S-SAP 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, S-SAP 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[8]. Although RBIP confers the power of codominance, developing flanking primers for nested retrotransposon insertions, which can constitute many of the insertions in cereals, is difficult and a method is therefore required for efficient mining of unique flanks. 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 methods can be generalized, furthermore, to other transposable element systems, such as miniature inverted-repeat transposable elements (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[41].



## MATERIALS

### REAGENTS

- Ethidium bromide solution in water, 0.5 mg ml$^{-1}$ **! CAUTION** Ethidium bromide is a carcinogenic agent and is irritating to the eyes, skin, mucous membranes and the upper respiratory tract. Latex gloves should be worn at all times while handling it
- DNA ladder for electrophoresis: GeneRules DNA ladder mix (cat. no. SM1173, Fermentas), 100–10,000-base range, or similar

### EQUIPMENT

- Power supply (minimum 300 V, 400 mA) for electrophoresis
- UV transilluminator with a viewing area of 20 × 20 cm, for visualization of ethidium bromide–stained nucleic acids

### REAGENT SETUP

**Agaroses** RESolute Wide Range (cat. no. 337100, BIOzym, http://www.biozymtc.com/), SERVA Premium (cat. no. 11381), Agarose MP (cat. no. A1091.0250, AppliChem, http://www.applichem.de/) or TopVision CG Agarose (cat. no. R0491, Fermentas, http://www.fermentas.com/). **▲ CRITICAL** Agaroses of the LE type are not effective for fine resolution of fingerprinting bands. Cambrex (http://www.cambrex.com/) NuSieve 3:1 (cat. no. 50094) and MetaPhor (cat. no. 50184) agaroses have low gel strength and make for difficult gel manipulation; 1% agaroses with gel strength >1,700 g cm$^{-2}$ can be used. An agarose mix consisting of 1 part Cambrex MetaPhor and 3 parts SeaKem LE (Cambrex, cat. no. 50004) is suitable for high resolution.

**Loading buffer** Prepare 10× buffer consisting of 30% (wt/vol) Ficoll 400 (Sigma), 100 mM Tris-HCl (pH 8.0), 10 mM EDTA, 0.01% (wt/vol) Orange G and 0.01% xylene cyanol FF in Milli–Q (Millipore) ultrapure deionized water.

**SBTE electrophoresis buffer** Prepare 20× SBTE buffer consisting of 0.4 M Tris-$H_3BO_3$, 0.1 M $Na_2B_4O_7$ and 20 mM EDTA, pH 8.6 in Milli–Q ultrapure deionized water. Although we get best results with STBE, standard TAE and TBE buffers may also be used.

**Thermostable polymerase** Many types and sources of recombinant thermostable polymerases, without 3′-to-5′ exonuclease proofreading activity, exist and could be applied for fingerprinting. Any *Thermus aquaticus* (*Taq*) DNA polymerase is applicable to PCR. We have tested several *Taq* DNA polymerases, including those of Epicentre (MasterAmp Taq DNA Polymerase, http://www.epibio.com/), Promega (http://www.promega.com/), Solis BioDyne (FIREPol, http://www.sbd.ee/en/) and SibEnzyme (http://www.sibenzyme.com/). Other thermostable polymerases, including DyNAzyme II (cat. no. F-501L, Finnzymes Oy, http://www.finnzymes.fi/) and DNA polymerase from *Thermus thermophilus* HB27 (cat. no. 1001, Biotools S.A., http://www.biotools.net/eng/index.htm), were also tested to determine whether the choice of polymerase enzyme has an effect on the products amplified. A polymerase mix consisting of 100 units *Taq* DNA polymerase and 2–4 units *Pfu* DNA polymerase (cat. no. EP0572, Fermentas; cat. no. M7745, Promega) improves amplification of long bands and the accuracy of the PCR.

**PCR reaction buffers** Several PCR reaction buffers are suitable for PCR, including Buffer 1, 1×: 75 mM Tris-HCl (pH 9.0 or 8.8), 2 mM $MgCl_2$, 50 mM KCl, 20 mM $(NH_4)_2SO_4$, 0.01% Tween 20; and Buffer 2, 1×: 10 mM Tris-HCl (pH 8.8), 1.5 mM $MgCl_2$, 50 mM KCl and 0.1% Triton X-100 (or 0.01% Tween 20). ▲ CRITICAL The PCR reaction and its efficiency depend on which buffer-enzyme combination is used. Most enzymes are supplied with their own recommended buffer; these are often suitable for other thermostable polymerases as well. The concentration of $MgCl_2$ 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 32 to 30 and can also help in PCR reactions containing somewhat impure DNA. Additional components such as (listed at their final concentration in the reaction buffer) 5% acetamide (Sigma cat. no. A6082), 0.5 M betaine (*N*,*N*,*N*-trimethylglycine; Sigma cat. no. B-2629), 3% DMSO (Sigma cat. no. D9170), 5% glycerol (Sigma cat. no. G8778), 5% PEG 8000 and 5–20 mM tetramethylammonium chloride (TMA, Sigma cat. no. T3411) can increase the PCR efficiency for multiple templates and PCR products[42].

**EQUIPMENT SETUP**

**Thermal cycler** Use a thermal cycler designed for 0.2-ml tubes or 96-well plates, with a rapid heating and cooling capacity between 4 °C and 99 °C so that the temperature can be changed by 3°C per second; for example, the Mastercycler Gradient (Eppendorf AG, http://www.eppendorf.com) or the PTC-100 Programmable Thermal Controller (MJ Research and Bio-Rad, http://www.bio-rad.com).

**Imaging system** Use a digital gel electrophoresis scanner for detection of ethidium bromide–stained nucleic acids by fluorescence, with a resolution of 50–100μm. Examples include the FLA-5100 imaging system (Fuji Photo Film (Europe) GmbH, http://www.fujifilm.com), or the Fujifilm LAS-3000 Lumi-Imager (CCD chip with 3.2-M pixels). Software such as the Aida Image Analyzer (http://www.raytest.com) and Adobe Photoshop (http://www.adobe.com) is required for image analysis and manipulation.

**Horizontal electrophoresis apparatus** No special cooling system is needed; most commercially available medium- or large-scale horizontal DNA gel electrophoresis systems are suitable, for example from such suppliers as Amersham or Pharmacia. These include the Amersham GNA-200 (http://www1.amershambiosciences.com), Hoefer HE 99X Max Submarine, BioExpress Wide Maxi Horizontal Gel System (cat. no. E-4123-1, http://www.bioexpress.com) and Sigma Maxi-Plus (http://www.sigmaaldrich.com). ▲ CRITICAL Small electrophoresis boxes and short gel trays are not suitable because of the large number of PCR products that need to be resolved. We routinely employ an apparatus with a run length of 20 cm.

**Gel comb** Use a 36-well comb, 1.0 mm in thickness, forming 4.0-mm-wide wells, with a 1.0-mm well spacing. ▲ CRITICAL This comb is ideal for analysis of any PCR amplification product or DNA restriction enzyme digests. The small space between the slots is important for analysis of banding patterns and for comparing lanes across the gel. Using this thickness of comb also improves band resolution.

## PROCEDURE

### Primer design

**1|** Design a PCR primer to match an LTR sequence near to either its 5′ or 3′ end, and orient the primer so that the amplification direction is toward 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 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 outward, 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. This can be tried in a second round of primer design, if the initial primer yields amplification products containing too many weak individual species 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. Microsatellite primers for REMAP or ISSR should be designed according to two principles: first, primer length should be between 19 - 22 bases; second, the last base at 3′-end of the primer is designed as selective base, which is absent in repeat unit itself. Examples of LTR conservation and consequent primer design for LTRs and microsatellites are shown in **Box 1** and **Box 2**.

▲ CRITICAL STEP We have designed LTR primers using the 'FastPCR' software: http://www.biocenter.helsinki.fi/bi/programs/fastpcr.htm. Database searches can sometimes be used to find un-annotated, 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 reverse transcriptase primer binding sites PBS and PPT is needed in order to define the LTR ends with confidence.

## BOX 1 | EXAMPLE OF IRAP AND REMAP PRIMER DESIGN: CONSERVED 3′ END OF *SUKKULA* LARD LTRS AND MATCHING PRIMER

```
TCCATTCTTGCGACACGACGAGATGCGCTTCTATCCCTGACGAGGCCCTCGTACCAAATTGAGGATAGGGTCGCATCTTGGGCGTGACA
TCCATTCTTGCGACAGGACGAGATGCGCTTCTATCCCTGACGAGGCCTTCGTGCCAAATTAAGGATAGGGTCGCATCTTGGGCGTGACA
TCCATTCTTGCGACACGACGAGATGCGCTTCTATCCCTGACGAGGCCTTCGTGCCAAATTGAGGATAGGGTCGCATCTTGGGCGTGACA
TCCATTCTTGCGACACGACGAGATGCGCTTCTATCCCTGACGAGGCCTTCGTGCCAAATTGAGGATAGGGTCGCATCTTGGGCGTGACA
TCCATTCTTGCGACACGACGAGATGCACTTCTATCCCTGTCGAGGCCCTCGTGCCAAAATAAGGATAGGGTCGCATCTTGGGCGTGACA
TCCATTCTTGCGACACGACGAGATGCGCTTCTACCCCTGTCGAGGCCCTCGTGCCAAAATAAGGATAGGGTCGCATCTTGGGCGTGACA
TCCATTCTTGCAACATGACGAGATGCGCTTCTATCCCTGTCGAGGCCCTCGTGCCAAAATAAGGATAGGGTCGCATCTTGGGCGTGACA
TCCATTCTTGCGACACGACGAGATGCGCTTCTATCCCTGACGAGGCCTTCGTGCCAAAGTGAGGATAGGGTCGCATCTTGGGCGTGACA
TCCATTCTTGCGACACGACGAGATGCGCTTCTATCCCTGACGAGGCCCTCGTGCCAGATTGAGGATAGGGCCGCATCTTGGGCGTGACA
TACTTGCTTGTGAAACGCTTAGATGCGCTTCTTTCCTATTCGGGGGCCTCGACCCCCAAATCGGAAAGGACCGCATCTTGGTCATTACA
* * * **** * * * ****** ***** ** ** ** * *** ** * *** *** ********** * * ***
TCCATTCTTGCGACACGACGAGATGCGCTTCTATCCCTGACGAGGCCCTCGTGCCAAATTGAGGATAGGGTCGCATCTTGGGCGTGACA
5′-tagggtcgcatcttgggcgtgaca
```

**Polymerase chain reaction (PCR)** ● TIMING ~ 2–2.5 h

**2|** Perform PCR in a 20-µl reaction mixture containing 20 ng DNA, 75 mM Tris-HCl (pH 9.0), 2 mM $MgCl_2$, 50 mM KCl, 20 mM $(NH_4)_2SO_4$, 0.2 µM primer(s), 200 µM dNTP and 1 U *Taq* polymerase. The standard PCR reaction (120 min) should consist of a 4-min denaturation step at 94 °C; 30–32 cycles of 40 s at 94 °C, 40 s at 60 °C and 120 s at 72 °C; and a 5-min final extension at 72 °C.

▲ CRITICAL STEP The amount of template DNA plays an important role in the quality of the resulting fingerprint. Most commonly, 1 ng DNA per 1 µl of reaction volume is ideal. Much higher DNA concentrations will produce smears between the bands, which is a sign of over-amplification.

▲ CRITICAL STEP DNA and primers should be stored in a 1× TE solution (10 mM Tris-HCl, 1 mM EDTA, pH 8.0).

▲ CRITICAL STEP The final primer concentration(s) in the reaction can vary from 200 to 400 nM. Although higher primer concentrations increase PCR efficiency and the rapidity of DNA amplification, they also produce over-amplified products, such as are shown in **Figure 2b**.

▲ CRITICAL STEP DNA quality is very important for obtaining good results. Standard DNA extraction methods are sufficient to yield DNA of high quality from most of samples. DNA should be free of polysaccharides, pigments and secondary metabolites. Some plant materials contain polysaccharides, pigments, oils or polyphenols, which can reduce the efficiency of PCR. Furthermore, contaminated DNAs will decline in PCR performance during prolonged (1 month or more) periods of storage, as a result of chemical modification. Such DNAs (for example, from *Brassica* spp.) should be extracted with methods involving sodium hydroxide, followed by two ethanol DNA precipitations. 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.

▲ CRITICAL STEP PCR thermal conditions can be varied without large effects on the resulting band pattern. The denaturation step in PCR can be carried out at 94 °C for 40 s or 98 °C for 10 s. The length of the annealing step can vary from 30 to 60 s at 60 °C. The annealing temperature varies with the melting temperature of the primer; it should be between 55 °C and 68 °C (60 °C is optimal for almost all primers and their combinations in IRAP and REMAP).

■ PAUSE POINT PCR reactions can be stored at 4 °C overnight with or without loading buffer.

## BOX 2 | MICROSATELLITE PRIMER EXAMPLES

Two-base-repeat microsatellites:
$(CT)_n$ microsatellites: 5′$(CT)_{10}$G, 5′$(CT)_{10}$T or 5′$(CT)_{10}$A
$(CA)_n$ microsatellites: 5′$(CA)_{10}$G, 5′$(CA)_{10}$T or 5′$(CA)_{10}$A
$(TG)_n$ microsatellites: 5′$(TG)_{10}$G, 5′$(TG)_{10}$C or 5′$(TG)_{10}$A
$(AG)n$ microsatellites: 5′$(AG)_{10}$G, 5′$(AG)_{10}$C or 5′$(AG)_{10}$T
$(AC)_n$ microsatellites: 5′$(AC)_{10}$G, 5′$(AC)_{10}$C or 5′$(AC)_{10}$T
Three-base-repeat microsatellites:
$(CTC)_n$ microsatellite: 5′$(CTC)_6$G, 5′$(CTC)_6$T or 5′$(CTC)_6$A
$(GTG)_n$ microsatellite: 5′$(GTG)_6$G, 5′$(GTG)_6$T or 5′$(GTG)_6$A
$(CAC)_n$ microsatellite: 5′$(CAC)_6$G, 5′$(CAC)_6$T or 5′$(CAC)_6$A
$(ACC)_n$ microsatellite: 5′$(ACC)_6$G, 5′$(ACC)_6$T or 5′$(ACC)_6$C
$(TCG)_n$ microsatellite: 5′$(TCG)_6$G, 5′$(TCG)_6$C or 5′$(TCG)_6$A

**Casting the agarose gel** ● TIMING ~2-3 h

**3|** Prepare 200 ml of 1.6–2% (wt/vol) agarose containing 1× STBE 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 60–65 °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 0.25 µg per ml (alternatively the gel can be stained at the end of the run as discussed in Step 8).

! CAUTION Take care not to boil over the agarose. Add ethidium bromide only after removing the agarose from the microwave oven to minimize risks from boil-over.

▲ CRITICAL STEP 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.

**4|** Pour the agarose into the gel tray (20 × 20 cm). Allow the agarose to solidify at room temperature for one hour minimum.

▲ CRITICAL STEP 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 this by the gel thickness.

**5|** Fill the chamber with 1× STBE running buffer until the buffer reaches about 3 mm above the surface of the gel.

**Sample preparation and loading** ● TIMING ~15 min

**6|** Add an equal volume of 2× loading buffer to the completed PCR reactions in tube or plate, 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 10 µl.

▲ CRITICAL STEP The DNA concentration has an important role in gel resolution. Overloaded lanes will result in poor resolution.

**Gel electrophoresis** ● TIMING ~5–10 h

**7|** Select running conditions appropriate to the configuration of your electrophoresis box. For a standard 20 × 20–cm gel, carry out electrophoresis at a voltage not exceeding a constant 80–85 V for 7 h (a total of 595 volt-hours). Electrophoresis may cause the gels to heat 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 50 V for 12 h (600 volt-hours). As the end of the run approaches, it is helpful to check the run with a UV transilluminator. Look to see if the target range of fragment sizes has been resolved and is in the desired gel position.

▲ CRITICAL STEP For samples with many or large (>1,000 bp) bands, perform the gel electrophoresis at a constant voltage of 70 V overnight (20 h) as shown in **Figure 3**. A useful guide for gel architectures differing from 20 × 20 cm is to first calculate the total volt-hours required for 500-bp fragments to migrate to the bottom of the gel and then adjust the voltage for subsequent runs so that the run time to achieve that total is at least 12 h.

**DNA visualization** ● TIMING ~15 min

**8|** DNA can be visualized directly by casting ethidium bromide into the gel as described above, or by incubating the gel in an ethidium bromide solution of equivalent strength following electrophoresis. Scan the gels on an FLA-5100 imaging system (Fuji Photo Film (Europe) GmbH) or equivalent scanner with a resolution of 50–100 μm.

● TIMING

Primer design (Step 1): ~1–2 h
Polymerase chain reaction (PCR; Step 2): ~2–2.5 h
Casting the agarose gel (Steps 3–5): ~2–3 h
Sample preparation and loading (Step 6): ~15 min
Gel electrophoresis (Step 7): ~5–10 h
DNA visualization (Step 8): ~15 min

**? TROUBLESHOOTING**

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

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 (such as Qiagen, http://www1.qiagen.com/) 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. Careful casting of gels is critical to success. Small, undissolved agarose inclusions in the gels will result in bands with spiked smears. Finally, a high-quality gel scanner with good sensitivity and resolution is also very important. Older still-video systems, which may be suitable for checking the success of restriction digests, cloning reactions or simple PCR reactions, are not suitable for analysis of complex banding patterns.

## ANTICIPATED RESULTS

Development of a new marker system for an organism in which retrotransposons have not been previously described generally takes 1–6 months. 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, with the pattern possessing a high degree of polymorphism owing 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 retrotransposon 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 matching 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[36,43], determination of essential derivation, marker-assisted selection and recombinational mapping[7,8,29,34,44]. In addition, the method can be used to fingerprint large genomic clones (for example, bacterial artificial chromosomes (BACs)) for the purpose of assembly. It can be extended, as well, to other prevalent repetitive genomic elements such as MITEs.

**ACKNOWLEDGMENTS** A.-M. Narvanto and S.U. Lönnqvist are thanked for their excellent technical assistance. The methods described here have been developed under grants from the Academy of Finland (ESGEMO, 6302016) and from the EU (TEBIODIV, GEDIFLUX, MMEDV), together with grants from CIMO of Finland, and in projects sponsored by Boreal Plant Breeding Ltd.

**COMPETING INTERESTS STATEMENT** The authors declare that they have no competing financial interests.

Published online at http://www.natureprotocols.com


1. Feschotte, C., Jiang, N. & Wessler, S. Plant transposable elements: where genetics meets genomics. *Nat. Rev. Genet.* **3**, 329–341 (2002).
2. Sabot, F. & Schulman, A.H. Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome. *Heredity* **97**, 381–388 (2006).
3. Schulman, A.H. & Kalendar, R. A movable feast: diverse retrotransposons and their contribution to barley genome dynamics. *Cytogenet. Genome Res.* **110**, 598–605 (2005).
4. Hedges, D.J. & Batzer, M.A. From the margins of the genome: mobile elements shape primate evolution. *Bioessays* **27**, 785–794 (2005).
5. Ostertag, E.M. & Kazazian, H.H. Genetics: LINEs in mind. *Nature* **435**, 890–891 (2005).
6. Jurka, J. Evolutionary impact of human Alu repetitive elements. *Curr. Opin. Genet. Dev.* **14**, 603–608 (2004).
7. Waugh, R. *et al.* 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 (1997).
8. Leigh, F. *et al.* Comparison of the utility of barley retrotransposon families for genetic analysis by molecular marker techniques. *Mol. Genet. Genomics* **269**, 464–474 (2003).
9. Queen, R.A., Gribbon, B.M., James, C., Jack, P. & Flavell, A.J. Retrotransposon-based molecular markers for linkage and genetic diversity analysis in wheat. *Mol. Genet. Genomics* **271**, 91–97 (2004).
10. Nagy, E.D., Molnar, I., Schneider, A., Kovacs, G. & Molnar-Lang, M. Characterization of chromosome-specific S-SAP markers and their use in studying genetic diversity in Aegilops species. *Genome* **49**, 289–296 (2006).
11. Yu, G.-X. & Wise, R.P. An anchored AFLP- and retrotransposon-based map of diploid *Avena*. *Genome* **43**, 736–749 (2000).
12. Venturi, S., Dondini, L., Donini, P. & Sansavini, S. Retrotransposon characterisation and fingerprinting of apple clones by S-SAP markers. *Theor. Appl. Genet.* **112**, 440–444 (2006).
13. Lanteri, S. *et al.* A first linkage map of globe artichoke (*Cynara cardunculus* var. *scolymus* L.) based on AFLP, S-SAP, M-AFLP and microsatellite markers. *Theor. Appl. Genet.* **112**, 1532–1542 (2006).
14. Syed, N.H. *et al.* A detailed linkage map of lettuce based on SSAP, AFLP and NBS markers. *Theor. Appl. Genet.* **112**, 517–527 (2006).
15. Jing, R. *et al.* Insertional polymorphism and antiquity of PDR1 retrotransposon insertions in *Pisum* species. *Genetics* **171**, 741–752 (2005).
16. Porceddu, A. *et al.* Development of S-SAP markers based on an LTR-like sequence from *Medicago sativa* L. *Mol. Genet. Genomics* **267**, 107–114 (2002).
17. Ellis, T.H.N., Poyser, S.J., Knox, M.R., Vershinin, A.V. & Ambrose, M.J. *Ty1-copia* class retrotransposon insertion site polymorphism for linkage and diversity analysis in pea. *Mol. Gen. Genet.* **260**, 9–19 (1998).
18. Tam, S.M. *et al.* Comparative analyses of genetic diversities within tomato and pepper collections detected by retrotransposon-based SSAP, AFLP and SSR. *Theor. Appl. Genet.* **110**, 819–831 (2005).
19. Tahara, M. *et al.* Isolation of an active element from a high-copy-number family of retrotransposons in the sweetpotato genome. *Mol. Genet. Genomics* **272**, 116–127 (2004).
20. Vicient, C.M. *et al.* Retrotransposon *BARE*-1 and its role in genome evolution in the genus *Hordeum*. *Plant Cell* **11**, 1769–1784 (1999).
21. Ramsay, L. *et al.* Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. *Plant J.* **17**, 415–425 (1999).
22. Zietkiewicz, E., Rafalski, A. & Labuda, D. Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics* **20**, 176–183 (1994).
23. Flavell, A.J., Knox, M.R., Pearce, S.R. & Ellis, T.H.N. Retrotransposon-based insertion polymorphisms (RBIP) for high throughput marker analysis. *Plant J.* **16**, 643–650 (1998).
24. Flavell, A.J. *et al.* A microarray-based high throughput molecular marker genotyping method: the tagged microarray marker (TAM) approach. *Nucleic Acids Res.* **31**, e115 (2003).
25. Shedlock, A.M. & Okada, N. SINE insertions: powerful tools for molecular systematics. *Bioassays* **22**, 148–160 (2000).
26. Pearce, S.R. *et al.* Rapid isolation of plant Ty1-copia group retrotransposon LTR sequences for molecular marker studies. *Plant J.* **19**, 711–717 (1999).
27. Kalendar, R., Grob, T., Regina, M., Suoniemi, A. & Schulman, A.H. IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. *Theor. Appl. Genet.* **98**, 704–711 (1999).
28. Boyko, E., Kalendar, R., Korzun, V., Gill, B. & Schulman, A.H. Combined mapping of *Aegilops tauschii* by retrotransposon, microsatellite, and gene markers. *Plant Mol. Biol.* **48**, 767–790 (2002).
29. Manninen, O., Kalendar, R., Robinson, J. & Schulman, A.H. Application of *BARE*-1 retrotransposon markers to map a major resistance gene for net blotch in barley. *Mol. Gen. Genet.* **264**, 325–334 (2000).
30. Tanhuanpää, P. *et al.* Generation of SNP markers for short straw in oat (*Avena sativa* L.). *Genome* **49**, 282–287 (2006).
31. Antonius-Klemola, K., Kalendar, R. & Schulman, A.H. TRIM retrotransposons occur in apple and are polymorphic between varieties but not sports. *Theor. Appl. Genet.* **112**, 999–1008 (2006).
32. Teo, C.H. *et al.* Genome constitution and classification using retrotransposon-based markers in the orphan crop banana. *J. Plant Biol.* **48**, 96–105 (2005).
33. Breto, M.P., Ruiz, C., Pina, J.A. & Asins, M.J. The diversification of *Citrus clementina* Hort. ex Tan., a vegetatively propagated crop species. *Mol. Phylogenet. Evol.* **21**, 285–293 (2001).
34. Pereira, H.S., Barao, A., Delgado, M., Morais-Cecilio, L. & Viegas, W. Genomic analysis of Grapevine Retrotransposon 1 (Gret 1) in *Vitis vinifera*. *Theor. Appl. Genet.* **111**, 871–878 (2005).
35. Smykal, P. Development of an efficient retrotransposon-based fingerprinting method for rapid pea variety identification. *J. Appl. Genet.* **47**, 221–230 (2006).
36. Baumel, A., Ainouche, M., Kalendar, R. & Schulman, A.H. Inter-retrotransposon amplified polymorphism (IRAP), and retotransposon-microsatellite amplified polymorphism (REMAP) in populations of the young allopolyploid species *Spartina angelica* Hubbard (Poaceae). *Mol. Biol. Evol.* **19**, 1218–1227 (2002).
37. Chadha, S. & Gopalakrishna, T. Retrotransposon-microsatellite amplified polymorphism (REMAP) markers for genetic diversity assessment of the rice blast pathogen (*Magnaporthe grisea*). *Genome* **48**, 943–945 (2005).
38. Díez, J., Béguiristain, T., Le Tacon, F., Casacuberta, J.M. & Tagu, D. Identification of Ty1-copia retrotransposons in three ectomycorrhizal basidiomycetes: evolutionary relationships and use as molecular markers. *Curr. Genet.* **43**, 34–44 (2003).
39. Schulman, A.H., Gupta, P.K. & Varshney, R.K. Organisation of retrotransposons and microsatellites in cereal genomes. in *Cereal Genomics* (eds. P.K. Gupta & R.K. Varshney) 83–118 (Kluwer, Dordrecht, The Netherlands, 2004).
40. Schulman, A.H., Flavell, A.J. & Ellis, T.H.N. The application of LTR retrotransposons as molecular markers in plants. *Methods Mol. Biol.* **260**, 145–173 (2004).
41. Tang, J.Q. *et al.* Alu-PCR combined with non-Alu primers reveals multiple polymorphic loci. *Mamm. Genome* **6**, 345–349 (1995).
42. Kovarova, M. & Draber, P. New specificity and yield enhancer of polymerase chain reactions. *Nucleic Acids Res.* **28**, E70 (2000).
43. Kalendar, R., Tanskanen, J., Immonen, S., Nevo, E. & Schulman, A.H. Genome evolution of wild barley (*Hordeum spontaneum*) by *BARE*-1 retrotransposon dynamics in response to sharp microclimatic divergence. *Proc. Natl. Acad. Sci. USA* **97**, 6603–6607 (2000).
44. Asins, M.J. *et al.* QTL analysis of citrus tristeza virus-citradia interaction. *Theor. Appl. Genet.* **108**, 603–611 (2004).