npg

REVIEW

Analysis of plant diversity with retrotransposon-based molecular markers

R Kalendar[1], AJ Flavell[2], THN Ellis[3], T Sjakste[4], C Moisy[1] and AH Schulman[1,5]

[1]MTT/BI Plant Genomics Laboratory, Institute of Biotechnology, University of Helsinki, Helsinki, Finland; [2]Division of Plant Sciences, University of Dundee at SCRI, Dundee, UK; [3]John Innes Centre, Colney Lane, Norwich, UK; [4]Genomics and Bioinformatics Laboratory, Institute of Biology of the University of Latvia, Salaspils, Latvia and [5]Biotechnology and Food Research, MTT Agrifood Research Finland, Jokioinen, Finland

Retrotransposons are both major generators of genetic diversity and tools for detecting the genomic changes associated with their activity because they create large and stable insertions in the genome. After the demonstration that retrotransposons are ubiquitous, active and abundant in plant genomes, various marker systems were developed to exploit polymorphisms in retrotransposon insertion patterns. These have found applications ranging from the mapping of genes responsible for particular traits and the management of backcrossing programs to analysis of population structure and diversity of wild species. This review provides an insight into the spectrum of retrotransposon-based marker systems developed for plant species and evaluates the contributions of retrotransposon markers to the analysis of population diversity in plants.
Heredity (2011) **106**, 520–530; doi:10.1038/hdy.2010.93; published online 4 August 2010

Introduction

Markers have a key role in the study of genetic variability and diversity, in the construction of linkage maps and in the tracking of individuals or lines carrying particular genes. The emergence of marker systems has closely followed developments in biochemistry and molecular biology for the past 40 years (Hubby and Lewontin, 1966). The shortcomings of biochemically derived markers, such as isozymes, drove the development of markers based on DNA polymorphisms (Kan and Dozy, 1978). A DNA molecular marker in essence detects nucleotide sequence variation at a particular location in the genome. The variation must be found between the parents of the chosen cross for the marker to be informative among their offspring and to allow its pattern of inheritance to be analyzed. DNA markers can generate 'fingerprints,' which are distinctive patterns of DNA fragments resolved by electrophoresis and detected by staining or labeling. The advent of the PCR was a breakthrough for molecular marker technologies, and made possible many fingerprinting methods. These fall into two broad categories, namely methods that detect single loci and multiplex methods that detect multiple loci simultaneously.

The first multiplex methods to be developed were named randomly amplified polymorphic DNA (Williams *et al.*, 1990; Welsh and McClelland, 1990) and DNA amplification fingerprinting (Caetano-Anollés *et al.*, 1991), respectively, and involve amplification of random repetitious sites in the genome using short primers, typically 8–12 nt in length. The approaches involve quick and easy reaction set-up and no genome sequence information is needed to design the primers. However, problems in reproducibility due to the presence of many potential priming sites in the genome and the low annealing temperatures in the reactions, because of the nature of the primers themselves, have led to the disappearance of these systems from the molecular marker toolkit today. The amplified fragment length polymorphism (AFLP) method, introduced in the mid-1990s, also generates anonymous markers. It detects restriction sites by amplifying a subset of all the sites for a given enzyme pair in the genome by PCR between ligated adapters (Vos *et al.*, 1995).

Interspersed repetitive sequences comprise a large fraction of the genome of many eukaryotic organisms and they predominantly consist of transposable elements (TEs). It is therefore not surprising that many DNA marker techniques that are based on these repeats have been devised. In an early example, restriction fragment length polymorphism (RFLP) probes derived from repetitive sequences were hybridized to Southern blots of restriction-digested genomic DNA to produce a highly variable pattern (Lee *et al.*, 1990). The RFLP technique was used extensively in the past, but has been replaced by PCR-based methods because of the slowness of Southern blotting.

Nucleotide sequences matching repetitive sequences showing polymorphism in RFLP analyses have also been used as PCR primers for the inter-repeat amplification polymorphism marker method (Meyer *et al.*, 1993; Salimath *et al.*, 1995). Such repetitive sequences include

Correspondence: *Professor AH Schulman, MTT/BI Plant Genomics Laboratory, Institute of Biotechnology, University of Helsinki, PO Box 65, Viikinkaari 1, Helsinki, FIN-00014, Finland.*
E-mail: alan.schulman@helsinki.fi

microsatellites, such as $(TG)_n$ or $(AC)_n$, which are distributed throughout the genome. A derived approach was developed to generate PCR markers based on amplification of microsatellites near the 3′ end of the *Alu* (short interspersed repetitive element (SINE)) TEs, called Alu-PCR or SINE-PCR (Chariieu *et al.*, 1992).

TEs are divided into two major classes. Class II transposons, which were first discovered by McClintock (1984), move by a cut-and-paste mechanism as double-stranded DNA. In contrast, Class I retrotransposons transpose through an RNA intermediate, and hence the original copy remains in the genome (Finnegan, 1989). Retrotransposons are separated into two major sub-classes that differ in their structure and transposition cycle. These are the long terminal repeat (LTR) retro-transposons and the non-LTR retrotransposons (LINE and SINE elements), which are distinguished by the respective presence or absence of LTRs) at their ends. All groups are complemented by their respective non-autonomous forms that lack one or more of the genes essential for transposition: MITEs (miniature inverted-repeat terminal elements) for class II, SINEs for non-LTR retrotransposons and TRIMs (terminal-repeat retrotransposons in miniature) and LARDs (large retrotransposon derivatives) for LTR retrotransposons.

LTR retrotransposons and genome organization

Retrotransposons are abundant throughout the genomes of virtually all eukaryotes and are particularly numerous in plants (Finnegan, 1989; Flavell *et al.*, 1992; Voytas *et al.*, 1992; Suoniemi *et al.*, 1998). In plants, the LTR retro-transposons are typically more plentiful and active than their non-LTR relatives (see, for example, Arabidopsis Genome Initiative, 2000; Rice Chromosome 10 Sequencing Consortium, 2003; Hill *et al.*, 2005; Macas *et al.*, 2007; Paterson *et al.*, 2009). In many crop plants, between 40 and 70% of the total DNA comprises LTR retro-transposons (Pearce *et al.*, 1996; SanMiguel *et al.*, 1996; Shirasu *et al.*, 2000). Although most prevalent retro-transposons are dispersed throughout the genome, at least in the cereals and citrus they are often locally nested one into another and in extensive domains that have been referred to as 'retrotransposon seas' surrounding gene islands (SanMiguel *et al.*, 1996; Shirasu *et al.*, 2000; Ramakrishna *et al.*, 2002; Bernet and Asins, 2004; Gu *et al.*, 2004; Kong *et al.*, 2004; Sabot *et al.*, 2005). Bertioli *et al.* (2009) have shown that retrotransposon-deficient gene islands correspond to highly conserved blocs in legume evolution.

LTR retrotransposons are transcribed from one LTR of an integrated element to produce a nearly full-length RNA transcript containing a single copy of the LTR split between its two ends (the LTR provides both the start site and polyadenylation signal for the element; Figure 1). This RNA is then reverse-transcribed into an extrachro-mosomal complementary DNA, reconstituting the full-length element that is ultimately integrated back into the genome. Immediately internal to the LTRs are the priming sites for reverse transcription. The large central part of the retrotransposon encodes the structural components of a virus-like particle into which the RNA is inserted, together with reverse transcriptase and integrase enzymes.

The structural features described above, as well as the basic stages of the life cycle, are shared by the retro-transposons and the retroviruses (Frankel and Young, 1998; Kim *et al.*, 2004; Wicker *et al.*, 2007). However, rather than escaping the genome to infect new individuals as do retroviruses, retrotransposons insert the new copies only into their host genomes. If the integration takes place within a cell lineage from which pollen or egg cells are ultimately derived, then a new polymorphism is contributed to the gene pool.

Retrotransposon-based marker systems

General considerations: The dynamism and dispersion of the various groups of TEs have led to their widespread exploitation as molecular markers. Direct comparisons of retrotransposon methods with AFLP indicate that the
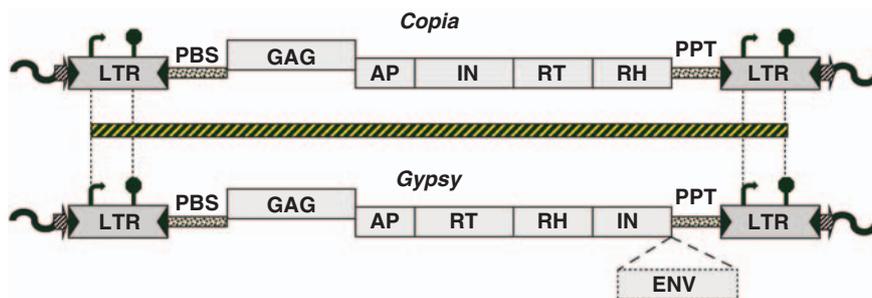

Figure 1 Organization of the LTR retrotransposon genome. The order of coding domains differs between the *Copia* and *Gypsy* superfamilies as shown. Retrotransposons are bounded by long terminal repeats (LTRs), which contain the transcriptional promoter and terminator (indicated diagrammatically by a bent arrow and stop sign, respectively). The resultant transcript is indicated as a hatched box between the *Gypsy* and *Copia* diagrams. The LTRs contain short inverted repeats at either end (shown as filled triangles). Reverse transcription is primed at the PBS and PPT domains, respectively, for the (−) and (+) strands of the complementary DNA (cDNA). The internal region of the retrotransposon codes for the proteins necessary for the retrotransposon life cycle and is generally divided into two open reading frames: GAG, for the capsid protein, which packages the transcript into a virus-like particle, and POL, for the other proteins. The POL contains: aspartic proteinase (AP), which cleaves the polyprotein; integrase (IN), which inserts the cDNA copy into the genome; reverse transcriptase (RT) and RNaseH (RH), which together copy the transcript into cDNA. An additional open reading frame for the envelope protein (*ENV)*, found in some groups of *Gypsy* elements, is indicated. The LTRs are generally well conserved within families, and can serve for the design of primers to generate DNA footprints (Figure 2). Direct repeats in the flanking genomic DNA are generated upon retrotransposon integration: these are depicted as short, hatched arrows. The flanking genomic DNA is shown as a wavy line. The apposition of a long element bearing conserved sequences within genomic DNA of random sequence is the basis for retrotransposon marker methods.

522

retrotransposon markers are more informative in a variety of crops (Waugh *et al.*, 1997; Ellis *et al.*, 1998; Yu and Wise, 2000; Porceddu *et al.*, 2002; Tam *et al.*, 2005). Polymorphism in AFLP markers is based on single-nucleotide polymorphism (SNP) changes in restriction sites and on indels throughout the genome, whereas retrotransposon markers incorporate integration events in the generation of detectable variation. Hence, it seems that retrotransposon insertions tend to be more dynamic than SNPs and indels as a whole.

Unlike Class II TEs, retrotransposons cannot excise themselves from their insertion locations (Finnegan, 1989). This unidirectionality of integration confers great advantages in reconstructing pedigrees and phylogenies because the ancestral state is obvious—it is the empty site, whereas for almost all other genetic polymorphisms upon which markers are built, directionality cannot be inferred. In this way, phylogeny can be explored using retrotransposon insertions; for example, SINE elements have been used to trace human roots to Africa (Batzer *et al.*, 1994), to establish the relationship of whales to even-toed ungulates (Shimamura *et al.*, 1997) and to infer the evolutionary relationships between wild rice species (Cheng *et al.*, 2002).

Most of the retrotransposon marker methods take advantage of two basic properties, namely that they cause large insertions by their transpositional activity and they contain conserved domains from which PCR primers can be designed. Some other methods target the small insertions and deletions found within otherwise conserved TE domains to generate fingerprints. Most of the techniques are also anonymous, producing fingerprints from multiple sites of retrotransposon insertion in the genome (Schulman *et al.*, 2004) by using PCR primed on conserved motifs in the element and on some widespread and conserved motif in the surrounding DNA. For LTR retrotransposons, the primers are generally designed from the LTRs near to the insertion site, in LTR sub-domains that are conserved within retrotransposon families and differ between families (Figure 2). Although regions internal to the LTR containing conserved segments can be used for this purpose, generally the LTRs are chosen to minimize the size of the target to be amplified and to assay insertion site polymorphism rather than events internal to the element.

The various methods for revealing insertion polymorphism of retrotransposons differ in the nature of the feature external to the TE that is used for primer design (Figure 2; Schulman *et al.*, 2004). Because the LTRs are direct repeats, a primer facing outward from the left or 5′ LTR will necessarily face inward from the right, or 3′ LTR. Depending on the nature of the second primer, the inward facing primer will either not amplify a product or produce a band from the TE interior that is typically of low usefulness because it represents the total TE subfamily rather than a particular copy. The internal amplicon can be removed by judicious use of an infrequently cutting enzyme or designing the primer to overlap the LTR end and adding bases that do not match the LTR-interior junction (Waugh *et al.*, 1997; Vershinin *et al.*, 2003). For retrotransposons with relatively short LTRs the transposon-specific primer can derive from an internal sequence present only once per element, simplifying this process (Ellis *et al.*, 1998).
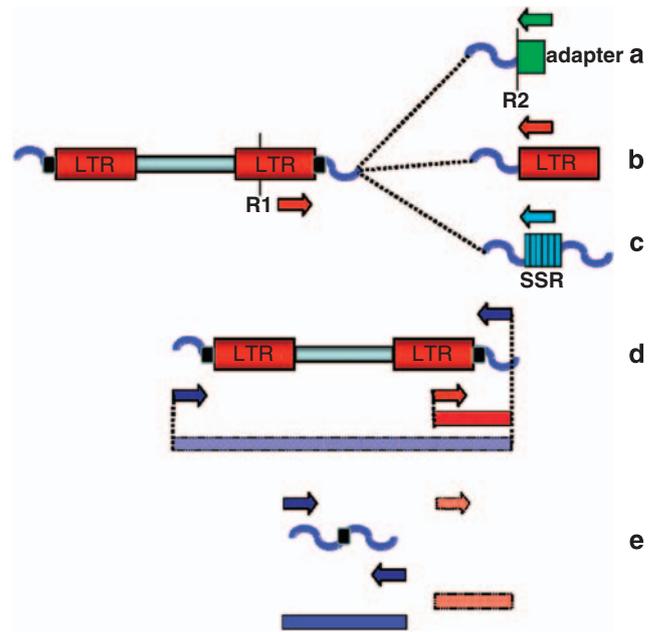
Figure 2 Retrotransposon-based molecular marker methods. (**a**–**c**) Alternative priming sites in the genome paired with a priming site in a retrotransposon. (**a**) The SSAP method. Amplification is carried out from genomic DNA cut with two restriction enzymes (R1 and R2), containing a retrotransposon and ligated to an adapter (shown only for R2). Primers are indicated as arrows; the LTR generally serves as the retrotransposon priming site. (**b**) The IRAP method. The second priming site is also a retrotransposon. (**c**) The REMAP method. Amplification takes place between a microsatellite domain (labeled simple sequence repeat (SSR)) and a retrotransposon, using a primer anchored to the proximal side of the microsatellite and a retrotransposon primer. (**d**, **e**) RBIP. (**d**) Full sites are scored by amplification between a primer in the flanking genomic DNA (shown as a blue wavy line) and a retrotransposon primer. The single product is shown as a red bar beneath the diagram. The alternative reaction between the primers for the left and right flanks (light blue bar beneath the diagram) is inhibited in the occupied site by the length of the retrotransposon. (**e**) The flanking RBIP primers are able to amplify the empty site, depicted as a deep blue bar beneath the diagram, but amplification from the retrotransposon primer does not occur (missing product shown as a light red bar) because the TE insert is missing.

Below, we present an overview of the various retrotransposon-based molecular marker methods that have been developed and discuss their use to visualize the genetic diversity generated by TE activity.

Sequence-specific amplified polymorphism (SSAP)/transposon display

Waugh *et al.* (1997) exploited the dispersion and prevalence of the *BARE*1 LTR retrotransposon in barley through modification of the AFLP technique (Figure 2). Their method is based on the digestion of genomic DNA with two different enzymes to generate a template for amplification between retrotransposons and adaptors ligated at the restriction sites (usually *Mse*I and *Pst*I although any pair of restriction enzymes could in principle be used), using selective bases in the adaptor primer (Syed and Flavell, 2007). Usually, SSAP shows more polymorphism and more co-dominance than AFLP. Nevertheless, SSAP also requires restriction digestion of genomic DNA to provide sites for adapter ligation as

does the AFLP method. The sensitivity of commonly used restriction enzymes to DNA methylation could provide false genotyping results and needs to be taken into account in experimental design; nevertheless, this can also be used as a design to sample DNA methylation.

In SSAP, nonselective primers can be used when enzymes used for digestion cut infrequently, or when the copy number of the TE is low. For high-copy-number families, the number of selective bases may be increased (Schulman et al., 2004). The use of two enzymes in SSAP correspondingly reduces genomic complexity, as does the use of selective bases on the primers associated with the adapters. TEs with low numbers of copies are not well suited to methods to reduce genomic complexity. However, the use of single enzyme digests with a systematic series of all possible selective bases allows the survey of all insertion sites for a given TE.

SSAP has been optimized for multiple plant species and protocols for rapidly obtaining retrotransposon sequence information for SSAP primer design have been developed (Syed and Flavell, 2007; Kalendar et al., 2010). The same technique was named transposon display when applied to DNA transposons rather than retrotransposons (van den Broeck et al., 1998). Rim2/Hipa transposon display produced highly polymorphic profiles with ample reproducibility within a species as well as between species in the Oryza genus (Kwon et al., 2005).

Inter-retrotransposon amplified polymorphism (IRAP)

As discussed above, although retrotransposons are dispersed, they can also be found clustered in the genome. It is the phenomenon of clustering that makes possible the IRAP method, which detects insertional polymorphisms by amplifying the portion of DNA between two retrotransposons (Figure 2; Kalendar et al., 1999; Kalendar and Schulman, 2006). If all retrotransposons were dispersed equally throughout the genome, even for an abundant family such as BARE1, individual elements would be 50 kb apart, and could not yield IRAP amplification templates.

A virtue of IRAP is its experimental simplicity. All that is needed is simple PCR followed by electrophoresis to resolve the PCR products. 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. Nearby TEs may be found in different orientations in the genome (head-to-head, tail-to-tail or head-to-tail) increasing the range of tools available to detect polymorphism depending on the method and primer combinations. If two primers are used, they may be from the same retrotransposon 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 pattern obtained will be related to the TE copy number, insertion pattern and size of the TE family.

IRAP fingerprints with single primers often generate bands from 500 to 3000 bases, lengths that are not convenient for capillary electrophoresis. To reduce the size of the DNA products to be separated and visualized, fluorescent primers may be used in the PCR reaction and the amplicon DNA digested with a four-base-specific restriction enzyme such as TaiI or TaqI after the PCR

reaction. In this way, IRAP can be adapted to analyses on capillary sequencing platforms.

Retrotransposon-microsatellite amplified polymorphism (REMAP)

The REMAP method is similar to IRAP, but one of the two primers matches a microsatellite motif (Figure 2; Kalendar et al., 1999; Kalendar and Schulman, 2006). Abundant in most genomes, microsatellites or simple sequence repeats seem to be associated with retrotransposons and have high mutation rates due to polymerase slippage. Therefore, they may show much variation at individual loci within a species (Schulman et al., 2004). In REMAP, anchor nucleotides are used at the 3′ end of the simple sequence repeat primer to avoid slippage of the primer between the individual simple sequence repeat motifs. An anchored primer also prevents the detection of variation in repeat numbers within the microsatellite.

Banding patterns are completely different if REMAP primers are used individually or in combination, indicating that the majority of bands are derived from sequences bordered by a microsatellite on one side, and by an LTR on the other. Usually, the REMAP pattern is more variable than the corresponding inter-simple sequence repeat pattern; and often (but not always, depending on the LTR sequence) the IRAP pattern with primer combinations shows more variability than with a single primer (Leigh et al., 2003; Kalendar et al., 1999, 2004; Kalendar and Schulman, 2006).

Retrotransposon-based insertional polymorphism (RBIP)

In addition to SSAP, IRAP and REMAP, a fourth method based on the polymorphic integration pattern of retrotransposons, RBIP, has been developed (Flavell et al., 1998). RBIP is the sole retrotransposon method designed to detect polymorphism for the integration of an element at a particular locus (Figure 2). The RBIP method uses primers flanking retrotransposon insertions and scores the presence and absence of insertions at individual sites. The method has also been called insertion sequence-based polymorphism (Paux et al., 2010).

Using three primers, RBIP yields co-dominant marker scores, which are particularly useful for phylogenetic studies because retrotransposon insertions are irreversible. In the case of a retrotransposon, a primer designed in the LTR is used together with a primer designed in the flanking region to allow the amplification of an insertion site, whereas primers specific to both the 5′ and 3′ flanking regions are used to score the corresponding empty site. TE insertions are usually more than thousands of bases long, and hence the flanking primers do not generate an amplicon from the occupied site. Hence, RBIP detects both the presence and absence of the insertion but requires that the sequence of the 5′ and 3′ flanking sequences of the TE insertions are known. RBIP has been used in rice (Vitte et al., 2004) to address the issue of the evolution of Indica and Japonica rice varieties.

Tagged microarray marker (TAM)

The basic RBIP method has been developed for high-throughput applications by replacing gel electrophoresis with array hybridization to a filter (Flavell et al., 1998; Jing et al., 2007). Initially, PCR reactions detecting the occupied
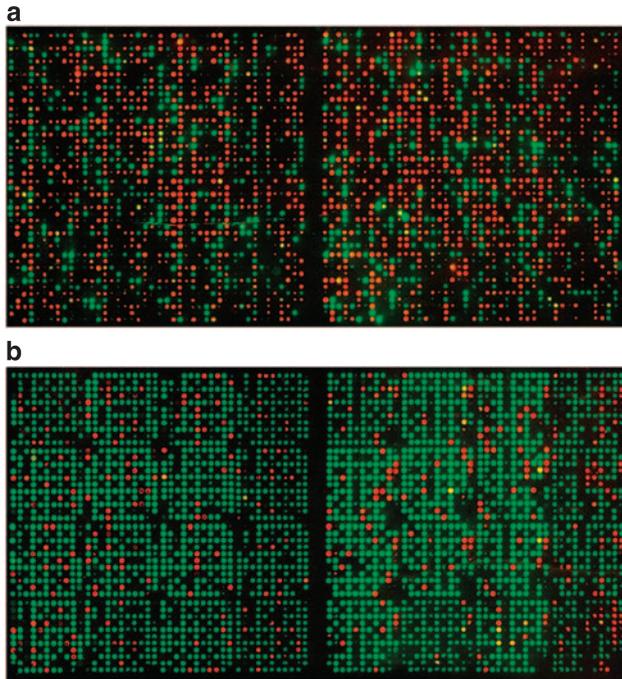
Figure 3 TAM fingerprinting of two RBIP markers. A total of 3263 *Pisum* lines were scored for the RBIP markers (**a**) Birte-B1 and (**b**) 1794-2 (Jing *et al.*, 2005) by the TAM approach (Flavell *et al.*, 2003; Jing *et al.*, 2007, 2010). Each spot represents a single sample (sample locations in the array are conserved between slides) and in these two cases a red spot indicates an occupied (retrotransposon insertion present) locus and the green spot an unoccupied locus. Yellow spots indicate an individual heterozygous for the retrotransposon insertion.

sites and unoccupied sites carried out together were spotted onto membranes, and probed with a locus-specific probe. TAM is an extension of this to a microarray format (Jing *et al.*, 2007). TAM based on the *PDR1*, *Cyclops* and *Tpv* LTR retrotransposons of pea has been developed for scoring thousands of DNAs for a co-dominant molecular marker on a glass microarray slide (Figure 3).

Scaling the dot blot approach to microarrays has given attendant advantages in throughput, efficiency and data collection. RBIP also works well with SNP markers (Flavell *et al.*, 2003; Jing *et al.*, 2007). In this approach, biotin-terminated allele-specific PCR products are spotted unpurified onto streptavidin-coated glass slides and visualized by hybridization of fluorescent detector oligo-nucleotides to tags attached to the allele-specific PCR primers. Two tagged primer oligonucleotides are used per locus and each tag is detected by hybridization to form a concatameric DNA probe labeled with multiple copies of a fluorochrome. This method has recently been used to study the diversity of a complete pea germplasm collection containing thousands of samples (Jing *et al.*, 2010). The DArT (diversity array technology) array (Jaccoud *et al.*, 2001), similar to TAM, is also array based. In contrast to TAM, it is an anonymous method (the sequences of the loci are not known beforehand) and scores multiple loci per array for relatively few accessions.

Retrotransposon UTL polymorphism fingerprinting (RUP)
A multilocus molecular marker based on length variability within the internal untranslated leader (UTL)

region named RUP has been described (Pelsy, 2007). It uses PCR amplification between primers from the LTR and the *gag* domain that lies near the 5′ LTR. Unlike the previously described methods, which visualize insertional polymorphisms, RUP detects UTL size variation in a variable region within individual members of a particular retrotransposon family. The RUP method shows these size differences as a fingerprint. The individual UTL size classes are stable enough to permit phylogenetic resolution of the plant accessions in which they are found. Pelsy (2007) used RUP to describe a genotype specific to each of the 94 *Vitis* accessions analyzed. Previously, a minimum of six standard microsatellite markers chosen for their high degree of allelic polymorphism were necessary to identify a grape variety (This *et al.*, 2004).

In a related approach, Vershinin and Ellis (1999) used primers throughout the internal region of the pea retrotransposon *PDR1* and compared this method with SSAP, which assesses insertion site polymorphism in pea. The two methods gave similar results especially when comparing the *gag* domain with the insertion sites. Nevertheless, the analysis of internal domain markers in addition to the insertional polymorphisms gave a more detailed picture of the evolution of element families in germplasm as well as their population history.

Development of retrotransposon marker projects
A major disadvantage of all the methods described above is the need for retrotransposon sequence information to design family-specific primers. However, related species have similar TE sequences (retroelements or transposons), meaning that primers for the anonymous marker methods described above (SSAP, IRAP and REMAP) from one species can be used in another. In our experience, LTR primers can be readily used across species lines, among closely related genera and even sometimes between plant families (Lou and Chen, 2007; Sanz *et al.*, 2007; Figure 4). In this case, primers designed to conserved TE sequences are advantageous. Moreover, TEs are dispersed throughout the genome and often interspersed with other elements and repeats. By combining PCR primers from different classes of repeats and families of LTRs, PCR fingerprints can be improved.

Deployment of a retrotransposon marker system into a species, in which the methods have not been previously used, requires PCR primers that recognize a retrotransposon and, in the case of RBIP, the flanking sequences. The retrotransposon targets that can be amplified by heterologous primers developed in a different species tend to be members of old families of elements present before the divergence of the plant clades in question. Jing *et al.* (2005) estimated the average age of segregating retrotransposon insertion sites in *Pisum* as being approximately 2 Myr. This result is roughly similar to estimates made by SanMiguel *et al.* (1998) and Vitte *et al.* (2004) in maize and rice, respectively, but may be biased toward younger elements because structural disruptions make old insertions harder to characterize. Nevertheless, these ages are comparable to divergence times between some closely related species (or even the genera *Homo* and *Pan*), suggesting that some retroelements may be useful in recently diverged clades.
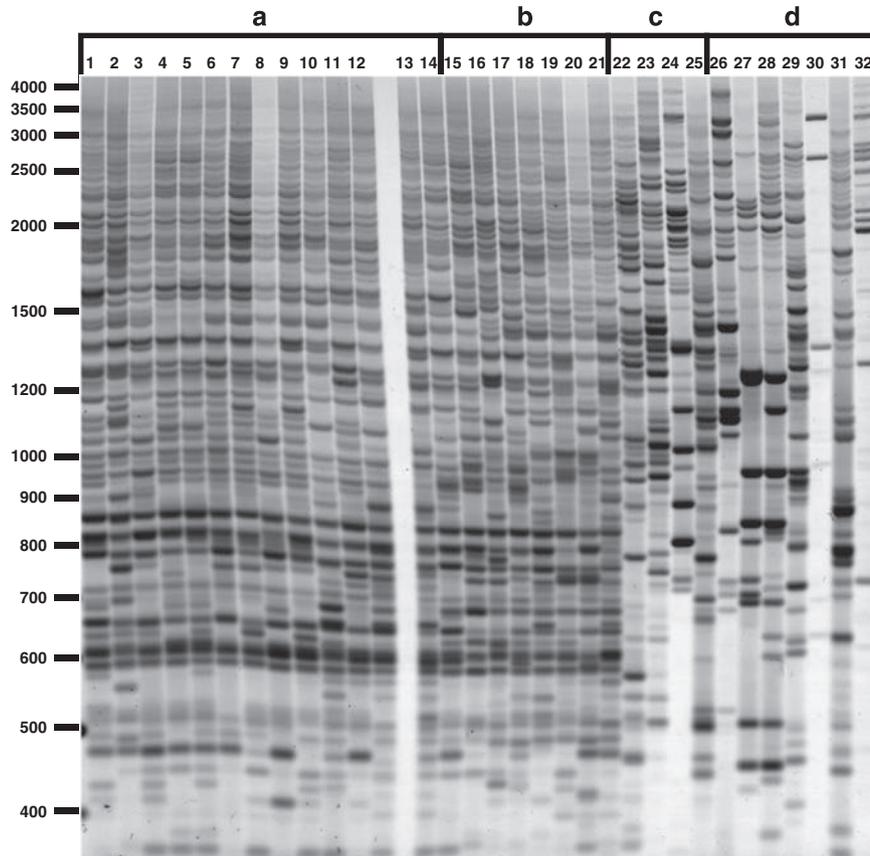
Figure 4 IRAP fingerprints for Triticeae species. A barley *BARE1* LTR primer (5′-GCCTCTAGGGCATAATTCCAACAC-3′) served in the reaction. (**a**) *Hordeum vulgare* cultivars and landraces: 1, Rolfi; 2, CI 9819; 3, Pallidum107; 4, Odesskij 31; 5, Odesskij 17; 6, Sonja; 7, Sultan; 8, Ingri; 9, Beka; 10, Djau Kabutak; 11, W1991; 12, 408; 13, 688; and 14, 1354. (**b**) *Hordeum spontaneum* lines: 15, T1 (Turkey); 16, T11(Turkey); 17, J31(Jordan); 18, IN68 (Iran); 19, IN80 (Iran); 20, IS112 (Israel); and 21, IS147 (Israel). (**c**) Other *Hordeum* species: 22, *H. murinum* ssp. *glaucum*; 23, *H. brachyantherum* ssp. *californicum*; 24, *H. erectifolium*; and 25, *H. marinum* ssp. *gussoneanum*. (**d**) Other Triticeae: 26, *Aegilops peregrina*; 27, *Triticum diccocoides*; 28, *Triticum aestivum* (cv. Bogdarka); 29, *Psathyrostachys fragilis* ssp. *fragilis*; 30, *Phleum pratense*; 31, *Avena sativa*; and 32, *Secale strictum*. Marker sizes in bp are indicated on the left axis.

A balance between monomorphism, allowing the alignment of fingerprint patterns between samples, and polymorphism, yielding information on relationships between samples, needs to be achieved. Phylogenetic analysis can only be made on individual fingerprinting bands of the same size that correspond to the same sequence and loci. Within a species, bands of the same length and amplification intensity usually contain the same sequence. Between species, it is best to confirm the identity for at least some of the bands to be scored. Similarly, for RBIP and TAM, the possibility that insertion flanks may not be single copy within the genome, or may not correspond to the same linkage group in interspecific comparisons, needs to be considered.

The IRAP, REMAP and SSAP methods are all dominant, as are randomly amplified polymorphic DNAs, restriction-fragment length polymorphism and AFLP, although some implementations permit the identification of heterozygotes by measuring amplicon amounts (Knox et al., 2009). For dominant retrotransposon markers, the absence of an amplicon may be the consequence of mutation at the locus carrying the insertion. The mutation could affect the binding site for the retrotransposon primer or, in SSAP, lead to the gain or loss of a restriction site. In practice, this problem does not arise in the application of retrotransposon markers to segregating populations generated by deliberate crosses, because the alleles are determined by the parents. Similarly, alleles that share a recent common ancestor will be unlikely to carry second site mutations. Although it may be possible to map two polymorphisms to the same site, and show by sequencing that they correspond to two allelic states differing by the insertion of a retrotransposon, this is very tedious in practice. It may also be possible to detect the presence of one or two doses of a retrotransposon marker band (Knox et al., 2009), thereby scoring both heterozygotes and homozygotes, although band intensity may also vary because of the reaction conditions.

In general, if one wishes to study closely related varieties or breeding lines, one should develop a retrotransposon marker system based upon the most polymorphic TE available. This process begins with amplification and sequencing of variable regions close to the outer termini of the TE, development of primers specific for the retrotransposon families found and testing these for their efficacy as markers (Pearce et al., 1999; Jing et al., 2005; Kalendar et al., 2010). It may be necessary to clone and sequence hundreds of clones to obtain a few good primers. In practice, up to one person-year of time is needed to develop and apply a fully

functioning novel retrotransposon-based marker system in a new species. However, this is a one-time investment that can be applied thereafter to the corresponding species and its close relatives.

The most polymorphic retrotransposons are likely to include those that are currently active. To identify these, one could amplify and sequence unconserved regions between conserved domains (for example, within the integrase or reverse transcriptase domains) in RNA, and then use a primer from the unconserved region and an adapter primer in a genome-walking approach to isolate the corresponding LTR (Kalendar et al., 2010).

The most general approach for acquiring new TE sequence information is to use shotgun sequencing of the whole genome, for example, with a 454 Genome Sequencer FLX Instrument (454 Life Sciences, Branford, CT, USA), followed by clustering of the repeats (Macas et al., 2007). Putative LTRs may then be identified by the presence of PBS and PPT motifs, with primers designed to match the most prevalent groups of similar putative LTRs. Given the long contiguous sequences spanning retrotransposons, active ones will likely contain open reading frames (indicating translational capacity), identical LTRs (indicating recent integration) and identical target site duplications (indicating recent integration). Alignments of elements meeting these criteria in comparison to those that do not may allow the design of primers specific to active retrotransposons for use in marker methods.

In principle, a marker from any of the multilocus, anonymous systems (SSAP, IRAP and REMAP) can be converted into a corresponding RBIP marker and vice versa. Markers from the former methods are easy to harvest and they can be quickly examined for their informativeness before taking on the investment of developing a corresponding RBIP marker. SSAP, IRAP and REMAP bands are derived from one side of a retrotransposon insertion and sequencing of them enables the design of a flanking genome PCR primer, provided that the sequence is not repetitive and therefore unusable. However, the genomic sequence flanking the other side of the element needs to be found to score the empty site. This can be obtained by screening germplasm accessions that are polymorphic for the original band, and then carrying out a SSAP reaction on these, in which the LTR primer is replaced with a primer designed to the known flank that is facing toward the insertion site (Jing et al., 2005).

As stated above, the only true co-dominant retrotransposon marker system is RBIP. But even here, the independent detection of the occupied and empty sites is really the detection of two cosegregating dominant alleles linked in repulsion. RBIP requires appreciable investment to develop for a new species, and is difficult in cases in which most retrotransposon insertions are nested, because of the issue with flanking repetitious DNA mentioned above. In the one situation in which it has been fully tested, pea, 80 functioning RBIP markers were developed and applied to ca 3000 lines in 6 years.

One way to overcome the dominant nature of the other marker systems is to use genetically homozygous material. For mapping populations, this can be achieved using double-haploid or recombinant inbred lines. In principle, the same result could be achieved by mapping in haploid tissues, for example, in the endosperm of gymnosperms. Populations consisting of double-haploid lines have been used in a wide variety of mapping efforts, using a diversity of marker systems after the early publications on their advantages (Choo, 1981; Dunn et al., 1991). The efficacy of double-haploid populations for the mapping of retrotransposon markers, and in the mapping of genes with retrotransposon markers, has been well established (Waugh et al., 1997; Manninen et al., 2000). The availability of methods for gametophytic embryogenesis goes hand in hand with effective deployment of retrotransposon markers in mapping efforts.

In species in which wide crosses are used to introduce novel traits, dominant markers developed from the wild species are useful in subsequent backcrosses needed to generate material for further breeding (Messmer et al., 1999). Wide crosses are similar, regarding the introgressed chromosome segment, to the genomic constitution of allopolyploid species, such as groundnut, in which the two constituent genomes have different retroelement populations. In this case, dominant markers corresponding to each genome behave in a co-dominant manner (Bertioli et al., 2009).

Application of retrotransposons to the analysis of plant populations

Quantitative genome diversification

Retrotransposons are known to be activated by biotic stresses such as pathogen attack (Grandbastien et al., 1997) and abiotic stress such as drought (Kalendar et al., 2000) as well as by tissue culture (Hirochika et al., 1996) and genome methylation status (Kaeppler et al., 2000; Liu and Wendel, 2000; Kubis et al., 2003). Retrotransposon transcriptional activation will lead to an increase in copy number and genome size if the newly transposed copies survive selection. These new copies will contribute to insertion site polymorphisms that can be detected using the methods described above. However, because of practical limitations on amplifying and showing every genomic copy, generally only a subset of the retrotransposon insertions can be surveyed by any given marker method. Quantitative methods offer an overall view of the effect of retrotransposon insertion on the genome by estimating total copy number. The accumulation of non-excising retrotransposon insertions seems to be a process that would lead to an interminable increase in retrotransposon number. However, insertions are lost from genomes by recombination and from populations by genetic drift or active selection. If the rates of gain and loss of TE insertions are balanced, then copy number will be stable (The International Brachypodium Initiative, 2010).

Measurement of retrotransposon copy number at first was based on Southern hybridization and its variants, dot and slot blotting (Vershinin et al., 1990; Pearce et al., 1996; Vicient et al., 1999; Kalendar et al., 2000, 2004). However, hybridization-based quantification is sensitive to small nuances in the protocol and to the affinity between a probe and its target sequence. Several factors need to be carefully controlled, including DNA loading, DNA hybridization efficiency, probe concentration and labeling and the linearity of signal detection, to avoid under- or overestimation in the number of copies.

An alternative to blotting is real-time PCR, an approach that allows quantification of PCR products during the exponential phase of amplification. In plants, the first time real-time PCR method was successfully used by Soleimani et al. (2006) for quantification of BARE1 in five cultivars of barley (Hordeum vulgare) and two accessions of its wild relative, Hordeum spontaneum. Examining the LTR and reverse transcriptase domains revealed significant differences in BARE1 copy number among cultivars, providing further evidence that BARE1 is active and has a major role in shaping the barley genome as a result of breeding and selection. More than 90% of the elements were truncated (solo LTRs), suggesting high levels of recombination. Intracultivar variation in BARE1 content was found in this study to be statistically insignificant. A real-time PCR approach was adopted and optimized by Pagnotta et al. (2009) to estimate and compare the copy number of WIS2-1A and BARE1 retrotransposons in Triticum and Aegilops species. Great variation was detected in copy number both within and among species. A nonlinear relationship was found between the copy number of retrotransposons and ploidy level. Both these studies provide significant insights into the microevolution of the repetitive part of genome.

Analysis of plant population structure

It has been well established that various TE families have been active in different temporal periods through evolution. This means that TE marker systems based on different TEs can show different levels of resolution and can be chosen to fit with the required analysis (Pearce et al., 2000; Leigh et al., 2003; Schulman and Kalendar, 2005; Teo et al., 2005; Antonius-Klemola et al., 2006; Vukich et al., 2009). Recently active retrotransposons with a short half-life for the persistence of individual genomic insertions are best suited for comparisons of breeding lines, whereas older families with slower turnover are better for analyses at the species and genus levels. We have described earlier, in the section 'Development of retrotransposon marker projects', how to go about identifying currently or recently active retrotransposon families.

On the genus level, examination of the SSAP insertion patterns of eight retrotransposon families in ten Vitis accessions show that these families are present across the Vitis genus and only a few insertion sites are fixed in all accessions, which seem to have been maintained during speciation (Moisy et al., 2008). Most of the scored bands are polymorphic, indicating that these families have been active after speciation across the genus. IRAP and REMAP have been used to track genome evolution in sawgrass (Spartina spp.) after hybridization (Baumel et al., 2002), to analyze the genome constitution in various banana (Musa) polyploids (Teo et al., 2005), and to examine genomic stability in Helianthus (Vukich et al., 2009).

All of the available retrotransposon marker systems have been used to examine genome evolution within species. Various combinations of SSAP, IRAP and REMAP has been used to measure diversity, similarity and cladistic relationships within Pisum (Pearce et al., 2000; Smýkal, 2006), Hordeum (Kalendar et al., 2000; Vicient et al., 2001), Citrus (Bretó et al., 2001), Malus (Antonius-Klemola et al., 2006), Oryza (Branco et al.,

2007), Cucumis (Lou and Chen, 2007) and Aegilops (Nagy et al., 2006; Saeidi et al., 2008). The marker systems have also been applied to plant pathogens, such as the rice blast pathogen (Magnaporthe grisea spp.; Chadha and Gopalakrishna, 2005). The locus-specific RBIP method has been used to analyze evolutionary history in Pisum (Flavell et al., 1998; Vershinin et al., 2003; Jing et al., 2005, 2010) and Oryza (Vitte et al., 2004).

Retrotransposon insertional polymorphism is sufficiently great to support not only analyses on the whole-genome level within species, but also gene mapping projects within the generally narrower germplasm of cultivated varieties (Queen et al., 2004). The anonymous methods (SSAP, IRAP and REMAP) have been used for variety fingerprinting and to develop genetic maps in barley (Waugh et al., 1997), pea (Ellis et al., 1998), bread wheat and its wild relatives (Gribbon et al., 1999), oat (Yu and Wise, 2000), Medicago sativa L. (Porceddu et al., 2002), tomato (Tam et al., 2005), apple (Venturi et al., 2006), globe artichoke (Lanteri et al., 2006) and lettuce (Syed et al., 2006). These methods have also been applied in gene mapping projects in barley (Manninen et al., 2000), oat (Tanhuanpää et al., 2006, 2007) and in the D-genome of bread wheat (Aegilops tauschii Coss.; Boyko et al., 2002).

Conclusions

Markers based on LTR retrotransposons, in one or other of the manifestations described above, often generically referred to as 'transposon display,' have come of age since their introduction over 13 years ago. At least 99 studies using these marker systems had been published by the end of 2009, covering the gamut from cereals and grasses to cashew and coconut, tomato and pepper, multiple legumes species, fungi, birds and insects. The applications range from investigations of retrotransposon activation and mobility to studies of biodiversity and genome evolution, to the mapping of genes and the estimation of genetic distance, to assessment of essential derivation of varieties and detection of somaclonal variation and to food traceability and purity. Because LTR retrotransposons (or their relatives, the endogenous retroviruses) are ubiquitous, these methods are generic.

Similar approaches have been applied to the non-LTR retrotransposons in the plants and animals, in particular to the SINE elements (Cheng et al., 2002, 2003; Prieto et al., 2005). The insertion pattern of Alu, a SINE and the most prevalent transposable element in the human genome has been especially useful. It has not only served as a tool in many studies of human population structure and origins (Batzer et al., 1994; Watkins et al., 2003), but has also been linked to various heritable diseases (Deininger and Batzer, 1999; Jurka, 2004). In other mammals, the SINEs have served to determine the relationship of whales to even-toed ungulates (Shimamura et al., 1997) and, for plants, to clarify the relationships between wild rice species (Cheng et al., 2003).

Recently, commercial platforms for SNP detection (for example, Illumina, San Diego, CA, USA) have been developed and have garnered much popularity for major crops, domestic animals and humans. Development of SNPs depends on having abundant sequence data. The costs of acquiring these data, as well as of applying commercial assays, represent a barrier for research on

underfunded tropical crops and wild species. Furthermore, evolutionary studies with SNPs are affected by the problems of homoplasy in SNP state, the lack of neutrality of genic markers and the uneven chromosomal distribution of the highly expressed genes that are used to generate SNPs. Although genetic analysis by shotgun sequencing remains a tantalizing possibility, the cost is still prohibitive. For these reasons, cheap, generic, easily applied retrotransposon marker systems will remain a viable choice for genetic markers for the foreseeable future.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

References

Antonius-Klemola K, Kalendar R, Schulman AH (2006). TRIM retrotransposons occur in apple and are polymorphic between varieties but not sports. *Theor Appl Genet* **112**: 999–1008.

Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**: 796–815.

Batzer MA, Stoneking M, Alegria-Hartman M, Bazan H, Kass DH, Shaikh TH et al. (1994). African origin of human-specific polymorphic *Alu* insertions. *Proc Natl Acad Sci USA* **91**: 12288–12292.

Baumel A, Ainouche M, Kalendar R, Schulman AH (2002). Retrotransposons and genomic stability in populations of the young allopolyploid species *Spartina anglica* C.E. Hubbard (Poaceae). *Mol Biol Evol* **19**: 1218–1227.

Bernet GP, Asins MJ (2004). Identification and genomic distribution of gypsy like retrotransposons in *Citrus* and *Poncirus*. *Theor Appl Genet* **108**: 121–130.

Bertioli DJ, Moretzsohn MC, Madsen LH, Sandal N, Leal-Bertioli SC, Guimarães PM et al. (2009). An analysis of synteny of *Arachis* with *Lotus* and *Medicago* sheds new light on the structure, stability and evolution of legume genomes. *BMC Genomics* **10**: 45.

Boyko E, Kalendar R, Korzun V, Korol A, Schulman AH, Gill BS (2002). A high-density cytogenetic map of the *Aegilops tauschii* genome incorporating retrotransposons and defense related genes: insights into cereal chromosome structure and function. *Plant Mol Biol* **48**: 767–790.

Branco CJ, Vieira EA, Malone G, Kopp MM, Malone E, Bernardes A et al. (2007). IRAP and REMAP assessments of genetic similarity in rice. *J Appl Genet* **48**: 107–113.

Bretó MP, Ruiz C, Pina JA, Asíns MJ (2001). The diversification of *Citrus clementina* Hort. ex Tan., a vegetatively propagated crop species. *Mol Phylogenet Evol* **21**: 285–293.

van den Broeck D, Maes T, Sauer M, Zethof J, De Keukeleire P, D'hauw M et al. (1998). Transposon Display identifies individual transposable elements in high copy number lines. *Plant J* **13**: 121–129.

Caetano-Anollés G, Bassam BJ, Gresshoff PM (1991). DNA amplification fingerprinting using very short arbitrary oligonucleotide primers. *Biotechnology* **9**: 553–557.

Chadha S, Gopalakrishna T (2005). Retrotransposon-microsatellite amplified polymorphism (REMAP) markers for genetic diversity assessment of the rice blast pathogen (*Magnaporthe grisea*). *Genome* **48**: 943–945.

Chariieu J-P, Laurent A-M, Carter DA, Bellis M, Roizeès G (1992). 3′ *Alu* PCR: a simple and rapid method to isolate human polymorphic markers. *Nucl Acids Res* **20**: 1333–1337.

Cheng C, Motohashi R, Tsuchimoto S, Fukuta Y, Ohtsubo H, Ohtsubo E (2003). Polyphyletic origin of cultivated rice: based on the interspersion pattern of SINEs. *Mol Biol Evol* **20**: 67–75.

Cheng C, Tsuchimoto S, Ohtsubo H, Ohtsubo E (2002). Evolutionary relationships among rice species with AA genome based on SINE insertion analysis. *Genes Genet Syst* **77**: 323–334.

Choo T (1981). Doubled haploids for studying the inheritance of quantitative characters. *Genetics* **99**: 525–540.

Deininger PL, Batzer MA (1999). *Alu* repeats and human disease. *Mol Genet Metab* **67**: 183–193.

Dunn MA, Hughes MA, Zhang L, Pearce RS, Quigley AS, Jack PL (1991). Nucleotide sequence and molecular analysis of the low temperature induced cereal gene, BLT4. *Mol Gen Genet* **229**: 389–394.

Ellis THN, Poyser SJ, Knox MR, Vershinin AV, Ambrose MJ (1998). *Tyl–copia* class retrotransposon insertion site polymorphism for linkage and diversity analysis in pea. *Mol Gen Genet* **260**: 9–19.

Finnegan DJ (1989). Eukaryotic transposable elements and genome evolution. *Trends Genet* **5**: 103–1077.

Flavell AJ, Bolshakov VN, Booth A, Jing R, Russell J, Ellis TH et al. (2003). A microarray-based high throughput molecular marker genotyping method: the tagged microarray marker (TAM) approach. *Nucleic Acids Res* **31**: e115.

Flavell AJ, Dunbar E, Anderson R, Pearce SR, Hartley R, Kumar A (1992). *Tyl–copia* group retrotransposons are ubiquitous and heterogeneous in higher plants. *Nucl Acids Res* **20**: 3639–3644.

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

Frankel AD, Young JA (1998). HIV-1: fifteen proteins and an RNA. *Annu Rev Biochem* **67**: 1–25.

Grandbastien MA, Lucas H, Morel JB, Mhiri C, Verbhettes S, Casacuberta JM (1997). The expression of the tobacco *Tnt1* retrotransposon is linked to plant defence response. *Genetica* **100**: 241–252.

Gribbon BM, Pearce SR, Kalendar R, Schulman AH, Paulin L, Jack P et al. (1999). Phylogeny and transpositional activity of *Ty1-copia* group retrotransposons in cereal genomes. *Mol Genet Genomics* **261**: 883–891.

Gu YQ, Coleman-Derr D, Kong X, Anderson OD (2004). Rapid genome evolution revealed by comparative sequence analysis of orthologous regions from four Triticeae genomes. *Plant Physiol* **135**: 459–470.

Hill P, Burford D, Martin DMA, Flavell AJ (2005). Retrotransposon populations of *Vicia* species with varying genome size. *Mol Genet Genomics* **273**: 371–381.

Hirochika H, Sugimoto K, Otsuki Y, Kanda M (1996). Retrotransposons of rice involved in mutations induced by tissue culture. *Proc Natl Acad Sci USA* **93**: 7783–7788.

Hubby JL, Lewontin RC (1966). A molecular approach to the study of genic heterozygosity in natural populations. I. The number of alleles at different loci in *Drosophila pseudoobscura*. *Genetics* **54**: 577–594.

Jaccoud D, Peng K, Feinstein D, Kilian A (2001). Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Res* **29**: E25.

Jing R, Bolshakov V, Flavell AJ (2007). The tagged microarray marker (TAM) method for high-throughput detection of single nucleotide and indel polymorphisms. *Nat Protocol* **2**: 168–177.

Jing R, Knox MR, Lee JM, Vershinin AV, Ambrose M, Ellis TH et al. (2005). Insertional polymorphism and antiquity of PDR1 retrotransposon insertions in *Pisum* species. *Genetics* **171**: 741–752.

Jing R, Vershinin A, Grzebyta J, Shaw P, Smýkal P, Marshall D et al. (2010). The genetic diversity and evolution of field pea (*Pisum*) studied by high throughput retrotransposon based insertion polymorphism (RBIP) marker analysis. *BMC Evol Biol* **10**: 44.

Jurka J (2004). Evolutionary impact of human *Alu* repetitive elements. *Curr Opin Genet Dev* **14**: 603–608.

Kaeppler SM, Kaeppler HF, Rhee Y (2000). Epigenetic aspects of somaclonal variation. *Plant Mol Biol* **43**: 179–188.

Kalendar R, Antonius K, Smykal P, Schulman AH (2010). iPBS: A universal method for DNA fingerprinting and retrotransposon isolation. *Theor Appl Genet* (doi:10.1007/s00122-010-1398-2).

Kalendar R, Grob T, Regina M, Suoniemi A, Schulman AH (1999). IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. *Theor Appl Genet* **98**: 704–711.

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

Kalendar R, Tanskanen J, Immonen S, Nevo E, Schulman A (2000). 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.

Kalendar R, Vicient CM, Peleg O, Anamthawat-Jónsson K, Bolshoy A, Schulman AH (2004). Large retrotransposon derivatives: abundance, conserved and non-autonomous retroelements of barley and related genomes. *Genetics* **166**: 1437–1450.

Kan YW, Dozy AM (1978). Polymorphism of DNA sequence adjacent to the human 3-globin structural gene: relationship to sickle mutation. *Proc Natl Acad Sci USA* **75**: 5631–5635.

Kim FJ, Battini JL, Manel N, Sitbon M (2004). Emergence of vertebrate retroviruses and envelope capture. *Virology* **318**: 183–191.

Knox M, Moreau C, Lipscombe J, Baker D, Ellis N (2009). High-throughput retrotransposon-based fluorescent markers: improved information content and allele discrimination. *Plant Methods* **5**: 10.

Kong XY, Gu YQ, You FM, Dubcovsky J, Anderson OD (2004). Dynamics of the evolution of orthologous and paralogous portions of a complex locus region in two genomes of allopolyploid wheat. *Plant Mol Biol* **54**: 55–69.

Kubis SE, Castilho MMF, Vershinin AV, Heslop-Harrison JS (2003). Retroelements, retrotransposons and methylation status in the genome of oil palm (*Elaeis guineensis*) and the relationship to somaclonal variation. *Plant Mol Biol* **52**: 69–79.

Kwon SJ, Park KC, Kim JH, Lee JK, Kim NS (2005). Rim 2/Hipa CACTA transposon display: a new genetic marker technique in *Oryza* species. *BMC Genet* **6**: 15.

Lanteri S, Acquadro A, Comino C, Mauro R, Mauromicale G, Portis E (2006). A first linkage map of globe artichoke (*Cynara cardunculus* varscolymus L) based on AFLP, S-SAP, M-AFLP and microsatellite markers. *Theor Appl Genet* **112**: 1532–1542.

Lee DL, Ellis TH, Turner L, Hellens RP, Cleary WG (1990). A *copia*-like element in *Pisum* demonstrates the uses of dispersed repeated sequences in genetic analysis. *Plant Mol Biol* **15**: 707–722.

Leigh F, Kalendar R, Lea V, Lee D, Donini P, Schulman AH (2003). Comparison of the utility of barley retrotransposon families for genetic analysis by molecular marker techniques. *Mol Genet Genomics* **269**: 464–474.

Liu B, Wendel JF (2000). Retrotransposon activation followed by rapid repression in introgressed rice plants. *Genome Res* **14**: 860–869.

Lou Q, Chen J (2007). *Ty1-copia* retrotransposon-based SSAP marker development and its potential in the genetic study of cucurbits. *Genome* **50**: 802–810.

Macas J, Neumann P, Navrátilová A (2007). Repetitive DNA in the pea (*Pisum sativum* L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and Medicago truncatula. *BMC Genomics* **8**: 427.

Manninen O, Kalendar R, Robinson J, Schulman AH (2000). Application of BARE-1 retrotransposon markers to map a major resistance gene for net blotch in barley. *Mol Gen Genet* **264**: 325–334.

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

Meyer W, Mitchell TG, Freedman EZ, Vilgalys R 1993. Hybridization probes for conventional DNA ngerprinting used as single primers in the polymerase chain reaction to distinguish strains of *Cryptococcus neoformans*. *J Clin Microbiol* **31**: 2274–2280.

Messmer MM, Keller M, Zanetti S, Keller B (1999). Genetic linkage map of a wheat X spelt cross. *Theor Appl Genet* **98**: 1163–1170.

Moisy C, Garrison KE, Meredith CP, Pelsy F (2008). Characterization of ten novel Ty1/copia-like retrotransposon families of the grapevine genome. *BMC Genomics* **9**: 469.

Nagy ED, Molnár I, Schneider A, Kovács G, Molnár-Láng M (2006). Characterization of chromosome-specific S-SAP markers and their use in studying genetic diversity in *Aegilops* species. *Genome* **49**: 289–296.

Pagnotta MA, Mondini L, Porceddu E (2009). Quantification and organization of WIS2-1A and BARE1 retrotransposons in different genomes of *Triticum* and *Aegilops* species. *Mol Genet Genomics* **282**: 245–255.

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H et al. (2009). The *Sorghum bicolor* genome and the diversification of grasses. *Nature* **457**: 551–556.

Paux E, Faure S, Choulet F, Roger D, Gauthier V, Martinant JP et al. (2010). Insertion site-based polymorphism markers open new perspectives for genome saturation and marker-assisted selection in wheat. *Plant Biotechnol J* **8**: 196–210.

Pearce SR, Harrison G, Li D, Heslop-Harrison JS, Kumar A, Flavell AJ (1996). The *Tyl-copia* group of retrotransposons in *Vicia* species: copy number, sequence heterogeneity and chromosomal localisation. *Mol Gen Genet* **205**: 305–315.

Pearce SR, Knox M, Ellis TH, Flavell AJ, Kumar A (2000). Pea Ty1-copia group retrotransposons: transpositional activity and use as markers to study genetic diversity in *Pisum*. *Mol Gen Genet* **263**: 898–907.

Pearce SR, Stuart-Rogers C, Knox MR, Kumar A, Ellis THN, Flavell AJ (1999). Rapid isolation of plant Ty1-copia group retrotransposon LTR sequences for molecular marker studies. *Plant J* **19**: 711–717.

Pelsy F (2007). Untranslated leader region polymorphism of Tvv1, a retrotransposon family, is a novel marker useful for analyzing genetic diversity and relatedness in the genus *Vitis*. *Theor Appl Genet* **116**: 15–27.

Porceddu A, Albertini E, Baracaccia G, Marconi G, Bertoli FB, Veronesi F (2002). Development of S-SAP markers based on an LTR-like sequence from *Medicago sativa* L. *Mol Genet Genomics* **267**: 107–114.

Prieto JL, Pouilly N, Jenczewski E, Deragon JM, Chevre AM (2005). Development of crop-specific transposable element (SINE) markers for studying gene flow from oilseed rape to wild radish. *Theor Appl Genet* **111**: 446–455.

Queen RA, Gribbon BM, James C, Jack P, Flavell AJ (2004). Retrotransposon-based molecular markers for linkage and genetic diversity analysis in wheat. *Mol Genet Genomics* **271**: 91–97.

Ramakrishna W, Dubcovsky J, Park YJ, Busso C, Emberton J, SanMiguel P et al. (2002). Different types and rates of genome evolution detected by comparative sequence analysis of

530

orthologous segments from four cereal genomes. *Genetics* **162**: 1389–1400.

Rice Chromosome 10 Sequencing Consortium (2003). In depth view of structure, activity, and evolution of rice chromosome 10. *Science* **300**: 1566–1569.

Sabot F, Sourdille P, Bernard M (2005). Advent of a new retrotransposon structure: the long form of the Veju elements. *Genetica* **125**: 325–332.

Saeidi H, Rahiminejad MR, Heslop-Harrison JS (2008). Retro-element insertional polymorphisms, diversity and phylogeography within diploid, D-genome *Aegilops tauschii* (Triticeae, Poaceae) sub-taxa in Iran. *Ann Bot* **101**: 855–861.

Salimath SS, de Oliveira AC, Godwin ID, Bennetzen JL (1995). Assessment of genome origins and genetic diversity in the genus *Eleusine* with DNA markers. *Genome* **38**: 757–763.

SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL (1998). The paleontology of intergene retrotransposons of maize. *Nat Genet* **20**: 43–45.

SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berhan A et al. (1996). Nested retrotransposons in the intergenic regions of the maize genome. *Science* **274**: 765–768.

Sanz AM, Gonzalez SG, Syed NH, Suso MJ, Saldaña CC, Flavell AJ (2007). Genetic diversity analysis in *Vicia* species using retrotransposon-based SSAP markers. *Mol Genet Genomics* **278**: 433–441.

Schulman AH, Flavell AJ, Ellis THN (2004). The application of LTR retrotransposons as molecular markers in plants. *Methods Mol Biol* **260**: 145–173.

Schulman AH, Kalendar R (2005). A movable feast: diverse retrotransposons and their contribution to barley genome dynamics. *Cytogenet Genome Res* **110**: 598–605.

Shimamura M, Yasue H, Ohshima K, Abe H, Kato H, Kishiro T et al. (1997). Molecular evidence from retroposons that whales form a clade within even-toed ungulates. *Nature* **388**: 666–670.

Shirasu K, Schulman AH, Lahaye T, Schulze-Lefert P (2000). A contiguous 66 kb barley DNA sequence provides evidence for reversible genome expansion. *Genome Res* **10**: 908–915.

Smýkal P (2006). Development of an efficient retrotransposon-based fingerprinting method for rapid pea variety identification. *J Appl Genet* **47**: 221–230.

Soleimani VD, Baum BR, Johnson DA (2006). Quantification of the retrotransposon BARE1 reveals the dynamic nature of the barley genome. *Genome* **49**: 389–396.

Suoniemi A, Tanskanen J, Schulman AH (1998). Gypsy-like retrotransposons are widespread in the plant kingdom. *Plant J* **13**: 699–705.

Syed NH, Flavell AJ (2007). Sequence specific amplification polymorphisms (SSAP): a multi-locus approach for analysing transposon insertions. *Nat Protoc* **1**: 2746–2752.

Syed NH, Sorensen AP, Antonise R, Wiel C, Linden CG, van't Westende W et al. (2006). A detailed linkage map of lettuce based on SSAP, AFLP and NBS markers. *Theor Appl Genet* **112**: 517–527.

Tam SM, Mhiri C, Vogelaar A, Kerkveld M, Pearce SR, Grandbastien MA (2005). Comparative analyses of genetic diversities within tomato and pepper collections detected by retrotransposon-based SSAP, AFLP and SSR. *Theor Appl Genet* **110**: 819–831.

Tanhuanpää P, Kalendar R, Laurila J, Schulman AH, Manninen O, Kiviharju E (2006). Generation of SNP markers for short straw in oat (*Avena sativa* L). *Genome* **49**: 282–287.

Tanhuanpää P, Kalendar R, Schulman AH, Kiviharju E (2007). A major gene for grain cadmium accumulation in oat (*Avena sativa* L). *Genome* **50**: 588–594.

Teo CH, Tan SH, Ho CL, Faridah QZ, Othman YR, Heslop-Harrison JS et al. (2005). Genome constitution and classification using retrotransposon-based markers in the orphan crop banana. *J Plant Biol* **48**: 96–105.

The International Brachypodium Initiative (2010). Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* **463**: 763–768.

This P, Jung A, Boccacci P, Borrego J, Botta R, Costantini L et al. (2004). Development of a standard set of microsatellite reference alleles for identification of grape cultivars. *Theor Appl Genet* **109**: 1448–1458.

Venturi S, Dondini L, Donini P, Sansavini S (2006). Retrotransposon characterisation and fingerprinting of apple clones by S-SAP markers. *Theor Appl Genet* **112**: 440–444.

Vershinin AV, Allnutt TR, Knox MR, Ambrose MJ, Ellis THN (2003). Transposable elements reveal the impact of introgression, rather than transposition, in *Pisum* diversity, evolution and domestication. *Mol Biol Evol* **20**: 2067–2075.

Vershinin AV, Ellis THN (1999). Heterogeneity of the internal structure of PDR1, a family of *Ty1/copia*-like retrotransposons in pea. *Mol Gen Genet* **262**: 703–713.

Vershinin AV, Salina EA, Solovyov VV, Timofeyeva LL (1990). Genomic organization, evolution and structural peculiarities of high repetitive DNA of *Hordeum vulgare*. *Genome* **33**: 441–449.

Vicient CM, Kalendar R, Schulman AH (2001). Envelope-class retrovirus-like elements are widespread, transcribed and spliced, and insertionally polymorphic in plants. *Genome Res* **11**: 2041–2049.

Vicient CM, Suoniemi A, Anamthawat-Jónsson K, Tanskanen J, Beharav A, Nevo E et al. (1999). Retrotransposon *BARE-1* and its role in genome evolution in the genus *Hordeum*. *Plant Cell* **11**: 1769–1784.

Vitte C, Ishii T, Lamy F, Brar D, Panaud O (2004). Genomic paleontology provides evidence for two distinct origins of Asian rice (*Oryza sativa* L). *Mol Genet Genomics* **272**: 504–511.

Vos P, Hogers R, Bleeker M, Reijans M, van der Lee T, Hornes M et al. (1995). AFLP: a new technique for DNA fingerprinting. *Nucl Acids Res* **21**: 4407–4414.

Voytas DF, Cummings MP, Konieczny AK, Ausubel FM, Rodermel SR (1992). *Copia*-like retrotransposons are ubiquitous among plants. *Proc Natl Acad Sci USA* **89**: 7124–7128.

Vukich M, Schulman AH, Giordani T, Natali L, Kalendar R, Cavallini A (2009). Genetic variability in sunflower and in the *Helianthus* genus as assessed by retrotransposon-based molecular markers. *Theor Appl Genet* **19**: 1027–1038.

Watkins WS, Rogers AR, Ostler CT, Wooding S, Bamshad MJ, Brassington AM et al. (2003). Genetic variation among world populations: inferences from 100 *Alu* insertion polymorphisms. *Genome Res* **13**: 1607–1618.

Waugh R, McLean K, Flavell AJ, Pearce SR, Kumar A, Thomas BBT 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.

Welsh J, McClelland M (1990). Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res* **18**: 7213–7218.

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B et al. (2007). A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* **8**: 973–982.

Williams JG, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucl Acids Res* **18**: 6531–6535.

Yu G-X, Wise RP (2000). An anchored AFLP- and retrotransposon-based map of diploid *Avena*. *Genome* **43**: 736–749.

Supplementary Information accompanies the paper on Heredity website (http://www.nature.com/hdy)