Theor Appl Genet (2010) 121:1419–1430
DOI 10.1007/s00122-010-1398-2

ORIGINAL PAPER

iPBS: a universal method for DNA fingerprinting and retrotransposon isolation

Ruslan Kalendar · Kristiina Antonius · Petr Smýkal · Alan H. Schulman

Received: 21 February 2010 / Accepted: 19 June 2010 / Published online: 10 July 2010

Abstract Molecular markers are essential in plant and animal breeding and biodiversity applications, in human forensics, and for map-based cloning of genes. The long terminal repeat (LTR) retrotransposons are well suited as molecular markers. As dispersed and ubiquitous transposable elements, their "copy and paste" life cycle of replicative transposition leads to new genome insertions without excision of the original element. Both the overall structure of retrotransposons and the domains responsible for the various phases of their replication are highly conserved in all eukaryotes. Nevertheless, up to a year has been required to develop a retrotransposon marker system in a new species, involving cloning and sequencing steps as well as the development of custom primers. Here, we describe a novel PCR-based method useful both as a marker system in its own right and for the rapid isolation of retrotransposon termini and full-length elements, making it ideal for "orphan crops" and other species with underdeveloped marker systems. The method, iPBS amplification, is based on the virtually universal presence of a tRNA complement as a reverse transcriptase primer binding site (PBS) in LTR retrotransposons. The method differs from earlier retrotransposon isolation methods because it is applicable not only to endogenous retroviruses and retroviruses, but also to both *Gypsy* and *Copia* LTR retrotransposons, as well as to non-autonomous LARD and TRIM elements, throughout the plant kingdom and to animals. Furthermore, the inter-PBS amplification technique as such has proved to be a powerful DNA fingerprinting technology without the need for prior sequence knowledge.

Communicated by T. Komatsuda.

Accession numbers for the sequences resulting from this study: AF538603, AF538607–AF538610, AF538612–AF538617, AJ489246, AY078073–AY078075AY271961, AY643843, AY860307, AY860308, DQ094839–DQ094843, DQ663704–DQ663717, EF19100, EF191001, EF191002, EF191007, EF191008, EF191009, EF191011, EF191012, EF439837, EF555576–EF555582, EU009616–EU009622, EU009624, EU090224, EU105454–EU105459, EU177767–EU177770, EU180578, EU180581–EU180586.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-010-1398-2) contains supplementary material, which is available to authorized users.

R. Kalendar · A. H. Schulman (✉)
MTT/BI Plant Genomics Laboratory, Institute of Biotechnology, Viikki Biocenter, University of Helsinki, P.O. Box 65, Helsinki, Finland
e-mail: alan.schulman@helsinki.fi

K. Antonius · A. H. Schulman
Biotechnology and Food Research, MTT Agrifood Research Finland, Jokioinen, Finland

P. Smýkal
Department of Plant Biotechnology, Agritec Plant Research Ltd., Zemědělská 2520/16, 787 01 Šumperk, Czech Republic

Abbreviations

LTR	Long terminal repeat
PBS	Primer binding site

Introduction

Retrotransposons as a class of repetitive and mobile sequences are ubiquitous and abundant components of virtually all known eukaryotic genomes (Flavell et al. 1992; Voytas et al. 1992; Wicker et al. 2007). In higher plants, they can constitute more than half of the repetitive DNA (Schnable et al. 2009), and are dynamic genome components with the ability to integrate new copies and facilitate

intra-chromosomal recombination (Belyayev et al. 2010; Kalendar et al. 2000). They show widespread chromosomal dispersion, and variable but generally high copy number (Alix and Heslop-Harrison 2004; Boyko et al. 2002; Ellis et al. 1998; Hedges and Batzer 2005; Schnable et al. 2009). Retrotransposons move to new chromosomal locations via an RNA intermediate, and insert new cDNA copies back into the genome. This mode of replication increases genome size, and contributes significantly to the total DNA of higher plants. Different retrotransposon families, each with its own lineage and structure, have the potential to be active in different tissues and evolutionary epochs (Jing et al. 2005). In plant genomes, the Class I retrotransposons containing long terminal repeats (LTRs) predominate (e.g., International Rice Genome Sequencing Project 2005; Schnable et al. 2009; The International Brachypodium Initiative 2010). The ubiquitous distribution, high copy number, and widespread chromosomal dispersion of retrotransposons in plants have provided excellent potential for the development of multiplex DNA-based marker systems (Kalendar and Schulman 2006; Schulman et al. 2004; Shedlock and Okada 2000).

Retrotransposon-based fingerprinting systems detect the insertion of elements hundreds to thousands of nucleotides long. Retrotransposon integration sites represent joints between the conserved LTR ends and flanking, essentially random, genomic DNA. Most retrotransposon-based marker systems use PCR to amplify a segment of genomic DNA surrounding this joint. One primer is designed to match a segment of the LTR conserved within a given family of elements, oriented outwards and away from the LTR. The second primer is designed to a general feature of the genome. The second primer matches a restriction site adapter in Sequence-Specific Amplified Polymorphism (SSAP; Schulman et al. 2004; Waugh et al. 1997), another retrotransposon in Inter-Retrotransposon Amplification Polymorphism (IRAP; Kalendar et al. 1999; Kalendar and Schulman 2006), and a microsatellite in Retrotransposon-Microsatellite Amplification Polymorphism (REMAP; Kalendar et al. 1999; Kalendar and Schulman 2006). The RBIP and TAM systems rely on either product size variations or differential amplification success across full compared with empty insertion sites, and require sequence information for design of primers matching the flanking genomic DNA at each specific site (Flavell et al. 1998, 2003).

The LTR retrotransposon marker systems can be generalized to other transposable element types, such as MITEs and SINEs, and to other organisms. For example, the SINE element *Alu* of humans has been used in a method called Alu-PCR, conceptually similar to IRAP, for the study of animal genomes (Chariieu et al. 1992; Jurka 2004; Tang et al. 1995).

As a result of their general applicability, simplicity of implementation, and genotype resolution, retrotransposon marker systems have been widely applied in evolutionary and genetic diversity studies as well as for marker-assisted selection (Feschotte et al. 2002; Kalendar and Schulman 2006; Schulman et al. 2004). The limiting factor, however, in the development of molecular marker systems based on LTR retrotransposons for new plant species is availability of retrotransposon sequences. If extensive genome sequences are not available, LTR ends must be cloned and sequenced, then trialled for their usefulness as markers. Previous methods for doing this have relied on amplification with degenerate primers matching conserved domains in retrotransposon polyproteins, particularly integrase or reverse transcriptase (Pearce et al. 1999), followed by walking to the LTR ends. The process can be tedious and time-consuming due to the distance between the conserved domains and the LTRs, the difficulty of defining the 3′ LTR, closer to the conserved sequences, in the absence of a matching 5′ LTR, and the repetitiveness of both LTR retrotransposons and frequently their integration sites, which interferes with primer walking.

Here, we describe a method that overcomes these difficulties and can both isolate LTR retrotransposons in virtually any organism as well as serve as a general marker system in its own right. It is based on the nearly universal use by both retroviruses and LTR retrotransposons of cellular tRNAs as primers for reverse transcription during their replication cycles. The tRNA binds to the primer binding site (PBS) adjacent to the 5′ LTR and primes synthesis of minus-strand cDNA by reverse transcriptase (Kelly et al. 2003; LeGrice 2003; Mak and Kleiman 1997; Marquet et al. 1995). A handful of exceptional LTR retrotransposons, including the *Tf1/sushi* group of fungi and vertebrates and *Fourf* in maize, is able to self-prime cDNA synthesis rather than use a tRNA primer (Hizi 2008; LeGrice 2003). Using a single PBS primer, we cloned (Kalendar et al. 2008) *Cassandra* TRIM retrotransposons for 50 seed plant species belonging to *Magnoliophyta* (flowering plants) and *Pteridophyta* (ferns). Here, we show that using a full set of the conserved parts of PBS sequences, we are able both to directly visualize polymorphism between individuals, rapidly clone LTR segments from genomic DNA, and carry out in silico database searches. The method is applicable to any organism with retrotransposons containing PBS sites complimentary to tRNA.

Materials and methods

Plant materials and DNA isolation

The plant and animal materials used as DNA sources are described in detail in the supplemental materials (Online

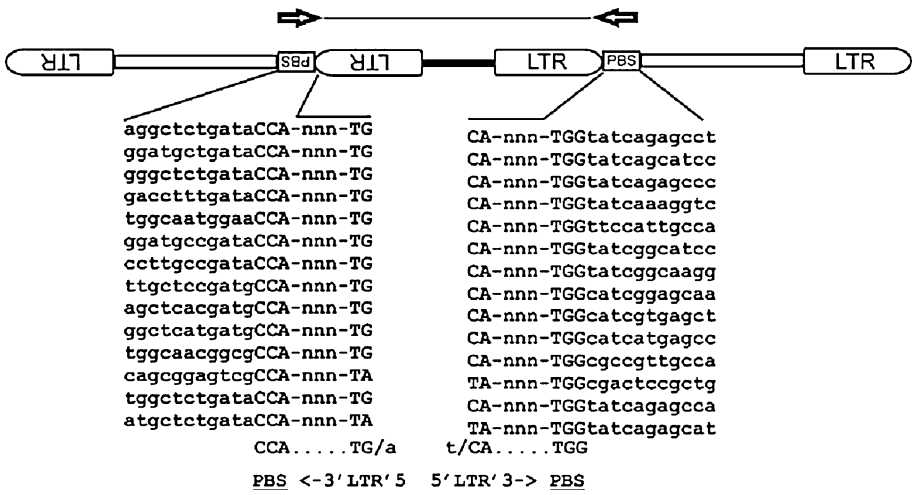

Fig. 1 The iPBS scheme. For iPBS, two retrotransposons must be in opposite orientation and either near enough for efficient amplification, as shown, or nested. The diagram depicts two key structural features of retrotransposons, the LTR (long terminal repeat) and PBS (primer binding site). The internal domain is shown as a *thick bar*, the intervening genomic DNA as *thick line*. The predicted product is show above, together with the orientation of the PBS amplification primers. The PCR product contains both LTRs and PBS sequences together with the genomic sequence between the LTRs. *Below the diagram* The sequence of a set of PBS domains, the 0–5 base spacer and the universal 5′ TG of LTRs is shown

Resource 1). Plant DNA was prepared as reported previously (Vicient et al. 1999). DNA from *Bos bovis* (cow), *B. grunniens* (yak), *Gallus gallus* (chicken) and *Ovis aries* (sheep) was obtained from MTT Agrifood Research Finland (Online Resource 1).

PBS primer design

Annotated sequences of LTR retrotransposons were collected from public databases (NCBI, RetrOryza, and TREP). Additional matching sequences were found by BLASTn searches of the non-redundant, non-human, non-mouse NCBI database. All sequences were then clustered using FastPCR software (http://primerdigital.com/fastpcr.html), the PBS sequences were extracted, and multiple alignments for each PBS motif were carried out. For each group, we designed matching PCR primers, oriented for amplification toward the LTR and starting with 5′ TGG (Fig. 1; Tables 1, 2). Two sets of primers were prepared, of 12 or 13 nt and of 18 nt. The primer Tm was calculated using FastPCR. Optimal anneal-ing temperature (Ta) was calculated as the Tm of the primer plus 9–10°C. PCR efficiency was verified with a gradient of annealing temperatures (Online Resource 2, Fig. S1) in a MasterCycler Gradient PCR machine (Eppendorf). For primer combinations with very different Tm, the optimal annealing temperature was chosen according to the primer with the lowest Tm. The experimentally determined optimal annealing Tm for 18-mer PBS primers is shown in Table 2. In practice, a Ta between 50 and 60°C suited all primers.

iPBS PCR amplification

The PCR was performed in a 25 µl reaction mixture containing 25–50 ng DNA, 1× DreamTaq PCR buffer, 1 µM of primer for 12–13 nt primers (for primer combinations, 1 µM total concentration) or 0.6 µM for 18 nt primers, 0.2 mM dNTPs, 1 units Taq DNA polymerase (DreamTaq, Fermentas) and 0.04 units Pfu DNA Polymerase (Fermentas). The PCR program consisted of: 1 cycle at 95°C for 3 min; 28–30 cycles of 95°C for 15 s, 50–60°C (see above) for 60 s, and 68°C for 60 s; a final extension step of 72°C for 5 min. Amplification was performed in PTC-100 Programmable Thermal Controller (MJ research Inc., Bio-Rad Laboratories, USA) or a MasterCycler Gradient (Eppendorf AG) in 0.2 ml tubes or 96-well plates. Products were analyzed by gel electrophoresis in 1.7% (w/v) agarose gels (RESolute Wide Range, BIOzym) with 1× STBE electrophoresis buffer (10× STBE: 0.25 M Tris–H_3BO_3, 40 mM $Na_2B_4O_7$, 10 mM EDTA, pH 8.6) at 80 V for 7 h and visualized by staining with ethidium bromide. Gels were scanned on a FLA-5100 imaging system (Fuji Photo Film (Europe) GmbH) scanner at a resolution of either 50 or 100 µm.

LTR cloning, data analysis and design of species-specific primers

To identify LTRs, PCR products from each set of primers were mixed and purified (MinElute PCR Purification Kit, Qiagen). The purified PCR products were ligated into pGEM-T. For ligation, 20 µl of PCR products were mixed

Table 1 12–13-mer PBS primers and their efficacy in single-primer iPBS amplification

	Sequence	Tm (°C)[a]	CG (%)	Optimal annealing, Ta (°C)	PCR efficiency[b]	
					Barley	Cow
2074	GCTCTGATACCA	40.5	50.0	49.6	5	5
2075	CTCATGATGCCA	42.1	50.0	51.2	5	4
2076	GCTCCGATGCCA	50.4	66.7	59.2	5	5
2077	CTCACGATGCCA	46.1	58.3	55.1	5	5
2078	GCGGAGTCGCCA	54.2	75.0	62.8	5	5
2079	AGGTGGGCGCCA	56.6	75.0	65.2	5	5
2080	CAGACGGCGCCA	54.6	75.0	63.3	3	5
2081	GCAACGGCGCCA	56.5	75.0	65.0	3	5
2083	CTTCTAGCGCCA	45.7	58.3	54.6	4	5
2085	ATGCCGATACCA	43.8	50.0	52.8	4	4
2087	GCAATGGAACCA	43.5	50.0	52.5	4	4
2095	GCTCGGATACCA	44.8	58.3	53.7	5	0
2374	CCCAGCAAACCA	47.1	58.3	53.5	5	5
2375	TCGCATCAACCA	45.1	50.0	52.5	5	5
2376	TAGATGGCACCA	43.1	50.0	52.0	5	5
2377	ACGAAGGGACCA	47.2	58.3	53.0	5	5
2378	GGTCCTCATCCA	44.2	58.3	53.0	5	5
2379	TCCAGAGATCCA	41.5	50.0	49.2	3	4
2380	CAACCTGATCCA	41.4	50.0	50.5	3	5
2381	GTCCATCTTCCA	40.9	50.0	50.0	4	5
2382	TGTTGGCTTCCA	44.9	50.0	50.5	4	5
2383	GCATGGCCTCCA	50.5	66.7	53.0	5	5
2384	GTAATGGGTCCA	40.9	50.0	50.0	4	5
2385	CCATTGGGTCCA	45.7	58.3	51.2	5	5
2386	CTGATCAACCCA	41.4	50.0	50.1	5	5
2387	GCGCAATACCCA	47.3	58.3	51.5	4	5
2388	TTGGAAGACCCA	43.4	50.0	51.0	4	5
2389	ACATCCTTCCCA	43.0	50.0	50.0	5	5
2390	GCAACAACCCCA	47.6	58.3	56.4	5	5
2391	ATCTGTCAGCCA	43.6	50.0	52.6	4	5
2392	TAGATGGTGCCA	43.1	50.0	52.2	4	5
2393	TACGGTACGCCA	47.1	58.3	51.0	5	4
2394	GAGCCTAGGCCA	48.5	66.7	56.5	5	5
2270	ACCTGGCGTGCCA	56.9	69.2	65.0	5	5
2271	GGCTCGGATGCCA	54.3	69.2	60.0	5	5
2272	GGCTCAGATGCCA	50.5	61.5	55.0	5	5
2273	GCTCATCATGCCA	47.6	53.8	56.5	5	5
2274	ATGGTGGGCGCCA	57.1	69.2	65.8	4	5
2276	ACCTCTGATACCA	42.7	46.2	51.7	4	1
2277	GGCGATGATACCA	46.2	53.8	52.0	5	4
2278	GCTCATGATACCA	42.3	46.2	51.0	4	2
2279	AATGAAAGCACCA	43.0	38.5	52.0	4	5

[a] Oligonucleotide concentration of 1 μM

[b] PCR efficiency rating scale (0, useless; 5, excellent): 0, no bands; 1, few and weak bands; 2, a few strong bands; 3, ≈10 strong bands; 4, many bands (good primer); 5, many strong and equally amplifying bands

with 5 μl of 10× ligation buffer (Fermentas), 5 μl 50% (w/v) PEG 4000, 250 ng pGEM-T, and 15 units T4 DNA ligase (Fermentas). The solution was incubated at 16°C overnight (12 h). 2 μl of a ligation reaction was transformed into *Escherichia coli* JM109 cells (40 μl). A total of 96 random colonies were chosen and sequenced using a standard M13 primer. Sequences were clustered and common regions were identified with the FastPCR LTR clustering tool

Table 2 18-mer PBS primers and their efficacy in single-primer iPBS amplification

	Sequence	Tm (°C)[a]	CG (%)	Optimal annealing, Ta (°C)	Average PCR efficiency[b]	
					Barley	Cow
2217	ACTTGGATGTCGATACCA	52.5	44.4	51.4	3	1
2218	CTCCAGCTCCGATTACCA	56.1	55.6	51.0	3	1
2219	GAACTTATGCCGATACCA	51.5	44.4	53.0	3	1
2220	ACCTGGCTCATGATGCCA	59.0	55.6	57.0	4	3
2221	ACCTAGCTCACGATGCCA	58.0	55.6	56.9	4	5
2222	ACTTGGATGCCGATACCA	55.7	50.0	53.0	5	2
2224	ATCCTGGCAATGGAACCA	56.6	50.0	55.4	5	5
2225	AGCATAGCTTTGATACCA	50.5	38.9	55.0	4	1
2226	CGGTGACCTTTGATACCA	54.2	50.0	53.1	3	1
2228	CATTGGCTCTTGATACCA	51.9	44.4	54.0	5	4
2229	CGACCTGTTCTGATACCA	53.5	50.0	52.5	5	3
2230	TCTAGGCGTCTGATACCA	54.0	50.0	52.9	5	3
2231	ACTTGGATGCTGATACCA	52.9	44.4	52.0	4	4
2232	AGAGAGGCTCGGATACCA	56.6	55.6	55.4	5	5
2237	CCCCTACCTGGCGTGCCA	65.0	72.2	55.0	5	5
2238	ACCTAGCTCATGATGCCA	55.5	50.0	56.0	5	5
2239	ACCTAGGCTCGGATGCCA	60.4	61.1	55.0	5	5
2240	AACCTGGCTCAGATGCCA	58.9	55.6	55.0	4	4
2241	ACCTAGCTCATCATGCCA	55.5	50.0	55.0	4	5
2242	GCCCCATGGTGGGCGCCA	69.2	77.8	57.0	5	5
2243	AGTCAGGCTCTGTTACCA	54.9	50.0	53.8	4	4
2244	GGAAGGCTCTGATTACCA	53.7	50.0	49.0	4	2
2245	GAGGTGGCTCTTATACCA	53.1	50.0	50.0	4	3
2246	ACTAGGCTCTGTATACCA	50.9	44.4	49.0	3	2
2249	AACCGACCTCTGATACCA	54.7	50.0	51.0	5	3
2251	GAACAGGCGATGATACCA	54.3	50.0	53.2	5	4
2252	TCATGGCTCATGATACCA	52.7	44.4	51.6	5	3
2253	TCGAGGCTCTAGATACCA	53.4	50.0	51.0	5	3
2255	GCGTGTGCTCTCATACCA	57.1	55.6	50.0	4	1
2256	GACCTAGCTCTAATACCA	49.6	44.4	51.0	5	2
2257	CTCTCAATGAAAGCACCA	52.4	44.4	50.0	5	3
2295	AGAACGGCTCTGATACCA	55.0	50.0	60.0	5	4
2298	AGAAGAGCTCTGATACCA	51.6	44.4	60.0	5	4
2373	GAACTTGCTCCGATGCCA	57.9	55.6	51.0	5	5
2395	TCCCCAGCGGAGTCGCCA	66.0	72.2	52.8	5	4
2398	GAACCCTTGCCGATACCA	57.1	55.6	51.0	5	4
2399	AAACTGGCAACGGCGCCA	63.4	61.1	52.0	4	4
2400	CCCCTCCTTCTAGCGCCA	61.6	66.7	51.0	5	2
2401	AGTTAAGCTTTGATACCA	47.8	33.3	53.0	4	2
2402	TCTAAGCTCTTGATACCA	49.0	38.9	50.0	3	2
2415	CATCGTAGGTGGGCGCCA	62.5	66.7	61.0	5	5

[a] Oligonucleotide concentration of 1 μM

[b] PCR efficiency as per Table 1

(http://primerdigital.com/fastpcr.html). The universal 5′… CA 3′ terminus of retrotransposon LTRs was searched for 0–4 nt from the 5′ end of the PBS seg-ment. Clustering and alignment of the other PCR products served to identify repetitive putative LTRs. Using this as a guide, conserved segments of the identified LTRs were used to design species- and retrotransposon-family-specific PCR primers using the FastPCR program.

In addition, the defined clusters' consensus sequences were screened against the non-redundant DNA database at

NCBI (using BLASTn or tBLASTx) and then clustered and aligned with the multiple alignment procedure from MULTALIN (Corpet 1988). Outward facing primers were designed to match the putative LTRs and used in single-primer IRAP. Primers designed to true LTRs of moderate or high copy number produced IRAP bands; those designed to low-copy LTRs or to other sequences did not. Several primer pairs, oriented away from each other as for inverse PCR, were designed for each identified element. Complete LTR retrotransposons were identified and extracted with long distance PCR with these primers, running low numbers of cycles (10–15) to select for abundant elements.

Long distance PCR to isolate complete LTR retrotransposons

PCR primers of 25–30 nt were designed with high Tm (>60°C) and a Ta of 65–68°C. The 100 µl reaction volume contained: 1× Phusion HF buffer, 100 ng DNA, 300 nM of each primer, 200 µM dNTP, 2 units Phusion DNA Polymerase (Finnzymes). The reaction cycle consisted of: 98°C, 30 s; 15–20 cycles of 10 s at 98°C, 60 s at 65°C, 200 s at 72°C; a final extension of 10 min at 72°C. After PCR, the Phusion DNA Polymerase was inactivated by freezing at −20°C. Non-templated 3′ adenosine was added to the blunt-end PCR fragments by adding Taq polymerase (5 units per 100 µl) and dATP (0.2 µM) and incubating for 20 min at 68°C. Products were purified from 1% agarose gels and identified using SYBR Green I in the loaded DNA samples and visualization under a Dark Reader (Clare Chemical Research). After gel extraction, PCR fragments were cloned into pGEM-T.

In silico LTR searches

We used the repeat searching algorithm of the FastPCR software to search for retrotransposons in plant and animal genomic sequences. The search consisted of the following steps: (1) identification of PBS sequences; (2) identification of LTR segments by the presence within 0–4 nt from the 5′ end of the PBS of the universal 5′…CA 3′ terminus of retrotransposon LTRs; (3) clustering of LTR sequences and the adjacent PBS; (4) identification of complete left LTR sequences by the presence of paired terminal inverted repeats (TIRs) matching the 5′ TG…CA 3′ consensus for retrotransposons, the CA being adjacent to the PBS; (5) identification of a matching right LTR with the same TIRs and assembly of the intervening internal domain; (6) construction of consensus sequences for each pair of LTRs and intervening internal domain. As parameters for clustering, we used a minimum LTR length of 100 bp and 75% similarity. Consensus sequences of complete retrotransposons were submitted to the NCBI database. In addition, genome sequences in databases were analyzed for LTR sequences using the "LTR search" tool of Fast PCR. The complete genomes of *Oryza sativa* (rice; Yu et al. 2005; http://rgp.dna.affrc.go.jp/IRGSP/), *Vitis vinifera* (grape; http://www.ebi.ac.uk/genomes/eukaryota.html; Jaillon et al. 2007; Pereira et al. 2005), *Arabidopsis thaliana* (http://www.arabidopsis.org/), and *Drosophila melanogaster* (fruitfly; http://www.ebi.ac.uk/genomes/eukaryota.html), as well as the shotgun sequence of *Solanum lycopersicum* (tomato; Solanaceae Genome Project; http://www.ncbi.nlm.nih.gov/genomes/leuks.cgi?p3=11:Plants&taxgroup=11:Plants|12%3A) were analyzed.

Results

PBS primers to clone plant LTR retrotransposons

In retroviruses, the PBS is complementary to 18 nt at the 3′ terminus of the primer tRNA; in retrotransposons, the complementary region varies from 8 to 18 nt (Kelly et al. 2003; LeGrice 2003; Mak and Kleiman 1997; Marquet et al. 1995). The PBS domains of LTR retrotransposons and all retroviruses match a limited set of tRNAs: $tRNA^{iMet}$, $tRNA^{Lys}$, $tRNA^{Pro}$, $tRNA^{Trp}$, $tRNA^{Asn}$, $tRNA^{Ser}$, $tRNA^{Arg}$, $tRNA^{Phe}$, $tRNA^{Leu}$, or $tRNA^{Gln}$. Depending on the group, one tRNA may predominate: lentiviruses, including HIV-1, use $tRNA^{Lys}$ (Mak and Kleiman 1997), whereas $tRNA^{iMet}$ is commonly, though not exclusively, used by plant retrotransposons. A set of primers were designed to match PBS sequences found in both general nucleic acid and transposable element databases. These were used in a PCR method that we call iPBS for inter-PBS amplification (Fig. 1), with genomic DNA derived from a wide range of plant and animal species (Online Resource 1). The iPBS method produced on average 15–50 bands from 100 to 5,000 bp in length that could be detected by ethidium bromide staining on agarose gels (Fig. 2), in all investigated species, both plants and animals. Figures 3 and 4 show typical iPBS products for apple and maize. The performance of single primers varied according to species. Some primers are effective in all species investigated (Table 1). In general, primers that show good efficiency in iPBS in one species are also efficient in many others. Primers matching similar PBS sequences (e.g., primers 2080 and 2081, with the same 9 nt at the 3′ ends) produced similar but not identical band patterns.

A set of iPBS reaction products was cloned and sequenced. Between 50 and 70% of the sequences from plant samples contained LTR sequences adjacent to the priming sites, in some cases even complete TRIM retrotransposons. The LTR sequences generally started 0–5 bp after PBS primer sequence, as expected from the structure of known LTR retrotransposons (Fig. 1). After determining

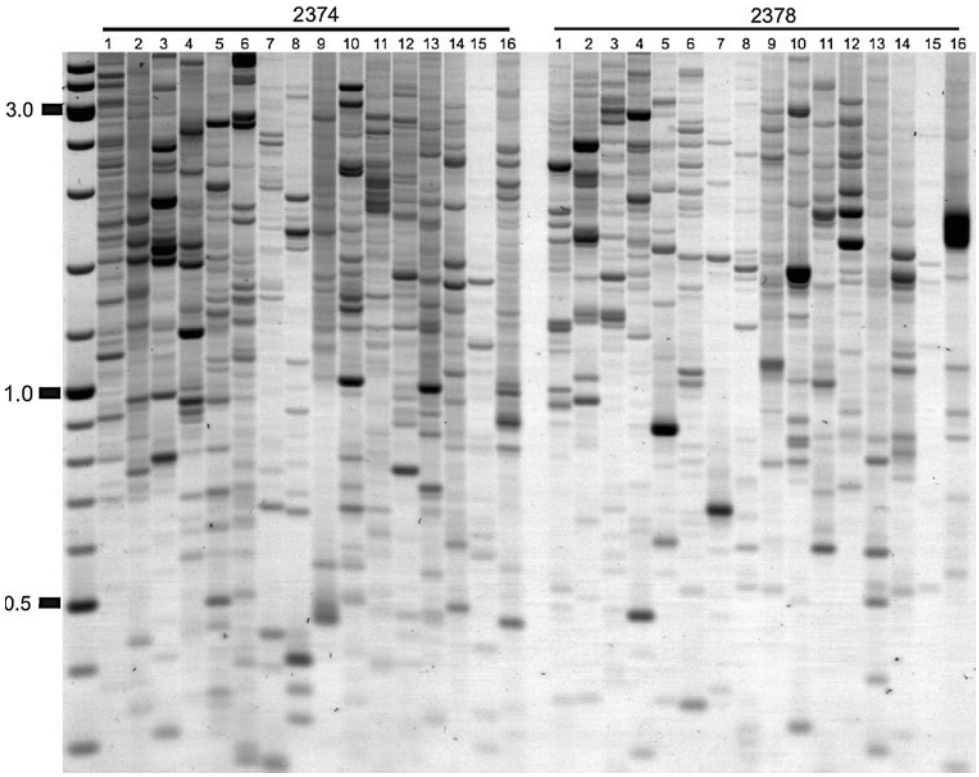

Fig. 2 Utility of iPBS for a diversity of plant species. iPBS fingerprints are shown as negative images of ethidium bromide-stained agarose gels following electrophoresis. Results for two primers 2374 and 2378 (Table 1) are shown. Lanes are: *1*, *H. vulgare* (cv. Rolfi); *2*, *Phleum pratense*; *3*, *Spartina alterniflora*; *4*, *Avena sativa* (cv. Salo); *5*, *Deschampsia antarctica*; *6*, *Brassica napus* (cv. Wildcat); *7*, *Vaccinium* sp.; *8*, *Glycine max*; *9*, *Pinus sylvestris*; *10*, *Gnetum gnemon*; *11*, *Ginkgo biloba*; *12*, *Cycas micholitzii*; *13*, *Equisetum arvense*; *14*, *Nephrolepis exaltata*; *15*, *Sphagnum* sp.; *16*, *Lycopodium* sp. A 100-bp DNA ladder is present on the *left*

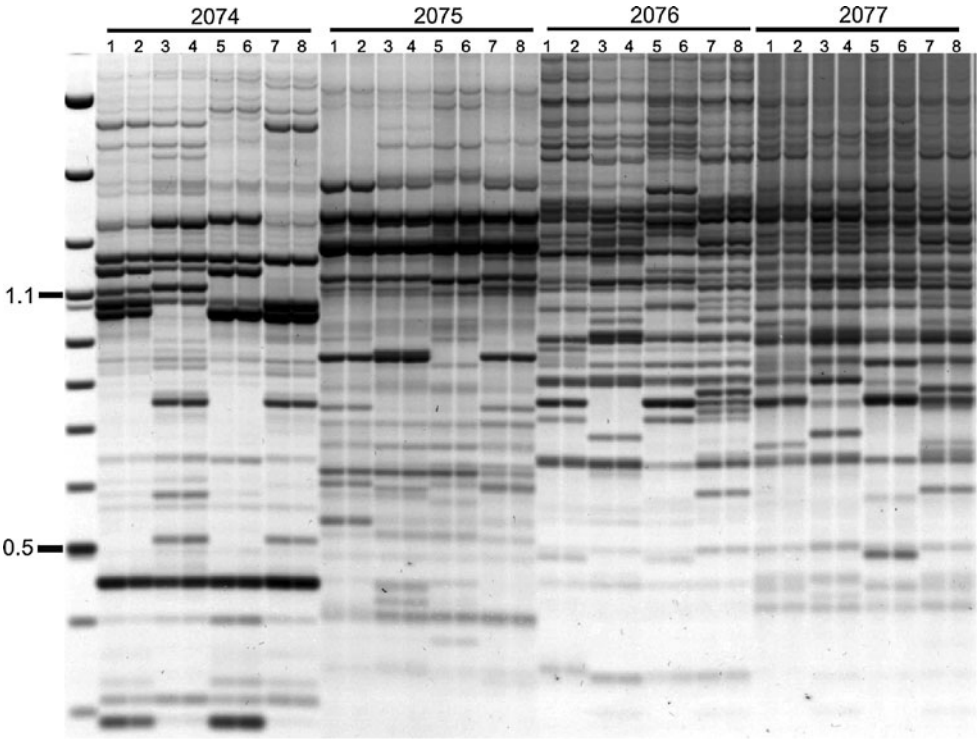

Fig. 3 iPBS fingerprinting of apple cultivars and their sports. Fingerprints are as for Fig. 2 and are grouped by PBS primer (Tables 1, 2). Lanes are: *1*, Atlas; *2*, its sport Red Atlas; *3*, Sävstaholm; *4*, its red sport Bergius; *5*, *Syysjuovikas*; *6*, its sport Luotsi; *7*, *Melba*; *8*, its sport *Melba Red Pate*

the LTR sequences of a selected family of retrotransposons, we aligned them, determined the most conserved regions, and designed new primers either for extended PCR to isolate entire retroelements or for the generation of IRAP, REMAP or SSAP fingerprints. The effectiveness of the method was established in several plant species; examples of full-length LTR retrotransposons isolated by PBS amplification are summarized in Table S1 (Online Resource 2), and examples of families of retrotransposons identified by their LTRs, from both *Gypsy* and *Copia* superfamilies as

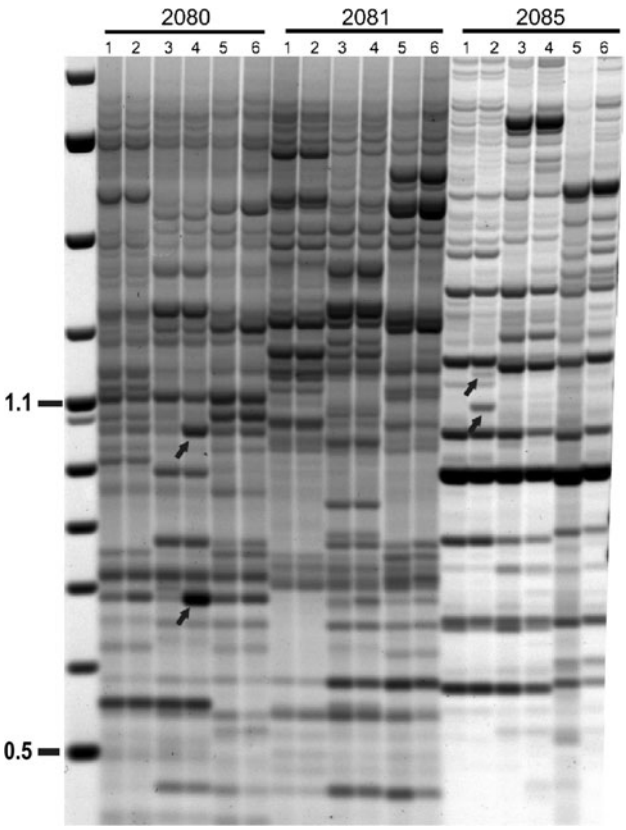

Fig. 4 iPBS fingerprinting of maize lines and regenerants. Fingerprints are as for Fig. 2. Lanes are grouped by PBS primer (Tables 1, 2). Lanes are: *1*, line P346; *2*, its regenerant; *3*, line UKCH6; *4*, its regenerant; *5*, line UKCH9; *6*, its regenerant. *Arrows* indicate new bands appearing in the regenerants

well as TRIMs and LARDs, are listed in Table S2 (Online Resource 2).

PBS primers for polymorphism detection

The iPBS method yielded polymorphic fingerprints for species throughout the plant kingdom, including angiosperms, gymnosperms, and lower plants (Fig. 2). Profiles were produced from the animal samples investigated. Both single PBS primers and two different PBS primers in combination were effective. Due to the high copy number of PBS sequences in the genomes examined, iPBS amplification generated many clear bands, which were easy to score on standard agarose gels stained with ethidium bromide. The background signal detected was low and the distribution of bands through the profile was even, both highly desirable features if primers are to be used for extensive or routine analyses. The products were predominantly 100–5,000-bp long. The iPBS amplification products behaved as dominant markers, as is the case for IRAP, REMAP, SSAP and other anonymous PCR-based molecular marker systems. For all species that we investigated in this study (Fig. 2; Online Resource 1), we were able to find optimal primers or combinations from among the set we present (Tables 1, 2) to produce a large number of clear bands with a high level of polymorphism.

The iPCR method was useful for display of diversity in wild germplasm accessions, even for the small genome of the model plant *Brachypodium distachyon* (Supplemental Fig. S2). Most of the primers tested on *B. distachyon* gave a large number of clear bands, of which a high percentage was polymorphic. The method was tested for its power to detect polymorphisms in closely related or clonally derived material, including apple (*Malus domestica*; Fig. 3) and maize (*Zea mays*; Fig. 4) as well as (data not shown) potato (*Solanum tuberosum*), flax (*Linum usitatissimum*), *Vaccinium* spp., and watermelon (*Citrullus lanatus*). Apple cultivars, but not their somatic mutations (sports) derived by clonal propagation, could be distinguished. For maize, not only could lines be distinguished, but also some new polymorphisms were apparent in lines regenerated after 2 months of tissue culture (Fig. 4). Likewise, the iPBS profiles of *Vaccinium* spp. made with several primers (2076, 2077, 2079, 2080, 2081, 2085) displayed several polymorphisms between somatic lines and their source (data not shown).

For comparison of the various techniques based on retrotransposon polymorphism, we chose three barley (*Hordeum vulgare*) varieties: Wanubet, Ingrid, and Golden Promise, which we had earlier analyzed with IRAP, REMAP, and SSAP (Leigh et al. 2003).When iPBS was carried out for all PBS primers, 12-mer, 13-mer, and 18-mer (Online Resource 2, Table S3), between 5 and 25% of the bands, were polymorphic. In this experiment, each PBS primer in single-primer iPBS visualized 23.3 ± 0.7 (SEM) bands of which $15.1 \pm 0.6\%$ (SEM) were polymorphic; there was no significant difference in either the number of bands or their degree of polymorphism between the long and short primers. The IRAP technique for these three barley varieties demonstrated polymorphism between 10 and 60%, depending on the chosen primer.

In view of the presence of endogenous retroviruses (ERVs) in animal genomes, we tested the PBS primer sets with cow, yak, sheep, and chicken samples (Fig. 5). All the PBS primers tested were effective on each of the species. For the bovine samples, two to ten polymorphic bands were produced from the individuals studied. The fingerprints showed some invariant bands between cattle and yak. Moreover, cattle progeny derived from crosses between different parental lines show that the primers can also differentiate closely related animals, e.g. a parent and its offspring (Fig. 5).

Use of PBS sequences for in silico genome analysis

The rapid growth in genome sequencing is generating an increasing need for rapid and sensitive identification and

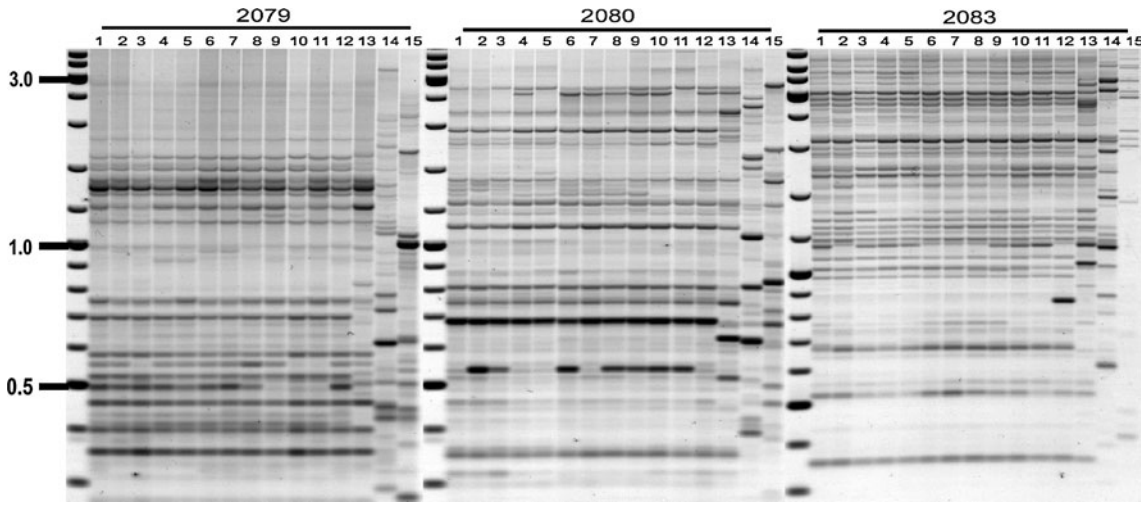

Fig. 5 iPBS fingerprinting of animal samples. Fingerprints are as for Fig. 2. Lanes are grouped by PBS primer (Table 1). Lanes are: *1–12*, cow, *lane 1* displaying individual P14; *lanes 2*, *3*, P70 and P69 (son and mother); *4*, *5*, P34 and P88 (half siblings); *6*, *7*, P167 and P169 (two cows from an inbred herd, exact degree of relatedness not known); *8*, *9*, P212 and P213 (two cows from an inbred herd); *10*, *11*, P134 and P137 (half siblings), *12*, P16; *13*, yak; *14*, sheep; *15*, chicken

annotation of transposable elements (Wicker et al. 2007). The PBS domain can be useful for the identification of retrotransposons, especially for non-autonomous groups such as TRIMs and LARDs that lack protein coding domains, and for finding the LTR immediately adjacent to the PBS. To explore this use, we carried out similarity searches against sequence databases. In the rice genome, we discovered 143,163 sequences adjacent to the PBS which were repeated from 2 to at least 1,000 times, which is 2.5 times greater than the total number of annotated retrotransposons in this genome. These were clustered together with the LTRs of the RetrOryza database (Chaparro et al. 2007). Applying a stringency of 75% identity, we found 14,488 clusters for 55,796 LTR sequences (query strings and complete LTRs), about the same as the total number of LTR retrotransposons (*Copia*, *Gypsy*, and "other") annotated for this genome (53,302), the rest being singletons. Many of the large clusters are not currently annotated in the rice genome and absent in RetrOryza database. Likewise, the searches in the *S. lycopersicum* and *V. vinifera* databases produced evidence for numerous new retrotransposon elements (Online Resource 2, Table S4). Analyses of the *D. melanogaster* and *A. thaliana* genomes revealed both known and novel retrotransposons; these will be presented separately.

Discussion

PBS primers for LTR retrotransposon isolation

Primers designed to match the conserved regions of the PBS domains of LTR retrotransposons and retroviruses proved to be very effective for the amplification of retroelements from both plant and animal genomes. The yield of correct products, 50–70%, was sufficiently high to make the technique useful for the cloning of new retrotransposons from unsequenced genomes. Alternatively making and screening of bulk shotgun reads from high-throughput sequencers such as on the GS FLX Titanium platform of 454 Life Sciences (Macas et al. 2007) is considerably less efficient by comparison, as the following calculation reveals. A single FLX read of a large, highly repetitive plant genome will yield approximately 4×10^8 nt, of which 3.2×10^8 (80%) might be derived from retrotransposons. The PBS and LTR termini (TIR and conserved flanking region) comprise only 40 nt, or 0.4% of each retrotransposon and 0.32% of the total read, which is further partitioned into many thousands of individual families. In order to identify the individual retrotransposon families and their PBS and LTRs, clustering of the reads must be carried out, an approach in silico use of iPBS in any case facilitates. From the shotgun sequences of modest genome portions (10% or less), moderately repeated or rare retrotransposons (500 copies per genome or less) will be hard to recognize reliably (Macas et al. 2007). The PCR products of iPBS, however, can serve as templates for high-throughput sequencing for global retrotransposon discovery.

We defined a set of universal primers for iPBS, of which the 18-mer primers (Table 2) were best suited. Because only PBS primers are used, the method relies on the tendency of retrotransposons to cluster near each other in relatively gene-poor domains of the genome (Liu et al. 2007; Shirasu et al. 2000; Wicker et al. 2005). Furthermore, for iPBS to function, two retroelements must be in inverse

orientation with respect to each other. They may either be separated by intervening genomic sequences or be adjacent to another in nests. Because the PBS is adjacent to the LTR and LTRs that are often less than 2 kb, the retroelements do not need to be adjacent to one other for efficient iPBS amplification. However, the method may be generalized by combining a PBS primer with an adapter primer in a modification of SSAP (Syed and Flavell 2007; Waugh et al. 1997). The iPBS method has been successfully applied for cloning retrotransposons from a large range of plant species and retrotransposon families (Online Resource 2, Tables S1 and S2). Furthermore, the method has been used to design effective IRAP and REMAP primers for a range of species including flax (Smýkal et al. submitted) as well as for watermelon, *Brassica* spp., *B. distachyon*, mangosteen (*Garcinia mangostana*), and cocoyam (*Xanthosoma sagittifolium*), the latter two being regionally important tropical "orphan crops."

iPBS for direct display of polymorphism

Unlike methods for retrotransposon isolation that rely on conserved protein coding domains (Pearce et al. 1999), the PBS primers also directly visualize polymorphisms for retrotransposon loci in the genome. Furthermore, this is the only retrotransposon-based method to our knowledge that has been shown to visualize polymorphism throughout the plant kingdom and for animals as well. Of the sequences matching tRNA in the genome, the greatest proportion consists of retroelements. The tRNA genes themselves comprise small families for each isoacceptor. For example, the rice genome contains 737 tRNA genes altogether, compared with 53,302 retrotransposons (Itoh et al. 2007). Moreover, the iPBS primers contain CCA at their 3′ termini, which is complementary to the 5′ TGG motif in PBS sites but which is not found in eukaryotic tRNA genes. In eukaryotes, 3′ terminal CCA is added post-transcriptionally by ATP(CTP):tRNA nucleotidyltransferase (Shi et al. 1998). Hence, given the difference in the number of tRNA genes and retrotransposons and their genomic position, the lack of tRNA mobility, and the specificity conferred by the 3′ CCA of iPBS, primers, iPBS selectively displays polymorphism in retrotransposon insertion sites.

Overall, most PBS primers were useful for iPBS (Tables 1, 2). Of the 12- and 13-mers tested, 31 out of 42 yielded many strong, even bands and only one primer (2379) gave poor results. For animal samples, 32 of the 13-mer primers gave excellent results and 4 of the primers poor outcomes. The iPBS can be carried out with single primers, for which the analyses above have been given, or with combinations of two primers. The banding patterns obtained when more than one primer is used will depend on the relative abundance of different retrotransposon families as well as on their distribution with respect to one another. The PBS primers can also be combined with microsatellite primers as in REMAP and with adapter primers as in SSAP (data not shown) for generation of additional scorable polymorphisms. In analyses of the three barley varieties, the iPBS method proved to be as informative as those obtained using IRAP, REMAP or SSAP, and about an equal level of polymorphism compared to IRAP and REMAP.

The iPBS method showed utility of a wide range of animals, including sheep, cow, yak and chicken. The yak and cow, both belonging to the genus *Bos*, displayed some invariant bands. The phenomenon of invariant retrotransposon display bands has been observed before, e.g. between barley and wheat (*Triticum* spp.; Kalendar et al. 2010), which are both members of the tribe Triticeae of the family Poaceae. Cow and yak are estimated to have diverged 74–78 thousand years ago (Ho et al. 2008); ERVs, targets for iPBS, have been components of mammalian genomes for millions of years and have been described in *Bos* (Xiao et al. 2008). The presence of invariant bands could be explained by the presence of ERV insertion sites in *Bos*, which have been maintained since before the radiation of the genus. This is consistent with the many conserved, orthologous ERV insertions in the primates, which have been preserved for millions of years (Lebedev et al. 2000).

Retrotransposons are known to be often activated by tissue culture (Cheng et al. 2006; Tadege et al. 2008; Wessler 1996). We were interested in the efficacy of iPBS to detect genomic polymorphism in plants regenerated from tissue culture, and examined regenerants from *Vaccinium* spp., apple, and maize. In *Vaccinium* and particularly in maize, iPBS polymorphisms appeared in the regenerated plants, consistent with retrotransposon activation. These data are likewise consistent with observations on particular groups of retrotransposons (Antonius-Klemola et al. 2006). Nevertheless, like all anonymous transposable element display methods, iPBS is not strictly quantitative; only insertions sufficiently close to one another to yield amplification products will be visualized.

In silico searches with PBS

The need for sensitive and accurate detection and annotation of retrotransposons grows commensurately with the speed of genome sequencing (Wicker et al. 2007). The LTR retrotransposons generally have been identified by similarity to diagnostic protein coding domains, particularly reverse transcriptase and integrase, or to LTRs of known retrotransposons. However, protein coding domains are absent from LARD and TRIM retrotransposons (Kalendar et al. 2004, 2008; Witte et al. 2001). Structure-based identification of LTRs, by the presence of long direct repeats and overabundance of short query strings in retrotransposons, is an alternative approach. The identification of LTRs is

greatly strengthened by the presence of an adjacent PBS sequence. Here, we found that we could identify new LTR retrotransposons without relying on a library of known LTRs or coding sequences by searching for the PBS domains. We were able to identify both superfamilies of LTR retrotransposons (*Gypsy* and *Copia*) in silico, as well as non-autonomous LARD and TRIM elements. The method was tested and verified on the databases for the genomes of rice, grape, and tomato, identifying previously un-annotated retrotransposons, particularly LARDs and TRIMs (Online Resource 2, Table S4), as well as for *A. thaliana* and *D. melanogaster* (data not shown). This suggests that searching PBS/LTR junctions is a sensitive method for de novo annotation.

Conclusions

The almost universal presence of a sequence complementary to tRNA, the PBS, in LTR retrotransposons has long been recognized but not previously exploited, to our knowledge, either for isolation of retrotransposons or directly displaying insertion site polymorphisms. The use of the PBS offers several advantages. First, it is universal, in that virtually all retrotransposons, including TRIMs and LARDs, which do not have internal coding domains, serve as targets. Second, it can be used both for cloning retrotransposons and ERVs as well as for displaying their insertion sites. Other methods, relying on conserved protein coding domains of retrotransposons, require walking to the LTRs in order to produce a marker system. Third, the PBS is adjacent to the LTR, facilitating its isolation. Last, a fairly small set of primers, once produced, can be used in any organism with a complement of matching elements. These advantages offer savings of time and money in the development of new marker systems for "orphan crops," wild species, and other species for which resources are limited or for which the development of SNP-based markers is impractical. The simple analytical systems for retrotransposon markers only strengthen this benefit.

Acknowledgments We thank the many collaborators listed in Online Resource 1 for gifts of plant materials. Eeva-Marja Turkki, Anne-Mari Narvanto, and Ursula Lönnqvist are thanked for their always excellent technical assistance. The work was supported by Ministry of Education of Czech Republic project MSM2678424601 and by the Academy of Finland, Grant 120810, Project Exbardiv of the ERA-NET Plant Genomics program.

References

Alix K, Heslop-Harrison JS (2004) The diversity of retroelements in diploid and allotetraploid *Brassica* species. Plant Mol Biol 54:895–909

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

Belyayev A, Kalendar R, Brodsky L, Nevo E, Schulman AH, Raskina O (2010) Transposable elements in a marginal plant population: temporal fluctuations provide new insights into genome evolution of wild diploid wheat. Mobile DNA 1:6

Boyko E, Kalendar R, Korzun V, Gill B, Schulman AH (2002) Combined mapping of *Aegilops tauschii* by retrotransposon, microsatellite, and gene markers. Plant Mol Biol 48:767–790

Chaparro C, Guyot R, Zuccolo A, Piegu B, Panaud O (2007) RetrOryza: a database of the rice LTR-retrotransposons. Nucleic Acids Res 35:D66–D70

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. Nucleic Acids Res 20:1333–1337

Cheng C, Daigen M, Hirochika H (2006) Epigenetic regulation of the rice retrotransposon Tos17. Mol Genet Genomics 276:378–390

Corpet F (1988) Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res 16:10881–10890

Ellis THN, Poyser SJ, Knox MR, Vershinin AV, Ambrose MJ (1998) Ty1-copia class retrotransposon insertion site polymorphism for linkage and diversity analysis in pea. Mol Gen Genet 260:9–19

Feschotte C, Jiang N, Wessler S (2002) Plant transposable elements: where genetics meets genomics. Nat Rev Genet 3:329–341

Flavell AJ, Dunbar E, Anderson R, Pearce SR, Hartley R, Kumar A (1992) Ty1-copia group retrotransposons are ubiquitous and heterogeneous in higher plants. Nucleic 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

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

Hedges DJ, Batzer MA (2005) From the margins of the genome: mobile elements shape primate evolution. BioEssays 27:785–794

Hizi A (2008) The reverse transcriptase of the Tf1 retrotransposon has a specific novel activity for generating the RNA self-primer that is functional in cDNA synthesis. J Virol 82:10906–10910

Ho SYW, Larson G, Edwards CJ, Tim H, Heupink TH, Lakin KE, Holland PWH, Shapiro B (2008) Correlating Bayesian date estimates with climatic events and domestication using a bovine case study. Biol Lett 4:370–374

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436:793–800

Itoh T, Tanaka T, Barrero RA et al (2007) Curated genome annotation of *Oryza sativa* ssp. *japonica* and comparative genome analysis with *Arabidopsis thaliana*. Genome Res 17:175–183

Jaillon O, Aury J-M, Noel B, Policriti A, Clepet C, French-Italian Public Consortium for Grapevine Genome Characterization et al (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463–467

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

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

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

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, Tanskanen J, Immonen S, Nevo E, Schulman AH (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-Jonsson K, Bolshoy A, Schulman AH (2004) Large retrotransposon derivatives: abundant, conserved but nonautonomous retroelements of barley and related genomes. Genetics 166:1437–1450

Kalendar R, Tanskanen J, Chang W, Antonius K, Sela H, Peleg O, Schulman AH (2008) Cassandra retrotransposons carry independently transcribed 5S RNA. Proc Natl Acad Sci USA 105:5833–5838

Kalendar R, Flavell AJ, Ellis THN, Sjakste T, Moisy C, Schulman AH (2010) Analysis of plant diversity with retrotransposon-based molecular markers. Heredity (in press)

Kelly NJ, Palmer MT, Morrow CD (2003) Selection of retroviral reverse transcription primer is coordinated with tRNA biogenesis. J Virol 77:8695–8701

Lebedev YB, Belonovitch OS, Zybrova NV, Khil PP, Kurdyukov SG, Vinogradova TV, Hunsmann G, Sverdlov ED (2000) Differences in HERV-K LTR insertions in orthologous loci of humans and great apes. Gene 247:265–277

LeGrice SFJ (2003) "In the beginning": initiation of minus strand DNA synthesis in retroviruses and LTR-containing retrotransposons. Biochemistry 42:14349–14355

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 R, Vitte C, Ma J, Mahama AA, Dhliwayo T, Lee M, Bennetzen JL (2007) A GeneTrek analysis of the maize genome. Proc Natl Acad Sci USA 104:11844–11849

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 trunculata*. BMC Genomics 8:427

Mak J, Kleiman L (1997) Primer tRNAs for reverse transcription. J Virol 71:8087–8095

Marquet R, Isel C, Ehresmann C, Ehresmann B (1995) tRNAs as primer of reverse transcriptases. Biochimie 77:113–124

Pearce SR, Stuart-Rogers C, Knox MG, 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

Pereira HS, Barao A, Delgado M, Morais-Cecilio L, Viegas W (2005) Genomic analysis of grapevine retrotransposon 1 (Gret 1) in *Vitis vinifera*. Theor Appl Genet 111:871–878

Schnable PS, Ware D, Fulton RS et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115

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

Shedlock AM, Okada N (2000) SINE insertions: powerful tools for molecular systematics. Bioassays 22:148–160

Shi P-Y, Maizels N, Weiner AM (1998) CCA addition by tRNA nucleotidyltransferase: polymerization without translocation? EMBO J 17:3197–3206

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

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

Tadege M, Wen J, He J, Tu H, Kwak Y, Eschstruth A, Cayrel A, Endre G, Zhao PX, Chabaud M, Ratet P, Mysore KS (2008) Large-scale insertional mutagenesis using the Tnt1 retrotransposon in the model legume *Medicago truncatula*. Plant J 54:335–347

Tang JQ, Korab-Laskowska M, Jarnik M, Cardinal G, Vanasse M, Melançon SB, Labuda D (1995) Alu-PCR combined with non-Alu primers reveals multiple polymorphic loci. Mamm Genome 6:345–349

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

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

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

Waugh R, McLean K, Flavell AJ, Pearce SR, Kumar A, Thomas BB, Powell W (1997) Genetic distribution of BARE-1-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). Mol Gen Genet 253:687–694

Wessler SR (1996) Turned on by stress. Plant retrotransposons. Curr Biol 6:959–961

Wicker T, Zimmermann W, Perovic D, Paterson AH, Ganal M, Graner A, Stein N (2005) A detailed look at 7 million years of genome evolution in a 439 kb contiguous sequence at the barley Hv-eIF4E locus: recombination, rearrangements and repeats. Plant J 41:184–194

Wicker T, Sabot F, Hua-Van A, Bennetzen J, Capy P, Chalhoub B, Flavell AJ, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH (2007) A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8:973–982

Witte CP, Le QH, Bureau T, Kumar A (2001) Terminal-repeat retrotransposons in miniature (TRIM) are involved in restructuring plant genomes. Proc Natl Acad Sci USA 98:13778–13783

Xiao R, Park K, Lee H, Kim J, Park C (2008) Identification and classification of endogenous retroviruses in cattle. J Virol 82:582–587

Yu J, Wang J, Lin W, Li S, Li H, Zhou J, Ni P et al (2005) The genomes of *Oryza sativa*: a history of duplications. PLoS Biol 3:e38