Rapid isolation of plant Ty1-copia group retrotransposon LTR sequences for molecular marker studies

Stephen R. Pearce1,2,*, Caroline Stuart-Rogers2, Maggie R. Knox3, Amar Kumar1, T.H. Noel Ellis3 and Andrew J. Flavell2
1Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, UK, 2Department of Biochemistry, University of Dundee, Dundee DD1 4HN, UK, and 3John Innes Centre, Colney, Norwich NR4 7UH, UK

Summary
The terminal sequences of long-terminal repeat (LTR) retrotransposons are a source of powerful molecular markers for linkage mapping and biodiversity studies. The major factor limiting the widespread application of LTR retrotransposon-based molecular markers is the availability of new retrotransposon terminal sequences. We describe a PCR-based method for the rapid isolation of LTR sequences of Ty1-copia group retrotransposons from the genomic DNA of potentially any higher plant species. To demonstrate the utility of this technique, we have identified a variety of new retrotransposon LTR sequences from pea, broad bean and Norway spruce. Primers specific for three pea LTRs have been used to reveal polymorphisms associated with the corresponding retrotransposons within the Pisum genus.

Introduction
Retrotransposons are ubiquitous in plant genomes (Flavell et al., 1992a; Kunze et al., 1997; Voytas et al., 1992). They are classified into two major families, based on the presence or absence of long-terminal repeats (LTRs) at the transposon ends (Boeke and Corces, 1989). LTR retrotransposons are present as large heterogeneous populations within plants (Flavell et al., 1992b; Matsuoka and Tsunewaki, 1996; Pearce et al., 1997; Vanderwiel et al., 1993). These retrotransposon populations are highly plastic, showing great variation in copy number and genomic localization between even closely related species (Casacuberta et al., 1997; Kumar et al., 1997; Pearce et al., 1996a; Pearce et al., 1996b). This variation has been exploited in retrotransposon-based molecular marker systems in pea and barley (Ellis et al., 1998; Flavell et al., 1998; Kalendar et al., 1999; Waugh et al., 1997). Several of these methods yield significant improvements over current marker technologies (Ellis et al., 1998; Waugh et al., 1997).

Retrotransposon-based markers require sequence information from the terminal regions of the mobile element. This is only available for a few of these mobile elements in a small number of plant species (reviewed by Kunze et al., 1997). The scarcity of LTR terminal sequence is therefore the main obstacle preventing the extension of this approach to all plants. In this report, we describe a method which rapidly delivers Ty1-copia retrotransposon LTR sequences directly from the genomic DNA of any higher plant. Using this technique, we have isolated 23 novel LTR sequences from two dicotyledonous plants (pea and broad bean) and a gymnosperm (Norway spruce). Primers specific to four of these new pea elements have been used to generate polymorphic SSAP markers in Pisum.

Results
General outline of the technique
This technique was devised to enable the rapid isolation of Ty1-copia retrotransposon terminal sequences. The LTR sequences of Ty1-copia retrotransposons are present at each end of the retrotransposon and are identical (Figure 1). The LTRs contain no conserved motifs which would allow their direct amplification by PCR. The nearest conserved sequence motifs to the 3′ LTR is within the adjacent RNaseH gene. We isolate and characterize PCR fragments between the RNaseH motif and a restriction site in the adjacent LTR sequence. Sequence analysis of these fragments identifies LTR terminal sequences which can then be used to design primers for retrotransposon-based marker analysis (Ellis et al., 1998; Kalendar et al., 1999; Waugh et al., 1997).

The method is shown schematically in Figure 2. Genomic DNA is digested partially with MseI, a four-cutter restriction endonuclease, to generate DNA fragments of an average size of 1 kb. The LTR of most Ty1-copia retrotransposons is within 500 bp of the RNaseH motif (Figure 1). Partial digestion therefore increases the number of fragments within this size range. Adapters are ligated to the MseI fragments and PCR carried out using a biotinylated RNaseH.
motif 1 primer and a primer homologous to the adapter (Figure 2). The adapters are unphosphorylated; this means that only one strand of the adapter is ligated to each restriction fragment. An initial denaturation step removes the unligated strand of the adapter, with the result that the amplification of non-specific products is greatly reduced. The RNaseH motif can lie at any position on the 1 kb template and so the biotinylated RNaseH products are on average 500 bp long. These fragments are isolated by affinity selection using streptavidin paramagnetic beads and recovered by another round of PCR with a nested primer to a second RNaseH motif (Figure 2). This produces a mixed PCR product which is subcloned and sequenced to identify the population of RNaseH-LTR terminal sequences.

**Identification of Ty1-copia retrotransposon RNaseH-LTR sequences**

Figure 3 shows the sequences obtained from three plant species, pea, broad bean and Norway spruce, using this method. Each sequence has features characteristic of Ty1-copia retrotransposons (Figure 1). Following motif 2, the RNaseH open reading frame, containing some or all of the seven conserved amino acid residues, continues for approximately 60 bp. This is followed by a polypurine tract (PPT) and a (T)TG-TG sequence motif characteristic of the 5’ end of a retrotransposon LTR. Beyond the LTR end there is very little sequence conservation except between very closely related sequences such as Tpa2 and Tpa3 and Tv5 and Tv6. The variation within the LTRs is shown clearly by the way in which two similar sequences from the RNaseH sequence, Tpa5 and Tpa6, are quite different in the LTR sequence.
structural features (Figure 3). Few duplicates or near duplicates were obtained, and so it is likely that many more new Ty1-copia group retrotransposon LTRs could be obtained by the characterization of more subclones. We conclude that this technique is highly efficient at revealing the LTR sequences of a diverse range of new Ty1-copia retrotransposon sequences from a wide range of higher plant species.

Molecular markers from LTRs of pea Ty1-copia group retrotransposons

The pea LTR retrotransposon PDR1 has been used successfully as a source of highly polymorphic molecular markers using sequence-specific amplification polymorphism (SSAP) (Ellis et al., 1998). To test the potential of the new P. sativum LTRs identified in this study, primers were designed from the LTR terminal sequences of three P. sativum elements (Tps3, Tps6 and Tps7) and SSAP analysis carried out on six Pisum accessions (Figure 4).

Each of the three primers produce bands on the SSAP gel for each accession, showing that each transposon is present within all the species. The banding pattern produced in each accession relates to the number and position of each retrotransposon insertion in each accession. The number of shared bands between each accession is related to the phylogenetic relationships of these accessions. The first six lanes in Figure 4 represent Tps7 insertions. Lanes (a) and (b) are identical, indicating that JI157 and JI281 are closely related. When each of the other accessions is compared, we see that as well as a number of shared bands there are also a number which are unique to each species. The two other retrotransposon primers Tps6 and Tps3 each give a distinct pattern, but the banding patterns in each set show more similarity between the P. sativum accessions (lanes a and b) than between the other Pisum species (lanes c–f). Although the data set is too small to draw any detailed conclusions, these findings are consistent with existing amplified fragment length polymorphisms (AFLP) and SSAP data based on the pea retrotransposon PDR1 (Ellis et al., 1998). The overall polymorphism level (the relative proportion of unshared bands) for each primer in these accessions is 77% for Tps3, 85% for Tps6 and 70% for Tps7. The differences in percentage indicate that each primer has a characteristic level of transposition in these accessions. The comparable polymorphism level for PDR1 in these same accessions is 93% (Ellis et al., 1998). This demonstrates that each of the three newly isolated retrotransposon markers has distinctive polymorphism characteristics, resulting from its individual transposition properties. The potential usefulness of using several markers simultaneously is discussed below.

Distribution and inheritance of Tps10 insertions in pea

To test the distribution and inheritance of the LTR markers, Tps10 SSAP analysis was performed on a recombinant inbred cross of two P. sativum accessions J115 and J1399 and 42 progeny (Figure 5). Tps10 was chosen as of all the markers it was most...
polymorphic between these lines. A number of bands are polymorphic between the two parental lines, four of these bands (Tps10/1–4) all behave as dominant markers within the progeny (lanes 1–42), and map to dispersed locations in the pea genome (data not shown).

Discussion

Molecular markers based on retrotransposons have proved to be more informative than non-transposon-based marker methods in the few cases tested to date (Ellis et al., 1998; Kalendar et al., 1999; Waugh et al., 1997). Unfortunately, retrotransposon-based markers are only available for a small number of well-characterized mobile elements in relatively few species groups. The method described here allows the rapid extension of this approach to include every higher plant species, including those for which no genomic information is available. We demonstrate this by the generation of heterogeneous populations of LTR sequences from two dicot angiosperms (pea and broad bean) and a gymnosperm (Norway spruce); additionally this approach has also yielded positive results in wheat, a monocot (B. Gribbon, P. Jack, A. Kumar and A.J. Flavell, unpublished data). As it is PCR-based, the method can yield sequence data for primer design in several days, enabling retrotransposon-based marker technologies to be employed immediately.

This technique not only makes retrotransposon-based marker analysis accessible for all plant species, it also

Figure 5. Distribution and inheritance of Tps10 insertions in pea.
Segregation of Tps10 LTR + T(ATT)-derived sequence-specific amplification polymorphisms in 42 progeny of a cross between two parental lines JI15 and JI399 (lanes 43 and 44). Four bands (Tps10/1–4, arrowed) which are polymorphic between the parental lines segregate in the progeny (lanes 1–42).
Rapid isolation of plant Ty1-copia LTR sequences

can increase the power of this approach by providing several different retrotransposon markers per species. A great variety of different Ty1-copia retrotransposons exist in most higher plants (Flavell et al., 1992a; Flavell et al., 1992b; Matsuoka and Tsunewaki, 1996; Pearce et al., 1996a; Pearce et al., 1996b; Pearce et al., 1997; SanMiguel et al., 1996; Vanderwiel et al., 1993; Voytas et al., 1992). Some of these, which have been transposing in the recent past, are extremely polymorphic within species and may be used as markers for linkage analysis and intra-specific diversity studies (Ellis et al., 1996; Waugh et al., 1996). Recombinant inbred (RI) lines from the cross JI15 × JI399 have been described previously (Ellis et al., 1992). Pirus DNA was isolated as described by Ellis (1994). Vicia faba (cv. Troy) was from the Vicia germplasm collections of the Scottish Crop Research Institute. V. faba DNA was isolated by the method of Saghai-Maroof et al. (1984). Picea abies (Norway spruce) was obtained from the University of Dundee Botanical Garden and DNA was isolated using Phytopure (Amersham Pharmacia Biotech, Amersham, UK).

Computer analysis

Computer analysis was carried out using the SEQNET facilities at the BBSRC Computing Centre, Daresbury, UK, using the GCG program (Version 8, August 1994; Genetics Computer Group, Madison, Wisconsin, USA).

Generation and characterization of retrotransposon sequences

Pirus lines JI157, JI281, JI15 and JI399 (P. sativum), JI225 (P. abyssinicum), JI1006 (P. fulvum), JI055 (P. elatus) and JI1794 (P. humile) were selected from the John Innes Germplasm core collection (Matthews and Ambrose, 1994). Recombinant inbred (RI) lines from the cross JI15 × JI399 have been described previously (Ellis et al., 1992). Pirus DNA was isolated as described by Ellis (1994). Vicia faba (cv. Troy) was from the Vicia germplasm collections of the Scottish Crop Research Institute. V. faba DNA was isolated by the method of Saghai-Maroof et al. (1984). Picea abies (Norway spruce) was obtained from the University of Dundee Botanical Garden and DNA was isolated using Phytopure (Amersham Pharmacia Biotech, Amersham, UK).

**Experimental procedures**

**Plant materials**

Pirus lines JI157, JI281, JI15 and JI399 (P. sativum), JI225 (P. abyssinicum), JI1006 (P. fulvum), JI055 (P. elatus) and JI1794 (P. humile) were selected from the John Innes Germplasm core collection (Matthews and Ambrose, 1994). Recombinant inbred (RI) lines from the cross JI15 × JI399 have been described previously (Ellis et al., 1992). Pirus DNA was isolated as described by Ellis (1994). Vicia faba (cv. Troy) was from the Vicia germplasm collections of the Scottish Crop Research Institute. V. faba DNA was isolated by the method of Saghai-Maroof et al. (1984). Picea abies (Norway spruce) was obtained from the University of Dundee Botanical Garden and DNA was isolated using Phytopure (Amersham Pharmacia Biotech, Amersham, UK).
washed in 0.2× SSC, resuspended in 30 μl 0.2× SSC and 20 μl (50%) of eluate added. This was incubated for 10 min at 20°C with occasional shaking. The magnespheres were washed in 100 μl 0.2× SSC twice, then in 100 μl 1× Pfu reaction buffer (Stratagene) twice. The magnespheres with the bound PCR product were resuspended in 50 μl 1× Pfu buffer, and 2 μl of the suspension was PCR-amplified with 0.8 μg of RNaseH motif 2 primer (5'-GCNGGATYNNTACNA-3'), 0.15 μg MseI adapter primer, in 50 μl Pfu reaction buffer containing 0.2 mM dNTPs and 1 unit Pfu polymerase (Stratagene). The product was purified by Qiaquick PCR purification column (Qiagen) and digested with BamHI restriction endonuclease. The digest was again purified on a Qiageld column and the product 5' phosphorylated using T4 polynucleotide kinase then ligated into a phosphatase-treated M13mp18 vector which had been digested with BamHI and Hincll. The recombinants were transformed into XLIBlue supercompetent cells (Stratagene), the insert sizes confirmed by PCR and sequences determined using Sequenase (Amersham).

**Sequence-specific amplification polymorphism (SSAP)**

marker analysis

SSAP analysis was carried out as described by Waugh *et al.* (1997), except that the template DNA was digested with MseI and adapters ligated. The template was pre-amplified with primer M homologous to the adapter sequence (M = 5'-GATGAG-TGCCTGAGTAA) with 30 cycles of 94°C for 30 sec, 45°C for 30 sec and 72°C for 60 sec.

Selective amplification was carried out on 1 μl of diluted pre-amplification with 1 μl of either Tps3, Tps6 or Tps7-labelled LTR oligonucleotide (Tps3 = 5'-CTCTTGGGATATTAACCACAC-3'; Tps6 = 5'-GTGAGATGGTTATATGTC-3'; Tps7 = 5'-CTATAATAACATACAACAACG-3'), with either 50 ng primer M, M(C) (5'-GATGAG-TGCCTGAGTAA-3') or M(A)C (5'-GATGAGTGCCTGAGTAAAC-3') as shown in Figure 1 with 20 cycles of 94°C for 30 sec, 55°C (reducing by 1°C for each successive cycle) for 30 sec and 72°C for 605 sec. This was followed by 20 cycles of 94°C for 30 sec, 45°C for 30 sec and 72°C for 60 sec.

Tps10 SSAP was carried out using a Tps10 LTR primer (5'-GGAATGATAGGCCTTGCC-3') and a TaqI adapter primer with three bases of selection T(ATT) under conditions as described by Ellis *et al.* (1998).

**Acknowledgements**

S.R.P. is supported by EC Biotechnology project BIO4 CT96-0508 under Framework IV. C.S.-R. is funded by a BBSRC earmark studentship. A.K. is funded by the Scottish Office Agriculture, Environment and Fisheries Department. We wish to thank Mark Taylor and Barbara Gibbon for useful discussions.

**References**


Rapid isolation of plant Ty1-copia LTR sequences 717


EMBL accession numbers PSA243035–PSA243043 (Tps2–Tps10), AJ243099–AJ243104 (Tvf1–Tvf6) and AJ243312–AJ243319 (Tpa1–Tpa8).