Mol Gen Genet (2000) 263: 898–907

ORIGINAL PAPER

S. R. Pearce · M. Knox · T. H. N. Ellis
A. J. Flavell · A. Kumar

Pea Ty1-*copia* group retrotransposons: transpositional activity and use as markers to study genetic diversity in *Pisum*

Received: 23 January 2000 / Accepted: 24 March 2000

Abstract The variation in transposition history of different Ty1-*copia* group LTR retrotransposons in the species lineages of the *Pisum* genus has been investigated. A heterogeneous population of Ty1-*copia* elements was isolated by degenerate PCR and two of these (*Tps12* and *Tps19*) were selected on the basis of their copy number and sequence conservation between closely related species for further in-depth study of their transpositional history in *Pisum* species. The insertional polymorphism of these elements and the previously characterised *PDR1* element was studied by sequence-specific amplification polymorphism (SSAP). Each of these elements reveals a unique transpositional history within 55 diverse *Pisum* accessions. Phylogenetic trees based on the SSAP data show that SSAP markers for individual elements are able to resolve different species lineages within the *Pisum* genus. Finally, the SSAP data from all of these retrotransposon markers were combined to reveal a detailed picture of the intra and inter-species relationships within *Pisum*.

Key words Retrotransposon · Genetic diversity · Molecular taxonomy · Pea · DNA markers

Communicated by R. Hagemann

S. R. Pearce (✉)
Plant Stress Unit, School of Biology, University of Sussex, Falmer, Brighton BN1 9QG, UK
e-mail: S.R.Pearce@sussex.ac.uk
Tel.: +44-1273-678872; Fax: +44-1273-678433

M. Knox · T. H. N. Ellis
Department of Applied Genetics, John Innes Centre, Norwich NR4 7UH, UK

A. J. Flavell
Department of Biochemistry, MSI/WTB, University of Dundee, Dundee DD1 5EH, UK

A. Kumar
Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, UK

Introduction

Retrotransposons are mobile genetic elements, which transpose via an RNA intermediate and are classified by the presence or absence of long terminal repeats (LTRs). The LTR retrotransposons can be divided into two groups, the Ty1-*copia* group and the Ty3-*gypsy* group, based on their sequence similarity to the prototype elements in *Drosophila* and yeast (Boeke and Corces 1989). Both LTR and non-LTR retrotransposons are present in plants, usually as high-copy-number dispersed sequences (Schmidt 1999; reviewed by Kunze et al. 1997; Kumar and Bennetzen 1999).

The Ty1-*copia* retrotransposons are the best studied retrotransposon group in plants. They are usually present in plant genomes as high-copy-number heterogeneous populations of sequences that are distributed throughout the genome (Schmidt et al. 1995; Pearce et al. 1996; Heslop-Harrison et al. 1997; Kumar et al. 1997). As the mechanism of retrotransposition is replicative, the parental copy is preserved, with transpositional activity resulting in new insertions of the element at distant sites in the genome. Transposition of retrotransposons therefore results in an increase in the copy number of the elements, sometimes to the point where they account for large percentages of the plant genome (Pearce et al. 1996; SanMiguel et al. 1996).

The high activity and stable inheritance of these elements makes them very informative as molecular markers. Several retrotransposon-based marker systems have been devised, which reveal insertional polymorphism generated by the transposition of these elements. These have proved useful for linkage, phylogeny and diversity studies (Purugganan and Wessler 1995; Waugh et al. 1997; Ellis et al. 1998; Flavell et al. 1998; Kalendar et al. 1999; Pearce et al. 1999).

Although retrotransposon sequences have been described in many plant species, the transpositional activity of retrotransposons within species lineages

is largely unknown. The *BARE-1* element of barley and *PDR1* of pea have both been active in several related species (Pearce et al. 1997; Ellis et al. 1998; Gribbon et al. 1999), and there is some evidence that *PDR1* activity has varied between individual *Pisum* lineages (Ellis et al. 1998). To address the question of whether other Ty1-*copia* group elements transpose at different rates in these lineages and investigate whether this influences the resolution of the corresponding retrotransposon-based molecular marker systems, we have selected new pea retrotransposons from the diverse population of Ty1-*copia* group retrotransposons in this species and developed a molecular marker system for each element. The insertional polymorphism of these transposons has been determined within geographically diverse *P. sativum* genotypes and related wild *Pisum* species. This study reveals that individual elements have different transpositional histories in individual *Pisum* lineages. The combination of sequence-specific amplification polymorphism (SSAP) data from four different retrotransposon markers improves the resolution of SSAP, relative to that obtained using a single element, resulting in the derivation of a phylogeny for *Pisum* which is the most detailed reported to date.

Materials and methods

Plant materials

Pea lines were selected from the John Innes *Pisum* Germplasm core collection (Matthews and Ambrose 1994) *Vicia narbonensis* and *V. bithynica* were from the SCRI *Vicia* germplasm collection (SCRI, Dundee, UK). DNAs from *V. narbonensis* and *V. bithynica* were prepared by the method of Saghai-Maroof et al. (1984). Pea DNAs were prepared from leaves as described by Ellis (1994).

Generation of reverse transcriptase sequences by PCR

Conserved elements of the reverse transcriptase domain were amplified by PCR using primers and conditions described previously (Flavell et al. 1992a). The PCR products were subcloned into M13mp18 vector and individual subclones were sequenced as described previously (Flavell et al. 1992a). Accession numbers of *Tps11–43* are AJ405217 to AJ405247, respectively.

Computer analysis

Sequence analysis was carried out on the SEQNET facilities at the HGMP Resource Centre, Hinxton, CB10 1SB, UK (http://www.hgmp.mrc.ac.uk/). DNA sequences were compared using the TREE/PAPA package of programmes (Doolittle and Feng 1990; Feng and Doolittle 1990), or CLUSTALV (a modification of Clustal; Higgins and Sharp 1988). Trees were constructed directly using CLUSTALV or TREE (a neighbour-joining method; Doolittle and Feng 1990) and PAPA3 (a parsimony-based method; Feng and Doolittle 1990). TREE produced a phylogenetic tree with the lowest standard deviation and this was adopted. Average branch lengths were derived from trees generated from nucleotide sequences using PAPA3. Sequence similarities were calculated using GAP, nucleotide substitution rates were calculated using DIVERGE and NEWDIVERGE, the protein alignment and tree construction used CLUSTAL V.

Isolation of lambda clones containing retrotransposon sequences

Reverse transcriptase fragments corresponding to individual retrotransposon sequences were generated from individual M13 clones by PCR using primers and conditions as above. Radiolabelled PCR fragments were used to screen a *P. sativum* (cv. Birte) genomic lambda library (a gift from E. Hitchin, John Innes Centre, Norwich UK). Strongly hybridising clones were selected under stringent conditions (0.l× SSC at 65 °C; Sambrook et al. 1989).

Determination of LTR sequences for Ty1-*copia* group retrotransposons from lambda clones

DNA (2.5 μg) from individual strongly hybridising lambda clones was digested with 2 U of *Mse*I (New England Biolabs) in a 50-μl volume of RL buffer (Waugh et al. 1997) for 2 h. Then 12.5 pmol of ds *Mse*I adapter (Table 1), l U of T4 DNA ligase, and rATP to 0.l mM was added and the mixture was incubated at 37 °C for 3 h. The reaction was then diluted by the addition of 450 μl of TE. PCR was carried out in 40 μl of l× *Taq* polymerase buffer (Promega) containing 10 ng of γ-^{33}P-labelled RNase H-specific primer (RNH, Table 1, labelled as in Ellis et al. 1998), 60 ng of *Mse*I adapter primer MP (Table 1), 0.2 mM each dNTP, 0.5 U of *Taq* polymerase and 1 μl of the ligation (above). PCR was carried out in a Techne Genius instrument for 25 cycles of 94 °C for 60 s, 50 °C for 60 s, 72 °C for 60 s. Following PCR, 40 μl of gel loading buffer was added to each sample and they were subjected to electrophoresis on a denaturing 6% polyacrylamide sequencing gel (Sambrook et al. 1989). Electrophoresis was carried out at 70 W for 120 min, and the gel was dried unfixed onto Whatman 3MM paper. The labelled fragments were visualised by autoradiography using Konica X-ray film and the corresponding bands were excised from the gel. These gel fragments were rehydrated for 5 h in 100 μl of 1× TE and the products recovered by re-amplification in a 50-μl volume of PCR buffer containing 0.8 μg of RNH primer (Table 1), 0.15 μg of MP primer, 0.2 mM dNTPs, l μl of gel eluate and 0.5 U of *Pfu* DNA polymerase (Stratagene). PCR was carried out using 25 cycles of 94 °C for 60 s, 50 °C for 60 s, 72 °C for 60 s. The DNA products were purified using Qiaquick PCR purification columns (Qiagen), digested with *Bam*HI and subcloned into the M13mpl8 vector digested with *Hin*cII and *Bam*HI. The nucleotide sequences of single-stranded M13 DNAs were determined using Sequenase.

Analysis of diversity by sequence-specific amplification polymorphism (SSAP)

Using the methodology described previously for *PDR1* (Ellis et al. 1998) SSAP data were generated from 55 *Pisum* accessions with three new retrotransposon LTR-derived primers, *Tps12a*, *Tps12b* and *Tps19* (Table 1). *Pisum* DNA was digested with *Taq*I and ds *Taq* adapter was ligated to the fragments (Table 1). SSAP was

Table 1 Adapter and primer sequences

Adapter/primer	Sequence (5′ → 3′)
Ds *Mse*I adapter	GACGATGGATCCTGAG/TACTCAGGA TCCTA
Ds *Taq* adapter	ATGAGTCCTGAA/CGTTCAGGACTCAT
MP	GATGGATCCTGAGTAA
PDR1	ATTCACCAGCTTGAGGGGAG
RNH	MGNACNAASCAYATHGA
T + T	ATGAGTCCTGAACGAT
Tps12a	GGGCTTTGACTAATGGACCTC
Tps12b	GCCGTTATCTCACCGGCGCTG
Tps12bL	GCCTCGAGGTCGACTGCG
Tps19	GGAGTAGAAGTAGAGAGCC

carried out with labelled retrotransposon-specific primers and adapter primers in the following combinations: *Tps12a*/T + ATT; *Tps12b*/T + ATT; *Tps19*/T + T (Table 1). The frequencies of band sharing were determined, and unrooted trees were constructed using the neighbour-joining (NJ) algorithm (Saitou and Nei 1987) with simple matching. Principal coordinate analysis was performed using the Genstats package, and band sharing was scored according to Jaccard (Numerical Algorithms Group, Oxford). The UPGMA (unweighted pair group method with arithmetic averages) was applied and Nei tree comparison was made using the Phylip package (http://evolution.genetics.washington.edu/phylip/software.html).

Results

Isolation, phylogenetic analysis and copy number determination of the Ty1-*copia* group retrotransposons of pea

The first requirement for this study was to characterise the population of Ty1-*copia* retrotransposon sequences from pea. Retrotransposon sequences were generated by PCR from *P. sativum* (JI15) using primer regions corresponding to two highly conserved domains of the reverse transcriptase gene (Flavell et al. 1992a). The PCR product was subcloned and individual subclones sequenced. The phylogenetic relationships between these sequences were determined and are shown in Fig. 1. The population of Ty1-*copia* group retrotransposons in pea is heterogeneous, but within this heterogeneity there are a few subgroups of closely related sequences, such as the *PDR1*, *Tps12*, *Tps25* and *Tps27* subgroups; this is typical of most higher plant genomes.

The copy numbers of a representative selection of individual element subgroups were estimated by hybridising RT sequences from individual PCR clones to a *P. sativum* genomic library. The proportion of hybridising plaques in the library after stringent washing of filters was used to estimate the copy number of the corresponding element subgroups (Fig. 1). The library has an average insert size 1.5×10^4 bp. As the pea genome contains 4.3×10^9 bp, 2.9×10^5 plaques represent one pea genome equivalent. The different transposons investigated have a wide range of copy numbers, varying from approximately 30 for *Tps12* to 200 for *Tps15*, 600 for *Tps19* and 3000 for *Tps26* (Fig. 1).

The reverse transcriptase sequences from pea were compared to existing Ty1-*copia* group sequences from pea (Lee et al. 1990; Flavell et al. 1992b), soybean (Laten et al. 1998), and *Vicia* species (Pearce et al. 1996; this study, Fig. 2). Many of the pea sequences are related to *Vicia* retrotransposons, for example, the *Tps12* subgroup of pea (*Tps12*, *Tps17* and *Tps39*) is related to the *Fab64* sequence from *V. faba*, *Mel31* from *V. melanops* and *Sat49* from *V. sativa*. A number of elements are represented both in *Vicia* and *Pisum* (Fig. 2). This suggests that the common ancestor of these species also had a heterogeneous population of Ty1-*copia* group elements.

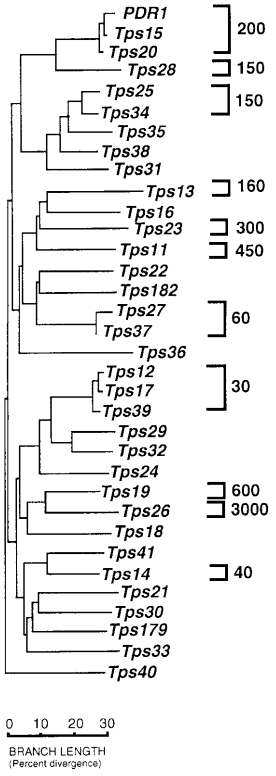
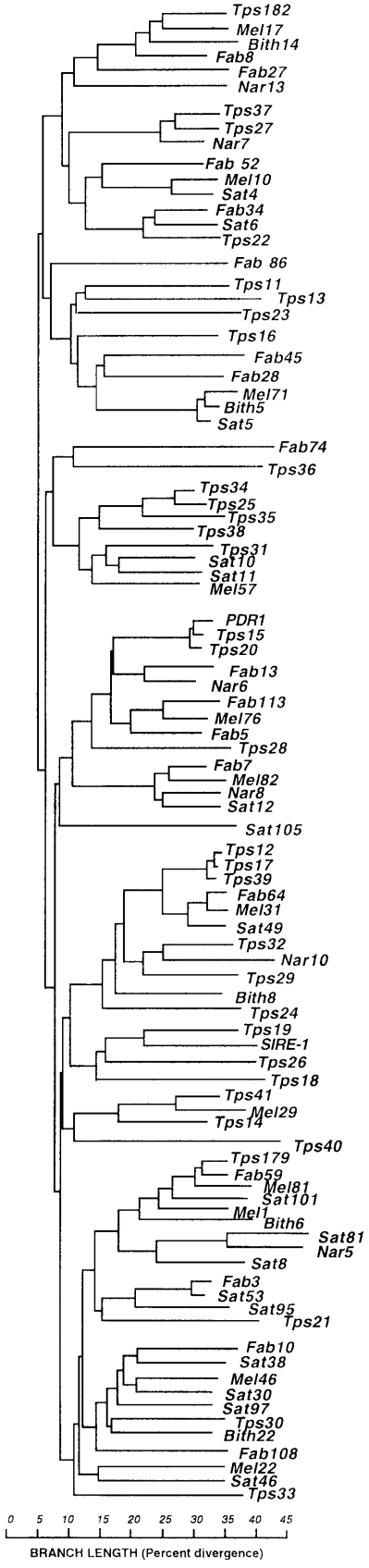


Fig. 1 Pea Ty1-*copia* group retrotransposons. Phylogenetic analysis of nucleotide sequences representing fragments of the reverse transcriptase genes of *Pisum sativum*. Sequences were analysed with the TREE/PAPA set of programs (Feng and Doolittle 1990; Doolittle and Feng 1990). Vertical branch lengths are of no significance. Divergences are indicated by *horizontal branch lengths*; as the tree is unrooted no significance attaches to the relative lengths of the two deepest branches of the tree. The sum of these two lengths, however, is equal to the distance separating these from the other sequences. Subgroups are named after the lowest integer sequence within. Copy number estimates for selected subgroups, obtained by genomic library screening of subcloned RT domains, are indicated in *brackets*

Characterisation of new retrotransposon LTRs for transposition analysis

The transpositional activity of the many retrotransposons shown in Fig. 2 is unknown, with the single exception of *PDR1* in *Pisum* (Ellis et al. 1998). We chose two elements – *Tps12* and *Tps19* – for further investigation, on the basis of their contrasting characteristics. *Tps12*, as determined from genomic library screening, is a low-copy-number element, which is extremely well conserved between *Pisum* and *Vicia* species (Fig. 1, Table 2). The Ks/Ka ratio (synonymous versus non-synonymous nucleotide substitution rate) of *Tps12* homologues between these species is very high (Table 2), showing that the reverse transcriptase sequences have been under strong selection, as would be the case for elements which are transpositionally active. *Tps19* in contrast has no detectable relatives in any of the *Vicia* species (Fig. 2) but is present in pea in high copy number (Fig. 1).

One of the best methods for studying transpositional activity of individual retrotransposons is by analysing



◀

Fig. 2 Comparison of Ty1-*copia* group retrotransposons in *Pisum, Vicia* and *Glycine* species. Phylogenetic analysis of nucleotide sequences representing fragments of the reverse transcriptase genes. Divergences between retrotransposon sequences are indicated in distance units (Feng and Doolittle 1990) by the *horizontal branch lengths*, vertical lengths have no significance. All trees are unrooted, so the relative lengths of the two horizontal sections of the deepest branches have no significance. However, the sum of the two lengths is equal to the distance separating these from the other sequences

their insertional polymorphisms by SSAP (Waugh et al. 1997; Ellis et al. 1998; Gribbon et al. 1999; Pearce et al. 1999). SSAP amplifies the DNA between a primer corresponding to the terminal sequence of the retrotransposon and a nearby restriction site in the flanking DNA sequence. Each retrotransposon insertion is represented as a distinct band on a sequencing gel. Sequence data from the terminal regions of each retrotransposon are required in order to design a primer for SSAP analysis. LTR terminal sequences were obtained for *Tps12* and *Tps19* by isolating genomic lambda clones containing these elements. The terminal sequences were then identified within these clones by a modification of the SSAP technique (see Materials and methods). In brief, clone DNA was digested with a frequently cutting restriction endonuclease and adapters were ligated to the resulting fragments. PCR was carried out with a primer homologous to the adapter primer and a radio-labelled degenerate primer RNH (Table 1), corresponding to a conserved motif within the RNaseH domain of the retrotransposon, just upstream of the 3′ LTR. The labelled RNase H-LTR fragments produced in this reaction were resolved by gel electrophoresis, recovered from the gel, re-amplified and sequenced. As the 3′ LTR sequence contained within these DNA fragments is identical to the extreme 5′ terminus of the element, this sequence can be used to design SSAP primers (Waugh et al. 1997; Gribbon et al. 1999; Pearce et al. 1999).

The RNaseH-LTR terminal sequences determined for *Pisum* Ty1-*copia* group retrotransposons *Tps12* and *Tps19* are shown in Fig. 3. They contain the conserved QIADIFTK motif followed by a polypurine tract and the (T)TG sequence characteristic of an RNaseH-LTR junction. (Pearce et al. 1999). Surprisingly, the *Tps12* lambda clone gave two different LTR sequences, *Tps12a* and *Tps12b*. The RNaseH domains and polypurine tracts of the two *Tps12* sequences are virtually identical, up to 6 bp inside the LTR, from which point onwards

Table 2 Sequence conservation within the *Tps12* subgroup

Species (sequence)[a]	*P. sativum*	*V. faba*	*V. melanops*	*V. sativa*
P. sativum	7.7 (10.10)	25.1 (6.1)	19.0 (4.8)	15.4 (1.7)
V. faba	–	–	5.8 (5.2)	1.0
V. melanops	–	–	2.0 (3.0)	3.9 (1.7)

[a] Numerical values are averages of Ks/Ka for all possible combinations of the sequences. Ks is the number of synonymous substitutions per synonymous site and Ka is the number of non-synonymous substitutions per non-synonymous site. Standard deviations are given in *parentheses*

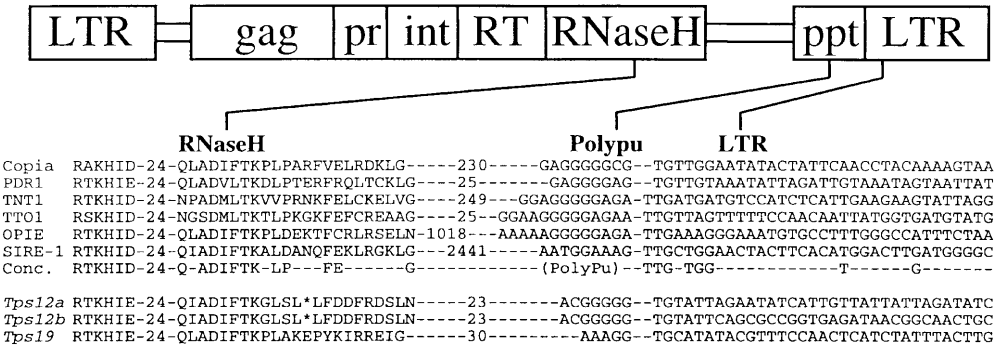


Fig. 3 Ty1-*copia* group retrotransposons: structural features of RNase H, polypurine tract (PPT) and LTR sequences. A consensus of a selection of previously characterised Ty1-*copia* retrotransposon RNase H, PPT and 3′ LTR sequences is presented, based on a selection of existing elements. Similar sequences for *Tps19*, *Tps12a* and *Tps12b* are shown. The *numbers* indicate intervening sequence of the length indicated; *hyphens* indicate gaps

they show no similarity (Fig. 3). To determine which of these RNaseH-LTR sequences connect with the *Tps12* RT fragment a PCR was performed on the *Tps12* clone from the genomic library, using a primer corresponding to the TAFLHG motif of the RT gene (Flavell et al. 1992a) and primers specific for each LTR (Tps12a, Tps12bL, Table 1). PCR fragments of 1300 bp (the expected size of these fragments for a Ty1-*copia* group retrotransposon) were obtained in both cases. Sequence analysis of these products confirmed that both LTRs are connected to a *Tps12* RT sequence (data not shown). This implies that two *Tps12* RT/RNaseH domains are present on the same genomic lambda clone, but each is connected to a different LTR sequence, the implications of this surprising result are discussed later.

Insertional polymorphism of new retrotransposons in *Pisum*

We have assayed the pattern of insertion site polymorphism for *Tps12a*, *Tps12b* and *Tps19* using SSAP analysis with LTR-specific primers for a set of 55 *Pisum* accessions; these accessions have been characterised previously using *PDR1* SSAPs (Ellis et al. 1998). *Tps12a* and *Tps12b* generated 55 and 77 informative markers, respectively. *Tps19* generated 38 markers, while *PDR1* generated an average of 69 informative markers. These numbers indicate that there is a high level of polymorphism for all these sequences.

All of the SSAP data from the four retroelements was combined and the 446 informative markers used to generate the dendrogram shown in Fig. 4. This dendrogram is our most accurate estimate of the phylogenetic relationships between these accessions. The inclusion of the theoretical ancestral condition "all zero" has little overall influence on the tree (not shown) but indicates the position of the presumed ancestor to all of these accessions which existed prior to any of the transposition events that gave rise to the observed SSAP bands.

Different retrotransposon families have different transpositional histories

The dendrogram shown in Fig. 4 has many of the features of the previous *PDR1*-based analysis (Ellis et al. 1998; Vershinin and Ellis 1999), but some differences are discernible, suggesting that the patterns of activity, and hence the degrees of variation, of the newly identified retrotransposons in these plant lineages are not identical to those of *PDR1*. We attempted to describe the differences in transposition histories between elements using two measures, the first is a measure of fixation within groups (F_{ST}; calculated as the ratio between the observed and expected variance in band frequency; Gonzáles-Candelas and Palacios 1997) and the second is Nei's genetic distance (D; Nei 1973) between groups of accessions (a to g) indicated in Fig. 5. For *PDR1*, five independent subsamples of 50 randomly selected markers are plotted, and for *Tps12b* two similar distinct samples of 30 markers are plotted. The F_{ST} estimates (Fig. 5a) overlap to some degree, suggesting that the majority of the variation within these groups is a consequence of the set of accessions which has been used. However, for two groups, *P. fulvum* (b) and the subset 'c' of *P. elatius*, a difference between *Tps12b* and *PDR1* appears to be well supported from the subsampling. For both *Tps12a* and *Tps19* in *P. fulvum* the F_{ST} values are higher than, and outside the range of, those calculated for *PDR1* and *Tps12b*. Similarly *Tps12a* appears to have reached a lower level of fixation than the other elements in *P. abyssinicum*, while *Tps19* has a higher F_{ST} value in both the 'a' and 'e' groups of *P. sativum*. These observations are consistent with there being small but significant differences in the transposition rates of these different elements within selected subgroups of extant *Pisum*.

Estimates of Nei's genetic distance have been made for the same samples of accessions, and of the 21 possible pairwise distances there are many overlapping distributions of estimates of D. Again this suggests that a large component of the D value is attributable to the accessions sampled. Nevertheless, seven pair-wise values show clear differences between the behaviour of *PDR1* and *Tps12b* (Fig. 5b). Note that groups b, c and e differ

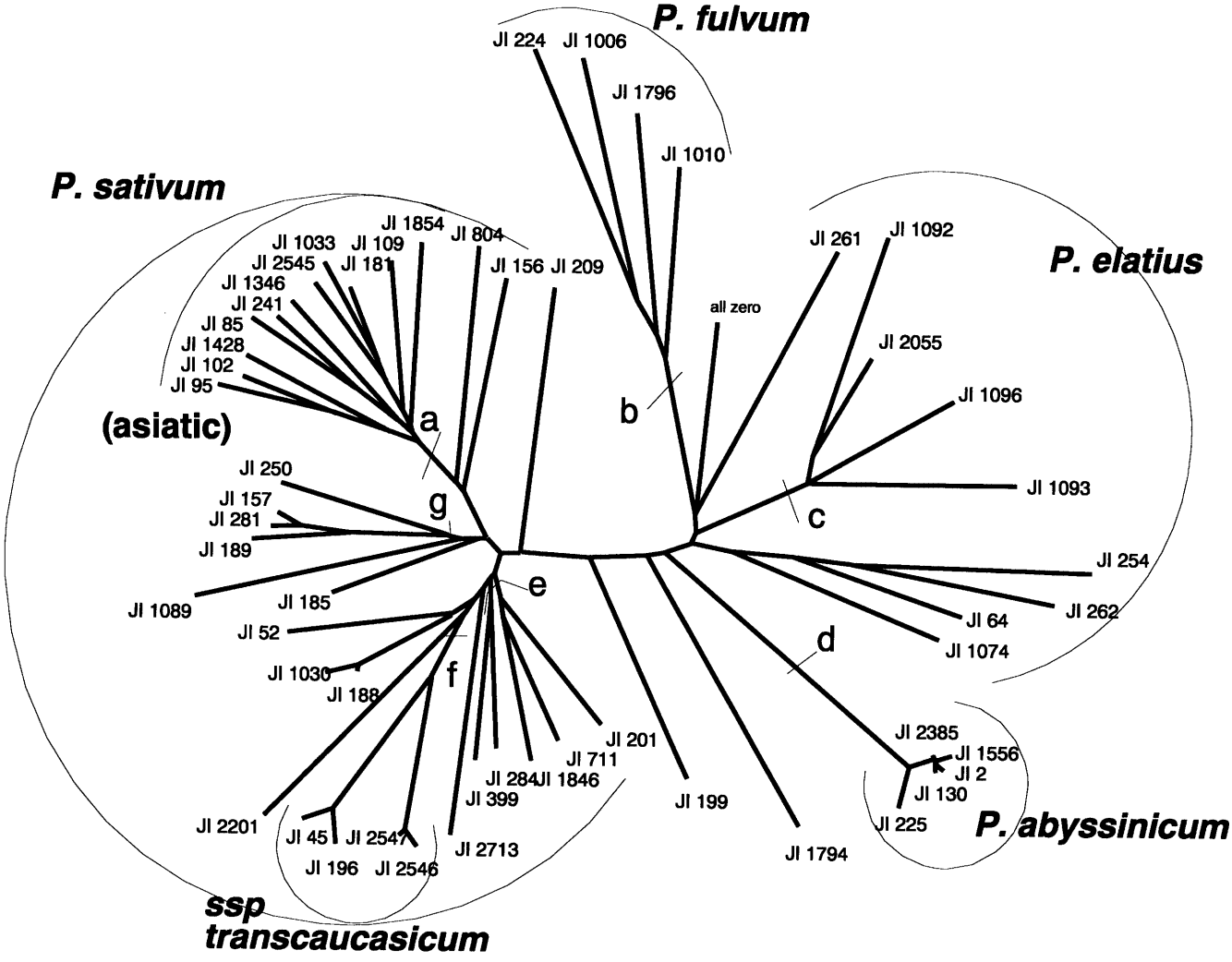


Fig. 4 Neighbour-joining tree of *Pisum* accessions. An SSAP-based NJ tree has been constructed from 446 informative markers as previously described (Ellis et al. 1998). The dendrogram was constructed from a distance matrix calculated by simple matching using the Neighbour Joining (Saitou and Nei 1987) algorithm in the Phylip package (http://evolution.genetics.washington.edu/phylip/ software.html). Accessions are identified by their numbers in the John Innes Germplasm Collection. The main groups corresponding to species and subspecies are indicated by *arcs*. Note that some accessions are placed on this tree in a way which is not consistent with these designations (e.g. JI2201 and JI199 have been classified as *P. elatius*; see Ellis et al. 1998) for further details. The groups marked a–g are discussed in the text and provide the basis for the analysis in Fig. 5

in their F_{ST} estimates for *PDR1* and *Tps12b*, but these are not the major groups exhibiting differences in the estimates of D. This suggests that there were differences in transposition rates in the lineages leading to the selected groups of accessions we have examined.

It is clear from Fig. 5b that, where different, the estimates of D for *Tps12b* are consistently lower than those for *PDR1*, presumably reflecting a much earlier bout of *Tps12b* transposition. Figure 5a shows less marked differences in recent transposition, and the *Tps12b* F_{ST} estimates are consistently higher than those for *PDR1*. *Tps12b* thus gives less discrimination than *PDR1*. Taken together, these data suggest that the maximal rate of *Tps12b* transposition preceded that of *PDR1*.

The accuracy of SSAP markers in phylogenetic studies is determined by the transpositional history of each marker

To visualise the differences in transposition properties between the four elements in *Pisum*, dendrograms were calculated using SSAP data for each retrotransposon in a 24-accession subset of the 55 *Pisum* accessions (Fig. 6). The 24 accessions represent six previously described species, sub-species or regional variants (four accessions each). Similar numbers of markers were used per tree to ensure that the accuracies of the trees were comparable.

The four trees share many similarities; for example, all four trees agree that the *P. abyssinicum* and *P. fulvum* accessions each form well-separated and robust clades. As well as the overall similarities between the trees there are a number of clear differences. *PDR1* groups European *P. sativum* with *P. abyssinicum* but *Tps12a*, in

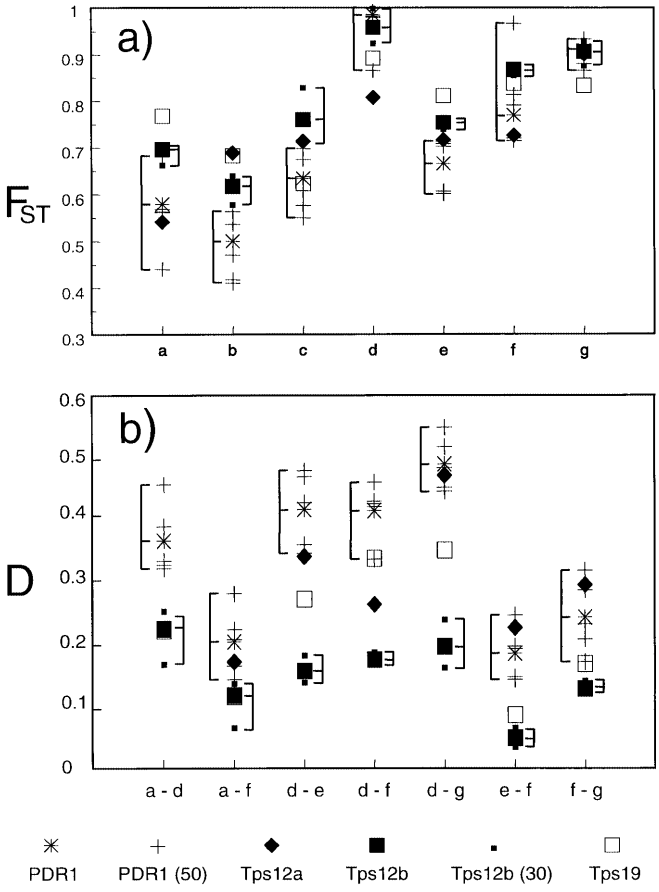


Fig. 5a, b F_{ST} and D values estimated using different retrotransposons. **a** F_{ST} estimates for seven *Pisum* groups (indicated a–g in Fig. 4) are plotted. The Asiatic *P. sativum* group (a) comprises accessions JI85, JI95, JI102, JI109, JI181, JI241, JI1033, JI1346, JI1428, JI1854, and JI2545. The *P. fulvum* accessions (b) were JI224, JI1006, JI1010, and JI1796. The *P. elatius* subset of accessions (c) were JI1092, JI1093, JI1096 and JI2055. *P. abyssinicum* (d) was represented by JI2, JI130, JI225, JI1556, and JI2385. The *P. sativum* cultivars (e) were JI201, JI284, JI399, JI711, and JI1846. The *P. sativum* ssp transcaucasicum accessions (f) were JI45, JI196, JI2546 and JI2547. The *P. sativum* African (g) accessions were JI157, JI189, JI250 and JI281 (note that JI250 has been called *P. jomardii*). On the graph the lines to the *right* mark the extremes and midpoint of the range of F_{ST} values from five independent *PDR1* data sets of 50 markers. The lines on the *left* mark the two subsets of 38 *Tps12b* markers and the mid point marks the 76 *Tps12b* markers that are polymorphic among these lines, i.e., no markers monomorphic on all the accessions studied have been included in this analysis. **b** Estimates of D. *Pisum* groups a–g are as above

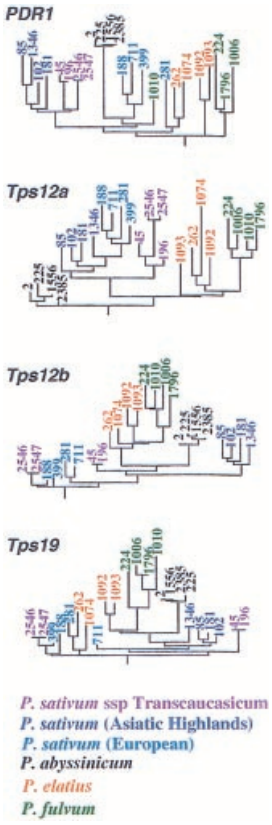


Fig. 6 Phylogenetic trees of *Pisum* accessions based on individual retrotransposon markers. Phylogenetic trees derived for a set of 24 *Pisum* accessions, using four different SSAP-based retrotransposon marker systems as indicated. For *PDR1*, *Tps12a* and *Tps12b*, 50 markers were used, while for *Tps19* only 38 markers were available

contrast, groups all *P. sativum* accessions together and almost resolves each of the three distinct geographical isolates. Each primer shows that the *P. sativum* (transcaucasicum) accessions consist of two distinct groups, one containing JI2546 and JI2547, and the other containing JI45 and JI196. All elements show that there is similarity between some *P. elatius* accessions and *P. fulvum*, although *Tps19* and the *PDR1* SSAP data less reliably group JI262 and JI1074. As well as differences in the branching pattern of the tree, there are also significant differences in branch length; for example, the branch connecting *P. abyssinicum* accessions to the rest of the tree is longer for *PDR1* than for the other markers. In conclusion, the four trees show strong similarities to each other, with pronounced differences which reflect the unique transpositional history of each element, in agreement with the corresponding D and F_{ST} values (Fig. 5).

Discussion

We have isolated a large heterogeneous population of previously uncharacterised Ty1-*copia* retrotransposons from pea (Fig. 1). Each major element group in pea has relatives in the more distantly related *Vicia* species, indicating that heterogeneous populations of these elements have been present throughout the evolution of the *Pisum* and *Vicia* genera from their common ancestor (Fig. 2). Individual pea Ty1-*copia* retrotransposons have widely varying copy numbers, which suggests that these elements have been active to differing extents during pea evolution.

We have investigated the activity of individual pea elements in more detail by isolating three contrasting elements – *Tps12a*, *Tps12b* and *Tps19* – and using their specific terminal sequences to study their insertional

polymorphism within *Pisum* species by SSAP. When these data are compared with those previously obtained for the pea retrotransposon *PDR1* (Ellis et al. 1998) they reveal that each newly isolated element has a high level of insertional polymorphism. The phylogenetic trees deduced from SSAP data for each retrotransposon show broad agreement, indicating that each retrotransposon has been active during *Pisum* species evolution.

The number and position of SSAP bands generated from the LTR of each element is unique in each *Pisum* species lineage. This makes each element potentially useful for determining diversity in these species. D values calculated from these data (Nei's genetic distance; Nei 1973) successfully differentiate the three major species groups in *Pisum*: *P. sativum*, *P. fulvum* plus *P. elatius*, and *P. abyssinicum*. The D values for each individual element agree with this species grouping, but the D values for each element are unique and do not necessarily overlap with similar data from each of the other elements. This shows that each element has behaved differently in each of the *Pisum* species lineages.

The different transpositional history of each element in *Pisum* is confirmed by the calculation of F_{ST} values (an index of the degree of fixation of each element, Fig. 5b) from the same data. *Tps12b* and *Tps19* have been less active in *P. sativum* and therefore have a higher fixation rate than *PDR1* or *Tps12a* (Fig. 5a), whereas in *P. elatius* and *P. fulvum* the four retrotransposons behave quite similarly. In contrast, *Tps12a* has a lower level of fixation in *P. abyssinicum*, indicating it has been more active within this lineage than the other elements. Despite the differences in transpositional activity of each of the element families studied, they all appear to have been active in the relatively recent past (as shown by F_{ST} values lower than 1), but also show low rates of transposition over prolonged evolutionary periods.

Relatively few of the SSAP markers are unique to, and fixed within, the subgroups we have examined. This suggests that introgression of these markers between the subgroups is fairly frequent relative to the rate of transposition. We cannot be certain that band sharing is not a consequence of homoplasy, but this is highly unlikely, as both the priming site and distance to the adjacent restriction site need to be duplicated exactly. Shared band absence is more likely to be caused by homoplasy, for some species lineages band absence is likely to be ancestral, while in others it may be a result of point mutation or deletion at the priming or restriction enzyme sites. The resolution of these issues requires a characterisation of individual element insertions and flanking sequences using the RBIP method (Flavell et al. 1998), which was beyond the scope of the present study.

Pea Ty1-*copia* group retrotransposons as markers for diversity in *Pisum*

All these data show that the behaviour of an individual element within a host lineage is not predetermined, but that each element may proliferate with varying levels of success as each lineage diversifies through evolution. This diversity of activity between different retrotransposon populations in the same host may potentially improve the usefulness of retrotransposons as molecular markers. We have tested this by constructing phylogenetic trees from SSAP data for each element using a representative number of *Pisum* accessions (Fig. 6). As SSAP-based phylogenetic trees are based on retrotransposon activity, each element provides high resolution in lineages in which the element has been active, and relatively low resolution in those lineages where there has been little activity. For example, there is almost no polymorphism associated with *PDR1* and *Tps19* in the four *P. abyssinicum* accessions (Fig. 6) due to the lack of activity of these elements within this species. *Tps12a* and *Tps12b* have been more active in this lineage and so improve the resolution of this area of the tree. Likewise, *Tps12a* has been active throughout *P. sativum* evolution and therefore reveals the subtle relationships within and between the European, Asiatic highland and Transcaucasian *P. sativum* regional variants better than any of the other elements.

A detailed phylogeny of *Pisum* based on a combination of SSAP markers

Retrotransposon-based marker systems provide high-resolution markers for phylogeny studies. We have combined four pea retrotransposon markers with contrasting transpositional histories to provide the most detailed data for *Pisum* diversity to date (Fig. 4). The tree broadly confirms a previous study based on SSAP data for *PDR1* (Ellis et al. 1998; Vershinin and Ellis1999) but there are some notable differences. The use of internal variation as a diversity marker (Vershinin and Ellis 1999) attempts to add "phylogenetic depth" to the resolution for a single element. The consistency observed between that approach and the multi-element analysis is further evidence that we are exploiting events which occurred at different periods. The tree is divided into three main branches, corresponding to *P. sativum*, *P. fulvum* together with *P. elatius*, and *P. abysinicum*. The *P. elatius* accessions consist of two separate clusters, each of which is as different from the other as each is from *P. fulvum*. This analysis reveals several possible mis-classifications. For example, the *P. elatius* accession JI2201 has more similarity to *P. sativum* than other *P. elatius* accessions. Similarly, on the basis of this analysis two of the *P. humile* accessions – JI241 and JI1854 – should be reclassified as *P. sativum*, as should JI250 (*P. jomardii*).

The Tps12 element has distinct LTR sequences

The isolation of two *Tps12* RNase H genes connected to different LTRs within the same genomic library clone is

intriguing (Fig. 3). Both elements are connected to *Tps12* RT domains, as shown by long-range PCR, but the sequence identity which is preserved throughout the RT and RNase H domains of *Tps12a* and *Tps12b* is lost a few bases after the start of the 3′ LTR sequence (Fig. 3). The LTR sequences that follow in the two *Tps12* variants are unrelated, but both are polymorphic and present in similar copy numbers, as judged by the number of SSAP bands (data not shown). The change of sequence within the LTR strongly suggests that one of the *Tps12* LTR sequences has been introduced from another retrotransposon. This is likely to have occurred either by recombination between LTRs of two diverse elements at the RNA or DNA level, or alternatively by transposition of another element into the LTR sequence of the original element. Both insertion and recombination processes are common occurrences amongst retrotransposons (SanMiguel et al. 1996; Jordan and McDonald 1998; Vicient et al. 1999) but the lack of sequence similarity between the *Tps12a* and *Tps12b* LTR sequences makes transpositional insertion a more likely explanation. As the LTR sequences control the element's activity (Casacuberta and Grandbastien 1993) changes in their sequence will affect the evolution of the retrotransposon in the host (Jordan and McDonald 1998). The exchange of LTR sequences between elements may partly explain how the behaviour of elements may vary between different species lineages.

In conclusion, this study confirms that retrotransposon-based markers are highly informative for genetic diversity and phylogenetic studies. It also shows that a multi-retrotransposon approach to phylogeny determination is more informative than one based upon a single retrotransposon, as each element has its own transpositional history which reveals different aspects of the evolution of the species lineage containing it. One of the major hurdles in using the retrotransposon-based SSAP method has been the limited numbers of retrotransposon terminal sequences available. This study, together with the recent development of a more powerful technique for rapid isolation of plant Ty1-*copia* group retrotransposon LTR sequences directly from genomic DNA (Pearce et al. 1999), strengthens the case for applying the multi-retrotransposon-based SSAP approach to phylogenetic analysis in other plant species.

Acknowledgements We thank Ed Hitchin for the pea lambda library. This work was supported by an EC Grant No. 960508 under the Framework IV Biotechnology Program. AK also acknowledges the SOAEFD for financial support.

References

Boeke JD, Corces VG (1989) Transcription and reverse transcription of retrotransposons. Annu Rev Microbiol 43: 403–434

Casacuberta JM, Grandbastien MA (1997) Characterisations of LTR sequences involved in the protoplast specific expression of the tobacco *Tnt1* retrotransposon. Nucleic Acids Res 21: 2087–2093

Doolittle RF, Feng DF (1990) Nearest neighbour procedure for relating progressively aligned amino acid sequences. Methods Enzymol 183: 659–669

Ellis THN (1994) Approaches to genetic mapping of pea. In: Linskens HF, Jackson JF (eds) Vegetables and vegetable products (Modern methods of plant analysis, vol 16). Springer, Berlin Heidelberg New York, pp 117–160

Ellis THN, Poyser MR, Knox MR, Vershinin AV, Ambrose MJ (1998) Polymorphism of insertion sites of Ty1-*copia* class retrotransposons and its use for linkage and diversity analysis in pea. Mol Gen Genet 260: 9–19

Feng DF, Doolittle RF (1990) Progressive alignment and phylogenetic tree construction of protein sequences. Methods Enzymol 183: 375–387

Flavell AJ, Smith DB, Kumar A (1992a) Extreme heterogeneity of Ty1-*copia* group retrotransposons in plants. Mol Gen Genet 231: 233–242

Flavell AJ, Dunbar, E, Anderson R, Pearce SR, Hartley R, Kumar A (1992b) 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

Gonzáles-Candelas F, Palacios C (1997) Analyzing molecular data for studies of genetic diversity. In: Ayad WG, Hodgkin T, Jaradat A, Rao VR (eds) Molecular genetic techniques for plant genetic resources. International Plant Genetic Resources Institute, Rome, pp 55–81

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

Heslop-Harrison JS, Brandes A, Taketa S, Schmidt T, Vershinin AV, Alkhimova EG, Kamm A, Doudrick RL, Schwarzacher T, Katsiotis A, Kubis S, Kumar A, Pearce SR, Flavell AJ, Harrison GE (1997) The chromosomal distributions of Ty1-*copia* group retrotransposable elements in higher plants and their implications for genome evolution. Genetica 100: 197–204

Higgins DG, Sharp PM (1988) CLUSTAL – a package for performing multiple sequence alignment on a microcomputer. Gene 73: 237–244

Jordan IK, McDonald JF (1998) Evidence for the role of recombination in the regulatory evolution of *Saccharomyces cerevisiae* Ty elements. J Mol Evol 47: 14–20

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

Kumar A, Bennetzen JL (1999) Plant retrotransposons. Annu Rev Genet 30: 479–532

Kumar A, Pearce SR, McLean K, Harrison G, Heslop-Harrison JS, Waugh R, Flavell AJ (1997) The Ty1-*copia* group of retrotransposons in plants: genomic organisation, evolution, and use as molecular markers. Genetica 100: 205–217

Kunze R, Saedler H, Lönnig WE (1997) Plant transposable elements. Adv Bot Res 27: 331–470

Laten HM, Majumdar A, Gaucher EA (1998) SIRE-1, a *copia*/Ty1-like retroelement from soybean, encodes a retroviral envelope-like protein. Proc Natl Acad Sci USA 95: 6897–902

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

Matthews P, Ambrose MJ (1994) Development and use of a core collection for the John Innes *Pisum* collection. Proceedings of the Genetic Resources Section Meeting of Eucarpia. Evaluation and exploitation of genetic resources for breeding. Clermont-Ferrand, France, pp 99–107

Nei M (1973) The theory and estimation of genetic distance. In: Morton NE (ed) Genetic structure of populations. University Press of Hawaii, Honolulu pp 45–54

Pearce SR, Harrison G, Li D, Heslop-Harrison JS, Kumar A, Flavell AJ (1996) The Ty1-*copia* group retrotransposons in

Vicia species: copy number, sequence heterogeneity and chromosomal localisation. Mol Gen Genet 250: 305–315

Pearce SR, Harrison G, Heslop-Harrison JS, Flavell AJ, Kumar A (1997) Characterisation and genomic organization of Ty1-*copia* group retrotransposons in rye (*Secale cereale*). Genome 40: 617–625

Pearce SR, Stuart-Rogers C, Kumar A, Flavell AJ (1999) Rapid isolation of plant Ty1-*copia* group retrotransposons LTR sequences for molecular marker studies. Plant J 19: 1–7

Purugganan MD, Wessler SR (1995) Transposon signatures-species-specific molecular markers that utilize a class of multiple-copy nuclear-DNA. Mol Ecol 4: 265–269

Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci USA 81: 8014–8018

Saitou N, Nei M (1987) The neighbor joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406–425

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual (2nd edn). Cold Spring Harbor Laboratory press, Cold Spring Harbor, N.Y.

SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, MelakeBerhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, Bennetzen JL (1996) Nested retrotransposons in the intergenic regions of the maize genome. Science 274: 765–768

Schmidt T (1999) LINE, SINEs and repetitive DNA: non-LTR retrotransposons in plant genomes. Plant Mol Biol 40: 903–910

Schmidt T, Kubis S, Heslop-Harrison JS (1995) Analysis and chromosomal localization of retrotransposons in sugar-beet (*Beta vulgaris*) – LINES and Ty1-*copia*-like elements as major components of the genome. Chromosome Res 3: 335–345

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

Vicient CM, Suoniemi A, Anamthawat-Jonsson 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

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