

Transposable Elements Reveal the Impact of Introgression, Rather than Transposition, in *Pisum* Diversity, Evolution, and Domestication

Alexander V. Vershinin,¹ Theodore R. Allnutt, Maggie R. Knox, Mike J. Ambrose, and T. H. Noel Ellis

Department of Crop Genetics, John Innes Centre, Norwich, United Kingdom

The genetic structure and evolutionary history of the genus *Pisum* were studied exploiting our germplasm collection to compare the contribution of different mechanisms to the generation of diversity. We used sequence-specific amplification polymorphism (SSAP) markers to assess insertion site polymorphism generated by a representative of each of the two major groups of LTR-containing retrotransposons, *PDR1* (Ty1/*copia*-like) and *Cyclops* (Ty3/*gypsy*-like), together with *Pis1*, a member of the *En/Spm* transposon superfamily. The analysis of extended sets of the four main *Pisum* species, *P. fulvum*, *P. elatius*, *P. abyssinicum*, and *P. sativum*, together with the reference set, revealed a distinct pattern of the NJ (Neighbor-Joining) tree for each basic lineage, which reflects the different evolutionary history of each species. The SSAP markers showed that *Pisum* is exceptionally polymorphic for an inbreeding species. The patterns of phylogenetic relationships deduced from different transposable elements were in general agreement. The retrotransposon-derived markers gave a clearer separation of the main lineages than the *Pis1* markers and were able to distinguish the truly wild form of *P. elatius* from the antecedents of *P. sativum*. There were more species-specific and unique *PDR1* markers than *Pis1* markers in *P. fulvum* and *P. elatius*, pointing to *PDR1* activity during speciation and diversification, but the proportion of these markers is low. The overall genetic diversity of *Pisum* and the extreme polymorphism in all species, except *P. abyssinicum*, indicate a high contribution of recombination between multiple ancestral lineages compared to transposition within lineages. The two independently domesticated pea species, *P. abyssinicum* and *P. sativum*, arose in contrasting ways from the common processes of hybridization, introgression, and selection without associated transpositional activity.

Introduction

The molecular mechanisms that generate genetic diversity are bewilderingly complex, making it difficult to distinguish between evolutionary hypotheses. The advent of DNA markers and whole, or extensive, genome sequencing have facilitated genome comparison; however, these approaches are very expensive and time consuming. The use of genetic markers is an alternative and less costly surveying method.

The differences between closely related genomes are often associated with repetitive DNAs (Britten and Kohne 1968; Walker 1968). Mobile genetic elements (or transposable elements, TEs) are a major component of plant genomes where they may comprise more than 50% of the nuclear DNA (Kunze, Saedler, and Lonning 1997; SanMiguel and Bennetzen 1998; Bennetzen 2000a) and contribute to diversity through both insertion site polymorphism and small structural rearrangement (Bennetzen 2000a). Transposable elements are classified into two groups according to their transposition mechanism and mode of propagation (Finnegan 1992): retrotransposons (class I elements) transpose via an RNA intermediate, whereas transposons (class II elements) move by excision and reintegration (“cut-paste”). The ubiquity and distribution of these TEs suggest that they should be potentially useful as diagnostic tools conforming to the requirement for abundant and reliable markers.

The relationship between transposable elements and chromosomal rearrangement in plants was first shown by McClintock (1946); deletions, insertions, frameshifts, inversions, duplications, translocations, and the generation of intron-like sequences have all been associated with TEs (Kunze, Saedler, and Lonning 1997). Because of their ability to move from place to place within a genome, or to produce new copies of themselves at any genomic location, TEs possess an extraordinary potential for the alteration of genome structure and adjacent gene function (Nekrutenko and Li 2001). These characteristics have promoted the idea that TEs are important players in plant evolution (Wessler, Bureau, and White 1995).

The genus *Pisum* L. has characteristics that make it particularly interesting for the application of TE-based markers to the analysis of intragenus genetic diversity. Pea (*Pisum sativum* L.) is an Old World legume crop thought to have been among the first cultivated in the Middle East, about 10,000 years ago (Blixt 1972; Zohary 1996). The modern gene pool of cultivated *Pisum* is diverse, reflecting this early domestication and subsequent widespread cultivation. However, in spite of the extensive phenotypic and genetic variability, existing taxonomic classifications distinguish few *Pisum* species, from two (Ben-Ze’ev and Zohary 1973) to six (Govorov 1937).

Pisum has a large genome (about 4,000 Mb), which is stable in size between species (Greilhuber and Ebert 1994). Pea is one of a few plant species in which the representatives of different TE groups have been isolated and characterized, and as an inbreeding species, it contrasts with maize that has extensively characterized retroelement families (SanMiguel et al. 1996). We have selected three elements from different TE groups for this study: *PDR1* is a Ty1-*copia* group retrotransposon present in about 200 copies per haploid genome; it is about 4 kb in length, and its LTRs (long terminal repeats), at 156 bp (Lee et al.

¹ Present address: Institute of Cytology and Genetics, Novosibirsk, Russia.

Key words: transposable elements, *Pisum* evolution, molecular markers, recombination, domestication, retrotransposon activity.

E-mail: alexander.vershinin@bbsrc.ac.uk.

Mol. Biol. Evol. 20(12):2067–2075. 2003

DOI: 10.1093/molbev/msg220

Molecular Biology and Evolution, Vol. 20, No. 12,

© Society for Molecular Biology and Evolution 2003; all rights reserved.

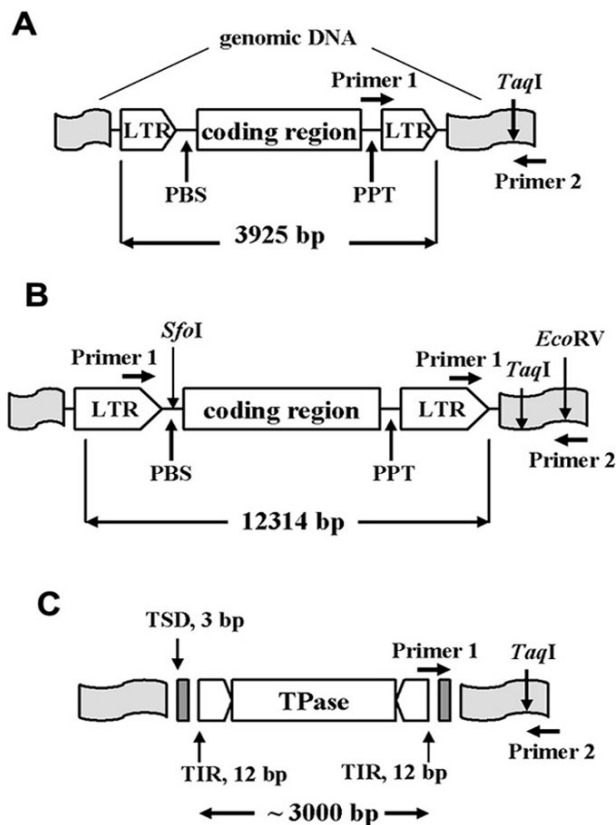


FIG. 1.—Maps of transposable elements. Maps show the positions of primers (1 and 2) used in SSAP analysis and restriction sites. A. *PDR1*; B. *Cyclops*; C. *Pis1*. LTR: long terminal repeats; TIR: terminal inverted repeats; TSD: target site duplication; TPase: transposase. The figure is not drawn to scale.

1990), are exceptionally short. In contrast, *Cyclops* has the typical *pol* region of the Ty3/*gypsy* group and is present in about 5,000 copies (Chavanne et al. 1998). *Cyclops* elements are approximately 12 kb long, including very long LTRs of about 1,500 bp. The reading frame of the *pol* region of the canonical *Cyclops* element is disrupted by several mutations, suggesting that it is nonfunctional. The third element we have studied was *Pis1*, a representative of pea class II mobile elements (Shirsat 1988), which has the copy number comparable to that of *Cyclops* as determined in pilot sequence-specific amplification polymorphism (SSAP) experiments. Its terminal inverted repeats contain the sequence (5'-CACTA-3') of the *En/Spm* superfamily, and it is abundant in the genome.

DNA markers have been powerful in phylogenetic diversity analysis and several methods have been shown to be efficient for comparison of different plant genotypes (Powell et al. 1996; Russell et al. 1997; Pejic et al. 1998), including pea (Lu et al. 1996). Amplified fragment length polymorphism (AFLP) and SSAP have a high multiplex ratio (Lu et al. 1996; Powell et al. 1996; Waugh et al. 1997; Ellis et al. 1998) and offer a distinct advantage when genome coverage is a major issue. For an extensively inbreeding system, such as pea, dominance of the markers is not such an important consideration and it has been shown that SSAP markers produced by TEs are more informative

than AFLP and restriction fragment length polymorphism (RFLP) markers (Ellis et al. 1998). This SSAP approach reveals insertion site polymorphism and sequence variation in the flanking DNA or primer binding site.

We have applied the SSAP approach to study the genetic structure and evolutionary history of *Pisum*. Our analysis is based on a wide range of *Pisum* accessions selected from John Innes Centre germplasm collection, which contains more than 3,000 accessions. We aimed to: (1) characterize the phylogenetic diversity of *Pisum* and to compare the patterns revealed by TEs with large differences in abundance and contrasting transposition mechanisms; (2) examine the relative contribution of different molecular mechanisms to genetic diversity within *Pisum*; and (3) compare the genetic diversity associated with the domestication of two domesticated species, *P. sativum* and *P. abyssinicum*, Braun.

Materials and Methods

Plant Material

Accessions from the John Innes *Pisum* Germplasm are designated JIx, where *x* is a number (<http://www.jic.bbsrc.ac.uk/germplas/pisum/Zgc4b.htm>). The analysis of this collection was carried out in three steps, as described in *Results*. DNAs were extracted from leaves with the DNeasy 96 Plant Kit (Qiagen).

Template Preparation, Polymerase Chain Reaction Condition, Adapters, and Primers

The structure of TEs with the position of restriction sites and primers used in this study is shown in figure 1. The short *PDR1* long terminal repeats (LTRs) allowed the use of the polypurine tract (PPT) primer and digestion of the genomic DNA (about 0.5 μ g) by one restriction enzyme, *TaqI*. For *Cyclops* SSAP, the primer was designed to the very 3'-end of the LTR, and DNA was cut by two restriction enzymes, *EcoRV* and *SfoI*, the latter to avoid the production of internal *Cyclops* fragments from the 5' LTR.

Other procedures for *PDR1* SSAP (ligation, PCR conditions and gel analysis of PCR products) were as described previously (Ellis et al. 1998). For the SSAP analysis of the high copy number *Cyclops* element, the template DNA after ligation with the *EcoRV* adapter was preamplified in 25 μ l using primers homologous to the adapter sequence (left primer: 5'-CCACAGCGA-TACCTTGATC-3'; right primer: 5'-CGAGTAAGTCCT-GATCGAT-3'). The 25 μ l polymerase chain reaction (PCR) (PCT-200, MJ Research Inc.) comprised 20 cycles of 94° (30 s), 56° (60 s), 72° (60 s). Then, 55 μ l T0.1E (10 mM Tris-HCl, pH 8.0, 0.1 mM EDTA) was added, and 3 μ l of this mixture was used in a second PCR with ³²P-end-labeled *Cyclops* primer (primer 1 in figure 1A; 5'-CACAATCCCTGTGGAGACG-3') and the "right" *EcoRV* primer (sequence see above) with a different three-nucleotide extension (primer 2 in figure 1B). The PCR temperature cycling with ³²P *Cyclops* primer was as described for *BARE-1* (Waugh et al. 1997). The *EcoRV* adapter primer had a different-three nucleotide extension. In addition, SSAP analysis of transposon *Pis1* (Shirsat

1988) used the primers 1 and 2 shown in figure 1C and was performed as previously described (Knox and Ellis 2002). Gel analysis of the PCR products was the same as for *PDR1* (Ellis et al. 1998). Replicated digestion-ligations from the same genomic DNAs and repeated amplification from the same digestion-ligation reaction resulted in consistent banding patterns.

Data Analysis

A table of presence/absence for each band in an SSAP profile was generated. Bands with the same migration were considered to be identical (Casa et al. 2002), whereas others were treated as independent insertions. The total number of pairwise mismatches was used to calculate pairwise distances from the SSAP band scores. The resulting matrix was used as input for principal component analysis (PCA) using the GenStat 5 package and analysis of molecular variance (AMOVA) (Excoffier, Smouse, and Quattro 1992). Phylogenetic trees were generated from this matrix using the Neighbor-Joining (NJ) algorithm (Saitou and Nei 1987) of the Neighbor program in the PHYLIP package (Felsenstein 1993). Neighbor-Joining dendrograms were also constructed from pairwise Φ_{ST} values between accession groups generated by AMOVA. Trees were plotted using Treeview. The genetic variation was measured in terms of gene diversity (Nei 1987) as H_{mean} .

Results

PDR1 Markers Showed a Different Evolutionary History of Each *Pisum* Species

We generated markers corresponding to genomic DNA neighboring the 3'-end of different classes of TEs (fig. 1). The marker data were analyzed in three steps as follows. Step 1 was based on the results of a previous survey of a sample of 56 accessions, selected mostly on geographical and morphological criteria (Ellis et al. 1998). From these earlier data a small subset of accessions that carried most (94%) of the markers was identified, and these lines were used in further analysis as a reference set. In step 2, extended sets of each of the four main *Pisum* species, *P. fulvum*, Sibth. et Sm., *P. elatius*, (M.B.) Stev., *P. abyssinicum*, and *P. sativum* were analyzed separately but including the reference set. This analysis comprised 43 accessions of *P. fulvum*, 42 accessions of *P. elatius*, 40 accessions of *P. sativum*, and 32 accessions of *P. abyssinicum*. From these analyses, a "final set" of 52 accessions that included representatives of the main clades of the NJ trees was selected for a combined analysis in step 3. In this final set there were 10 accessions of *P. fulvum*, 12 of *P. elatius*, 5 of *P. abyssinicum*, 2 of *P. humile*, Boiss. et Noe, and 23 of *P. sativum*.

The *PDR1* family has a copy number well suited to SSAP analysis and genome mapping because no preamplification is required in the PCR and we can examine a high proportion of the total number of insertion sites. Therefore, we selected this family for the analysis of extended sets of main *Pisum* lineages (step 2) and produced NJ trees for each species, together with corresponding reference accessions (fig. 2). In the NJ tree

of the *P. fulvum* extended set (fig. 2A), all *P. fulvum* accessions are quite well separated from other groups and are characterized by long branches, suggesting remote common ancestry. Within *P. fulvum* group one can distinguish three sub-groups with shorter distances between branch points. A very different pattern of branching was found in the extended set of *P. abyssinicum* (fig. 2B). Of the 32 accessions, 30 are placed on one major branch with extremely short distances between accessions, but two accessions are slightly separated from all others. While on the basis of morphology JI1937 is typical for *P. abyssinicum*, JI2674 carries some *P. sativum* morphological characteristics and probably represents *P. abyssinicum* introgressed with *P. sativum*.

According to taxonomic classifications based on the morphology, ecology, cytogenetics, and species distribution, *P. fulvum* represents the most ancient lineage divergence (Govorov 1937; Blixt 1972; Ben-Ze'ev and Zohary 1973; Zohary 1996). Although it is well separated, another ancient lineage, *P. elatius*, showed a number of internal branches widely spread and interspersed with the branches of *P. humile* and *P. abyssinicum*, as well as with some accessions of *P. sativum* (fig. 2C). After *P. abyssinicum*, the *P. sativum* lineage (fig. 2D) has a highest number of short branches. Again, some *P. elatius* accessions are intermixed with the *P. sativum* accessions, indicating close phylogenetic relationships between these lineages.

Analysis of the extended sets of different *Pisum* species revealed a high number of new markers for each species (table 1) that were absent in initial reference accessions of given species. However, for *P. fulvum* and particularly *P. elatius*, a significant proportion of these new markers were absent from the initial reference set while for *P. abyssinicum* and *P. sativum* there were a very few markers of this type.

Genetic diversity estimates for the extended sets are shown in table 2. These values are high (e.g., see Ghebru, Schmidt, and Bennetzen 2002) for all lineages, except *P. abyssinicum*, pointing to a significant level of heterozygosity at some time within pea species, despite their self-fertility.

Taxonomy of *Pisum* as Assessed by TE Markers

All TEs studied revealed a high degree of insertion site polymorphism with few monomorphic bands for each element. *PDR1* (Ellis et al. 1998; Knox and Ellis 2002), *Cyclops* (data not shown) and *Pis1* element (Knox and Ellis 2002) derived markers are distributed throughout the pea genome. We have followed nearly 200 insertion sites of each marker type, but, given the relative copy numbers, we analyzed a smaller fraction of *Cyclops* and *Pis1* insertion sites.

Principal component analysis (PCA) of these data, obtained with the final set of accessions (step 3), identified three major groups, *P. fulvum*, *P. abyssinicum*, and a *P. elatius*-*P. humile*-*P. sativum* complex for both *PDR1* and *Cyclops* (fig. 3A and 3B). The percentages of the total variation, explained by the first two axes for *Cyclops* and *PDR1*, were 64% and 67%, respectively. Within the *P. elatius*-*P. humile*-*P. sativum* complex, two subclusters

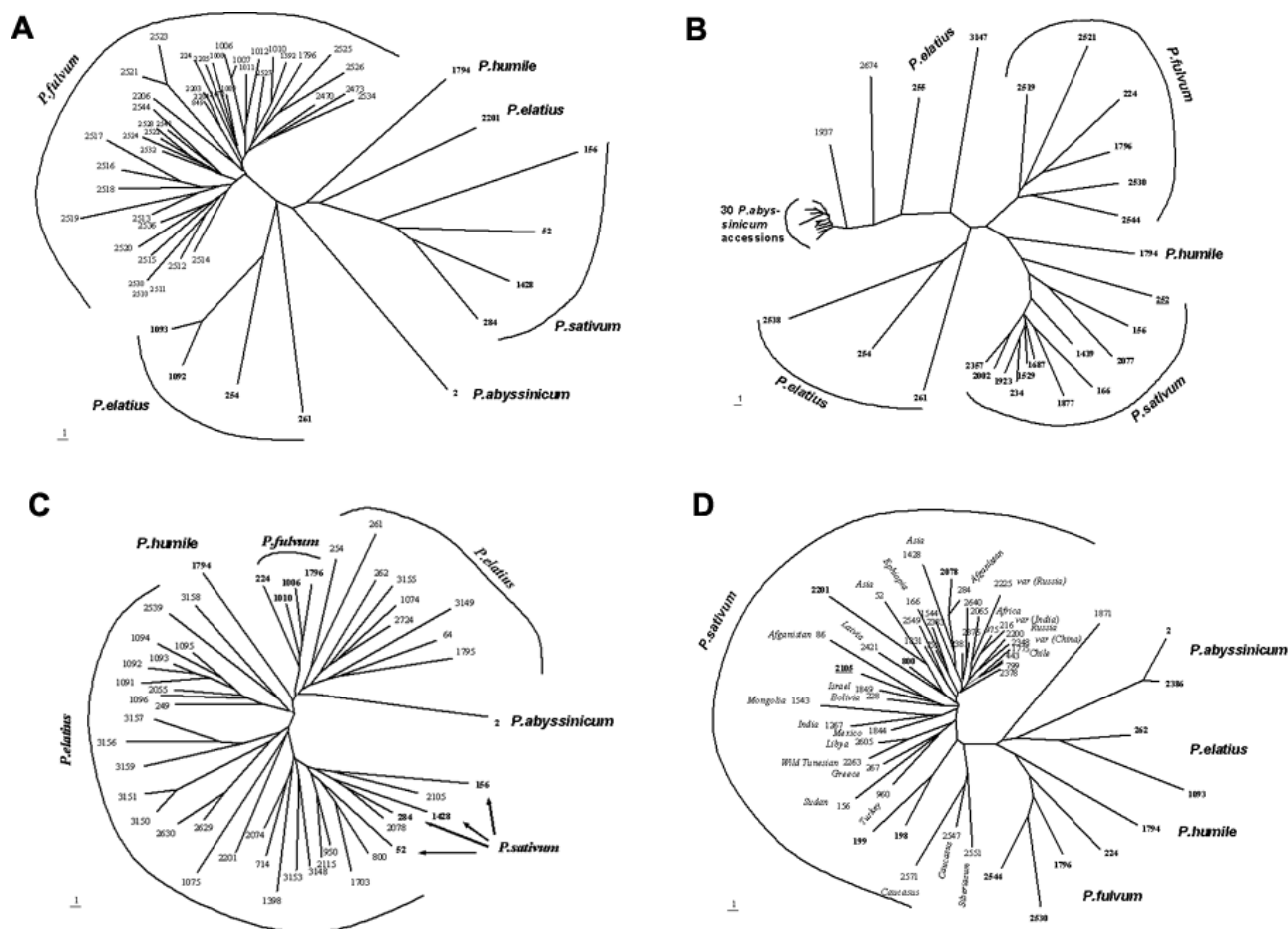


FIG. 2.—NJ *Pisum* trees for *PDR1* SSAP markers derived from the extended sets of species accessions. A. *P. fulvum*; B. *P. abyssinicum*; C. *P. elatius*; D. *P. sativum*.

of accessions were evident. One subcluster (group 1) comprised eight *P. elatius* and one *P. humile* accession, and the other (group 2) consisted of four *P. elatius* and all *P. sativum* accessions. For the *Pis1* data (fig. 3C) the *P. fulvum* accessions were closer to the accessions of *P. elatius* complex, but *P. abyssinicum* was as distinct as for the retroelement data. In this data set, the first two axes made up 56.5% of the total variation, with nearly 45% in PCA1.

The markers derived from both retrotransposons showed very similar patterns of interspecies relationships, with clear separation of the main lineages. The *Pis1*-derived markers gave similar results, but the discrimina-

tion was not as clear. For a comprehensive picture, and to visualize the relationships among the different accessions, we combined all the retrotransposon data and constructed an NJ tree (fig. 4). The main characteristics of the tree reiterated the conclusions from the analyses of the four groups independently and are these: (1) clear separation of the *P. fulvum* lineage with long internal branches; (2) extremely low diversity within *P. abyssinicum*; (3) a broad distribution of the main *P. elatius* branches, several of which (group 2) are intermingled with *P. humile* and *P. sativum* accessions; (4) independent grouping of *P. sativum* accessions of African and Asiatic origin.

An AMOVA was conducted to test the significance of differences within and between species. The AMOVA

Table 1
The Distribution of Novel *PDR1* Markers for One Primer Combination

Species	Number of Markers		
	Total	Absent in Initial Reference Accession of Given Species	Totally Absent in Reference Set
<i>P. fulvum</i>	78	43	12
<i>P. elatius</i>	129	70	26
<i>P. abyssinicum</i>	45	12	1
<i>P. sativum</i>	77	29	2

Table 2
Estimates of Genetic Diversity (H_{mean}) for *Pisum* Species Assessed Using *PDR1*

Species	Number of Accessions	H_{mean}	Conf.
<i>P. fulvum</i>	43	0.393	0.063
<i>P. elatius</i>	42	0.597	0.044
<i>P. abyssinicum</i>	29	0.156	0.042
<i>P. sativum</i>	40	0.513	0.068

NOTE.—Conf. = 95% confidence limits.

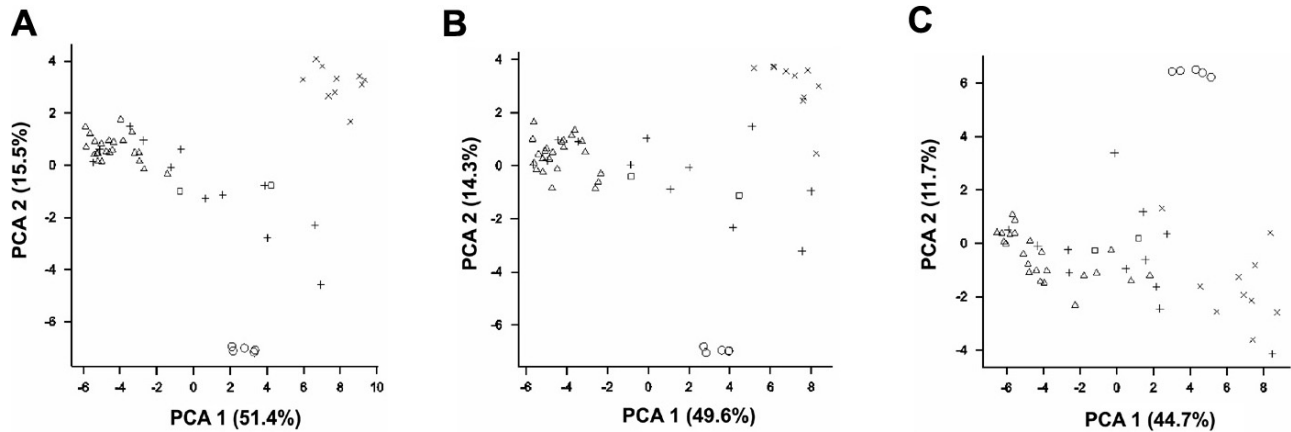


FIG. 3.—Patterns of relationship of 52 *Pisum* accessions of the “final set” revealed by the principal component analysis (PCA) of SSAP marker data derived from three different TEs. Proportion of the total variation explained by the first two axes (PCA1 and PCA2) is shown. × = *P. fulvum*; ○ = *P. abyssinicum*; + = *P. elatius*; □ = *P. humile*; △ = *P. sativum*. A. *PDR1*; B. *Cyclops*; C. *PisI*.

program generates Φ statistics (Excoffier, Smouse, and Quattro 1992) analogous to the F statistics in population genetics (Wright 1951). Pairwise Φ_{ST} values derived from AMOVA reflect a number of significant ($P < 0.05$) differences between accession groups when individual pairs of accession groups were compared. Initially, we divided the final set of 52 accessions into five groups corresponding to the main pea species. In this analysis Φ_{ST} between *P. elatius* and *P. humile* was not significant ($P > 0.05$) and AMOVA failed to recognize them as separate groups, so they were united in further treatments. The separation into four groups was statistically significant for all three TEs studied ($P < 0.001$ at 1,000 replicate bootstrap), with the variance among accession groups being 32% for both retrotransposons and 24% for *PisI*.

Phylogenetic trees constructed with the NJ algorithm, from pairwise Φ_{ST} values, were similar for the three TE data sets (fig. 5). However, the *P. fulvum* branch was noticeably shorter in the tree generated with *PisI*-derived markers than in the one generated with the retrotransposon data, suggesting a higher proportion of *P. fulvum* *PisI*-markers shared with other species. The *P. elatius*–*P. sativum* branch has the same length in all trees; however, the shape of the branch in the *PisI* tree is different when compared to the retrotransposon trees; for these, the position of *P. elatius* is closer to the point of divergence from *P. fulvum* and *P. abyssinicum*. These differences in the *PisI* tree are consistent with its lower value for the variance among accession groups and are likely reflect differences in the evolutionary history of these elements and their relative contribution in genetic diversity.

TE Markers Are a Tool for Study of the Evolutionary Dynamics of Genome Microstructure

Sequence-specific amplification polymorphism can be used to measure fluctuations in copy number between accessions and the appearance of unique insertions. The variation in insertion site copy number that we observed between accessions was similar for all of the TEs, and no accession was remarkable. We identified unique bands

(that are present in only one accession) and species-specific bands that are present in accessions of only one species, regardless of their frequency. All species, except *P. abyssinicum*, showed an extremely high percentage of polymorphic bands (table 3). However, although the percentage of species-specific (including unique) markers makes up a noticeable part of the overall polymorphism of *P. fulvum* and *P. elatius*, within *P. sativum* the proportion of such markers is very low and similar to that of *P. abyssinicum*. Of the *PDR1* species-specific markers in *P. abyssinicum*, none were unique (table 3). Most of the *P. elatius* species-specific markers and all the unique markers were found in the group 1 accessions. *PDR1* showed more than a twofold higher percentage of species-specific markers than *PisI* (table 3), suggesting that retrotransposition might have occurred during speciation of *P. fulvum* and *P. elatius*. The low proportion of species-specific and

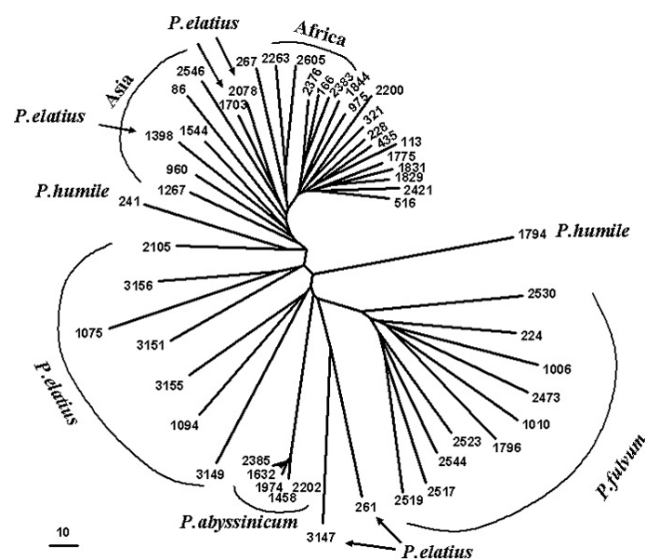


FIG. 4.—Combined NJ *Pisum* tree for *PDR1* and *Cyclops*-derived SSAP markers. Numbers refer to JI *Pisum* germplasm accession numbers. Main groups are outlined.

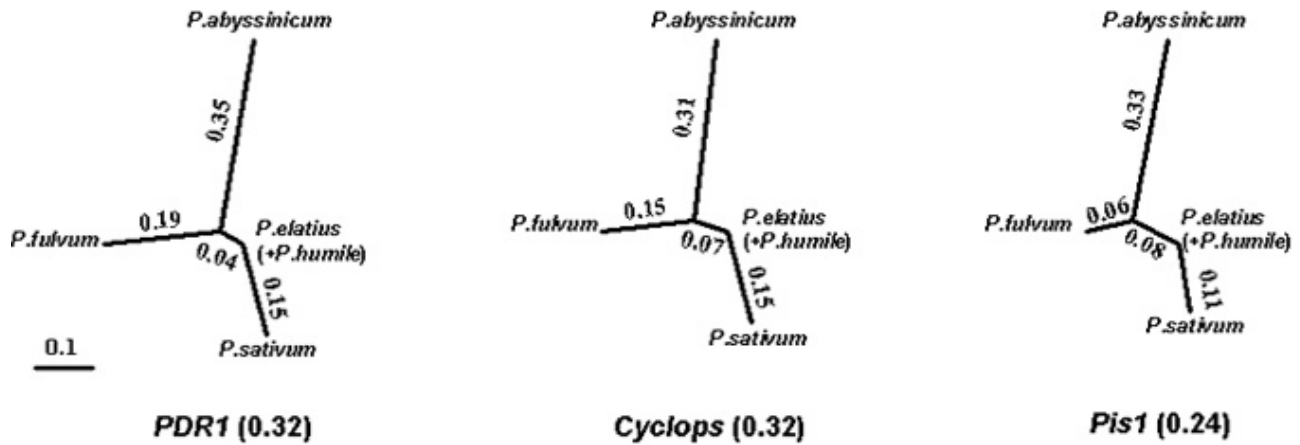


FIG. 5.—NJ dendrograms for different TEs constructed from pairwise Φ_{ST} values between accession groups generated by AMOVA. Variance among accession groups for each TE is shown in brackets. Figures on branches indicate NJ branch lengths calculated from pairwise Φ_{ST} . All pairwise Φ_{ST} values were significant ($P < 0.001$). Branch length for *P. elatius* is effectively zero.

unique *Cyclops* markers may indicate that this element has been passive throughout most of *Pisum* evolution, but we cannot be sure that the high number of both *Cyclops* and *Pis1* contributes to this effect.

The accessions studied have a very similar *PDR1* copy number, well suited to SSAP analysis, and the insertion sites are distributed throughout the pea genome (Ellis et al. 1998; Knox and Ellis 2002), suggesting that these are comprehensive diagnostic markers. We scored the number of markers shared by two, three, and all species. *P. fulvum* has more markers in common with *P. abyssinicum* than *P. elatius*, whereas the latter has the highest proportion of markers in common with *P. sativum* (fig. 6A). Furthermore, *P. abyssinicum* and *P. sativum* do not have shared markers that are absent from the other lineages, supporting the idea that these lineages are of recent origin and were brought into cultivation independently. The general pattern of marker sharing is illustrated in figure 6B, where the overlapping circles depict the proportions of the corresponding gene pool of *Pisum*. The different segments reflect the proportions of species-specific and shared markers. The proportion of markers in common to all *Pisum* is 14%, but these are not fixed markers, so the remaining 86% of the *Pisum* genome underestimates *Pisum* diversity. These results strongly suggest that *Pisum* is a species complex.

Discussion

Retrotransposon Contribution to *Pisum* Genetic Diversity

Transposable elements may be involved in genome rearrangement and produce genetic diversity by transposition itself or by other activities they themselves promote. By comparing the behavior of TEs with different mechanism of propagation, we intended to gauge the relative importance of transposition, especially with respect to recombination and segregation, but recognizing the potential for other types of rearrangement (McClintock 1946; Shalev and Levy 1997; Bennetzen 2000a).

At the level of SSAP analysis we cannot distinguish between different mechanisms that rearrange TE flanking DNA. Transposon (*Pis1* in our analysis) propagates by

a cut-and-paste mechanism; this and other rearrangements of flanking DNA would result in the replacement of one SSAP band by another. Retrotransposition based on amplification and the insertion of new copies will produce new SSAP markers in addition to preexisting ones. If the DNA that flanking different TEs is affected by rearrangement to a similar extent, and if each transposition event is unique, then the relative proportion of unique and species-specific markers for each element will reflect the relative contribution of that element's transposition to the observed diversity. The longer *P. fulvum* branches in the NJ tree constructed from the AMOVA analysis of retrotransposon markers, as compared to transposon markers (fig. 5), results from a higher value of the variance among the accession groups. This implies that there is a higher proportion of retrotransposon markers (rather than transposon markers) that are restricted to *P. fulvum*. This tendency is well supported by the data of table 3, where *P. fulvum*, *P. elatius*, and *P. sativum* shared a similar high level of polymorphism, but in *P. sativum* alone the proportion of species-specific and unique bands was close to zero. The *PDR1* markers were the most variable in *P. fulvum* and *P. elatius*, more than twofold higher than the *Pis1* markers, a finding consistent with retrotransposition in these two lineages.

Despite the high level of polymorphism, element copy number is remarkably constant. The evolutionary history of the reverse transcriptase gene in 11 plant species, including pea, has indicated strong patterns of purifying selection (Navarro-Quezada and Schoen 2002) with high rates of element loss balanced by transposition. Our data suggest that segregation and drift could be effective mechanisms accounting for the loss of insertion sites.

Recombination Is a Dominant Feature of *Pisum* Evolution and Domestication

Virtually all SSAP bands are polymorphic (table 3), with unique and species-specific markers making up only a small proportion of them. This situation is consistent with the possibility that introgression, segregation, and small rearrangement, rather than transposition itself, are

Table 3
Characteristics of Markers Derived from Different TEs

Species	Transposable Element	Total Number of Markers per Species	Number of Polymorphic Markers Within Species	Number of Species-Specific Markers	
				Total	Unique
<i>P. fulvum</i>	<i>PDR1</i> (259)*	167 (64.5%)	160 (95.8%)	35 (21.0%)	17 (10.2%)
	<i>Cyclops</i> (343)*	275 (80.2%)	266 (96.7%)	20 (7.3%)	3 (1.1%)
	<i>Pisl</i> (215)*	159 (73.9%)	157 (98.7%)	16 (10.1%)	6 (3.8%)
<i>P. elatius</i>	<i>PDR1</i>	189 (73.0%)	183 (96.8%)	25 (13.2%)	16 (8.5%)
	<i>Cyclops</i>	312 (91.0%)	304 (97.4%)	7 (2.2%)	2 (0.6%)
	<i>Pisl</i>	158 (73.5%)	158 (100%)	12 (7.6%)	6 (3.8%)
<i>P. abyssinicum</i>	<i>PDR1</i>	75 (28.9%)	16 (21.3%)	2 (2.7%)	0
	<i>Cyclops</i>	122 (35.5%)	5 (4.1%)	0	0
	<i>Pisl</i>	35 (16.3%)	5 (14.3%)	0	0
<i>P. sativum</i>	<i>PDR1</i>	159 (61.4%)	149 (93.7%)	4 (2.5%)	3 (1.9%)
	<i>Cyclops</i>	259 (75.5%)	240 (92.7%)	0	0
	<i>Pisl</i>	133 (61.9%)	131 (98.5%)	5 (3.7%)	2 (1.5%)

NOTE.—Asterisk indicates total number of markers identified in all species.

the dominant modes of diversity generation, even for the most ancient *Pisum* lineages, *P. fulvum* and *P. elatius*. Comparative sequence analysis has revealed a much higher degree of diversity at the microstructural level than was predicted by genetic mapping studies of closely related plant species (Bennetzen 2000b; Bennetzen and Ramakrishna 2002). Sequencing of the *bz* genomic region of two maize lines revealed dramatic differences between them in both retrotransposon clusters and genes, demonstrating that genetic microcolinearity can be violated even within the same species (Fu and Dooner 2002).

The taxonomy of *Pisum* has been much disputed (Govorov 1937; Blixt 1972; Ben-Ze'ev and Zohary 1973; Zohary 1996). Conventionally one cultivated species *P. sativum* and three wild taxa, *P. elatius*, *P. humile*, and *P.*

fulvum, are recognized. Based on the analysis of morphology, ecology, cytogenetics, and hybrid performance, Ben Ze'ev and Zohary (1973) concluded that *P. fulvum* is a fully divergent species, whereas *P. humile*, *P. elatius*, and *P. sativum* form a single-species complex comprised of two main races, weedy forms (*elatius* and *humile*), and cultivated derivatives (*sativum*); *P. fulvum* together with *P. elatius* were recognized as the ancient lineages (Govorov 1937; Blixt 1972; Ben-Ze'ev and Zohary 1973; Zohary 1996).

Phylogenetic relationships revealed by SSAP analysis are generally congruent with those derived from traditional taxonomic studies. The NJ trees constructed for extended sets of each of the major *Pisum* lineages illustrate very well the main characteristics of *Pisum* phylogeny summarized in the combined NJ tree (fig. 4). The *P. fulvum*

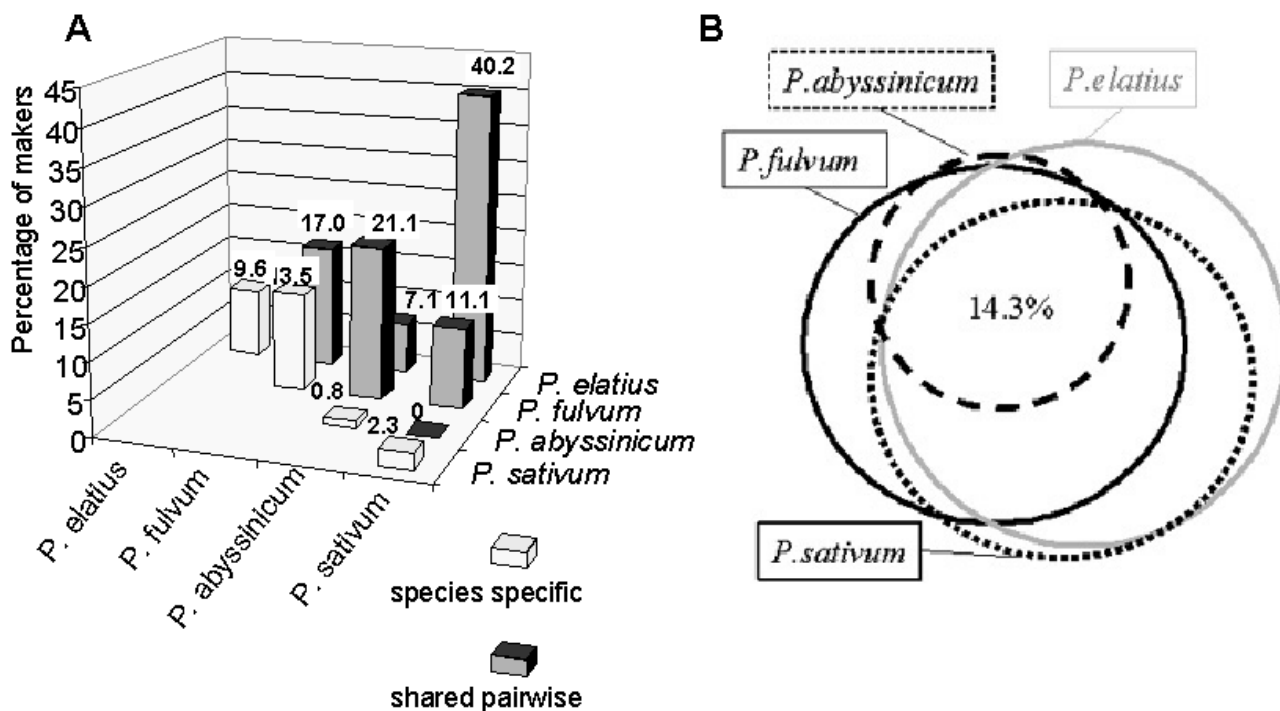


FIG. 6.—Evolutionary dynamics of different *Pisum* species demonstrated by the proportions of *PDR1* SSAP markers. A, markers shared exclusively by species pairs. Diagonal columns show the proportions of species-specific markers. B, circles illustrating the proportions of markers occurring in different species combinations.

lineage with long internal branches forms a distinct group; *P. abyssinicum* also forms a distinct group, but with extremely short internal branches and, correspondingly low within-lineage diversity. The *P. elatius* accessions are widely distributed, and several of them are intermingled with *P. humile* and *P. sativum* accessions. Some authors considered *P. fulvum* as species having insignificant intraspecific differentiation of morphological features (Govorov 1937). In contrast, the distribution pattern of our molecular markers produced by different TEs among *P. fulvum* accessions did not show any monomorphic bands and H_{mean} of *P. fulvum* is closer to the *P. elatius* and *P. sativum* values than to the *P. abyssinicum* value (table 2). A similar disparity between phenotypic and molecular estimation of diversity has been reported for other species, such as maize (Burstin and Charcosset 1997; Smith et al. 1997) and tomato (Noli, Salvi, and Tuberosa 1997).

The close relationships among *P. elatius*, *P. humile*, and *P. sativum* have been corroborated by PCA analysis (fig. 3), where *Pis1*-derived markers placed these species in one broad cluster. *PDR1* and *Cyclops* markers subdivided this cluster into two groups. Group 1 comprises eight accessions of *P. elatius* plus both *P. humile* accessions, whereas group 2 consists of four *P. elatius* accessions and all accessions of *P. sativum*. Analysis of the *P. elatius* accessions showed that the two groups carried contrasting alleles associated with domestication at the following loci: *Np* (neoplastic pods), *Gty* (gritty testa), and *Dp* (dehiscent pod). Thus *P. elatius* group 1 can be considered the wild forms of *P. elatius* (and *P. humile*), whereas *P. elatius* accessions from group 2 likely represent the section of *P. elatius* from which the antecedents of cultivated *P. sativum* were drawn, or alternatively that have had recent introgression with *P. sativum*.

One of the most discussed problems in *Pisum* taxonomy is the position and status of *P. abyssinicum*, namely whether this lineage has diverged far enough from other taxa to be considered a separate species or whether it should be placed within *P. sativum* as a subgroup or ecotype. Govorov (1937) found slightly differentiated morphological features among different *P. abyssinicum* accessions, as for *P. fulvum*, and judged that these were sufficiently distinct to consider this lineage a separate cultivated species. However, more often, *P. abyssinicum* has been regarded as an ecotype or subgroup of *P. sativum* (Makasheva 1984, p.40). All our data demonstrate the clear distinction of *P. abyssinicum* from all other lineages. The transposon and retrotransposon markers showed the same level of resolution in PCA (fig. 2), and, for both, the *P. abyssinicum* branch was the longest in the NJ trees generated by AMOVA (fig. 4). *P. abyssinicum* accessions are very similar, so for the final set of 52 accessions, we selected fewer *P. abyssinicum* accessions than accessions from other species (except *P. humile*). The high genetic homogeneity and exceptional distinction of this taxon was also revealed for 48 loci controlling morphological characters and allozymes (Weeden and Wolko 2001). We conclude that *P. abyssinicum* is distinct from all other lineages, including *P. sativum*. The extreme homogeneity of *P. abyssinicum* can be explained by the possibility that this taxon has gone through a bottleneck, and that this

passage may have been associated with a hybridization event. One of the consequences of a bottleneck for population is a low percentage of polymorphic loci (Ledig et al. 1999). Although *P. abyssinicum* showed a very low percentage of polymorphic markers compared to other pea species (table 3), their frequency distribution also has the U-shape typical of almost all natural populations. The bottleneck may be associated with the cultivation of this taxon, which is endemic to the extremely hot parts of Ethiopia (Lamprecht 1974).

Pisum abyssinicum has a low number of polymorphic, species-specific, and unique markers, as well as a low proportion of total markers from the *Pisum* gene pool (table 3, fig. 6A), suggesting that few progenitors gave rise to this lineage and that *P. fulvum* was a major contributor (fig. 6). In contrast, in *P. sativum* the low proportion of unique and species-specific markers is combined with a high level of total and polymorphic markers, similar to those in *P. fulvum* and *P. elatius* (table 3). This finding is consistent with the origin of *P. sativum* from the hybridization of significantly more progenitors, with the major contribution from *P. elatius* (fig. 6A). The absence of common markers shared exclusively by *P. abyssinicum* and *P. sativum* strongly supports the idea that both species were brought into cultivation independently, and it is not consistent with the widely accepted view of *P. abyssinicum* as an ecotype or subspecies of *P. sativum* (Makasheva 1984, p.40).

Although *Pisum* is well known as an inbreeder, a significant level of heterogeneity is maintained within pea species (table 2). Our data highlight extensive introgression and intermixing among all lineages. Even the highly homogeneous *P. abyssinicum* appears to have a hybrid origin. We estimate 14.3% of markers as common among lineages (fig. 6B); however, even this fraction of insertion sites is not fixed, but its polymorphism is shared between lineages. Most of the genome exists in a state of presence-absence polymorphism. Despite the many observed differences between the main lineages, shared polymorphism is common, so the term *species complex* can be applied to *Pisum*. We conclude that, as expected for a species complex, recombination, introgression, and segregation between pea inbred lineages is common, although this may be rare per plant generation.

Acknowledgments

The John Innes Centre is supported by the Biotechnology and Biological Sciences Research Council. This work was also supported by the European Union project TEGERM (QLK-CT-2000-01502). We thank G. Moore for his helpful comments on the manuscript.

Literature Cited

- Bennetzen, J. L. 2000a. Comparative sequence analysis of plant nuclear genomes: microcolinearity and its many exceptions. *Plant Cell* **12**:1021–1029.
- . 2000b. Transposable element contributions to plant gene and genome evolution. *Plant Mol. Biol.* **42**:251–269.
- Bennetzen, J. L., and W. Ramakrishna. 2002. Numerous small rearrangements of gene content, order and orientation differentiate grass genomes. *Plant Mol. Biol.* **48**:821–827.

- Ben-Ze'ev, N., and D. Zohary. 1973. Species relationships in the genus *Pisum* L. Israel J. Bot. **22**:73–91.
- Blixt, S. 1972. Mutation genetics in *Pisum*. Agric. Hort. Genet. **30**:1–293.
- Britten, R. J., and D. E. Kohne. 1968. Repeated sequences in DNA. Science **161**:529–540.
- Burstin, J., and A. Charcosset. 1997. Relationships between phenotypic and marker distance: theoretical and experimental investigations. Heredity **78**:477–483.
- Casa, A. M., S. E. Mitchell, O. S. Smith, J. C. Register III, S. R. Wessler, and S. Kresovich. 2002. Evaluation of *Hbr* (MITE) markers for assessment of genetic relationships among maize (*Zea mays* L.) inbred lines. Theor. Appl. Genet. **104**:104–110.
- Chavanne, F., D.-X. Zhang, M.-F. Liaud, and R. Cerff. 1998. Structure and evolution of *Cyclops*: a novel giant retrotransposon of the *Ty3/Gypsy* family highly amplified in pea and other legume species. Plant Mol. Biol. **37**:363–375.
- Ellis, T. H. N., S. J. Poyser, M. R. Knox, A. V. Vershinin, and M. J. Ambrose. 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.
- Excoffier, L., P. E. Smouse, and J. M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics **131**:479–491.
- Felsenstein, J. 1993. PHYLIP (Phylogeny Inference Package). Department of Genetics, University of Washington, Seattle.
- Finnegan, D. J. 1992. Transposable elements. Curr. Opin. Genet. Dev. **2**:861–867.
- Fu, H., and H. K. Dooner. 2002. Intraspecific violation of genetic colinearity and its implications in maize. Proc. Natl. Acad. Sci. USA **99**:9573–9578.
- Ghebru, B., R. J. Schmidt, and J. L. Bennetzen. 2002. Genetic diversity of Eritrean sorghum landraces assessed with simple sequence repeat (SSR) markers. Theor. Appl. Genet. **105**:229–236.
- Govorov, L. I. 1937. *Pisum*. Pp. 231–336 in N. I. Vavilov, and E. V. Wulff, eds. Flora of cultivated plants. IV. Grain leguminosae. State Agricultural Publishing Company, Moscow, Leningrad.
- Greilhuber, J., and I. Ebert. 1994. Genome size variation in *Pisum sativum*. Genome **37**:646–655.
- Knox, M. R., and T. H. N. Ellis. 2002. Excess heterozygosity contributes to genetic map expansion in pea recombinant inbred populations. Genetics **162**:861–873.
- Kunze, R., H. Saedler, and W.-E. Lonning. 1997. Plant transposable elements. Adv. Bot. Res. **27**:331–470.
- Lamprecht, H. 1974. Monographie der Gattung *Pisum*. Steiermarkische Landesdruckerei, Graz.
- Ledig, F. T., M. T. Conkle, B. Bermejo-Velazquez, T. Eguiluz-Piedra, P. D. Hodgskiss, D. R. Johnson, and W. S. Dvorak. 1999. Evolution **53**:91–99.
- Lee, D., T. H. N. Ellis, L. Turner, R. P. Hellens, and W. G. Cleary. 1990. A *copia*-like element in *Pisum* demonstrates the uses of dispersed repeat sequences in genetic analysis. Plant Mol. Biol. **15**:707–722.
- Lu, J., M. R. Knox, M. J. Ambrose, J. K. M. Brown, and T. H. N. Ellis. 1996. Comparative analysis of genetic diversity in pea assessed by RFLP- and PCR- based methods. Theor. Appl. Genet. **93**:1103–1111.
- Makasheva, R. K. 1984. The pea. A. A. Balkema, Rotterdam.
- McClintock, B. 1946. Maize genetics. Carnegie Inst. Washington Year Book **45**:176–186.
- Navarro-Quezada, A., and D. J. Schoen. 2002. Sequence evolution and copy number of *Ty1-copia* retrotransposons in diverse plant genomes. Proc. Natl. Acad. Sci. USA **99**:268–273.
- Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.
- Nekrutenko, A., and W.-H. Li. 2001. Transposable elements are found in a large number of human protein-coding genes. Trends Genet. **17**:619–621.
- Noli, G., S. Salvi, and R. Tuberosa. 1997. Comparative analysis of genetic relationships in barley based on RFLP and RAPD markers. Genome **40**:607–616.
- Pejic, I., P. Ajmone-Marsan, M. Morgante, V. Kozumplick, P. Castiglioni, G. Taramino, and M. Motto. 1998. Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs, RAPDs, SSRs, and AFLPs. Theor. Appl. Genet. **97**:1248–1255.
- Powell, W., M. Morgante, C. Andre, M. Hanafey, J. Vogel, S. Tingey, and A. Rafalski. 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol. Breeding **2**:225–238.
- Russell, J. R., J. D. Fuller, M. Macaulay, B. G. Hatz, A. Jahoor, W. Powell, and R. Waugh. 1997. Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. Theor. Appl. Genet. **95**:714–722.
- Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.
- SanMiguel, P., and J. L. Bennetzen. 1998. Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. Ann. Bot. **82**:37–44.
- SanMiguel, P., A. Tikhonov, Y. K. Jin, N. Motchoulskaia, D. Zakharov, A. Melake-Berhan, P. S. Springer, K. J. Edwards, M. Lee, Z. Avramova, and J. L. Bennetzen. 1996. Nested retrotransposons in the intergenic regions of the maize genome. Science **274**:765–768.
- Shalev, G., and A. A. Levy. 1997. The maize transposable element *Ac* induces recombination between the donor site and an homologous ectopic sequence. Genetics **146**:1143–1151.
- Shirsat, A. H. 1988. A transposon-like structure in the 5' flanking sequence of a legumin gene from *Pisum sativum*. Mol. Gen. Genet. **212**:129–133.
- Smith, J. S. C., E. C. L. Chin, H. Shu, O. S. Smith, S. J. Wall, M. L. Senior, S. E. Mitchell, S. Kresovich, and J. Ziegler. 1997. An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.): comparisons with data from RFLPs and pedigree. Theor. Appl. Genet. **95**:163–173.
- Walker, P. M. B. 1968. How different are the DNAs from related animals? Nature **219**:228–232.
- Waugh, R., K. McLean, A. J. Flavell, S. R. Pearce, A. Kumar, B. T. Thomas, and W. Powell. 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.
- Weeden, N. F., and B. Wolko. 2001. Allozyme analysis of *Pisum sativum* ssp. *abyssinicum* and the development of a genotype definition for this subspecies. Pisum Genet. **33**:21–25.
- Wessler, S. R., T. E. Bureau, and S. E. White. 1995. LTR-retrotransposons and MITE: important players in the evolution of plant genomes. Curr. Opin. Genet. Dev. **5**:814–821.
- Wright, S. 1951. The genetical structure of populations. Ann. Eugenics. **15**:323–354.
- Zohary, D. 1996. The mode of domestication of the founder crops of near east agriculture. Pp. 142–158 in D. R. Harris, ed. The origin and spread of agriculture and pastoralism in Eurasia. University College London Press, London.

Pierre Capy, Associate Editor

Accepted July 14, 2003