The Evolution of Ty1-copia Group Retrotransposons in Gymnosperms

Caroline Stuart-Rogers¹ and Andrew J. Flavell
Department of Biochemistry, University of Dundee, Scotland

A diverse collection of Ty1-copia group retrotransposons has been characterized from the genome of Picea abies (Norway spruce) by degenerate PCR amplification of a region of the reverse transcriptase gene. The occurrence of these retrotransposable elements in the gymnosperms was investigated by Southern blot hybridization analysis. The distribution of the different retrotransposons across the gymnosperms varies greatly. All of the retrotransposon clones isolated are highly conserved within the Picea (spruce) genus, many are also present in Pinus (pine) and/or Abies (fir) genera, and some share strongly homologous sequences with one or more of cedar, larch, Sequoia, cypress, and Ginkgo. Further subclones of one of the most strongly conserved retrotransposon sequences, Tpa28, were obtained from Ginkgo and P. abies. Comparisons of individual sequence pairs between the two species show nucleotide cross-homologies of around 80%-85%, corresponding to nucleotide substitution rates similar to those of nuclear protein-coding genes. Analysis of Tpa28 consensus sequences reveals that strong purifying selection has acted on this retrotransposon in the lineages connecting Ginkgo and Picea. Collectively, these data suggest, first, that the evolution of the Ty1-copia retrotransposon group in the gymnosperms is dominated by germ line vertical transmission, with strong selection for reverse transcriptase sequence, and, second, that extinction of individual retrotransposon types has been comparatively rare in gymnosperm species lineages compared with angiosperms. If this very high level of sequence conservation is a general property of the retrotransposons, then their extreme sequence diversity implies that they are extremely ancient, and the major element lineages seen today may have arisen early in eukaryote evolution. The data are also consistent with horizontal transmission of particular retrotransposons between species, but such a mechanism is unnecessary to explain the results.

Introduction

Retrotransposons are the dominant class of dispersed repetitive DNA in many eukaryote nuclear genomes. These mobile genetic elements transpose recirculatively through RNA intermediates, and this can result in their accumulation to very high copy numbers (Pearce et al. 1996; SanMiguel et al. 1996; Kumar and Bennettzen 1999). For at least one eukaryote, maize, retrotransposons have expanded to occupy around one half of the entire genome in a few million years, causing large-scale changes in genome size (SanMiguel et al. 1996, 1998).

Ty1-copia group retrotransposons are present as large, highly heterogeneous populations within all higher plant divisions (Flavell et al. 1992; Flavell, Smith, and Kumar 1992; Voytas et al. 1992). Many phylogenetic comparisons among these populations of mobile elements have been carried out, both within and between species. These studies have shown the existence of homologous retrotransposon pairs in divergent species and discontinuous distributions of many retrotransposons among species groups (Flavell et al. 1992; Flavell, Smith, and Kumar 1992; Voytas et al. 1992). This led to suggestions that retrotransposons may have been transmitted horizontally between plant species. However, these early studies were inconclusive, because the relatively modest levels of sequence similarity detected are also consistent with the alternative hypothesis of vertical transmission, combined with low germ line transmission rates and strong selection for retrotransposon sequence conservation (Vanderwiel, Voytas, and Wendel 1993; Capy, Axolabéhere, and Langin 1994; Flavell et al. 1997). Similarly, the discontinuities in element distributions do not prove horizontal transfer and can be explained by extinctions in multiple species lineages (Capy, Axolabéhere, and Langin 1994). Subsequent phylogenetic studies have failed to provide convincing evidence for horizontal transfer of plant retrotransposons (Vanderwiel, Voytas, and Wendel 1993; Gribbon et al. 1999; Matsuoka and Tsuchewaki 1999), and it is now believed to be either rare or nonexistent. The prevailing view of plant retrotransposon evolution is that these genetic elements already existed early in plant evolution and diverged into heterogeneous subgroups before modern plant orders arose, and the descendants of these subgroups have been transmitted vertically to the descendant species.

Recently, however, strong evidence for horizontal transmission of animal retrotransposons has been reported (Gonzales and Lessios 1999; Jordan, Matyunina, and McDonald 1999), and this suggests a reappraisal of the above model for the evolution of plant retrotransposons. In this study, the heterogeneous population of Ty1-copia group retrotransposons in Norway spruce (Picea abies) was characterized and the distribution of the different retrotransposons within the branch Ginkgoes was investigated. The picture that emerged was of a very diverse family of mobile elements within P. abies which shows a complex distribution pattern within the gymnosperms, varying from genus-specific retrotransposons to extremely widely distributed and very well conserved elements. These data support the hypothesis that two counteracting processes have dominated the evolution of this transposon group in the gymnosperms. The first of these is vertical transmission down lineages,
Fig. 1.—a, Phylogenetic tree of *Picea abies* retrotransposons. Divergences in distance units (Thompson, Higgins, and Gibson 1994) between retrotransposon sequences are indicated by horizontal branch lengths. Percentage bootstrap values are shown (1,000 trials). The tree is unrooted, so the relative lengths of the two horizontal sections of the deepest branch have no significance, but the total of the two lengths is equal to the distance separating *Tpa42* from the other sequences. b, Summary of Southern data showing *P. abies* retrotransposon distributions in gymnosperms. Clusters of probes generating similar Southern hybridization patterns are bracketed. The relative hybridization signals are from subjective assessment and are not quantitative (see fig. 2 for actual data for six representative probes).

with strong selection for the sequence of at least the reverse transcriptase, and the second is extinction of retrotransposons in particular lineages. The data do not exclude the possibility of horizontal transmission of retrotransposons between species, but they do not require it.

Materials and Methods
Isolation and Characterization of Ty1-copia Group Retrotransposons

Ty1-copia group retrotransposon reverse transcriptase gene fragments were ampliﬁed by degenerate PCR of *P. abies* DNA as previously described (Flavell et al. 1992). The PCR products were subcloned into HinII-digested M13mp19 vector and sequenced using Sequenase. *Tpa9—Tpa19* correspond to EMBL database accession numbers AJ288018—AJ288028, respectively, *Tpa21—Tpa44* to numbers AJ288030—AJ288053, respectively, and *Tpa45—Tpa55* to numbers AJ288171—AJ288191, respectively. Sequences *Tgb1—Tgb19* correspond to EMBL database accession numbers AJ288182—AJ288200, respectively.

Sequence Comparisons and Phylogenetic Trees

Computing was carried out at the SEQNET facility of the Daresbury Laboratory and the HGMP facility at Hinxton, CB10 1SB, United Kingdom (http://www.hgmp.mrc.ac.uk/). Nucleotide sequences were conceptually translated in three reading frames, which were compared visually with a reference set of plant retrotransposon peptide sequences to identify frame-shifts. The corrected peptide sequences were aligned, and a phylogenetic tree was generated using CLUSTAL W (Thompson, Higgins, and Gibson 1994). Ratios of synonymous to nonsynonymous nucleotide substitutions were calculated using DIVERGE, version 10.0 (Wisconsin Package, Genetics Computer Group [GCG], Madison, Wis.). Nucleotide substitution rates were calculated by the method of Jukes and Cantor (1969). Consensus
FIG. 2.—Distribution of particular Picea retrotransposon homologs in the Gymnospermae shown by Southern analysis at moderate wash stringency (2 × SSC, 64°C). Lane 1, marker; lanes 2–7, Picea samples; lane 2, P. abies; lane 3, P. glauca; lane 4, P. obovata (partially degraded input DNA sample); lane 5, P. orientalis; lane 6, P. rubens; lane 7, P. sitchensis; lanes 8–12, Pinus samples; lane 8, P. cembra; lane 9, P. coulterii; lane 10, P. elliottii; lane 11, P. sylvestris var. Scotica; lane 12, P. thunbergii; lane 13, Abies alba (failed digestion in j); lane 14, Cedrus libani breviflora; lane 15, Larix europaea; lane 16, Ginkgo biloba; lane 17, Sequoia giganteum; lane 18, Taxodium distichum; lane 19, Welwitschia mirabilis; lane 20, marker.
nucleotide sequences were obtained using CONSENSUS (part of the GCG package).

Southern Analysis

DNAs were isolated from fresh needles or leaves using Nucleon Phytopure kits. In some cases, DNA samples which could not be restriction-digested after purification were banded by CsCl density gradient ultracentrifugation. Southern blotting to Biodyne A nylon membranes (Pall) used the manufacturer's conditions. Blot hybridization to radioactively labeled retrotransposon probes prepared by the random oligo-primed labeling method was at Tm — 20°C in 0.75 M Na+. Final washing of the blots was in 2 × SSC (0.3 M Na+) at 64°C (fig. 3) or in 2 × SSC at 64°C or 0.1 × SSC (0.015 M Na+) at 65°C (fig. 4). Blots were stripped by the manufacturer's recommended method and re-exposed to check stripping efficiency before rehybridization.

PCR Analysis of Tpa28 Reverse Transcriptase Fragments from Picea and Ginkgo

DNA sequence comparisons of Tpa28 and its closest homologs identified in P. abies identified a primer just upstream of the conserved YVDDM motif of the RT gene (CCACAATGAAGTATAGGTTGG) which is highly specific for Tpa28 alone. PCR with this primer and the conserved upstream primer (encoding TAFLHG; Flavell, Smith, and Kumar 1992) generated bands of the expected size (~220 bp) from both P. abies and Ginkgo biloba. These bands were subcloned and sequenced as above. The possibility of cross-contamination artifacts between the Picea and the Ginkgo DNA preparations was excluded by PCR using primer pairs specific for P. abies and G. biloba phytochrome genes. PCR primers selective for the P. abies phytochrome gene (EMBL ac-
Ty1-copia Retrotransposon Evolution in Gymnosperms

Ty1-copia group retrotransposon sequences were amplified from P. abies by degenerate PCR of a part of the reverse transcriptase gene using primers which have been successfully used for a variety of higher plant species (Flavell et al. 1992; Flavell, Smith, and Kumar 1992; Pearce et al. 1996; Gibbon et al. 1999). The PCR products were subcloned, inserted were sequenced, and the inferred amino acid sequences were aligned. A phylogenetic tree for these sequences is shown in figure 1a. A highly heterogeneous collection of Ty1-copia group retrotransposons was evident, as is typical for plants with large genomes (Flavell et al. 1992; Flavell, Smith, and Kumar 1992; Vanderwiel, Voytas, and Wendel 1993; Pearce et al. 1996; Gibbon et al. 1999).

To determine the distribution of these retrotransposable elements across the Gymnospermae, each of the 35 retrotransposon clones was hybridized to Southern blots (figs. 1b and 2) of DNAs from 18 species which represent the diversity of the gymnosperm branch but are skewed in favor of Picea (six species) and its sister genus Pinus (five species). For both of these genera, species from the Old and New Worlds were represented. The other seven species comprise one representative each from the Abies (fir), Cedrus (cedar), Larix (larch), Ginkgo (maidenhair tree), Sequoiadendron (redwood), Taxodium (cypress), and Welwitschia (welwitschia) genera.

The different retrotransposon probes produced a wide variety of Southern patterns. Some of these are shown in figure 2, and the overall results are summarized in figure 1b. The distribution of the different retrotransposons across the gymnosperms varies greatly. For example, Tpa43 is strictly confined to the genus Picea (fig. 2a), where it is present in high copy numbers (300 copies per genome would appear as a faint band under the conditions used here), suggesting that most of these sequences are present in very high copy numbers in these large genomes (ca. $2 \times 10^{10}$ bp). Tpa12 shows additional very weak hybridization in Pinus (fig. 2b), despite the much lower signal in Picea. Tpa17 and Tpa35 are also strongly represented across all tested Picea species (fig. 2c and d), where the Southern patterns are strongly conserved, and faint signals are discernible in Pinus. The within-genus conservation of Southern pattern is shared by all of the Ty1-copia group retrotransposons studied here (fig. 2 and data not shown), in agreement with previous results for the Tpe1 element in Pinus (Kamm et al. 1996).

Other retrotransposons show higher conservation across larger evolutionary gaps. For example, Picea Tpa29 homologs are detected strongly in the genera Picea and Abies and weakly in Ginkgo (fig. 2e). Tpa27 homologs also share this species distribution, with the addition that strongly homologous sequences are also seen in Pinus species and weaker signals are seen in Cedrus and Larix (fig. 2f). Tpa22 shows similarities in species distribution to Tpa29, except for the absence of a signal in Abies alba and Ginkgo (fig. 2g). The most extreme examples of strong conservation of Ty1-copia group retrotransposons across the range of gymnosperms tested are found for Tpa28 and Tpa13 (fig. 2h and i). Tpa28-homologous sequences are present in every gymnosperm tested except Taxodium and Welwitschia, and Tpa13-homologous sequences are found in all of the species tested. There are many cases in which particular retrotransposon restriction fragments are conserved across large evolutionary gaps (fig. 2 and data not shown), and this is particularly apparent for Tpa28 and Tpa13 (fig. 2h and i), showing that two six-base sequences and the size gap between them are conserved between plants which are separated by at least 250 Myr (see below).

The Southern blots in figure 2 had been washed at moderate stringency ($2 \times $SSC, 64°C). To get a better idea of the level of homology shared between one of the best conserved retrotransposons, Tpa28, in the gymnosperms, high-stringency Southern wash conditions ($0.1 \times $SSC, 65°C) were used (fig. 3b). These conditions are 2°C below the calculated Tm for a perfect Tpa28 hybrid. This treatment reduced the Picea hybridization signals roughly 10-fold relative to a moderate-stringency wash (fig. 3a; the exposure time for fig. 3b is threefold longer than that used for fig. 3a). The strongest Ginkgo-specific bands were still just visible after this treatment, showing that close homologs of Tpa28 were present in the Ginkgo DNA sample.

Characterization and Comparison of Ty1-copia Group Retrotransposons of P. abies and G. biloba

The phylogenetic relationship between Picea and Ginkgo homologs of one of the best conserved retrotransposons, Tpa28, was studied further. PCR primers, which specifically amplify Tpa28 alone out of all 35 P. abies elements cloned, were designed (see Materials and Methods). PCR with either of these primers and the degenerate upstream primer used in the original cloning of all of these elements (Flavell et al. 1992) was carried out to isolate Tpa28 homologs from both P. abies and G. biloba. Multiple subclones of these were sequenced as before, and the phylogenetic relationships among these sequences are shown in figure 4. The Ginkgo Tpa28 homologs form a well-defined clade with a bootstrap value of 97 which is related to a clade of three P. abies retrotransposons, Tpa52, Tpa53, and Tpa55.

The strong similarity among Tpa28 homologs from widely diverged species and that which is evident from
the Southern data in figure 2 for other retrotransposons can be explained by two contrasting models. The first proposes that retrotransposons have been transmitted vertically down the species lineages, with very strong selection acting on at least the reverse transcriptase gene. The second model proposes that some retrotransposons have been transmitted horizontally between gymnosperm species in the relatively recent evolutionary past. Vertical transmission is the normal way in which eukaryotes acquire their retroelements, but extreme sequence conservation across large evolutionary distances has shown that horizontal transmission of retrotransposons can occur (Gonzales and Lessios 1999; Jordan, Matyunina, and McDonald 1999).

The vertical-transmission model rests on the hypothesis that strong selective pressure has acted on the retrotransposon sequence in the species lineages. We therefore looked more closely at the nucleotide substitution rates of the retrotransposon sequence in the species lineages. We were generated for the five groups of sequences shown in figure 4, and substitution rates were calculated between these consensus sequences (table 1). The \(K_s/K_a\) ratio between the consensus sequence for the clade containing Tgb12 (clade D) and the consensus of all Picea Tpa28 homologs (group A) is 15.9, and the average \(K_s/K_a\) ratio between clade D and the other three groupings (A–C) is 12.8 (SD = 2.5), showing that the pseudogene effect has indeed partially masked the strong conservation of these two retrotransposon lineages and there has been strong selection acting on the reverse transcriptase of transposing homologs of Tpa28. This result is consistent with the vertical-transmission model, and, although horizontal transmission could also explain the high level of sequence conservation between the Tpa28 homologs of Picea and Ginkgo, such a mechanism is unnecessary and less parsimonious.

Strong conservation for the encoded protein sequence of clade C, the other Ginkgo Tpa28 variant, with the Picea group A and clade D consensuses was also seen (\(K_s/K_a\) values of 10.2 and 12.9, respectively). Therefore, both Ginkgo Tpa28 variants have been strongly conserved against both each other and their homologs in Picea. Paradoxically, there was little pseudogene effect apparent for clade B, the group of three Ginkgo homologs sequenced in this study (data not shown) suggest that the majority are pseudogenes.

To correct for the pseudogene effect, it is necessary to identify the particular Tpa28 retrotransposons which are being selected for transposition, i.e., those copies which are transposing replicatively. This is a very difficult task for a high-copy-number retrotransposon, but the sequence(s) of transposing elements can be inferred from the consensus of the genomic copies, which derive from the transposing elements. Therefore, consensuses were generated for the five groups of sequences shown in figure 4, and substitution rates were calculated between these consensus sequences (table 1). The \(K_s/K_a\) ratio between the consensus sequence for the clade containing Tgb12 (clade D) and the consensus of all Picea Tpa28 homologs (group A) is 15.9, and the average \(K_s/K_a\) ratio between clade D and the other three groupings (A–C) is 12.8 (SD = 2.5), showing that the pseudogene effect has indeed partially masked the strong conservation of these two retrotransposon lineages and there has been strong selection acting on the reverse transcriptase of transposing homologs of Tpa28. This result is consistent with the vertical-transmission model, and, although horizontal transmission could also explain the high level of sequence conservation between the Tpa28 homologs of Picea and Ginkgo, such a mechanism is unnecessary and less parsimonious.

### Discussion

The results presented here show that in some ways the Ty1-copia group retrotransposons of gymnosperms...
are typical of other plants with large genome sizes. They are a complicated heterogeneous population of interrelated sequences, most of which are present in very high overall copy numbers, and all individual retrotransposons transcend species boundaries. However, the degree of conservation of some of these retrotransposons is the highest reported to date for plants (Flavell et al. 1992, 1997; Flavell, Smith, and Kumar 1992; Voytas et al. 1992; Vanderwiel, Voytas, and Wendel 1993; Matsuoka and Tsunewaki 1997; Kumar and Bennetzen 1999). In plants, it is common to see Southern blot cross-hybridization for retrotransposons between related genera (Matsuoka and Tsunewaki 1997; Pearce et al. 1997; Chavanne et al. 1998), and such signals occasionally persist between angiosperm tribes such as the Triteceae and the Poaceae (Matsuoka and Tsunewaki 1997), which are separated by ca. 60 Myr, but the data presented here comprise, to our knowledge, the first reported case in which strong cross-hybridization between retroelements extends to the taxonomic level of order, across hundreds of millions of years.

Two striking examples of high sequence conservation presented here are the Tpa28 and Tpa13 elements (fig. 2h and i). These conifer-derived retrotransposons detect strongly homologous sequences across evolutionary gaps up to and including Ginkgo, the only remaining representative of the order Ginkgoales, which is first recorded definitively in the lower Permian, approximately 260 MYA (Stewart and Rothwell 1993). Modern conifers arose in the Triassic (>200 MYA). Virtually all phylogenetic analyses place Ginkgo basal to the conifer lineage (Crand 1985; Nixon et al. 1994; Qi et al. 1999; Solitis, Solitis, and Chase 1999). There is therefore a minimum separation of 260 Myr between the modern representatives of the two orders. Further investigation of the Ginkgo and Picea homologs of Tpa28 shows a pair of nucleotide sequences from these species which are 85.6% identical, corresponding to an overall nucleotide substitution rate of 6.1 × 10⁻¹⁰ per site per year. The comparable rate for the phytochrome gene of these plants is 9.9 × 10⁻¹⁰ per site per year (EMBL accession numbers GB08162 and X80298). A more comprehensive comparison between four genes (from Pinus and Ginkgo; Qi et al. 1999) shows substitution rates between 2.0 × 10⁻¹⁰ and 3.6 × 10⁻¹⁰ per site per year (data not shown). The conservation of the RT gene fragment described here is thus broadly similar to that seen for nuclear protein-coding genes.

For animals, nucleotide substitution rates similar to those seen here have been interpreted to be possibly due to horizontal transfer (ca. 3 × 10⁻¹⁰ to 8 × 10⁻¹⁰ per site per year for gypsy group retrotransposons of echnooids; Gonzales and Lessios 1999), and 100% conservation of retrotransposons in Drosophila and echinoid species has provided conclusive evidence for horizontal transmission of these retrotransposons between insect species (Gonzales and Lessios 1999; Jordan, Matyunina, and McDonald 1999). It is tempting to suggest that horizontal transmission has been responsible for the closely related Ty1-copia group retrotransposons in the gymnosperms. By such a model, Tpa28 and the other strongly conserved elements were introduced horizontally into one or both of the Ginkgo or Picea lineages long after these two lineages split but still millions of years ago, and in the ensuing period, they have become amplified by retrotransposition, with subsequent mutation yielding the many defective copies seen today.

To accept the horizontal-transmission model, it is necessary to exclude beyond reasonable doubt the alternative model that vertical transmission has been responsible for the data presented here. We are not yet convinced that this is the case, for the following reasons. First, studies on clades of echinoid retrotransposons, for which the retrotransposon phylogeny resembles the phylogenies of the corresponding organism lineages, show that the retrotransposon sequences can be conserved at rates between 1.3 × 10⁻⁹ to 2.6 × 10⁻⁸ per site per year (Gonzales and Lessios 1999). The lowest of these rates is only double that seen for Tpa28. Second, detailed study of the ratios of synonymous substitutions to nonsynonymous substitutions (K₁/Kₐ) between the Ginkgo and Picea Tpa28 homologs in figure 5 shows that they have been under strong purifying selection (see Results and table 1). The third reason for favoring the vertical-transmission model at present is the distribution of most of the retrotransposons across the gymnosperms (fig. 1b). For the majority of the Ty1-copia group retrotransposons studied here, the hybridization signal is very roughly proportional to the evolutionary distance of a given species from Picea, and there are few examples of major signal discontinuities across the species tested here. Such a result is entirely consistent with vertical transmission, but horizontal transmission would be expected to lead to departures from this situation. Furthermore, most of the retrotransposons in this study are conserved between conifers and Ginkgo (fig 1b), and if horizontal transmission has been responsible for the wide distribution of one of these (Tpa28), then presumably it has been involved in the wide distributions of the others; i.e., it has dominated the evolution of the majority of gymnosperm Ty1-copia group retrotransposons. This seems unlikely, because vertical transmission is far better established in all other eukaryote branches. In conclusion, our data do not rule out the possibility of horizontal transmission of Tpa28 or other retrotransposons between gymnosperm species, but they do not require it either, and the most parsimonious explanation is that vertical transmission, with strong sequence conservation, has dominated the evolution of this retrotransposon group in the gymnosperms.

There is no obvious reason why such strong sequence conservation should not act on the Ty1-copia group retrotransposons of all eukaryotes. The basic structures and life cycles of long-terminal-repeat (LTR) retrotransposons are shared between fungi, plants, and animals, and, as mentioned above, previous studies of animal LTR retrotransposons have revealed strong conservation for retrotransposon reverse transcriptase sequence (McAllister and Werren 1997; Gonzales and Lessios 1999). However, the large majority of the plant retrotransposons studied previously show much lower conservation across large evolutionary divides (Flavell
et al. 1992; Flavell, Smith, and Kumar 1992; Voytas et al. 1992; Matsuoka and Tsunewaki 1997, 1999; Gribbon et al. 1999). We propose that this apparent contradiction derives from the fact that the large majority of the plant sequences studied previously are from angiosperms.

Why should retrotransposons show lower rates of sequence conservation in angiosperms than they do in gymnosperms? There does not seem to be any significant difference between the overall rates of nucleotide change of nuclear genes in angiosperms and gymnosperms (Qiu et al. 1999; unpublished data). One possible explanation is that true angiosperm retrotransposon homologs, descended from a unique element in the nearest common ancestor, are rarely found, and paralogous sequences have usually been compared instead. A possible gymnosperm example of this is Tpa29 (fig. 2e). The ancestral element of Tpa29 of Picea, which shows weak homology in Ginkgo, may have arisen long before the Ginkgoales-Coniferales split, and at the time of the split, it had diverged considerably into two variants (let us call them variants 1 and 2). If variant 1 would eventually become Tpa29 in P. abies and variant 2 would become its Ginkgo homolog, then the sequence divergence between them would date back to the last common ancestor, long before the Ginkgoales-Coniferales split. Paralogy has previously been invoked (Capy, Anxolabéhere, and Langin 1994) to explain phylogenetic inconsistencies in transposon occurrence between species.

In the above example, if variant 1 survived in both lineages we would see close homology between the Ginkgo and the Picea Tpa29 elements; i.e., variant 1 must have become extinct in the Ginkgo lineage. In general, if true retrotransposon homologs between widely diverged species are much rarer in angiosperms than in gymnosperms, and this is not due to a difference in substitution rates, then extinction of retrotransposon lineages must have been much more common in angiosperms than in gymnosperms. This is an attractive model, because a potential mechanism for extinction is very active in angiosperms but is less so in gymnosperms. Drastic differences in genome sizes within a genus are common in angiosperms (http://www.rbgkew.org.uk/ceval/database1.html), and such changes can occur over a few million years (SanMiguel et al. 1998), but the genomes of gymnosperms appear to be more stable. For example, the Pinus genus is approximately 200 Myr old, all species are diploid, containing the same chromosome number, and the genome sizes are all huge (between 21 and 30 pg per haploid genome; Wakamiya et al. 1993). A chronically large genome can tolerate multiple different retrotransposon types, but a small genome size places strong constraints on retrotransposon copy number (Charlesworth, Sniegowski, and Stephan 1994), and competition between different low-copy-number elements should lead to frequent extinctions. If most of the angiosperm lineages extant today had ancestors with small genomes, retrotransposon extinction in those ancestors may have eliminated many of the different ancestral retrotransposons which appear to have survived so well in the gymnosperms.

Our findings also suggest a reevaluation of the well-known extreme heterogeneity seen among retrotransposons (Flavell et al. 1992; Flavell, Smith, and Kumar 1992; Voytas et al. 1992). For example, comparison of the reverse transcriptase regions (studied here) between the recently transposed (and presumably active) white-apricot copia retrotransposon insertion of Dro sophila and clade D consensus gives a $K_s$ value (rate of nonsynonymous substitutions per nonsynonymous site) of 0.51, compared with 0.050 for clade D versus all Picea homologs. If these two retrotransposon lineages have been subject to levels of selection similar to that of Tpa28 in the gymnosperms, then the divergence time between copia and clade D is on the order of 2.5 billion years, which is around the beginning of eukaryote life. Similar comparisons between clade D and a consensus of eight BARE-1 retrotransposon sequences of barley (Gribbon et al. 1999) or the yeast Saccharomyces cerevisiae Ty1 element give $K_s$ values of around 0.60, which is close to saturation. An alignment illustrating the relative sequence homologies between these retrotransposons is shown in figure 6.

We propose that the ancestors of clade D, copia, BARE-1, Ty1, and the many elements with similar diversity levels (Flavell et al. 1992; Flavell, Smith, and Kumar 1992; Voytas et al. 1992) split in the early phase of eukaryote life. The Ty1-copia group is only one of the three retrotransposon groups, and the oldest of these was most likely the non-LTR retrotransposons (Xiong and Eickbush 1991). Malik, Burke, and Eickbush (1999) deduced from phylogenetic data that the non-LTR retrotransposons (the probable ancestral retrotransposon group) arose and diversified early in eukaryote evolution. If our proposition for the Ty1-copia group is confirmed, it may well turn out to be the case that the antiquity of non-LTR retrotransposons has been underestimated. Finally, all of these propositions depend on the assumption that retrotransposons have evolved at roughly similar rates in multiple diverse lineages. This is a very big assumption, but it should be testable.

Acknowledgments

C.S.-R. was funded by a BBSRC earmark studentship. This work was supported by grant 960508 (TE-
BIODIV) from the European Commission under the Framework IV Biotechnology Program. We thank Steve Pearce for much helpful advice with the work; Alan Schulman for constructive criticism of the manuscript; Steve Pearce, Amar Kumar, Noel Ellis, Pat Heslop-Harrison, and Alan Schulman for many interesting discussions; and Michele Morgante for P. abies DNA samples.

LITERATURE CITED

CAPY, P. D. ANXOLABÉHERE, and T. LANGIN. 1994. The strange phylogenies of transposable elements: are horizontal transfers the only explanation? Trends Genetics 10:7–12.

CHARLESWORTH B., P. SNIEGOWSKI, and W. STEPHAN. 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371:215–220.

CHAVANNE, F., D.-X. ZHANG, M.-F. LIAUD, and R. CERRF. 1998. Structure and evolution of Cyclops: a novel giant retrotransposon of the Ty3gypsy family highly amplified in pea and other legume species. Plant Mol. Biol. 37:363–375.

CRANE, P. R. 1985. Phylogenetic relationships in seed plants. Cladistics 1:329–348.

FLAVELL, A. J., E. DUNBAR, R. ANDERSON, S. R. PEARCE, R. HARTLEY, and A. KUMAR. 1992. Ty1-copia group retrotransposons are ubiquitous and heterogeneous in higher plants. Nucleic Acids Res. 20:3639–3644.

FLAVELL, A. J., S. R. PEARCE, J. S. HESLOP-HARRISON, and A. KUMAR. 1997. The evolution of Ty1-copia group retrotransposons in eukaryote genomes. Genetica 100:185–195.

FLAVELL, A. J., D. B. SMITH, and A. KUMAR. 1992. Extreme heterogeneity of TyTy1-copia group retrotransposons in plants. Mol. Gen. Genet. 231:233–242.

GONZALEZ, P., and H. A. LESSIOS. 1999. Evolution of sea urchin retroviral-like (SURL) elements: evidence from 40 Echinoid species. Mol. Biol. Evol. 16:938–952.

GRIBBON, B. M., S. R. PEARCE, R. KALENDAR, A. H. SCHULMAN, L. PAULIN, P. J. JACK, and A. J. FLAVELL. 1999. Phylogeny and transpositional activity of Ty1-copia group retrotransposons in cereal genomes. Mol. Gen. Genet. 261:883–891.

JORDAN, I. K., L. V. MATYUNINA, and J. F. McDONALD. 1999. Evidence for the recent horizontal transfer of long terminal repeat retrotransposon. Proc. Natl. Acad. Sci. USA 96:12621–12625.

JUKES, T. H., and C. R. CANTOR. 1969. Evolution of protein molecules. Pp. 21–132 in H. N. MUNRO, ed. Mammalian protein metabolism. Academic Press, New York.

KAMM, A., R. L. DOUDRICK, J. S. HESLOP-HARRISON, and T. SCHMIDT. 1996. The genomic and physical organisation of Ty1-copia-like sequences as a component of large genomes in Pinus elliottii and other gymnosperms. Proc. Natl. Acad. Sci. USA 93:2708–2713.

KONIECZNY, A., D. E. VOYTAS, M. P. CUMMINGS, and F. M. AUSUBEL. 1991. A superfamily of Arabidopsis thaliana retrotransposons. Genetica 127:801–809.

KUMAR, A., and J. L. BENNETZEN. 1999. Plant retrotransposons. Annu. Rev. Genet. 33:479–532.

MCALLISTER, B. F., and J. H. WERREN. 1997. Evolutionary analysis of a retrotransposon with implications for strong evolutionary constraints on reverse transcriptase. Mol. Biol. Evol. 14:69–80.

MALIK, H. S., W. D. BURKE, and T. H. EICKBUSH. 1999. The age and evolution of non-LTR retrotransposable elements. Mol. Biol. Evol. 16:793–805.

MATSUOKA, Y., and K. TSUNEWAKI. 1997. Presence of wheat retrotransposons in Graminae species and the origin of wheat retrotransposon families. Genes Genet. Syst. 72:335–343.

——. 1999. Evolutionary dynamics of Ty1-copia group retrotransposons in grass shown by reverse transcriptase domain analysis. Mol. Biol. Evol. 16:208–217.

NIXON, K. C., W. L. CREPET, D. STEVENSON, and E. M. FRIS. 1994. A reevaluation of seed plant phylogeny. Ann. Mo. Bot. Gard. 81:484–533.

PEARCE, S. R., G. HARRISON, J. S. HESLOP-HARRISON, A. J. FLAVELL, and A. KUMAR. 1997. Characterisation and genomic organisation of Ty1-copia group retrotransposons in rye (Secale cereale). Genome 40:617–625.

PEARCE, S. R., G. HARRISON, D. LI, J. S. HESLOP-HARRISON, A. KUMAR, and A. J. FLAVELL. 1996. The Ty1-copia group retrotransposons in Vicia species: copy number, sequence heterogeneity and chromosomal localisation. Mol. Gen. Genet. 250:305–315.

QU, Y. L., J. LEE, E. BERNASCONI-QUADRONI, D. E. SOLTIS, P. S. SOLTIS, M. ZANIS, E. A. ZIMMER, Z. CHEN, V. SAVOLAINEN, and M. W. CHASE. 1999. The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. Nature 402:404–407.

SANMIGUEL, P., B. S. GAUT, A. TIKHONOV, Y. NAKJAMA, and J. L. BENNETZEN. 1998. The paleontology of intergene retrotransposons of maize. Nat. Genet. 20:43–45.

SANMIGUEL, P., A. TIKHONOV, Y.-K. JIN et al. (11 co-authors). 1996. Nested retrotransposons in the intergenic regions of the maize genome. Science 274:765–768.

SOLTIS, P. S., D. E. SOLTIS, and M. W. CHASE. 1999. Nature 402:402–404.

STEWART, W. N., and G. W. ROTHWELL. 1993. Paleobotany and the evolution of plants. 2nd edition. Cambridge University Press, Cambridge, England.

THOMPSON, J. D., D. G. HIGGINS, and T. J. GIBSON. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673–4680.

VANDERWIEL, P. L., D. VOYTAS, and J. F. WENDEL. 1993. Copia-like retrotransposable element evolution in diploid and polyploid cotton (Gossypium L.). J. Mol. Evol. 36:429–447.

VOYTAS, D. F., M. P. CUMMINGS, A. KONIECZNY, F. M. AUSUBEL, and S. R. RODERMEL. 1992 Copia-like retrotransposons are ubiquitous among plants. Proc. Natl. Acad. Sci. USA 89:7124–7128.

WAKAMIYA, I., R. J. NEWTON, J. S. JOHNSTON, and H. J. PRICE. 1993. Genome size and environmental factors in the genus Pinus. Am. J. Bot. 80:1235–1241.

XIONG, Y., and T. H. EICKBUSH. 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. 9:3353–3362.

PIERRE CAPY, reviewing editor

Accepted October 4, 2000