Diversity, Origin, and Distribution of Retrotransposons (*gypsy* and *copia*) in Conifers

Nikolai Friesen, Andrea Brandes,* and John Seymour (Pat) Heslop-Harrison†*

*Botanical Garden of the University of Osnabrück, Osnabrück, Germany; and
†Department of Biology, University of Leicester, Leicester, England

We examined the diversity, evolution, and genomic organization of retroelements in a wide range of gymnosperms. In total, 165 fragments of the reverse transcriptase (RT) gene domain were sequenced from PCR products using newly designed primers for *gypsy*-like retrotransposons and well-known primers for *copia*-like retrotransposons; representatives of long interspersed nuclear element (LINE) retroposons were also found. *Gypsy* and *copia*-like retroelements are a major component of the gymnosperm genome, and in situ hybridization showed that individual element families were widespread across the chromosomes, consistent with dispersion and amplification via an RNA intermediate. Most of the retroelement families were widely distributed among the gymnosperms, including species with wide taxonomic separation from the Northern and Southern Hemispheres. When the gymnosperm sequences were analyzed together with retroelements from other species, the monophyletic origin of plant *copia, gypsy,* and LINE groups was well supported, with an additional clade including badnaviral and other, probably virus-related, plant sequences as well as animal and fungal *gypsy* elements. Plant retroelements showed high diversity within the phylogenetic trees of both *copia* and *gypsy* RT domains, with, for example, retroelement sequences from *Arabidopsis thaliana* being present in many supported groupings. No primary branches divided major taxonomic clades such as angiosperms, monocotyledons, gymnosperms, or conifers or (based on smaller samples) ferns, Gnetales, or Sphenopsida (Equisetum), suggesting that much of the existing diversity was present early in plant evolution, or perhaps that horizontal transfer of sequences has occurred. Within the phylogenetic trees for both *gypsy* and *copia,* two clearly monophyletic gymnosperm/conifer clades were revealed, providing evidence against recent horizontal transfer. The results put the evolution of the large and relatively conserved genome structure of gymnosperms into the context of the diversity of other groups of plants.

Introduction

Gymnosperms, considered a sister group to the angiosperms, represent an important component of the world's plants, being the dominant vegetation type in many ecosystems and a major crop for construction materials and paper. Substantial progress has been made in understanding the structure and organization of the genomes of gymnosperms (Hizume, Ishida, and Murata 1992; Brown et al. 1993; Hizume et al. 1993; Sederoff and Stomp 1993; Doudrick et al. 1995; Lubaretz et al. 1996; Brown and Carlson 1997; Schmidt et al. 2000; Scotti et al. 2000), although our understanding is much less complete for gymnosperms than for angiosperms, many animal groups, and fungi. Most species of gymnosperms have very large genome sizes, typically with more than 20,000 Mb in the Pinaceae (Murray 1998), compared with 130–140 Mb for *Arabidopsis thaliana* or 5,500 Mb for barley (*Hordeum vulgare*). Polyploidy has played little part in the evolution of gymnosperms, and chromosome numbers are all relatively similar (Khoshoo 1959, 1961), typically 2n = 18–24, although a few species have 14 chromosomes. The geographic distribution of conifer taxa is uneven: some families, such as the Cupressaceae, are distributed in both hemispheres; the Pinaceae and Taxaceae are essentially found only in the Northern Hemisphere; the Araucariaceae and Podocarpaceae occur only in the Southern Hemisphere. Such a pattern of distribution presumably developed as the continents separated 50–135 MYA.

Retroelements and their derivatives are ubiquitous and abundant components of plant genomes (Flavell, Smith, and Kumar 1992; Voytas et al. 1992; Hirochika and Hirochika 1993; Matsuoka and Tsunewaki 1996, 1999), often representing 50% of all the DNA (Pearce et al. 1996; SanMiguel et al. 1996). Based on their structure, the retrotransposons are divided into two groups: those that are flanked by long terminal repeats (LTRs), and non-LTR retrotransposons, or long interspersed nuclear elements (LINEs; see reviews by Kumar and Bennetzen 1999; Schmidt 1999). LTR retrotransposons are further divided, most importantly into the two groups Ty1 or *copia,* and Ty3 or *gypsy*. The major structural difference between *copia* and *gypsy* groups is in the order of the reverse transcriptase (RT) and integrase domains in their pol genes. *Gypsy* group elements have similarities to retroviruses (see reviews in Bennetzen 1996, 2000; Kumar and Bennetzen 1999). The RT genes have conserved amino acid domains, some of which are characteristic of each retroelement group (Xiong and Eickbush 1990; Eickbush 1994). In plants, degenerate oligonucleotide primers have been designed to amplify these domains by PCR and used for detection and assessment of their distribution and evolution. The detailed characterization of different plant taxa with respect to the content, variability, and physical distribution of retrotransposons makes a major contribution to our understanding of host genome organization and evolution.

Abbreviations: LTR, long terminal repeat; MP, maximum parsimony; RT, reverse transcriptase.

Key words: *Picea abies,* Pinus, pine, spruce, retroelements, biodiversity, evolution, genome organization, gymnosperms, phylogenetics.

Address for correspondence and reprints: Pat Heslop-Harrison, Department of Biology, University of Leicester, Leicester LE1 7RH, United Kingdom. E-mail: phh4@le.ac.uk.

Mol. Biol. Evol. 18(7):1176–1188. 2001
© 2001 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

Copia group sequences have been found in diverse species, including single-cell algae, bryophytes, gymnosperms, and angiosperms (Voytas and Ausubel 1988; Grandbastien et al. 1998; Flavell, Smith, and Kumar 1992; Hirochika, Fukuchi, and Kikuchi 1992; Voytas et al. 1992; Kamm et al. 1996; Heslop-Harrison et al. 1997). *Gypsy*-like elements have been reported from major taxonomic groups of plants (pine [IFG7; Kossack and Kinlaw 1999], lily [del1; Smyth et al. 1989], maize [magellan; Purugganan and Wessler 1994], tomato [Su and Brown 1997], pineapple [Tomson, Thomas, and Dietzgen 1998], rice [Kumekawa et al. 1999], several angiosperms and gymnosperms [e.g., Brandes et al. 1997; Suoniemi, Tanskanen, and Schulman 1998]). It is likely that the detection methods using heterologous primers in polymerase chain reactions (PCRs) are less efficient due to relatively higher sequence heterogeneity among *gypsy*-like elements (Su and Brown 1997).

The ubiquity of plant retrotransposons, their extant sequence heterogeneity, and the function including reverse transcription suggest that their major functions were present in the first eukaryotes (see Heslop-Harrison 2000), although retrotransposons may have originated after the creation of the first eukaryotes and reached their current wide dispersal by a combination of vertical and horizontal transmission (Kumar and Bennetzen 1999). An important question in retroelement research centers on the contribution of vertical or horizontal transmission to retroelements' sequence evolution and species dispersion.

In this study, we isolated, cloned, and sequenced part of the RT gene of *gypsy*-like and *copia*-like retrotransposons from different species, focusing on the major taxonomic groups of the gymnosperms and comparing them with published sequences. We aimed to reveal the lineages of *gypsy*- and *copia*-like retrotransposons in gymnosperms and the relationships between taxonomic groups. We also aimed to characterize the content and distribution of retroelement-related DNA sequences to identify the locations of such sequences within and between chromosomes and examine their chromosomal conservation in the Pinaceae family.

Materials and Methods

Plant Material

The taxonomic classification, origin of gymnosperm species examined, and isolated clones of the retroelements are listed in table 1.

Isolation of DNA

DNA for PCR amplification was mostly isolated from young leaves, and sometimes seeds, using the Qiagen DNAeasy Plant mini kit. Isolated DNA was used directly in PCR amplifications. Isolation from seeds was particularly effective for recalcitrant material which could not be germinated easily.

PCR Assay for Reverse Transcriptases

Degenerate oligonucleotides for *gypsy*-like retrotransposons were newly designed by inspection of conserved amino acid sequences in the RT domains of different published *gypsy*-like retrotransposons (see *Results*): CyRT1 = MRNATGTGYGTNGAYTAYMG, encoding the peptide RMCVDYR, and GyRT4 = RCAYTTNSWNARYTTNGCR, encoding YAKLSKC, where R = A + G, Y = C + T, M = A + C, S = G + C, W = A + T, and N = A + G + C + T. PCR was carried out in 50 μl including 100–200 ng genomic DNA, 50 pmol of each primer, 2 U *Taq* DNA polymerase, 5 μl buffer (Life Technologies), and 3.5 μl 50 mM MgCl for amplification of *copia*-like retrotransposons following Flavell et al. (1992). PCR products were gel-purified and cloned in pGEM T-Easy vector (Promega).

DNA Sequencing

Clones were amplified by PCR with M13 primers, and 40 ng of the product was used in a 10-μl cycle sequencing reaction with the ABI BigDye Terminator Kit on an ABI 377 DNA sequencer (ABI, Foster City, Calif.). Most clones were sequenced in both strands.

Sequence Analysis

For both the *copia* and the *gypsy* data sets, initial sequence alignments and neighbor-joining trees were constructed with CLUSTAL X (Thompson et al. 1997) and improved manually. Generalized parsimony analyses were performed with PAUP*, version 3.1, with the branch-and-bound search option, MULPARS, ACCTRAN, TBR branch swapping, and gaps treated as missing. For bootstrap support (Felsenstein 1985), the same settings as in the initial tree searches were used.

DNA Labeling and Membrane Hybridization

The nonradioactive chemiluminescence method Alk-Phos Direct (Amersham) was used for DNA labeling, hybridization, and detection. Southern blots were prepared using standard protocols (Sambrook, Fritsch, and Maniatis 1989). Five micrograms, or in some cases 10 μg, of digested genomic DNA (*Hin*dIII, *Hae*III) from different conifer species was used for each row. Digested fragments were separated on 1.2% agarose gels, blotted, and probed with different clones of *Picea abies, Pinus pinaster,* and *Ginkgo biloba.*

In Situ Hybridization

Methods for chromosome preparation and in situ hybridization essentially followed Schwarzacher and Heslop-Harrison (2000). Briefly, seedling root tips were placed in ice water overnight, followed by a 3-h pretreatment with 0.05% colchicine prior to fixation in alcohol : acetic acid (3:1). Roots were partially digested with enzymes, and cells were spread on glass slides. Clones were labeled by PCR with biotin or digoxigenin. The clone pTa71, containing rDNA from *Triticum aestivum* (Gerlach and Bedbrook 1979), was used for detection of 45S rDNA sites, and clone pTa794, containing a 410-bp *Bam*H1 fragment of the 5S rDNA from of *Triticum aestivum* (Gerlach and Dyer 1980), was used

Table 1
Gymnosperm Species Used and Isolated Clones

Species and Taxonomic Position	Common Name	Origin	*Gypsy* Clones	*Copia* Clones
Pinaceae Lindl.				
Picea A. Dietr.				
P. abies (L.) Carst.	Norway spruce	Sweden, M2022	pPagy5, pPagy7, pPagy9, pPagy11, pPagy12, pPagy13, pPagy14, pPagy16, pPagy51, pPagy52, pPa15ty	pPaty1, pPaty3, pPaty5, pPaty6, pPaty7, pPaty8, pPaty11, pPaty12, pPaty13, pPaty14, pPaty16, pPaty18, pPaty19, pPaty20, pPaty21, pPaty22, pPaty23, pPaty24, pPaty25, pPaty26, pPaty27, pPaty28, pPaty29
P. omorika (Pancic) Purk.	Serbian spruce	Gatersleben, Germany	pPom5li, pPom7li	pPoty1
Pinus L.				
P. cembra L.	Arolla pine	Gatersleben, Germany		pPicty1, pPicty2, pPicty4
P. echinata Engelm.	Shortleaf pine	Harrison City, Miss.		pPecty3, pPecty4, pPecty5, pPecty6, pPecty7, pPecty8, pPecty9, pPecty10, pPecty11, pPecty12
P. elliottii Engelm.	Slash pine	Harrison City, Miss.	pPegy2	pPety1, pPety2, pPety3, pPety4, pPety6
P. palsutris Mill. ...	Longleaf pine	Harrison City, Miss.	pPipagy1, pPipagy2, pPipagy3, pPipagy4, pPipagy7	
P. maritima Poir. (syonomym *P. pinaster*)	Maritime pine		pPpgy1, pPpgy3, pPpgy5, pPp2ty, pPp4ty, pPp5ty, pPp8ty, pPp9ty, pPp10ty	pPpty1, pPpty7, pPpty13, pPpty15, pPpty16, pPpty18
P. pumila (Pall.) Regel	Small pine	East Siberia, Russia	pPipgy2, pPipgy4, pPipgy7	pPipty1, pPipty2, pPipty3, pPipty4, pPipty6, pPipty8, pPipty9, pPipty11
P. sibirica Du Tour	Siberian pine	Altai, Russia	pPsbgy1, pPsbgy2, pPsbgy3, pPsbgy4, pPsbgy5, pPsbgy6, pPsbgy, pPsb1li, pPsb2li, Psb3li	
P. strobus L.	Weymouth pine	St. Louis, Mo.	pPsgby1, pPs1ty, pPs10ty, pPs11ty	pPsty2, pPsty6, pPsty7, pPsty8, pPsty12, pPsty13
P. taeda L.	Loblolly pine	Harrison City, Miss.	pPitagy1, pPitagy5, pPitagy7	pPitaty2, pPitaty4, pPitaty8
Abies Mill.				
A. concolor Lindl. et Gord.	White fir	Gatersleben, Germany	pAcogy1, pAcogy6, pAcogy7	
A. koreana Wilson	Korean fir	Norwich, U.K.		
L. decidua Mill.	European larch	Norwich, U.K.	pLdgy5, pLd9ty, pLd12ty, pLd20ty, pLd21ty	pLdty3, pLdty6, pLdty7, Pldty8
Podocarpaceae Neger				
Podocarpus L'Herit. ex Pers.				
P. totara C. Benn. ex D. Don	Totara	New Zealand	pPot1li, pPot3li, pPot5li	
P. nivalis Hook. f.	Alpine totara	New Zealand	pPongy2	pPonty3
Dacricarpus (Endl.) de Laub.				
D. dacridiodes (A. Rich.) de Laub ..	Kahikatea	New Zealand	pDdgy1	

Table 1
Continued

Species and Taxonomic Position	Common Name	Origin	*Gypsy* Clones	*Copia* Clones
Taxodiaceae Neger				
Sequoiadendron Buchh.				
S. giganteum (Lindl.) Buchh.	Giant sequoia	Norwich, U.K.		pSdty1, pSdty2, pSdty3, pSdty5, pSdty6, pSdty7, pSdty8, pSdty9
Taxodium L.C.M. Rich.				
T. distichum L.C.M. Rich.	Bald cypress	Harrison City, Miss.	pTxgy3, pTxgy4, pTxgy6, pTxgy7	
Taxaceae S.F. Gray				
Taxus L.				
T. baccata L.	Common yew	Norwich, U.K.	pTbgy1, pTbgy4, pTbgy5, pTb2li, pTb5li	pTbty2, pTbty3, pTbty4
Ginkgoaceae Engl.				
Ginkgo L.				
G. biloba L.	Ginkgo	Gatersleben, Germany	pGigy1, pGigy6, pGi7ty, pGi11i	pGity2, pGity3, pGity4, pGity5, pGity6
Araucariaceae Strasb.				
Araucaria Juss.				
A. araucana (Mol.) C. Koch	Monkey puzzle	Norwich, U.K.	pAr1li, pAr2li, pAr3li, pAr4li, Par5li, pAr6li	

to detect 5S rDNA sites. The hybridization mixture (40% formamide, 2 × SSC, 10% dextran sulfate, 1 μg salmon sperm DNA, 0.15% SDS, and about 100 ng of two different labeled probes) was denatured and applied to the slides, and probes and slides were denatured together at 85°C for 5 min. After overnight hybridization at 37°C, slides were washed, with the most stringent wash being at 42°C in 20% formamide, 0.1 × SSC. The hybridization sites were detected using antidigoxigenin conjugated to FITC (Roche) and Cy3 conjugated to streptavidin (Sigma). After detection, the slides were washed, counterstained with DAPI (4′,6-diamidino-2-phenylindole), and photographed with appropriate filters on a Nikon epifluorescence microscope. Negatives were scanned and printed from Adobe PhotoShop 5.0 using only contrast and brightness functions affecting the whole image equally.

Results

Identification and Heterogeneity of *gypsy* and *copia* Retrotransposons

Degenerate primers have been widely used to amplify a fragment of the *copia*-like RT domain and reveal multiple families of *copia* retrotransposons in many eukaryotic species, showing the universal nature of the primers (Flavell et al. 1992; Voytas et al. 1992). However, the greater heterogeneity of *gypsy*-like elements makes designing universal primers more difficult. Analysis of *gypsy*-like sequences available in databases showed that amino acid sequence domains II and VI of the RT gene (Xiong and Eickbush 1990) were among the most conserved. We designed an upstream primer to part of the highly conserved amino acid sequence domain II, present in all described plant *gypsy*-like elements. The amino acids in domain VI are less conserved, and a more selective downstream primer was constructed; the two degenerate primers were predicted to span 420 bp. The chosen primer sequences are present in *del1* (*Lilium henryi,* X13886), IPG7 (*Pinus radiata,* AJ004945), *magellan* (*Zea mays,* U03916), the element from different tomato species (Z95335–Z955351), and other plants, as well as in Ty3 in yeast (*Saccharomyces cerevisiae,* M34549).

Consistent with the spacing of the amino acid sequence domains, the degenerate oligonucleotide primers for *gypsy* and *copia* RT domain fragments amplified PCR products of ~420 bp and ~260 bp, respectively, from the gymnosperm species studied here (table 1; 79 clones of *gypsy*-like and 86 *copia*-like retrotransposons were isolated from 21 species of gymnosperm). With the *copia* primers, about 95% of the 260-bp PCR fragments were identified as *copia*-like retrotransposons by sequence analysis, with occasional products of ~420 bp and ~600 bp that were *gypsy*-like and LINE-like, respectively. After cloning of the 420-bp PCR product from the *gypsy* primers, only half of the clones were homologous to the RT domain of *gypsy*-like elements. Products of other sizes were also analyzed, and some were retroelement-related: the *gypsy* primers also amplified ~260-bp *copia*-like products and ~600-bp LINE-related fragments. PCR reactions were carried out more than once, and there was evidence that each reaction amplified different subsets of the target sequences: for example, in the amplification of *copia*-like ele-

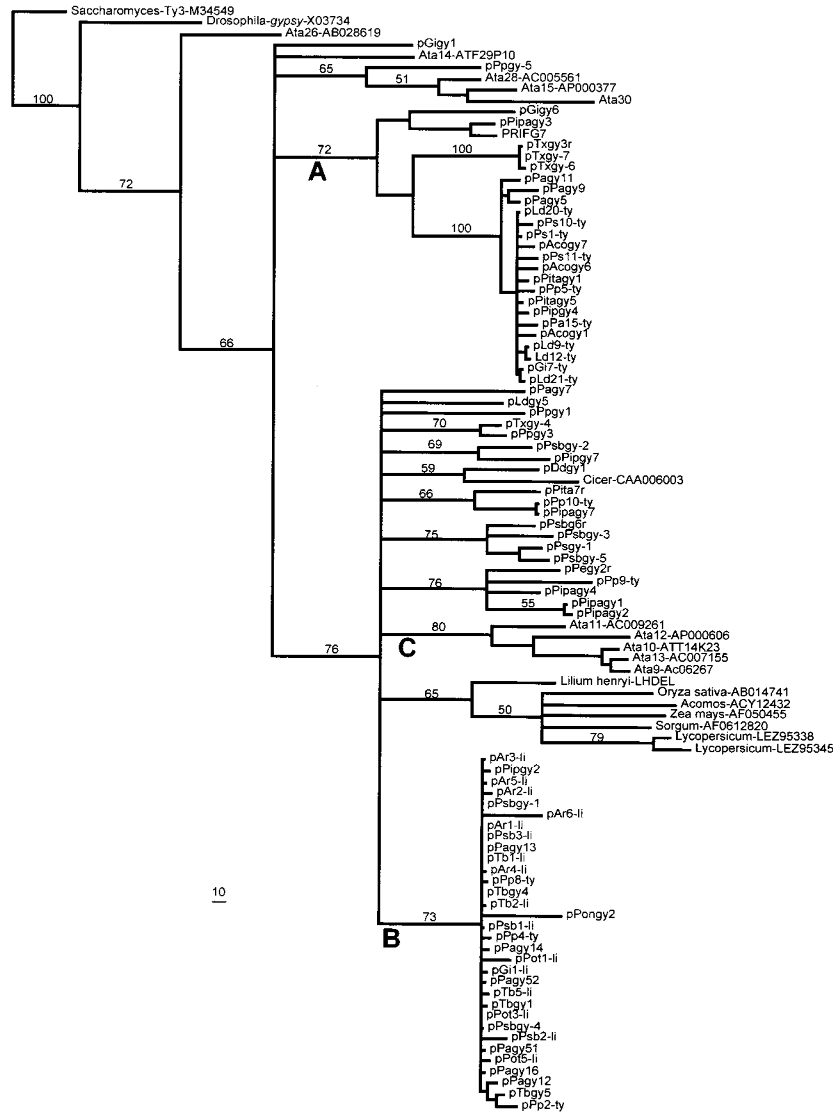


FIG. 1.—Phylogenetic analysis based on 100 sequences of part of the reverse transcriptase gene (ca. 420 bp) of *gypsy*-like retroelements; strict-consensus tree of 37,458 trees by maximum parsimony. Tree lengths = 4,007; consistency index = 0.247; homoplasy index = 0.753; retention index = 0.684. For clades with bootstrap support above 50% (calculated from 500 resamples) the values are given along the branches. Major branches are identified by the letters A, B, and C (see text). Sequence abbreviations for gymnosperm clones isolated here are given in table 1.

ments from *P. abies,* clones within each of the three reactions (*pPaty*1-*pPaty*11, *pPaty*12-*pPaty*18, and *p-Paty*19-*pPaty*29) were more similar than those between the reactions, suggesting that the degenerate primers sampled only part of the diversity of the RT genes in each amplification procedure.

Phylogenetic Analysis of Sequences *Gypsy*-like Retrotransposons

Many *gypsy*-like sequences fell into two clades (fig. 1; clades labeled A and C/B). Pairwise comparisons of the 76 *gypsy*-like retrotransposon fragments showed nucleotide homologies of 37.7% (*Ps*11ty/*Ar*6li) to 99.4% (*Psbgy*4/*Tbgy*4) between species, and a similar range of 41.5% (*Pagy*9/*Pagy*52) to 99.0% (*Pagy*14/*Pagy*16) within species (see EMBL database accessions). More detailed analysis of gymnosperm *gypsy*-like elements (420 bp) used maximum-parsimony (MP) analysis with sequences from table 1 and the database (accession numbers shown on the tree) and Ty3 as the outgroup, as shown in figure 1. MP phylogeny provided strong bootstrap support for a monophyletic origin of plant *gypsy*-like elements but showed high diversity within all species. The tree showed a clade of 32 sequences (fig. 1, branch B) with 92% homology isolated from 10 divergent gymnosperm species (*G. biloba, Araucaria araucana, Taxus baccata, Podocarpus totara* and *Podocarpus nivalis,* Pinus species, and *P. abies*). The second well-supported clade (branch A) also represented the gymnosperm species, including *G. biloba, Taxodium distichum,* and most studied species from the family Pinaceae. A larger number of well-supported small clades were not resolved by a strict-consensus tree (fig. 1), while other major phylogenetic divisions were at best weak.

There were a few monophyletic clusters representing sequences from only one species (five *A. thaliana* sequences, branch C, two clones from *Lycopersicon esculentum,* and another of three from *T. distichum*). Some reasonably well-supported clades were formed from distantly related species (*G. biloba* and *T. distichum,* or *Dacrycarpus dacridioides* and *Cicer arietinum*). Retrotransposons from *A. thaliana,* with many sequences from the genome-sequencing program in the database, were widely distributed over the tree.

Copia-like Retrotransposons

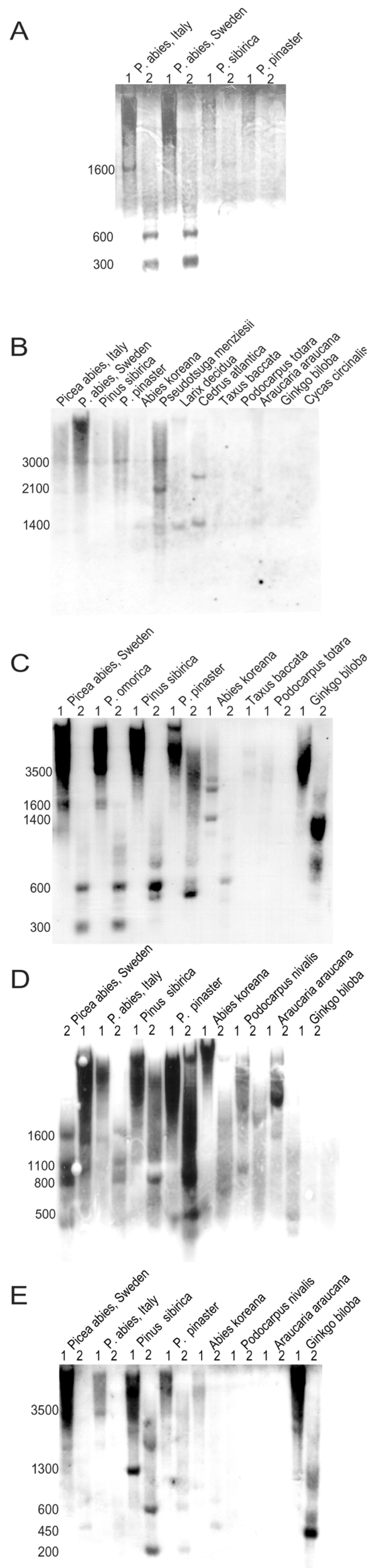
For the gymnosperm sequences analyzed, within-species similarity ranged from 40.1% (*Paty*14/*Paty*23) to 98.1% (*Paty*24/*Paty*26) in 24 sequences of *copia*-like elements from *P. abies.* Between-species homology ranged from 28.2% (*Pitaty*2/*Pipty*1) to 85.8% (*Ppty*18/*Gity*4), thus suggesting slightly higher similarity within than between species, in agreement with the suggestion and results obtained by Flavell et al. (1992) for angiosperms.

The phylogenetic picture emerging from the MP analysis of *copia*-like element sequences (260 bp, giving lower character numbers and hence lower bootstrap support values than the *gypsy* tree) is slightly different from that for *gypsy* elements (fig. 2). A monophyletic origin of plant sequences is not supported by bootstrap support values of over 50% with respect to either Ty1 from Saccharomyces or *copia* from *Drosophila melanogaster.* However, more clades of two to four sequences from single or related taxa were well supported for the *copia*-like sequences than for the *gypsy*-like sequences. As with the *gypsy* tree, similar sequences were frequently found in divergent taxa: *Cajanus cajan* (Angiospermae, Dicotyledonae) and *Gnetum montanum* (Gnetophyta); *Dicranum scoparium* (Bryophyta) and *A. thaliana* (Angiospermae; Dicotyledonae); *Equisetum arvense* (Sphenopsida) and *L. esculentum* (Angiospermae; Dicotyledonae); and *Pinus coulteri* (Coniferopsida) and *Secale cereale* (Angiospermae; Liliopsida) all showed more than 60% bootstrap support in the MP tree. Two conifer-specific branches existed in the MP *copia* tree. The first clade had 100% bootstrap support (branch A), with sequences only from different species of the genus Pinus, and the second clade had bootstrap support of 65% (branch B) and was divided into two sister subgroups: the first included 31 sequences from 12 Pinaceae species and *P. nivalis* (Podocarpaceae), with strong support (bootstrap support 100%), and the second included four sequences from another family of conifers (Cryptomeria and Sequoiadendron, Taxodiaceae). As with the *gypsy* sequences, *A. thaliana copia* sequences were found throughout the tree. The monophyly of the third supported main gymnosperm branch (branch C) in the MP tree (bootstrap support 85%) was destroyed by inclusion of a sequence from *Equisetum scirpoides* (Sphenopsida). Branch C was divided into the Equisetum branch and two others, one including sequences from only Pinus, and the second including sequences from *G. biloba, P. abies, P. pinaster,* and *Larix decidua.*

FIG. 2.—Phylogenetic analysis based on 121 sequences of part of the reverse transcriptase (ca, 260 bp) of *copia*-like retroelements; strict-consensus tree of 13,452 trees by maximum parsimony. Tree lengths = 3,876; consistency index = 0.184; homoplasy index = 0.816; retention index = 0.612. For clades with bootstrap support above 50% (calculated from 500 resamples), the values are given along the branches. Major branches are identified by the letters A, B, and C (see text). Sequence abbreviations for gymnosperm clones isolated here are given in table 1.

Membrane Hybridization

Hybridization of genomic DNA from Pinus, Picea, and Ginkgo to *gypsy*- and *copia*-like retroelement clones isolated from diverse gymnosperms revealed the diversity in copy number and species distribution of the different retroelement families represented. Consistent with the diversity of sequences present in all genera that was seen in the sequence and phylogenetic analyses, most clones showed some hybridization to each of the species, indicating that each retroelement family is present in most species (figs. 1 and 2; see also *copia* alignment



for sequence conservation, EMBL accession number DS43492). Sequences from Sequoiadendron showed weaker hybridization to the genomic DNA of the Pinaceae species, also consistent with the sequences lying in a separate clade in the trees. The divergence of the Ginkgo genome from Pinus and Picea was supported by its weaker hybridization to most clones from the conifer species.

Six *gypsy* clones, representative of their diversity in *P. abies* (*Pagy*5, *Pagy*7, *Pagy*9, *Pagy*14, *Pagy*16, and *Pa*15*ty*), and one clone (*Gigy*6) from *G. biloba* were used for Southern hybridization to DNA digests from Pinus, Picea, and other representative gymnosperms (fig. 3).

The clones *Pagy*7, *Pagy*9, and *Pa*15*ty* showed distinct hybridization patterns, with differences between genera and a greater strength of hybridization to the Picea species compared with Pinus. *Pagy*7 (fig. 3*A*) was abundant in both *P. abies* accessions, showing two bands, presumably from internal fragments of the retroelement. *Pagy*9 was abundant in most species from the family Pinaceae and showed no hybridization to taxa outside the family Pinaceae. *Pagy*7 and *Pagy*9 had weaker hybridization signal than did the abundant *Pa15ty,* suggesting that these families have lower copy numbers than *Pa*15*ty*. *Pagy*14 (fig. 3*D*) showed weaker hybridization than the other sequences, but increased DNA loading (10 μg on each lane) enabled detection of multiple bands, with similar hybridization strengths, in most species studied. This result is consistent with the clade of 32 elements with 92%–99% similarity representing many gymnosperm species (figs. 1 and 2, branch B). *Pagy*9 and *Pa*15*ty* were more abundant in species from the family Pinaceae, with weak or no hybridization to the species from other families except for *Pa*15*ty* hybridizing strongly to DNA from *G. biloba,* outside the Pinaceae. A *gypsy* clone, *Gigy*6, from *G. biloba* showed strong hybridization to digests from Pinus, Picea, and Abies as well as *G. biloba* (fig. 3*E*).

Three *copia* clones, representative of their diversity in *P. abies,* were used as probes: *Paty*5 and *Paty*14 belong to one clade, and *Paty*11 belongs to another (fig. 2). *Paty*11 showed a different hybridization pattern from *Paty*5 and *Paty*14. All three probes were present in Picea, Pinus, and Abies DNA but showed differences in genomic organization (pattern of bands) and abundance between the genera (fig. 4). It is notable that probe *Paty*11 was abundant in *G. biloba* but weak in *P. nivalis* and *A. araucana.*

No differences in genomic organization were detected between Italian and Swedish accessions of *P. abies,* or even between *P. abies* and *P. omorika.* However, small differences in genomic organization were detected between *Pinus sibirica* and *P. pinaster,* representing two different subgenera (figs. 3*C,* 3*E,* 4*A,* and 4*B*).

←

FIG. 3.—Southern hybridization patterns of *gypsy* clones (*A*) Pagy7, (*B*) Pagy9, (*C*) Pa15ty, (*D*) Pagy14, and (*E*) Gigy6 to DNA of different gymnosperm species. 1—DNA digested with *Hin*dIII; 2—DNA digested with *Hae*III.

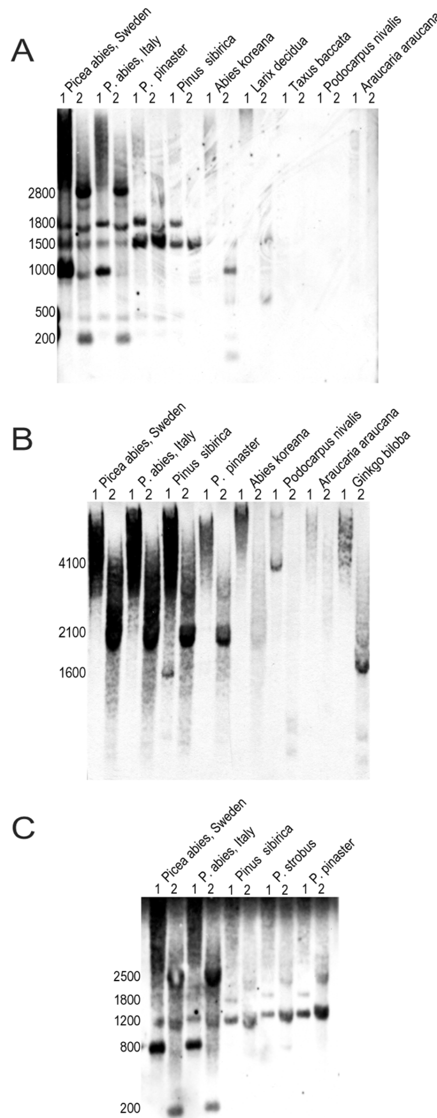


FIG. 4.—Southern hybridization patterns of *copia* clones (*A*) Paty5, (*B*) Paty11, and (*C*) Paty14. 1—DNA digested with *Hin*dIII; 2—DNA digested with *Hae*III.

In Situ Hybridization

rDNA probes were used to assist with chromosome identification by localizing the blocks of repetitive sequences and to provide a comparison with the dispersed signals seen with retroelement probes. On the *P. abies* (2n = 24) chromosomes, we detected 12 major and 2 minor sites of 45S rDNA, all at intercalary positions (fig. 5*A*), and two major intercalary pairs of 5S rDNA sites, with additional pairs of terminal sites (fig. 5*I*).

The individual retroelement probes revealed different and characteristic hybridization patterns in *P. abies.* The three *gypsy* clones *Pagy5, Pagy7,* and *Pagy11* (fig. 5*C, D,* and *F*) were distributed over the chromosomes, with particular bands showing stronger hybridization. *Pagy*11 was clustered toward the ends of all chromosome arms, while *Pagy*7 and *Pagy*9 showed more uniform distribution over the chromosomes with different intense clusters, which were stronger with *Pagy*9. In situ hybridization with clone *Pagy*14 (result not shown) gave no hybridization signal, showing that this clone has a low copy number in the genome of *P. abies,* consistent with Southern hybridization results.

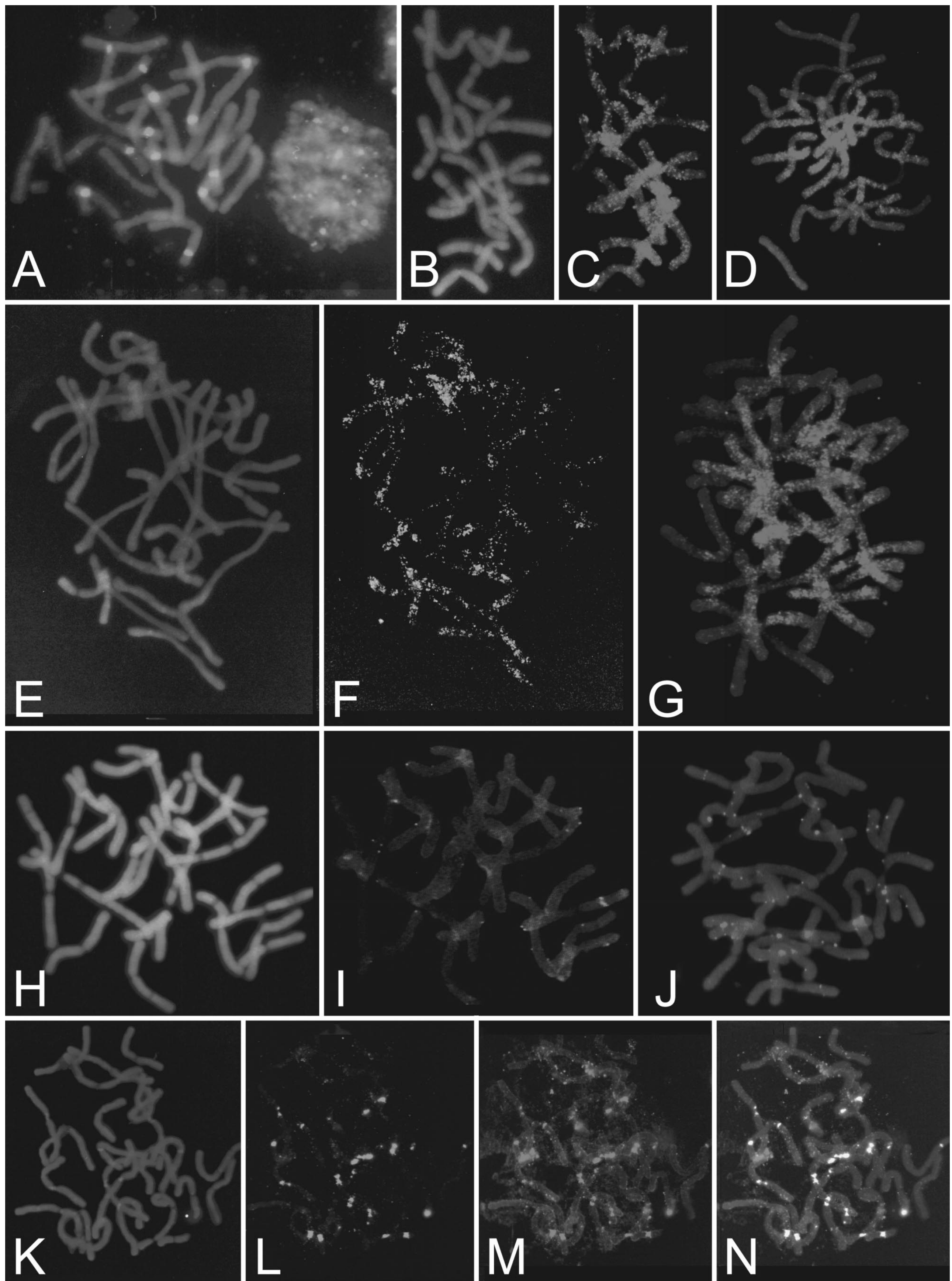
The *copia* probes *Paty*5, *Paty*11, and *Paty*14 (fig. 5*G* and *J*) also showed a dispersed distribution over the chromosomes, but *Paty*14 showed multiple, more intense, bands in the intercalary regions of many chromosome arms. Regions of weaker hybridization at rDNA sites were also revealed, particularly with *Paty*11 (fig. 5*G*).

After hybridization with a 390-bp probe of 18S rDNA genes from *P. pinaster,* chromosomes of *P. pinaster* (2n = 24) showed five chromosome pairs with strong intercalary hybridization signals and one with a weak intercalary hybridization signal, and additional clear strong signals at the centromeric regions of some chromosome pairs (fig. 5*L*). With 45S rDNA probe (pTa71), hybridization to the centromeric regions was very weak (data not shown). Hizume, Ishida, and Murata (1992) and Lubaretz et al. (1996) also reported weak hybridization signals with the rDNA probe to the centromeric regions of *Pinus thunbergii* and *Pinus sylvestris,* respectively. The *gypsy*-like element *Ppgy1,* which is isolated in the trees (e.g., fig. 1*B*), is collocalized with the 18S rDNA probe, with additional signals on the other parts of the chromosomes, including centromeric regions (fig. 5*M* and *N*). The other *gypsy*-like retroelements (*Ppgy3* and *Ppgy5*) are dispersed on all chromosomes (data not shown).

Discussion

Gypsy- and *copia*-like retroelements are a major component of the gymnosperm genome, and multiple families are present, many related to those present in other plant species. Degenerate primers designed for *gypsy* and *copia* elements amplified some members of all types of retrotransposons (figs. 1 and 2), supporting the suggestion that the RT genes of all retrotransposons are related by their common, monophyletic, origin (Xiong and Eickbush 1990; Flavell 1992; Eickbush 1994). Both membrane (figs. 3 and 4) and in situ hybridization showed that families of retroelements in gymnosperms were abundant genomic components and that major families were present in all taxa. For example, we detected strong hybridization of clones from *P. abies* (Pa15ty and Paty11) to digested DNA of *G. biloba* (figs. 3*C* and 4*B*) and vice versa (Gigy6 from *G. biloba* to DNA of Picea and Pinus species; fig. 3*E*). Furthermore, different families showed characteristic genomic distributions along chromosomes, generally being dispersed (Heslop-Harrison et al. 1997) as expected from their mode of amplification. One of the *gypsy*-like retroelements (pPgy1; fig. 5*N*) was localized in the centromeric region, as reported for an element in barley (Presting et al. 1998).

Although the sequencing of retroelements from different species is far from complete and is largely based on PCR amplification using at least partially selective primers, there are enough data that useful phylogenetic


A
B
C
D
E
F
G
H
I
J
K
L
M
N

inferences can be made. For examination of the relationships between all groups of retroelements, an unrooted neighbor-joining tree was used with fragments of the RT gene for 26 *gypsy*-like and badnaviral retroelements, 28 *copia*-like elements, and three LINEs (fig. 6). The sequences included retroelement fragments from gymnosperms (selected to sample the diversity present based on the trees; figs. 1 and 2) and published retroelement sequences, including *gypsy* and *copia* from *D. melanogaster* and Ty1 and Ty3 from *S. cerevisiae.* The alignment of individual RT domains, each typically 260 bp long, spanned only 278 bp, showing relatively high similarity. The tree supported the monophyletic origin of the *copia* and LINE clades. At the base of the *gypsy* clade, there was a grouping including banana streak badnavirus (BSV), two retroelements from legumes (*Cyclops—Pisum sativum*; broad bean element—*Vicia faba*), *gypsy* from *D. melanogaster,* and Ty3. Most of the plant *gypsy* elements also showed a monophyletic origins, and the Ty3 retrotransposon from *S. cerevisiae* is the next relative to the plant *gypsy* lineage. The broad bean and pea *Cyclops* retroelements (also placed in an anomalous position by Miller et al. [1999] and Chavanne et al.]1998[), along with the BSV sequence (Harper et al. 1999), are an exception, representing another lineage of retroelements.

The sequences of both the *copia* and the *gypsy* RT domains of plants were separated from other kingdoms with bootstrap support. Below the kingdom level, there is no dichotomy representing gymnosperms, angiosperms (figs. 1, 2, and 6), or (based on more limited data for *copia* only) other higher taxa (e.g., Ferns, Gnetales, Sphenopsida; figs. 2 and 6). It is notable that retroelements from *A. thaliana,* with numerous and unselected sequences coming from the genome sequencing program in the database, are widely distributed over all trees (*copia* and *gypsy*); the *A. thaliana* sequences have been grouped into 23 families of *gypsy*-like and 27 families of *copia*-like retrotransposons (Le et al. 2000).

Among the plant branches, there were several supported groupings representing only gymnosperms or conifers. The hybridization of genomic DNA from three species to clones representing many gymnosperm retrotransposons supported the suggestion that all species have similar diversities of retrotransposons but major differences in copy number.

The evidence from the trees allows us to explore support for three suggestions, not mutually exclusive, about the evolution of the RT domain of retroelements in gymnosperms and plants. First, there may have been an explosive radiation of retroelements of each group in the common ancestor of all plants, with most of these families having been maintained with only limited further divergence in sequence in the subsequent 350 Myr. Second, multiple horizontal transfers of the retroelements may have occurred between physically and phylogenetically distant populations or taxa. Finally, the constraints on retroelement evolution may be such that retroelements have reached near-identical ranges of sequence diversity in widespread modern plant taxa (convergent evolution) which are different from the sequence diversity ranges in the animal and fungal kingdoms. The contribution of the three models can also be examined within the gymnosperm-specific elements: were the retrotransposon families present in all gymnosperms before their divergence, becoming differentially amplified in different genomes? Or are the families showing horizontal transmission between species?

The monophyletic grouping of all types of retroelements in figure 6 and the monophyletic grouping of some families of gymnosperm-specific clones in figures 1 and 2, taken with the hybridization results, give some evidence that common ancestry and rapid radiation are the major factors in the current diversity of retroelements. Perhaps the stress events during early evolution of the kingdoms led to widespread activation of retroelements, known to be a stress response (Grandbastien 1998). As the conserved regions are over 30–40 amino acids long and many elements are nonfunctional due to the presence of stop codons and frameshifts, it is unlikely that similarities are due to convergence of function and then of sequence. Such a phenomenon may explain the similarities between two sequences, but it cannot explain the similarities between several sequences belonging to different superfamilies and classes of elements (Capy et al. 1997). If the model of early generation of the different RT domains of the retroelement families is correct, then differential amplification (and perhaps loss) of different families, without lineage-specific diversification, must have occurred. However, given that most domains are apparently nonfunctional, it is unclear why there should be this lack of diversification of the sequences (fig. 2, e.g., branch B).

There are widespread viruses and biting insects that could be responsible for the horizontal transfer of retroelements between plants. There is good evidence for horizontal transfer of *gypsy* in Drosophila (Robertson and Lampe 1995; Jordan, Matyunina, and McDonald 1999; Terzian et al. 2000), and the model of widespread horizontal transmission of elements between plants (Hir-

←

FIG. 5.—In situ hybridization to metaphase chromosomes of *Picea abies* (*A–J*) and *Pinus pinaster* (*K–N*) counterstained light blue (seen as gray) with DAPI. *A,* Metaphase chromosomes and an interphase nucleus of *P. abies* showing 12 major and 2 minor sites of the 45S rDNA (green signal, seen lighter superimposed on blue counterstained chromosomes). *B, E,* and *H,* Metaphase chromosomes and nuclei of *P. abies* counterstained with DAPI. *C,* Metaphase chromosomes of *P. abies* probed with *gypsy*-like retroelement Pagy7 (red signal). *D,* Metaphase chromosomes of *P. abies* probed with of *copia*-like retroelement Paty14 (red signal). *F,* Metaphase chromosomes of *P. abies* probed with of *gypsy*-like retroelement Pagy11 labeled with digoxigenin (green signal). *I,* Metaphase chromosomes of *P. abies* probed with 5S rDNA (green signal). *K,* Metaphase chromosomes of *P. pinaster* counterstained with DAPI. *L,* Metaphase chromosomes of *P. pinaster* probed with a 300-bp fragment of 18S rDNA from *P. pinaster* (green signal). *M,* Metaphase chromosomes of *P. pinaster* probed with *gypsy*-like retroelement Ppgy1 (red signal). *N,* Metaphase chromosomes of *P. pinaster* with rDNA and Ppgy1 sites overlaid. See www.molcyt.com for color version.

FIG. 6.—Phylogenetic analysis of nucleotide sequences of the common part of the reverse transcriptase genes from different retrotransposons; unrooted neighbor-joining tree. The branch lengths indicate the numbers of substitutions per 100 sites. Numbers associated with branches are bootstrap percentages (1,000 resamples). Main branches are indicated. Sequence abbreviations for gymnosperm clones isolated here are given in table 1.

ochika and Hirochika 1993), which would lead to the homogeneous population structure seen in large plant groupings, has been considered. But the clear monophyletic structure of the trees suggests that cross-kingdom transfer has not occurred: perhaps barriers such as methylation, codon usage, specificity of polymerases and other enzymes, or integration site specificity act to prevent such transfers.

Determination of the phylogeny of any group of retroelements comparison with the phylogeny of their hosts would suggest the validity of either model: if they were identical, then the explosive evolution model would be most likely (Eickbush 1994). However, it would be impossible to obtain unequivocal phylogenies because of the difficulty in obtaining and distinguishing members of a retroelement lineage and in determining an independent plant phylogeny. Nevertheless, there is a possibility that Ginkgo is more closely related to the family Pinaceae than is normally recognized.

The data presented here add to our knowledge of genome structure of the Pinaceae, allowing construction of molecular karyotypes of Pinus and Picea and integrating information about major classes of repetitive DNA sequences with the morphology of chromosomes. The rDNA sequences show chromosome-specific distribution patterns and hence allow identification of individual chromosomes, as was found in *Pinus elliottii* (Doudrick et al. 1995). The information about chromosome identification will be useful for the integration of genetic and physical maps and for comparative analysis of conserved synteny between the various conifer species as detailed maps of agronomic traits become available. The sequence data from the complete *A. thaliana* genome, with retroelement sequences distributed widely in our trees, shows the importance of nonselective genomic sequencing and how it can be used in understanding genome evolution in diverse species. The addition of information about biodiversity present within the retroelements, as one of the most abundant genomic components, is a valuable addition to data about gymnosperm genomes, in which molecular diversity at the DNA level seems more limited than in angiosperms, and adds another tool to those available to reconstruct the evolutionary history of the gymnosperms and the plant kingdom.

Supplementary Material

The nucleotide sequence data reported in this paper will appear in the EMBL nucleotide sequence databases under the accession numbers AJ228323–AJ228325, AJ224363–AJ224368, and AJ290587–AJ290741. The EMBL sequence alignment numbers are DS43491, DS43492, and DS43493.

Acknowledgments

We are grateful for EU Framework IV grant PL 962125 for support of this project, and we thank Trude Schwarzacher and Alex Vershinin for valuable discussions.

LITERATURE CITED

BENNETZEN, J. L. 1996. The contributions of retroelements to plant genome organisation, function and evolution. Trends Microbiol. **4**:347–353.

———. 2000. Transposable element contributions to plant gene and genome evolution. Plant Mol. Biol. **42**:251–269.

BRANDES, A., J. S. HESLOP-HARRISON, A. KAMM, R. L. DOUDRICK, and T. SCHMIDT. 1997. Comparative analysis of the chromosomal and genomic organization of *Ty1-copia*-like retrotransposons in pteridophytes, gymnosperms and angiosperm. Plant Mol. Biol. **33**:11–21.

BROWN, G. R., V. AMARASINGHE, G. KISS, and J. E. CARLSON. 1993. Preliminary karyotype and chromosomal location of rDNA sites in white spruce using fluorescence *in situ* hybridization. Genome **36**:310–316.

BROWN, G. R., and J. E. CARLSON. 1997. Molecular cytogenetics of the genes encoding 18s–5. 8s–26s rRNA and 5s rRNA in two species of spruce *Picea.* Theor. Appl. Genet. **95**:1–9.

CAPY, P., T. LANGIN, D. HIGUET, P. MAURER, and C. BAZIN. 1997. Do the integrases of LTR-retrotransposons and class II element transposases have a common ancestor? Genetica **100**:63–72.

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. Evol. **37**:363–375.

DOUDRICK, R. L., J. S. HESLOP-HARRISON, C. D. NELSON, T. SCHMIDT, W. L. NANSE, and T. SCHWARZACHER. 1995. Karyotyping slash pine *Pinus elliottii* var. *elliottii* using patterns of fluorescence *in situ* hybridization and fluorochrome banding. J. Hered. **86**:289–296.

EICKBUSH, T. 1994. Origin and evolutionary relationships of retroelements. Pp. 121–157 *in* S. MORSE, ed. The evolutionary biology of viruses. Raven Press, New York.

FELSENSTEIN, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39**:783–791.

FLAVELL, A. J. 1992. *Ty1-copia* group retrotransposons and the evolution of retroelements in eukaryotes. Genetica **86**:203–214.

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., D. B. SMITH, and A. KUMAR. 1992. Extreme heterogeneity of *Ty1-copia* group retrotransposons in plants. Mol. Gen. Genet. **231**:233–242.

GERLACH, W. L., and J. R. BEDBROOK. 1979. Cloning and characterization of ribosomal RNA genes from wheat and barley. Nucleic Acids Res. **7**:1869–1885.

GERLACH, W. L., and T. A. DYER. 1980. Sequence organization of the repeated units in the nucleus of wheat which contain 5s-rRNA genes. Nucleic Acids Res. **8**:4851–4865.

GRANDBASTIEN, M. A. 1998. Activation of plant retrotransposons under stress conditions. Trends Plant Sci. **3**:181–187.

HARPER, G., J. O. OSUJI, J. S. HESLOP-HARRISON, and R. HULL. 1999. Integration of banana streak badnavirus into the *Musa* genome: molecular and cytogenetic evidence. Virology **255**:207–213.

HESLOP-HARRISON, J. S. 2000. RNA, genes, genomes and chromosomes: repetitive DNA sequences in plants. Chromosomes Today **13**:45–56 + plate VI.

HESLOP-HARRISON, J. S., A. BRANDES, S. TAKETA et al. (15 co-authors). 1997. The chromosomal distribution of *Ty1-copia* group retrotransposable elements in higher plants and their implications for genome evolution. Genetica **100**:197–204.

HIROCHIKA, H., A. FUKUCHI, and F. KIKUCHI. 1992. Retrotransposon families in rice. Mol. Gen. Genet. **233**:209–216.

HIROCHIKA, H., and R. HIROCHIKA. 1993. *Ty1-copia* group retrotransposons as ubiquitous components of plant genomes. Jpn. J. Genet. **68**:35–46.

HIZUME, M., F. ISHIDA, and M. MURATA. 1992. Multiple location of the rRNA genes in chromosomes of pines, *Pinus densiflora* and *P. thunbergii.* Jpn. J. Genet. **67**:389–396.

HIZUME, M., H. H. TOMINAGA, K. KONDO, Z. GU, and Z. YUE. 1993. Fluorescent chromosome banding in six taxa of Eurasian *Larix.* Pinaceae. Khromosomo **69**:2342–2354.

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 **26**:12621–12625.

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

KHOSHOO, T. N. 1959. Polyploidy in gymnosperms. Evolution **13**:24–39.

———. 1961. Chromosome numbers in gymnosperms. Silvae Genet. **1**:1–9.

KOSSACK, D. S., and C. S. KINLAW. 1999. IFG, a *gypsy*-like retrotransposon in *Pinus,* Pinaceae, has an extensive history in pines. Plant Mol. Biol. **39**:417–426.

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

KUMEKAWA, N., H. OHTSUBO, T. HORIUCHI, and E. OHTSUBO. 1999. Identification and characterization of novel retrotransposons of the *gypsy* type in rice. Mol. Gen. Genet. **260**:593–602.

LE, Q. H., S. WRIGHT, Z. YU, and T. BUREAU. 2000. Transposon diversity in Arabidopsis thaliana. Proc. Natl. Acad. Sci. USA **97**:7376–7381.

LUBARETZ, O., J. FUCHS, R. AHNE, A. MEISTER, and I. SCHUBERT. 1996. Karyotyping of three Pinaceae species via fluorescent *in situ* hybridization and computer-aided chromosome analysis. Theor. Appl. Genet. **92**:411–416.

MATSUOKA, Y., and K. TSUNEWAKI. 1996. Wheat retrotransposon family identified by reverse transcriptase domain analysis. Mol. Biol. Evol. **13**:1384–1392.

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

MILLER, K., C. LYNCH, J. MARTIN, E. HERNIOU, and M. TRISTEM. 1999. Identification of multiple *gypsy* LTR-retrotransposon lineages in vertebrate genomes. J. Mol. Evol. **49**:358–366.

MURRAY, B. G. 1998. Nuclear DNA amount in gymnosperms. Ann. Bot. **82**(Suppl. A):3–15.

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. **25**:305–315.

PRESTING, G. G., L. MALYSHEVA, J. FUCHS, and I. SCHUBERT. 1998. A Ty3/*gypsy* retrotransposon-like sequence localizes

to the centromeric regions of cereal chromosomes. Plant J. **16**:721–728.

PURUGGANAN, M. D., and S. R. WESSLER. 1994. Molecular evolution of *magellan,* a maize *Ty3/gypsy*-like retrotransposon. Proc. Natl. Acad. Sci. USA **91**:11674–11678.

ROBERTSON, H. M., and D. J. LAMPE. 1995. Recent horizontal transfer of a *mariner* transposable element among and between *Diptera* and *Neuroptera.* Mol. Biol. Evol. **12**:850–862.

SAMBROOK, J., E. F. FRITSCH, and T. MANIATIS. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

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

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

SCHMIDT, A., R. L. DOUDRICK, J. S. HESLOP-HARRISON, and T. SCHMIDT. 2000. The contribution of short repeats of low sequence complexity to large conifer genomes. Theor. Appl. Genet. **101**:7–14.

SCHMIDT, T. 1999. LINEs, SINEs and repetitive DNA: non-LTR retrotransposons in plant genomes. Plant Mol. Biol. **4**: 903–910.

SCHWARZACHER, T., and J. S. HESLOP-HARRISON. 2000. Practical *in situ* hybridization. Bios, Oxford, England.

SCOTTI, I., F. MAGNI, R. FINK, W. POWELL, G. BINELLI, and P. E. HEDLEY. 2000. Microsatellite repeats are not randomly distributed within Norway spruce (*Picea abies* K.) expressed sequences. Genome **43**:41–46.

SEDEROFF, R. R., and A. M. STOMP. 1993. DNA transfer to conifers. Pp. 241–254 *in* M. R. AHUJA and W. J. LIBBY, eds. Clonal forestry. I. Genetics and biotechnology. Springer Verlag, Berlin/Heidelberg.

SMYTH, D. R., P. KALITSIS, J. L. JOSEPH, and J. W. SENTRY. 1989. Plant retrotransposon from *Lilium henryi* is related to Ty3 of yeast and the *gypsy* group of *Drosophila.* Proc. Natl. Acad. Sci. USA **86**:515–519.

SU, P. Y., and T. A. BROWN. 1997. Ty3/*gypsy*-like retrotransposon sequences in tomato. Plasmid **38**:148–157.

SUONIEMI A., J. TANSKANEN, and A. H. SCHULMAN. 1998. *Gypsy*-like retrotransposons are widespread in the plant kingdom. Plant J. **13**:699–705.

SWOFFORD, D. L. 1991. PAUP: phylogenetic analysis using parsimony. Version 3.1.1. Smithsonian Institution, Washington, D.C.

TERZIAN, C., C. FERRAZ, J. DEMAILLE, and A. BUCHETON. 2000. Evolution of the *gypsy* endogenous retrovirus in the *Drosophila melanogaster* subgroup. Mol. Biol. Evol. **17**: 908–914.

THOMPSON, J. D., Y. J. GIBSON, F. PLEWNIAK, F. JEANMOUGIN, and D. G. HIGGINS. 1997. The CLUSTAL-X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25**: 4876–4882.

TOMSON, K. G., J. E. THOMAS, and R. G. DIETZGEN. 1998. Retrotransposon-like sequence integrated into the genome of pineapple, *Ananas comosus.* Plant Mol. Biol. **38**:461–465.

VERNHETTES, S., M. A. GRANDBASTIEN, and J. M. CASACUBERTA. 1998. The evolutionary analysis of the Tnt1 retrotransposon in *Nicotiana* species reveals the high variability of its regulatory sequences. Mol. Biol. Evol. **15**:827–836.

VOYTAS, D. F., and F. M. AUSUBEL. 1988. A *copia*-like transposable element family in *Arabidopsis thaliana.* Nature **336**:242–244.

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

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

PIERRE CAPY, reviewing editor

Accepted March 9, 2001