32 Heidstra, R. and Bisseling, T. (1996) Nod factor-induced host responses and mechanisms of nod factor perception, *New Phytol.* 133, 25–43

33 Hirsch, A.M. *et al.* (1989) Early nodulin genes are induced in alfalfa root outgrowths elicited by auxin transport inhibitors, *Proc. Natl. Acad. Sci. U. S. A.* 86, 1244–1248

34 Cooper, J.B. and Long, S.R. (1994) Morphogenetic rescue of *Rhizobium meliloti* nodulation mutants by *trans*-zeatin secretion, *Plant Cell* 6, 215–225

35 Röhrig, H. *et al.* (1995) Growth of tobacco protoplasts stimulated by synthetic lipo-chitooligosaccharides, *Science* 269, 841–843

36 Walden, R. and Lubenow, H. (1996) Genetic dissection of auxin action: more questions than answers? *Trends Plant Sci.* 1, 335–339

37 Bauer, P. *et al.* (1996) Nod factors and cytokinins induce similar cortical cell division, amyloplast deposition and *MsEnod12A* expression patterns in alfalfa roots, *Plant J.* 10, 91–105

38 Silver, D.L. *et al.* (1996) Posttranscriptional regulation of the *Sesbania rostratao* early nodulin gene *SrEnod2* by cytokinin, *Plant Physiol.* 112, 559–567

39 van de Sande, K. *et al.* (1996) Modification of phytohormone response by a peptide encoded by *ENOD40* of legumes and a nonlegume, *Science* 273, 370–373

40 Van Rijn, P. *et al.* Signal transduction pathways in forming arbuscular-mycorrhizae and *Rhizobium*-induced nodules may be conserved based on the expression of early nodulin genes in alfalfa mycorrhizae, *Proc. Natl. Acad. Sci. U. S. A.* (in press)

41 Li, Y., Hagen, G. and Guilfoyle, T.J. (1991) An auxin-responsive promoter is differentially induced by auxin gradients during tropisms, *Plant Cell* 3, 1167–1175

42 Boerjan, W. *et al.* (1995) *superroot*, a recessive mutation in *Arabidopsis*, confers auxin overproduction, *Plant Cell* 7, 1405–1419

43 Baskin, T.I. *et al.* (1995) *STUNTED PLANT 1*, a gene required for expansion in rapidly elongating but not in dividing cells and mediating root growth responses to applied cytokinin, *Plant Physiol.* 107, 233–243

44 Deikman, J. and Ulrich, M. (1995) A novel cytokinin-resistant mutant of *Arabidopsis* with abbreviated shoot development, *Planta* 195, 440–449

Catharina Coenen is at Universität Freiburg, Institut für Biologie II, Zellbiologie, Schänzlestraße 1, 79104, Germany; Terri Lomax* is at the Dept of Botany and Plant Pathology and Center for Gene Research and Biotechnology, Oregon State University, Corvallis, OR 97331-2902, USA.

*Author for correspondence (tel +1 541 737 5278; fax +1 541 737 3573; e-mail lomaxt@bcc.orst.edu).

Complex gene families in pine genomes

Claire S. Kinlaw and David B. Neale

The genome structures of extant species suggest that conifer and angiosperm genomes have evolved by different mechanisms. For example, in the evolution of the pine genome, the amplification and dispersal of genes to form complex families appears to have been especially prominent. An analysis of the structure and organization of pine gene families is critical for understanding the organization and evolution of pine genomes, and may help explain adaptation.

Photo supplied by H.D.B.

Conifer genomes are remarkable for their large size. For example, haploid pine nucleii contain between 21 and 31 pg DNA[1]. Reassociation kinetic analysis has demonstrated the presence of repeated sequences whose copy numbers vary over a broad range[2]. Among the repeated sequences of pine are very high copy number sequences found in the genomes of other plants[3]. These include ribosomal genes[4] and noncoding intergenic DNA regions such as microsatellites[5] and retrotransposons[6].

In addition to highly repeated DNA sequences, conifer genomes also contain multiple copies of sequences that hybridize to cDNA probes in Southern (DNA) hybridizations. When results for anonymous pine cDNA sequences[7] are compared with those for angiosperm cDNA sequences[8–11], there is evidence that amplification of gene sequences has been more active during pine genome evolution, creating numerous complex families (Table 1). Similar results are seen for a wide range of conifer genomes[12].

In general, the degree of observable gene family complexity correlates with plant genome size (Table 1). The smallest and simplest genomes, which have the least repetitive DNA, include *Arabidopsis*[13], rice[14] and tomato[14]. These are all angiosperms in which most proteins are encoded by simple gene families[8,9,13]; larger angiosperm genomes show a higher percentage of complex gene families. However, even when compared with relatively complex genomes such as that of maize[11], pine genomes appear to have fewer simple gene families and more multicopy families (Table 1).

There are two possible explanations for the Southern hybridization patterns observed for anonymous pine cDNAs. Gene families may have evolved that are composed of many members whose sequences have diverged. Alternatively, individual genes may have evolved to be large, and these contain either many introns or introns of a large size. The structure of the few conifer genes that have been characterized suggests that it is not size alone that is responsible for these complex Southerns. For example, an alcohol dehydrogenase gene characterized from loblolly pine (*Pinus taeda*)

PII S1360-1385(97)01099-6

Table 1. Southern hybridization patterns using anonymous cDNAs

Species	Genome size (pg haploid^{-1})	Low complexity (percentage with 1–3 bands)	Medium complexity (percentage with 4–10 bands)	High complexity (percentage with >10 bands)	Refs
Rice	0.45[a]	66	17	17	8
Tomato	1.0[a]	78	18	Not reported	9
Lettuce	2.7[a]	66	34	Not reported	10
Maize	2.6[a]	50	50	Not reported	11
Loblolly pine (*Pinus taeda*)	22.0[b]	27	45	28	7

[a]Data from Ref. 14. [b]Data from Ref. 1.

has nine introns, each <500 bp in size, and all are in conserved locations relative to their maize counterparts[15]. The independent segregation of multiple loci from many complex gene families in pine mapping populations[7] further supports the idea of multilocus gene families as opposed to individual large loci.

Further evidence that gene family complexity is more prominent in pines than angiosperms comes from an analysis of specific pine gene families whose putative biochemical functions have been identified (Table 2). Identification is through cDNA sequencing and sequence comparisons to public databases (http://www.cbc.med.umn.edu). Although some gene families are similarly complex for both angiosperms and conifers, many examples exist of genes with few copies in angiosperms[16–22], but many copies in pines (Table 2).

Mechanisms of gene amplification

Gene amplification events appear to have been frequent throughout pine evolution, and some events appear to have occurred in recent geological time. Southern hybridizations of loblolly pine cDNAs to a variety of pine genomic DNA sequences, including the closely related species slash pine (*P. elliottii*)[12], revealed five out of 30 sequence families that are more complex in the pine species other than loblolly pine (Table 3). The amplification and dispersal of gene family members might be expected to result in a significant loss of gene order between species if such mechanisms were ongoing and random as species evolved.

Evidence of stable gene order over time comes from recent mapping studies comparing loblolly pine and Monterey pine (*P. radiata*)[23]. Thus, the mechanisms proposed to explain the generation of complex gene families in pine must reconcile the dispersed nature of these gene families with the apparent paradox that the order of genes along pine chromosomes may be well conserved.

The comparative mapping studies of Devey *et al.*[23] emphasized the very similar Southern banding patterns of pine genes, and this emphasis improved confidence that genomic locations of orthologs (the same locus in two species), and not paralogs (genes duplicated in the same species), were being compared from each pine species. However, because genes with conserved Southern patterns have highly conserved DNA sequences, it is unclear whether this conservation of gene order holds true for all genes. Southern patterns with anonymous cDNAs have revealed that many pine gene sequences are not so highly conserved[12], and it is possible that genes whose sequences have diverged more may also show more divergent gene order.

The question as to whether the amplification of pine gene sequences results in functional genes or pseudogenes (or both) is also unresolved. Mechanisms that might generate complex gene families with functional members include duplications of whole genomes (polyploidy), duplications of whole chromosomes (aneuploidy), duplications of large chromosome segments or duplications of

Table 2. Highly complex pine gene families with simple angiosperm counterparts

Pine gene family identified by partial cDNA sequence	Pine Southern pattern	Angiosperm species	Angiosperm family complexity	Refs
Chaperonin 60 beta[a]	>10 bands[b]	*Arabidopsis*	Low copy number	16
Thiolase[a]	>10 bands[b]	Cucumber	Single copy	17
Elongation factor 1α[a]	>10 bands[b]	Tomato	Single copy	18
Acid phosphatase[a]	>10 bands[b]	Tomato	Single copy	19
Actin-depolymerizing factor[a]	>10 bands[b]	Rape	Low copy number	20
Heat shock polypeptide HSP90[a]	>10 bands[b]	Madagascar periwinkle (*Catharanthus roseus*)	Single copy	21
Alcohol dehydrogenase[c]	>10 bands[c]	Maize	Two loci	22

[a]Data from http://www.cbc.med.umn.edu. [b]Data from Ref. 7. [c]Data from Ref. 30.

Table 3. Gene family complexity differences revealed in pines using loblolly pine (*Pinus taeda*) cDNAs as probes

Putative function or cDNA clone number	Family complexity observed in loblolly pine with loblolly pine cDNA	Family complexity observed in other pines with loblolly pine cDNA
Clone 0147[a]	2–3 bands[b]	4–10 bands (slash pine)[c]
Protein kinase[a]	1 band[b]	4–10 bands (slash pine)[c]
Membrane protein[a]	4–10 bands[b]	>10 bands [Scots pine (*Pinus sylvestris*)][c]
Peroxidase[a]	4–10 bands[b]	>10 bands [sugar pine (*Pinus lambertiana*)][c]
Glutamine synthetase[a]	2–3 bands[b]	4–10 bands [western white pine (*Pinus monticola*)][c]

[a]Data from http://www.cbc.med.umn.edu. [b]Data from Ref. 7. [c]Data re-evaluated from Ref. 12.

small chromosome regions containing complete genes. The dispersal of genes generated by such duplications would require random crossover events, and the degree of dispersal of family members would depend upon the time at which such duplications occurred.

The duplication of whole genomes has played an important role in the evolution of angiosperm genomes, and thus in the generation of gene families in such species[3]. For example, maize is an ancient tetraploid, and wheat is hexaploid. Polyploidy also appears to have played an important role in the evolution of at least one group of 'primitive' vascular plants, the ferns, which have high chromosome numbers[24]. In contrast, pine gene families are unlikely to have arisen by duplication of either whole genomes or individual chromosomes. There is no cytogenetic evidence that pine genomes are polyploid[25], and all extant pine species, of which there are more than 100, are diploid, with a diploid chromosome number of 24. Mapping data[7] have revealed no evidence for large duplicated linkage groups suggestive of polyploidy or aneuploidy.

The duplication of functional genes has played an important role in the generation of gene families in all higher organisms[26]. Duplicated genes evolve new regulatory sequences, and these provide new patterns of gene expression in different tissues at different developmental stages or in response to different environmental signals. Related but distinct protein functions are also thought to evolve from the shuffling of coding regions from duplicated genes[27]. Thus, functional gene duplications have the potential strongly to influence the direction of evolution and adaptation.

The pine alcohol dehydrogenase gene family is an example of a complex pine gene family that has more functional gene family members than its angiosperm counterparts. In jack pine (*P. banksiana*), at least seven linked functional alcohol dehydrogenase loci have been identified by polymerase chain reaction amplification of mRNA from the haploid nutritive seed tissue, the megagametophyte[28]. The clustering of seven loci into two linked groups may reflect the occurrence of duplications at varying times during pine evolution that have not yet had time to disperse throughout the genome by random crossover events. Southern hybridizations with alcohol dehydrogenase cDNA probes reveal many more than seven bands (Table 2), and the possibility remains that some of the Southern bands result from pseudogenes.

In contrast to the duplication of genomes, chromosomes, chromosome segments or complete genes, retrotranscription of RNA molecules via reverse transcriptase provides a mechanism for generating nonfunctional gene family members that can be integrated into sites that are not linked to the original gene. There is evidence for the existence of such retropseudogenes in Norway spruce (*Picea abies*)[29]. Do some of the many Southern bands seen for pine alcohol dehydrogenase reflect dispersed nonfunctional copies? Will highly complex pine gene families turn out to have both functional and pseudogene members? Because of the high copy number of retrotransposons in the genomes of extant pine species, it is tempting to invoke the mechanism of reverse transcription in the generation of complex pine gene families.

Evolutionary significance of complex gene families

The prevalence of complex gene families in the genomes of extant pine species suggests that the evolution of conifer and angiosperm genomes has proceeded along different paths, and this raises intriguing questions. Do long-lived and slowly growing species such as pine simply tolerate the presence of gene sequence families along with highly repetitive noncoding intergenic DNA, or do such sequences have adaptive value? Before answering this question, it will be necessary to determine whether pine gene amplifications create functional or nonfunctional copies. Also, are there multiple mechanisms, some of which generate functional gene copies and others that produce nonfunctional copies? How frequently have gene amplifications occurred? If, as genetic maps suggest, dispersed gene families are a central feature of pine genome structure and evolution, how has dispersal ensued, and how have gene order and chromosome stability been maintained as new copies arise and disperse across the genome? The answers to such questions may provide interesting insights into the evolution of pine genomes and may also have implications for plant adaptation.

Acknowledgements

The authors gratefully acknowledge editing by Sandra Young and thoughtful comments from John Brabson and Bohun Kinloch.

References

1 Wakamiya, I. *et al.* (1993) Genome size and environmental factors in the genus *Pinus*, *Amer. J. Bot.* 80, 1235–1241

2 Kriebel, H.B. (1985) DNA sequence components of the *Pinus strubus* nuclear genome, *Can. J. For. Res.* 15, 1–4

3 Dean, C. and Schmidt, R. (1995) Plant genomes: a current molecular description, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 46, 395–418

4 Cullis, C.A. *et al.* (1988) The 25S 18S and 5S ribosomal RNA genes from *Pinus radiata* D. Don, in *Petawawa National Forestry Institute Information Report P1-X-80* (Cheliak, W.M. and Yapa, A.C., eds), pp. 34–40, Canadian Forestry Service

5 Echt, C.S. and May-Marquardt, P. (1997) Survey of microsatellite DNA in pine, *Genome* 40, 9–17

6 Kamm, A. *et al.* (1996) The genomic and physical organization of *Ty1-copia*-like sequences as a component of large genomes in *Pinus elliotti* var. *elliotti* and other gymnosperms, *Proc. Natl. Acad. Sci. U. S. A.* 93, 2708–2713

7 Devey, M.E. *et al.* (1994) An RFLP linkage map for loblolly pine based on a three-generation outbred pedigree, *Theor. Appl. Genet.* 88, 273–278

8 Causse, M.A. *et al.* (1994) Saturated molecular map of the rice genome based on an interspecific backcross population, *Genetics* 138, 1251–1274

9 Bernatzky, R. and Tanksley, S.D. (1986) Majority of random cDNA clones correspond to single loci in the tomato genome, *Mol. Gen. Genet.* 203, 8–14

10 Landry, B.S. *et al.* (1987) Comparison of restriction endonucleases and sources of probes for their efficiency in detecting restriction fragment length polymorphisms in lettuce (*Lactuca sativa* L.), *Theor. Appl. Genet.* 74, 646–653

11 Shen, B. *et al.* (1994) Partial sequencing and mapping of clones from two maize cDNA libraries, *Plant Mol. Biol.* 26, 1085–1101

12 Ahuja, M.R. *et al.* (1994) Mapped DNA probes from loblolly pine can be used for restriction fragment length polymorphism mapping in other conifers, *Theor. Appl. Genet.* 88, 279–282

13 Meyerowitz, E.M. and Pruitt, R.E. (1985) *Arabidopsis thaliana* and plant molecular genetics, *Science* 229, 1214–1218

14 Arumuganathan, K. and Earle, E.D. (1991) Nuclear DNA content of some important plant species, *Plant Mol. Biol. Rep.* 9, 208–218

15 Harry, D. *et al.* (1989) DNA sequence diversity in ADH genes from pines, in *The Proceedings of the 20th Southern Forest Tree Improvement Conference*, pp. 373–380

16 Zabaleta, E. *et al.* (1992) Isolation and characterization of genes encoding chaperonin 60 beta from *Arabidopsis thaliana*, *Gene* 111, 175–181

17 Preisig-Müller, R. and Kindl, J. (1993) Thiolase mRNA translated *in vitro* yields a peptide with a putative N-terminal sequence, *Plant Mol. Biol.* 22, 59–66

18 Shewmaker, C. *et al.* (1990) Nucleotide sequence of an EF1-alpha genomic clone from tomato, *Nucleic Acids Res.* 18, 4276

19 Tanaka, H. *et al.* (1992) Partial sequence of acid phosphatase-1 gene (*Aps*-1) linked to nematode resistance gene (*Mi*) of tomato, *Biosci. Biotechnol. Biochem.* 56, 583–587

20 Kim, S-R., Kim, Y. and An, G. (1993) Molecular cloning and characterization of anther-preferential cDNA encoding a putative actin-depolymerizing factor, *Plant Mol. Biol.* 21, 39–45

21 Schröder, G. *et al.* (1993) HSP90 homologue from Madagascar periwinkle (*Catharanthus roseus*): cDNA sequence, regulation of protein expression and location in the endoplasmic reticulum, *Plant Mol. Biol.* 23, 583–594

22 Dennis, E.S. *et al.* (1985) Molecular analysis of the alcohol dehydrogenase 2 (*Adh2*) gene of maize, *Nucleic Acids Res.* 13, 727–743

23 Devey, M.E., Sewell, M.M. and Neale, D.B. (1996) A comparison of loblolly and radiata pine genomes using RFLP markers, in *Tree Improvement for Sustainable Tropical Forestry* (Dieters, M.J. *et al.*, eds), pp. 478–480

24 Gastony, G.J. (1991) Gene silencing in a polyploid homosporous fern: paleopolyploidy revisited, *Proc. Natl. Acad. Sci. U. S. A.* 88, 1602–1605

25 Mirov, N.T. (1967) *The Genus Pinus*, Ronald Press

26 Ohno, S. (1970) *Evolution by Gene Duplication*, Springer

27 Gilbert, W. (1978) Why genes in pieces? *Nature* 271, 501

28 Perry, D.J. and Furnier, G.R. (1996) *Pinus banksiana* has at least seven expressed alcohol dehydrogenase genes in two linked groups, *Proc. Natl. Acad. Sci. U. S. A.* 93, 13020–13023

29 Kvarnheden, A., Tandre, K. and Engrö, M.P. (1995) A *cdc2* homologue and closely related processed retropseudogenes from Norway spruce, *Plant Mol. Biol.* 27, 391–403

30 Kinlaw, C.S., Harry, D.E. and Sederoff, R.R. (1990) Isolation and characterization of alcohol dehydrogenase cDNAs from *Pinus radiata*, *Can. J. For. Res.* 20, 1343–1350

Claire Kinlaw is at the Institute of Forest Genetics, Pacific Southwest Research Station, United States Dept of Agriculture Forest Service, PO Box 245, Berkeley, CA 94701, USA; David Neale is at the Institute of Forest Genetics, Pacific Southwest Research Station, United States Dept of Agriculture Forest Service, PO Box 245, Berkeley, CA 94701, USA and the Dept of Environmental Horticulture, University of California (Davis), Davis, CA 95616, USA.*

**Author for correspondence (tel +1 510 559 6429; fax +1 510 559 6499; e-mail csk@s27w007.pswfs.gov).*

Articles of interest in recent issues of other *Trends* magazines

Viral nucleic acid sequence transfer between fungi and plants, *J.R. Marienfeld, M. Unseld, P. Brandt* and *A. Brennicke*, **Trends in Genetics** 13, 260–261

Origin of genes encoding multi-enzymatic proteins in eukaryotes, *J.N. Davidson* and *M.L. Peterson*, **Trends in Genetics** 13, 281–285

Novel protein serine/threonine phosphatases: variety is the spice of life, *P.T.W. Cohen*, **Trends in Biochemical Sciences** 22, 245–251

Palaeontology in a molecular world: the search for authentic ancient DNA, *J.J. Austin, A.B. Smith* and *R.H. Thomas*, **Trends in Ecology & Evolution** 12, 303–306