

- HEWITT, G. M. (2001): Speciation, hybrid zones and phylogeography – or seeing genes in space and time. *Molecular Ecology* **10**: 537–549.
- ISODA, K., S. SHIRAIISHI and H. KISANUKI (2000): Classifying *Abies* species (Pinaceae) based on the sequence variation of a tandemly repeated array found in the chloroplast DNA *trnL* and *trnF* intergenic spacer. *Silvae Genetica* **49**: 161–165.
- KORMUTAK, A., B. VOOKOVA and B. ZIEGENHAGEN (2002): Reproductive isolation between Colorado white fir (*Abies concolor*) and the Mediterranean firs. *Biologia Bratislava* **57**: 527–532.
- KORMUTAK, A., B. VOOKOVA, B. ZIEGENHAGEN, H. Y. KWON and Y. P. HONG (2004): Chloroplast DNA variation in some representatives of the Asian, North American and Mediterranean Firs (*Abies* spp.). *Silvae Genetica* **53**: 99–104.
- LIEPELT, S., R. BIALOZYT and B. ZIEGENHAGEN (2002): Wind-dispersed pollen mediates postglacial gene flow among refugia. *Proceedings of the National Academy of Sciences of the U.S.A.* **99**: 14590–14594.
- MATTFELD, J. (1926): Die europäischen und mediterranen *Abies*-Arten. *Die Pflanzenareale* **1**: 22–29.
- PALMER, J. D. (1987): Chloroplast DNA evolution and biosystematic uses of chloroplast DNA variation. *American Naturalist* **130**: 6–29.
- PARDUCCI, L. and A. E. SZMIDT (1999): PCR-RFLP analysis of cpDNA in the genus *Abies*. *Theoretical and Applied Genetics* **98**: 802–808.
- PETIT, R. J., C. BODÉNÈS, A. DUCOUSO, G. ROUSSEL and A. KREMER (2003): Hybridization as a mechanism of invasion in oaks. *New Phytologist* **161**: 151–164.
- PETIT, R. J. and G. G. VENDRAMIN (2005): Plant phylogeography based on organelle genes: an introduction, in *Phylogeography of Southern European Refugia* edited by S. WEISS and N. FERRAND, Kluwer, Dordrecht (in press).
- TABERLET, P., L. GIELLY, G. PAUTOU and J. BOUVET (1991): Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* **17**: 1105–1109.
- TSUMURA, Y. and Y. SUYAMA (1998): Differentiation of mitochondrial DNA polymorphisms in populations of five Japanese *Abies* species. *Evolution* **52**: 1031–1042.
- WU, J., K. KRUTOVSKII and S. H. STRAUSS (1998): Abundant mitochondrial genome diversity, population differentiation and convergent evolution in pines. *Genetics* **150**: 1605–1614.

## Evolution of Genome Size in Conifers

By M. RAJ AHUJA<sup>1</sup> and DAVID B. NEALE<sup>1,2</sup>

(Received 9<sup>th</sup> June 2005)

### Abstract

Conifers are the most widely distributed group of gymnosperms in the world. They have large genome size (1C-value) compared with most animal and plant species. The genome size ranges from ~6,500 Mb to ~37,000 Mb in conifers. How and why conifers have evolved such large genomes is not understood. The conifer genome contains ~75% highly repetitive DNA. Most of the repetitive DNA is composed of non-coding DNA, including ubiquitous transposable elements. Conifers have relatively larger rDNA repeat units, larger gene families generated by gene duplications, larger nuclear volume, and perhaps larger genes, as compared to angiosperm plants. These genomic components may partially account for the large genome size, as well as variation in genome size, in conifers. One of the major mechanisms for genome size expansion and evolution of species is polyploidy, which is widespread in angiosperms, but it is rare in conifers. There are only a few natural polyploids in one family of conifers, Cupres-

saceae. Other conifers, including well-studied pines, are nearly all diploids. Whether ancient polyploidy has played a role in the evolution of genome size in conifers still remains an open question. The mechanisms that account for the variation and evolution of genome size in conifers are addressed in this review.

*Key words:* genome size, conifers, pines, polyploidy, paleopolyploidy, duplicate genes, repetitive DNA, retrotransposons, introns, genome evolution.

### Introduction

Conifers, belonging to the order Coniferales, are the largest and the most diverse group of cone-bearing gymnosperms that are distributed widely, albeit unevenly, throughout the world (FARJON, 1998). While conifers mostly dominate the temperate zone forests in the northern hemisphere, they are rather scattered in the southern hemisphere. They are wind-pollinated, highly heterozygous, and long-lived trees with vegetative phases extending from one to several decades, and even centuries in the case of bristle cone pine (*Pinus longaeva*). Conifers are economically the most important plant group of wood and fiber crops. Fossil records suggest

<sup>1</sup> Institute of Forest Genetics, USDA Forest Service.

<sup>2</sup> Department of Plant Sciences, University of California, 1 Shields Avenue, Davis, CA 95616, USA.

Correspondence: E-mail: mrahuja@hotmail.com

that conifers originated from protogymnosperms during the Carboniferous Period (325 mya), and were most diverse and abundant during the Jurassic Period (195 mya) (MILLER, 1977; STEWART and ROTHWELL, 1993).

Modern Conifers are placed in eight families, 68 genera, and 630 species (FARJON, 1998). These families include Araucariaceae (3 genera), Cephalotaxaceae (1 genus), Cupressaceae sensu lato (s.l.) (28 genera, including 9 genera of Taxodiaceae), Phyllocladaceae (1 genus), Pinaceae (11 genera), Podocarpaceae (18 genera), Sciadopityaceae (1 genus), and Taxaceae (5 genera). The haploid chromosome numbers in conifers range between 9 and 19: 11 and 12 being the most common numbers (KHOSHOO, 1961). Most species are diploids, and polyploidy is rare in conifers. Recent molecular phylogeny studies based on chloroplast, mitochondrial, nuclear genes, and intron loss suggest that gymnosperms are divided into five different groups, namely: Cycadales, Ginkgoales, Gnetales (Gnetaceae, Ephedraceae, Welwitschiaceae), Pinaceae, and Coniferales II (comprising of all conifer families except Pinaceae) (BOWE *et al.*, 2000; CHAW *et al.*, 2000; GUGERLI *et al.*, 2001). These molecular studies have also identified Cycadales as the most basal group of gymnosperms, and Gnetales and Pinaceae clade form a sister group to Coniferales II. For the purposes of our discussion on genome size we shall treat Coniferales II and Pinaceae under one group Coniferales.

Conifers have relatively large genomes compared to most other land plant species (LEITCH *et al.*, 2005). How and why conifers have evolved such large genomes is not understood. Increase in genome size may occur by several different mechanisms. These include genome duplication (polyploidy), gene duplications, amplification of transposable elements, and increase in number and size of introns. Which one of these mechanisms has predominantly contributed to the evolution of large genome size in conifers? Even though extant polyploidy is rare in conifers, has ancient polyploidy played a role in the evolution of genome size in conifers? We examine these questions in this review.

### Genome Size in Plants

Genome size, referred to as the C-value, is the amount of DNA in an unreplicated gametic nucleus of an organism (BENNETT *et al.*, 1998; SOLTIS *et al.*, 2003a). It varies ~200,000-fold in eukaryotes, ranging from 9.2 Mb in the fungus *Ashbya gossypii* (DIETRICH *et al.*, 2004) to 680,000 Mb in *Amoeba dubia* (GREGORY, 2001). Genome sizes exhibit enormous variation in both plants and animals: it ranges >2000-fold between different land plant species (BENNETT and LEITCH, 2003), and >3000-fold in different animal species (GREGORY, 2001). In plants genome size has been estimated in about 4000 species, which ranges in size from 50 Mb in *Cardamine amara* (Brassicaceae) to 125,000 Mb in *Fritillaria assyriaca* (Liliaceae) (BENNETT and SMITH, 1991; BENNETT and LEITCH, 2003; SOLTIS *et al.*, 2003a). However, recent estimates of genome size of *Cardamine amara* have given a higher 1C-value (~220 MB), and that the previous 1C value (50 Mb) is an error (BENNETT and LEITCH, 2005).

In view of this finding, the 1C genome in *Arabidopsis* (157 Mb) estimated by BENNETT *et al.* (2003) may be considered perhaps the smallest genome in angiosperms. Most angiosperms (~50%) have relatively small 1C-values (1C <3,400 Mb) within five-fold the mode (600 Mb), and very large genomes are represented in only two distantly related groups (monocots and Santalales) (LEITCH and BENNETT, 2002; SOLTIS *et al.*, 2003a). But the range of genome size is the greatest (~1,000-fold) in angiosperms of all land plant groups (SOLTIS *et al.*, 2003a).

Conifers have relatively large modal genome sizes as compared to most other plants species, including many angiosperm tree species (Table 1). The genome size ranges from ~6,500 Mb to ~37,000 Mb in conifers, showing a 5-fold variation. By contrast, the genome size in several representative genera of the angiosperm trees ranges between 540 Mb to 2,000 Mb (Table 1). Although some angiosperm trees may have larger genomes (>2,000 Mb), for example in *Camelia*, by and large the modal 1C genome in angiosperm trees is <1,000 MB, as compared to the modal 1C genome in conifers (>15,000). Genome size in gymnosperms, other than conifers, ranges from 3,820 Mb in *Gnetum* to 18,000 Mb in *Ephedra*, also showing about 5-fold variations (LEITCH *et al.*

Table 1. – Genome size in some conifers and angiosperm tree species.

Species	Chromosome Number (n)	Genome size <sup>a</sup> 1 C (Mb) <sup>b</sup>	Family
<b>Conifers</b>			
<i>Pinus gerardiana</i>	12	37,000	Pinaceae
<i>Pinus lambertiana</i>	12	33,500	Pinaceae
<i>Pinus strobus</i>	12	31,200	Pinaceae
<i>Pinus pinaster</i>	12	30,300	Pinaceae
<i>Pinus longaeva</i>	12	28,500	Pinaceae
<i>Pinus resinosa</i>	12	28,000	Pinaceae
<i>Pinus radiata</i>	12	26,500	Pinaceae
<i>Pinus sylvestris</i>	12	24,600	Pinaceae
<i>Pinus caribaea</i>	12	24,300	Pinaceae
<i>Pinus elliotii</i>	12	24,200	Pinaceae
<i>Pinus taeda</i>	12	24,000	Pinaceae
<i>Pinus banksiana</i>	12	21,500	Pinaceae
<i>Pseudotsuga menziesii</i>	13	18,700	Pinaceae
<i>Picea abies</i>	12	18,200	Pinaceae
<i>Sequoia sempervirens</i>	33	31,500	Cupressaceae
<i>Calocedrus decurrens</i>	11	15,400	Cupressaceae
<i>Metasequoia glyptostroboides</i>	11	10,800	Cupressaceae
<i>Sequoiadendron giganteum</i>	11	9,700	Cupressaceae
<i>Taxodium distichum</i>	11	9,700	Cupressaceae
<i>Sciadopitys verticillata</i>	10	20,400	Sciadopityaceae
<i>Taxus baccata</i>	12	10,900	Taxaceae
<i>Araucaria cunninghamii</i>	13	10,700	Araucariaceae
<i>Lepidothamnus intermedium</i>	15	6,500	Podocarpaceae
<b>Angiosperms</b>			
<i>Acacia victoriae</i>	13	2,000	Leguminosae
<i>Ulmus glabra</i>	14	1,100	Ulmaceae
<i>Quercus robur</i>	12	800	Fagaceae
<i>Eucalyptus grandis</i>	11	600	Myrtaceae
<i>Populus deltoides</i>	19	540	Salicaceae

a. Data in Pinaceae, Araucariaceae, Taxaceae, and Podocarpaceae from MURRAY (1998), SILJAK-YAKOVLEV *et al.* (2002); except the genus *Pinus* from GROTKOPP *et al.* (2004); Cupressaceae, and Sciadopityaceae from HIZUME *et al.* (2001); *Quercus* from OHRI and AHUJA (1990); *Eucalyptus* (GRATTAPAGLIA and BRADSHAW (1994); *Populus* from DHILLON (1987); and *Acacia* and *Ulmus* from BENNETT and SMITH (1991). In all these studies, the nuclear DNA amounts were given in picograms (pg).

b. Conversion basis: 1 pg = 980 Mb (BENNETT *et al.*, 2003).

al., 2001) The genome size in pteridophytes shows enormous variation (>450-fold) in genome size: it ranges from 157 Mb to 71,250 MB (OBERMAYER *et al.*, 2002).

Based on the genome size in eukaryotes, it clear that there is little relationship between the genome size and the degree of organismal complexity, giving rise to the phenomenon known as 'C-value paradox' (THOMAS, 1971). The C-value paradox has been a driving force for searching for mechanisms that would account for the variation in genome size in eukaryotes. A partial explanation to the C-value paradox may lay in the differential amounts of non-coding, repetitive, DNA in the genome that may lead to the genome size variation in eukaryotes (PETROV, 2001; GREGORY, 2001; KIDWELL, 2002). However, relative contributions of different types of non-coding repetitive DNA and other genomic components to genome size still remains unanswered. (KIDWELL, 2002; GREGORY, 2005).

#### *Intraspecific and Interspecific Genome Size Variation*

Variation in genome size within a species has been reported in a number of plant species (PRICE, 1988; OHRI, 1998). For example, genome size varies 1.5-fold within provenances/populations of *Pinus banksiana*, 1.6-fold in *Picea glauca* (MIKSCH, 1968), 2.2-fold in *Pinus resinosa* (DHIR and MIKSCH, 1974), 2.5-fold in *Zea mays* (RAYBURN *et al.*, 1985; BENNETT and SMITH, 1991), 1.3-fold in *Pisum sativum* (CAVALLINI *et al.*, 1993), and 1.15 in *Glycine max* (GRAHAM *et al.*, 1994). The intraspecific genome variation seems to be associated with the environmental conditions or growth parameters (PRICE, 1988; KNIGHT *et al.*, 2005). Seed size and duration of tree development appear to be positively correlated with increase of genome size in conifers (OHRI and KHOSHOO, 1986; WAKAMIYA *et al.*, 1993; NEWTON *et al.*, 1999).

Plant species from the northern regions of the northern hemisphere tend to have larger nuclear volumes and higher DNA content as compared to those from the southern regions (STEBBINS, 1966). This correlation has been observed within the populations (intraspecific variation) of a number of conifer species, including *Picea glauca*, *Picea sitchensis*, *Pinus sylvestris*, *Pinus banksiana* and *Pseudotsuga menziesii* (MERGEN and THIELGES, 1967; EL-LAKANY and SZIKLAI, 1971; MIKSCH, 1971). Populations of *Picea rubens*, with relatively high DNA content, seem to be less tolerant to high altitude as compared to those of *Picea mariana*, with a low nuclear DNA amount (BERLYN *et al.*, 1990). However, there are exceptions to such relationships. A south-to-north increasing DNA gradient was not observed in the populations of *Pinus resinosa* (DHIR and MIKSCH, 1974). Stability in the nuclear DNA content was also reported in the disjunct populations of Fraser fir (*Abies fraseri*) in the Appalachian Mountains (AUKLAND *et al.*, 2001).

There is an association between a relatively larger genome size (28,200 Mb) of *Pinus gerardiana* with temperate and xeric habitat, as compared to pines from tropical habitats, for example *Pinus caribaea* (19,200 Mb) (OHRI and KHOSHOO, 1986). Recent estimates of genome size in pines by GROTKOPP *et al.* (2004), however, give much higher values for these pine species (*Pinus gerardiana* 37,000 Mb and *Pinus caribaea* 24,300 Mb,

see Table 1), but the overall genome size ratio remains the same. In conifers, the large genome size tends to correlate with increase in the nuclear volume (BURLEY, 1965; MERGEN and THIELGES, 1967). In the interspecific genome size comparisons in conifers, species with smaller nuclear volume and DNA content seem to display greater geographical distribution than species with larger volumes and nuclear DNA amounts (NEWTON *et al.*, 1999).

Association between genome size and several other growth traits and ecological factors have also been observed in a number of pine species (WAKAMIYA *et al.*, 1993; JOYNER *et al.*, 2001) According to a recent large genome constraint hypothesis (KNIGHT *et al.*, 2005), the ecological and evolutionary constraints may affect the phenotype and physiological processes of large genome species and confer a selective disadvantage by restricting distribution. How the environmental stresses shape the genome size in the populations of a species is not well understood. One possibility is that the variability in the genome size may be caused by the modulated of retrotransposons, discussed in a later section.

#### **Mechanisms of Genome Size Expansion and Contraction**

Genome size can expand by several mechanisms. These include: 1) whole genome duplication (polyploidy), 2) gene duplication, 3) modulation of repetitive DNA sequences, and 4) increase in intron size. On the other hand, reduction in genome size may occur by illegitimate recombination and deletion of genetic material, mostly non-coding DNA (WENDEL, 2000; PETROV, 2001; BENNETZEN, 2002; BENNETZEN *et al.*, 2005). Here we shall focus on the genome expansion mechanisms for an understanding of genome size evolution in conifers.

##### *1. Whole Genome Duplication (Polyploidy)*

Polyploidy occurs both in plants and animals. It occurs at a relatively high frequency in plants as compared to animals. Polyploidy has been reported in animals that reproduce by parthenogenetic means, for example insects and amphibians, but occurs rarely in sexually reproducing animals, such as mammals (OTTO and WHITTON, 2000). But polyploidy provided a rapid means for the evolution of new genes and speciation during the early evolution of both animals and plants, and still continues to be an important mechanism for speciation of plants (WENDEL, 2000; RAMSEY and SCHEMSKE, 2002; SOLTIS *et al.*, 2003b; BLANC and WOLFE, 2004).

##### *Polyploidy and evolution of species*

The role of polyploidy as a mechanism for the evolution of eukaryotes has attracted a lot of attention during the past several decades (STEBBINS, 1950; WENDEL, 2000; SOLTIS *et al.*, 2003b). OHNO (1970) in his seminal classic "Evolution by Gene Duplication", proposed that it is much easier to create new genes by duplication than to produce them *de novo*, and that genome duplication via polyploidy was a quicker way to generate a vast number of duplicate genes. OHNO postulated that two or possibly three rounds of polyploidizations, later dubbed as 2R

hypothesis (HUGHES, 1999), occurred during the early vertebrate evolution. Comparative genomic (map-based) and phylogenetic (tree-based) studies have suggested that a number of animals and plants are paleopolyploids or ancient polyploids (WOLFE, 2001; BLANC and WOLFE, 2004). These paleopolyploids include humans (GIBSON and SPRING, 2000; McLYSAGHT *et al.*, 2002), fishes (VAN DE PEER *et al.*, 2003), maize (GAUT and DOEBLEY, 1997), rice (PATERSON *et al.*, 2004; although VANDERPOELE *et al.*, 2003 have presented evidence to suggest that rice and other cereals are ancient aneuploids), *Arabidopsis* (The Arabidopsis Genome Initiative, 2000), yeast (WOLFE, 2001), tomato, cotton, and soybean, (BLANC and WOLFE, 2004), that later diploidized by sequence divergence between duplicated chromosomes (WOLFE, 2001). These species have probably undergone ancient rounds of chromosome doubling followed by sequence divergence between duplicated chromosomes and deletions leading to gene loss. Further reduction or increase in genome size in these organisms may have occurred during the course of evolution by the interplay between the non-coding repetitive DNA and coding sequences. The 2R hypothesis has generated a lot of interest and debate during the last decade regarding the role of polyploidy and re-establishment of diploidy in eukaryotes (MARTIN, 1999; HOLLAND, 1999; GIBSON and SPRING, 2000; WOLFE, 2001; McLYSAGHT *et al.*, 2002; PRINCE and PICKETT, 2002). There are questions whether one round (1R) or two rounds (2R) of polyploidizations occurred, or whether large-scale segmental duplications could account for the evolution of animals and plants (SKRABANEK and WOLFE, 1998; HUGHES, 1999; WOLFE, 2001; SANKOFF, 2001; MARTIN, 2001; MAKALOWSKI, 2001; ZHANG, 2003; VANDERPOELE *et al.*, 2003).

#### Polyploidy in conifers

Polyploidy, that is presence of more than two genomes per nucleus, is widespread in plants. Recent estimates suggest that 50 to 80% of all angiosperms are polyploids (MASTERSON, 1994; OTTO and WHITTON, 2000). Many angiosperms may have experienced one or more episodes of polyploidization during their evolution (SOLTIS and SOLTIS, 1999; WENDELL, 2000). In the pteridophytes the incidence of polyploidy may be close to 95% (GRANT, 1981). Although polyploidy is relatively common in the angiosperm trees, it is rather infrequent among gymnosperms (KHOSHOO, 1959; WRIGHT, 1976; AHUJA, 2001, 2005). The frequency of polyploidy may be close to 5% in the gymnosperms and about 1.5% in the conifers (KHOSHOO, 1959).

There are only a few naturally occurring polyploids among conifers (KHOSHOO, 1959; DELEVORYAS, 1980; AHUJA, 2005). These include: two tetraploids *Fitzroya cupressoides* ( $2n=4x=44$ ) (HAIR, 1968), and *Juniperus chinensis* 'Pfitzeriana' ( $2n=4x=44$ ) (SAX and SAX, 1933), and a hexaploid *Sequoia sempervirens* ( $2n=6x=66$ ) (STEBBINS, 1948; SAYLOR and SIMONS, 1970; SCHLARBAUM and TSUCHIYA, 1984). Interestingly, all three genera belong to the family Cupressaceae, but show different types of polyploidy: *Fitzroya* is very likely an autotetraploid (PREMOLI *et al.*, 2000), *Juniperus* is an allotetraploid (DE LUC *et al.*, 1999), and *Sequoia* remains an

enigmatic hexaploid (either an autoallohexaploid, or a segmental allohexaploid, or a partially diploidized autohexaploid (STEBBINS, 1948; AHUJA and NEALE, 2002; AHUJA, 2005). Polyploidy is conspicuously absent in other families of conifers, including well-studied pines.

#### Are pines ancient polyploids?

Nearly all the genera in the family Pinaceae (for example, *Pinus*, *Picea*, and *Larix*) are diploid ( $2n=24$ ; exception *Pseudotsuga menziesii* also a diploid, but with  $2n=26$ ). The karyotype of *Pinus* species has been studied more extensively than other genera of conifers (SAYLOR, 1972, 1983; PEDERICK, 1970; HIZUME *et al.*, 2002). Cytogenetic studies have not detected polyploidy in pines (KHOSHOO, 1961; MIROV, 1967). However, based on similar Giemsa bands on different chromosomes in the pine genome (*Pinus resinosa*), DREWRY (1988) suggested that hidden polyploidy has played a role in the evolution of the pine genome. But superficial homology of Giemsa bands on different chromosomes may not necessarily be indicative of ancient polyploidy, without the genomic sequence analysis. The question is how have pines achieved such large chromosomes and genome size during their evolution?

Is it possible that ancient polyploidy has played a role in the evolution of pines and other gymnosperms? In view of high incidence of polyploidy in angiosperms, it has been suggested that many if not all plant species have had at least one polyploid ancestor at some point during their evolution (WENDEL, 2000; BLANC and WOLFE, 2004). Are pines and other conifers exception to this rule?

The origin of the genus *Pinus* is thought to be in early to middle Mesozoic (MILLAR, 1998). Fossil record suggests that ancient species of *Pseudoaraucaria* and *Pityostrobus*, closely related to pines, may have provided the ancestral gene pool of pines (MILLAR, 1998). Although the genome size in prehistoric *Pseudoaraucaria* is not known, arbitrarily we have assumed that the genome size in *Pseudoaraucaria* and *Pityostrobus* may be ~10,000 Mb to present three different models for the origin of pines from their putative ancestors (models based on higher or lower genome sizes of the ancestral genomes can also be constructed). Is it possible that the pines are ancient polyploids derived by either: 1) hybridization between some ancient species of *Pseudoaraucaria*-like, *Pityostrobus*-like, or another ancient conifer, followed by one round (1R) of polyploidization and subsequent diploidization, or 2) one round of autopolyploidization (1R) in a putative pine ancestor, followed by diploidization, or 3) large segmental duplications in a putative pine ancestor, leading to enlargement of genome size, followed by sequence divergence? Of the three different hypothetical scenarios presented, two are based on the assumption that ancient polyploidy may have played a role in the pine evolution. While the third scenario does not involve polyploidy *per se*, instead it invokes large-scale segmental duplications (both gene and chromosomal segments) for the evolution of the pine genome (AHUJA, 2005). Studies in genomic research in pines would be necessary to discriminate between these postulates.

Even though there are a number of plant species that are now considered paleopolyploids (WOLFE, 2001; BLANC and WOLFE, 2004), sometimes, it is rather difficult to detect paleopolyploidy because: 1) time erases the traces of duplication, 2) majority (70–90%) of duplicated genes formed during the millions of years of polyploid evolution may return to single copy state, thus reestablishing disomic segregation, for example, as in *Arabidopsis* and yeast, and 3) chromosomal rearrangements relocate duplicate segments around the genome, which further scramble the intragenomic synteny (OTTO and WHITTON, 2000; BLANC and WOLFE, 2004). The current consensus map of loblolly pine, *Pinus taeda*, has not provided convincing evidence for the presence of duplicated syntenic regions (SEWELL *et al.*, 1999) to support ancient polyploidy in pines. Nevertheless, comparative genomic studies in pines along with other plant species would be necessary to resolve the issue of paleopolyploidy in pines and other conifers.

Although polyploidy is generally accompanied by an increase in genome size, this is not always the case. In some polyploidy is accompanied by the additive sum of the parental species (for example, *Nicotiana*, *Gossypium*, *Triticum*), while in others (for example, *Betula*, *Brassica*, *Ranunculus*) there was a reduction in genome size relative to the parental contribution (LEITCH and BENNETT, 2004). In other words, polyploidy is accompanied by downsizing of the genome size in some plants. Therefore, polyploidy must be considered in conjunction with other genomic components for evaluating evolution of genome size in plants.

## 2. Gene Duplication

Another mechanism of genome expansion involves gene duplications. Many genes exist in two or multiple copies. There are a number of different mechanisms by which gene duplications can arise. These include, in addition to chromosome and genome duplication discussed above, tandem duplications, unequal crossing over, gene conversion, duplication of chromosome segments. Duplicate genes provide the raw genetic material on which mutations and natural selection could operate for the evolution of novel gene functions and associated phenotypes. Accumulation of duplicate genes (mostly nonfunctional) in tandem or dispersed arrays may account for the genome size variation.

### Divergence and maintenance of gene function

Regardless their origin, duplicate genes may have at least three different kinds of evolutionary fates: 1) non-functionalization, 2) neo-functionalization, and 3) sub-functionalization (HALDANE, 1933; FISHER, 1935; FORCE *et al.*, 1999; WALSH, 2003). Since deleterious mutations occur much more frequently than the beneficial ones, the fate of the duplicated gene, in a process of non-functionalization, results in the loss of function of the gene leading to a pseudogene or deletion due to chromosome rearrangements, or gene silencing. However, less frequently, in neo-functionalization, one of the pair of duplicate genes may acquire a new function due to a beneficial mutation.

According to sub-functionalization, the duplicate copies of an ancestral gene acquire complimentary function due to mutations in independent sub-functions, so that both partially defective genes produce the full function of the ancestral gene, but in somewhat different way. Recent studies, based on the available genomic databases in a number of eukaryotic species (animals, plants and yeast), have shown that a surprisingly large number of duplicate genes are present in the sequenced genomes, indicating evolutionary conservation of genes through global DNA duplication events (LYNCH and CONERY, 2000; PRINCE and PICKETT, 2002; ZHANG, 2003). The duplicate genes apparently arise at a very high rate, an average 0.01 per gene per million years (LYNCH and CONERY, 2000; LYNCH, 2002).

Duplicate genes (Table 2) identified by fairly selective criteria account for approximately 38% of the genes in humans, *Homo sapiens*, 30% for yeast, *Saccharomyces cerevisiae*, 65% for *Arabidopsis*, 41% for fruit fly, *Drosophila*, and 49% for nematode, *Caenorhabditis elegans* (RUBIN *et al.*, 2000; ZHANG, 2003), and 35% for maize, *Zea mays*, (GAUT, 2001).

### Gene duplication and multigene families

Although precise estimates of gene duplications are lacking in conifers, a large number of multigene families have been detected in the loblolly pine, *Pinus taeda*, (DEVY *et al.*, 1994) and other conifer species, including several *Pinus* species, *Douglas fir* (*Pseudotsuga menziesii*), *Norway spruce* (*Picea abies*), and coast redwood (*Sequoia sempervirens*) by Southern hybridization to pine cDNA probes (AHUJA *et al.*, 1994). The multigene families may be genetically linked, or dispersed in the genome. Southern hybridization patterns using cDNAs have revealed complex band patterns suggesting that the large conifer genome contains relatively larger gene families and/or larger genes (AHUJA *et al.*, 1994; KINLAW and NEALE, 1997) as compared to other angiosperms with smaller genomes, including those of rice (CAUSEE *et al.*, 1994) and maize (SHEN *et al.*, 1994). One of the best-characterized gene families in plants, the alcohol dehydrogenase (*Adh*) gene family, is larger in pines than in angiosperms (KINLAW *et al.*, 1990). Numerous duplications, as large as 217 bp, were detected within the non-coding regions of *Pinus banksiana Adh* genes, and may be common feature of the conifer genome (PERRY and FURNIER, 1996).

In general, the degree of multigene family complexity seems to be correlated with the plant genome size. For example, plant with smallest and simplest genomes, such as *Arabidopsis*, rice, and tomato (BENNETT *et al.*, 2003), not only have least amounts of repetitive DNA sequences, but also most proteins in these plant species are encoded by relatively simple gene families (The Arabidopsis Genome Initiative, 2000). As compared to a paleopolyploid-derived genome of maize (SHEN *et al.*, 1994), the pine genome appears to have relatively more complex multigene families than simple gene families (KINLAW and NEALE, 1997). The multigene families are not only confined in pines, but have also been observed in a wide range of conifer genome (AHUJA *et al.*, 1994). The multigene families may consist of functional gene

families, along with a large array of pseudogenes (KVARNHEDEN *et al.*, 1998; GILL *et al.*, 2003; BALAKIREV and AYALA, 2003), and some of these sequences may be subject to gene silencing, loss or deletion from the genome. The multigene families in pines are not always genetically linked, but may also be dispersed in the genome. In spite of the dispersed nature of multigene families, it is paradoxical that the gene order of such gene families seems to be preserved in the pine genome. It is, however, possible that some of these duplicate genes have diverged from the original gene during the course of evolution and are highly conserved, and therefore may show a divergent gene order (BROWN *et al.*, 2001; NEALE and KRUTOVSKY, 2004; KRUTOVSKY *et al.*, 2004).

### 3. Modulation of Repetitive DNA Sequences

Earlier studies by MIKSCHE and HOTTA (1973) showed that some conifers (*Pinus resinosa*, *Pinus banksiana*, and *Picea glauca*) contained a relatively large proportion of repetitive DNA. Subsequent studies have shown that about 75% of the conifer (*Picea glauca* and *Pinus strobus*) genome consists of repetitive DNA sequences and 20–30% of the DNA contained “unique”, that is single copy, sequences (RAKE *et al.*, 1980; KRIEBEL, 1985). The high proportion (20–30%) of unique DNA sequences in the large conifer genome, according to THOMPSON and MURRAY (1981), that behave as single-copy DNA under generally used kinetic conditions are actually repeated DNA sequences that have become highly diverged in the course of evolution. These unique DNA sequences probably consist of ancient diverged repeated DNA sequences, which probably originated from retroelements (KRIEBEL, 1993; ELSIK and WILLIAMS, 2000). However, these earlier studies did not discriminate between different classes of repetitive DNA sequences.

The repetitive DNA sequences may be broadly classified into randomly and dispersed repeated sequences. There are several classes of tandemly repetitive DNA, which include telemeric, subtelomeric, and centromeric repeats, satellite DNA, and ribosomal RNA genes (rDNA) (FLAVELL, 1986). Here we shall focus on the role of rDNA genes and the dispersed repeated sequences, the transposable elements, in the evolution of genome size in conifers.

#### Ribosomal DNA genes and genome size

Ribosomal RNA genes belong to a multigene family that arose by gene duplication and recombination (OHTA, 1990). The number of rDNA repeat units varied from 500 to 40,000 copies per genome in plants (LONG and DAWID, 1980; ROGERS and BENDICH, 1987). The transcribed region of the repeat unit consists of 18s-5.8s-26s rDNA that is highly conserved, while the intergenic spacer is highly variable in sequence and length (FLAVELL, 1986). The overall length of the rDNA repeat unit varies from 6-14 kb in angiosperms. The length variation in the rDNA repeat is due to different amounts of intergenic DNA, which separates the adjacent transcription units in the tandem arrays. Ribosomal DNA repeat units are considerably longer in conifers. In loblolly pine, *Pinus taeda*, rDNA repeat unit seems to be

20-24 kb long, of which 20 kb are estimated to be the spacer region (SEDEROFF *et al.*, 1987). The length of the entire rDNA repeat unit is close to 27 kb in *Pinus radiata* (CULLIS *et al.*, 1988), and *Pinus sylvestris* (KARVONEN *et al.*, 1993), and between 32 and 40 kb in *Picea rubens* and *Picea mariana* (BOBOLA *et al.*, 1992), which is more than twice as long as in most angiosperms rDNA genes. The copy number of the rDNA repeat units varies within a species depending on the environmental conditions: ranging from 355 to 7356, and approximately 12-fold variation among individuals within a populations in *Pinus rigida* (GOVINDRAJU and CULLIS, 1982) 770 to 3850 in *Picea rubens* and *Picea mariana*, showing 3 to 6-fold within the populations of the two species (BOBOLA *et al.*, 1992). The number of rDNA copies is lower among populations of *Pinus rigida* subjected to stress (GOVINDRAJU and CULLIS, 1982).

Although rDNA copy number is affected by environmental stress factors, there seems to be an association between rDNA copy number and genome size. A recent study based on 162 species of eukaryotes, including 68 species of plants and 94 species of animals have provided the first convincing evidence for a positive relationship between the rDNA copy number and genome size in eukaryotes (PROKOPOWICH *et al.*, 2003). Since the protein coding genes remain more or less constant in the genome, they are unlikely to affect the genome size *per se*. The rDNA genes are also coding genes. Therefore, in terms of genome size increase/decrease, the rDNA genes play a role in the genome size as repetitive elements rather than coding genes (PROKOPOWICH *et al.*, 2003). However, the mechanism for the rDNA-mediated variation in genome size still remains unclear.

Another class of tandem repeats, the microsatellites or simple sequence repeats, are an abundant class of repeats in conifers. In particular GA and CA motifs are highly amplified and are major components of the conifer genome (SMITH and DEVEY, 1994; ECHT and MAY-MARQUADT, 1997; SCHMIDT *et al.*, 2000; ELSIK and WILLIAMS, 2001). In addition, a 142 bp tandem repeat DNA sequence (~20,000 copies in the genome) has been identified in a number of *Picea* species, but not in other members of Pinaceae (*Pinus radiata*, *Pseudotsuga menziesii*, and *Thuja plicata*) (BROWN *et al.*, 1998). It seems there is tendency to accumulate abundant simple sequence repeats motifs in larger genomes (HANCOCK, 2002), including conifers.

#### Transposable elements

The second major group of repetitive DNA elements consist of dispersed repeats, which are relatively more abundant than the tandemly repeated DNA sequences in the genome. The transposable elements (TEs) are divided into two classes according to the mechanism of transposition: Class 1 TEs have an RNA intermediary for transposition, while Class 2 TEs use DNA-mediated mode of transposition (FINNEGAN, 1989; FESCHOTTE *et al.*, 2002). Retrotransposons (Class 1) are ubiquitous in plants and play an important role in gene and genome evolution. The retrotransposons are members of a larger group of Retroid agents, which also include retroviruses (MCCLURE, 1999). The retroelements exhibit a large

variation in copy number in the genomes of eukaryotes (KIDWELL, 2002).

### Retrotransposons

There are two subclasses of retrotransposons in the genome of eukaryotes: with long terminal repeats (LTRs), and others that lack terminal repeats (non-LTRs). There are two distinct groups of LTR retrotransposons: the *Ty1-copia* and *Ty3-gypsy*, which are widely distributed in plants and animals, and do not encode any known proteins. The non-LTR retrotransposons include long interspersed nuclear elements (*LINES*s), short interspersed nuclear elements (*SINES*s), and miniature inverted repeat transposable elements (*MITE*s) (KUMAR and BENNETZEN, 1999; FESCHOTTE *et al.*, 2002; KIDWELL, 2002). The retrotransposons content of the genome varies among eukaryotes (Table 2). It seems that there has been explosion of retrotransposon activity in the genomes of maize (~60%) (SANMIGUEL *et al.*, 1998) and barley during the last several million years (~55%) (KUMAR and BENNETZEN, 1999; VICIENT *et al.*, 1999). It appears that in the several genera of the grass family, Gramineae (Poaceae), retrotransposons might have contributed to genome size obesity during their evolution (BENNETZEN and KELLOGG, 1997). Whether the same phenomenon has occurred in other higher organisms remains unclear at the present time. Nevertheless, it appears that in organisms with large genome size (>500 Mb), the contribution of transposable elements to genome size variation, relative to other sources of variation, is greater than organisms with relatively smaller genome size (500 Mb) (KIDWELL, 2002).

### Retrotransposons in conifer genome evolution

Retrotransposons have been isolated and characterized in several genera of conifers. The *Ty1-copia* like retrotransposons have been detected in several conifers, including *Pinus coulteri*, *Picea glauca*, *Metasequoia glyptostroboides*, *Cedrus deodara*, and *Taxus baccata* (VOYTAS *et al.*, 1992). Subsequently, retroelements of both groups *Ty1-copia* and *Ty3-gypsy* have been characterized in conifers. A *Ty1-copia*-like sequence called *TPE1*, about 4.8 kb long, has been isolated from *Pinus elliottii* (KAMM *et al.*, 1996). Genomic analysis revealed that *TPE1* carries partial reverse transcriptase and integrase gene sequences, and is highly amplified in *Pinus elliottii*. This retrotransposon was also detected in a number of other *Pinus* species (including *P. strobus*, *P. resinosa*, *P. banksiana*, and *P. palustris*), and *Picea abies* and *P. glauca*. The *TPE1* seems to be inactive as no transcription of this retroelement was detected in the conifers tested (KAMM *et al.*, 1996).

A *Ty3-gypsy*-like retrotransposon, called *IFG* (named after Institute of Forest Genetics) has been isolated in *Pinus radiata* (KOSSACK and KINLAW, 1999). The retroelement *IFG7* has also been detected in a number of *Pinus* species. *IFG7* is about 6 kb long, and there are at least 10,000 copies of this retrotransposon in the genome of *Pinus* (0.5% of the genome). Although, *IFG7* is not transcriptionally active in pines, it appears to have an extensive history in pines. *IFG7* has not been detected in the families Cupressaceae and Taxodoaceae

(KOSSACK and KINLAW, 1999). Subsequently, both *Ty1-copia*-type and *Ty3-gypsy*-type retrotransposons have been identified in spruce, *Picea glauca* and *Picea mariana* (L'HOMME *et al.*, 2000). These LTR retrotransposons are around 5-10 kb long, and comprise about 10% of the spruce genome. Both types of retroelements are inactive in the spruce genome.

Retrotransposon amplification has led to doubling of maize genome during the past 6 million years, and account for more than 60% of the maize genome (SANMIGUEL *et al.*, 1996, 1998). Although retrotransposons have been reported in pines and spruces, their share of the genome size has not been fully investigated. In spruces two types of retrotransposons (*Ty1-copia* type and *Ty3-gypsy* type) represent ~10% of the genome (L'HOMME *et al.*, 2000), which is close to all the retroelements that are present in a small genome of *Arabidopsis* (Table 2). *Ty1-copia*-like sequences are uniformly dispersed on all 12-chromosome pairs, and are highly amplified within the genome of *Pinus elliottii*, and several other pines and spruces analyzed (KAMM *et al.*, 1996). On the other hand, the *IFG* (*Ty3-gypsy* type) retrotransposons represent only 0.5% of the pine genome (KOSSACK and KINLAW, 1999). This does not necessarily represent the entire spectrum of TEs in the conifer genome, as shown by recent studies that suggest that divergent retrotransposon families have contributed to the expansion of pine genome (ELSIK and WILLIAMS, 2000; FRIESEN *et al.*, 2001; STUART-ROGERS and FLAVELL, 2001; MURRAY *et al.*, 2002). But to what extent is not known at the present time.

Variation in genome size within the populations of a species under different environmental conditions may be associated, among other things, with retrotransposon

Table 2. – Genome size, coding and duplicate genes, and transposable elements (TE) in some eukaryotes.

Species	Genome size 1C (Mb)	# coding genes (≈)	Duplicate genes (%)	TE (%)	Reference
<i>Saccharomyces cerevisiae</i>	12	5,800	30	3	KIM <i>et al.</i> 1998; WOLFE, 2001.
<i>Caenorhabditis elegans</i>	100	19,000	49	6	WATERSTON & SULSTON, 1995; RUBIN <i>et al.</i> , 2000.
<i>Arabidopsis thaliana</i>	157	25,000	65	10	The Arabidopsis Genome Initiative, 2000; BENNETT <i>et al.</i> 2003.
<i>Drosophila melanogaster</i>	175	13,600	41	15	VIEIRA <i>et al.</i> 1999; ADAMS, <i>et al.</i> , 2000; BENNETT <i>et al.</i> , 2003.
<i>Oryza sativa</i>	430	50,000		14	TURCOTTE <i>et al.</i> , 2001
<i>Zea mays</i>	2,500	50,000	35	60	GOFF <i>et al.</i> , 2002
<i>Homo sapiens</i>	3,000	30,000	38	44	SANMIGUEL <i>et al.</i> , 1998 WALBOT & PETROV, 2001 International Human Genome Sequencing Consortium, 2001.
<i>Hordeum vulgare</i>	5,000	n/a	n/a	55	VICIENT <i>et al.</i> , 1999.
<i>Picea glauca</i>	19,800	n/a	n/a	10	L'HOMME <i>et al.</i> , 2000.
<i>Pinus radiata</i>	26,500	n/a	n/a	0.5	KOSSACK & KINLAW, 1999.

\* Only a fraction of retroelements detected so far. More recent studies (STUART-ROGERS and FLAVELL, 2001; FRIESEN *et al.*, 2001) indicate that retroelements are a major component of the conifer genome. But the total proportion of such retroelements in the conifer genome is still unknown.

activity (WENDEL and WESSLER, 2000). In natural populations of wild barley, *Hordeum spontaneum*, *BARE-1* long terminal repeat (LTR) retrotransposon, abundant in the barley genome, displayed nearly three-fold variation in copy number at intraspecific level in different habitats (KALENDAR *et al.*, 2000). The genome size was relatively larger in the wild barley populations' resident of higher and drier slopes, suggesting a possible connection between adaptive genetic evolution and variation in genome size. Whether such variation in conifers is also caused by a burst of retrotransposon activity remains to be investigated.

#### 4. Increase in Intron Size

In recent years it has been suggested that there is an association between genome size and intron size in a broad range of eukaryotes (DEUTSCH and LONG, 1999; VINOGRADOV, 1999). For example, comparisons of 199 introns in 22 orthologous genes in puffer fish (*Fugus rubripes*) and humans showed that intron size was on average eight times smaller in pufferfish than humans, consistent with the ratio of their genome sizes (400/3,000 Mb) (MCLYSAGHT *et al.*, 2000). A similar correlation, based on 115 orthologous genes, was also observed in *Drosophila* species: *D. virilis* with a genome size of 350 Mb had introns significantly larger (~394 bp) as compared to *D. melanogaster* (~283 bp), which has a genome size of 175 Mb (MORIYAMA *et al.*, 1998). Among the 10 animal genera studied, humans had the longest intron size (3,400 bp) showing a general but weak relationship between intron size and genome size (DEUTSCH and LONG, 1999). However, recent studies by WENDEL *et al.* (2002) did not find an association between intron size and genome size in different species of *Gossypium* and other plant species. For example, intron size/genome size in several plant species is as follows: cotton (150 bp/1,960 Mb) (WENDEL *et al.*, 2002), Arabidopsis (168/157 Mb) (The Arabidopsis Genome Initiative, 2000), rice (356/430 Mb) (YU *et al.*, 2000), pine (350/20,000) (NEALE, unpublished), and human (3,400/3,000Mb) (International Human Genome Sequencing Consortium, 2001). Therefore, intron-genome size relationship seems weak in the plant taxa.

#### Concluding Statement

Why do conifers have such large genome sizes? Does the large genome size have any relationship to long life span or functional genomics in conifers? Does the large genome in conifers provide a reservoir of adequate genetic material for responding to adverse changes in environment during their long life cycles, in some cases extending to more than 2,000 to 4,000 years (*Sequoia sempervirens*, *Fitzroya cupressoides*, and *Pinus longaeva*)? We do not know satisfactory answers to these questions. At the present time, we can only partially account for the acquisition of large genome size in conifers. It would appear that the large conifer genome may predominantly consists of non-coding repetitive DNA, particularly due to retrotransposon amplification, and the presence of a larger rDNA repeat units, large number of multigene families, larger nuclear volume,

and perhaps larger genes, and abundant pseudogenes. The genome size in conifers has undergone both increases and decreases during its evolution to account for the variation in genome size in different genera, families and species. In the final analysis increase or decrease in genome size would be determined by the global action of mutations, environmental stresses, and natural selection acting on individual, but more likely on several genomic components simultaneously, during the genome size evolution of conifers. Whether ancient polyploidy has tinkered with the genome size in conifers remains enigmatic? But future research in molecular biology and chromosome painting may shed light on this important question.

#### Acknowledgements

We thank the Institute of Forest Genetics, USDA Forest Service, at Placerville and Davis, and Department of Plant Sciences, University of California, Davis, for research and library facilities for this review. We thank two anonymous reviewers for critical reading of the manuscript and constructive suggestions.

#### References

- ADAMS, M. C., S. E. CELNIKER and R. A. HOLT *et al.* (2000): The genome sequence of *Drosophila melanogaster*. *Science* **287**: 2185–2195.
- AHUJA, M. R. (2001): Recent advances in molecular genetics of forest trees. *Euphytica* **121**: 173–195.
- AHUJA, M. R. (2005): Polyploidy in gymnosperms: Revisited. *Silvae Genet.* **54**: 59–69.
- AHUJA, M. R., M. E. DEVEY, A. T. GROVER, K. D. JERMSTAD and D. B. NEALE (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.
- AHUJA, M. R. and D. B. NEALE (2002): Origins of polyploidy in coast redwood (*Sequoia sempervirens* (D. Don) Endl.) and relationship of coast redwood to other genera of Taxodiaceae. *Silvae Genet.* **51**: 93–100.
- AUKLAND, L. D., J. S. JOHNSTON, H. J. PRICE and F. E. BRIDGEWATER (2001): Stability of nuclear DNA content among divergent and isolated populations of Fraser fir. *Can. J. Bot.* **79**: 1375–1378.
- BALAKIREV, E. and F. J. AYALA (2003): Pseudogenes: Are they “junk” or functional DNA? *Annu. Rev. Genet.* **37**: 123–151.
- BENNETT, M. D. and I. J. LEITCH (2003): Angiosperm DNA C-values database. <http://www.rbgbkew.org.uk/cval/homepage.html>.
- BENNETT, M. D. and I. J. LEITCH (2005): Plant genome size research: A field in focus. *Ann. Bot.* **95**: 1–6.
- BENNETT, M. D. and J. B. SMITH (1991): Nuclear DNA amounts in angiosperms. *Phil. Trans. R. Soc. Lond. B.* **334**: 309–345.
- BENNETT, M. D., I. J. LEITCH and L. HANSON (1998): DNA amounts in two samples of angiosperm weeds. *Ann. Bot.* **82** (Supplement A): 121–134.
- BENNETT, M. D., I. J. LEITCH, H. J. PRICE and J. S. JOHNSTON (2003): Comparisons with *Caenorhabditis* (~100 Mb) and *Drosophila* (~175 Mb) using flow cytometry show genome size in Arabidopsis to be ~157 Mb and thus ~25 % larger than the Arabidopsis Genome Initiative estimate of ~125 MB, *Ann. Bot.* **91**: 547–557.

- BENNETZEN, J. L. (2002): Mechanisms and rates of genome expansion and contraction in flowering plants. *Genetica* **115**: 29–36.
- BENNETZEN, J. L. and E. A. KELLOGG (1997): Do plants have a one-way ticket to genomic obesity? *Plant Cell* **9**: 1509–1514.
- BENNETZEN, J. L., J. MA and K. M. DEVOS (2005): Mechanisms of recent genome size variation in flowering plants. *Ann. Bot.* **95**: 127–132.
- BERLYN, G. P., J. L. ROYTE and A. O. ANOROU (1990): Cytophotometric differentiation of high elevation spruces: physiological and ecological implications. *Stain Tech.* **65**: 1–14.
- BLANC, G. and K. H. WOLFE (2004): Widespread paleopolyploidy in model plant species inferred from age distribution of duplicate genes. *Plant Cell* **16**: 1667–1678.
- BOBOLA, M. S., D. E. SMITH and A. S. KLEIN (1992): Five major nuclear ribosomal repeats represent a large and variable fraction of the genomic DNA of *Picea rubens* and *P. mariana*. *Mol. Biol. Evol.* **9**: 125–137.
- BOWE, L. M., G. COAT and C. W. DEPAMPHILIS (2000): Phylogeny of seed plants based on all three genomic compartments: Extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. *Proc. Nat. Acad. Sci. USA* **97**: 4092–4097.
- BROWN, G. R., C. H. NEWTON and J. E. CARLSON (1998): Organization and distribution of a *Sau3A* tandem repeated DNA sequence in *Picea* (Pinaceae) species. *Genome* **41**: 560–565.
- BROWN, G. R., E. E. KADEL and D. I. BASSONI *et al.* (2001): Anchored reference loci in loblolly pine (*Pinus taeda* L.) for integrating pine genomics. *Genetics* **159**: 799–809.
- BURLEY, J. (1965): Karyotype analysis of Sitka spruce, *Picea sitchensis* (Bong.) Carr. *Silvae Genet.* **14**: 127–132.
- CAVALLINI, A., I. NATALI, G. CIONINI and D. GENNAI (1993): Nuclear DNA variability within *Pisum sativum* (Leguminosae): nucleotypic effects on plant growth. *Heredity* **70**: 561–565.
- CAUSSE, M. A., T. M. FULTON and Y. G. CHO *et al.* (1994): Saturated molecular map of rice genome based on an interspecific backcross population. *Genetics* **138**: 1251–1274.
- CHAW, S.-M., C. L. PARKINSON, Y. CHENG, T. M. VINCENT and J. D. PALMER (2000): Seed plant phylogeny inferred from all three plant genomes: Morphology of extant gymnosperms and origin of Gnetales from conifers. *Proc. Nat. Acad. Sci. USA* **97**: 4086–4091.
- CULLIS, C. A., G. P. GRIESSEN, S. W. GORMAN and R. D. TEASDALE (1988): The 25S, 18S, and 5S ribosomal RNA genes from *Pinus radiata* D. Don. *In: Molecular Genetics of Forest Trees. Proc. 2<sup>nd</sup> Workshop IUFRO Working Party s2.04.06.* CHELIAK, W. M. and YAPA, A. C. (Eds). Canadian Forestry Service PNFI Inf. Rep. PI-X-80, pp. 34–40.
- DELEVORYAS, T. (1980): Polyploidy in gymnosperms. *In: Polyploidy – Biological Relevance.* LEWIS, W. H. (Ed). Plenum Press, New York, pp. 215–218.
- DE LUC, A., R. A. ADAMS and M. ZHANG (1999): Using random amplification of polymorphic DNA for taxonomic evaluation of Pfitzer *Juniperus*. *HortScience* **34**: 1123–1125.
- DEVEY, M. E., T. A. FIDDLER, B.-H. LIU, S. J. KNAPP and D. B. NEALE (1994): An RFLP linkage map for loblolly pine based on three generation outbred pedigree. *Theor. Appl. Genet.* **88**: 273–278.
- DEUTSCH, M. and M. LONG (1999): Intron-exon structure of eukaryotic model organisms. *Nucleic Acid Res.* **27**: 3219–3228.
- DHILLON, S. S. (1987): DNA in tree species. *In: Cell and Tissue Culture in Forestry. Vol. 1.* BONGA, J. M. and DURZAN, D. J. (Eds). Martinus Nijhoff Publishers, Dordrecht, pp. 298–313.
- DHIR, N. K. and J. P. MIKSCH (1974): Intraspecific variation of nuclear DNA content in *Pinus resinosa* Ait. *Can. J. Genet. Cytol.* **16**: 77–83.
- DIETRICH, F. S., S. VOEGELI, S. BRACHAT *et al.* (2004): The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. *Science* **304**: 304–307.
- DREWRY, A. (1988): The G-banded karyotype of *Pinus resinosa* Ait. *Silvae Genet.* **37**: 218–221.
- ECHT, C. S. and P. MAY-MARQUARDT (1997): Survey of microsatellite DNA in pine. *Genome* **40**: 9–17.
- ELSIK, C. G. and C. G. WILLIAMS (2000): Retroelements contribute to the excess of low-copy number DNA in pine. *Mol. Genet. Genomics* **264**: 47–55.
- ELSIK, C. G. and WILLIAMS, C. G. (2001): Families of clustered microsatellites in a conifer genome. *Mol. Genet. Genomics* **265**: 535–542.
- FARJON, A. (1998): World Checklist and Bibliography of Conifers. The Royal Botanic Garden, Kew.
- FESCHOTTE, C., N. JIANG and S. R. WESSLER (2002): Plant transposable elements: where genetics meets genomics. *Nature Rev. Genet.* **3**: 329–341.
- FINNEGAN, D. J. (1989): Eukaryotic transposable elements and genome evolution. *Trends Genet.* **5**: 103–107.
- FISHER, R. A. (1935): The sheltering of lethals. *Am. Nat.* **69**: 446–455.
- FLAVELL, R. (1986): The structure and control of expression of ribosomal RNA genes. *Oxford Surv. Plant Mol. Biol.* **3**: 251–274.
- FORCE, A., M. LYNCH, F. B. PICKETT, A. AMORES, Y. YAN and J. POSTLETHWAIT (1999): Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**: 1531–1545.
- FRIESEN, N., A. BRANDES and J. S. HESLOP-HARRISON (2001): Diversity, origin and distribution of retrotransposons (*gypsy* and *copial*) in conifers. *Mol. Biol. Evol.* **18**: 1176–1188.
- GAUT, B. S. (2001): Patterns of chromosomal duplication in maize and their implications for comparative maps of grasses. *Genome Res.* **11**: 55–66.
- GAUT, B. S. and J. F. DOEBLEY (1997): DNA sequence evidence for the segmental allotetraploid origin of maize. *Proc. Natl. Acad. Sci. USA* **94**: 6809–6814.
- GIBSON, T. J. and J. SPRING (2000): Evidence in favor of ancient octoploidy in the vertebrate genome. *Biochem Soc. Tans.* **28**: 259–264.
- GILL, G. P., G. R. BROWN and D. B. NEALE (2003): A sequence mutation in the cinamyl alcohol dehydrogenase gene associated with altered lignification in loblolly pine. *Plant Biotech. J.* **1**: 253–258.
- GOFF, S. A., D. RICKE and T.-H. LAN *et al.* (2002): A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* **296**: 92–100.
- GOVINDRAJU, D. R. and C. A. CULLIS (1992): Ribosomal DNA variation among populations of *Pinus rigida* Mill. (patch pine) ecosystem. I. Distribution of copy numbers. *Heredity* **69**: 133–140.
- GRAHAM, M. J., C. D. NICKELL and A. L. RAYBURN (1994): Relationship between genome size and maturity group in soybean. *Theor. Appl. Genet.* **88**: 429–432.
- GRANT, V. (1981): *Plant Speciation.* (Second Edition). Columbia University Press, New York.

- GRATTAPAGLIA, D. and H. D. BRADSHAW (1994): Nuclear DNA amounts of commercially important *Eucalyptus* species. *Can. J. For. Res.* **24**: 1074–1078.
- GREGORY, T. R. (2001): Animal genome size database. <http://www.genomesiz.com>.
- GREGORY, T. R. (2005): The C-value enigma in plants and animals: A review of parallels and an appeal for partnership. *Ann. Bot.* **95**: 133–146.
- GROTKOPP, E., M. REJÁNEK, M. J. SANDERSON and T. L. ROST (2004): Evolution of genome size in pines (*Pinus*) and its life-history correlates: supertree analyses. *Evolution* **58**: 1705–1729.
- GUGERLI, F., C. SPERISON and U. BÜCHLER *et al.* (2001): The evolutionary split of Pinaceae from other conifers: Evidence from an intron loss and a multigene phylogeny. *Mol. Phylogenet. Evol.* **21**: 167–175.
- HAIR, J. B. (1968): The chromosomes of the Cupressaceae. I. Tetraclineae and Actinostrobaeae (Callitroideae). *New Zealand J. Bot.* **6**: 277–284.
- HALDANE, J. B. S. (1933): The part played by recurrent mutations in evolution. *Am. Nat.* **67**: 5–9.
- HANCOCK, J. M. (2002): Genome size and accumulation of simple sequence repeats: Implications of new data from genome sequencing projects. *Genetica* **115**: 93–103.
- HIZUME, M., T. KONDO, F. SHIBATA and R. ISHIZUKU (2001): Flow cytometric determination of genome size in the Taxodiaceae, Cupressaceae *sensu stricto* and Sciadopityaceae. *Cytologia* **66**: 307–311.
- HIZUME, M., F. SHIBATA, Y. MATSUSAKI and Z. GARAJOVA (2002): Chromosome identification and comparative karyotype analysis of four Pinus species. *Theor. Appl. Genet.* **105**: 491–497.
- HUGHES, A. L. (1999): Phylogenies of developmentally important proteins do not support the hypothesis of two rounds of duplication early in vertebrate history. *J. Mol. Biol.* **48**: 565–578.
- International Human Genome Sequencing Consortium. (2001): Initial sequencing and analysis of human genome. *Nature* **409**: 860–921.
- JOYNER, K. L., X.-R. WANG, J. S. JOHNSTON, H. J. PRICE and C. G. WILLIAMS (2001): DNA content for Asian pines parallels new world relatives. *Can. J. Bot.* **79**: 192–196.
- KALENDAR, R., J. TANKKANEN, S. IMMONEN, E. NEVO and A. H. SCHULMAN (2000): Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. *Proc. Natl. Acad. Sci. USA* **97**: 6603–6607.
- 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. *elliottii* and other gymnosperms. *Proc. Natl. Acad. Sci. USA* **93**: 2708–2713.
- KARVONEN, P., M. KARJALAINEN and O. SOVOLAINEN (1993): Ribosomal RNA genes in Scots pine (*Pinus sylvestris* L.): chromosomal organization and structure. *Genetica* **88**: 59–68.
- KAVARNHEDEN, A., V. A. ALBERT and P. ENGSTROM (1998): Molecular evolution of *cdc2* pseudogene in spruce (*Picea*). *Plant Mol. Biol.* **36**: 767–774.
- KHOSHOO, T. N. (1959): Polyploidy in gymnosperms. *Evolution* **13**: 24–39.
- KHOSHOO, T. N. (1961): Chromosome numbers in gymnosperms. *Silvae Genet.* **10**: 1–9.
- KIDWELL, M. G. (2002): Transposable elements and evolution of genome size in eukaryotes. *Genetica* **115**: 49–63.
- KIM, J. M., S. VANGURI, J. D. BOEKE and D. F. VOYTAS (1998): Transposable elements and genome organization: A comprehensive survey of retrotransposons revealed by the complete Saccharomyces cerevisiae genome sequence. *Genome Res.* **8**: 464–478.
- KINLAW, C. S., D. E. HARRY and R. R. SEDEROFF (1990): Isolation and characterization of alcohol dehydrogenase cDNA from *Pinus radiata*. *Can. J. For. Res.* **20**: 1343–1350.
- KINLAW, C. S. and D. B. NEALE (1997): Complex gene families in pine genomes. *Trends Plant Sci.* **2**: 356–359.
- KNIGHT, C. A., N. A. MOLINARI and D. A. PETROV (2005): The large genome constraint hypothesis: Evolution, ecology and phenotype. *Ann. Bot.* **95**: 177–190.
- 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.
- KRIEBEL, H. B. (1985): DNA sequence components of the *Pinus strobus* nuclear genome. *Can. J. For. Res.* **15**: 1–4.
- KRIEBEL, H. B. (1993): Molecular structure of forest trees. *In: Clonal Forestry I. Genetics and Biotechnology.* AHUJA, M. R. and LIBBY, W. J. (Eds). Springer Verlag, Berlin, pp. 224–240.
- KRUTOVSKY, K. V., M. TROGGIO, G. R. BROWN, K. D. JERMSTAD and D. B. NEALE (2004): Comparative mapping in Pinaceae. *Genetics* **168**: 447–461.
- KUMAR, A. and J. L. BENNETZEN (1999): Plant retrotransposons. *Annu. Rev. Genet.* **33**: 479–532.
- EL-LAKANY, M. H. and O. SZIKLAI (1971): Intraspecific variation in nuclear characteristics of Douglas-fir. *Advan. Front. Plant Sci.* **28**: 363–378.
- LEITCH, I. J. and M. D. BENNETT (2002): New insights into patterns of nuclear genome size evolution in plants. *Current Genomics* **3**: 551–562.
- LEITCH, I. J. and M. D. BENNETT (2004): Genome downsizing in polyploid plants. *Biol. J. Linnean Soc.* **82**: 651–663.
- LEITCH, I. J., L. HANSON, M. WINFIELD, J. PARKER and M. D. BENNETT (2001): Nuclear DNA C-values complete familial representation in gymnosperms. *Ann. Bot.* **88**: 843–849.
- LEITCH, I. J., D. E. SOLTIS, P. S. SOLTIS and M. D. BENNETT (2005): Evolution of DNA amounts across land plants (Embryophyta). *Ann. Bot.* **95**: 207–217.
- L'HOMME, Y., A. SÉGUIN and F. M. TREMBLAY (2000): Different classes of retrotransposons in coniferous spruce species. *Genome* **43**: 1084–1089.
- LONG, E. O. and I. B. DAWID (1980): Repeated genes in eukaryotes. *Annu. Rev. Biochem.* **49**: 727–764.
- LYNCH, M. (2002): Gene duplication and evolution. *Science* **297**: 945–947.
- LYNCH, M. and J. S. CONERY (2000): The evolutionary fate and consequences of duplicate genes. *Science* **290**: 1151–1155.
- MAKALOWSKI, W. (2001): Are we polyploids? A brief history of one hypothesis. *Genome Research* **11**: 667–670.
- MARTIN, A. P. (1999): Increasing genomic complexity by gene duplication and origin of vertebrates. *Am. Nat.* **154**: 111–128.
- MARTIN, A. (2001): Is tetralogy true? Lack of support for the 'one-to-four' rule. *Mol. Biol. Evol.* **18**: 89–93.
- MASTERTSON, J. (1994): Stomatal size in fossil plants: Evidence for polyploidy in majority of angiosperms. *Science* **264**: 421–423.
- MCCLURE, M. A. (1999): The retroid agents: disease, function and evolution. *In: Origin and Evolution of Viruses.* DOMINGO, E., WEBSTER, R. and HOLLAND, J. (Eds). Academic Press, London, pp. 163–195.

- McLYSAGHT, A., L. ENRIGHT, L. SKRABANEK and K. H. WOLFE (2000): Estimation of synteny conservation and genome compaction between pufferfish (*Fugu*) and human. *Yeast* **17**: 22–36.
- McLYSAGHT, A., K. HOKAMP and K. H. WOLFE (2002): Extensive genomic duplication during early chordate evolution. *Nature Genetics* **31**: 200–204.
- MERGEN, F. and B. A. THIELGES (1967): Intraspecific variation in nuclear volume in four conifers. *Evolution* **21**: 720–724.
- MIKSCH, J. P. (1968): Quantitative study of intraspecific variation of DNA per cell in *Picea glauca* and *Pinus banksiana*. *Can. J. Genet. Cytol.* **10**: 590–600.
- MIKSCH, J. P. (1971): Intraspecific variation of DNA per cell between *Picea sitchensis* (Bong.) Carr. provenances. *Chromosoma* **32**: 343–352.
- MIKSCH, J. P. and Y. HOTTA (1973): DNA base composition and repetitive DNA in several conifers. *Chromosoma* **41**: 29–36.
- MILLAR, C. I. (1998): Early evolution of pines. In: *Ecology and Biogeography of Pinus*. RICHARDSON, D. M. (Ed). Cambridge University Press, Cambridge, pp. 69–91.
- MILLER, C. N. (1977): Mesozoic conifers. *Bot. Rev.* **43**: 217–280.
- MIROV, N. T. (1967): *The Genus Pinus*. Ronald Press, New York.
- MORIYAMA, E. N., D. A. PETROV and D. L. HARTL (1998): Genome size and intron size in *Drosophila*. *Mol. Biol. Evol.* **15**: 770–773.
- MURRAY, B. G. (1998): Nuclear DNA amounts in gymnosperms. *Ann. Bot.* **82** (Supplement A): 3–15.
- MURRAY, B. G., N. FRIESEN and J. S. HESLOP-HARRISON (2002): Molecular cytogenetic analysis of *Podocarpus* and comparison with other gymnosperm species. *Ann. Bot.* **89**: 483–489.
- NEALE, D. B. and K. V. KRUTOVSKY (2004): Comparative genome mapping in trees: The group of conifers. In: *Biotechnology in Agriculture and Forestry*. Vol. 55. *Molecular Marker Systems*. LÖRZ, H. and WENZEL, G. (Eds). Springer Verlag, Berlin, pp. 267–277.
- NEWTON, R. J., M. G. MESSINA, H. J. PRICE and I. WAKAMIYA-NOBORI (1999): DNA content, water relations, and environmental stress in gymnosperms. In: *Handbook of Plant and Crop Stress*. Second Edition. PRESSARAKLI, M. (Ed). Marcel Dekker, New York, pp. 659–673.
- OBERMAYER, R., I. J. LEITCH, L. HANSON and M. D. BENNETT (2002): Nuclear DNA C-values in 30 species double the estimated familial representation in pteridophytes. *Ann. Bot.* **90**: 209–217.
- OHNO, S. (1970): *Evolution by Gene Duplication*. Springer Verlag, Berlin.
- OHRI, D. (1998): Genome size variation and plant systematic. *Ann. Bot.* **82** (Supplement A): 75–83.
- OHRI, D. and M. R. AHUJA (1990): Giemsa C-banded karyotype in *Quercus* L. (oak). *Silvae Genet.* **39**: 216–219.
- OHRI, D., and T. N. KHOSHOO (1986): Genome size in gymnosperms. *Pl. Syst. Evol.* **153**: 119–132.
- OHTA, T. (1990): How gene families evolve. *Theor. Pop. Biol.* **37**: 213–219.
- OTTO, S. P. and WHITTON, J. (2000): Polyploidy incidence and evolution. *Annu. Rev. Genet.* **34**: 401–437.
- PATERSON, A. H., J. E. BOWERS and B. A. CHAPMAN (2004): Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc. Natl. Acad. Sci. USA* **101**: 903–998.
- PERRY, D. L. and G. R. FURNIER (1996): *Pinus banksiana* has at least seven expressed alcohol dehydrogenase genes in two linked groups. *Proc. Natl. Acad. Sci. USA* **93**: 13020–13023.
- PETROV, D. A. (2001): Evolution of genome size: New approaches to an old problem. *Trends Genet.* **17**: 23–28.
- PREMOLI, A. C., T. KITZBERGER and T. T. VEBELEN (2000): Conservation genetics of the endangered conifer *Fitzroya cupressoides* in Chile and Argentina. *Conservation Genet.* **1**: 57–66.
- PRICE, H. J. (1988): DNA content variation among higher plants. *Ann. Missouri Bot. Garden* **75**: 1248–1257.
- PRINCE, V. E. and F. B. PICKETT (2002): Splitting pairs: The diverging fates of duplicate genes. *Nature Rev. Genet.* **3**: 827–837.
- PROKOPOWICH, C. D., T. R. GREGORY and T. J. CREASE (2003): The correlation between rDNA copy number and genome size in eukaryotes. *Genome* **46**: 48–50.
- RAKE, A. V., J. P. MIKSCH, R. B. HALL and K. M. HANSEN (1980): DNA reassociation kinetics of four conifers. *Can. J. Genet. Cytol.* **22**: 69–79.
- RAMSEY, J. and D. W. SCHEMSKE (2002): Neopolyploidy in flowering plants. *Annu. Rev. Ecol. Syst.* **33**: 589–639.
- RAYBURN, A. L., H. J. PRICE, J. D. SMITH and J. R. GOLD (1985): C-band heterochromatin and DNA content in *Zea mays*. *Am. J. Bot.* **72**: 1610–1617.
- ROGERS, S. O. and A. J. BENDICH (1987): Ribosomal RNA genes in plants: variability in copy number and in the intergenic spacers. *Plant Mol. Biol.* **9**: 509–520.
- RUBIN, G. M., M. D. YANDELL and J. R. WORTMAN *et al.* (2000): Comparative genomics of the eukaryotes. *Science* **287**: 2204–2215.
- SANKOFF, D. (2001): Gene and genome duplication. *Curr. Opin. Genet. Dev.* **11**: 681–684.
- SANMIGUEL, P., A. TIKHONOV and Y.-K. JIN *et al.* (1996): Nested retrotransposons in the intergenic regions of the maize genome. *Science* **274**: 765–768.
- SANMIGUEL, P. and J. L. BENNETZEN (1998): Evidence that a recent increase in maize genome size was caused by the massive amplification of intergenic retrotransposons. *Ann. Bot.* **82** (Supplement A): 37–44.
- SAX, K. and H. J. SAX (1933): Chromosome number and morphology in the conifers. *J. Arnold Arboretum* **14**: 356–375.
- SAYLOR, L. C. and H. A. SIMONS (1970): Karyology of *Sequoia sempervirens*; karyotype and accessory chromosomes. *Cytologia* **35**: 294–303.
- SCHLARBAUM, S. E. and T. TSUCHIYA (1984): A chromosome study of coast redwood, *Sequoia sempervirens* (D. Don.) Endl.). *Silvae Genet.* **33**: 56–62.
- 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.
- SEDEROFF, R. R., A.-M. STOMP and B. GWYNN *et al.* (1987): Application of DNA recombinant techniques in pines: A molecular approach to genetic engineering in forestry. In: *Cell and Tissue Culture in Forestry*. Vol. 1. BONGA, J. M. and DURZAN, D. J. (Eds). Martinus Nijhoff Publishers, Dordrecht, pp. 314–329.
- SEWELL, M. M., B. K. SHERMAN and D. B. NEALE (1999): A consensus map for loblolly pine (*Pinus taeda* L.). I. Construction and integration of individual linkage maps from two outbred three-generation pedigrees. *Genetics* **151**: 321–330.
- SHEN, B., N. CARNEIRO and L. TORRES-JEREZ *et al.* (1994): Partial sequencing and mapping of clones from two maize cDNA libraries. *Plant Mol. Biol.* **26**: 1085–1101.

- SILJAK-YAKOVIEV, S., M. CERBAH and J. COULAUD *et al.* (2002): Nuclear DNA content, base composition, heterochromatin and rDNA in *Picea amara* and *Picea abies*. *Theor. Appl. Genet.* **104**: 505–512.
- SKRABANEK, L. and K. H. WOLFE (1998): Eukaryote genome duplication – where’s the evidence? *Curr. Opin. Genet. Dev.* **8**: 694–700.
- SMITH, D. N. and M. E. DEVEY (1994): Occurrence and inheritance of microsatellite loci in *Pinus radiata*. *Genome* **37**: 977–983.
- SOLTIS, D. E. and P. S. SOLTIS (1999): Polyploidy: recurrent formation and genome evolution. *Trends Ecol. Evol.* **14**: 348–352.
- SOLTIS, D. E., P. S. SOLTIS, M. D. BENNETT and I. J. LEITCH (2003a): Evolution of genome size in angiosperms. *Am. J. Bot.* **90**: 1596–1603.
- SOLTIS, D. E., P. S. SOLTIS and J. TATE (2003b): Advances in the study of polyploidy since plant speciation. *New Phytologist* **161**: 173–191.
- STUART-ROGERS, C. and A. J. FLAVELL (2001): The evolution of Ty1- *copia* group retrotransposons in gymnosperms. *Mol. Biol. Evol.* **18**: 155–163.
- STEBBINS, G. L. (1948): The chromosomes and relationship of *Metasequoia* and *Sequoia*. *Science* **108**: 95–98.
- STEBBINS, G. L. (1950): *Variation and Evolution in Plants*. Columbia University Press, New York.
- STEBBINS, G. L. (1966): Chromosomal variation and evolution. *Science* **152**: 1463–1469.
- STEWART, W. N. and G. W. ROTHWELL (1993): *Paleobotany and the Evolution of Plants*. Second Edition. Cambridge University Press, Cambridge.
- The Arabidopsis Genome Initiative (2000): Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**: 796–815.
- THOMAS, C. A. (1970): The genetic organization of chromosomes. *Annu. Rev. Genet.* **5**: 237–256.
- THOMSON, W. F. and M. G. MURRAY (1981): The nuclear genome: structure and function. *In: The Histochemistry of Plants*. Vol. 6. STUMPF, P. K. and CONN, E. E. (Eds). Academic Press, London, pp. 1–81.
- TURCOTTE, K., S. SRINIVASAN and T. BUREAU (2001): Survey of transposable elements from rice genome sequences. *Plant J.* **25**: 169–179.
- VAN DE PEER, Y., J. S. TATLOR and A. MEYER (2003): Are all fishes ancient polyploids? *J. Structural and Functional Genomics* **2**: 65–73.
- VANDERPOEPE, K., C. SIMILLION and Y. VAN DE PEER (2003): Evidence that rice and other cereals are ancient aneuploids. *Plant Cell* **15**: 2192–2202.
- VICENT, C. M., A. SUONIEMI, ANAMTHAWAT-JÓNSSON, J. TANSKANEN, A. BEHARAV, E. NEVO and A. H. SCHULMAN (1999): Retrotransposon *BARE-1* and its role in genome evolution in the genus *Hordium*. *Plant Cell* **11**: 1769–1784.
- VOYTAS, D. F., M. P. CUMMINGS, A. KONIECZNY, F. M. ASUBEL and S. RODERMEL (1992): *Copia*-like retrotransposons are ubiquitous among plants. *Proc. Natl. Acad. Sci. USA* **89**: 7124–7128.
- VIEIRA, C., D. LEPETIT, S. DUMONT and C. BIEMONT (1999): Make up of transposable elements following *Drosophila simulans* worldwide colonization. *Mol. Biol. Evol.* **16**: 1251–1255.
- VINOGRADOV, A. E. (1999): Intron-genome size relationship on a large evolutionary scale. *J. Mol. Evol.* **49**: 376–384.
- 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.
- WALBOT, V. and D. A. PETROV (2000): Gene galaxies in the maize genome. *Proc. Natl. Acad. Sci. USA* **98**: 8163–8164.
- WALSH, B. (2003): Population-genetic models of the fates of duplicate genes. *Genetica* **118**: 279–294.
- WATERSTON, R. and J. SULSTON (1995): The genome of *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* **92**: 10836–10840.
- WENDEL, J. F. (2000): Genome evolution in polyploids. *Plant Mol. Biol.* **42**: 225–249.
- WENDEL, J. F., R. C. CRONN, I. ALVAREZ, B. LIU, R. L. SMALL and D. S. SENCHINA (2002): Intron size and genome size in plants. *Mol. Biol. Evol.* **19**: 2346–2352.
- WENDEL, J. F. and S. R. WESSLER (2000): Retrotransposon-mediated genome evolution on a local ecological scale. *Proc. Natl. Acad. Sci. USA* **97**: 6250–6252.
- WOLFE, K. H. (2001): Yesterday’s polyploids and the mystery of diploidization. *Nature Rev. Genet.* **2**: 333–341.
- WRIGHT, J. W. (1976): *Introduction to Forest Genetics*. Academic Press, New York.
- YU, Z., S. J. WRIGHT and T. E. BUREAU (2000): Mutator-like elements in *Arabidopsis thaliana*. Structure, diversity and evolution. *Genetics* **156**: 2019–2031.
- ZHANG, J. (2003): Evolution by gene duplication: An update. *Trends Ecol. Evol.* **18**: 292–298.