Variation and Evolution of Genome Size in Gymnosperms

Deepak Ohri

Research Cell, Amity University Uttar Pradesh, Lucknow Campus, Malhaur, (Near Railway Station), P.O. Chinchhat, Lucknow-226028, U.P., India, E-mail: ohri_deepak@rediffmail.com

Abstract

Gymnosperms show a significantly higher mean ($1C=18.16$, $1Cx=16.80$) and a narrow range (16.89-fold) of genome sizes as compared with angiosperms. Among the 12 families the largest ranges of $1C$ values is shown by Ephedraceae (4.73-fold) and Cupressaceae (4.45-fold) which are partly due to polyploidy as $1Cx$ values vary 2.41 and 1.37-fold respectively. In rest of the families which have only diploid taxa the range of $1C$ values is from 1.18-fold (Cycadaeae) to 4.36-fold (Podocarpaceae). The question is how gymnosperms acquired such big genome sizes despite the rarity of recent instances of polyploidy. A general survey of different families and genera shows that gymnosperms have experienced both increase and decrease in their genome size during evolution. Various genomic components which have accounted for these large genomes have been discussed. The major contributors are the transposable elements particularly LTR-retrotransposons comprising of *Ty3gypsy*, *Ty1copia* and *gymny* superfamilies which are most widespread. The genomes of gymnosperms have been acquiring diverse LTR-RTs in their long evolution in the absence of any efficient mechanism of their elimination. The epigenetic machinery which silences these large tracts of repeat sequences into the stretches of heterochromatin and the adaptive value of these silenced repeat sequences need further investigation.

Keywords: *genome size evolution, gymnosperms, cycads, conifers, polyploidy, duplicate genes, tandem repeats, retrotransposons, introns*

Introduction

Gymnosperms are the most ancient spermatophytes bearing naked seeds unprotected by ovary, and are ancestral to angiosperms (Kenrick and Crane 1997). The first evidence of gymnosperm fossils dates back to Carboniferous (c.290 million years ago (mya)), while angiosperms appear on Lower-Cretaceous (c. 130 mya) which subsequently achieved ecological dominance in mid-Cretaceous (c.100mya) (Bateman et al. 2006, Friis et al. 2010) and presently are represented by c.352,000 species (Paton et al. 2008). Extant gymnosperms form a much smaller group, and have been classified into four subclasses Cycadidae, Ginkgoideae, Gnetales and Pinidae under the class Equisetopsida (Chase and Reveal 2009) comprising 12 families and 83 genera (Christenhusz et al. 2011) and 88 genera with 1104 recognized species according to The Plant List (www.theplantlist.org). There is however, still some controversy about monophyletic or paraphyletic origin of the group (Hill 2005). Besides, the species diversity is very unevenly distributed among different families e.g. Pinaceae and Zamiaceae are represented by more than 200 species and Cupressaceae, Podocarpaceae, Cycadaceae, Ephedraceae, Gnetales, Taxaceae, Araucariaceae by 166, 191, 92, 70, 41, 31, 39 species respectively while Welwitschiaceae, Sciadopityaceae and Ginkgoaceae are monospecific (The Plant List). However, despite their very restricted species diversity (only 0.4 % that of angiosperms), gymnosperms are widely distributed in both northern and southern Hemispheres occupying 39 % of the total forest area (Armenise et al. 2012, Wang and Ran 2014) in nearly all biogeographical regions (Fragniere et al. 2015). In fact only few angiosperm families can compete with large conifer families in terms of productivity or biomass in certain ecological niches (Brodribb et al. 2012). Therefore being a small group of immense evolutionary and economic importance all the 12 families (Leitch et al. 2001) and 514 taxa included in 471 species (42.7 % of total species) are

represented for their genome sizes making it the most well studied group in this respect. Except for Gnetaceae the gymnosperm species generally show large genomes despite the rarity of recent instances of polyploidy (Khoshoo 1959, Ahuja 2005, Ohri 2021a). The present review presents the range (Tables 1 and 2, Fig.1) and possible adaptive value of genome size variation in different taxa, and describes various DNA components which account for the genomic architecture of this group of plants.

Table 1
Range and mean of 1C and 1Cx values in 12 families of gymnosperms

Sr. No.	Family	Range of holoploid genome size (1C-value, pg)	Mean	Ratio (max./min.)	Range of monoploid genome size (1Cx-value, pg)	Mean	Ratio (max./min.)
1.	Cycadaceae	12.55-14.76	13.42	1.18	12.55-14.76	13.42	1.18
2.	Zamiaceae	16.85-32.35	22.67	1.92	16.85-32.35	22.67	1.92
3.	Ginkgoaceae	11.75	11.75	-	11.75	11.75	-
4.	Welwitschiaceae	7.20	7.20	-	7.20	7.20	-
5.	Gnetaceae	2.27-4.20	3.60	1.85	2.27-4.20	3.60	1.85
6.	Ephedraceae	8.10-38.34	16.84	4.73	7.58-10.38	9.24	1.37
7.	Pinaceae	12.30-37.68	24.70	3.06	12.30-37.68	24.70	3.06
8.	Araucariaceae	10.82-22.75	15.86	2.10	10.82-22.75	15.86	2.10
9.	Podocarpaceae	4.15-18.10	8.36	4.36	4.15-18.10	8.36	4.36
10.	Sciadopityaceae	20.80	20.80	-	20.80	20.80	-
11.	Cupressaceae	8.27-36.80	14.25	4.45	8.27-20.00	12.55	2.41
12.	Taxaceae	11.45-30.20	18.07	2.64	11.45-30.20	18.07	2.64

Table: 2
Representation of species in 80 genera of gymnosperms with respect to the range and mean of 1C and 1Cx-values

Sr. No.	Taxon	Species with C-values/ No. of species in the genus	Representation (%)	Range of holoploid genome size (1C- value, pg)	Mean	Ratio (max./ min.)	Range of monoploid genome size (1Cx- value, pg)	Mean	Ratio (max./ min.)
1.	<i>Cycas</i>	8/92	8.69	12.55-14.76	13.42	1.18	12.55-14.76	13.42	1.18
2.	<i>Dioon</i>	2/13	15.38	24.35-25.60	24.98	1.05	24.35-25.60	24.98	1.05
3.	<i>Bowenia</i>	2/2	100.00	20.70-21.95	21.33	1.06	20.70-21.95	21.33	1.06
4.	<i>Macrozamia</i>	2/39	5.13	26.50-26.90	26.70	1.02	26.50-26.90	26.70	1.02
5.	<i>Lepidozamia</i>	2/2	100.00	27.65-30.15	28.90	1.09	27.65-30.15	28.90	1.09
6.	<i>Encephalartos</i>	16/68	20.59	21.09-32.35	29.12	1.53	21.09-32.35	29.12	1.53
7.	<i>Stangeria</i>	1/1	100.00	19.50	19.50	-	19.50	19.50	-
8.	<i>Ceratozamia</i>	1/25	4.00	31.60-31.65	31.58	1.00	31.60-31.65	31.58	1.00
9.	<i>Microcycas</i>	1/1	100.00	20.60	20.60	-	20.60	20.60	-
10.	<i>Zamia</i>	62/65	95.38	16.85-22.85	19.86	1.36	16.85-22.85	19.86	1.36
11.	<i>Ginkgo</i>	1/1	100.00	11.75	11.75	-	11.75	11.75	-
12.	<i>Welwitschia</i>	1/1	100.00	7.20	7.20	-	7.20	7.20	-
13.	<i>Gnetum</i>	4/41	9.76	2.27-4.20	3.60	1.85	2.27-4.20	3.60	1.85
14.	<i>Ephedra</i>	49/70	70.00	8.10-38.34	16.84	4.73	7.58-10.38	9.24	1.37
15.	<i>Cedrus</i>	3/3	100.00	15.90-16.50	16.24	1.04	15.90-16.50	16.24	1.04
16.	<i>Pinus</i>	82/130	63.08	19.45-37.68	27.76	1.94	19.45-37.68	27.76	1.94
17.	<i>Cathaya</i>	1/1	100	24.75	24.75	-	24.75	24.75	-
18.	<i>Picea</i>	7/40	17.50	17.45-21.05	19.43	1.21	17.45-21.05	19.43	1.21
19.	<i>Pseudotsuga</i>	1/4	25.00	17.60	17.60	-	17.60	17.60	-
20.	<i>Larix</i>	7/14	50.00	9.50-13.99	13.10	1.47	9.50-13.99	13.10	1.47
21.	<i>Pseudolarix</i>	1/1	100.00	26.10	26.10	-	26.10	26.10	-
22.	<i>Tsuga</i>	7/10	70.00	17.05-20.95	19.17	1.23	17.05-20.95	19.17	1.23

Table 2: Continued

Sr. No.	Taxon	Species with C-values/ No. of species in the genus	Representation (%)	Range of holoploid genome size (1C- value, pg)	Mean	Ratio (max./ min.)	Range of monoploid genome size (1Cx- value, pg)	Mean	Ratio (max./ min.)
23.	<i>Keteleeria</i>	1/3	33.33	24.20	24.20	-	24.20	24.20	-
24.	<i>Abies</i>	12/48	25.00	15.80-19.85	17.80	1.26	15.80-19.85	17.80	1.26
25.	<i>Araucaria</i>	19/19	100.00	15.65-22.75	17.31	1.45	15.65-22.75	17.31	1.45
26.	<i>Wollemia</i>	1/1	100.00	14.20	14.20	-	14.20	14.20	-
27.	<i>Agathis</i>	3/18	16.67	10.82-15.80	13.41	1.46	10.82-15.80	13.41	1.46
28.	<i>Phyllocladus</i>	3/4	75.00	7.95-11.40	10.08	1.43	7.95-11.40	10.08	1.43
29.	<i>Lepidothamnus</i>	3/3	100.00	4.75-6.70	6.00	1.41	4.75-6.70	6.00	1.41
30.	<i>Prumnopitys</i>	3/9	33.33	5.60-7.95	7.15	1.42	5.60-7.95	7.15	1.42
31.	<i>Sundacarpus</i>	1/1	100.00	6.80	6.80	-	6.80	6.80	-
32.	<i>Halocarpus</i>	3/3	100.00	7.85-11.70	8.95	1.49	7.85-11.70	8.95	1.49
33.	<i>Lagarostrobos</i>	1/1	100.00	15.20	15.20	-	15.20	15.20	-
34.	<i>Manoao</i>	1/1	100.00	17.15	17.15	-	17.15	17.15	-
35.	<i>Saxegothaea</i>	1/1	100.00	5.10	5.10	-	5.10	5.10	-
36.	<i>Microcachrys</i>	1/1	100.00	4.15	4.15	-	4.15	4.15	-
37.	<i>Pherosphaera</i>	2/2	100.00	4.20-4.30	4.25	1.02	4.20-4.30	4.25	1.02
38.	<i>Acmopyle</i>	2/2	100.00	6.90-8.75	7.83	1.27	6.90-8.75	7.83	1.27
39.	<i>Dacrycarpus</i>	2/9	22.22	5.35-18.10	11.73	3.38	5.35-18.10	11.73	3.38
40.	<i>Dacrydium</i>	6/22	27.27	6.55-18.00	8.70	2.75	6.55-18.00	8.70	2.75
41.	<i>Falcatifolium</i>	1/6	16.67	11.20	11.20	-	11.20	11.20	-
42.	<i>Retrophyllum</i>	1/5	20.00	5.90	5.90	-	5.90	5.90	-
43.	<i>Nageia</i>	1/6	16.67	5.60	5.60	-	5.60	5.60	-
44.	<i>Afrocarpus</i>	3/5	60.00	5.10-6.40	5.77	1.25	5.10-6.40	5.77	1.25
45.	<i>Podocarpus</i>	7/108	6.48	7.95-11.25	9.17	1.42	7.95-11.25	9.17	1.42
46.	<i>Sciadopitys</i>	1/1	100.00	20.80	20.80	-	20.80	20.80	-
47.	<i>Cunninghamia</i>	2/2	100.00	12.75	12.75	-	12.75	-	-
48.	<i>Taiwania</i>	1/1	100.00	13.40	13.40	-	13.40	13.40	-
49.	<i>Athrotaxis</i>	3/3	100.00	10.10-11.11	10.74	1.10	10.10-11.11	10.74	1.10
50.	<i>Metasequoia</i>	1/2	50.00	11.05	11.05	-	11.05	11.05	-
51.	<i>Sequoia</i>	1/1	100.00	32.14	32.14	-	10.71	10.71	-
52.	<i>Sequoiadendron</i>	1/1	100.00	10.40	10.40	-	10.40	10.40	-
53.	<i>Cryptomeria</i>	1/1	100.00	11.05	11.05	-	11.05	11.05	-
54.	<i>Glyptostrobus</i>	1/1	100.00	9.99	9.99	-	9.99	9.99	-
55.	<i>Taxodium</i>	2/2	100.00	9.05-9.95	9.38	1.10	9.05-9.95	9.38	1.10
56.	<i>Papuacedrus</i>	1/1	100.00	12.00	12.00	-	12.00	12.00	-
57.	<i>Austrocedrus</i>	1/1	100.00	10.90	10.90	-	10.90	10.90	-
58.	<i>Libocedrus</i>	2/5	40.00	19.50-20.00	19.75	1.03	19.50-20.00	19.75	1.03
59.	<i>Pilgerodendron</i>	1/1	100.00	14.05	14.05	-	14.05	14.05	-
60.	<i>Widdringtonia</i>	3/4	75.00	10.75-14.40	12.01	1.34	10.75-14.40	12.01	1.34
61.	<i>Diselma</i>	1/1	100.00	9.05	9.05	-	9.05	9.05	-
62.	<i>Fitzroya</i>	1/1	100.00	17.50	17.50	-	8.75	8.75	-
63.	<i>Callitris</i>	2/15	13.33	10.40-11.17	10.78	1.07	10.40-11.17	10.78	1.07
64.	<i>Actinostrobus</i>	1/3	33.33	10.65	10.65	-	10.65	10.65	-
65.	<i>Neocallitropsis</i>	1/1	100.00	12.75	12.75	-	12.75	12.75	-
66.	<i>Thujopsis</i>	1/1	100.00	13.97	13.97	-	13.97	13.97	-
67.	<i>Thuja</i>	5/5	100.00	11.95-13.43	12.53	1.12	11.95-13.43	12.53	1.12
68.	<i>Fokienia</i>	1/1	100.00	12.53	12.53	-	12.53	12.53	-
69.	<i>Chamaecyparis</i>	5/5	100.00	8.55-15.05	11.75	1.76	8.55-15.05	11.75	1.76
70.	<i>Cupressus</i>	10/19	52.63	10.75-22.33	12.73	2.08	10.75-14.18	11.98	1.32
71.	<i>Juniperus</i>	67/75	89.33	10.96-36.8	14.71	3.36	10.96-14.94	12.54	1.36
72.	<i>Calocedrus</i>	2/4	50.00	15.25-17.00	16.13	1.11	15.25-17.00	16.13	1.11
73.	<i>Tetraclinis</i>	1/1	100.00	12.85	12.85	-	12.85	12.85	-
74.	<i>Platycladus</i>	1/1	100.00	10.46	10.46	-	10.46	10.46	-
75.	<i>Microbiota</i>	1/1	100.00	10.17	10.17	-	10.17	10.17	-
76.	<i>Pseudotaxus</i>	1/1	100.00	17.30	17.30	-	17.30	17.30	-
77.	<i>Taxus</i>	5/9	55.55	11.45-12.15	11.80	1.06	11.45-12.15	11.80	1.06
80.	<i>Cephalotaxus</i>	1/8	12.50	26.25	26.25	-	26.25	26.25	-
81.	<i>Amentotaxus</i>	1/6	16.67	30.20	30.20	-	30.20	30.20	-
82.	<i>Torreya</i>	3/6	50.00	21.60-22.30	21.98	1.03	21.60-22.30	21.98	1.03

Fig.1

Fig. 1

Comparison of the mean and range (in parentheses) of 1C holoploid genomes of 12 families of gymnosperms included in four subclasses indicated by the following abbreviations: Cy (Cycadidae), Gi (Ginkgoidae), Gn (Gnetidae), Pi (Pinidae) (Christenhusz et al. 2011). The number in parentheses following family names indicates the number of species for which genome size data are available.

Variation of genome size among families

The genome size estimates are now available for 80 genera comprising 471 species (514 taxa) of gymnosperms (Table 2), therefore representing 91.0 % of genera and 42.6 % of the total recognized species. The 1C values range from 2.27 pg (*Gnetum ula*) to 38.34 pg (*Ephedra antisiphilitica*) while 1Cx values range from 2.27 pg (*Gnetum ula*) to 37.68 pg (*Pinus gerardiana*) therefore showing a 16.89 and 16.60-fold variation respectively (Tables 1 and 2). This is a rather narrow range compared with 2400-fold variation in angiosperms with 1C values varying from 0.063 pg (*Genlisea margaretae*) to 152.23 pg (*Paris japonica*) (Leitch and Leitch 2012). The mean 1C and 1Cx values for 514 taxa of gymnosperms considered presently are 18.16 pg and 16.80 respectively, compared with mean 1C value of 5.12 pg obtained from 10768 species of angiosperms (Pellicer et al. 2018).

Cycadaceae: The genome size is known for eight species out of 92 accepted species of *Cycas*. The range of 1C values is 12.55-14.76 pg with a mean of 13.42 pg (Tables 1 and 2, Fig. 1).

Zamiaceae: The genome size has been studied of 89 out of 216 species (41.20 %) and 1C and 1Cx values range from 16.85-32.35 pg (1.92-fold) with mean value of 22.67 pg (Table 1, Fig. 1). The maximum 1.53-fold range is shown by *Encephalartos* and the highest mean 1C and 1Cx value (31.58 pg) is depicted by *Ceratozamia* (Table 2). *Zamia* has a 1.36-fold variation with a mean 1C and 1Cx value of 19.86 pg in 62 species studied (Tables 2).

Ginkgoaceae: *Ginkgo biloba* has a 1C and 1Cx value of 11.75 pg (Tables 1 and 2, Fig. 1).

Welwitschiaceae: *Welwitschia mirabilis* shows a 1C and 1Cx value of 7.20 pg (Tables 1 and 2, Fig. 1).

Gnetaceae: Four species of *Gnetum* show a range of 1C and 1Cx values from 2.27-4.20 pg therefore showing a 1.85-fold variation (Tables 1 and 2, Fig. 1).

Ephedraceae: In *Ephedra* 49 species have been studied out of 70 accepted species and the 1C values range from 8.10 (*E. regeliana*, 2x) to 38.34 pg (*E. antisiphilitica*, 8x) a 4.73-fold variation and a mean value of 16.84 pg (Fig. 1). On the other hand 1Cx values vary from 7.58 (*E. likiangensis*, 4x) to 10.38 pg (*E. pachyclada* subsp. *sinaica*, 4x), a 1.37-fold variation and a mean value of 9.24 pg (Tables 1 and 2).

Pinaceae: Genome size is known for 122 species of this family the 1C and 1Cx values range from 12.30 (*Larix sibirica*)-37.68 pg (*Pinus gerardiana*) therefore showing a 3.06-fold variation (Table 1, Fig. 1). The most variable genus is *Pinus* with a 1.94-fold variation with the highest 1C mean value of 27.76 pg as compared to the rest of the genera (Table 2). Members of subg. *Strobos* have an average genome size 5.2 Gbp larger than the subg. *Pinus* (Grotkopp et al. 2004).

Araucariaceae: 1C and 1Cx values of 23 species studied range from 10.82 (*Agathis robusta*)-22.75 pg (*Araucaria araucana*) (mean=15.86 pg) therefore showing a 2.10-fold variation (Tables 1 and 2, Fig. 1).

Podocarpaceae: 1C and 1Cx values of 42 species studied range from 4.15 pg (*Microcachrys tetragona*) to 18.10 pg (*Manoao colensoi*) (mean=8.36 pg) therefore showing a 4.36-fold variation (Tables 1 and 2, Fig. 1).

Sciadopityaceae: *Sciadopitys verticillata* the only species shows 1C and 1Cx value of 20.80 pg (Tables 1 and 2, Fig. 1).

Cupressaceae: 1C values of 120 species range from 8.27 (*Callitris columellaris*, 2x)-36.80 pg (*Juniperus phoenicea*, 6x) (mean=14.25 pg) and a 4.45-fold variation, and 1Cx values range from 8.27 (*Callitris columellaris*, 2x) -20.00 pg (*Libocedrus plumosa*, 2x) (mean=12.55 pg) and a 2.10-fold variation (Tables 1 and 2, Fig. 1). *Cupressus* shows mean 1C and 1Cx values of 12.73 pg and 11.98 pg respectively (Table 2). The largest variation is shown by *Juniperus* with 1C values ranging from 10.96 (*Juniperus communis* var. *nipponica*, 2x) -36.80 pg (*Juniperus phoenicea*, 6x) (3.36-fold, mean=14.71) and 1Cx values show 1.36-fold variation ranging from 10.96 (*Juniperus communis* var. *nipponica*, 2x) to 14.94 pg (*Juniperus foetidissima*, 4x) (mean=12.54) (Table 2).

Taxaceae: 1C and 1Cx values of 11 species studied show a range of 11.45 (*Taxus cuspidata*)-30.20 pg (*Amentotaxus yunnanensis*) (mean=36.13) and a 2.64-fold variation (Tables 1 and 2, Fig. 1).

Intraspecific variation

Intraspecific variation has been deduced in earlier studies mostly because of inappropriate methodology, taxonomic riddles and lack of proper cytological analysis of the taxa studied (Griehuber 2005, Murray 2005). Many such examples of refutable intraspecific variation mainly because of methodological errors have been reported earlier (Murray 1998, Ohri 1998, Ahuja and Neale 2005). However, some recent reports of intraspecific variability in conifers need to be mentioned.

A significant variation in 1C DNA amount of 2.18 pg has been reported in four provenances of *Pinus tecunumanii* ranging from 19.25-21.43 pg (Hall et al. 2000). Similarly Nunes et al. (2009) also found a difference of 1.16 pg in four provenances of *P. tecunumanii* ranging from 20.33-21.49 pg in the same species. According to Hall et al. (2000) higher genome size is shown by the provenances found in the primary areas of distribution, an observation not supported by Nunes et al. (2009). Significant variation has also been observed in 1C DNA contents between two *P. nigra* subspecies (subsp. *dalmatica* 22.25 pg and subsp. *nigra* 22.75-23.23 pg) and in continental and sub-Mediterranean populations of *P. nigra* subsp. *nigra* (Bogunic et al. 2003). However, no significant variation of 1C DNA content was found in a later study on 20 populations of five subspecies of *P. nigra* (subsp. *dalmatica* 23.79 pg; *mauretanica* 23.24 pg; *nigra* 23.63 pg; *pallasiana* 23.80 pg; *salzmanni* 23.55 pg) (Bogunic et al. 2007). Intraspecific variation in 1C DNA amounts has also been reported in some tropical Mexican pine species belonging to subg. *Pinus* e.g. *P. oocarpa* (21.11-22.38 pg) and *P. tecunumanii* (19.25-21.43 pg) (Hall et al. 2000). Similarly *P. patula* (21.92-21.94pg) is significantly different from *P. patula* subsp. *longipedunculata* (20.90 pg) (Hall et al. 2000). Significant intraspecific differences have also been observed in two populations of *Manoao colensoi* originating from two main islands of New Zealand. One population from Gremouth shows 1C values of 13.90 and 13.70 pg in female and male plants respectively the other population from Ruahine Range and Coromandel State forest shows 16.40 and 17.15 pg in male and

female plants respectively which according to Davies et al. (1997) might justify their segregation as two different taxa.

A recent study shows that genome size may also be affected under different growth conditions. Genome size measurement has been made in 19 conifer species growing as wild, potted and bonsai plants. It has been generally observed that 1Cx values decrease by 8.48 % in potted plants and increase by 26.86 % in bonsai plants as compared to the wild plants (Hidalgo et al. 2015). Most significantly in *Juniperus thurifera* wild, potted and bonsai plants show 1C values of 18.84, 20.58, and 22.99 pg respectively, this range according to Hidalgo et al. (2015) cannot be attributed to the natural variation, being larger than the range (19.92-21.33pg) observed in natural wild populations (Romo et al. 2013).

There are in fact a number of cases where intraspecific stability in genome size has been noticed (Murray 1998, Ohri 1998). For example, *Abies fraseri* shows no significant variation in its six disjunct populations in the Appalachian mountains in which 1C values vary from 16.90 to 17.50 pg (Auckland et al. 2001).

Adaptive significance of genome size variation

Nucleotypic effects of genome size variation as related to nucleus and cell size and cell cycle duration further manifesting in various phenotypic and developmental characteristics are now well known across various taxonomic groups (Greilhuber and Leitch 2013). Various studies done till now in conifers indeed indicate the correlation of genome size with geographical distribution, climatic adaptation and development.

Ohri and Khoshoo (1986) showed a general trend of correlation of genome size with climate, aridity, seed size, etc. in *Pinus*. Wakamiya et al. (1993) studied the relationship between genome size and climatic and growth indices in 18 North American and one exotic species of *Pinus*. The genome size showed negative correlation with lowest mean annual precipitation as well as mean monthly precipitation and with highest mean monthly spring air temperature, while a positive correlation was observed with seed weight and volume and minimum seed bearing age (Wakamiya et al. 1993). Subsequently, a study on six species showed that though hydraulic conductivity was not correlated with genome size, yet the species with larger genome size show better ability to withstand low water potential because of thicker cell walls and smaller ratio of lumen radii to cell wall thickness (Wakamiya et al. 1996). However, a comparison of genome sizes of 28 species belonging to subg. *Strobus* (soft pines) and subg. *Pinus* (hard pines) originating from both temperate and tropical zones of New World does not bring out any latitudinal cline rather they segregate according to the taxonomic division of the genus as the genome sizes of soft pines (subg. *Strobus*) are generally larger than those of hard pines (subg. *Pinus*) (Hall et al. 2000). The authors have therefore surmised that earlier latitudinal correlations were because of inadequate sampling. In a later study on the combined data set of 46 species of both Asian and New World origin again the genome size followed the subgeneric classification rather than geographical distribution or the latitude (Joyner et al. 2001).

A detailed analysis in this respect has been made by Grotkopp et al. (2004) based on the total dataset of 85 species of *Pinus*. The complete dataset including 60 taxa studied by them supports the division of *Pinus* into two subgenera, where subg. *Strobus* has a significantly greater mean genome size (1C=32.08 pg) as compared to subg. *Pinus* (1C=27.00 pg). The authors have included many variables concerning geographical distribution and climate and many life history traits to derive robust correlations with genome size. Therefore genome size has been shown to be negatively correlated with latitude which is more strongly supported in subg. *Pinus*, and a weak negative relationship with average annual rain fall, while no relationship is evident in relation to elevation. Similarly, while a marginal positive relationship occurs between genome size and minimum generation time in the whole of genus and in subg. *Pinus* but this relationship is not present within the subg. *Strobus*. In the same way no relationship exists with respect to interval between large seed crops, serotiny and rarity. However, a negative correlation of genome size and invasiveness exists in the subg. *Pinus* (subg. *Strobus* has only one invasive species in the sample) and because invasive species require faster growth, a negative correlation also exists with respect to seedling relative growth rate (RGR) across the whole genus. Among the components of RGR, leaf area ratio (LAR), specific leaf area (SLA) and relative leaf production rate (RLPR) are negatively correlated with genome size (Grotkopp et al. 2002, 2004). The strongest positive correlation has been observed between genome size and seed mass and mode of dispersal (animal vs. wind). Interestingly the seed mass shows a significantly negative correlation with RGR, LAR, and SLA in subg. *Pinus* while this relationship in subg. *Strobus* is shown only with respect to LAR and SLA (Grotkopp et al. 2004). Interestingly an unequivocal positive correlation of seed mass with aridity has been observed across the whole genus based on 99 species where subg. *Strobus* which inhabits arid regions has heavier seeds and this correlation is also conspicuous when analyzed within the subg. *Pinus* where species of subsection *Pinaster* possess heavier seeds coupled with aridity as compared to those of subsection *Pinus* (Salazar-Tortosa et al. 2020). Remarkably, subg. *Strobus* which inhabits more arid conditions has an average larger genome size than subg. *Pinus* and the trend of larger genome size associated with arid conditions is also apparent within the subg. *Pinus* (Grotkopp et al. 2004). This association between seed mass and aridity has also been reported in 88 North American conifer species including 36 *Pinus* species (Rueda et al. 2017). These studies therefore present an equivocal evidence for the adaptive value of genome size in gymnosperms as the positive correlation is seen only in few characters and weak or neutral relationship in rest of the cases.

Components of genome expansion

Modulation of Repetitive DNA Sequences

It is now well established that gymnosperms are generally characterized by large average genome sizes (Ohri and Khoshoo 1986, Murray 1998, Pellicer and Leitch 2019) and their

chromosome complements are marked by AT or GC rich regions at centromeres, telomeres and interstitial positions (Murray 2013, Rastogi and Ohri 2020b, Ohri 2021b). Early studies have shown that these massive genomes consist of large proportion of repetitive DNA as in case of *Pinus resinosa*, *P. banksiana* and *Picea glauca* (Miksche and Hotta 1973). In *Cycas revoluta* the DNA reassociation study showed highly repetitive, fast repeats, slow repeats and single copy sequences comprising 15 %, 24 %, 44 % and 17 % respectively of the complete genome (Kurdi-Haidar et al. 1983). Some later studies also showed a repetitive and 'unique DNA' sequences comprising 75 % and 20-30 % respectively of the total DNA in *Picea glauca*, *Pinus banksiana*, *P. lambertiana*, *P. resinosa* (Rake et al. 1980) and *P. strobus* (Kriebel 1985). A later reanalysis of Cot fractions in *Pinus strobus* showed that earlier estimate for unique sequences was an overestimate as shown by 14 % single-copy, 20 % low-repetitive, 45 % middle-repetitive and 15 % highly-repetitive components (Elsik and Williams 2000). Moreover, a recent study on Cot analysis of *Pinus taeda* has resulted in highly repetitive, moderately repetitive and single/low-copy fractions accounting for 57.2, 23.9, 10.2 respectively and rest of the 9.0 % being fold back and damaged DNA (Perera et al. 2018). This study shows a much lower percentage of 'unique' sequences in *P. taeda* compared to the earlier reports in some conifers (Rake et al. 1980), which according to Perera et al. (2018) may be highly inflated and flawed values because of technical reasons in accounting for fold back and damaged DNA. Similarly, *Taxodium distichum* also shows highly repetitive (47.0 %), moderately repetitive (41.15 %) and single/low-copy (7.4 %), therefore repetitive fraction makes up 90.0 % of the total genome (Liu et al. 2011). The mean repetition frequencies of moderately repetitive and highly repetitive sequences have been calculated as 61 and 2054 respectively (Liu et al. 2011).

The repetitive DNA elements which make up a major component of the genome can be grouped broadly into tandemly repetitive elements comprised by different types of satellite DNAs, the telomeric repeat and rDNA and dispersed repeated sequences constituted by retroelements (Kubis et al. 1998). Tandem repeats, in particular, have been shown to constitute about 3.0 % of the pine genomes (Wegrzyn et al. 2014).

Tandemly repetitive DNA

The DNA reassociation study done on *Cycas revoluta*, having 1C amount of 10.30 pg shows short interspersed pattern of repetitive and single copy sequences and undefined repetitive sequences ranging from 100-350 bp, which showed a bimodal distribution with one class around 150 bp and the other at 300 bp, rest of the repetitive sequences have a broad size range of 400bp to 2kb (Kurdi-Haidar et al. 1983).

In *Zamia paucijuga* a highly repetitive DNA family (*ZpS1*) organized as unit repeats of GC rich 323bp was isolated and cloned. This highly repetitive DNA which makes up 5 % of the genome and is localized at subtelomeric regions of one end of most of the chromosomes, shows two types of organization as tandem arrays of typical satellite DNA and as dispersed elements composed mostly of 2-4 satellite DNA tandem repeats flanked by two direct AT-rich repeats of 0.6 kb length (Cafasso

et al. 2003). Further study in 30 other species of *Zamia* showed that the dispersed pattern of *ZpS1* repeat differed among various species, being in low copy number in some species while in others they are present in high copy numbers due to secondary amplification events, (Cafasso et al. 2009). Further detailed characterization of *ZpS1* repeat in 12 species of *Zamia* showed sequence similarity to a great extent indicating that *ZpS1* satellite has originated from a common ancestor which has been maintained in original structure to the present time (Cafasso and Chinali 2014).

A similar genus specific tandem repeat has been reported in *Picea*. This 142 bp tandem repeat identified by restriction enzyme *Sau3A* is present in 20,000 copies in *Picea glauca* and is located in the centromeric region of some of the chromosomes (Brown et al. 1998, Hizume et al. 1999, Shibata and Hizume 2008). Furthermore, the Southern hybridization showed that *Sau3A* elements are found in 17 more *Picea* species but not in *Pinus radiata*, *Pseudotsuga menziesii* and *Thuja plicata* therefore showing specificity to only *Picea* (Brown et al. 1998). Another tandem repeat unit specific to *Larix* has been cloned from HindIII- digested genomic DNA of *L. leptolepis*. Two AT rich repeat units 170bp and 220 bp and coinciding with DAPI bands located in the proximal region of all the chromosomes were identified and named as *LPD* family (*Larix* proximal DAPI band specific repeat sequence family) is conserved in *Larix* species and constitutes 2.8-3.4 % of the genomes of most of the species (Hizume et al. 2002b).

Microsatellites or simple sequence repeats are characterized by single sequence motif (1-6 bp) tandemly repeated for varying number of times. These are abundantly found in conifers and being a rich source of genetic variation, have been used as valuable markers for mapping, fingerprinting and population genetic studies (Christiakov 2006, Oliveira 2006, Echt et al. 2011). Microsatellites or simple sequence repeats abundantly occur in conifers and dimeric AG and AC motifs are highly amplified forming major component of the genome (Smith and Devey 1994, Echt and May-Marquadt 1997, Pfeiffer et al. 1997, Soranzo et al. 1998, Schmidt et al. 2000, Elsik and Williams 2001, Yazdani et al. 2003, Hung et al. 2012, Herbinger et al. 2011). Moreover, SSRs are also abundant in the coding fractions of the genome called EST-SSRs (Morgante et al. 2002, Ranade et al. 2014). Here they occur in 5'- or 3'- Untranslated Regions or in Open reading Frames (Li et al. 2004). The in silico analysis of EST-SSRs in seven gymnosperm genera shows that their length is positively related to the genome size irrespective of their density and it further shows that dimer is the most abundant class while the proportion of trimers increases in ORF's (Ranade et al. 2014). The most abundant dimer motif is AT/AT in *Pinus* (Berube et al. 2007, Chagne et al. 2004, Von Stackelberg et al. 2006, Victoria et al. 2011, Maomao et al. 2012) *Picea* (Zane et al. 2002, Rungis et al. 2004, Von Stackelberg et al. 2006) and *Ginkgo* (Von Stackelberg et al. 2006) while AG/CT is most amplified in *Cycas* (Von Stackelberg et al. 2006) and *Gnetum* (Victoria et al. 2011). Minisatellite repeats having longer repeat unit also have been shown to constitute substantial part of conifer genome being highly amplified in two and three-needle pines (*P. echinata*, *P. elliotii* var. *elliotii*, *P.*

palustris, *P. caribaea*, *P. oocarpa*, *P. banksiana*, *P. massoniana*, *P. resinosa*) and moderately abundant in *Pinus strobus*, *Picea glauca*, *P. abies*, *Taxodium dsitchicum* and *Ginkgo biloba* (Schmidt et al. 2000).

Telomeres are short tandemly arranged minisatellite sequences present at the ends of chromosomes however, in many gymnosperms they have been found to be located in the interstitial positions also (Fuchs et al. 1995, Hizume et al. 2000, Schmidt et al. 2000, Murray et al. 2002, Leitch and Leitch 2012, Rastogi and Ohri 2020b, Ohri 2021b). *Arabidopsis* type (A-type) of telomeric repeat (TTTAGGG)_n has been found to be common in gymnosperms (Fuchs et al. 1995). Telomeric repeats have been reported to be highly amplified in two and three needle pines of the section *Pinus* as shown by the presence of these repeats at different genomic loci with a conserved organization (Schmidt et al. 2000). Similar AT rich telomere sequences located by DAPI bands have been observed in proximal, interstitial and terminal regions in *Pinus densiflora* (Shibata et al. 2005). Four positive clones of these sequences were found to possess A-type telomeric sequences as well as degenerate telomeric sequences and FISH analysis shows that these sequences are located at both proximal and interstitial positions (Shibata et al. 2005). Another tandem repeat of 27 bp repeat unit has been located on the proximal CMA bands of 20 chromosomes of *P. densiflora* and named as *PCSR* (proximal CMA band specific repeat) later it was physically mapped in three more *Pinus* species (Hizume et al. 2001, 2002a). Pines are generally known to possess long telomeres up to 57 Kbp (Flanary and Kletetschka 2005, Aronen and Ryyanen 2012).

Ribosomal DNA is present as tandem repeats on chromosomal loci whose number in gymnosperms can vary from 2-38 in the somatic complement (Garcia et al. 2017, Rastogi and Ohri 2020b, Ohri 2021b). Each repeat unit comprises a transcribed unit separated from the next by an intergenic spacer (IGS). The transcribed region consists of an external transcribed spacer (ETS), the 18S gene, an internal transcribed spacer (ITS1), the 5.8S gene, a second ITS (ITS2) and the 26S gene. The length of the repeat unit depends on intergenic spacers as the transcribed region is highly conserved. This repeat unit is present in thousands of copies and occurs in one or many loci on one or several pairs of chromosomes in the somatic complement. In contrast to angiosperms with a narrow range and smaller size (565-700 base pairs) (Baldwin et al. 1995), the ITS region in gymnosperms have a larger range and size (975-3125 bp) (Liston et al. 1996). This is evident from 32 genera sampled from different families e.g. Pinaceae with the largest range (1550-3125 bp), Cupressaceae, Taxodiaceae, Cephalotaxaceae, and Taxaceae with a relatively narrow range (975-1125 bp), followed by Sciadopityaceae (1250 bp), Araucariaceae (1325-1350 bp), Podocarpaceae (2000-2100 bp), Cycadales (1150- 1450 bp), Ginkgoales (1200 bp), *Ephedra* (1500 bp), *Gnetum* (1200 bp) and *Welwitschia* (750 bp) from Gnetales. Remarkably the smallest size of ITS region as found in *Welwitschia* is still larger by 50 bp than the largest size seen in angiosperms (Liston et al. 1996). Some earlier reports on certain *Pinus* and *Picea* species also show the size of ITS region to be ca. 3000 bp in size (Cullis et al. 1988, Bobola et al. 1992, Maggini and Baldassini 1995). In

gymnosperms ITS1 is mainly responsible for this large range of length variation caused by variable repeated elements among different taxa (Maggini and Baldassini, 1995; Liston et al., 1996; Gernandt et al., 2001; Ickert-Bond and Wojciechowski, 2004, Campbell et al. 2005). The ITS1 in 11 genera of Pinaceae has been shown to range from 944 bp (*Pseudolarix amabilis*) to 3271 bp (*Picea abies*) and this largest intergeneric length variation among the eukaryotes is attributed to the number variation in subrepeats (Kan et al. 2007). Furthermore, the remarkable variability is evident in a study of 16 *Gnetum* species from South America, Africa and Asia in which ITS1 ranges from 225-986 bp and ITS2 from 259-305 bp (Won and Renner 2005).

Transposable elements

The other major contributor to the genome is dispersed repeats represented by transposable elements. Transposable elements (TEs) are characterized into two classes by the mechanism of transposition, class1 TEs transpose by an RNA intermediary and class2 have DNA-mediated transposition. Class 1 TEs have two subclasses; long terminal repeat retrotransposons (LTRs) and those which lack terminal repeats (non-LTRs). LTR retrotransposons are more prevalent in plant and animal taxa and have two distinct groups Ty1-copia and Ty3-gypsy. Non-LTR transposons have long interspersed nuclear elements (LINEs), short interspersed nuclear elements (SINEs), and miniature inverted repeat transposable elements (MITEs) (Kejnovsky et al. 2012). These retrotransposons, showing sequence divergence and varying copy numbers, are distributed widely in the genomes of various gymnosperm taxa and are therefore primarily responsible for the diversity in genome sizes (Voronova et al. 2017). Ty1-copia like retrotransposons have been identified in *Agathis robusta*, *Cephalotaxus fortunei*, *Callitris roei*, *Pinus coulteri*, *Cedrus deodara*, *Picea glauca*, *Podocarpus gracilior*, *Metasequoia glyptostroboides*, *Taxus baccata* (Voytas et al. 1992). A ≥4.8 kbp repeat sequence called TPE1 forming an internal part of Ty1-copia retrotransposon has been isolated from the genomic library of *Pinus elliotii*. This sequence, which is evenly distributed in the chromosome complement, is highly amplified in all the nine *Pinus* species from both the subgenera *Pinus* and *Strobus* besides *Picea abies* and *P. glauca*, but shows much less amplification in *Taxodium distichum* and *Ginkgo biloba* (Kamm et al. 1996). A 6 kb long Ty3-gypsy like retrotransposon named *IFG-7* (after Institute of Forest Genetics) has been isolated from *Pinus radiata* and has been detected in species belonging to both the subgenera of *Pinus* (Kossack and Kinlaw 1999). Though this element is also present in seven other genera of Pinaceae, it is absent in Cupressaceae and Taxodiaceae (Kossack and Kinlaw 1999). A *gymny* (Ty3-gypsy) element has been reported to be amplified (16 MYA) in pine genomes. These elements show a dispersed distribution in the genome and contribute equivalent of *Arabidopsis thaliana* genome to the genome of *Pinus taeda* (Morse et al. 2009). The retroelement *IFG-7* with 2,10,557 copies (Magbanua et al. 2011) and *Gymny* with 21,700 copies account for 5.8 % and 1.3 % respectively of the *P. taeda* genome (Morse et al. 2009), therefore together constituting 1.54 Gbp of *Pinus taeda* genome, a value equivalent to 9.8 times that of *Arabidopsis thaliana* (Magbanua et al.

2011, Morse et al. 2009). Both *Ty1-copia* and *Ty3-gypsy* like elements have been observed in *Picea glauca* and *P. mariana*. These elements are 5–10 kb long and constitute 10 % of the *Picea* genome (L'Homme et al. 2000). Furthermore, a large component of single copy sequences in *Pinus taeda* has been shown to originate from diverged retroelements which form a major part of the excess of the low-repetitive kinetic fraction of the genome (Elsik and Williams 2000). Such divergent retroposon families contributing to the genome expansion have been reported in a number of conifer species (Stuart-Rogers and Flavell 2001, Friesen et al. 2001, Murray et al. 2002, Kovach et al. 2010, Liu et al. 2011). An *Au SINE* element has also been reported in *Ephedra ciliata* which is 15 bp shorter than the consensus sequence and is similar to that of *Glycine*, therefore indicating its distribution in both angiosperms and gymnosperms (Yagi et al. 2011). Recently an interesting case of horizontal transposon transfer (HTT) of arthropod *Penelope*-like elements (PLEs) lineage EN(+) PLEs called *Dryads* to a common ancestor of conifers has been reported. These *Dryads* are highly diversified in conifers but are absent from other gymnosperms (Lin et al. 2016).

Recent genome annotations (Morse et al. 2009, Nystedt et al. 2013, Neale et al., 2014, Wegrzyn et al., 2014, Guan et al. 2016, Wan et al. 2018) and BAC sequencing (Kovach et al. 2010, Magbanua et al. 2011, Stival Sena et al. 2014, Seoane-Zonjick et al. 2016, Perera et al. 2018) show that three families i.e. the *Ty3/Gypsy*, *Ty1/Copia* and *Gymny* superfamilies make up the bulk of LTR-RTs in conifers. Repetitive elements constitute a large portion of the genomes, estimated at 40 % in *Picea glauca* (Hamberger et al 2009), 69 % in *Picea abies* (Nystedt et al., 2013), 79 % in *Pinus lambertiana* (Stevens et al. 2016), 76.58 % in *Ginkgo biloba* (Guan et al. 2016), 78 % in *Abies alba* (Mosca et al. 2019), 80 % in *Pinus taeda* (Wegrzyn et al. 2014), 82.0 % in *Larix sibirica* (Kuzmin et al. 2019) and up to 85.9 % in *Gnetum montanum* (Wan et al. 2018). The LTR-RT sequences in particular make up 58 % of the genome both in *Picea abies* and in *Pinus taeda* (Nystedt et al. 2013, Neale et al. 2014, Wegrzyn et al. 2013, 2014), 67 % in *Pinus lambertiana* (Stevens et al. 2016) and only *gypsy*-like elements account for 77.4 % of the genome of *Gnetum montanum* (Wan et al. 2018). Besides, pseudogenes also occur abundantly in conifer genomes (Bautista et al. 2007, Kovach et al. 2010, Magbanua et al. 2011, Nystedt et al. 2013).

Gene duplication and multigene families

The gene families formed as a result of gene duplications, are much larger and with more members in conifers than the corresponding gene families in angiosperms. Multigene families have been indicated in *Pinus taeda* by some cDNA probes used for RFLP mapping which revealed more than one segregating locus (Devey et al. 1994). Some of the 32 such probes cross hybridized by Southern hybridization to 12 other conifer species show more than ten bands which suggests that these transcripts originated from members of multigene families (Ahuja et al. 1994). Many gene families showing differential expansion in gymnosperms as compared to angiosperms have been reported (De La Torre et al. 2020). Some of the better known gene families in conifers are alcoholdehydrogenase

(Adh) gene family (Perry and Furnier 1996), protochlorophyllide (POR) family (Skinner and Timko 1998), lipid transfer protein (nsLtp) genes (Kinlaw et al. 1994), Phenylalanine ammonia lyase (PAL) genes (Bagal et al. 2012), gene families concerning multiple defense mechanisms in *Ginkgo* (Guan et al. 2016), cellulose synthase genes in *Gnetum* (Wan et al. 2018).

Intron size

A positive direct relationship has been observed between intron length and genome size on a broad evolutionary scale (Vinogradov 1999) however, it may not be true when comparisons are made within narrow taxonomic groups (Wendel et al. 2002). Generally the studies done recently show that gymnosperm species have longer introns than angiosperms (Stival Sena et al. 2014) however, the impact of longer introns on the overall genome size is not well understood. Conifers are characterized by abundance of long introns. Introns of more than 20 kbp and upto 68 kbp in length are observed in *Picea abies* (Nystedt et al. 2013) and introns of average length of 2.4 kbp with a maximum length of 158 kbp have been found in *Pinus taeda* (Wegrzyn et al. 2014). Similarly average intron length of 2.3 kb in *Pseudotsuga menziesii* (Neale et al. 2017), 8.0 kb and 12.8 kb in *Pinus taeda* and *P. lambertiana* respectively have been reported (Stevens et al. 2016). The largest (892 kb) and second largest (578 kb) introns yet found in any plant genome are found in *Pinus taeda* and *P. lambertiana* respectively (Stevens et al. 2016). Similarly the largest intron size of 370 kb has been reported in *Picea glauca* (Warren et al. 2015). However, the mean and maximum intron lengths in *Abies alba* (mean=0.31 kb; maximum=36.01 kbp) and *Larix sibirica* (mean=0.35 kb; maximum=10.15 kb) are comparable and much smaller than in other conifers (Mosca et al. 2019). Furthermore, total intronic sequences per gene vary insignificantly between *Picea glauca* (3.13 kbp) and *Pinus taeda* (3.17 kbp) therefore suggesting that intron length is largely conserved in these species (Stival Sena et al. 2014). Other gymnosperm species such as *Ginkgo* also possesses long introns, where 12.45 % of total introns are more than 20 kb which is significantly higher than that of conifers (Guan et al. 2016). In fact LTR-RTs account for 59 % of the large intron sequences in *Ginkgo* (Guan et al. 2016). One species from Gnetales, *Gnetum montanum* with a much smaller genome size of 4.11 Gb/1C also shows significantly smaller minimum and maximum intron sizes as compared to conifers and *Ginkgo*. Another peculiarity is that the intron sequences of *G. montanum* are dominated by both long interspersed nuclear elements (LINEs) and LTR-RTs, unlike conifers and *Ginkgo* where LINEs are absent (Wan et al. 2018). When orthologous introns of *Pinus taeda* and *Ginkgo biloba* are compared the long introns (≥ 6 kb) of *P. taeda* are comparable with the orthologues of *G. biloba* suggesting that the amplification of LTR-RTs took place in some most recent ancestor. In contrast to this the comparison of such orthologues of *G. montanum* shows a remarkable reduction in intron size and a high proportion of LINEs similar to those *Amborella trichopoda*, which makes it different from other gymnosperms (Wan et al. 2018). Now the question arises whether there is any relationship between average intron length and genome size. *Gnetum*

montanum shows a significant reduction in average intron length, which is commensurate with its small genome size of 4.10 Gbp/1C (Wan et al. 2018). However, among conifers studied the mean and maximum intron lengths do not show a straight positive correlation with the genome size (Mosca et al. 2019). Similarly *Ginkgo biloba* which has a genome size of 10.0 pg/1C has an average intron size longer than that of *Picea abies* with a genome size of 20.35 pg/1C (Guan et al. 2016).

Whole genome duplication (Polyploidy)

The occurrence and role of polyploidy in gymnosperms has been addressed in another publication (Ohri 2021a). Nevertheless, we shall discuss the extent to which polyploidy has contributed to genome size evolution in this group of plants. The present data on chromosome numbers of 714 taxa including 628 species (Rastogi and Ohri 2020a, Ohri 2021a) shows that only 9.80 % of the species are recent polyploids which are present only in two families i.e. Ephedraceae and Cupressaceae.

Dysploid variation is present in *Zamia* as the chromosome number varies $2n=16-28$. As this range does not represent a geometric increase and is the result of chromosomal fission/fusion, therefore the presence of any whole genome duplications can be discarded (Gorelick et al. 2014, Zonneveld and Lindstrom 2016, Rastogi and Ohri 2020b). Furthermore, *Zamia* shows a very narrow range of 1C DNA values of 8.43-11.43, therefore showing a merely 1.36-fold variation which confirms that large diversity of chromosomes numbers in this genus is not correlated with genome size (Gorelick et al. 2014).

Recently it has been shown that polyploidy is viable and well maintained in *Ginkgo biloba* under cultivated condition as out of the 200 cultivars studied, thirteen were haploid, three triploid and eight tetraploid. The 1C DNA values varied proportionately to the ploidy levels ($1x=5.10$ Gbp, $2x=9.76$ Gbp, $3x=14.60$ Gbp, $4x=19.06-19.70$ Gbp) (Smarda et al. 2018). The slight variation in genome size in tetraploids has been explained to be due to some aneuploidy (Smarda et al. 2018).

Ephedra shows a high percentage of polyploidy as out of the 55 species studied, 18 (30.90 %) species possess both diploid and polyploid cytotypes and 24 (43.63 %) are exclusively polyploid and these species range from $2x-8x$. The largest 4.73-fold range in 1C DNA values among gymnosperms shown by *Ephedra* is due to polyploidy which extends up to octoploidy as this is further confirmed by only 1.37-fold difference in 1Cx values (Leitch et al. 2001, Ickert-Bond et al. 2020, Ohri 2021a) (Tables 1 and 2, Fig. 1).

In Cupressaceae, besides the classic examples of *Fitzroya* ($4x$) and *Sequoia* ($6x$) which show genome sizes commensurate with their ploidy levels, the genus *Juniperus* shows a very high frequency of polyploidy. Studies done on 97 taxa (total 75 species and 40 varieties) show that 22.30 % are polyploid and 11.60 % are exclusively polyploidy while 10.6 % show intraspecific polyploidy cytotypes (Ohri 2021a). The hexaploid cytotypes of two species i.e. *J. foetidissima* and *J. phoenicea* show 1C values of 35.66 pg and 36.80 pg respectively, and like *Ephedra* there is only 1.36-fold variation in 1Cx values which shows that the genome size has increased primarily by polyploidy (Farhat

et al. 2019 a,b, Ohri 2021a) (Table 1 and 2). Recently another natural tetraploid ($2n=44$) *Cupressus vietnamensis* (Farjon & T.H.Nguyễn) Silba (= *Xanthocyparis vietnamensis* Farjon & T.H.Nguyễn) has been reported which has 1C value of 22.33 pg (Farhat et al. 2021).

Conclusions

In contrast to the hardwoods which show extensive inter- and intraspecific polyploid series and palaeoploidy (Ohri 2015), gymnosperms possess a very narrow range of somatic chromosome numbers ($2n=14-66$), conserved karyotypes within various families and genera (Rastogi and Ohri 2020 a,b, Ohri 2021b) and rare instances of recent polyploidy (Khoshoo 1959, Ahuja 2005, Ohri, 2021a). Only some genera like *Zamia*, *Pseudotsuga*, *Pseudolarix*, *Prumnopitys*, *Halocarpus*, *Podocarpus* exhibit dysploid numbers (Rastogi and Ohri 2020a, Ohri 2021b). Out of 628 species (711 taxa) belonging to 81 genera known cytologically, multiple polyploid cytotypes and auto- or allopolyploids are known from *Ephedra*, *Ginkgo*, *Juniperus*, *Cupressus*, *Fitzroya*, and *Sequoia* with maximum ploidy reaching up to octoploidy in *Ephedra* (Rastogi and Ohri 2020a, Ohri 2021a). Nevertheless, gymnosperms show a much higher mean and a conserved narrow range of genome sizes as compared with angiosperms (Ohri and Khoshoo 1986, Murray 1998, Ohri 2005, Leitch and Leitch 2012, Pellicer et al. 2018). This difference in the range of genome sizes is reflected in the species diversity of angiosperms and gymnosperms (Puttik et al. 2015).

The question now arises how gymnosperms have acquired so big genomes and what evolutionary selective advantage they have contributed in their survival over a long geological period? From the above account it is evident now that the size and variation of the genomes of gymnosperms has been largely the result of transposon activity and their continual accretion, especially that of diverse LTR-RTs followed by the absence of efficient elimination mechanisms, resulting in low-copy soloLTRs as seen in conifer genomes (Kovach et al. 2010). The similarity of transposable elements in the sequenced conifer genomes suggests that there has been a massive expansion of TEs in the past during conifer evolution (Neale et al. 2014) and that these, mostly inactive elements, accumulate at a rate faster than their removal resulting in the massive genomes (Bennetzen et al. 2005, Nystedt et al. 2013, Kelly et al. 2015). So much so that this may seem like a one way ticket to genomic obesity (Bennetzen et al. 2005). It has been deduced that the lack of DNA loss is brought about by the epigenetic suppression of recombination processes. However, further studies show that various gymnosperm lineages have undergone expansion and contraction of the genome many times in their evolutionary past. This is best exemplified by *Gnetum* which has a much smaller genome size (2.25-4.11 Gb/1C) compared to rest of gymnosperms. Obviously *Gnetum* while having a similar profile of repeat accumulation, has lost most of the genomic features present in conifers and *Ginkgo* as is evident from the high ratio of soloLTR/LTR-RTs as a result of the elimination of

sequences by recombination based DNA break repair, reduction of the intron sizes, distinct patterns of proliferation of functional protein domains different from other gymnosperms and differential expansion and contraction of specific gene families (Wan et al. 2018). Schubert and Vu (2016) have in fact hypothesized that double strand break (DSB) (-mis)repair with bias toward deletions and insertions (retroelement dispersion) may lead to genome contraction in small and expansion in large genomes respectively, of these plants.

This is now clear that the unusually large genomes of conifers and *Ginkgo* have come into being because of the lack of efficient mechanisms to purge repetitive DNA sequences. But what factors and epigenetic machinery maintain the structure of these large genomes by silencing large stretches of repeats as heterochromatin. Further studies are needed to clarify the issues regarding the adaptive role of these silenced repeat sequences in gymnosperms (Liu et al. 2019).

Acknowledgements

This article is dedicated to my respected teacher Late Dr T.N. Khoshoo who introduced me to this subject. Thanks are due to the Pro-VC, Amity University U.P., Lucknow Campus for providing necessary facilities.

References

- Ahuja MR (2005) Polyploidy in gymnosperms revisited. *Silvae Genetica* 54: 59–69. <https://doi.org/10.1515/sg-2005-0010>
- Ahuja MR, Neale DB (2005) Evolution of Genome Size in Conifers. *Silvae Genetica* 54: 126–137. <https://doi.org/10.1515/sg-2005-0020>
- Ahuja MR, Devey ME, Grover AT, Jermstad KD, Neale DB (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. <https://doi.org/10.1007/bf00223632>
- Armenise L, Simeone M, Piredda R, Schirone B (2012) Validation of DNA barcoding as an efficient tool for taxon identification and detection of species diversity in Italian conifers. *Eur. J. Forest Res.* 131: 1337–1353. <https://doi.org/10.1007/s10342-012-0602-0>
- Aronen T, Ryyanen L (2012) Variation in telomere repeats of scots pine (*Pinus sylvestris* L.). *Tree Genetics and Genomes* 8: 267–275. <https://doi.org/10.1007/s11295-011-0438-7>
- Auckland L, Johnston J, Price H, Bridgewater F (2001) Stability of nuclear DNA content among divergent and isolated populations of Fraser fir. *Can. J. Bot.* 79: 1375–1378. <https://doi.org/10.1139/b01-104>
- Bagal UR, Leebens-Mack JH, Lorenz WW, Dean JF (2012) The phenylalanine ammonia lyase (PAL) gene family shows a gymnosperm-specific lineage. *BMC Genomics* 13: S1. <https://doi.org/10.1186/1471-2164-13-s3-s1>
- Baldwin BG, Sanderson MJ, Porter JM, Wojciechowski MF, Campbell CS, Donoghue MJ (1995) The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Ann. Mo. Bot. Gard.* 82: 247–277. <https://doi.org/10.2307/2399880>
- Bateman RM, Hilton J, Rudall PJ (2006) Morphological and molecular phylogenetic context of the angiosperms: Contrasting the 'top-down' and 'bottom-up' approaches used to infer the likely characteristics of the first flowers. *J. Exp. Bot.* 57: 3471–3503. <https://doi.org/10.1093/jxb/erl128>
- Bautista R, Villalobos DP, Díaz-Moreno S, Cantón FR, Cánovas FM, Claros MG (2007) Toward a *Pinus pinaster* bacterial artificial chromosome library. *Ann. For. Sci.* 64:855–864. <https://doi.org/10.1051/forest:2007060>
- Bennetzen JL, Ma J, Devos KM (2005) Mechanisms of recent genome size variation in flowering plants. *Ann. Bot.* 95:127–132. <https://doi.org/10.1093/aob/mci008>
- Berube Y, Zhuang J, Rungis D, Ralph S, Bohlmann J, Ritland K (2007) Characterization of EST SSRs in loblolly pine and spruce. *Tree Genet Genomes* 3:251–259. <https://doi.org/10.1007/s11295-006-0061-1>
- Bogunic F, Muratovic E, Brown SC, Siljak-Yakovlev S (2003) Genome size and base composition of five *Pinus* species from the Balkan region. *Plant Cell Reports* 22:59–63. <https://doi.org/10.1007/s00299-003-0653-2>
- Bogunic F, Muratovic E, Ballian D, Siljak-Yakovlev SS, Brown SC (2007) Genome size stability among five subspecies of *Pinus nigra* Arnold s.l. *Environmental and Experimental Botany* 59: 354–360. <https://doi.org/10.1016/j.envexpbot.2006.04.006>
- Bobola MS, Smith DE, Klein AS (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. <https://doi.org/10.1093/oxfordjournals.molbev.a040702>
- Brodribb TJ, Pitterman J, Coomes DA (2012) Elegance versus speed: examining the competition between conifer and angiosperm trees. *International Journal of Plant Sciences* 173: 673–694. <https://doi.org/10.1086/666005>
- Brown GR, Newton CH, Carlson JE (1998) Organization and distribution of a *Sau3A* tandem repeated DNA sequence in *Picea* (Pinaceae) species. *Genome* 41: 560–565. <https://doi.org/10.1139/g98-054>
- Cafasso D, Chinali G (2014) An ancient satellite DNA has maintained repetitive units of the original structure in most species of the living fossil plant genus *Zamia*. *Genome* 57: 125–135. <https://doi.org/10.1139/gen-2013-0133>
- Cafasso D, Cozzolino S, De Luca P, Chinali G (2003) An unusual satellite DNA from *Zamia paucijuga* (Cycadales) characterised by two different organizations of the repetitive unit in the plant genome. *Gene* 311: 71–79. [https://doi.org/10.1016/s0378-1119\(03\)00555-9](https://doi.org/10.1016/s0378-1119(03)00555-9)
- Cafasso D, Cozzolino S, Vereecken NJ, De Luca P, Chinali G (2009) Organization of a dispersed repeated DNA element in the *Zamia* genome. *Biologia Plantarum* 53: 28–36. <https://doi.org/10.1007/s10535-009-0005-3>
- Campbell CS, Wright WA, Cox M, Vining TF, Majorc CS, Arsenault MP (2005) Nuclear ribosomal DNA internal transcribed spacer 1 (ITS1) in *Picea* (Pinaceae): sequence divergence and structure. *Molecular Phylogenetics and Evolution* 35: 165–185. <https://doi.org/10.1016/j.ympev.2004.11.010>
- Chagne D, Chaumeil P, Ramboer A, Collada C, Guevara A, Cervera MT, Vendramin GG, Garcia V, Frigerio JMM, Echt C, Richardson T, Plomion C (2004) Cross-species transferability and mapping of genomic and cDNA SSRs in pines. *Theor. Appl. Genet.* 109:1204–1214. <https://doi.org/10.1007/s00122-004-1683-z>
- Chase MW, Reveal JL (2009) A phylogenetic classification of the land plants to accompany APGIII. *Bot. J. Linnean Soc.* 161: 122–127. <https://doi.org/10.1111/j.1095-8339.2009.01002.x>
- Christenhusz MJM, Reveal JL, Farjon A, Gardner MF, Mill RR, Chase MW (2011) A new classification and linear sequence of extant gymnosperms. *Phytotaxa* 19: 55–70. <https://doi.org/10.11646/phytotaxa.19.1.3>
- Christiakov DA, Hellemans B, Volckaert FAM (2006) Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. *Aquaculture* 255: 1–29. <https://doi.org/10.1016/j.aquaculture.2005.11.031>
- Cullis CA, Griessem GP, Gorman SW, Teasdale RD (1988) The 25S, 18S, and 5S ribosomal RNA genes from *Pinus radiata* D. Don. In: *Molecular Genetics of Forest Trees. Proc. 2nd Workshop IUFRO Working Party s2.04.06*. Chelak W M, Yapa A C (Eds). Canadian Forestry Service PNFI Inf. Rep. PI-X-80, pp.34–40.
- Davies BJ, O'Brien IEW, Murray BG (1997) Karyotypes, chromosome bands and genome size variation in New Zealand endemic gymnosperms. *Plant Systematics and Evolution* 208: 169–185. <https://doi.org/10.1007/bf00985440>
- De La Torre AR, Piot A, Liu B, Willhite B, Weiss M, Porth I (2020) Functional and morphological evolution in gymnosperms: A portrait of implicated gene families. *Evolutionary Applications* 13:210–227. <https://doi.org/10.1111/eva.12839>
- Devey ME, Fiddler TA, Liu BH, Knapp SJ, Neale DB (1994) An RFLP linkage map for loblolly pine based on three generation outbred pedigree. *Theor. Appl. Genet.* 88: 273–278. <https://doi.org/10.1007/bf00223631>

- Echt CS, May-Marquardt TP (1997) Survey of microsatellite DNA in pine. *Genome* 40: 9–17. <https://doi.org/10.1139/g97-002>
- Echt CS, Saha S, Krutovsky KV, Wimalanathan K, Erpelding JE, Chun Liang C, Nelson CD (2011) An annotated genetic map of loblolly pine based on microsatellite and cDNA markers. *BMC Genetics* 12:17. <https://doi.org/10.1186/1471-2156-12-17>
- Elsik CG, Williams CG (2000) Retroelements contribute to the excess of low-copy number DNA in pine. *Mol. Genet. Genomics* 264: 47–55. <https://doi.org/10.1007/s004380000279>
- Elsik CG, Williams CG (2001) Families of clustered microsatellites in a conifer genome. *Mol. Genet. Genomics* 265: 535–542. <https://doi.org/10.1007/s004380100443>
- Farhat P, Hidalgo O, Robert T, Siljak-Yakovlev S, Leitch I, Adams RP, Daghar Kharrat MB (2019a) Polyploidy in the genus *Juniperus*: an unexpectedly high rate. *Frontiers in Plant Science* 10: Article 676. <https://doi.org/10.3389/fpls.2019.00676>
- Farhat P, Siljak-Yakovlev S, Adams RP, Daghar Kharrat MB, Robert T (2019b) Genome size variation and polyploidy in the geographical range of *Juniperus sabina* L. (Cupressaceae). *Botany Letters*. <https://doi.org/10.1080/23818107.2019.1613262>
- Farhat P, Siljak-Yakovlev S, Hidalgo O, Rushforth K, Bartel JA, Valentin N, Leitch IJ, Adams RP (2021) Polyploidy in Cupressaceae: Discovery of a new naturally occurring tetraploid, *Xanthocyparis vietnamensis*. *Journal of Systematics and Evolution*. <https://doi.org/10.1111/jse.12751>
- Flanary BE, Kletetschka G (2005) Analysis of telomere length and telomerase activity in tree species of various life-spans, and with age in the bristlecone pine *Pinus longaeva*. *Biogerontology* 6:101–111. <https://doi.org/10.1007/s10522-005-3484-4>
- Fragniere Y, Betrisey S, Cardinaux L, Stoffe IM, Kozłowski L (2015) Fighting the last strand? A global analysis of the distribution and conservation status of gymnosperms. *Journal of Biogeography* 42: 809–820. <https://doi.org/10.1111/jbi.12480>
- Friesen N, Brandes A, Heslop-Harrison JS (2001) Diversity, origin, and distribution of retrotransposons (gypsy and copia) in conifers. *Molecular Biology and Evolution* 18: 1176–1188. <https://doi.org/10.1093/oxfordjournals.molbev.a003905>
- Friis EM, Pedersen KR, Crane PR (2010) Diversity in obscurity: Fossil flowers and the early history of angiosperms. *Philos. Trans. R. Soc. B Biol. Sci.* 365: 369–382. <https://doi.org/10.1098/rstb.2009.0227>
- Fuchs J, Brandes A, Schubert I (1995) Telomere sequence localization and karyotype evolution in higher plants. *Plant Systematics and Evolution* 196: 227–241. <https://doi.org/10.1007/bf00982962>
- García S, Ales Kova A, Leitch AR, Garnatje T (2017) Cytogenetic features of rRNA genes across land plants: analysis of the Plant rDNA database. *The Plant Journal* 89: 1020–1030. <https://doi.org/10.1111/tpj.13442>
- Germandt DS, Liston A, Pinerod (2001) Variation in the nrDNA ITS of *Pinus* subsection *Cembroides*: implications for molecular systematic studies of pine species complexes. *Molecular. Phylogenetics and Evolution* 21: 449–467. <https://doi.org/10.1006/mpev.2001.1026>
- Gorelick R, Fraser D, Zonneveld BJM, Little DP (2014) Cycad (Cycadales) chromosome numbers are not correlated with genome size. *Int. J. Plant Sci.* 175:986–997. <https://doi.org/10.1086/678085>
- Greilhuber J (2005) Intraspecific variation in genome size in angiosperms: identifying its existence. *Ann. Bot.* 95: 91–98. <https://doi.org/10.1093/aob/mci004>
- Greilhuber J, Leitch I (2013) Genome size and the phenotype. In: I.J. Leitch et al. (eds.), *Plant Genome Diversity Volume 2*, pp 323–344. https://doi.org/10.1007/978-3-7091-1160-4_20
- Grotkopp E, Rejmanek M, Rost TL (2002) Toward a causal explanation of plant invasiveness: seedling growth and life-history strategies of 29 pine (*Pinus*) species. *Am. Nat.* 159:396–419. <https://doi.org/10.2307/3079249>
- Grotkopp E, Rejmanek M, Sanderson MJ, Rost TL (2004) Evolution of genome size in pines (*Pinus*) and its life-history correlates: supertree analysis. *Evolution* 58: 1705–1729. <https://doi.org/10.1111/j.0014-3820.2004.tb00456.x>
- Guan R, Zhao Y, Zhang H, Fan G et al. (2016) Draft genome of the living fossil *Ginkgo biloba*. *GigaScience* 5: 49. <https://doi.org/10.1186/s13742-016-0154-1>
- Hall SE, Dvorak WS, Johnston JS, Price HJ, Williams CG (2000) Flow cytometric analysis of DNA content for tropical and temperate New World pines. *Ann. Bot. (Lond.)* 86:1081–1086. <https://doi.org/10.1006/anbo.2000.1272>
- Hamberger B, Hall D, Yuen M, Oddy C, Hamberger B et al. (2009) Targeted isolation, sequence assembly and characterization of two white spruce (*Picea glauca*) BAC clones for terpenoid synthase and cytochrome P450 genes involved in conifer defence reveal insights into a conifer genome. *BMC Plant Biol.* 9: 106. <https://doi.org/10.1186/1471-2229-9-106>
- Herbinger CM, Gordost K, Allen H (2011) Tetranucleotide and dinucleotide microsatellite markers for red spruce (*Picea rubens*). *The Americas Journal of Plant Science and Biotechnology* 5 (Sp. Issue 2): 105–111.
- Hidalgo O, Vallès J, Romo A, Canela MA, Garnatje T (2015) Genome size variation in gymnosperms under different growth conditions. *Caryologia*. 68: 92–96. <https://doi.org/10.1080/00087114.2015.1024546>
- Hill K (2005) Diversity and evolution of gymnosperms. In: Henry RJ (ed.), *Plant Diversity and evolution : Diversity and phenotypic Variation in Higher Plants* CABI Publishing Wallingford, Oxfordshire UK. <https://doi.org/10.1079/9780851999043.0025>
- Hizume M, Shibata F, Matsusaki Y, Kondo T (2000) Chromosomal localization of telomere sequence repeats in five gymnosperm species. *Chromosome Science* 4: 39–42
- Hizume M, Shibata F, Murayama Y, Kondo T (2001) Cloning of DNA sequences localized on proximal fluorescent chromosome bands by microdissection in *Pinus densiflora* Sieb. *Zucc. Chromosoma* 110: 345–351. <https://doi.org/10.1007/s004120100149>
- Hizume M, Shibata F, Matsusaki Y, Garajova Z (2002a) Chromosome identification and comparative karyotype analysis of *Pinus* species. *Theoretical and Applied Genetics* 105: 491–497. <https://doi.org/10.1007/s00122-002-0975-4>
- Hizume M, Shibata F, Matsumoto A, Maruyama Y, Hayashi E, Kondo T, Kondo K, Zhang S (2002b) Tandem repeat DNA localising on the proximal DAPI bands of chromosomes of *Larix*, Pinaceae. *Genome* 45:777–783. <https://doi.org/10.1139/g02-041>
- Hizume M, Shibata F, Kondo K, Hoshi Y, Kondo T, Ge S, Yang Q, Hong D (1999) Identification of chromosomes in two Chinese spruce species by multicolor fluorescence in situ hybridization. *Chromosome Sci.* 3: 37–41.
- Hung KH, Lin CY, Huang CC, Hwang CC, Hsu TW, Kuo YL, Wang WK, Hung CY, Chiang TY (2012) Isolation and characterization of microsatellite loci from *Pinus massoniana* (Pinaceae). *Botanical Studies* (2012) 53: 191–196.
- Ickert-Bond SM, Wojciechowski MF (2004) Phylogenetic Relationships in *Ephedra* (Gnetales): Evidence from Nuclear and Chloroplast DNA Sequence Data. *Systematic Botany* 29: 834–849. <https://doi.org/10.1600/0363644042451143>
- Ickert-Bond SM, Sousa A, Min Y, Loera I, Metzgar J, Pellicer J, Hidalgo O, Leitch I (2020) Polyploidy in gymnosperms-Insight into the genomic and evolutionary consequences of polyploidy in *Ephedra*. *Molecular Phylogenetics and Evolution* 147: 106786. <https://doi.org/10.1016/j.ympev.2020.106786>
- Joyner KL, Wang X-R, Johnston JS, Price HJ, Williams CG (2001) DNA content for Asian pines parallels New World relatives. *Canadian Journal of Botany* 79:192–196. <https://doi.org/10.1139/b00-151>
- Kamm A, Doudrick RL, Heslop-Harrison JS, Schmidt T (1996) The genomic and physical organization of *Ty1-Copia*-like sequences as a component of large genomes in *Pinus elliotii* var. *elliottii* and other gymnosperms. *Proceedings National Academy of Sciences USA* 93: 2708–2713. <https://doi.org/10.1073/pnas.93.7.2708>
- Kan XZ, Wang SS, Ding X, Wang XQ (2007) Structural evolution of nrDNA ITS in Pinaceae and its phylogenetic implications. *Molecular Phylogenetics and Evolution* 44:765–477. <https://doi.org/10.1016/j.ympev.2007.05.004>
- Kelly LJ, Renny-Byfield S, Pellicer J, Macas J, Novak P, Neumann P, Lysak MA, Day PD, Berger M, Fay MF, Nichols RA, Leitch AR, Leitch IJ (2015) Analysis of the giant genomes of *Fritillaria* (Liliaceae) indicates that a lack of DNA removal characterizes extreme expansions in genome size. *New Phytol* 208:596–607. <https://doi.org/10.1111/nph.13471>
- Kenrick P, Crane PR (1997) The origin and early evolution of plants on land. *Nature* 389: 33–39. <https://doi.org/10.1038/37918>
- Kejnovsky E, Hawkins J, Feschotte C (2012) Plant transposable elements: biology and evolution. In: Wendel JF (ed) *Plant genome diversity*, vol 1, *Plant genomes, their residents, and their evolutionary dynamics*. Springer, Wien, New York, pp 17–34. https://doi.org/10.1007/978-3-7091-1130-7_2

- Khoshoo TN (1959) Polyploidy in gymnosperms. *Evolution* 13: 24–39. <https://doi.org/10.1111/j.1558-5646.1959.tb02991.x>
- Kinlaw CS, Gertulla SM, Carter MC (1994) Lipid transfer protein genes of loblolly pine are members of a complex gene family. *Plant Molecular Biology* 26:1213–1216. <https://doi.org/10.1007/bf00040702>
- Kossack DS, Kinlaw CS (1999) IFG, a gypsy-like retrotransposon in *Pinus* (Pinaceae) has an extensive history in pines. *Plant Mol. Biol.* 39: 417–426. <https://doi.org/10.1023/a:1006115732620>
- Kovach A, Wegrzyn JL, Parra G, Holt C, Bruening GE et al. (2010) The *Pinus taeda* genome is characterized by diverse and highly diverged repetitive sequences. *BMC Genomics* 11: 420. <https://doi.org/10.1186/1471-2164-11-420>
- Kriebel HB (1985) DNA sequence components of the *Pinus strobus* nuclear genome. *Can. J. For. Res.* 15:1–4. <https://doi.org/10.1139/x85-001>
- Kubis S, Schmidt T, Heslop-Harrison JS (1998) Repetitive DNA elements as a major component of plant genomes. *Ann. Bot.* 82:45–56. <https://doi.org/10.1006/anbo.1998.0779>
- Kurdi-Haider B, Shalhoub V, Dib-Hajj S, Deeb S (1983) DNA sequence organization in the genome of *Cycas revoluta*. *Chromosoma* 88: 319–327. <https://doi.org/10.1007/bf00285854>
- Kuzmin D, Feranchuk S, Sharov VV, Krutovsky KV (2019) Stepwise large genome assembly approach: A case of Siberian larch (*Larix sibirica* Ledeb). *BMC Bioinformatics* 20(Suppl 1):37. <https://doi.org/10.1186/s12859-018-2570-y>
- L'Homme Y, Segun A, Trembley FM (2000) Different classes of retrotransposons in coniferous spruce species. *Genome* 43: 1084–1089. <https://doi.org/10.1139/g00-077>
- Leitch AR, Leitch IJ (2012) Ecological and genetic factors linked to contrasting genome dynamics in seed plants. *New Phytologist* 194:629–646. <https://doi.org/10.1111/j.1469-8137.2012.04105.x>
- Leitch IJ, Hanson L, Winfield M, Parker J, Bennett MD (2001) Nuclear DNA C-values complete familial representation in gymnosperms. *Annals of Botany* 88: 843–849. <https://doi.org/10.1006/anbo.2001.1521>
- Li YC, Korol AB, Fahima T, Nevo E (2004) Microsatellites within genes: structure, function, and evolution. *Mol Biol Evol* 21:991–1007. <https://doi.org/10.1093/molbev/msh073>
- Lin X, Faridi N, Casola C (2016) An Ancient Transkingdom Horizontal Transfer of Penelope. Like Retroelements from Arthropods to Conifers. *Genome Biol. Evol.* 8:1252–1266. <https://doi.org/10.1093/gbe/evw076>
- Liston A, Robinson WA, Oliphant JM, Alvarezbuilla ER (1996) Length variation in the nuclear ribosomal DNA internal transcribed spacer region of non-flowering seed plants. *Systematic Botany* 21: 109–120. <https://doi.org/10.2307/2419742>
- Liu W, Thummasuwan S, Sehgal SK, Chouvarine P, Peterson DG (2011) Characterization of the genome of bald cypress. *BMC Genomics* 12:553. <https://doi.org/10.1186/1471-2164-12-553>
- Liu Y, El-Kassaby YA (2019) Novel insights into plant genome evolution and adaptation as revealed through transposable elements and non-coding RNAs in conifers. *Genes* 10:228. <https://doi.org/10.3390/genes10030228>
- Magbanua ZV, Ozkan S, Bartlett BD, Chouvarine P, Saski CA et al. (2011) Adventures in the enormous: a 1.8 million clone BAC library for the 21.7 Gb genome of loblolly pine. *PLoS ONE* 6: e16214. <https://doi.org/10.1371/journal.pone.0016214>
- Maggini F, Baldassini S (1995) Ribosomal RNA genes in the genus *Pinus*. I. *Caryologia* 48: 17–25. <https://doi.org/10.1080/00087114.1995.10797314>
- Maomao Yan XD, Shuxian L, Tongming Y (2012) A meta-analysis of EST-SSR sequences in the genomes of pine, poplar and eucalyptus. *Tree Genetics and Molecular Breeding* 2:1–7. <https://doi.org/10.5376/tgmb.2012.02.0001>
- Miksche JP, Hotta Y (1973) DNA base composition and repetitious DNA in several conifers. *Chromosoma* 41: 29–36. <https://doi.org/10.1007/bf00284072>
- Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nature Genetics* 30: 194–200. <https://doi.org/10.1038/ng822>
- Morse AM, Peterson DG, Islam-Faridi MN, Smith KE, Magbanua Z, Garcia SA, Kubisiak TL, Amerson HV, Carlson JE, Nelson CD et al. (2009) Evolution of genome size and complexity in *Pinus*. *PLoS ONE* 4: e4332. <https://doi.org/10.1371/journal.pone.0004332>
- Mosca E, Cruz F, Gómez-Garrido J, Bianco L, Rellstab C, Brodbeck S, Csilléry K, Fady B, Fladung M et al. (2019) A Reference Genome Sequence for the European Silver Fir (*Abies alba* Mill.): A Community-Generated Genomic Resource. *G3 Genes/Genomes/Genetics* 9:2039. <https://doi.org/10.1534/g3.119.400083>
- Murray BG (1998) Nuclear DNA amounts in gymnosperms. *Annals of Botany* 82: 3–14. <https://doi.org/10.1006/anbo.1998.0764>
- Murray BG (2005) When does intraspecific C-value variation become taxonomically significant? *Annals of Botany* 95: 119–126. <https://doi.org/10.1093/aob/mci007>
- Murray BG (2013) Karyotype Variation and Evolution in Gymnosperms. In: I.J. Leitch et al. (eds.), *Plant Genome Diversity Volume 2*, pp. 231–242. https://doi.org/10.1007/978-3-7091-1160-4_14
- Murray BG, Friesen N, Heslop-Harrison JS (2002) Molecular cytogenetic analysis of Podocarpus and comparison with other gymnosperm species. *Annals of Botany* 89: 483–489. <https://doi.org/10.1093/aob/mcf047>
- Neale DB, Wegrzyn JL, Stevens KA, Zimin AV, Puiu D et al. (2014) Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biology* 2014, 15:R59. doi:10.1186/gb-2014-15-3-r59. <https://doi.org/10.1186/gb-2014-15-3-r59>
- Neale DB, McGuire PE, Wheeler NC, Stevens KA, Crepeau MW, Cardeno C, Zimin AV, Puiu D, Pertea GM, Sezen UU, Casola C, Koralewski TE, Paul R, Gonzalez-Ibeas D, Zaman S, Cronn R, Yandell M, Holt C, Langley CH, Yorke JA, Steven L, Salzberg SL, Jill L, Wegrzyn JL (2017) The Douglas-Fir Genome sequence reveals specialization of the photosynthetic Apparatus in Pinaceae. *G3 Genes/Genomes/Genetics* 7:3157. <https://doi.org/10.1534/g3.117.300078>
- Nunes JD, Torres GA, Davide LC, de Campos JMS (2009) Chromosome banding and DNA content in tropical *Pinus* species. *Scientia Forestalis*, Piracicaba 37: 213–218.
- Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin YC, Scofield DG, Vezzi F, Delhomme N, Giacomello S, Alexeyenko A et al. (2013) The Norway spruce genome sequence and conifer genome evolution. *Nature* 497: 579–584. <https://doi.org/10.1038/nature12211>
- Ohri D (1998) Genome size variation and plant systematic. *Annals of Botany* 82: 75–84. <https://doi.org/10.1006/anbo.1998.0765>
- Ohri D (2005) Climate and growth form: the consequences for genome size in plants. *Plant Biology* 7: 449–458. <https://doi.org/10.1055/s-2005-865878>
- Ohri D (2015) How small and constrained is the genome size of angiosperm woody species. *Silvae Genetica* 64: 20–32. <https://doi.org/10.1515/sg-2015-0002>
- Ohri D (2021a) Polyploidy in gymnosperms – a reappraisal. *Silvae Genetica* 70: 22–38. <https://doi.org/10.2478/sg-2021-0003>
- Ohri D (2021b) Karyotype evolution in conifers. *Feddes Repertorium*. <https://doi.org/10.1002/fedr.202100014>
- Ohri D, Khoshoo TN (1986) Genome size in gymnosperms. *Plant Systematics and Evolution* 153: 119–132. <https://doi.org/10.1007/bf00989421>
- Oliveira EJ, Pádua JG, Zucchi MI, Vencovsky R, Vieira MLC (2006) Origin, evolution and genome distribution of microsatellites. *Genetics and Molecular Biology* 29: 294–297. <https://doi.org/10.1590/s1415-47572006000200018>
- Paton AJ, Brummitt N, Govaerts R, Harman K, Hinchcliffe S, Allkin B, Lughadha EN (2008) Towards target 1 of the global strategy for plant conservation: A working list of all known plant species—progress and prospects. *Taxon* 57: 602–611.
- Pellicer J, Leitch IJ (2019) The Plant C-value data base (Release 7.1): an updated repository of plant genome size data for comparative studies. *New Phytologist* 226: 301–305. <https://doi.org/10.1111/nph.16261>
- Pellicer J, Hidalgo O, Dodsworth S, Leitch IJ (2018) Genome Size Diversity and Its Impact on the Evolution of Land Plants. *Genes* 9: 88. <https://doi.org/10.3390/genes9020088>
- Perera D, Magbanua ZV, Thummasuwan S, Mukherjee D, Ila MA, Chouvarine P, Nairn CJ, Schmutz J, Grimwood J, Deang JFD, Peterson DG (2018) Exploring the loblolly pine (*Pinus taeda* L.) genome by BAC sequencing and Cot analysis. *Gene* 663: 165–177. <https://doi.org/10.1016/j.gene.2018.04.024>
- Perry DL, Furnier GR (1996) *Pinus banksiana* has at least seven expressed alcohol dehydrogenase genes in two linked groups. *Proceedings National Academy of Sciences USA* 93: 13020–13023. <https://doi.org/10.1073/pnas.93.23.13020>
- Pfeiffer A, Olivieri AM, Morgante M (1997) Identification and characterization of microsatellites in Norway spruce (*Picea abies* K.). *Genome* 40:411–419. <https://doi.org/10.1139/g97-055>

- Puttick MN, Clark J, Donoghue PCJ (2015) Size is not everything: rates of genome size evolution, not C-value, correlate with speciation in angiosperms. *Proc. R. Soc. B* 282: 20152289. <http://dx.doi.org/10.1098/rspb.2015.2289>.
- Rake AV, Miksche JP, Hall RB, Hansen KM (1980) DNA reassociation kinetics of four conifers. *Canadian Journal of Genetics and Cytology* 22: 69–79. <https://doi.org/10.1139/g80-010>
- Ranade SS, Lin YC, Zuccolo A, Van de Peer Y, García-Gil M del R (2014) Comparative in silico analysis of EST-SSRs in angiosperm and gymnosperm tree genera. *BMC Plant Biology* 14:220. <https://doi.org/10.1186/s12870-014-0220-8>
- Rastogi S, Ohri D (2020a) Chromosome numbers in gymnosperms—An update. *Silvae Genetica* 69: 13–19. <https://doi.org/10.2478/sg-2020-0003>
- Rastogi S, Ohri D (2020b) Karyotype evolution in cycads. *Nucleus* 63: 131–141. <https://doi.org/10.1007/s13237-019-00302-2>
- Romo A, Hidalgo O, Boratynski A, Sobierajska K, Jasinka A, Vallès J, Garnatje T (2013) Genome size and ploidy levels in highly fragmented habitats: the case of western Mediterranean *Juniperus* (Cupressaceae) with special emphasis on *J. thurifera* L. *Tree Genetics and Genomes* 9: 587–599. <https://doi.org/10.1007/s11295-012-0581-9>
- Rueda M, Godoy O, Hawkins BA (2017) Spatial and evolutionary parallelism between shade and drought tolerance explains the distributions of conifers in the conterminous United States. *Global Ecology and Biogeography* 26: 31–42. <https://doi.org/10.1111/geb.12511>
- Rungis D, Berube Y, Zhang J, Ralph S, Ritland CE, Ellis BE, Douglas C, Bohlmann J, Ritland K (2004) Robust simple sequence repeat markers for spruce (*Picea* spp.) from expressed sequence tags. *Theoretical and Applied Genetics* 109:1283–1294. <https://doi.org/10.1007/s00122-004-1742-5>
- Salazar-Tortosa D, Castro J, Saladin B, Zimmermann NE, De Casas RR (2020) Arid environments select for larger seeds in pines (*Pinus* spp.). *Evolutionary Ecology* 34:11–26. <https://doi.org/10.1007/s10682-019-10016-1>
- Schmidt A, Doudrick RL, Heslop-Harrison JS, Schmidt T (2000) The contribution of short repeats of low sequence complexity to large conifer genomes. *Theoretical and Applied Genetics* 101: 7–14. <https://doi.org/10.1007/s001220051442>
- Schubert I, Vu GT (2016) Genome stability and evolution attempting a holistic view. *Trends in Plant Science* 21: 749–757. <https://doi.org/10.1016/j.tplants.2016.06.003>
- Seoane-Zonjic P, Cañas RA, Bautista R, Gómez-Maldonado J, Arrillaga I, Fernández-Pozo N, Claros MG, Cánovas FM, Ávila C (2016) Establishing gene models from the *Pinus pinaster* genome using gene capture and BAC sequencing. *BMC Genomics* 17:148. <https://doi.org/10.1186/s12864-016-2490-z>
- Shibata F, Hizume M (2008) Comparative FISH karyotype analysis of 11 *Picea* species. *Cytologia* 73: 203–211. <https://doi.org/10.1508/cytologia.73.203>
- Shibata F, Matsusaki Y, Hizume M (2005) AT-rich sequence containing *Arabidopsis* type telomere sequence and the chromosomal distribution in *Pinus densiflora*. *Theoretical and Applied Genetics* 110: 1253–1258. <https://doi.org/10.1007/s00122-005-1960-5>
- Skinner JS, Timko MP (1998) Loblolly pine (*Pinus taeda* L.) contains multiple expressed genes encoding light-dependent NADPH: protochlorophyllide oxidoreductase (POR). *Plant Cell Physiol.* 39: 795–806. <https://doi.org/10.1093/oxfordjournals.pcp.a029437>
- Smarda P, Horova L, Knapek O, Dieck H, Dieck M, Razna K, Hrubik P, Orloci L, Papp L, Vesela K, Vesely P, Bures P (2018) Multiple haploids triploids and tetraploids found in modern day 'living fossil' *Ginkgo biloba*. *Horticulture Research* 5:55. <https://doi.org/10.1038/s41438-018-0055-9>
- Smith DN, Devey ME (1994) Occurrence and inheritance of microsatellite loci in *Pinus radiata*. *Genome* 37: 977–983. <https://doi.org/10.1139/g94-138>
- Soranzo N, Provan J, Powell W (1998) Characterisation of microsatellite loci in *Pinus sylvestris* L. *Mol. Ecol.* 7:1260–1261
- Stevens KA et al. (2016) Sequence of the sugar pine megagenome. *Genetics* 204,1613–1626. <https://doi.org/10.1534/genetics.116.193227>
- Stival Sena J, Giguère I, Boyle B, Rigault P, Birol I, Zuccolo A, Ritland K, Bohlmann J, Jones S, Bousquet J, Mackay J (2014) Evolution of gene structure in the conifer *Picea glauca*: a comparative analysis of the impact of intron size. *BMC Plant Biology* 14:95. <http://www.biomedcentral.com/1471-2229/14/95>.
- Stuart-Rogers C, Flavell AJ (2001) The evolution of Ty1-copia group retrotransposons in gymnosperms. *Mol. Biol. Evol.* 18: 155–163. <https://doi.org/10.1093/oxfordjournals.molbev.a003789>
- The Plant List (2010) Version 1 Published on the Internet; <http://www.theplantlist.org/>
- Victoria FC, da Maia LC, de Oliveira AC (2011) In silico comparative analysis of SSR markers in plants. *BMC Plant Biology* 11:15.
- Vinogradov AE (1999) Intron-genome size relationship on a large evolutionary scale. *J. Mol. Evol.* 49: 376–384. <https://doi.org/10.1007/pl00006561>
- Von Stackelberg M, Rensing SA, Reski R (2006) Identification of genic moss SSR markers and a comparative analysis of twenty-four algal and plant gene indices reveal species-specific rather than group-specific characteristics of microsatellites. *BMC Plant Biology* 6:9. <https://doi.org/10.1186/1471-2229-6-9>
- Voronova A, Belevich V, Korica A, Rungis D (2017) Retrotransposon distribution and copy number variation in gymnosperm genomes. *Tree Genetics & Genomes* 13:88. <https://doi.org/10.1007/s11295-017-1165-5>
- Voytas DF, Cummings MP, Konieczny A, Ausubel FM, Roderick SR (1992) Copia-like retrotransposons are ubiquitous among plants. *Proceedings National Academy of Sciences USA* 89:7124–7128. <https://doi.org/10.1073/pnas.89.15.7124>
- Wakamiya I, Newton RJ, Johnston JS, Price HJ (1993) Genome size and environmental factors in the genus *Pinus*. *American Journal of Botany*. 80: 1235–1241. <https://doi.org/10.1002/j.1537-2197.1993.tb15360.x>
- Wakamiya I, Price HJ, Messina MG, Newton RJ (1996) Pine genome diversity and water relations. *Physiologia Plantarum* 96: 13–20. <https://doi.org/10.1034/j.1399-3054.1996.960103.x>
- Wan T, Liu ZM, Li LF, Leitch AR, Leitch IJ et al. (2018) A genome for gnophytes and early evolution of seed plants. *Nature Plants* 4: 82–89. <https://doi.org/10.1038/s41477-017-0097-2>
- Wang SQ, Ran JH (2014) Evolution and biogeography of gymnosperms. *Molecular Phylogenetics and Evolution* 75: 24–40. <https://doi.org/10.1016/j.ympev.2014.02.005>
- Warren RL, Keeling CI, Yuen MMS, Raymond A, Taylor GA et al. (2015) Improved white spruce (*Picea glauca*) genome assemblies and annotation of large gene families of conifer terpenoid and phenolic defense metabolism. *The Plant Journal* 83: 189–212.
- Wegrzyn J, Lin B, Zieve J, Dougherty W, Martinez-Garcia P, Koriabine M, Holtz-Morris A, de Jong P, Crepeau M, Langley C et al. (2013) Insights into the loblolly pine genome: Characterization of BAC and fosmid sequences. *PLoS ONE* 8: e72439. <https://doi.org/10.1371/journal.pone.0072439>
- Wegrzyn JL, Liechty JD, Stevens KA et al. (2014) Unique features of the loblolly pine (*Pinus taeda* L.) megagenome revealed through sequence annotation. *Genetics* 196: 891–909. <https://doi.org/10.1534/genetics.113.159996>
- Wendel JF, Cronn RC, Alvarez I, Liu B, Small RL, Sanchina DS (2002) Intron size and genome size in plants. *Mol. Biol. Evol.* 19: 2346–2352F <https://doi.org/10.1093/oxfordjournals.molbev.a004062>
- Won H, Renner SS (2005) The internal transcribed spacer of nuclear ribosomal DNA in the gymnosperm *Gnetum*. *Molecular Phylogenetics and Evolution* 36 : 581–597. <https://doi.org/10.1016/j.ympev.2005.03.011>
- Yagi E, Akita T, Kawahara T (2011) A novel Au SINE sequence found in a gymnosperm, *Genes Genet. Syst.* 86: 19–25. <https://doi.org/10.1266/ggs.86.19>
- Yazdani R, Scott I, Jansson G, Plomion C, Mathur G (2003) Inheritance and diversity of simple sequence repeat (SSR) microsatellite markers in various families of *Picea abies*. *Hereditas* 138: 219–227. <https://doi.org/10.1034/j.1601-5223.2003.01524.x>
- Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. *Mol Ecol* 11:1–16. <https://doi.org/10.1046/j.0962-1083.2001.01418.x>
- Zonneveld BJM, Lindstrom AJ (2016) Genome sizes for 71 species of Zamia (Cycadales: Zamiaceae) correspond with three different biogeographic regions. *Nordic J. Bot.* 34: 744–751. <https://doi.org/10.1111/njb.01094>