Genes & Genomics
https://doi.org/10.1007/s13258-020-00941-9

Online ISSN 2092-9293
Print ISSN 1976-9571

REVIEW

The C- and G-value paradox with polyploidy, repeatomes, introns, phenomes and cell economy

Ik-Young Choi[1] · Eun-Chae Kwon[2] · Nam-Soo Kim[2,3]

Received: 7 April 2020 / Accepted: 12 May 2020

Abstract

Background The apparent disconnection between biological complexity and both genome size (C-value) and gene number (G-value) is one of the long-standing biological puzzles. Gene-dense genomic sequences in prokaryotes or simple eukaryotes are highly constrained during selection, whereas gene-sparse genomic sequences in higher eukaryotes have low selection constraints. This review discusses the correlations of the C-value and G-value with genome architecture, polyploidy, repeatomes, introns, cell economy and phenomes.

Discussion Eukaryotic chromosomes carry an assortment of various repeated DNA sequences (repeatomes). Expansion of copies of repeatomes together with polyploidization or whole-genome duplication (WGD) are major players in genome size (C-value) bloating, but genomes are equipped with counterbalancing systems such as diploidization, illegitimate recombination, and nonhomologous end joining (NHEJ) after double-strand breaks (DSBs). The lack of these efficient purging systems allowed the accumulation of repeat DNA, which resulted in extremely large genomes in several species. However, the correlation between chromosome number and genome size is not clear due to inconsistent results with different sets of species. Positive correlations between genome size and intron size and density were reported in early studies, but these proposals were refuted by the results with increased numbers of species, in which genome-wide features of introns (size, density, gene contents, repeats) were weakly associated with genome size. The assumption of the correlations between C-value and gene number (G-value) and organismal complexity is acceptable in general, but this assumption is often violated in specific lineages or species, suggesting C- and G-value paradoxes. The C-value paradox is partly explained by noncoding repeatomes. The G-value paradox can also be explained by several genomic features: (1) one gene can produce many mature mRNAs by alternative splicing, and eukaryotic gene expression is highly regulated at both the transcriptional and translational levels; (2) many proteins exert multiple functions during development; (3) gene expansion/contraction are frequent events in the gene family among evolutionarily close species; and (4) sets of homeotic genes regulate development such that organismal complexity is sometimes not clear among organisms. A large genome must be burdensome in terms of cell economy, such that a large genome constraint results in the distribution of genome sizes skewed to small genomes. Moreover, the C-value can affect the phenome. A strong positive correlation has been recognized between genome size and cell size, but the relationship is weak or null with higher-level traits. Additional analyses of the relationship between the C-value and phenome should be carried out, because natural selection acts on the phenotype rather than the genotype.

Conclusions Dramatic advancement in genomics has given some answers to the C-value and G-value paradoxes. We know the mechanisms by which the current genomes have been constructed. However, basic questions have not yet been fully resolved. Why have some species retained small genomes yet some closely related species have large genomes? Random genetic drift and mutational pressure might have affected for genome size in the limited population size during evolution; thus, genome size may be quasiadaptable rather than the best adaptive trait.

Keywords C-value · G-value · Polyploidy · Repeatome · Intron · Cell economy · Phenome

✉ Nam-Soo Kim
kimnamsu@kangwon.ac.kr

Extended author information available on the last page of the article

Published online: 22 May 2020

Introduction

The amount of nuclear DNA, the C-value, is a characteristic of a species (Swift 1950). The C-value refers to the total amount of DNA in an unreplicated haploid or gametic nucleus of an organism (Greilhuber et al. 2005) and is often reported in picograms (pg) (1 pg ≅ 980 Mbp). In eukaryotes, the C-value is variable tremendously - as much as 66,000-fold - from the lowest of 0.0023 pg in *Encephalitozoon intestinalis*, a parasitic microsporodian, to 151.9 pg in *Paris japonica*, a monocotyledonous plant in the Liliales (Corradi et al. 2010; Pellicer et al. 2010). The major distinction between eukaryotic genomes and prokaryotic genomes involves genomic constraints. Genomic constraints in prokaryotes or simple eukaryotes are high, such that the genomes of these simple organisms are gene dense, whereas the genomic constraints of multicellular eukaryotes are low, such that their genomes are packed with repeated sequences, making them gene sparse (Koonin and Wolf 2010). Thus, the C-value is generally proportional to the organism's developmental complexity. However, this general propensity is often violated by a lack of apparent correlation between organismal complexity and genome size and large differences in the C-value among closely related species (Cavallier-Smith 1978; Gall 1981), which is referred to as the "C-value paradox" (Thomas 1971). For instance, the enormous genome of the whisk fern (*Tmesipteris obliqua*;1 C = 147.3 Gbp) (Hildago et al. 2017a) is approximately 46 times larger than the human genome size (1C = 3.2 Gbp) (Pennisi 2001). An example of genome size differences between closely related species is in the genus *Eleocharis* in the Poales of angiosperms. The genus *Eleocharis*, a sedge genus, contains approximately 250 species in which the genome of *E. acicularis* ($2n = 20$, 1C = 0.25 pg) is 20 times smaller than that of *E. palustris* ($2n = 16$, 1C = 5.5 pg) (Zedek et al. 2010). With accurate gene number estimation from whole-genomic sequencing of various eukaryotic organisms, Hahn and Wray (2002) coined the term "G-value" to designate the number of genes in a haploid genome and the term "I-value" for the amount of information encoded in a genome, which includes the number of genes and complexity added as a result of gene expression and interacting genes. They also coined the term "G-value paradox" to explain the explicit disconnections between the number of protein-coding genes and organismal complexity.

Next-generation sequencing (NGS) and analytically efficient bioinformatics tools have generated entire genomic sequences with highly accurate genetic information for many species (Park and Kim 2016; Straiton et al. 2019). One of the striking findings from NGS projects is that eukaryotic genomes are highly loaded with so-called "junk" sequences, which partly resolved the C-value paradox. However, unlocking the biological functions of the junk sequences is still a challenging project for understanding the evolutionary significance of genome evolution (Adelman and Egan 2017; Bernardi 2019). Genome size information has been acquired for more than 15,000 eukaryotic species, including plants (Plant C-value Database, www.cvalues.science.kew.org) (Pellicer and Leitch 2020), animals (Animal Genome Size Database, www.genomesize.com), and fungi (Fungal Genome Database, www.zbi.ee/fungal-genomesize), in recent decades. Although our understanding of genome architecture has dramatically increased because of both whole-genome sequence database information (https://www.ncbi.nlm.nih.gov/genome/browse#!/eukaryotes/; https://genome.jgi.doe.gov/portal/) and a wealth of genome size information, the fundamentals of genome evolution are not fully understood. Some species have streamlined genomes, but their closely related species have enormous genomes with high amounts of noncoding sequences. We can posit genomic theories to explain the C-value paradox with the knowledge of genome architecture in various types of organisms (Elliott and Gregory 2015). However, do contemporary genomes have evolutionarily inevitable outcomes? If so, do the genome sizes represent the best adaptive features for the extant species over evolutionary history? The current review provides recent updates on C-value genomics and evolutionary perspectives on eukaryotic genome size biology with an emphasis on plant genomes.

Eukaryotic chromosome architecture

The prominent Japanese geneticist Hitoshi Kihara coined the striking aphorism "*The history of the earth is recorded in the layers of its crust. The history of all organisms is inscribed in the chromosomes*" in the early 20th century (Crow 1994). His foresight without knowledge of molecular details on the chromosomes holds true even today in the genomic era. Eukaryotic chromosomes are now finely dissected at various molecular levels to enhance our understanding of the evolutionary history of organisms. Chromosomes are dynamic architectural structures to ensure that they pass their genetic integrity to daughter nuclei and regulate gene expression for cellular function (Bickmore 2001). To maintain genetic integrity generation after generation of cell division, chromosomes must have three basic elements: centromeres, telomeres, and replication origins.

Chromosomes consist of DNA and proteins that are collectively called chromatin. Genes are not evenly distributed along the chromosomes; genes are present in the loosely condensed euchromatic regions between highly condensed

heterochromatin blocks (Schimidt and Heslop-Harrison 1998; King 2002). Along with euchromatin and heterochromatin, chromosomes have other chromosomal landmarks, including centromeres, telomeres, and nucleolar organizing regions (NORs) (Fig. 1). Each chromosome is distinct in its shape by the location of the centromere and the distribution of euchromatin and heterochromatin. Moreover, heterochromatin is composed of a mixture of elements of repeated DNAs, such as minisatellites, simple sequence repeats (SSRs), and transposable elements (TEs) (Heslop-Harrison 2000). While highly repeated satellite DNA sequences and *Ty3/gypsy* long terminal repeat (LTR)-retrotransposons are packed in the centromeric regions, class 2 DNA transposons, *Ty1/copia* LTR retrotransposons, and SSRs are dispersed and often present in clusters (Schimidt and Heslop-Harrison 1998; Heslop-Harrison and Schimidt 2001). NORs are chromosomal sites that appear during secondary constriction in cytological preparations and are the sites where 18S, 5.8S, and 25S rRNA genes reside in tandem arrays of thousands of copies (Heslop-Harrison 2000). Another type of ribosomal RNA repeat is the 5S rRNA gene repeat, which is separately or closely located to NORs (Nguyen et al. 2016). The 5S rDNA genes are also repeated in tandem arrays of hundreds or thousands of copies (Cloix et al. 2002). Eukaryotic chromosomes are capped with telomeric repeats at both ends with many thousands of simple TTAGGG telomeric repeats whose main function is protecting chromosome integrity during cell division (McKnight and Shippen 2004). Other types of intercalary tandem or dispersed repeats are also scattered throughout the chromosomes.

If chromosomal DNA is stretched, the human genome (1C ≅ 3 Gb) is approximately 1.5 m, and the largest eukaryotic genome (that of *P. japonica*; 1C ≅ 148.8 Gb) is as long as 100 m; however, the eukaryotic nucleus is approximately 10 μm in diameter (Huber and Gerace 2007). Thus, packaging long DNA molecules into the small nucleus is highly challenging for eukaryotes; this is the primary function of chromatin. The chromosome structure is uneven in chromatin packaging such that the gene-rich euchromatic regions are relatively loosely packaged, but the heterochromatic gene-sparse regions are tightly packaged. The chromatin structure must also be able to be unpackaged during replication and gene transcription and then packaged again during cell division to be passed to daughter cells; thus, the dynamic regulation of chromatin structure is vital for successful survival throughout evolution of the species. The process of packaging and unpackaging chromosomal DNA is finely regulated by epigenetic mechanisms, which is beyond the scope of this review.

C-value and chromosome numbers

The haploid chromosome number is designated as *n*, which is a genetic characteristic of eukaryotes, and ranges from $n = 01$ in jack juniper ant (*Myrmecia pilosula*) (Crosland and Crozier 1986) to $n = 720$ in the monilophyte fern *Ophioglossum reticulum* (Khandelwal 2008). Chromosomes of polyploids will be addressed more thoroughly in relation to the C-value in the next section. Reports on the relationship between genome size and chromosome numbers are available with inconsistency. There was no clear relation between *n* number and C-value in the analysis of 343 taxa of Balkan flora by Siljak-Yakovlev and Pustahija

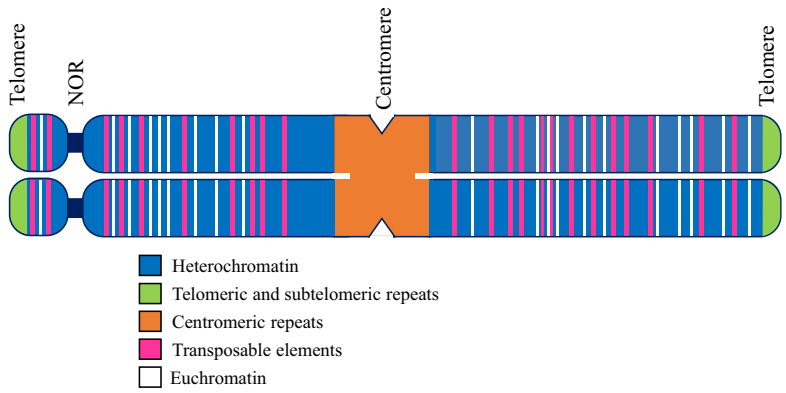

Fig. 1 Illustration of a mitotic metaphase chromosome. The chromosomes have reached their maximum condensed state and consist of two genetically identical sister chromatids. The centromere is the site of kinetochore formation where microtubules are attached to pull the sister chromatids to each pole. The centromeric region is highly saturated with *Ty1*-gypsy retroelements and other satellite DNA. Both ends of each chromosome are capped with thousands of copies of TTAGGG simple repeats whose function is protecting chromosome integrity during cell division. NOR is the site where hundreds to thousands of copies of 45S rDNA reside in a tandem array. Hundreds to thousands of 5S rDNA repeats can also be located elsewhere (not shown in the illustration). Heterochromatin contains various mixtures of repeat DNAs. Both class 1 and class 2 TEs are distributed along the chromosome. Euchromatin DNA is distributed between heterochromatin along the chromosomes, such as on islands in heterochromatic oceans

(2010). Pellicer et al. (2014) analyzed the genome size and chromosome evolution in the Melanthiaceae family of monocots; the haploid chromosome number ranged from 5 to 27, but the C-value was highly variable, as much as 230-fold among the species. For instance, the genera *Paris* and *Trillium* are $n = 05$, whereas the n numbers are variable in the genera *Heonias* ($n = 17$), *Stenanthium* ($n = 10$), and *Xerophyllum* ($n = 15$). The 1C value of the species in the genera *Paris* and *Trillium* ranged from 31.21 ~ 56.59 pg and 27.51 ~ 54.56 pg (excluding the tetraploid), respectively, whereas the 1C values of the species in the last three genera were approximately 3 pg, indicating that the chromosome number and genome size are negatively associated among the species in the family Melanthiaceae. Nishikawa et al. (1984) also reported a strong negative correlation between chromosome number and genome size among species in the genus *Carex* in the *Cyperaceae* family and argued that species with many small chromosomes are derived from a small number of large holocentric chromosomes by chromosome fragmentation followed by DNA loss. In combination with phylogenetic analysis with the chromosome numbers and C-values among *Carex* species, Chung et al. (2012) reported that the correlation was nearly zero or weakly positive or weakly negative at deeper phylogenetic scales. Thus, the authors postulated that the highly labile chromosome numbers might have affected reduced selection pressure for chromosome numbers in the *Carex* genus indirectly. However, the genus *Eleocharis* in the same *Cyperaceae* family showed a strong positive correlation between chromosome number and genome size in another study (Zedek et al. 2010). The chromosomes of both *Carex* and *Eleocharis* are holocentric and are easily breakable during cell division. The authors posited that the genome size variations in the *Eleocharis* species were derived from the occurrence of polyploidy and aneuploidy/symploidy with the amplification of LTR retrotransposons. The correlation between chromosome number and genome size was weakly negative ($r = -0.0187$) in their study. However, a weak but significantly positive correlation between the C-value and chromosome number was reported in a study involving more than 500 eukaryotic species ($r = 0.1456$, $p = 0.0076$) (Elliott and Gregory 2015). We retrieved the data of chromosome numbers and 1C values of diploids from the angiosperm genome size database (Bennett and Leitch 2011) and then analyzed the correlations between chromosome numbers and genome size. Of the 868 diploid species analyzed, the haploid chromosome number (n) ranged from 3 in *Hypochaeris oligocephala* to 43 in *Ceiba pentandra*, and the C-value ranged from 0.2 pg (*Arabidopsis thaliana* and 16 other species) to 77.4 pg in *Fritillaria koidzumiana*. The relationship between chromosome number and C-value revealed a relatively weak negative relationship ($r = -0.019$) in our analysis.

C-value and polyploidization

Whole-genome duplication (WGD; polyploidy) is an important driver during genome evolution. Polyploidization can cause doubling of the genome in autopolyploidization and the addition of two parental genomes in allopolyploidization. The WGD is followed by subsequent diploidization processes, including gene loss, genome fractionation, genome downsizing, and chromosome rearrangement (Wendel 2000). Polyploidy is common in plants but rare in animals such that only a few polyploid species exist (insects and reptiles) (Otto 2007), and ancestral vertebrates have undergone two rounds of WGD (Dehal and Boore 2005). However, polyploids are frequent in the plant kingdom; as much as approximately 70% of the extant plants are polyploids (Soltis et al. 2003). Virtually all plants have experienced one or more rounds of WGD, affecting both genome size and gene contents (Soltis and Soltis 2016; Clark and Donoghue 2018). While many rounds of WGD have been recorded in angiosperms, WGD in gymnosperms is rare, with only a few available reports, such as WGD in cycad and gingko (Roodt et al. 2017). *Amborella trichocarpa* is a basal angiosperm, and its genome has undergone at least two rounds of WGD (AGP 2013). In addition, the small genome of *Arabidopsis thaliana* has undergone at least two additional rounds of WGD since the divergence of eudicots (Bowers et al. 2003; del Pozo and Ramirez-Parra 2015).

Repeated rounds of WGD shaped the contemporary genomes of angiosperms, leading to inevitable genome size increases throughout evolution (Wendel 2015; Soltis and Soltis 2016). Then, is the genome size proportional to the number of WGDs? Are the genome sizes the sum of those of both parental species of allopolyploids? Genome size increases are not directly proportional to polyploidization events, such that the genome sizes of polyploids are usually smaller than expected, which is termed "genome downsizing" (Leitch and Bennet 2004; Doyle and Coate 2019). Here is a theoretical inference. The number of episodes of WGD was estimated to be as high as 288 in *Brassica napus* ($2n = 4x = 38$) and 144 in *Gossypium hirsutum* ($2n = 4x = 26$) (Wendel 2015). However, the genome size of ancestral angiosperms was estimated to be very small ($1C \leq 1.4$ pg) (Soltis et al. 2003). If so, genome sizes of the current *B. napus* and *G. hirsutum* species should be as large as > 400 pg and > 200 pg, respectively, instead of the actual C-values of 1.5 pg in *B. napus* and 2.4 pg in *G. hirsutum* (Bennet and Letich 2011), implying that there must be some counterbalancing system for genome size. Reduction in repetitive DNA sequences was posited as a main mechanism for genome downsizing

(Doyle and Coate 2019). Renny-Byfield et al. (2011) demonstrated the elimination of repetitive DNA sequences in the genome of *Nicotiana tabacum* ($2n = 4x = 48$), which is an allotetraploid derived from interspecific hybridization between *N. sylvestris* and *N. tomentosiformis*. Large amounts of *Ty3-gypsy* long terminal repeat (LTR) retroelements and 35S rDNA were eliminated from the *N. tabacum* genome compared with that of its parental species. Illegitimate or unequal recombination between LTR sequences accounted for the purging of the *Ty3-gypsy* elements in the synthetic allotetraploid *N. tabacum*. Reduction in rDNA sequences was also reported in several other allopolyploid species, including those of the *Brassica*, *Festuca*, *Glycine*, and Triticeae (Wendel 2000). By contrast, genome size increased in the species of the sunflower genus *Helianthus* after allopolyploidization (Ungerer et al. 2006). The genome sizes of the hybrid taxa *H. deserticola*, *H. anomalus*, and *H. paradoxus* were larger than the expected sum of their diploid parents *H. annuus* and *H. petiolaris*. The proliferation of *Ty1-gypsy* LTR retrotransposons by genome shock via interspecific hybridization was attributed to the differences of the hybrids in the sunflowers (Ungerer et al. 2006; Staton et al. 2012). In contrast to that in allopolyploids, genome downsizing data are limited in autopolyploids (Parisod et al. 2010). Raina et al. (1994) reported 17% of total DNA loss in synthetic autotetraploids of *Phlox drummondii* immediately after tetraploidization and further reduction (up to 25%) in the third generation. However, synthetic autotetraploids of *A. thaliana* revealed no DNA loss from the expected amount (Ozkan et al. 2006). There are 34 autotetraploid species in the 2221 angiosperm genome size database (Bennet and Leitch 2011). We manually checked the genome sizes of these autotetraploids, which revealed that six species had exactly doubled values from their diploids, whereas 20 showed a reduction in genome size compared with that of their diploids, but eight autotetraploids showed genome size increases (data not shown). Transposable elements were again posited to be responsible for the genome changes of the autotetraploids by genome shock, and thus, modulation of genome size should be considered part of the response to genome duplication (Doyle and Coate 2019).

C-value and introns

Eukaryotic genes are interrupted by introns that have to be removed from the primary transcript to form mature messenger RNA. Introns that do not encode proteins remain a debated issue. One speculation is that having introns might have driven eukaryotic evolution by enhancing coding capacity by alternative splicing (Kim et al. 2007; Nielson and Graveley 2010). However, transcribing and splicing introns requires energetic and time costs. For instance, the human dystrophin gene is 2300 kb, which encodes only a 14 kb mRNA, but the rest (99.4%) of this gene are 78 intronic sequences. Transcription of this gene takes up to 16 h, and splicing of a high number of introns also requires high amounts of cellular energy (Tennyson et al. 1995). Nevertheless, because introns are evolutionarily less constrained than are coding sequences, they usually evolve faster than coding sequences (Chamary and Hurst 2004).

Positive correlations between intronic features (average size, total number and nucleotide contents) and genome size are available in several reports over large evolutionary scales (Vinogradov 1999; Lynch and Conery 2003; Suetsuga et al. 2013; Elliott and Gregory 2015). Lynch (2007) reported that the amount of DNA in introns was nearly equal to the amount of DNA in exons in small genomes (< 200 Mbp), whereas intronic DNA occupied approximately 95% of the total length of protein-coding genes in large-genome-sized (> 2,500 Mbp) mammals. In the analysis of variable genome sizes of 12 animal species ranging from *C. elegans* (1C = 100.3 Mbp) to *Bombyx mori* (1C = 431.7 Mbp) to humans (1C = 3101.8 Mbp), Suetsuga et al. (2013) reported a strong positive correlation between intron size and genome size, with a correlation coefficient (CC) of $r = 0.942$; this value was higher than that (0.558) between genome size and TE contents in the genome, whereas the CC value between average exon and genome size was negative (−0.487). Elliott and Gregory (2015) analyzed associations of genome size with gene content, chromosome number, TE content, and intron features in 502 species, comprising 148 animals, 81 land plants, 202 fungi, and 70 protists. In their study, intron content and genome size were positively correlated across eukaryotes in terms of average intron size ($r = 0.6065$, $p < 0.0001$, $n = 245$ contrasts), the number of introns per genome ($r = 0.4535$, $p < 10^{-07}$, $n = 121$ contrasts) and the total amount of intronic DNA present in the genome ($r = 0.6079$, $p < 10^{-11}$, $n = 115$ contrasts). Francis and Wörheide (2017) presented interesting results on the introns in 68 species across 12 animal phyla, in which both introns and intergenic fractions displayed a linear correlation to total genome size, and the ratio of introns:intergenic regions approached 1:1 ($r^2 = 0.8286$, p value: 5.6×10^{-27}). These studies support the idea that intronic features scale linearly with genome size; larger genomes have more and longer introns, and *vice versa* for small genomes. However, opposite opinions also exist in several studies. Proponents of the latter theory argue that the concerted evolution between genome size and introns is weak or null throughout the Eukarya because recurrent intron loss/gain occurred at the lineage-specific level (Wang et al. 2014), a large portion of eukaryotic genomes lacks an organism-level function (Wang et al. 2014), or intron densities are variable across a wide range eukaryotic lineages (Farlow et al. 2011). Recently, Lozada-Chávez et al. (2020) carried out correlative association

studies of various genome-wide features of introns (size, density, genome content, repeats), genome size, and multicellular complexity of 461 eukaryotes. In their study, the intronic features were weakly correlated between themselves and genome size at a broad phylogenetic scale. The strength of the associations was variable at the lineage-specific level, and the variations in intron length and abundance within the genome were largely independent throughout the Eukarya. Their findings might be reasonable from the presumption that various kinds of repeat sequences, including TEs, are major drivers and that TEs can rapidly increase in number in specific lineages (Oliver et al. 2013; Pellicer et al. 2014; Blommaert et al. 2019).

C-value and repeatomes

Eukaryotic genomes teem with menagerie of repeated sequences, which are collectively called repeatomes (Maumus and Quesneville 2014). Along with polyploidization, the expansion of repeatomes is the main contributor to the large genome size variation among eukaryotes. There are two main types of repeats: tandem repeats and dispersed repeats. The tandem repeats include centromeric repeats, telomeric repeats, and ribosomal RNA (rDNA) genes, while dispersed repeats include simple sequence repeats (SSRs), minisatellites, and various kinds of transposable elements (TEs) (Heslop-Harrison and Schmidt 2001). Except for rDNAs, these repeated sequences are usually selectively neutral in terms of their accumulation in the genome, without major effects on the phenomes of the host; they were once considered "junk" DNA or "selfish" DNA (Doolittle and Sapienza 1980; Orgel and Crick 1980). However, the biological roles of this selfish DNA have been revisited, armed with a plethora genomic information; the unwelcome moniker has been changed to "genomic treasure" because this DNA has played major roles in shaping the current genomes and biodiversity (Volff 2006; Maumus and Quesneville 2014, 2016).

Of the various repeat sequences, TEs are major players in genome size variations by constituting a variably large proportion of eukaryotes, especially plants genomes (Kumar and Bennetzen 1999, 2000). TEs are classified into two classes on the basis of transposition mechanisms: class 1 and class 2 (Finnegen 1989). Class 1 TEs are retrotransposons that retrotranspose semiconservatively via mRNA intermediates in a "copy-and-paste" manner, whereas class 2 TEs are DNA transposons that transpose conservatively in a "cut-and-paste" manner. The content of TEs is generally proportional to the genome size; a small proportion of TEs exists in small genomes; a large proportion of TEs exists in large genomes (Civán et al. 2011). For instance, TEs constitute approximately 3% of the minute genome of the carnivorous bladderwort *Utricularia gibba* (1C = 0.079 pg) (Ibarra-Laclette et al. 2013), whereas TEs constitute > 85% of the maize genome (1C = 2.55 pg) (Schnable et al. 2009). However, no congruence in phylogenetic context with TE content has been reported in the fully sequenced genomes of 24 crop species, and the success of different types of TEs differs in different species (Vitte et al. 2014). For instance, *Ty1-gypsy* LTR retrotransposons are predominant in the genomes of maize and grapevine, whereas non-LTR retrotransposons are prevalent TEs in the genomes of sorghum, barley, potato and tomato.

The conservative cut-and-paste transposition mode does not usually allow high copies of class 2 DNA TEs; thus, they are present in moderate numbers, and their impact on genome size is not great compared with that of class 1 retrotransposons (Lee and Kim 2014). Class 1 retrotransposons can proliferate via very large copy numbers to cause genome bloating because the original copies are left behind in the copy-and-paste retrotransposition process (Bennetzen and Kellogg 1997). In the 50 fully sequenced plant genomes, the increase in genome size is correlated with the abundance of repeatomes ($r^2 = 0.584$) and, more specifically, LTR retrotransposons ($r^2 = 0.68$) (Michael 2014). Thus, it is generally accepted that there is a positive linear function between genome size and the content of TEs in eukaryotes (plants) in which class 1 L retrotransposons are major contributors to C-value differences (Kim 2017). For instance, the 17,000 Mbp wheat genome comprises 63.7% class 1 TEs and 14.9% class 2 TEs, and 2,300 Mbp of the maize genome constitutes 75.6% class 1 TEs and 8.6% class 2 TEs. Similarly, TEs constitute 68% of the large genome of *Secale cereale* (1C = 8.093 pg), in which class 1 and class 2 TEs constitute 64.3% and 5%, respectively (Oliver et al. 2013). The small genome of *Arabidopsis thaliana* (125 Mbp) comprises 7.5% of class 1 TEs and 11% of class 2 TEs. TE expansion has caused genome size variations in animals as well. For instance, rotifers of the *Brachionus plicatilis* species complex exhibit severalfold differences in genome size due to genome doubling and transposon expansion (Blommaert et al. 2019). Kapusta et al. (2017) also reported that many mammal and bird lineages have experienced different rates of TE accumulation, resulting in substantial variation in genome size between species.

If the copy-and-paste retrotransposition allows accumulation of class 1 retrotransposons in the genome, do the genomes become large only by one way? (Bennetzen and Kellogg 1997). The answer is 'no' because maintaining a large genome may be a burden to cell physiology, which will be discussed more in the chapter below. Analysis of the C-values of more than 6000 plant species (6287 angiosperms, 204 gymnosperms) revealed that plant genomes are skewed to small sizes (Civán et al. 2011). The C-values of 95% of the angiosperms are less than 22 Gbp, with a mean

of 5.809 Gbp and a median 2.401 Gbp, whereas those of 95% of gymnosperms are in the range of 7–33 Gbp, with a mean 18.157 Gbp and a median 17.506 Gbp. Hildago et al. (2017b) showed a violin plot of genome size distributions in flowering plants, ferns and vertebrates, including mammals, in which the genome sizes of all these biological domains were streamlined except for a few extraordinarily large genomes. Eukaryotic cells are equipped with mechanisms to counterbalance increasing genome size, such as illegitimate recombination (Devos et al. 2002; Hawkins et al. 2009) and nonhomologous end joining (NHEJ) after double-strand breaks (DSBs) (Chen et al. 2013; Fawcett et al. 2012; Lynch 2007). Unequal crossing over between repeat sequences leads to sequence deletion. Illegitimate intrastrand homologous recombination between direct repeat LTR sequences results in the deletion of sequences between LTRs, leading to solo LTRs. Devos et al. (2002) demonstrated that there was fivefold more illegitimate intrastrand recombination than unequal crossing over, which led to the small genome of *A. thaliana*. *Gossypium* (cotton) species carry *Geroge3*, a *gypsy*-like LTR retrotransposon, with variably high copy numbers among species (Hawkins et al. 2006). The copy number of *George3* increased in lineages specific to A- and K-genome diploids that have approximately 3 times larger genome sizes than D-genome diploids do, which have many more solo-LTR *George3* than the A- and K-diploids do, implying that intrastrand recombination purged *George3* copies in the D-genome species (Hawkins et al. 2009). NHEJ after DSB can also lead to the purging of LTR retrotransposons. For instance, the *Oryza brachyantha* genome is approximately 60% smaller than its close relative *O. sativa*, in which the amplification and deletion of recent LTR retrotransposons account for the difference. Comparison of protein-coding genes between the two species revealed that only 70% of the *O. brachyantha* genes were collinear with those of *O. sativa*. In this respect, the low LTR retrotransposon activity and massive amount of internal deletions of LTRs by NHEJ after DSB were proposed to cause the genome reduction in *O. brachyantha* (Chen et al. 2013). Removal of repeatomes may be a safeguard system in preventing uncontrolled genome expansion in combination with epigenetic regulation of TE activities (Slotkin and Martienssen 2007).

Why did evolution lead some species askew to have excessive large genomes? If closely related species with small and large genomes have similar DNA deletion systems, then the old repeats must have purged from both large and small genomes equally, but the species with large genomes must have undergone recent amplification of a few LTR retrotransposons (i.e., *George3* in *Gossypium* species) (Hawkins et al. 2009). Genome size and phylogenetic analyses have revealed that the lack of an efficient DNA removal system resulted in extreme expansion of the large genome of *Fritillaria* (Liliaceae) (Kelly et al. 2015). Studies on species with extreme genomes, such as lungfish (Metcalf et al. 2012), black salamander (Sun et al. 2012), and loblolly pine (Wegrzyn et al. 2014), have also revealed the presence of highly heterogeneous repeated DNA sequences.

C- and G-value paradox

The C-value paradox can be partly resolved by the bloating of noncoding DNA and polyploidy of some eukaryotic genomes, as mentioned above. It is hard to deny the general perception that gene number is roughly correlated with organismal complexity; however, it is also hard to accept this obsessive perception of linear correlation, because the gene number is lower in those developmentally complex organisms (i.e., mammals) than in simple organisms (i.e., ciliates of protists, many species of plants, and zebrafish of vertebrates). Table 1 shows the C- and G-values of 35 fully sequenced species from prokaryotes to eukaryotes. Both C- and G-values generally increased according to organismal complexity: prokaryotes < single cellular eukaryotes < multicellular eukaryotes. For instance, both the C- and G-values of single-celled yeast (*S. cerevisiae* and *S. pombe*) and the microsporodian fungus (*E. cuniculi*) were smaller than those of multi-cellular fungi (*N. crassa* and *U. maydis*). The genome sizes and gene numbers of protists are smaller than those of plants and animals. The C- and G-values of the moss *P. patens* are larger than those of the plant *A. thaliana*. If we consider G-values, the simple ciliate *Tetrahymena* has more genes than do developmentally complex organisms (i.e., Amborella, fruit fly, medaka fish, zebrafish, silkworm, etc.). The human G-value is dwarfed by that zebrafish, loblolly pine, wheat, soybean, and even ciliates. Thus, in parallel with the term 'C-value' paradox, the 'G-value' paradox was coined to account for the disconnection between the number of genes and organismal complexity (Hahn and Wray 2002). Indeed, eukaryotic genome size variation is approximately 66,000-fold, whereas the transcriptome difference is approximately 17-fold (Cavalier-Smith 2005).

Organismal complexity is somehow an illusory definition. Does the complexity mean the number of proteins produced or the number of cell types or organs? The proteome refers to the entire set of proteins that are expressed by a genome, cell, tissue, or organism under certain conditions (Altelaar et al. 2013). The one-gene one-protein concept is obsolete in modern genetics. The number of genes underestimates the proteome and developmental complexity because alternative splicing can produce several mature mRNAs. Approximately 95% of human multiexonic genes are alternatively spliced, and the specific mRNA from the alternative splicing of a gene is developmental or cell specific (Pan et al. 2008; David and Manley 2008). Many proteins have several cellular functions, and these 'Swiss army knife'-style proteins can also account

Table 1 C- and G-values of sorted organisms from prokaryotes to eukaryotes that have been fully sequenced

Taxonomic group	Species name	Common name	C-value (Mbp)	G-value (no. genes)	References
Eubacteria	*Escherichia coli*	Bacterium	4.64	4288	Blattner et al. (1997)
	Bacillus subtilis	Bacillus	4.21	4100	Kunst et al. (1997)
	Mycoplasma genitalium	Mycoplasma	0.58	470	Fraser et al. (1995)
Archaea	*Nanoarchaeum equitans*	Archaea	0.49	536	Waters et al. (2003)
	Methanosarcina barkeri	Archaea	4.83	3680	Maeder et al. (2006)
Fungi (yeast)	*Saccharomyces cerevisiae*	Baker yeast	12.1	6300	Goffeau et al. (1996)
	Schizosaccharomyces pombe	Fission yeast	13.8	4824	Wood et al. (2002)
	Neurospora crassa	Mold	40	10,082	Galagan et al. (2003)
	Encephalitozoon cuniculi	Microsporodian	2.9	1997	Katinka et al. (2001)
	Ustilago maydis	Smut fungus	20.5	6902	Kámper et al. (2006)
Protists	*Chlamydomonas reinhardtii*	Green alga	120	15,143	Merchant et al. (2007)
	Dictyostelium discoideum	Ameba	34	12,500	Eichinger et al. (2005)
	Tetrahymena discoideum	Ciliate	104	27,000	Eisen et al. (2006)
	Plasmodium falciparum	Protozoan	22.85	5286	Gardner et al. (2002)
	Trypanosome brucei	Trypanosome	26	9068	Berriman et al. (2005)
Plant	*Physcomitrella patens*	Moss	480	35,938	Rensing et al. (2008)
	Selaginella moellendorffii	Lycophyte	106	22,285	Banks et al. (2011)
	Amborella trichopoda	Amborella	870	26,846	AGP (2013)
	Arabidopsis thaliana	Thalecress	125	26,000	AGI (2000)
	Glycine max	Soybean	1,115	46,430	Schmutz et al. (2010)
	Brachypodium distachyon	Grass	313.6	25,532	TIBI (2010)
	Oryza sativa	Rice	460	61,668	Goff et al. (2002)
	Zea mays	Corn	2,650	32,000	Schnable et al. (2009)
	Lotus japonicus	Lotus	472	26,000	Sato et al. (2008)
	Triticum aestivum	Wheat	17,000	124,201	IWGSC (2014)
	Pinus taeda	Loblolly pine	23,000	50,172	Neale et al. (2014)
Animals	*Caenorhabditis elegans*	Worm	100.26	19,000	CSC (1998)
	Ciona intestinalis	Sea squirt	156	16,000	Dehal et al. (2002)
	Drosophila melanogaster	Fruit fly	137	13,600	Adams et al. (2000)
	Oryzias latipes	Medaka fish	800	20,141	Kasahara et al. (2007)
	Danio rerio	Zebra fish	1700	26,000	Howe et al. (2013)
	Fugu rubripes	Pufferfish	390	18,093	Brenner et al. (1993)
	Bombyx mori	Silkworm	530	11,202	Mita et al. (2004)
	Mus musculus	Mouse	2500	25,000	Guénet (2005)
	Homo sapiens	Human	3088	24,200	IHGSC (2001)

for the smaller-than-expected G-values of multicellular species (Hahn and Wray 2002). The expression of eukaryotic genes is finely regulated by sophisticated machinery (Krebs et al. 2018), and the development of multicellular organisms is regulated by a specific set of homeotic genes (Popodi et al. 1996). Additionally, noncoding RNAs regulate gene expression at both the transcriptional and posttranscriptional levels (Hirota et al. 2008; Palazzo and Lee 2015). Expansion of genes in multigene families has occurred differently in evolutionarily close species. For instance, olfactory receptor genes were identified to be present as 339 copies in humans (Malnic et al. 2004) but there are 1,296 copies in mice (Zhang and Firestein 2002). Thus, gene number may be related to organismal complexity in general, but we have to accept many exceptions for this general dogma. To account for the paradox in the correlation between C- and G-values and organismal complexity, the I-value was posited as a measure of the total information contained in a genome (Hahn and Wray 2002).

C-value and cell economy

It is obvious that a dramatic increase in noncoding or repeated sequences would be a burden to the host not only in terms of cellular physiology but also in terms of packaging DNA within a limited nuclear space; thus, eukaryotic genomes have been streamlined as much as possible (Cavalier-Smith 2005; Hildago et al. 2017b). Metabolic expense may be important to maintain and replicate the bulk noncoding DNA whose function is mostly unknown, which might be costly to the fitness of the host. Nuclear volume space doubles with genome doubling, but the surface area of the nuclear envelope increases only 1.6-fold (Melaragno et al. 1993), which can cause an imbalance in cellular factors mediating the interactions between chromosomes and nuclear components (Comai 2009). There is a strong nucleotypic effect on the cell cycle regardless of ploidy level in 100 plant species in which the C-value is positively related to cell cycle time (Francis et al. 2008). Knight et al. (2005) proposed that species with relatively small genomes presented higher growth rates than did those species with large genomes by facilitating fast cell divisions, so he posited the 'large genome constraints' theory to explain the physiological and metabolic costs associated with maintaining large genomes with excessive amounts of repeat DNA. The 'large genome constraint' theory explains the disadvantages of large genomes in terms of evolution, ecology, and physiology such that large genomes have diversified more slowly by being constrained, being underrepresented in extreme environments, and presenting reduced maximum photosynthetic rates; consequently, species with large genomes were trimmed from evolutionary trees and restricted in ecological distribution.

Reducing the genome size is a reasonable inference from the perspective of "large genome constraint", and the distribution of genome size is actually skewed to small sizes in all domains of eukaryotes (Oliver et al. 2007; Pellicer and Leitch 2020). C-values are less than 2 pg except in a few species within the tail region of those with large genomes in a graphical distribution of 6287 plant species (Civán et al. 2011). A strong correlation between cell size and genome size was observed in early studies in the 1950 and 1960 s (Mirsky and Ris 1951; Vialli 1957; Baetke et al. 1967). Then, have genomes become larger or smaller? Cavalier-Smith (1978) proposed that nuclear volume and genome size must be adjusted according to cell volume to allow reasonable growth rates, because DNA has two additional major functions in addition to encoding proteins, such as controlling cell volume by the number of replication origins and determining nuclear volume by the overall bulk of DNA. Nucleotides are charged solutes, and a large genome size decreases the osmotic potential of plant cells to draw more water into the cell, resulting in larger cells requiring more cellular and metabolic resources (Knight and Beaulieu 2008). Because nuclear DNA is encapsulated within the nuclear architecture, which is dynamically dissolved and reformed during the cell cycle, the amount of nuclear DNA is positively correlated with the volume of nuclear architecture. The intracellular parasite *Plasmodium* (microsporidia) has two nuclei with a normal large nucleus and secondary micronuclei (nucleomorphs) (Archibold and Lane 2009). The normal large nucleus shows a typical positive correlation between genome size and cell volume, but the small nucleomorph nuclei did not display an obvious correlation between them (Cavalier-Smith 2005). While the main nucleus allowed expansion of repeat DNA, minute nucleomorphs strongly decreased the genome size even by reducing gene sizes; thus, the author argued that the nuclear dimorphism of *Plasmodium* strongly supported the skeletal DNA/karyoplasmic ratio interpretation of genome size evolution, as economy, speed and size matter for evolutionary forces driving nuclear genome miniaturization and expansion. Furthermore, he refuted the previous idea of the correlation between cell cycle and nuclear DNA contents from the inference of small cells and rapid growth rates (Commoner 1964; Bennett 1972), because the relation between genome size and cell cycle length was much weaker than the relation between cell and nuclear volume (Cavalier-Smith 1978, 2005). However, this is disputable because many contrasting reports were put forward with respect to genome size and growth rates of plants (Suda et al. 2015; Roddy et al. 2020). Nevertheless, it still remains to be resolved why phylogenetically closely related species display many-fold differences in genome size.

Cell economy has slowed genome expansion so that most eukaryotes possess small genomes. However, the size of some genomes has skewed and expanded to an extraordinarily large size; this has occurred for *P. japonica* (1C = 148.8 Gbp) among angiosperms (Pellicer et al. 2010), *Tmesipteris obliqua* (1C = 147.3 Gbp) among whisk ferns (Hildago et al. 2017a), and *Protopterus aethiopicus* (lungfish, 130.0 Gbp) (Metcalf et al. 2012) and *Necturus lewisi* (salamander, 118.0 Gbp) (Sun et al. 2012) among vertebrates. If so, what is the biological upper limit of genome size? Hildago et al. (2017b) suggested that ~ 150 Gbp might be the biological upper limit for genome size. For this theory, they suggested several basic constraining factors, including biochemical and energetic costs, the maintenance of genome integrity, geometric constraints from a decreasing surface area-to-volume ratio of the cell as the genome size increases, timing constraints from longer mitosis and meiosis, and evolutionary constraints.

C-value and the phenome

Natural selection acts on phenotypes rather than genotypes. The phenome is a collective term describing the set of all phenotypes expressed by a cell, tissue, organ, organism or even species (Furbank and Tester 2011; Bush et al. 2016). Does genome size affect phenomes? A good example of the genome size effect on phenomes is the cell size of autopolyploids, which is discussed in detail in the literatures (Tsukaya 2013; Orr-Weaver 2015). The effect of genome/cell size has been documented in the salamander family Plethodontidae, which exhibits large genome variation from 1C to 15 pg (i.e., the genus *Desmognathus*) to ~120 pg (i.e., the genus *Necturus*) (Gregory 2005), and strong positive correlations were observed between C-values and blood cells as well as nucleus sizes among salamander species (Mueller et al. 2008). Such cases also occur in other fishes, birds and mammals (Gregory 2001, 2005).

As discussed in the section on the G-value paradox, gene contents are not greatly variable among species of taxa with different levels of biological complexities. The C-value paradox can be explained by the fact that larger genomes are packed more with selectively neutral repeat DNA than small genomes are. Then, is there any relationship between the C-value and phenome? Plants grow annually, biennially, or perennially. Annual or perennial growth might be another good example of how the C-value can affect the phenome. Table 2 shows the C-values of 2000 species of annual, biannual, and perennial plants collected by Bennet and Leitch (2011). The C-values of perennial plants were distributed mainly from 0.2 to 77.4 pg, but the distribution of biennials narrowed to 0.2–3.5 pg. The C-values of annuals ranged from 0.2 to 20.2 pg, which means that the plants with more than 20.2 pg are obligate perennials. By using regression analysis with 110 plant species, Francis et al. (2008) confirmed the general assumption that larger genomes take more time to multiply; there is a strong positive relationship between cell cycle time and C-value of diploids and polyploids, including for both monocots and dicots. The limited C-value of the perennials (20.2 pg) may imply that the large genomes place some selective disadvantage for plants that develop within only one growing season.

Table 2 Genome sizes (pg) of annual, perennial, and biannual angiosperm species

	Annual	Perennial	Biennial	All
Average	2.63 ± 2.74	6.43 ± 11.33	1.34 ± 1.23	5.74 ± 10.47
Median	1.50	2.20	0.70	1.90
Range	0.2–20.2	0.2–77.4	0.2–3.5	0.2–77.4

*C-value data from Bennet and Leitch (2011)

Previous observations revealed that there was a positive correlation between genome size and seed mass and various metrics of growth and leaf morphology characteristics of plants (Bennett 1971, 1972, 1987). However, Knight and Beaulieu (2008) reported somewhat different results, in which genome size was a strong predictor of phenotypic traits at the cellular level, but the power decreased for the higher-level phenotypes. There was a strong positive correlation between genome size and guard cell length and epidermal cell area and a negative correlation with stomatal density. However, the relationship was weak for the traits of the higher-level phenotypes (i.e., seed mass, leaf mass per unit area, wood density). Plant height was interesting: an increasing genome size decreases plant height among angiosperms, but it was reversed in gymnosperms, as species with larger genomes were taller. Similarly, a contrasting effect between angiosperms and gymnosperms was found for the relationship between genome size and pollen size (Knight et al. 2010). De Baedemaeker et al. (2018) reported that tetraploid apple trees tolerated drought better than did diploids; the authors speculated that the higher water content in leafy shoots, higher amount of parenchyma cells, and larger vessel area and size resulted in significantly higher hydraulic cavitation of the tetraploid plants. This might be important because global climate change is widely accepted among the public as well as within the scientific community. Global warming is obvious, and dry areas are rapidly expanding in many areas. Then, are species with large genomes better able to cope with environmental change? We do not provide any solid answers to this question. Species with small genomes may have traits conferring a growth advantage, such as longer dispersal of small pollen or seeds and shorter generation times, owing to the higher rates of cell division and efficiency in cell metabolism (Suda et al. 2015; Roddy et al. 2020). Many reports are available describing that species with smaller genomes are more invasive and successful in new habitats (Bennett et al. 1998; Pandit et al. 2014; Pysek et al. 2018). However, species can experience cellular shock in new environments, which can unlock the epigenetic suppression systems of TEs to propagate class 1 retroelements. TEs and epigenetic components are important environment-sensitive molecular elements, and coupling these two elements allows fine-tuning to adjust the production of phenotypes and genetic variations, including genome size (Rey et al. 2016). Li et al. (2018) reported differential expansion and contraction of the number of TEs among worldwide collections of *A. thaliana*, which might have played a role in their adaptive evolution. The genomes of salamanders are most variable among vertebrates, having from 13.89 pg to 120.60 pg and a mean of 35.35 pg per 1C (Lertzman-Lepofsky et al. 2019). The larval habitat of salamanders is either permanent aquatic

or ephemeral aquatic, or direct development occurs. While small-genome species are distributed across a gradient of ephemeral habitats, species with a larger genome are almost exclusively associated with a permanent aquatic habitat. Moreover, smaller-genome species showed a higher rate of evolutionary transition between permanent and ephemeral larval habitats. Thus, the authors proposed that the evolutionary constraint on the ecological habitat was imposed by the genome size of salamanders such that the species with large genomes were restricted to the permanent aquatic habitat due to their slower development.

Concluding remarks

The genome is defined as the whole set of genetic information of a species. There are highly diverse life forms on Earth, and all of them have their own genome. Like the immense biological diversity of life, eukaryotic genomes are also highly variable among species. Genome size (C-value) and gene content (G-value) are generally proportional to organismal complexity, except for a few outliers. The disconnection between gene number and biological complexity may be derived from highly complex gene expression regulation, multifunctional proteins, alternative splicing, multigene families, and developmental regulation by homoeotic gene sets. Polyploidy and TE expansion are two major players in genome size expansion, but cell economy has restricted genome size growth by the use of counterbalancing systems such as illegitimate recombination and NHEJ after DSB. Thus, genome size evolution follows a simple proportional model in which distribution is skewed to smaller genomes without invoking strong selection against large genomes. Nevertheless, a few species with extremely large genomes exist in which heterogeneous groups of repeat sequences accumulate to very high numbers of copies because they did not have efficient systems to remove repeated sequences from their genomes. Evolution is a stochastic process, and genome size is no exception from the many probabilistic events during selection. Because random genetic drift is a prominent evolutionary force within populations with limited size, substantial deviations are expected with high possibility of specific phylogenetic lineages whose genome size is prone to contraction/expansion; thus, genome size may be quasiadaptable rather than the best adaptive trait. Genome size affects various levels of phenomes, and genome size variations exist among species from different niches. In this respect, genome size is an important subject because many species are driven to new habitats from climate change.

Acknowledgements This work was supported by the National Research Foundation of Korea Grant (NRF-2016M3C9A4923797).

Compliance with ethical standards

Conflict of interest All authors, Ik-Young Choi, Eun-Chae Kwon, and Nam-Soo Kim, declare that they have no competing of interest.

Ethical approval This study does not contain any performing with human and animals.

References

Adams MD, Celniker SE, Holt RA, Evans CA, Cocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF et al (2000) The genome sequence of *Drosophila melanogaster*. Science 287:185–195. https://doi.org/10.1126/scienec287.5461.2185

Adelman K, Egan E (2017) More uses for genomic junk. Nature 543:183–185. https://doi.org/10.1083/543183a

AGI (Arabidopsis genome initiative) (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815. https://doi.org/10.1038/35048692

AGP (Amborella genome project) (2013) The *Amborella* genome and evolution of flowering plants. Science 342:1241089. https://doi.org/10.1126/science.1241089

Altelaar AF, Munoz J, Heck AJ (2013) Next generation proteomics: towards an integrative view of proteomics: towards an integrative view of proteome dynamics. Nat Rev Genet 14:35–48. https://doi.org/10.1038/nrg3356

Archibold JM, Lane CE (2009) Going, going, not quite gone: nucleomorphs as a case study in unclear genome reduction. J Hered 100:582–590. https://doi.org/10.1093/jhered/esp055

Baetke KP, Sparrow AH, Naumann CH (1967) The relationship of DNA content to nuclear and chromosome volumes and to radiosensitivity (LD50). Proc Natl Acad Sci USA 58:533–540

Banks JA, Nishiyama T, Hasebe M, Bowman JL, Gribskov M, dePamphilis C, Albert VA, Aoyama T, Ambrose BA et al (2011) The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants. Science 332:960–963. https://doi.org/10.1126/science.1203810

Bennet MD, Leitch IJ (2011) Nuclear DNA amounts in angiosperms: targets, trend and tomorrow. Ann Bot 107:467–590. https://doi.org/10.1093/aob/mcq258

Bennett MD (1971) The duration of meiosis. Proc R Soc Lond Ser B 178:277–299

Bennett MD (1972) Nuclear DNA content and minimum generation time in herbaceous plants. Proc R Soc Lond Ser B 181:109–135

Bennett MD (1987) Variation in genomic forms in plants and its ecological implications. New Phytol 106:177–200

Bennett MD, Leitch IJ, Hanson L (1998) DNA amounts in two samples of angiosperm weeds. Ann Bot 82:121–134

Bennetzen JL, Kellogg EA (1997) Do plants have a one-way ticket to genome obesity? Plant Cell 9:1509–1914. https://doi.org/10.1105/tpc.9.9.1509

Bernardi G (2019) The genomic code: a pervasive encoding/molding of chromatin structures and a solution of the "non-coding DNA" mystery. BioEssays 41:12 e1900106. https://doi.org/10.1002/bies.1900106

Berriman M, Ghedin E, Hertz-Fowler C, Renauld H, Bartholomeu DC, Lennard NJ, Hamlin NE, Haas B et al (2005) The genome of the African trypanosome *Trypanosoma brucei*. Science 309:416–422. https://doi.org/10.1126/science.1112642

Bickmore WA (2001) Eukaryotic chromosomes. eLS. https://doi.org/10.1038/npg.els.0001153

Blattner FR, Plukett G III, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF et al

(1997) The complete genome sequence of *Escherichia coli* K-12. Science 277:1453–1462. https://doi.org/10.1126/science.277.5331.1453

Blommaert J, Riss S, Hecox-Lea B, Welch DNM, Stelzer CP (2019) Small, but surprisingly repetitive genomes: transposon expansion and not polyploidy has driven a doubling in genome size in a metazoan species complex. BMC Genom 20:466. https://doi.org/10.1186/s12864-019-5859-y

Bowers JE, Chapman BA, Rong J, Paterson AH (2003) Unraveling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature 422:433–438. https://doi.org/10.1038/nature01521

Brenner S, Elgar G, Sanford R, Macrae A, Venkatesh B, Aparicio S (1993) Characterization of the pufferfish (Fugu) genome as a compact model vertebrate genome. Nature 366:265–268. https://doi.org/10.1038/366265

Bush WS, Oetjens MT, Crawford DC (2016) Unravelling the human genome-phenome relationship using phenome-wide association studies. Nat Rev Genet 17:129–145. https://doi.org/10.1038/nrg.2015.36

Cavalier-Smith T (1978) Nuclear volume control by nucleoskeletal DNA, selection for cell volume and cell growth rate, and the solution of the DNA C-value paradox. J Cell Sci 34:247–178

Cavalier-Smith T (2005) Economy, speed and size matter: evolutionary forces driving nuclear genome miniaturization and expansion. Ann Bot 95:147–175. https://doi.org/10.1093/aob/mci010

Chamary JV, Hurst LD (2004) Similar rates but different modes of sequence evolution in introns and at exonic silent sites in rodents: evidence for selectively driven codon usage. Mol Biol Evol 21:1014–1023. https://doi.org/10.1093/molbev/msh087

Chen J, Huang Q, Gao D, Wang J, Lang Y, Liu T, Li B, Bai Z, Luis G, Liang C et al (2013) Whole-genome sequencing of *Oryza branchyantha* reveals mechanisms underlying *Oryza* genome evolution. Nat Commun 4:1595. https://doi.org/10.1038/ncomms2596

Chung KS, Hipp AL, Roalson EH (2012) Chromosome number evolves independently of genome size in a clade with nonlocalized centromeres (Carex: Cyperaceae). Evolution 66–9:2708–2722. https://doi.org/10.1111/j.1558-5646.2012.01624x

Civán P, Svec M, Hauptvogel P (2011) On the coevolution of transposable elements and plant genomes. J Bot. https://doi.org/10.1155/2011/893546

Clark JW, Donoghue PCJ (2018) Whole-genome duplication in plant macroevolution. Trends Plant Sci 23:933–945. https://doi.org/10.1016/j.tplants.2018.07.006

Cloix C, Tutois S, Yukawa Y, Mathieu O, Cuvillier C, Espagnol MC, Picard G, Tourmenete S (2002) Analysis of the 5S RNA pool in *Arabidopsis thaliana*: RNAs are heterogenous and only two of the genomic 5S loci produce 5S RNA. Genome Res 12:132–144. https://doi.org/10.1101/gr.181301

Comai L (2009) The advantages and disadvantages of being polyploidy. Nat Rev Genet 6:836–846. https://doi.org/10.1038/nrg1711

Commoner B (1964) Roles of deoxyribonucleic acid in inheritance. Nature 202:960–968

Corradi N, Pombert JF, Farinelli L, Didier ES, Keeling PJ (2010) The complete sequence of the smallest known nuclear genome from the microsporidian *Encephalotozoon intestinalis*. Nat Commun 1:77. https://doi.org/10.1038/ncomms1082

Crosland MW, Crozier RH (1986) *Myrmecia pilosula*, an ant with only one pair of chromosomes. Science 231:1278. https://doi.org/10.1126/science.231.4743.1278

Crow JF (1994) Hitoshi Kihara, Japan's pioneer Geneticist. Genetics 137:891–894

CSC (, C. elegans, Sequencing Consortium) (1998) Genome sequence of the nematode *C. elegans*: a platform or investigating biology. Science 282:2012–2018. https://doi.org/10.1126/science.282.5396.2012

David CJ, Manley J (2008) The search for alternative splicing regulators: new approaches off a path to a splicing code. Genes Dev 22:279–285. https://doi.org/10.1101/gad.1643108

De Baerdemaeker NJF, Hias N, van den Bulcke J, Keulemans W, Steppe K (2018) The effect of polyploidization on tree hydraulic functioning. Am J Bot 105:161–171. https://doi.org/10.1002/ajb2.1032

Dehal P, Satou Y, Campbell RK, Chapman J, Degnan B, De Tomaso A, Davidson DM, Gregorio AD, Gelpke M, Goodstein DM et al (2002) The Draft genome of *Ciona intestinalis*: Insights into chordate and vertebrate origins. Science 298:2157–2167. https://doi.org/10.1126/science.0180049

Dehal P, Boore JL (2005) Two rounds of whole genome duplication in the ancestral vertebrates. PLoS Biol 3:e314. https://doi.org/10.1371/journal.pbio,0030314.g001

del Pozo JC, Ramirez-Parra E (2015) Whole genome duplication in plants: an overview from *Arabidopsis*. J Exp Bot 66:6691–7003. https://doi.org/10.1093/jxb/erv432

Devos KM, Brown JK, Bennetzen JL (2002) Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. Genome Res 12:1075–1079

Doolittle WF, Sapienza C (1980) Selfish genes, the phenotypic paradigm and genome evolution. Nature 284:601–603. https://doi.org/10.1038/284601a0

Doyle JJ, Coate JE (2019) Polyploidy, the nucleotype, and novelty: The impact of genome doubling on the biology of the cell. Int J Plant Sci 180:1–52. https://doi.org/10.1086/700636

Eichinger L, Pachebat JA, Glockner G, Rajandream M-A, Sucgan R, Berriman M, Song J, Olsen R, Szafranski K, Xu Q et al (2005) The genome of the social amoeba *Dictyostelium discoideum*. Nature 435:43–57. https://doi.org/10.1038/nature03481

Eisen JA, Coyne RC, Wu M, Wu D, Thiagarajan M, Wortman JR, Bajer JH, Ren Q, Amedeo P, Jones KM (2006) Macronuclear genome sequence of the ciliate *Tetrahymena thermophile*, a model eukaryote. PLoS Biol 4(9):e286. https://doi.org/10.1371/journal.pbio.0040286

Elliott TA, Gregory TR (2015) What's in the genome? The C-value enigma and evolution of eukaryotic genome content. Phil Trans R Soc Lond B Biol Sci 370(1678):20140331. https://doi.org/10.1098/rstb.2014.0331

Farlow A, Meduri E, Schlötterer C (2011) DNA double strand break repair and the evolution of intron density. Trend Genet 27:1–6. https://doi.org/10.1016/j.tig.2010.10.004

Fawcett JA, Rouze P, van der Peer Y (2012) Higher intron loss rates in *Arabidopsis thaliana* that *A. lyrate* is consistent with stronger selection for a small genome. Mol Biol Evol 29:849–859. https://doi.org/10.1093/molbev/msr254

Finnegen DJ (1989) Eukaryotic transposable elements and genome evolution. Trends Genet 5:103–107. https://doi.org/10.1016/0168-9525(89)90039-5

Francis D, Davies MS, Barlow PW (2008) A strong nucleotype effect on the cell cycle regardless of ploidy level. Ann Bot 101:747–757. https://doi.org/10.1093/aob/mcn038

Francis WR, Wörheide G (2017) Similar ratios introns to intergenic sequence across animal genomes. Genome Biol Evol 9:1582–1598. https://doi.org/10.1093/gbe/evx103

Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, Bult CJ, Kerlavage AR, Sutton G, Kelly JM et al (1995) The minimal gene complement of *Mycoplasma genitalium*. Science 27:397–404. https://doi.org/10.1126/science.270.5235.397

Furbank RT, Tester M (2011) Phenomics-technologies to relieve the phenotyping bottleneck. Trends Plant Sci 16:635–644. https://doi.org/10.1016/j.tplants.2011.09.005

Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, Jaffe D, Fitzhugh W, Ma LJ, Smirnov S et al (2003) The genome sequence of the filamentous fungus *Neuropora crassa*. Nature 422:859–868. https://doi.org/10.1038/nature01554

Gall JG (1981) Chromosome structure and the C-value paradox. J Cell Biol 91:3–14. https://doi.org/10.1083/jcb.91.3.3s

Gardner M, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S et al (2002) Genome sequence of the human parasite Plasmodium falciparum. Nature. https://doi.org/10.1038/nature01097

Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp Japonica). Science 296:92–100. https://doi.org/10.1126/science.1068275

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M et al (1996) Life with 6000 genes. Science 274:546–567. https://doi.org/10.1126/science.274.5287.546

Gregory TR (2001) The bigger the C-value, the larger the cell: genome size and red blood cell size in vertebrates. Blood cells Mol Dis 27:830–843. https://doi.org/10.1006/bcmd.2001.0457

Gregory TR (2005) Genome size evolution in animals. In: Gregory TR (ed) The evolution of the genome. Elsevier, San Diego, pp 3–87

Greilhuber J, Dolozel J, Lysác MA, Bennet MD (2005) The origin, evolution and proposed stabilization of the term 'genome size' and 'C-value' to describe nuclear DNA contents. Ann Bot 95:255–260. https://doi.org/10.1093/aob/mci019

Guenét JL (2005) The mouse genome. Genome Res 15:1729–1740. https://doi.org/10.1101/gr.3728305

Hahn MW, Wray GA (2002) The g-value paradox. Evol Dev 4(2):73–75. https://doi.org/10.1046/j.1525-142X.2002.01069

Hawkins JS, Kim HR, Nason JD, Wing RA, Wendel JF (2006) Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. Genome Res 16:1252–1261. https://doi.org/10.1101/gr.5282906

Hawkins JS, Proulx SR, Rapp RA, Wendel JE (2009) Rapid DNA loss as a counterbalance to genome expansion through retrotransposon proliferation in plants. Proc Natl Acad Sci USA 106:17811–17816. https://doi.org/10.1073/pnas.0904339106

Heslop-Harrison JS (2000) Comparative genome organization in plants: from sequences and markers to chromatins to chromosomes. Plant Cell 12:617–635. https://doi.org/10.1105/tpc.12.5.617

Heslop-Harrison JS, Schimidt T (2001) Plant nuclear genome composition. eLS. https://doi.org/10.1002/9780470015902.a0002014.pub2

Hildago O, Pellicer J, Chrstenhusz MJM, Schneider H, Leitch IJ (2017) Genomic gigantism in the whisk-fern family (Psilotaceae): *Tmesiptepris obliqua* challenges record holder *Paris japonica*. Bot J Linn Soc 183:509–514. https://doi.org/10.1093/botlinnean/box003

Hildago O, Pellicer J, Christenhusz M, Schneider H, Leitch AR (2017) Is there an upper limit to genome size? Trends Plant Sci 22:567–573. https://doi.org/10.1016/j.tplants.2017.04.005

Hirota K, Miyoshi T, Kugou K, Hiffman CS, Shibata T, Ohta K (2008) Stepwise chromatin remodeling by a cascade of transcription initiation of non-coding RNAs. Nature 456:130–134. https://doi.org/10.1038/natur07348

Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, Collins JE, Humphray S, McLaren K, Mattews L et al (2013) The zebra fish reference genome sequence and its relationship to the human genome. Nature 496:498–503. https://doi.org/10.1038/nature12111

Huber MD, Gerace L (2007) The size-wise nucleus: nuclear volume control in eukaryotes. J Cell Biol 19:583–584. https://doi.org/10.1083/jcb.200710156

Ibarra-Laclette E, Lyons E, Hernández-Guzmán G, Pérez-Torres CA, Carretoro-Paulet L, Chang T-H, Lan T, Welch AJ, Jazmín M et al (2013) Architecture and evolution of a minute plant genome. Nature 498:94–98. https://doi.org/10.1038/nature12132

IHGSC (International human genome sequencing consortium) (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921. https://doi.org/10.1038/35057062

IWGSC (International Wheat Genome Sequencing Consortium) (2014) A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. Science 345:1251788. https://doi.org/10.1026/science.1251788

Kámper J, Kahmann R, Bölker M, Ma LJ, Brefort T, Saville BJ, Banuett F, Kronstad JW, Gold SE, Müller O, Perlin MH et al (2006) Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. Nature 444:97–101. https://doi.org/10.1038/nature05248

Kapusta A, Suh A, Feschotte (2017) Dynamics of genome size evolution in birds and mammals. Proc Natl Acad Svi USA. https://doi.org/10.1073/pnas.1616702114

Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamaa T, Nagayasu Y, Doi K, Kasai Y et al (2007) The medaka draft genome and insights into vertebrate genome evolution. Nature 447:714–719. https://doi.org/10.1038/nature05846

Katinka MD, Duprat S, Cornillot E, Méténier G, Thomarat F, Prensier G, Barbe V, Peyretallade EC, Brottier P, Wincker P et al (2001) Genome sequence and gene compaction of the eukaryote parasite. Encephalitozoon cuniculi Nature 414:450–453. https://doi.org/10.1038/35106579

Kelly LJ, Renny-Byfield S, Pellicer J, Macas J, Neumann P, Lysak MA, Day PD, Berger M, Fay MF, Nichols R et al (2015) Analysis of the giant genomes of *Fritillaria* (Liliaceae) indicates that a lack of DNA removal characterizes extreme expansions in genome sizes. New Phytol 208:596–607. https://doi.org/10.1111/nph.13471

Khandelwal S (2008) Chromosome evolution in the genus *Ophioglossum* L. Bot J Linn Soc 102:205–217. https://doi.org/10.1111/j.1095-8339.1990.tb01876.x

Kim E, Magen A, Ast G (2007) Different levels of alternative splicing among eukaryotes. Nucl Acids Res 35:125–131. https://doi.org/10.1093/nar/gkl924

Kim NS (2017) The genomes and transposable elements in plants: are they friends or foes. Genes Genom 39:359–370. https://doi.org/10.1007/s13258-017-0522-y

King GJ (2002) Through a genome, darkly: comparative analysis of plant chromosomal DNA. Plant Mol Biol 48:5–20

Knight CA, Molinari NA, Petrov DA (2005) The large genome constraint hypothesis: evolution, ecology and phenotype. Ann Bot 95:177–190. https://doi.org/10.1093/aob/mci011

Knight CA, Beaulieu JM (2008) Genome size scaling through phenotype space. Ann Bot 101:759–766. https://doi.org/10.1093/aob/mcm321

Knight CA, Clancy RB, Gotezenberger L, Dann L, Beaulieu JM (2010) On the relationship between pollen size and genome size. J Bot. https://doi.org/10.1155/2010/612017

Koonin EV, Wolf YI (2010) Constraints and plasticity in genome and molecular-phenome evolution. Nat Rev Gent 11:487–498. https://doi.org/10.1038/nrg2810

Krebs JE, Goldstein ES, Kilpatrick ST (2018) Lewin's Genes III. Burlington. ISBN 978-1-284-10449-3

Kumar A, Bennetzen JL (1999) Plant retrotransposons. Ann Rev Genet 33:479–532. https://doi.org/10.1146/annurev.genet.33.1.479

Kumar A, Bennetzen JL (2000) Retrotransposons: central players in the structure, evolution and function in plant genomes. Trends Plant Sci 5:509–510. https://doi.org/10.1016/s1360-1385(00)01760-x

Kunst F, Borcehrt S et al (1997) The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. Nature 390:249–256. https://doi.org/10.1038/36786

Lee SI, Kim NS (2014) Transposable elements and genome size variations in plants. Genomics Inform 12:87–97. https://doi.org/10.5808/GI.2014.12.3.87

Leitch IJ, Bennet MD (2004) Genome downsizing in polyploid plants. Biol J Linn Soc 82:651–663. https://doi.org/10.1111/j.1095-8312.2004.00349

Lertzman-Lepofsky G, Mooers A, Greenberg DA (2019) Ecological constrains associated with genome size across salamander lineages. Proc R Soc B 286:20191780. https://doi.org/10.1098/rspb.2019.1780

Li ZW, Hou XU, Chen JF, Xu YC, Wu Q, Gonzalez J, Guo YL (2018) Transposable elements contribute to the adaptation of Arabidopsis thaliana. Genome Biol Evol 10:2140–2150. https://doi.org/10.1093/gbe/evy171

Lozada-Chávez I, Stadeler PF, Prohaska SJ (2020) Genome-wide features of introns are evolutionary decoupled among themselves and from genome size throughout Eukarya. BioRxiv. https://doi.org/10.1101/283549

Lynch M, Conery JS (2003) The origin of genome complexity. Science 302:1401–1404. https://doi.org/10.1126/science.1089370

Lynch M (2007) Genome size and organismal complexity. 'The origin of genome architecture.' Sinauer, Sunderland, pp 29–42

Maeder DL, Anderson I, Brettin TS, Bruce DC, Gilna P, Han CS, Lapidus A, Metcalf WW, Saunders E, Tpoia R et al (2006) The *Methanosarcinabarkeri* genome: Comparative analysis with *Methaosarcina acetivorans* and *Methanosarcina mazei* reveals extensive rearrangement with Metahnosarcinal genome. J Bact 188:7922–7931. https://doi.org/10.1128/JB.00810-06

Malnic B, Godfrey PA, Buck LB (2004) The human olfactory receptor gene family. Proc Natl Acad Sci USA 101:2584–2589. https://doi.org/10.1073/pnas.0307882100

Maumus F, Quesneville H (2014) Deep investigation of *Arabidopsis thaliana* junk DNA reveals a continuum between repetitive elements and genomic dark matter. PLoS One 94(4):e94101. https://doi.org/10.1371/journal.pone.0094101

Maumus F, Quesneville H (2016) Impact and insights from ancient repetitive elements in plant genomes. Curr Opin Plant Biol 30:41–46. https://doi.org/10.1016/j.pbi.2-16.01.003

McKnight TD, Shippen DE (2004) Plant telomere biology. Plant Cell 16:794–803. https://doi.org/10.1105/tpc.160470

Merchant SS, Prochnik SE, Vallon O, Harfris EH, Karpowicz SJ, Witman GB, Terry A, Salamaov A, Fritz-Laylin LK, Marechal-Drouard L et al (2007) The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. Science 318:245–250. https://doi.org/10.1126/science.1143609

Merlaragno JE, Mehrotra B, Coleman AW (1993) Relationship between endopolyploidy and cell size epidermal tissue of Arabidopsis. Plant Cell 5:1661–1668. https://doi.org/10.1105/tpc.5.11.1661

Metcalf CJ, Filee J, Germon J, Joss J, Casane D (2012) Evolution of the Australian lungfish (*Neoceratodus forsteri*) genome: a major role for CR1 and L2 LINE elements. Mol Biol Evol 29:3529–3539. https://doi.org/10.1093/molbev/mss159

Michael TP (2014) Plant genome size variation: bloating and purging DNA. Brief Funct Genom 13:308–317. https://doi.org/10.1093/bfgp/elu005

Mirsky AE, Ris H (1951) The DNA content of animal cells and its evolutionary significance. J General Physiol 34:451–462

Mita K, Kasahara M, Sasaki S, Nagayasu Y, Yamada T, Kanamori H, Namiki N, Kitagawa M, Yamashita H, Yasukochi Y, Kadono-Okuda K et al (2004) The genome sequence of silkworm, Bombyx mori. DNA Res 29:27–35. https://doi.org/10.1093/dnares/11/1/27

Mueller RL, Gregory TR, Gregory SM, Hsieh A, Boore JL (2008) Genome size, cell size, and evolution of enucleated erythrocytes in attenuated salamanders. Zoology 111:218–230. https://doi.org/10.1016/j.zool.2007.01.010

Neale DB, Wegrgyn JL, Stevens K, Zimin AV, Puiu D, Crepeau MW, Careno C, Korabine M, Holtz-Morris AE, Liechty JD et al (2014) Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. Genome Biol 15:R59. https://doi.org/10.1186/gb-2014-15-3-r59

Nguyen TX, Lee SI, Rai R, Kim NS, Kim JH (2016) Ribosomal DNA locus variation and REMAP analysis of the diploid and triploid complexes of *Lilium lancifolium*. Genome 59:551–564. https://doi.org/10.1139/gen-2016-0011

Nielson TW, Graveley BR (2010) Expansion of the eukaryotic proteome by alternative splicing. Nature 463:457–463. https://doi.org/10.1038/nature08909

Nishikawa K, Furuta Y, Ishitobi K (1984) Chromosome evolution in the genus *Carex* as viewed from nuclear DNA content, with special reference to its aneuploidy. Jap J Genet 59:465–472

Oliver MJ, Petrov D, Ackerly D, Falkowski PF, Schofield OM (2007) The mode and tempo of genome size evolution in eukaryotes. Genome Res 17:594–601. https://doi.org/10.1101/gr.6096207

Oliver KR, McComb JA, Greene W (2013) Transposable elements: powerful contributors to angiosperm evolution and diversity. Genome Biol 5:1886–1901. https://doi.org/10.1093/gbe/evt141

Orgel LE, Crick FH (1980) Selfish DNA: the ultimate parasite. Nature 284:604–607. https://doi.org/10.1038/284604a0

Orr-Weaver TL (2015) When bigger is better: The role of polyploidy in organogenesis. Trends Genet 31:307–315. https://doi.org/10.1016/j.tog.2015.03.011

Otto SP (2007) The evolutionary consequences of polyploidy. Cell 131:452–462. https://doi.org/10.1016/j.cell.2007.10.022

Ozkan H, Tuna M, Galbraith DW (2006) No DNA loss in autotetraploids of *Arabidopsis thaliana*. Plant Breed 125:288–291

Palazzo AF, Lee ES (2015) Non-coding RNA: what is functional and what is junk? Front Genet 6:2. https://doi.org/10.3389/fgene.2015.00002

Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet 40:1413–1415. https://doi.org/10.1038/ng.259

Pandit MK, White SM, Pocock MJO (2014) The contrasting effects of genome size, chromosome number and ploidy level on plant invasiveness: a global analysis. New Phytol 203:697–703. https://doi.org/10.1111/nph.12799

Parisod C, Holdergger R, Brochmann C (2010) Evolutionary consequences of autopolyploids. New Phytol 186:5–17. https://doi.org/10.1111/j.1469-8137.2009.03142.x

Park ST, Kim J (2016) Trends in Next genration sequencing and a new era for whole genome sequencing. Int Neurorol J 20:S76–S83. https://doi.org/10.5213/inj.1632742.371

Pellicer J, Fay MF, Leitch IJ (2010) The largest eukaryotic genome of them all? Bot J Linn Soc 164:10–15. https://doi.org/10.1111/j.1095-8339.2010.01072.x

Pellicer J, Kelly LJ, Leitch IJ, Zomlefer WB, Fay MF (2014) A universe of dwarf and giants: genome size and chromosome evolution in the monocot family Melanthiaceae. New Phytol 201:1484–1497. https://doi.org/10.1111/nph.12617

Pellicer J, Leitch IJ (2020) The plant DNA C-values database (release 7.1): an updated online repository of plant genome size data for comprehensive studies. New Phytol 226:301–305. https://doi.org/10.1111/nph.16261

Pennisi E (2001) The human genome. Science 291:1177–1180. https://doi.org/10.1126/sciene291.5507.1177

Popodi E, Kissinger JC, Andrews ME, Raff RA (1996) Sea urchin *Hox* genes: insights into the ancestral *Hox* cluster. Mol Biol Evol

13:1078–1086. https://doi.org/10.1093/oxfordjournals.molbev.a025670
Pysek P, Skalova H, Cuda J, Guo WY, Suda J, Dolozel J, Kauzal O, Lambertini C, Lucanova M, Mandakova T et al (2018) Small genome separates native and invasive populations in an ecologically important cosmopolitan grass. Ecology 99:79–90. https://doi.org/10.1002/ecy.2068
Raina SN, Parida A, Koul KK, Salimath SS, Bisht MS, Raja V, Khoshoo TN (1994) Associated DNA changes in polyploids. Genome 37:560–564
Renny-Byfield S, Chester M, Kovarík A, Le Comber SC, Grandbastien MA, Deloger M, Nichols RA, Macas J, Novák P, Chase MW, Leitch AR (2011) Next generation sequencing reveals genome downsizing in allotetraploid *Nicotiana tabacum*, predominantly through the elimination of paternally derived repetitive DNAs. Mol Biol Evol 28:2843–2845. https://doi.org/10.1093/molbev/msr112
Rensing SA, Lang D, Zimmer AD, Terry S, Salamov A, Shapiro H, Nishiyama T et al (2008) The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. Science 319:64–69. https://doi.org/10.1126/science.1150646
Rey O, Danchin E, Mirouze M, Blanchet S (2016) Adaptation to global change: Transposable element-epigenetics perspective. Trends Ecol Evol 31:514–526. https://doi.org/10.1016/j.tree.2016.03.013
Roddy AB, Théroux-Rancourt G, Abbo T, Benedetti JW, Brodersen CR, Castro M, Castro S, Gilbride AB, Jensen B, Jiang GF et al (2020) The scaling of genome size ad cell size limits maximum rates of photosynthesis with implications for ecological strategies. Int J Plant Sci 181:75–87. https://doi.org/10.1086/706186
Roodt D, Lohaus R, Sterck R, Swanepoel RL, Van de Peer Y, Mizrachi E (2017) Evidence for an ancient whole genome duplication in the cycad. PLoS One 12:e0184454. https://doi.org/10.1371/journal.pone.0184454
Sato S, Nakamura Y, Kaneko T, Asamizu E, Kato T, Nakao M, Sasamoto S, Watanabe A, Ono A, Kawashima K et al (2008) Genome structure of the legume, *Lotus japonicus*. DNA Res 15:227–239. https://doi.org/10.1093/dnares/dsn008
Schimidt T, Heslop-Harrison JS (1998) Genomes, genes and junk: the large-scale organization of plant chromosomes. Trends Genet 3:195–199. https://doi.org/10.1016/S1360-1385(98)01223-0
Schmutz J, Cannon SB, Schuleter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J et al (2010) Genome sequence of the paleopolyploid soybean. Nature 463:178–183. https://doi.org/10.1038/nature08670
Schnable PS, Warre D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang S, Zhang J, Fulton L, Graves TA et al (2009) The B73 maize genome complexity, diversity, and dynamics. Science 326:1112–1115. https://doi.org/10.1126/science.1178534
Siljak-Yakovlev F, Pustahija P (2010) Towards a genome size and chromosome number database of Balkan flora: C-values in 343 taxa with novel values for 242. Adv Sci Lett 3:190–213. https://doi.org/10.1166/asl.2010.1115
Slotkin PK, Martienssen R (2007) Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet 8:272–285. https://doi.org/10.1038/nrg2072
Soltis DE, Soltis PS, Bennett MD, Leitch IJ (2003) Evolution of genome size in the angiosperms. Am J Bot 90:1596–1603. https://doi.org/10.3732/ajb.90.11.1596
Soltis PS, Soltis DE (2016) Ancient WGD events as drivers of key innovations in angiosperms. Curr Opin Plant Biol 30:159–165. https://doi.org/10.1016/j.pbi.2016.03.015
Staton SE, Bakken BH, Blackman BK, Chapman MA, Kane NC, Tang S, Ungerer MC, Knapp SJ, Rieseberg LH, Burke JM (2012) The sunflower (*Helianthus annus* L.) genome reflects a recent history of biased accumulation of transposable elements. Plant J 72:142–153. https://doi.org/10.1111/j.1365-313x.2012.05072x
Straiton J, Free T, Sawyer A, Martin J (2019) From Sanger sequencing to genome databases and beyond. Biotechniques 66:2. https://doi.org/10.2144/btn-2019-0011
Suda J, Meyerson LA, Leitch IJ, Pysek P (2015) The hidden side of plant invasions: The role of genome size. New Phytol 205:994–1007.https://doi.org/10.1111/nph.13107
Suetsuga Y, Futahashi R, Kanomori H, Kadono-Okuda K, Sasanuma S-I, Narukawa J, Ajimura M, Jouraku A, Namiki N, Shimomura M et al (2013) Large scale full-length cDNA sequencing reveals unique genomic landscape in a Lepitoptera model insect, *Bombyx mori*. Genes Genom Genet 3:1481–1492. https://doi.org/10.1534/g3.006239
Sun C, Shepard DG, Chong RA, López Arriaza JR, Hall K, Castoe TA, Feschotte C, Pollock DD, Mueller RL (2012) LTR retrotransposons contribute to genome gigantism in Plethodontid salamanders. Genome Biol Evol 4:168–183. https://doi.org/10.1093/gbe/evr139
Swift H (1950) The constancy of deoxyribose nucleic acid in plant nuclei. Proc Natl Acad Sci USA 36:643–654. https://doi.org/10.1073/pnas.36.11.643
Tennyson CN, KLammut HJ, Worton RG (1995) The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced. Nature Genet 9:184–190
Thomas CA (1971) The genetic organization of chromosomes. Ann Rev Genet 5:237–256. https://doi.org/10.1046/annurev.ge.05.120171.001321
TIBI (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature 463:763–768. https://doi.org/10.1038/nature08747
Tsukaya H (2013) Does ploidy level directly control cell size? Counterevidence from Arabidopsis genetics. PLoS One 8(12):e83729. https://doi.org/10.1371/journal.pone.0083729
Ungerer MC, Strakosh SC, Zhen Y (2006) Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. Curr Biol 16:R872–R873. https://doi.org/10.1016/j.cub.2006.09.020
Vialli M (1957) Volume et contenuen ADN par noyau. Experi Cell Res Suppl 4:283–293
Vinogradov AE (1999) Intron-genome size relationship on a large evolutionary scale. J Mol Evol 49:376–384
Vitte C, Fustier MA, Alix K, Tenallon M (2014) The bright side of transposons in crop evolution. Brief Funct Genom 13:276–295. https://doi.org/10.1093/bfgp/elu002
Volff JN (2006) Turning junk into gold: domestication of transposable elements and creation of new genes in eukaryotes. Bioessays 28:913–922. https://doi.org/10.1002/biwa.20452
Wang H, Devos KM, Bennetzen JL (2014) Recurrent loss of specific introns during angiosperm evolution. PLoS Genet 10:e1004843. https://doi.org/10.1371/journal.pgen.1004843
Waters E, Hohn MJ, Ahel I, Graham DE, Adams MD, Barnstead M, Beeson KY, Bibbs L, Bolanos R, Keller M et al (2003) The genome of *Nanoarchaeum equitans*: Insights into early archaeal evolution and derived parasitism. Proc Natl Acad Sci USA 100:12984–12988. https://doi.org/10.1073/pnas.1735403100
Wegrzyn JL, Liechty JD, Stevens KA, Wu L-S, Loopstra CA, Vasquez-Gross HA, Dougherty WM, Lin BY, Zieve JL, Martínez-García PJ et al (2014) Unique features of loblolly pine (*Pinus taeda* L.) megagenome revealed through sequence annotation. Genetics 196:891–909. https://doi.org/10.1534/genetics.113.159996
Wendel JF (2000) Genome evolution in polyploids. Plant Mol Biol 20:6–7
Wendel JF (2015) The wondrous cycles of polyploidy in plants. Am J Bot 102:1753–1766. https://doi.org/10.3732/ajb.1500320

Wood R, Gwilliam R, Rajandream M-A, Lyne M, Stewart A, Sgouros J, Peat N, Hayles J, Baker S, Basham D et al (2002) The Genome sequence of *Schzosaccharomyces pombe*. Nature 415:871–880. https://doi.org/10.1038/nature724

Zedek F, Šmerda J, Šmarda P, Bureš P (2010) Correlated evolution of LTR retrotransposon and genome size in the genus Eleocharis. BMC Plant Biol 10:265. https://doi.org/10.1186/1471-2229-10-265

Zhang X, Firestein S (2002) The olfactory receptor gene superfamily of the mouse. Nat Neurosci 5:124–133. https://doi.org/10.1038/nn800

Affiliations

Ik-Young Choi[1] · Eun-Chae Kwon[2] · Nam-Soo Kim[2,3]

[1] Department of Agriculture and Life Industry, Kangwon National University, Chuncheon 24341, Republic of Korea

[2] Department of Molecular Bioscience, Kangwon National University, Chuncheon 24341, Republic of Korea

[3] Institute of Biotechnology and Bioscience, Kangwon National University, Chuncheon 24341, Republic of Korea