BBA Advances 4 (2023) 100107

Contents lists available at ScienceDirect

BBA Advances

journal homepage: www.sciencedirect.com/journal/bba-advances

C-value paradox: Genesis in misconception that natural selection follows anthropocentric parameters of 'economy' and 'optimum'

Subhash C. Lakhotia

Cytogenetics Laboratory, Department of Zoology, Institute of Science, Banaras Hindu University, Varanasi 221005, India

ARTICLE INFO

Keywords:
Selfish DNA
Nucleoskeletal DNA
Biological complexity
Noncoding RNA
Endoreplication

ABSTRACT

C-value paradox refers to the lack of correlation between biological complexity and the intuitively expected protein-coding genomic information or DNA content. Here I discuss five questions about this paradox: i) Do biologically complex organisms carry more protein-coding genes? ii) Does variable accumulation of selfish/ junk/ parasitic DNA underlie the c-value paradox? iii) Can nucleoskeletal or nucleotypic function of DNA explain the enigma of orders of magnitude high levels of DNA in some 'lower' taxa or in taxonomically related species? iv) Can the newly understood noncoding but functional DNA explain the c-value paradox? and, v) Does natural selection uniformly apply the anthropocentric parameters for 'optimum' and 'economy'? Answers to Q.1–5 are largely negative. Biology presents numerous 'anomalous' examples where the same end function/ phenotype is attained in different organisms through astoundingly diverse ways that appear 'illogical' in our perceptions. Such evolutionary oddities exist because natural selection, unlike a designer, exploits random and stochastic events to modulate the existing system. Consequently, persistence of the new-found 'solution/s' often appear bizarre, uneconomic, and therefore, paradoxical to human logic. The unexpectedly high c-values in diverse organisms are irreversible evolutionary accidents that persisted, and the additional DNA often got repurposed over the evolutionary time scale. Therefore, the c-value paradox is a redundant issue. Future integrative biological studies should address evolutionary mechanisms and processes underlying sporadic DNA expansions/ contractions, and how the newly acquired DNA content has been repurposed in diverse groups.

Discovery of species-specific constant (c-) values of cellular DNA content in eukaryotes and genesis of the c-value paradox

Following the rediscovery of Mendel's laws in 1900 and the establishment of the chromosomal basis of inheritance, genetic studies initially focused on the genotype-phenotype relationships and mapping of the mutant alleles on the hypothetical linear genetic or linkage maps to represent locations of different genes on chromosomes [82]. Demonstration of the existence of filterable and transmissible agents with the ability to lyse bacteria [30] led H. J. Muller [83] to comment more than 100 years ago, "If these d'Hérelle bodies were really genes, fundamentally like our chromosome genes, they would give us an utterly new angle from which to attack the gene problem. They are filterable, to some extent isolable, can be handled in test-tubes, and their properties, as shown by their effects on the bacteria, can then be studied after treatment. It would be very rash to call these bodies genes, and yet at present, we must confess that there is no distinction known between the genes and them". This was the first indication that the material constituting genes could be subjected to qualitative and quantitative analysis.

An active search for the chromosomal material that could function as genes became possible during the next few decades following the development of i) Feulgen staining method in 1920s [40] for selective visualization of cellular DNA, ii) a method for isolation of cell nuclei [6], and iii) cyto-spectrophotometric quantification of cellular macromolecules [17]. Studies in the late 1940s using biochemical and cytophotometric approaches on Feulgen-stained cells [11,79,108] indicated that despite the variable amounts of DNA per nucleus in different species, the DNA content remains relatively constant in different cell types of a given species with a 1:2 ratio between gametes and diploid cells, while RNA and protein contents in chromosomes were highly variable, leading Mirsky to infer "that DNA is part of the gene substance" [78]. The observed constancy of DNA in different cell types of a given eukaryote, and studies on bacterial transformation and bacteriophage propagation [2,52] collectively established DNA as the genetic material in pro- and eukaryotes. The DNA content of the haploid gametes was referred to as 1C, while that of fertilized zygote as 2C; the c-values did not correlate with the haloid and diploid chromosome numbers (designated as 1 N and 2 N, respectively) seen in the metaphase stage cells of the species.

E-mail address: lakhotia@bhu.ac.in.

https://doi.org/10.1016/j.bbadva.2023.100107
Received 11 June 2023; Received in revised form 11 October 2023; Accepted 12 October 2023
Available online 13 October 2023
2667-1603/© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

The term "c-value paradox", introduced by C. A. Thomas in 1971 [110] while reviewing genetic organization of chromosomes, refers to the unexpectedly high variations in c-values of different species. In his [110] words, "different species contain different amounts of DNA in their nuclei. This harmless information caused some discomfort when it was learned that primitive amphibians and fish contained more than 20 times as much DNA per nucleus as did man. It was argued that mammals display a greater developmental complexity than primitive fish, therefore, they must have more genes, yet why should the lower forms have more DNA if DNA is the chemical basis of the gene?". He [110] noted that the c-value paradox reflected three unexplainable issues: i) why do many 'lower' organisms have significantly higher c-values than the more evolved "higher" organisms; ii) why the c-values between some related species with comparable morphology and body organization differ by one or more orders of magnitude, and iii) why is the proportion of DNA that does not code for proteins (noncoding or ncDNA) so high (up to ~98 %) even in genomes of species that carry the 'basal' c-value characteristic of the given group?

Some early explanations to resolve the c-value paradox included possible inaccuracies in the estimates of the genome sizes or unusual events of genomic multiplication, like polynemy or multi-stranded mitotic chromosomes in species with very high haploid DNA content. These explanations, however, were soon ruled out [16,81,110]. Another suggestion to explain the unusually high c-value content in some lower forms was that the additional DNA served as reserve for future evolutionary experiments. However, since "selection is applied to the organism as it is, not as it might be" [110], this explanation also was untenable.

In the following, I question the various possibilities that have been suggested during the past five decades to resolve the c-value paradox. The final question that I address is if the premises on which the c-value paradox was initially formulated and has been discussed during the past five decades are indeed valid? A summary of the various explanations and their current status is given in Table 1.

Question 1: do genomes of biologically complex organisms carry more protein-coding genes?

With the passage of time, evolution generally leads to the appearance of more complex biological organizations so that the later evolved 'higher' organisms are more complex than the earlier 'lower' or 'primitive' organisms. The commonly used empirical measures of 'biological complexity' are morphological intricacy and the number of cell or tissue types present in the organism [81]. Since the classical genetic and early molecular studies led to a widely accepted belief that proteins only determine structures and functions, and thus the organism's phenotype, the protein-coding genes were expected to be higher in biologically more complex groups. However, classical cytogenetic studies showed that the numbers of protein coding genes (g-value) in organisms varied within a narrow range so that the protein coding gene numbers were not correlated either with biological complexity or with the genome size. This led Thomas [110] to state "if 98 % of the DNA is irrelevant in flies, we can estimate that 99. 98 % is irrelevant in *Triturus*", where 'irrelevant' refers to the genome's ncDNA component. Such lack of correlation between protein coding gene number and genome size generated a related "g-value paradox" [14,23,49,81].

Discovery of variable amounts of repetitive (satellite, high or mid-repetitive) DNA sequences in diverse eukaryotic genomes in the 1960s and their association with the condensed heterochromatin, which was conventionally considered to be genetically inert (see below), led to the belief that such sequences are of no use for the organism. The next question, therefore, examines the possibility that the varying abundance of diverse repetitive and noncoding sequences, which were labelled as 'selfish' or 'junk' or 'parasitic' DNA, could resolve the c-value paradox.

Table 1
Summary of various explanations put forward to resolve the c-value paradox (for details, see text).

Explanation	Evidence	Counter evidence	Current status
Variable accumulation of 'selfish' or 'junk' or 'parasitic' DNA underlies the c-value paradox.	High variability in content of 'non-coding' intergenic DNA and constitutive heterochromatin, which was initially believed to i) lack typical protein-coding genes, ii) to be transcriptionally silent, and iii) to be enriched in highly repetitive, satellite, mid-repetitive and transposable element sequences. Together, these were considered 'selfish' or 'junk' or 'parasitic' DNA and were suggested to variably accumulate and persist in genomes resulting in the loss of correlation between biological complexity and c-value.	Studies during the later part of 20th century, and the subsequent progresses in genomics, have established essential functions of heterochromatin and other noncoding DNA sequences.	'Selfish' or 'junk' or 'parasitic' DNA sequences do not exist in genomes in quantities that can explain the enormous variations in c-values.
Besides the DNA associated with genetic functions, genome also includes nucleoskeletal or nucleotypic sequences to sustain larger nuclear and cell volumes. Necessity for such DNA may explain the orders of magnitude high levels of DNA in some 'lower' taxa or in taxonomically related species.	Larger cells with associated greater nuclear DNA permit greater synthetic and storage capability. There is a wide correlation between higher c-values and larger cell and nuclear volumes, longer cell cycle duration and generation time. The additional 'nucleoskeletal' or 'secondary' or 'nucleotypic' DNA provides a skeletal framework for sustaining larger nuclei and cell bodies but its variable quantum leads to the variable c-values.	Many instances exist where organisms with very large body size have small cells and lower genomic DNA content. In several instances where endoreplication generates larger nuclear and cell sizes, the hereochromatin regions, presumed to also function as 'nucleoskeletal' DNA, actually remain under-replicated.	Existence of 'nucleoskeletal' DNA may explain some instances of high c-values but does not satisfactorily answer the question why some species within a group or in some 'less' evolved taxa need higher cell size and, therefore, greater 'nucleoskeletal' DNA.
The newly understood non-coding but regulatory functions of genomic DNA explain the high proportion of non-coding component in genomes.	A very large proportion of non-coding DNA in any genome has diverse regulatory roles and is thus essential for biological complexity. Variable quantities of such regulatory	Species with lower biological complexity but with very high c-value or species with much higher c-value than their relatives having comparable biological complexity are not expected to need	Regulatory roles of the ncDNA account for its higher abundance than the coding DNA in organisms with taxa's basal genome size but cannot fully explain other

(*continued on next page*)

Table 1 (*continued*)

Explanation	Evidence	Counter evidence	Current status
	DNA determine the c-value.	enormously greater regulatory DNA.	features of the c-value paradox.

Question 2: can accumulation of 'selfish' or 'junk' or 'parasitic' DNA explain the c-value paradox?

The early view that proteins only determine the organism's phenotype found support in some early genetic and cell biological studies on heterochromatin, which suggested that the condensed constitutive heterochromatin [15] i) lacked typical protein-coding genes, ii) was transcriptionally inactive, iii) was enriched in highly repetitive, satellite, mid-repetitive and transposable element sequences, and iv) was even dispensable [27,51,62,63,88,98]. Like the c-values, the relative content of the condensed constitutive heterochromatin also varies widely even in related species. Believing that heterochromatic and repetitive DNA sequences were genetically inert, eukaryotic genomes were suggested to carry variable but high amounts of non-functional 'selfish' or 'junk' or 'parasitic' DNA (collectively referred to here as 'selfish' DNA), which, despite being irrelevant to the host, persist in genomes [33,91,93,94]. Although the proposals of 'selfish' DNA were quickly refuted by many [10,19,34,46,54,102], 'selfish' DNA became a common epithet for ncDNA, and a variable invasion of genomes by 'selfish' DNA remained a common explanation for c-value paradox for several decades.

Contrary to the earlier common belief, some studies during later part of the 20th century and later genomic studies identified increasing numbers of noncoding genes to be functional, leading to a wider and better appreciation of essential roles of the noncoding genomic components in organism's biology [22,39,51,60,62,63,68,99]. Proponents of the belief that constitutive heterochromatin is an unavoidable 'selfish' burden were apparently unaware of phenomenon of chromatin diminution described in 1890s by Theodore Boveri and later by others in several unrelated animal species, which showed that the genomes in these species had the necessary machinery to specifically eliminate heterochromatin from somatic cells, but nevertheless retained it in the germline for essential functions [105]. The application of 'selfish DNA' tag to heterochromatin also reflected ignorance about many pre-1980 studies that had shown heterochromatin i) to function as 'chromosome-engineering DNA' during evolutionary chromosome repatterning, ii) to be transcribed, and iii) to indeed have defined roles in development, fertility, reproductive isolation, and thus, in speciation [28,32,36,41,53,60,65,67,70,71,84,95,99,100,117].

Following an increasing awareness about diverse roles played by heterochromatin and the continuing unraveling of myriads of functional noncoding RNAs (see later), the concept of 'selfish' DNA is now largely 'junked' [24,29,62,63,75,77]. Consequently, varying accumulation of 'selfish' DNA in different genomes cannot be a satisfactory explanation for the c-value paradox.

Question 3: can nucleoskeletal or nucleotypic function of DNA explain the enigma of orders of magnitude high levels of DNA in some 'lower' taxa or in taxonomically related species?

The relative constancy of nucleo-cytoplasmic ratio across the biological systems implies that the cell size is constrained by its nuclear size, which in turn depends largely upon the genomic DNA content [4,9,18,107]. Larger cells provide better efficiency in producing and storing proteins and/or other cellular components and their larger nuclear volume facilitates increased rates of transcription and RNA processing. In view of the wide correlation existing between higher c-values and larger cell and nuclear volumes on one hand and longer cell cycle duration and generation time on the other, two distinct functional classes of eukaryotic genomic DNA have been suggested: i) the primary genic or 'g-DNA' that codes for proteins and/or is involved in regulation of replication, transcription, translation and recombination, and ii) the 'nucleoskeletal' or 'secondary' or 'nucleotypic' DNA, which provides a skeletal framework to sustain larger nuclei and therefore, larger cell bodies [9,18,20,84]. Genomes with unusually high c-values and cell sizes have higher contents of repetitive DNA, ribosomal genes, and transposable elements, which may also function as nucleoskeletal DNA [9,18,20,23,69,84,85,99].

Diverse mechanisms have generated enlarged DNA content during the evolutionary history of living forms. Ancestors of a few major taxa and many unrelated species in different taxa independently acquired larger genomes through one or more rounds of whole genome duplications (polyploidization), or chromosome or gene duplications, which besides increasing the nuclear and cell sizes also facilitated evolution of novel proteins and regulatory circuits [20,72,90]. Many animal and plant species with smaller genomes use developmental polyploidy or endoreplication or polyteny or endomitosis to produce larger cells in specific somatic tissue types [84]. Unicellular ciliates with smaller genomes but larger cell size use an entirely different mechanism involving dimorphic nuclei (micro- and macro-nuclei): the ciliate macronucleus becomes large through a unique system of endoreplication of individual gene size DNA molecules left after an orderly elimination of nearly 95 % of the intervening DNA sequences, whose function is restricted to the 'germline' micronucleus [89].

The varying genome and consequent cell sizes in different taxa have been suggested to serve diverse adaptative functions. For example, among the amphibians, a group with a wide range of genomes and body sizes, the small-bodied and more agile salamanders and frogs have smaller genomes, but the larger salamanders with very high c-values are more sluggish, and slow developing. The larger cell size in sluggish lung fishes with enormously high c-values may help them store high levels of glycogen required during their long estivation in cocoons. On the other hand, the smaller genomes and small cell size in the later evolved birds and bats has been suggested to help them in achieving high metabolic rates necessary for flight [20,47]. However, the estimated cell and genome sizes in extinct dinosaurs do not correlate with their huge body sizes [92]. The presence of small genomes in dinosaurs long before the first birds came into existence also questions if the reduced genome size in birds and bats is indeed related to their high metabolic rates necessary for flight [37].

Thus, while the expanded genomic DNA in many species with unusually high c-values may perform a nucleoskeletal rather than an 'informative' role, and support the large cell and body size, instances also exist where large body size is not dependent upon unusually high c-values and larger cell sizes. Further, in several instances where endoreplication generates larger nuclear and cell sizes, the hereochromatin regions, presumed to also function as 'nucleoskeletal' DNA, actually remain under-replicated [61]. Besides these anomalous situations, the idea of nucleotypic DNA also does not satisfactorily answer the question why some species within a group or some 'less' evolved taxa need higher cell size and, therefore, greater DNA.

Question 4: can the newly understood noncoding but functional DNA explain the c-value paradox?

A significant fraction of the rapidly labeled heterogeneous nuclear RNAs (hnRNAs), derived from unique as well as repetitive sequences, was found during the 1960s to remain confined to the nucleus [31,35,44,50,55,56,101,106,113] but reasons for their nuclear retention remained largely unexplored in the last century because of the wide belief that such DNA sequences were 'selfish' [64]. Establishment of specific functions of a few well-defined RNA pol II synthesized non-coding RNAs during the past century [60] and the recent genomic revolution have, however, led to a much better understanding and appreciation of the genome's noncoding components. It is now well established that, rather than being 'selfish', a large part, if not all, of the

genome is actually transcribed to execute diverse essential regulatory functions in all living organisms [1,5,12,21,43,45,59,66,74,97,111,114,119]. Earlier suggestions [14,60,76] that changes in regulation played more significant and pivotal roles in evolution than mutations in protein-coding 'structural' genes, have been amply confirmed by the contemporary genomic studies. For example, 30-fold more noncoding than protein-coding regions were found to be related with the rate of bill shape evolution in 72 bird species [118]. Likewise, most of the single nucleotide polymorphisms (SNP) associated with diverse human diseases are enriched in non-coding regions near the protein-coding genes [25]. The abundance of introns, adding to the ncDNA proportion, correlates with organismic complexity as multiple introns permit a greater diversity of regulated alternative splicing and thus production of novel RNAs, proteins, and functions [42,87,103,115,116]. Genomic analyses show that the noncoding regulatory DNA sequences associated with the protein coding genes are usually much larger than the protein coding regions [26]. The high c-value genomes also display greater frequency of 'orphan' genes [38], generated through duplication of protein-coding or non-coding genes [38,109]. In view of the increasingly better understanding of the diverse and essential functions of the noncoding component of genomic DNA, the concept of 'selfish' DNA has largely lost its relevance [60,64,120].

The puzzle of very high proportion of ncDNA in genomes of the basal size, characteristic of the given taxonomic group, thus appears nearly resolved when the transcribed noncoding, regulatory, and protein-coding DNA sequences in the organism's genome are considered together [73]. However, the 'c-value enigma' [47], associated with orders of magnitude higher c-values in some 'lower' organisms than the more evolved 'higher' organisms, or some species having significantly higher c-values than their close relatives of comparable biological complexity, remains unanswered.

The next question, therefore, is if nature really assesses 'optimum' with the same sense of 'economy' and 'purpose' as the human mind does, and does nature apply a common yardstick in every case?

Question 5: does natural selection uniformly apply the anthropocentric parameters for 'optimum' and 'economy'?

The genesis and perpetuation of the c-value paradox lies in the general human perception of cost and benefit, so that genetic information and biological complexity are expected to be correlated. The c-value paradox, primarily an outcome of the many instances where this expectation is not fulfilled, was further compounded by the historical emphasis only on protein coding function of genes. As discussed above, appreciation of non-coding DNA as functional component of genome and the nucleoskeletal role of DNA in generating larger nuclei/ cells can partially explain some aspects of the c-value paradox, but the diversity of the mechanisms underlying the quantitative change in the genomic DNA content, and the phylogenetically independent recurrences of disproportionate genome sizes defy the reductionist and cost-benefit considerations that are generally inherent in explanations based on human perspectives of cost, benefit, and purpose.

Biological systems present numerous paradoxical instances where the given end result/ phenotype has been successfully achieved in different species through diverse mechanisms that often appear 'illogical' to human analysis. A glaring example is the diversity of sex-determination mechanisms. Although bisexuality is a basic attribute of most eukaryotes, the mechanisms that determine sex and trigger the male or female modes of zygotic development are uncannily diverse, often even between related species [3,112]. Another example relates to the very different mechanisms and processes employed for somatic 'silencing' of the constitutive heterochromatin and the associated repetitive sequences, which have major roles in gametogenesis and reproductive isolation. These processes span from chromatin/ chromosome diminution in ciliates, and some unelated animal species from nematodes to vertebrates [18,20,86,89,104,105], under-replication during endoreplication cycles in many insects and plants [8,61] and the more widely used epigenetic silencing (heterochromatinization) of chromosome sets, individual chromosomes or chromosome regions [7,13,48]. If the chromosome/ chromatin diminution of non-functional heterochromatic regions or elimination of large number of specific genes in somatic cells of ciliate macronucleus, or the under-replication of heterochromatin during endoreplication cycles, were 'smart economic' solutions that save on the non-productive energy expenditure in carrying and replicating, and then silencing them in soma, why such a diversity of regulatory processes evolved, and more importantly, why have they evolved independently but recurringly rather than being maintained uniformly across phylogenetic lineages? Another paradoxical feature, perhaps unique to ciliate genomes, is the presence of 'scrambled' genes, requiring 'unscrambling' during macronucleus development [58,96]. Obviously, these and numerous other such instances are results of unplanned evolutionary accidents that have survived because the end-result remained functional, no matter how bizarre the underlying processes appear *post facto* to the human mind, which is trained to design cost-effective optimal systems. Apparently, natural selection does not have pre-set parameters to identify 'optimal' and 'economical'.

Concluding remarks and future perspectives

Mutations, point or chromosomal rearrangements, or whole genome or individual chromosome or gene duplications are random and stochastic irreversible accidental events. As stated by J. Monod in the book 'Chance and Necessity' [80] "once incorporated in the DNA structure, the accident – essentially unpredictable because always singular - will be mechanically and faithfully replicated and translated: that is to say, both multiplied and transposed into millions or billions of copies. Drawn out of the realm of pure chance, the accident enters into that of necessity, of the most implacable certainties. *For natural selection operates at the macroscopic level, the level of organisms*" (italic fonts added for emphasis). The myriads of regulatory networks operating in the biological systems at levels of replication, transcription, translation, compartmentalization, and turnover can buffer the initial 'disadvantage' of the mutation, including gain (or loss in some cases) of DNA, and let the organism carrying the newly acquired additional DNA persist in the prevailing environment, especially when the population size is small. Over the evolutionary time scale, the added DNA gets repurposed in diverse ways.

Although the human mind is an outcome of the action of natural selection, the anthropocentric logic of purpose, economy and optimum is not practiced by natural selection. Biological organisms are not created by design but are outcomes of natural selection exploiting random events to operate upon the existing system. Natural selection 'discovers' the adaptive value or otherwise of any change only after it has happened. Therefore, natural selection, being a tinkerer rather than a designer or engineer [57], lets any working system survive irrespective of whether it is the most optimal solution, as would be preferred by a human engineer. If the cumulative adaptive features permit a species to leave enough progeny, despite its acquiring huge genome or an unusual system of selective DNA elimination or endoreplication through evolutionary accident/s, natural selection can eliminate but not revert the system back to energetically more efficient state. Thus, persistence of the 'excess' DNA in a genome becomes paradoxical only when the cost and benefit ratio is applied with a human engineer's or economist's perspective. When viewed from nature's perspective, the c-value paradox becomes irrelevant. Nevertheless, the various discussions on these issues during the past five decades have been very stimulating and have indeed led to a much better understanding of genomes and genomic evolution.

It is expected that future genomics studies, comprehensively covering wider range of species belonging to same as well as distantly related taxa, would generate more extensive but precise and detailed data not only for the c-value changes but also with respect to evolution

of individual genes, transposable elements, and repetitive sequences. These would unravel a much clearer picture of long- as well as short-range trends in genome and gene evolution. Integration of the genomic data of a species and higher taxa with the corresponding morphological, physiological, and developmental features on one hand, and the environmental factors on the other, would unfold the wider canvas on which natural selection works. Genome analyses of species that have become extinct in recent times and comparisons with different populations of their close extant relatives are of great interest, especially for learning about recent changes, including gain or loss, in genomic DNA. Nature and natural selection provide endless diversities in their operational systems.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

I thank Prof. Rajiva Raman (Varanasi, India), Dr. Richa Arya (Varanasi, India), Prof. Pradeep Burma (New Delhi, India) and Prof. V. Nanjundiah (Bangalore, India) for helpful comments on the initial draft.

Funding

Department of Biotechnology (Govt. of India) Grant no. BT/PR32126/BRB/10/1775/2019. Science & Engineering Research Board, Govt. of India (SB/DF/009/2019). The author has applied a Creative Commons Attribution (CC BY) license (or an equivalent open license) to any AAM version arising from this submission, as required by the funders.

References

[1] P.P. Amaral, J.S. Mattick, Noncoding RNA in development, Mamm. Genome 19 (2008) 454–492.
[2] O.T. Avery, C.M. MacLeod, M. McCarty, Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III, J. Exp. Med. 79 (1944) 137–158.
[3] D. Bachtrog, J.E. Mank, C.L. Peichel, M. Kirkpatrick, S.P. Otto, T.L. Ashman, M. W. Hahn, J. Kitano, I. Mayrose, R. Ming, N. Perrin, L. Ross, N. Valenzuela, J. C. Vamosi, C. The Tree of Sex, Sex determination: why so many ways of doing it? PLoS Biol. 12 (2014), e1001899.
[4] S. Balachandra, S. Sarkar, A.A. Amodeo, The nuclear-to-cytoplasmic ratio: coupling DNA content to cell size, cell cycle, and biosynthetic capacity, Annu. Rev. Genet. 56 (2022) 165–185.
[5] C. Barrandon, B. Spiluttini, O. Bensaude, Non-coding RNAs regulating the transcriptional machinery. Biology of the cell /under the auspices of the, Eur. Cell Biol. Organization 100 (2008) 83–95.
[6] M. Behrens, In Handbuch der Biologischen Arbeitsmethoden, 5, Urban and Schwarzenberg, Berlin, 1938, p. 1363.
[7] C. Beisel, R. Paro, Silencing chromatin: comparing modes and mechanisms, Nat. Rev. Genet. 12 (2011) 123–135.
[8] E. Belyaeva, E. Andreyeva, S. Belyakin, E. Volkova, I. Zhimulev, Intercalary heterochromatin in polytene chromosomes of Drosophila melanogaster, Chromosoma 117 (2008) 411–418.
[9] M.D. Bennett, R. Riley, Nuclear DNA content and minimum generation time in herbaceous plants, Proc. R. Soc. Lond. B Biol. Sci. 181 (1972) 109–135.
[10] G. Bernardi, Genome instability and the selfish DNA issue, Folia Biol. 29 (1983) 82–92.
[11] A. Boivin, R. Vendrely, Sur le rôle possible des deux acides nucléiques dans la cellule vivante, Experientia 3 (1947) 32–34.
[12] R. Bonasio, R. Shiekhattar, Regulation of transcription by long noncoding RNAs, Annu. Review Genet. 48 (2014) 433–455.
[13] S. Bongiorni, G. Prantera, Imprinted facultative heterochromatization in mealybugs, Genetica 117 (2003) 271–279.
[14] R.J. Britten, E.H. Davidson, Gene regulation for higher cells: a theory, Science 165 (1969) 349–357.
[15] S.W. Brown, Heterochromatin, Science 151 (1966) 417–425.
[16] H.G. Callan, The organization of genetic units in chromosomes, J. Cell Sci. 2 (1967) 1–7.
[17] T. Caspersson, II.—Methods for the determination of the absorption spectra of cell structures, J. R. Microsc. Soc. 60 (1940) 8–25.
[18] T. Cavalier-Smith, 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 (1978) 247–278.
[19] T. Cavalier-Smith, How selfish is DNA? Nature 285 (1980) 617–618.
[20] T. Cavalier-Smith, Economy, speed and size matter: evolutionary forces driving nuclear genome miniaturization and expansion, Ann. Bot. 95 (2005) 147–175.
[21] J.J. Chan, Y. Tay, Noncoding RNA:RNA regulatory networks in cancer, Int. J. Mol. Sci. 19 (2018).
[22] L.L. Chen, L. Yang, ALUternative regulation for gene expression, Trends Cell Biol. 27 (2017) 480–490.
[23] I.Y. Choi, E.C. Kwon, N.S. Kim, The C- and G-value paradox with polyploidy, repeatomes, introns, phenomes and cell economy, Genes Genomics 42 (2020) 699–714.
[24] M.B. Clark, A. Choudhary, M.A. Smith, R.J. Taft, J.S. Mattick, The dark matter rises: the expanding world of regulatory RNAs, Essays Biochemistry 54 (2013) 1–16.
[25] T.E.P. Consortium, An integrated encyclopedia of DNA elements in the human genome, Nature 489 (2012) 57–74.
[26] T.E.P. Consortium, Expanded encyclopaedias of DNA elements in the human and mouse genomes, Nature 583 (2020) 699–710.
[27] J.E.K. Cooper, T.C. Hsu, Radiation-induced deletions and translocations of Microtus agrestis sex chromosomes in vivo, Exp.l Cell Res. 67 (1971) 343–351.
[28] K.W. Cooper, Cytogenetic analysis of major heterochromatic elements (especially Xh and Y) in Drosophila melanogaster, and the theory of "heterochromatin", Chromosoma 10 (1959) 535–588.
[29] F.F. Costa, Non-coding RNAs: meet thy masters, BioEssays 32 (2010) 599–608.
[30] F. D'Hérelle, Sur un microbe invisible antagoniste des bacilles dysentériques, Compt. Rend. Acad. Sci. 165 (1917) 373–375.
[31] J.E. Darnell Jr, Ribonucleic acids from animal cells, Bacteriol. Rev. 32 (1968) 262.
[32] C. Díaz-Castillo, Junk DNA contribution to evolutionary capacitance can drive species dynamics, Evol. Biol. 44 (2017) 190–205.
[33] W.F. Doolittle, C. Sapienza, Selfish genes, the phenotype paradigm and genome evolution, Nature 284 (1980) 601–603.
[34] G. Dover, Ignorant DNA? Nature 285 (1980) 618–620.
[35] J.E. Edstrom, Chromosomal RNA and other nuclear RNA fractions, in: M. Locke (Ed.), The Role of Chromosomes in Development, Academic Press, New York, 1964, pp. 137–152.
[36] J.F. Elder Jr., B.J. Turner, Concerted evolution of repetitive DNA sequences in eukaryotes, Quart. Rev. Biol. 70 (1995) 297–320.
[37] H. Ellegren, Evolutionary genomics: a dinosaur's view of genome-size evolution, Curr. Biol. 17 (2007) R470–R472.
[38] A.Z. Fakhar, J. Liu, K.M. Pajerowska-Mukhtar, M.S. Mukhtar, The lost and found: unraveling the functions of orphan genes, J Dev Biol 11 (2023) 27.
[39] D. Ferreira, S. Meles, A. Escudeiro, A. Mendes-da-Silva, F. Adega, R. Chaves, Satellite non-coding RNAs: the emerging players in cells, cellular pathways and cancer, Chromosome Res. 23 (2015) 479–493.
[40] R. Feulgen, H. Rossenbeck, Mikroskopisch-chemischer Nachweis einer Nucleinsäure vom Typus der Thymonucleinsäure und die- darauf beruhende elektive Färbung von Zellkernen in mikroskopischen, Präparaten 135 (1924) 203–248.
[41] M. Gallach, 1.688g/cm3 satellite-related repeats: a missing link to dosage compensation and speciation, Mol. Ecol. 24 (2015) 4340–4347.
[42] L. Gallego-Paez, M. Bordone, A. Leote, N. Saraiva-Agostinho, M. Ascensao-Ferreira, N. Barbosa-Morais, Alternative splicing: the pledge, the turn, and the prestige, Hum. Genet. 136 (2017) 1015–1042.
[43] S. Geisler, J. Coller, RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts, Nat. Rev. Mol. Cell Biol. 14 (2013) 699–712.
[44] L. Goldstein, O.H. Trescott, Characterization of RNAs that do and do not migrate between cytoplasm and nucleus, P. Natl. Acad. Sci. USA 67 (1970) 1367–1374.
[45] J. Goodrich, J. Kugel, From bacteria to humans, chromatin to elongation, and activation to repression: the expanding roles of noncoding RNAs in regulating transcription, Crit. Rev. Biochem. Mol. Biol. 44 (2009) 3–15.
[46] S.J. Gould, Darwinism and the expansion of evolutionary theory, Science 216 (1982) 380–387.
[47] T.R. Gregory, A bird's-eye view of the C-value enigma: genome size, cell size, and metabolic rate in the class Aves, Evolution (N Y) 56 (2002) 121–130.
[48] S. Grewal, D. Moazed, Heterochromatin and epigenetic control of gene expression, Science 301 (2003) 798–802.
[49] M.W. Hahn, G.A. Wray, The g-value paradox, Evolution Development 4 (2002) 73–75.
[50] H. Harris, Nuclear ribonucleic acid, in: J.N. Davidson, W.E. Cohn (Eds.), Progress in Nucleic Acid Research and Molecular Biology, Academic Press, 1963, pp. 19–59.
[51] W. Hennig, Heterochromatin and germ line-restricted DNA, Results Probl. Cell Differ. 13 (1986) 175–192.

[52] A.D. Hershey, M. Chase, Independent functions of viral protein and nucleic acid in growth of bacteriophage, J. Gen. Physiol. 86 (1952) 39–56.
[53] O. Hess, G.F. Meyer, Genetic activities of the Y chromosome in Drosophila during spermatogenesis, Adv. Genet. 14 (1968) 171–223.
[54] D.A. Hickey, Selfish DNA: a sexually-transmitted nuclear parasite, Genetics 101 (1982) 519–531.
[55] D.S. Holmes, J. Bonner, Interspersion of repetitive and single-copy sequences in nuclear ribonucleic acid of high molecular weight, Proc. Natl. Acad. Sci. USA 71 (1974) 1108–1112.
[56] D.S. Holmes, J.E. Mayfield, G. Sander, J. Bonner, Chromosomal RNA: its properties, Science 177 (1972) 72–74.
[57] F. Jacob, Evolution and tinkering, Science 196 (1977) 1161–1166.
[58] L.A. Katz, A.M. Kovner, Alternative processing of scrambled genes generates protein diversity in the ciliate Chilodonella uncinata, J. Experim. Zool. Part B 314B (2010) 480–488.
[59] F. Kopp, J.T. Mendell, Functional classification and experimental dissection of long noncoding RNAs, Cell 172 (2018) 393–407.
[60] S.C. Lakhotia, RNA polymerase II dependent genes that do not code for protein, Indian J. Biochem. Biophys. 33 (1996) 93–102.
[61] S.C. Lakhotia, Replication in Drosophila chromosomes. XII. Reconfirmation of underreplication of heterochromatin in polytene nuclei by cytofluorometry, Chromosoma 89 (1984) 63–67.
[62] S.C. Lakhotia, From heterochromatin to long noncoding RNAs in Drosophila: expanding the arena of gene function and regulation, in: M.R.S. Rao (Ed.), Long Non Coding RNA Biology, Springer Nature Singapore Pte Ltd, Singapore, 2017, pp. 75–118.
[63] S.C. Lakhotia, Non-coding RNAs demystify constitutive heterochromatin as essential modulator of epigenotype, Nucleus 60 (2017) 299–314.
[64] S.C. Lakhotia, Central dogma, selfish DNA and noncoding RNAs: a historical perspective, Proc. Indian Natn. Sci. Acad. 84 (2018) 315–427.
[65] S.C. Lakhotia, J. Jacob, EM autoradiographic studies on polytene nuclei of Drosophila melanogaster. II. Organization and transcriptive activity of the chromocentre, Exp. Cell. Res. 86 (1974) 253–263.
[66] S.C. Lakhotia, B. Mallick, R. Jyoti, Non-coding RNAs: ever-expanding diversity of types and functions, in: R. Sharma (Ed.), RNA-Based Regulation in Human Health and Disease, Elsevier, 2020, pp. 5–57.
[67] S.C. Lakhotia, T. Mukherjee, Absence of novel translation products in relation to induced activity of the 93D puff in Drosophila melanogaster, Chromosoma 85 (1982) 369–374.
[68] S. Lavrov, M. Kibanov, Noncoding RNAs and chromatin structure, Biochemistry. Biokhimiia 72 (2007) 1422–1438.
[69] N.F. Lunkova, N.V. Zhukovskaya, V.B. Ivanov, Relationship of the holoploid DNA content with the life form and duration of plants' life cycle, Russ. J. Dev. Biol. 51 (2020) 387–396.
[70] S. Luo, J. Lu, Silencing of transposable elements by piRNAs in Drosophila: an evolutionary perspective, Genomics Proteomics Bioinformatics 15 (2017) 164–176.
[71] R. MacIntyre, Mutation Driven Evolution, J. Heredity 106 (2015), 420-420.
[72] S. Magadum, U. Banerjee, P. Murugan, D. Gangapur, R. Ravikesavan, Gene duplication as a major force in evolution, J. Genet. 92 (2013) 155–161.
[73] J.M. Marcus, A Partial Solution to the C-Value Paradox, Springer Berlin Heidelberg, Berlin, Heidelberg, 2005, pp. 97–105.
[74] J.S. Mattick, P.P. Amaral, P. Carninci, S. Carpenter, H.Y. Chang, L.L. Chen, R. Chen, C. Dean, M.E. Dinger, K.A. Fitzgerald, T.R. Gingeras, M. Guttman, T. Hirose, M. Huarte, R. Johnson, C. Kanduri, P. Kapranov, J.B. Lawrence, J. T. Lee, J.T. Mendell, T.R. Mercer, K.J. Moore, S. Nakagawa, J.L. Rinn, D. L. Spector, I. Ulitsky, Y. Wan, J.E. Wilusz, M. Wu, Long non-coding RNAs: definitions, functions, challenges and recommendations, Nat. Rev. Mol. Cell Biol. (2023).
[75] J.S. Mattick, I.V. Makunin, Non-coding RNA, Hum. Mol. Genet. 15 (2006) R17–R29.
[76] E. Mayr, Populations, Species, and Evolution. An abridgment of Animal Species and Evolution, Harvard University Press, Cambridge (USA), 1970.
[77] P. Michalak, RNA world - the dark matter of evolutionary genomics, J. Evol. Biol. 19 (2006) 1768–1774.
[78] A.E. Mirsky, Some chemical aspects of the cell nucleus, in: L.C. Dunn (Ed.), Genetics in the 20th Century. Essays on the Progress of Genetics During Its First 50 Years, McMillan Co., New York, 1951, pp. 127–153, by.
[79] A.E. Mirsky, H. Ris, Variable and constant components of chromosomes, Nature 163 (1949) 666–667.
[80] J. Monod, Chance and necessity: an Essay On the Natural Philosophy of Modern Biology Alfred A, Knopf, New York, 1971.
[81] G.P. Moore, The C-value paradox, Bioscience 34 (1984) 425–429.
[82] T.H. Morgan, A.H. Sturtevant, H.J. Muller, C.B. Bridges, The Mechanism of Mendelian heredity, Henry Holt, New York, NY, 1915.
[83] H.J. Muller, Variation due to change in the individual gene, Am. Nat. 56 (1922) 32–50.
[84] W. Nagl, Endopolyploidy and Polyteny in Differentiation and Evolution, Elsevier/North-Holland Biomedical Press, Amsterdam, 1978.
[85] M. Naville, S. Henriet, I. Warren, S. Sumic, M. Reeve, J.N. Volff, D. Chourrout, Massive changes of genome size driven by expansions of non-autonomous transposable elements, Curr. Biol. 29 (2019) 1161–1168, e1166.
[86] J. Niedermaier, K. Moritz, Organization and dynamics of satellite and telomere DNAs in Ascaris: implications for formation and programmed breakdown of compound chromosomes, Chromosoma 109 (2000) 439–452.
[87] T.W. Nilsen, B.R. Graveley, Expansion of the eukaryotic proteome by alternative splicing, Nature 463 (2010) 457–463.
[88] G. Nishibuchi, J. Déjardin, The molecular basis of the organization of repetitive DNA-containing constitutive heterochromatin in mammals, Chromosome Res. 25 (2017) 77–87.
[89] M. Nowacki, V. Vijayan, Y. Zhou, K. Schotanus, T.G. Doak, L.F. Landweber, RNA-mediated epigenetic programming of a genome-rearrangement pathway, Nature 451 (2008) 153–158.
[90] S. Ohno, Evolution By Gene Duplication, Springer-Verlag, 1970.
[91] S. Ohno, Gene Duplication, Junk DNA, Intervening Swquences and the Universal signal for their removal, Rev. Brasil Genet. III (1980) 99–114.
[92] C.L. Organ, A.M. Shedlock, Palaeogenomics of pterosaurs and the evolution of small genome size in flying vertebrates, Biol. Lett. 5 (2009) 47–50.
[93] L.E. Orgel, F.H. Crick, Selfish DNA: the ultimate parasite, Nature 284 (1980) 604–607.
[94] L.E. Orgel, F.H. Crick, C. Sapienza, Selfish DNA, Nature 288 (1980) 645–646.
[95] C. Pelling, Ribonucleic acid synthesis in giant chromosomes. Autoradiographic investigations on Chironomus tentans, Chromosoma 15 (1964) 71.
[96] D.M. Prescott, The DNA of ciliated protozoa, Microbiol. Rev. 58 (1994) 233–267.
[97] A. Saxena, P. Carninci, Long non-coding RNA modifies chromatin: epigenetic silencing by long non-coding RNAs, BioEssays 33 (2011) 830–839.
[98] V.C. Shah, S.C. Lakhotia, S.R.V. Rao, Nature of heterochromatin, J. Sci. Ind. Res. 32 (1973) 467–480.
[99] J.A. Shapiro, R. von Sternberg, Why repetitive DNA is essential to genome function, Biol. Rev. 80 (2005) 227–250.
[100] T. Sharma, Chromosomal and molecular divergence in the Indian pygmy field mice Mus booduga-terricolor lineage of the subgenus Mus, Genetica 97 (1996) 331–338.
[101] R. Shearer, B. McCarthy, Evidence for ribonucleic acid molecules restricted to the cell nucleus, Biochemistry 6 (1967) 283–289.
[102] B.J. Shuter, J.E. Thomas, W.D. Taylor, A.M. Zimmerman, Phenotypic correlates of genomic DNA content in unicellular eukaryotes and other cells, Am. Nat. 122 (1983) 26–44.
[103] C.W. Smith, J. Valcarcel, Alternative pre-mRNA splicing: the logic of combinatorial control, Trends Biochem. Sci. 25 (2000) 381–388.
[104] J. Smith, C. Baker, E. Eichler, C. Amemiya, Genetic consequences of programmed genome rearrangement, Curr. Biol. 22 (2012) 1524–1529.
[105] J.J. Smith, Programmed DNA elimination: keeping germline genes in their place, Curr. Biol. 28 (2018) R601–R603.
[106] R. Soeiro, M. Vaughan, J. Warner, The turnover of nuclear DNA-like RNA in HeLa cells, J. Cell Biol. 39 (1968) 112–118.
[107] B. Sokoloff, The nucleo-cytoplasmic ratio and cancer, J. Cancer Res. 7 (1922) 395–415.
[108] H. Swift, The constancy of desoxyribose nucleic acid in plant nuclei, Proc. Natl. Acad. Sci. USA 36 (1950) 643–654.
[109] D. Tautz, T. Domazet-Lošo, The evolutionary origin of orphan genes, Nat. Rev. Genet. 12 (2011) 692–702.
[110] C.A.J. Thomas, The genetic organization of chromosomes, Annu. Rev. Genet. 5 (1971) 237–256.
[111] S. Tomita, M.O.A. Abdalla, S. Fujiwara, T. Yamamoto, H. Iwase, M. Nakao, N. Saitoh, Roles of long noncoding RNAs in chromosome domains, Wiley Interdisciplinary Reviews: RNA 8 (2017).
[112] B. Vicoso, Molecular and evolutionary dynamics of animal sex-chromosome turnover, Nat. Ecol. Evol. 3 (2019) 1632–1641.
[113] R.A. Weinberg, Nuclear RNA metabolism, Annu. Rev. Biochem. 42 (1973) 329–354.
[114] J.E. Wilusz, Long noncoding RNAs: re-writing dogmas of RNA processing and stability, Biochim. Biophys. Acta 1859 (2016) 128–138.
[115] W.M. Wojtowicz, J.J. Flanagan, S.S. Millard, S.L. Zipursky, J.C. Clemens, Alternative splicing of drosophila dscam generates axon guidance receptors that exhibit isoform-specific homophilic binding, Cell 118 (2004) 619–633.
[116] P. Yang, D. Wang, L. Kang, Alternative splicing level related to intron size and organism complexity, BMC Genomics [Electronic Resource] 22 (2021) 853.
[117] J.J. Yunis, W.G. Yasmineh, Heterochromatin, satellite DNA, and cell function. Structural DNA of eucaryotes may support and protect genes and aid in speciation, Science 174 (1971) 1200–1209.
[118] L. Yusuf, M.C. Heatley, J.P.G. Palmer, H.J. Barton, C.R. Cooney, T.I. Gossmann, Noncoding regions underpin avian bill shape diversification at macroevolutionary scales, Genome Res. 30 (2020) 553–565.
[119] M. Zaratiegui, D. Irvine, R. Martienssen, Noncoding RNAs and gene silencing, Cell 128 (2007) 763–776.
[120] E. Zuckerkandl, Why so many noncoding nucleotides? The eukaryote genome as an epigenetic machine, Genetica 115 (2002) 105–129.