

Evolution of genome size: new approaches to an old problem

Dmitri A. Petrov

Eukaryotic genomes come in a wide variety of sizes. Haploid DNA contents (C values) range >80 000-fold without an apparent correlation with either the complexity of the organism or the number of genes. This puzzling observation, the C-value paradox, has remained a mystery for almost half a century, despite much progress in the elucidation of the structure and function of genomes. Here I argue that new approaches focussing on the genetic mechanisms that generate genome-size differences could shed much light on the evolution of genome size.

Because DNA is the stuff of genes, it is natural to think that more complex organisms would require more genes and thus have more DNA. Paradoxically, however, even the initial observations¹ showed that many apparently simple organisms could have over a thousand times more DNA than presumably more complex multicellular organisms. It is difficult to see why some amoebas would need 200 times more DNA than humans², or why lillies would need over 200 times more DNA than rice.

Research into these differences soon revealed that in many organisms much of the DNA is noncoding and often repetitive³. This provided a solution to the original paradox in that it showed that apparently simple organisms probably do have fewer essential genes than more complex organisms, even though they sometimes have larger genomes because of larger amounts of noncoding DNA. However, these discoveries also generated a whole set of new questions. Which evolutionary forces could produce vast amounts of noncoding DNA? What is the adaptive function, if any, of the nongenic DNA? If nongenic DNA does not have an adaptive role, why would natural selection tolerate the burden of extra DNA?

'Adaptive' versus 'junk DNA' theories of genome-size evolution

Traditionally, theories of genome-size evolution primarily attempted to solve the puzzle of the apparent wastefulness of Nature; that is, why would many genomes have vast amounts of extra DNA considering the actual informational needs of the organism? There are two broad classes of explanations.

The 'adaptive' theories postulate an adaptive function for this extra DNA given that DNA abundance, rather than its information content, can have a direct and significant effect on phenotype⁴ (Box 1). For instance, a larger genome size could be

adaptive because it directly or indirectly increases nuclear and cellular volumes^{4,5}, helps to buffer fluctuations in the concentration of regulatory proteins⁶ or protects coding DNA from mutation⁷. According to these hypotheses, the observed variation in genome size reflects different adaptive needs or the efficacy of natural selection in different organisms.

Alternatively the 'junk DNA' theories propose that the extra DNA is indeed extra – that it is useless, maladaptive DNA fixed by random drift and carried passively in the chromosomes⁸. More recent versions of this theory propose that the junk DNA comprises parasitic transposable elements (TEs) (the 'selfish' DNA hypothesis)^{9,10}. According to these theories, purifying selection against the accumulation of useless DNA is often not strong enough completely to counteract the steady stream of DNA addition through transposition and pseudogene formation. The final genome size is then set at the highest tolerable maximum which depends on the particular ecological and developmental needs of the organism.

Evolutionary forces affecting genome size

The current dichotomy between adaptive and junk DNA theories places the focus squarely on the question of whether extra DNA benefits the organism. But could there be more to it? I propose that by focussing exclusively on the question of the function (or lack of thereof) of extra DNA, the current debate obscures differences between various evolutionary explanations that might also be relevant. The easiest way to see this is to cast the question of genome-size evolution in terms of population genetics.

Whatever the evolutionary scenarios of genome-size change might be, they must involve mutational mechanisms of addition and loss of DNA. Figure 1 shows some of these mechanisms, including the activity of TEs, spontaneous deletions and insertions, genome duplications and many others. The genome-size variants sometimes affect phenotype and thus have to go through natural selection before being fixed in the genome. It is also likely that, within a certain range, genome-size variants could be of such similar selective values that their ultimate fates are determined primarily by neutral drift (Box 2).

The essential point is that changes in genome size can occur through modulation of any evolutionary force in Fig. 1. Certainly, changes in the strength and direction of natural selection can do that. But also, unless natural selection is exceptionally strong and does not allow any genetic drift to take place, modulation of the strength of any of the mutational factors shown in Fig. 1 should also affect genome size. For instance, other things being equal, an increase in transposition rates should lead to an increase in genome size, even though the exact magnitude of the increase will depend on the strength of natural selection for or against genome-size growth, and on other factors, such as the availability of nondeleterious insertion sites. The same applies to changes in the average rates

D.A. Petrov
Department of Biological
Sciences, Stanford
University, Stanford,
CA 94035, USA.
e-mail:
dpetrov@stanford.edu

Box 1. Phenotypic correlates of genome size

Most evidence for the adaptive evolution of genome size comes from the numerous observations of correlations between genome size and various phenotypic traits of apparent selective significance. The strongest positive correlation of the genome size is with the cellular and nuclear sizes^{a,b}. Genome size also correlates (among many other characters) with:

- duration of mitosis and meiosis^c
- rate of basal metabolism in mammals and passerine birds (inversely)^{d,e}
- minimum generation time^{f,g}
- seed size^{g,h}
- response of annual plants to CO₂ (Ref. i)
- embryonic development time in plethodontid salamanders^j

• morphological complexity in the brains of frogs and salamanders (inversely)^k.
Such correlations strongly suggest that natural selection has had an important role in genome-size evolution, although the exact nature of these selective forces remains obscure.

References

- a Cavalier-Smith, T. (1982) Skeletal DNA and the evolution of genome size. *Annu. Rev. Biophys. Bioeng.* 11, 273–302
- b Gregory, T.R. and Hebert, P.D. (1999) The modulation of DNA content: proximate causes and ultimate consequences. *Genome Res.* 9, 317–324
- c Bennett, M.D. (1987) Variation in genomic form in plants and its ecological implications. *New Phytol.* 106, 177–200
- d Vinogradov, A.E. (1995) Nucleotypic effect in homeotherms: Body-mass-corrected basal metabolic rate of mammals is related to genome size. *Evolution* 49, 1249–1259
- e Vinogradov, A.E. (1997) Nucleotypic effect in homeotherms: Body-mass independent resting metabolic rate of passerine birds is related to genome size. *Evolution* 51, 220–225
- f Bennett, M.D. (1972) Nuclear DNA content and minimum generation time in herbaceous plants. *Proc. R. Soc. Lond. B Biol. Sci.* 181, 109–135
- g Wakamiya, I. (1993) Genome size and environmental factors in the genus *Pinus*. *Am. J. Bot.* 80, 1235–1241
- h Chung, J. *et al.* (1998) Relationships between nuclear DNA content and seed and leaf size in Soybean. *Theor. Appl. Gen.* 96, 1064–1068
- i Jasienski, M. and Bazzaz, F.A. (1995) Genome size and high CO₂. *Nature* 376, 559–560
- j Jockush, E.L. (1997) An evolutionary correlate of genome size change in plethodontid salamanders. *Proc. R. Soc. Lond. B Biol. Sci.* 264, 597–604
- k Roth, G. *et al.* (1994) Cell size predicts morphological complexity in the brains of frogs and salamanders. *Proc. Natl. Acad. Sci. U. S. A.* 91, 4796–4800

of deletion or insertion or to heterochromatin shrinkage or expansion.

The most common interpretations of both the adaptive and the junk DNA theories pick out only a subset of this range of possibilities, by assuming that genome size is set at the level where natural selection is strong enough to allow no or minimum genetic drift of genome size to take place. Of course, one might argue that these theories can easily accommodate the possibility of genetic drift and of

changes in mutational patterns in different organisms. That is true. By focusing only on the function of extra DNA, adaptive and junk DNA theories do not need to be explicit regarding these issues. But this is exactly my point – by not taking a stand on exactly which evolutionary forces produce changes in genome size and on the relative importance of these forces, the above theories fail to make explicit important distinctions between different scenarios of genome-size evolution.

I suggest that, in addition to thinking of genome-size evolution in terms of the adaptive and junk DNA theories, whenever possible we should also investigate directly all of the potential evolutionary scenarios, including all of the mutational and selective forces potentially affecting genome size. We should then try to estimate directly the strength of individual forces in different organisms to see whether the modulation in strength of the individual forces corresponds to changes in genome size. Such an approach should not only tell us which genetic mechanisms and selective forces affect genome size, but also give us a quantitative sense of their relative importance. For instance, if in a particular group of organisms, variation in the strength of natural selection for or against DNA addition explains 95% of the variation in genome size, whereas variation in the rate of transposition explains 5%, we will be able to tell which force is more important and do it quantitatively.

In the event that natural selection proves to be the dominant force, we would then need to determine which of the many phenotypic correlates of genome-size changes are selectively important. But it is important to know that natural selection is indeed modulating genome size before trying to find phenotypic reasons for this selection.

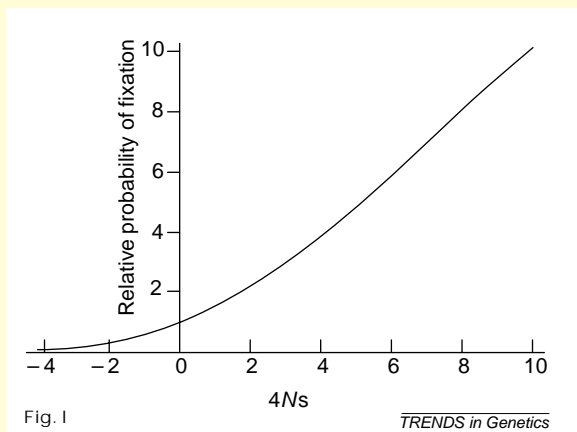
Of course, different evolutionary forces can be important in different organisms and across different

Box 2. Evolutionary fates of new mutations

New mutations have a long-term evolutionary effect only to the extent to which they become fixed in populations. The probability of fixation of a new mutation (present in the population at the frequency of $1/2N$) depends both on the population size (N) and the selective coefficient (s) of the new mutation and is given by the formula $2s/(1-e^{-4Ns})$. If the new mutation is neutral ($s = 0$) or effectively neutral ($Ns \ll 1$), its fate is governed by genetic drift and its probability of fixation is $1/2N$. The relative probability of

fixation of a non-neutral versus neutral mutation is then $4Ns/(1-e^{-4Ns})$ (see Fig. 1).

Note that even favorable mutations do not necessarily become fixed and that deleterious mutations are not always lost.



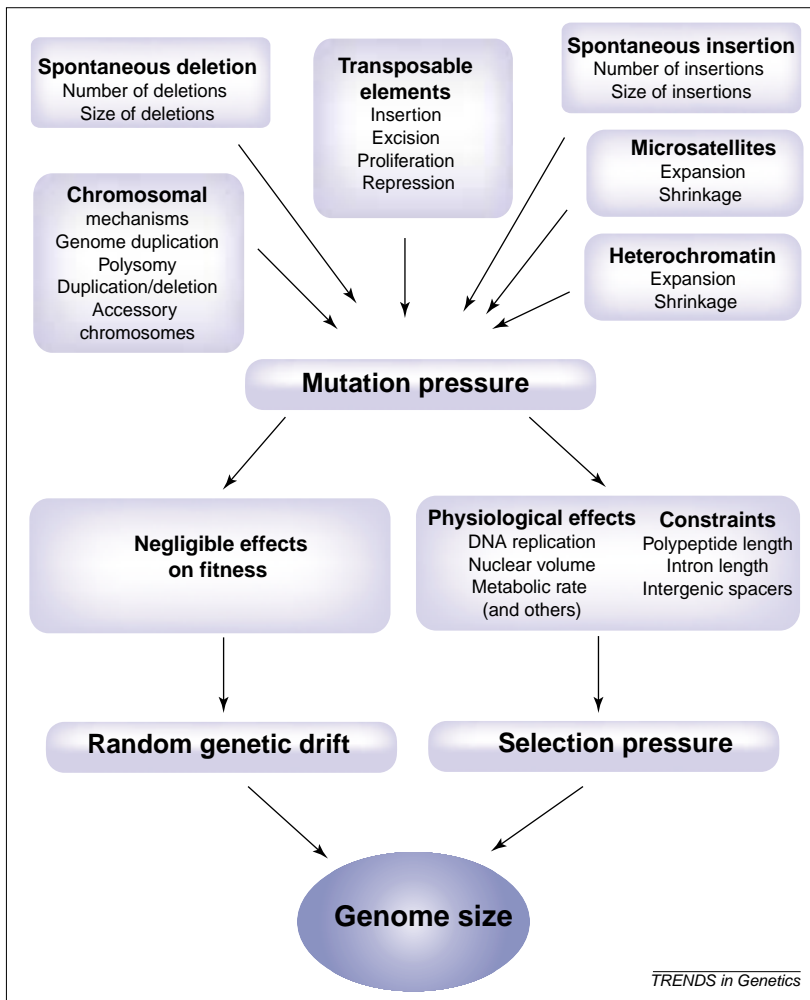


Fig. 1. The forces affecting genome-size evolution. DNA-length mutants are created through a variety of mechanisms shown at the top, producing mutational pressure either to expand or contract the genome size. Some of these mutations affect the phenotype and undergo natural selection. Some might have negligible selective effects and are governed primarily by genetic drift. The combined interplay of all these forces affects genome size.

time frames. It might turn out, for instance, that different forces are important in animals versus plants, or in long-term evolution of genome size versus genome-size differences between closely related species. But if we cast the net broadly enough, in the end this approach can help bring about a robust understanding of genome-size evolution.

Distinguishing evolutionary forces of genome-size change

Given the multitude of genetic and population processes involved, the main challenge of the approach I am advocating is to find a way of distinguishing among them and to studying them individually.

One way to distinguish among the forces acting on genome size is to consider the timescale over which these forces could be effective, as different mutational mechanisms act on very different timescales. The activity of TEs is relatively fast, potentially amplifying the transposable-element copy number by 20–100 copies (~0.1–1 Mbp) in a single generation^{11–16}. By contrast, changes of genome size through small spontaneous deletions and insertions are relatively slow, with, for example, the *Drosophila melanogaster* genome losing less than a single base pair per generation¹⁷.

Of course, the evolutionary impact of these changes depends on the probability of their becoming fixed in the

genome. If there is strong selection *against* increases in genome size, even strong mutational pressure to increase genome size would not affect the long-term evolution of genome size. However, strong selection *for* a change in genome size could substantially enhance the impact of slow mutational mechanisms by increasing the probability and the rate of fixation of length variants over the neutral expectations.

The above considerations could help us restrict our search for the mechanisms responsible for genome-size evolution if we know the timescale of genome-size divergence in a particular case. Specifically, we can assert that genome-size changes between closely related organisms must be due either to very fast (relative to the time of species divergence) mutational mechanisms or to natural selection. However, the long-term genome size, established over very long periods of time, is affected by all of the forces in Fig. 1, both fast and slow. In the long run, slow and steady forces could be just as powerful as the quick but sporadic ones.

Global nature of forces affecting genome size

Another way to differentiate among multiple forces acting on genome size is to consider the scope of their action. Some forces, such as natural selection acting on a trait correlated with total genome size, are global in the sense that they affect the size of all genomic sequences, provided that these sequences are free to vary in size and that size variation exists. For example, if a lineage experiences an increased selection for a shorter developmental time (and therefore for more rapid DNA replication and cell division), then all of the unconstrained or weakly constrained sequences should be reduced in size. In other cases, however, a force affecting genome size could have a more limited scope and would only affect some genomic components and not others: an increase in the rate of heterochromatin shrinkage through a mechanism specific to heterochromatic DNA should not affect the size of euchromatin, and an expansion of satellite DNA through a mechanism specific to satellite DNA should not affect the size of satellite-free sequences.

Two forces clearly have the capacity to affect the whole genome: natural selection and global deletion–insertion biases. The activity of TEs could also have global effects on genome size, at least in the long run when copies of TEs can no longer be recognized, yet continue to take up space. Other forces, such as satellite expansion, heterochromatic shrinkage or expansion, creation of accessory chromosomes and so forth have a much more limited scope of action. Polyploidization is also not expected to increase the size of all genomic compartments, but rather to make them more numerous.

Thus, a fundamental question in the study of genome-size evolution is whether different genomic components vary together in a correlated fashion. Data of this kind are unfortunately limited. However, a number of cytogenetic and molecular studies give us a glimpse of a general pattern.

The first key observations were made by cytogeneticists who showed that genome-size differences are scattered generally throughout (at least) the euchromatic portion of the genome^{18–23}. Subsequent molecular studies support these claims.

A comparison of orthologous mammalian introns revealed a correlation between the average size of introns and genome size²⁴. A comparison of 115 complete introns present in 42 homologous genes in *Drosophila virilis* and *D. melanogaster*²⁵ showed that the average length of introns is larger in *D. virilis* than in *D. melanogaster* (P value = 0.003). The mean intron lengths for the two species were 394 and 283 bp, respectively, a difference of 39%. Interestingly, the size of the euchromatic genome of *D. virilis* exceeds that of *D. melanogaster* by 36% (150 Mb versus 110 Mb)^{26,27}. The trend of the correlated change in intron sizes with the changes in euchromatic genome size was confirmed by the analysis of a very large number of introns in model organisms^{28,29}. Note that the change in intron length does not account for all the changes in genome size, implying that other sequences in the genomes grow or shrink together with introns.

Increases in genome size have also been associated with increases in the copy number of TEs^{13,16,30–33}, increases in the amount of simple repeated sequences³⁴, the presence of large numbers of pseudogenes³⁵, increases in the size of inter-enhancer spacers (C.M. Bergman, pers. commun.), and increases in the size or abundance of microsatellites^{30,36}. The tentative trend emerging from these studies is that, when genomes change in size, they do it across all genomic components, implicating a global force as the agent of genome-size change. This conclusion is at best preliminary and much additional work is required before it is firmly established.

Measuring individual forces

How can we assess the relative importance of particular mechanisms as forces of this genome-size change? As genome-size changes are reflected in all genomic components, we cannot answer this question simply by testing whether any one particular class of sequences is amplified in large genomes or reduced in compact ones. The action of any global force would produce very similar long-term effects.

What we need are direct experimental studies of the strength of the individual mechanisms of genome-size evolution in action. The lack of viable experimental approaches has long been an obstacle to such studies. However, two new approaches (discussed below)^{33,37} have changed this, at least for some of the potential mechanisms of genome-size change.

Long-term estimates of transposable element activity

The activity of TEs shapes much of eukaryotic genomes and has a major impact on the evolution of genome size. Because transposition rates are generally higher than excision rates, TEs increase genome size. Although one can easily envision how a

change in transposition rate could result in a change of genome size, the difficulty is that the change in the copy number of TEs in a genome cannot be taken as strong evidence of a change in transposition rates; modulation of any global genomic force would produce the same end result. However, a recent study^{33,38} demonstrates a way in which it is possible to estimate the average rate of transposition and fixation of at least some TEs in a predetermined genomic region. The approach relies on contiguous sequencing of a defined genomic region and identification of all long-terminal repeat (LTR)-containing retrotransposable elements in the region. The authors then use sequence divergence of LTRs in individual elements to estimate when their insertions into the region took place. The mechanism of transposition of LTR retroelements ensures that at the time of insertion the 5' and 3' LTRs are identical in sequence. However, after insertion they start evolving independently and diverging in sequence. The extent of the divergence can then be used to calibrate the age of each element.

SanMiguel and colleagues^{33,38} used this approach in maize to obtain a remarkable result. In the 240 kb of contiguous sequence around the *adh1* gene they found 23 copies of TEs belonging to 11 families of retrotransposons. These 23 copies of retrotransposons accounted for over 160 kb. Importantly, the LTR analysis demonstrated that all elements have transposed in the past 6 Myr, with most jumping in the past 3 Myr. Assuming that the *adh1* region is representative of the maize genome in general (and there is no reason to believe otherwise), this result implies that the maize genome has grown by 50%, from 1200 Mbp to 2400 Mbp, in the past 3 Myr.

How can we interpret these results? One straightforward explanation is that the transposition frequency in maize has increased substantially in the past 3 Myr. Alternatively, it is also possible that the fixation probability of retrotransposons has changed. For example, natural selection against genome-size growth might keep TEs from fixation in maize relatives, but not in maize itself. This study did not address this question, although in principle it can be approached by estimating population frequencies and ages of individual TEs in maize and its relatives (e.g. sorghum). A higher population frequency of TEs of similar age in maize than in sorghum would implicate natural selection, because it would mean that in maize each individual TE insertion has a higher chance of persisting and ultimately being fixed than in sorghum. Similar population frequencies of similarly aged TEs implicate differences in transposition rate. One complicating factor is that the more compact sorghum genome could have fewer nondeleterious insertion sites available for TEs than the larger genome of maize.

The use of a predefined contiguous genomic region is essential in this study as it allows the study of the number and ages of *all* elements present in the region and extrapolation of the results to the whole genome. However, this requirement makes it difficult to

implement in organisms where little genomic analysis has been done. The hope is that continuing technological advances in cloning and sequencing will mitigate this difficulty and allow the use of this method in a broad array of organisms.

It will be important to see whether genome-size differences across larger evolutionary distances than those in the studies of SanMiguel and colleagues^{33,38} are also due (at least partly) to changes in TE activity. Alternatively, it is possible that many lineages experience transient bursts of transposition and genome-size growth, but that over longer times most lineages experience a similar number of bursts without generating long-term divergences in genome size³⁹.

Deletion–insertion spectra

The spectrum (distribution) of size and frequency of small spontaneous nucleotide insertions and deletions (indels) is one of the important parameters in the long-term evolution of genome size. If spontaneous insertions are more frequent and longer on average than deletions, this would generate a persistent global pressure towards genome-size growth. The reverse would be true if deletions are more frequent and were longer on average than insertions. If the spectrum varies among lineages, this would provide a force for diversification of genome sizes.

The first time that the indel spectrum was investigated as a parameter in genome-size evolution was in the study of mammalian pseudogenes⁴⁰. There, spontaneous deletions outnumbered insertions and were longer on average. Intriguingly, the DNA loss was estimated to be faster in rodents than in humans, corresponding to the smaller rodent genomes. However, the overall rate of spontaneous DNA loss in mammals turned out to be so low as to appear a minor parameter in genome-size evolution^{40–42}.

Use of nonfunctional sequences, such as bona fide pseudogenes, is essential in the study of all spontaneous mutations, including indels. The reason is that occurrence of mutations in nonfunctional sequences is affected only by how frequently they are formed and not by natural selection for the information content, whereas varied selective effects of different kinds of mutation in functional sequences confounds meaningful inference. Unfortunately, pseudogenes are very rare in most model organisms (e.g. in *Drosophila*) and are not available in less well characterized genomes, preventing us from studying indel spectra (and other kinds of mutation) in diverse organisms.

Even though many eukaryotes do not have a large number of bona fide pseudogenes, practically all carry identifiable non-functional DNA. In particular, non-LTR retrotransposable elements are ubiquitous in eukaryotes (with the exception of yeast) and commonly generate non-functional, transpositionally defunct copies. We recently showed that it is possible to use maximum parsimony analysis to separate mutations that occur in active, master lineages (that are subject to selection) from those that occur in the

inactive, pseudogene-like copies (that are neutral)^{37,43}. The set of mutations in inactive copies can then be used to estimate mutational patterns.

The initial application of this method confirmed that spontaneous deletions are more frequent than insertions, but also revealed a striking difference in the indel spectra between *Drosophila* and mammals, with 60-fold faster DNA loss in *Drosophila* corresponding to its much more compact genome^{37,44}. The rate of DNA loss in *Drosophila* is also quite substantial in absolute terms, resulting in the loss of 50% of pseudogene DNA in approximately 14 Myr. To test further whether indel spectra affect genome size, we used the non-LTR approach in another group of insects, Hawaiian crickets of the genus *Laupala*, which have genomes that are 11 times larger on average than those of *Drosophila* species. Our study revealed a 40-fold lower rate of DNA loss in these crickets than in *Drosophila*, in agreement with the genome-size difference⁴⁵.

The use of pseudogenes avoids the problem of natural selection for information content of a sequence, but not the problem of selection acting on the bulk of DNA. Thus the difference in observed indel spectra could be due either to the differences in mutational mechanisms or, just as in the case of TE accumulation, to differences in natural selection for genome size⁴⁶. However, there is strong evidence that natural selection for genome size, acting directly on individual indels, cannot explain the observed differences in indel patterns^{43,45,47}. Deletions of different sizes do not appear to have different probabilities of persistence in the population as would be predicted if natural selection acted on them on the basis of their effect on genome size. Strong selection on small deletions and insertions would be expected to act even more strongly on much larger insertions of TEs in the same studies, which was not observed. In addition the deletion profiles change extremely abruptly (e.g. 3- to 5-bp deletions are found in equal frequencies in *Drosophila* and mammals, whereas deletions of 6- to 8-bp are 25-fold more frequent in *Drosophila*) – much more abruptly than predicted by selective scenarios.

Because non-LTR elements are ubiquitous in eukaryotes and can be cloned easily^{45,48}, the non-LTR-element-based method can be used to study mutational patterns in a comprehensive range of organisms. However, one should not ignore alternative sources of unconstrained DNA. Indeed, recent studies of indel frequency and size spectra in *Caenorhabditis elegans* using a large pseudogene family⁴⁹ and in brown mountain grasshoppers using nuclear insertions of mitochondrial DNA⁵⁰ show that it can be done. These studies add further evidence that differences in indel spectra could indeed underlie some changes in genome size. Clearly, it is not important exactly which unconstrained sequences are used to assay indel spectra. What is essential is to study indel patterns in a phylogenetically broad array of organisms to answer the question of how much of the variation in eukaryotic genome sizes is attributable to the variation in indel spectra.

Conclusion

The question of the C-value paradox has puzzled us for almost half a century. For much of this time the debate has centered on whether the vast amounts of noncoding DNA have any functional, adaptive role. This question needs to be settled if we are to understand fully the evolution of genome size. However, I believe the primary focus on this one question distracts us from other essential questions. How important is genetic drift in genome-size evolution? Do mutational rates of DNA addition and loss vary between different organisms independently of natural selection for small or large genome size? If yes, do mutational rates of DNA addition and loss correlate with genome size?

These questions are much more mechanistic and address the process of genome-size evolution rather than the functional significance of genome size. One possible reason for their exclusion from debate is that experimental approaches that can be used to address them were lacking until now. However, recent studies on the long-term rates of TE mobilization and on the rates of small deletions and insertions discussed in this review are just two examples of these new experimental approaches. I believe that we are getting closer to being able to study all possible evolutionary scenarios directly. Such studies are bound to disclose much information on the process of genome-size evolution and might even clarify the question of the possible adaptive significance of 'extra' DNA in eukaryotic genomes.

Acknowledgements

I thank M. Siegal, D. Hartl, D. Bensasson, T. Gregory and E. Zuckerkandl for very helpful comments. Comments by the two anonymous referees and the editor helped to improve the manuscript significantly.

References

- Mirsky, A.E. and H. Ris (1951) The DNA content of animal cells and its evolutionary significance. *J. Gen. Physiol.* 34, 451–462
- Thomas, C.A. (1971) The genetic organization of chromosomes. *Annu. Rev. Genet.* 5, 237–256
- John, B. and Miklos, G.L.G. (1988) *The Eukaryotic Genome in Development and Evolution*, Allen & Unwin
- Bennett, M.D. (1971) The duration of meiosis. *Proc. R. Soc. Lond. B Biol. Sci.* 178, 259–275
- 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, 247–278
- Vinogradov, A.E. (1998) Buffering: a possible passive-homeostasis role for redundant DNA. *J. Theor. Biol.* 193, 197–199
- Hsu, T.C. (1975) A possible function of constitutive heterochromatin: the bodyguard hypothesis. *Genetics* 79 (Suppl.), 137–150
- Ohno, S. (1972) So much 'junk' in our genomes. In *Evolution of Genetic Systems* (Smith, H.H. ed.), pp. 366–370, Brookhaven Symp. Biol.
- Orgel, L.E. and Crick, F.H.C. (1980) Selfish DNA: the ultimate parasite. *Nature* 284, 604–607
- Doolittle, W.F. and Sapienza, C. (1980) Selfish genes, the phenotype paradigm and genome evolution. *Nature* 284, 601–603
- O'Neill, R.J. *et al.* (1998) Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. *Nature* 393, 68–72
- Bingham, P.M. *et al.* (1982) The molecular basis of P-M hybrid dysgenesis: the role of the P element, a P-strain specific transposon family. *Cell* 29, 995–1004
- Finnegan, D.J. and Fawcett, D.H. (1986) Transposable elements in *Drosophila melanogaster*. In *Oxford Surveys of Eukaryotic Genes* (N. MacLean, ed.), pp. 1–62, Oxford University Press
- Petrov, D.A. *et al.* (1995) Diverse transposable elements are mobilized in hybrid dysgenesis in *Drosophila virilis*. *Proc. Natl. Acad. Sci. U. S. A.* 92, 8050–8054
- Walbot, V. *et al.* (1988) Regulation of mutator activities in maize. *Basic Life Sci.* 47, 121–135
- Kalendar, R. *et al.* (2000) From the cover: genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. *Proc. Natl. Acad. Sci. U. S. A.* 97, 6603–6607
- Petrov, D.A. and Hartl, D.L. (1998) High rate of DNA loss in the *Drosophila melanogaster* and *Drosophila virilis* species groups. *Mol. Biol. Evol.* 15, 293–302
- Keyl, H.G. (1965) A demonstrable local and geometric increase in the chromosomal DNA of *Chironomus*. *Experientia* 21, 191–193
- Jones, R.N. and Rees, H. (1968) Nuclear DNA variation in *Allium*. *Heredity* 23, 591–605
- Narayan, R.K.J. (1982) Constraints upon the organization and evolution of chromosomes in *Allium*. *Theor. Appl. Genet.* 75, 319–329
- Ohri, D. *et al.* (1998) Evolution of genome size in *Allium* (Alliaceae). *Plant Syst. Evol.* 1998, 57–86
- Ohri, D. and Khoshoo, T.N. (1986) Plant DNA contents and systematics. In *DNA Systematics*. (Dutta, S.K., ed.), pp. 2–19, CRC Press
- Uozu, S. *et al.* (1997) Repetitive sequences: cause for variation in genome size and chromosome morphology in the genus *Oryza*. *Plant Mol. Biol.* 35, 791–799
- Ogata, H. *et al.* (1996) The size differences among mammalian introns are due to the accumulation of small deletions. *FEBS Lett.* 390, 99–103
- Moriyama, E.N. *et al.* (1998) Genome size and intron size in *Drosophila*. *Mol. Biol. Evol.* 15, 770–773
- Ashburner, M. (1989) *Drosophila: A Laboratory Handbook*, Cold Spring Harbor Laboratory Press.
- Powell, J.R. (1997) *Progress and Prospects in Evolutionary Biology: the Drosophila Model*, Oxford University Press
- Deutsch, M. and Long, M. (1999) Intron-exon structures of eukaryotic model organisms. *Nucleic Acids Res.* 27, 3219–3228
- Vinogradov, A.E. (1999) Intron-genome size relationship on a large evolutionary scale. *J. Mol. Evol.* 49, 376–384
- Crollius, H.R. *et al.* (2000) Characterization and repeat analysis of the compact genome of the freshwater pufferfish *Tetraodon nigroviridis*. *Genome Res.* 10, 939–949
- Brenner, S. *et al.* (1993) Characterization of the pufferfish (*Fugu*) genome as a compact model vertebrate genome. *Nature* 366, 265–268
- Smit, A.F. (1999) Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr. Opin. Genet. Dev.* 9, 657–663
- SanMiguel, P. *et al.* (1998) The paleontology of intergene retrotransposons of maize. *Nat. Genet.* 20, 43–5
- Hancock, J.M. (1996) Simple sequences and the expanding genome. *BioEssays* 18, 421–425
- Weiner, A.M. *et al.* (1986) Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. *Annu. Rev. Biochem.* 55, 631–661
- Schlötterer, C. and Harr, B. (2000) *Drosophila virilis* has long and highly polymorphic microsatellites. *Mol. Biol. Evol.* 17, 1641–1646
- Petrov, D.A. *et al.* (1996) High intrinsic rate of DNA loss in *Drosophila*. *Nature* 384, 346–349
- SanMiguel, P. *et al.* (1996) Nested retrotransposons in the intergenic regions of the maize genome. *Science* 274, 765–768
- Ku, H.M. *et al.* (2000) Comparing sequenced segments of the tomato and *Arabidopsis* genomes: large-scale duplication followed by selective gene loss creates a network of synteny. *Proc. Natl. Acad. Sci. U. S. A.* 97, 9121–9126
- Graur, D. *et al.* (1989) Deletions in processed pseudogenes accumulate faster in rodents than in humans. *J. Mol. Evol.* 28, 279–285
- Gu, X. and Li, W.-H. (1995) The size distribution of insertions and deletions in human and rodent pseudogenes suggests the logarithmic gap penalty for sequence alignment. *J. Mol. Evol.* 40, 464–473
- Robertson, H.M. and Martos, R. Molecular evolution of the second ancient human mariner transposon, Hsma2, illustrates patterns of neutral evolution in the human genome lineage. *Gene* 205, 219–228
- Petrov, D.A. and Hartl, D.L. (1997) Trash DNA is what gets thrown away: High rate of DNA loss in *Drosophila*. *Gene* 205, 279–289
- Petrov, D.A. and Hartl, D.L. (1998) High rate of DNA loss in the *D. melanogaster* and *D. virilis* species groups. *Mol. Biol. Evol.* 15, 293–302
- Petrov, D.A. *et al.* (2000) Evidence for DNA loss as a determinant of genome size. *Science* 287, 1060–1062
- Charlesworth, B. (1996) The changing sizes of genes. *Nature* 384, 315–316
- Petrov, D.A. and Hartl, D.L. (2000) Pseudogene evolution and natural selection for a compact genome. *J. Hered.* 91, 221–227
- Wright, D.A. *et al.* (1996) Multiple non-LTR retrotransposons in the genome of *Arabidopsis thaliana*. *Genetics* 142, 569–578
- Robertson, H.M. (2000) The large srh family of chemoreceptor genes in *Caenorhabditis nematodes* reveals processes of genome evolution involving large duplications and deletions and intron gains and losses. *Genome Res.* 10, 192–203
- Bensasson, D. *et al.* Genomic gigantism: DNA loss is slow in mountain grasshoppers. *Mol. Biol. Evol.* (in press)