Full text provided by www.sciencedirect.com
SCIENCE DIRECT®

ELSEVIER

Evolution of genome size: multilevel selection, mutation bias or dynamical chaos?

Alexander E Vinogradov

In the past two years, new data on conceptual aspects of the evolution of eukaryotic genome size have appeared, including the adaptivity of genome enlargement, the mechanisms of genome size change and the relation of genome size to organismal complexity. New data on the hypotheses of 'selfish DNA' and 'mutational equilibrium' have been recently obtained. A relationship is emerging between the intragenomic distribution of noncoding DNA and differential gene expression, which suggests that noncoding DNA is involved in epigenetic organization of the genome and organismal complexity. The standpoint of dynamical chaos, which integrates multilevel selection and mutation biases, may provide a framework for studying the evolution of genome size.

Addresses
Institute of Cytology, Russian Academy of Sciences, Tikhoretsky Avenue 4, St Petersburg 194064, Russia
e-mail: aevin@mail.cytspb.rssi.ru

Current Opinion in Genetics & Development 2004, **14**:620–626

This review comes from a themed issue on
Genomes and evolution
Edited by David Haig and Steve Henikoff

Available online 5th October 2004

0959-437X/$ – see front matter
© 2004 Elsevier Ltd. All rights reserved.

DOI 10.1016/j.gde.2004.09.007

Abbreviations
LINE1 long interspersed nuclear element 1

Introduction

Elucidating the evolution of genome size in eukaryotes is, in most part, a problem of variation in the amount of noncoding DNA. The genome of yeast is about four orders of magnitude smaller than those of some plants, amphibians and lungfishes, but only one order of magnitude can be explained by differences in the number of genes. Several important questions remain to be answered in the field of genome size evolution. First, is the accumulation of noncoding DNA in the genome adaptive or just tolerated by selection at the organismal level? Second, if this accumulation is adaptive, what benefits does it confer on the organism? Third, what mechanisms underlie genome size changes (and especially, genome compaction)? Fourth, is genome size related to organismal complexity? Last, does the distribution of noncoding DNA in the genome have physiological significance? Recently, new insights have been obtained into all of these questions.

In this article, I review studies on the evolution of eukaryotic genome size that have been published mostly in the past two years. My emphasis is on conceptual aspects (rather than on new genome size measurements), such as the adaptivity of genome enlargement, the mechanisms of genome size change, and the relation of genome size to organismal complexity. I also discuss the relationship between the intragenomic distribution of noncoding DNA and differential gene expression, which suggests that noncoding DNA is involved in epigenetic organization of the genome and organismal complexity. Finally, I propose that dynamical chaos, which integrates multilevel selection and mutation biases, may provide a framework for elucidating issues of genome size evolution.

Adaptability and phenotypic traits associated with genome size

More than 20 years ago, it was suggested that the accumulation of redundant DNA in the genome might result from the activity of selfish intragenomic elements (behaving as Darwinian units) and might be merely tolerated by selection at the organismal level [1,2]. For a long time, this hypothesis of 'selfish DNA' remained untested, because it is difficult to disentangle selection acting at different levels. The appeal to selfish DNA has been even dismissed as ''a narrative scheme'' that is ''untestable and therefore not a hypothesis'' [3].

Recently, however, it has been shown that threatened plant species (those that are on the brink of extinction) have, on average, larger genomes than their more secure relatives. This finding suggests that redundant DNA in the plant genome might increase the likelihood of extinction [4•]. The effect has been found to be (at least partially) independent of the duration of plant life cycle. Moreover, polyploidy has been found not to be associated with an increased risk of extinction, suggesting that it is the accumulation of noncoding DNA in the genome that threatens species survival [4•]. At the same time, plant genomes seem generally to enlarge during the course of evolution (although there are exceptions in some lineages [5,6]).

A small genome size is ancestral for most major clades of angiosperms [7•]. There is a negative correlation between the mean genome size of angiosperm families and the

upper limit of their first appearance in geological time (i.e. more recent families have larger genomes) [4•]. Taken together, these data indicate that antagonism exists between different levels of selection, thereby supporting the concept of multilevel (hierarchical) selection [8,9]; however, they do not mean that the presence of all noncoding DNA in the eukaryotic genome can be explained by the activity of 'selfish DNA'. The picture is more complicated in vertebrates: in reptiles and birds, there is a correlation between genome size and increased risk of extinction at the within-order and within-family levels; in fishes and amphibians, this correlation is found only at higher taxonomic levels (with an insufficient number of degrees of freedom); and in mammals, it is not found at all [10].

Genome size is known to relate to several phenotypic traits. The most universal trait associated with genome enlargement is an increase in nucleus and cell size coupled with a retardation of the cell cycle and development [11,12]. Although it is easy to imagine an adaptive advantage of more rapid development, an adaptive function of retarded development caused by the accumulation of ballast seems doubtful [4•]. Probably, this accumulation can be adaptive only if it is associated with some compensating adaptation, for example, a reduction in metabolic rate (owing both to a lower cell surface to volume ratio and, according to the concept of 'buffering DNA', to an energy-independent attenuation of environmental fluctuations that can reach the nucleus), which allows a species to occupy an ecological niche with a lower energy supply [13–16].

A lower body-mass-corrected metabolic rate in animals with larger genomes has been recently demonstrated both in birds [17•], for which previously there were only limited data [15,18], and in reptiles [19•]. The relationship between genome size and metabolic rate has facilitated a novel explanation of the interspecies allometry of metabolic rate as a result of the evolutionary diversification of genome size in narrow taxonomic groups [20•]. The link between genome size and body-mass-corrected metabolic rate was recently confirmed in mammals taking into account a non-linearity of the relationships between parameters under study [21]. At the same time, a similar analysis of genome size in amphibians has proved inconclusive [22], probably because of the strong dependence of metabolic rate on temperature and the adaptation of different species to different temperatures.

Increased longevity, which can also be considered as an adaptive trait, has been recently reported for reptiles and fishes with larger genomes [19•,23] (see also Update). This was suggested to be a consequence of their lower metabolic rate. A positive correlation has been found between genome size and a number of genes encoding rRNA [24], which may reflect a need for augmented ribosomal machinery in organisms with larger genomes to provide for their larger cell sizes.

Mechanisms of genome size change

Transposable elements and indels

Although it is easy to explain a gain in noncoding DNA through the activity of transposable elements and segmental genomic duplications, the mechanisms underlying DNA loss remain vague. A hypothesis of 'mutational equilibrium' has been recently proposed [25•] on the grounds of previously reported observations (from a rather limited data set) that small (<400 bp) deletions occur more frequently than do small insertions and that the strength of this bias is negatively correlated with genome size [26,27]. According to this hypothesis (which is reminiscent of 'mutationism' by De Vries), genome size is determined by a balance between the propagation of transposable elements that cause genome enlargement and the bias in favor of small deletions that reduce the size of genome [25•].

Data obtained recently on the evolution of pufferfish genomes, however, contradict this hypothesis [28•]. Smooth pufferfish (Tetraodontidae) have the most compact genomes among vertebrates (~400 Mb), whereas spiny pufferfish (Diodontidae) and the closest relatives of both groups (Molidae) have genomes that are twice as large. It is therefore assumed that it is the genomes of smooth pufferfish that have experienced contraction rather than the genomes of the spiny pufferfish that have grown [28•]. According to the hypothesis of 'mutational equilibrium' [25•], genome compaction in smooth pufferfish should be due to a higher bias towards 'indels' (small insertion/deletions of less than 400 bp). Not only was the indel bias found to be similar in the two groups, but also the actual DNA loss in smooth pufferfish caused by small indels was found to be only a few per cent of the genome size [28•], which could not explain the twofold genome compaction in smooth pufferfish even in the absence of indel bias in spiny pufferfish and in the absence of transposable element propagation in smooth pufferfish (and suggests that some larger deletions were involved). At the same time, it is known that most families of transposable elements described in teleost fishes are present in smooth pufferfish and that some of them (e.g. *Zebulon*) were active after divergence of smooth pufferfish from spiny pufferfish [29•]. Similarly, transposable elements are active in the very compact genomes of *Arabidopsis* and *Drosophila* [30•,31,32]. (The hypothesis of 'mutational equilibrium' has been also criticized in [33].)

Larger deletions

Analysis of genome-wide alignments between human and mouse have indicated that most nonaligning regions that are not identifiable as insertions of lineage-specific transposable elements represent deletions in the other species that have occurred since divergence from the common

ancestor [34,35•]. It has been shown that such deletions in the mouse genome correspond to about a half of the human genome [35•]. Taking into account the fact that a significant part of the human genome has grown through the activity of primate-specific transposable elements (15–20%) [36•], and that transposition has been even more active in the mouse lineage [34], it can be concluded that there have been many more deletions in the mouse genome than is necessary to explain the roughly 15% difference in size between human and mouse genomes. Thus, it is clear that whatever the actual deletion mechanisms, the process of genome evolution is not short of them.

It has been shown that, in contrast to insertions, deletions require two breakpoints in the genomic sequence, which make them potentially more deleterious because they are more likely to involve a coding region [25•]. It remains to be seen whether there are special deletion mechanisms that are preferentially targeted at noncoding DNA. Unequal recombination between two repeats located in the same intergenic spacer or intron could provide such a mechanism. Recently, many examples of unequal recombination mediated by DNA repeats and leading to deletions have been reported [29•,37–40], but it remains to be elucidated whether the likelihood of such events being confined within the same segment of noncoding DNA is higher than of events involving a coding region. Illegitimate recombination involving very short repeats has been also shown to be significant in genome compaction in *Arabidopsis* [30•].

Unexpectedly, it has been found that in cultured human cells, retrotransposition of LINE1 (long interspersed nuclear) elements (a family of retroposons of ~6 kb that is very abundant in mammalian genomes) can cause deletions that are much larger (>70 kb) than the LINE1 elements themselves [41•]. Gilbert *et al.* [41•] concluded that LINE1 retrotransposition can act to reduce genome size. It is interesting that LINE1 elements are excluded from genic regions [42], which suggests that LINE1-mediated deletions might be preferentially located in noncoding DNA.

Multilevel selection

The concept of multilevel selection may be also relevant to explain DNA loss. In multicellular organisms, an additional selection level is presented by germ cells. Genome compaction might be favored during the proliferation of germ cells because proliferation accelerates the cell cycle. At the same time, benefits from a reduced metabolic rate caused by genome enlargement (which could have an advantage at the organismal level owing to the opportunity to occupy a niche with a lower energy supply) would not be important for selection during germ cell proliferation because these cells are equally supplied with energy. Moreover, the nucleus is a dead load in sperm. Sperm competition can lead to a decrease in sperm size and an increase in sperm number [43•], which suggests that this might be a genome-reducing force. Thus, selection at the level of germ cells could act to counterbalance the pressure of transposable elements that inflates the genomes of multicellular organisms, especially in large-body organisms with weakened organismal-level selection (owing to lower population sizes and longer life cycles).

'Dynamical chaos' model

Although the 'mutation equilibrium' hypothesis in its present form cannot explain most changes in genome size, the correlation of indel bias with genome size remains intriguing. It is possible that other mutation mechanisms involving genome size may have similarly correlated biases. Insertions or deletions in noncoding DNA that do not disrupt functionally significant regions are likely to be associated with very small selection coefficients (even if the indel is dozens of kilobases), because the indel accounts for only a tiny fraction of the whole genome. It is unlikely, however, that changes in genome size can be explained by purely mutational forces because there are clear indications that selection also has a role. For example, the distribution of genome sizes in birds and mammals is consistent with strong stabilizing selection [44•]. In plants there are many examples of ecological correlates of genome size [11].

It is thus tempting to suggest that mutation biases themselves can be determined by a selection for DNA-handling enzymes and thereby used to change genome size in an ecophysiologically relevant direction or to counterbalance the activity of selfish genetic elements in the case of stabilizing organismal-level selection. Such a combined evolutionary mechanism would invoke positive and negative feedbacks that are characteristic of dynamical chaos [45]. The positive feedback might be caused by a selection for genome size change enforced by a selection-formed mutation bias in the same direction, whereas the negative feedback would appear owing to stabilizing selection after reaching an optimum genome size. The dynamical chaos should differ both from a 'non-inertial' selection and from pure mutational mechanisms by oscillation that might occur after reaching an optimum genome size.

An analogy can be made with the GC content of the genome (another 'bulk' genomic trait). The isochores (genomic regions differing in GC content) in the genomes of mammals and birds are supposed to have physiological significance with regard to gene regulation [46]. At the same time, it is likely that some kind of mutation bias had a role in their formation (after all, selection for each separate nucleotide in noncoding DNA would incur a too high genetic load) [47]. It has been shown that the isochoric structure may have deteriorated in the most

recent period of mammalian evolution [47,48], which could be interpreted as a backward movement after reaching an optimum level [46]. Transposable elements of different age could be used as markers to determine whether there was a similar oscillation in genome size in a phylogenetic group that experienced a past genome size change caused by a general shift in ecophysiology (e.g. bats).

Organismal complexity and distribution of noncoding DNA

Genome size does not correlate closely with organismal complexity. This observation has been termed the 'C-value paradox'. However, the genomes of more complex organisms are, on average, larger than the genomes of less complex organisms, and it is therefore not certain that no part of additional noncoding DNA is involved in complexity. The evolution of multicellular organisms was accompanied by a great increase in the complexity of transcriptional regulation [49]. The regulatory sequences (promoters, enhancers, silencers and insulators) of a given gene can be scattered over distances of roughly 10 kb in the *Drosophila* and 100 kb in the human genome [49]. This difference agrees well with the difference in size between *Drosophila* and human genomes (117 Mb versus 2910 Mb, respectively, for euchromatic parts), taking into account the several-fold higher number of genes in the human genome.

It has been recently proposed that the genome complexity (caused by the retention of duplicate genes and transposable elements) of eukaryotes emerged by random genetic drift (non-adaptively) in response to the long-term population-size reductions that accompanied the increases in organism size [50]. In a test of this hypothesis, however, an outcome contrary to the authors' predictions has been so far observed, which by reverse inference supports an adaptive interpretation of genome enlargement [51]. Nevertheless, the accumulation of noncoding DNA in the genome, even if it first happened accidentally, might become a trap: in addition to the fact that transposable elements form 'landing pads' for other such elements, the longer life cycles and the lower population sizes caused by genome enlargement might attenuate purifying selection against larger genomes, thereby simulating a neutralist effect of 'permissive' evolution [4•].

The issue of the relationship between genome size and organismal complexity has been recently raised in the field of functional genomics. It has been found that highly and broadly expressed genes (i.e. genes that are expressed in many tissues) have shorter intronic and coding sequences than do genes expressed in a tissue-specific fashion [52•–55•]. Because transcription and translation are energetically costly, this shortness has been interpreted as a result of selection for economy [52•–54•]. The 'economy' hypothesis implicitly assumes a neutralist (permissive) interpretation of the accumulation of DNA in the eukaryotic genome. The incessantly expressed genes are supposed to 'slim down' (selection condition), whereas those that work less intensively are thought to 'get fat' (permissive condition). It has been argued, however, that housekeeping genes are not becoming shorter but rather tissue-specific genes are getting longer [55•]. The tissue-specific proteins have more complex architectures that explain the increase in their length [55•]. Not only introns, but also intergenic spacers around highly expressed and housekeeping genes are similarly (or even more regularly) shorter than those around tissue- and development-specific genes [54•,55•]. As a result, the 'gene nest' proportion (the ratio of intra- plus intergenic noncoding to coding DNA lengths) negatively correlates with the breadth and level of gene expression [55•]. The greater amount of intra- and intergenic noncoding DNA, in which the tissue- and development-specific genes are embedded, is supposed to be involved in the more complex regulation and chromatin-mediated suppression of these genes (especially when possible surges of chromatin decondensation caused by fluctuations of intracellular solute composition are taken into account), thereby participating in the epigenetic organization of the genome and transcriptional noise suppression [55•].

An exemplar intergenic spacer has been carefully studied in the *Drosophila melanogaster* species complex and has revealed a rapid DNA sequence turnover [56•]. The length of the spacer initially increased during the evolution of these species because of several large insertions but subsequently decreased owing to negative indel bias. Singh and Petrov [56•] interpreted these findings as evidence for unconstrained evolution of noncoding DNA owing to random mutations, which is consistent with the 'mutational equilibrium' model. It is, however, the expression of genes that changes most rapidly in closely related species [57,58]. In *Drosophila* species, the change of expression is especially strong for sex-dependent genes [58]. One of the genes around the exemplar intergenic spacer is sex-specific, another is embryo-specific (and both are coded in the opposite direction with location of their potential 5′-upstream regulatory regions within this spacer) [56•]. It is therefore possible that structural changes in this intergenic spacer were accompanied by physiologically relevant changes in the expression of adjacent genes.

Conclusions

Notwithstanding the recent advancements, all of the issues raised in this review need further investigation. The following points remain to be elucidated: first, the universality of reduction in metabolic rate as a consequence of genome enlargement, the mechanisms of this reduction (in particular, whether there are other effects in addition to the mere change in cellular geometrical

parameters), and whether this reduction is a cause for other possible adaptive traits (increased longevity and stress-tolerance, etc.); second, the generality of the negative correlation between genome size and insertion/deletion mutation bias (not only with regard to small indels), and whether there are deletion mechanisms preferentially confined within the noncoding DNA; third, the role of multilevel selection in genome size changes and in formation of mutation biases; and last, the relevance of noncoding DNA amount to organismal and genome complexity, especially the role of its intragenomic distribution in differential gene expression.

Update

The report on increased longevity in fishes with larger genomes [23] was criticized on the ground of a more careful statistical analysis [59].

Recently, the correlation between the number of B chromosomes, the degree of outbreeding, and genome size was found in angiosperm plants [60•]. The link between the number of B chromosomes (which are believed to be selfish elements) and genome size is consistent with a weakened selection against the accumulation of redundant DNA in species with larger genomes [60•]. As for the degree of outbreeding, it is more ambiguous. The role of the breeding system in regard to adaptive versus neutralist interpretation of accumulation of 'selfish' DNA is inconclusive [4•]. On the one hand, selection against deleterious genetic changes is promoted by recombination and is generally more effective in sexual breeding systems as compared to asexual; on the other hand, transposable elements also propagate more successfully in sexual populations than in asexual ones (reviewed in [4•]).

Acknowledgements

This work was supported by the Russian Foundation for Basic Research (RFBR) and by the Programme of the Presidium of the Russian Academy of Sciences 'Molecular and Cellular Biology'.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Doolittle WF, Sapienza C: **Selfish genes, the phenotype paradigm and genome evolution**. *Nature* 1980, **284**:601-603.

2. Orgel LE, Crick FH: **Selfish DNA: the ultimate parasite**. *Nature* 1980, **284**:604-607.

3. Sternberg R: **On the roles of repetitive DNA elements in the context of a unified genomic–epigenetic system**. *Ann N Y Acad Sci* 2002, **981**:154-188.

4. • Vinogradov AE: **Selfish DNA is maladaptive: evidence from the plant red list**. *Trends Genet* 2003, **19**:609-614.
Threatened plant species are shown to have larger genomes than their more secure relatives. The effect is (at least partially) independent of the duration of plant life cycle. Polyploidy is found not to be associated with an increased risk of extinction, suggesting that it is the accumulation of noncoding DNA in the genome that threatens species survival. In addition, using the fossil record database, the author finds a negative correlation between mean genome size of angiosperm families and the upper limit of their first appearance in geological time (i.e. more recent families have larger genomes). Taken together, these data support the hypothesis that 'selfish DNA' inflated the genomes of plants during the course of evolution and indicate an antagonism between different selection levels.

5. Bennetzen JL: **Mechanisms and rates of genome expansion and contraction in flowering plants**. *Genetica* 2002, **115**:29-36.

6. Wendel JF, Cronn RC, Johnston JS, Price HJ: **Feast and famine in plant genomes**. *Genetica* 2002, **115**:37-47.

7. • Soltis DE, Soltis PS, Bennett MD, Leitch IJ: **Evolution of genome size in the angiosperms**. *Am J Bot* 2003, **90**:1596-1603.
Using a molecular phylogenetic framework and genome size measurements for crucial basal angiosperms (which were not previously available), the authors show that a very small genome size (~1.4 pg) is ancestral not only for angiosperms in general, but also for most major clades of them, including monocots and eudicots.

8. Gould SJ, Lloyd EA: **Individuality and adaptation across levels of selection: how shall we name and generalize the unit of Darwinism?** *Proc Natl Acad Sci USA* 1999, **96**:11904-11909.

9. Gregory TR: **Macroevolution, hierarchy theory, and the C-value enigma**. *Paleobiology* 2004, **30**:179-202.

10. Vinogradov AE: **Genome size and extinction risk in vertebrates**. *Proc R Soc Lond B Biol Sci* 2004, **271**:1701-1705.

11. Bennett MD: **Plant genome values: how much do we know?** *Proc Natl Acad Sci USA* 1998, **95**:2011-2016.

12. Petrov DA: **Evolution of genome size: new approaches to an old problem**. *Trends Genet* 2001, **17**:23-28.

13. Szarski H: **Cell size and the concept of wasteful and frugal evolutionary strategies**. *J Theor Biol* 1983, **105**:201-209.

14. Cavalier-Smith T: **Cell volume and the evolution of eukaryote genome size**. In *The Evolution of Genome Size.* Edited by Cavalier-Smith T: Chichester: John Wiley & Sons; 1985:105-184.

15. Vinogradov AE: **Nucleotypic effect in homeotherms: body mass-corrected basal metabolic rate of mammals is related to genome size**. *Evolution* 1995, **49**:1249-1259.

16. Vinogradov AE: **Buffering: a possible passive-homeostasis role for redundant DNA**. *J Theor Biol* 1998, **193**:197-199.

17. • Gregory TR: **A bird's-eye view of the C-value enigma: genome size, cell size, and metabolic rate in the class Aves**. *Evolution* 2002, **56**:121-130.
A negative correlation between metabolic rate (corrected for body mass) and genome size is shown in birds.

18. Vinogradov AE: **Nucleotypic effect in homeotherms: body mass-independent resting metabolic rate of passerine birds is related to genome size**. *Evolution* 1997, **51**:220-225.

19. • Olmo E: **Reptiles: a group of transition in the evolution of genome size and of the nucleotypic effect**. *Cytogenet Genome Res* 2003, **101**:166-171.
A negative correlation between metabolic rate (corrected for body mass) and genome size is found in reptiles. This work also shows a positive correlation between genome size and longevity in reptiles.

20. • Kozlowski J, Konarzewski M, Gawelczyk AT: **Cell size as a link between noncoding DNA and metabolic rate scaling**. *Proc Natl Acad Sci USA* 2003, **100**:14080-14085.
The relationship between genome size and metabolic rate in mammals is used for a novel explanation of the interspecies allometry of metabolic rate as a result of evolutionary diversification of genome sizes within narrow taxonomic groups, which underlines the participation of cell size and cell number in body size optimization. This optimization is suggested to result in inverse relation between slopes of interspecific allometries of metabolic rates and genome size. The authors show that the correlation between allometric exponents of the basal metabolic rate and genome size is negative, in accordance with their suggestion.

21. Vinogradov AE, Anatskaya OV: **Phenological resonance and quantum life-history**. *J Theor Biol* 2004, **228**:417-420.

22. Gregory TR: **Variation across amphibian species in the size of the nuclear genome supports a pluralistic, hierarchical**

approach to the C-value enigma. *Biol J Linn Soc* 2003, **79**:329-339.

23. Griffith OL, Moodie GE, Civetta A: **Genome size and longevity in fish**. *Exp Gerontol* 2003, **38**:333-337.

24. Prokopowich CD, Gregory TR, Crease TJ: **The correlation between rDNA copy number and genome size in eukaryotes**. *Genome* 2003, **46**:48-50.

25. • Petrov DA: **Mutational equilibrium model of genome size evolution**. *Theor Popul Biol* 2002, **61**:531-544.
This paper presents the hypothesis of 'mutational equilibrium', which suggests that genome size is a result of mechanical balance between the propagation of transposable elements that causes genome enlargement and the negative bias of small indels that reduces the size of genome (without participation of selection). Although in its present form this concept cannot explain most changes in genome size, it could serve both as a useful null hypothesis against which new data can be checked and as a possible component of a more general theory.

26. Petrov DA, Sangster TA, Johnston JS, Hartl DL, Shaw KL: **Evidence for DNA loss as a determinant of genome size**. *Science* 2000, **287**:1060-1062.

27. Bensasson D, Petrov DA, Zhang DX, Hartl DL, Hewitt GM: **Genomic gigantism: DNA loss is slow in mountain grasshoppers**. *Mol Biol Evol* 2001, **18**:246-253.

28. • Neafsey DE, Palumbi SR: **Genome size evolution in pufferfish: a comparative analysis of diodontid and tetraodontid pufferfish genomes**. *Genome Res* 2003, **13**:821-830.
Indels and nucleotide substitutions are studied in the genomes of pufferfish. Smooth pufferfish (Tetraodontidae) have the most compact genomes among vertebrates (~400 Mb), whereas spiny pufferfish (Diodontidae), as well as the closest relatives of both groups (Molidae), have genomes that are twice as large. The authors therefore assume that it is the genomes of smooth pufferfish that have experienced contraction and not the genomes of spiny pufferfish that have grown. The indel bias is found to be similar in both groups. In addition, the actual DNA loss in the smooth pufferfish caused by indels is only a few per cent of the genome size.

29. • Bouneau L, Fischer C, Ozouf-Costaz C, Froschauer A, Jaillon O, Coutanceau JP, Korting C, Weissenbach J, Bernot A, Volff JN: **An active non-LTR retrotransposon with tandem structure in the compact genome of the pufferfish *Tetraodon nigroviridis***. *Genome Res* 2003, **13**:1686-1695.
This paper shows that fish retrotransposon *Zebulon* was active relatively recently in the genome of a smooth pufferfish, in which it contributed to the extension of intergenic and intronic sequences. These data indicate that even the compact genomes of smooth pufferfish are not evolutionarily inert and contain active retrotransposons. They also show that retrotransposable elements are accumulated in heterochromatin and excluded from euchromatic regions. Homologous recombination between partial tandem sequences of retrotransposons is supposed as a mechanism for eliminating retrotransposons from euchromatic regions.

30. • Devos KM, Brown JK, Bennetzen JL: **Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis***. *Genome Res* 2002, **12**:1075-1079.
The illegitimate recombination involving very short repeats was reported as a cause of genome reduction in the very compact *Arabidopsis* Genome

31. Wright SI, Agrawal N, Bureau TE: **Effects of recombination rate and gene density on transposable element distributions in *Arabidopsis thaliana***. *Genome Res* 2003, **13**:1897-1903.

32. Petrov DA, Aminetzach YT, Davis JC, Bensasson D, Hirsh AE: **Size matters: non-LTR retrotransposable elements and ectopic recombination in *Drosophila***. *Mol Biol Evol* 2003, **20**:880-892.

33. Gregory TR: **Is small indel bias a determinant of genome size?** *Trends Genet* 2003, **19**:485-488.

34. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P *et al.*: **Initial sequencing and comparative analysis of the mouse genome**. *Nature* 2002, **420**:520-562.

35. • Hardison RC, Roskin KM, Yang S, Diekhans M, Kent WJ, Weber R, Elnitski L, Li J, O'Connor M, Kolbe D *et al.*: **Covariation in frequencies of substitution, deletion, transposition, and recombination during eutherian evolution**. *Genome Res* 2003, **13**:13-26.
This paper shows that deletions in the mouse genome (defined as nonaligning regions in human–mouse genome-wide alignments that are not identifiable as insertions of lineage-specific transposable elements and are thought to represent deletions in the other species since divergence from the common ancestor) correspond to about a half of the human Genome

36. • Liu G, Zhao S, Bailey JA, Sahinalp SC, Alkan C, Tuzun E, Green ED, Eichler EE: **Analysis of primate genomic variation reveals a repeat-driven expansion of the human genome**. *Genome Res* 2003, **13**:358-368.
This paper presents a large-scale comparison of about 10 Mb of genomic sequence from lemur, baboon and chimpanzee with genomic sequence from human. It shows that there has been a 15–20% expansion of human genome over the past 50 million years of primate evolution, 90% of which is due to new retroposon insertions.

37. Babcock M, Pavlicek A, Spiteri E, Kashork CD, Ioshikhes I, Shaffer LG, Jurka J, Morrow BE: **Shuffling of genes within low-copy repeats on 22q11 (LCR22) by *Alu*-mediated recombination events during evolution**. *Genome Res* 2003, **13**:2519-2532.

38. Salem AH, Kilroy GE, Watkins WS, Jorde LB, Batzer MA: **Recently integrated *Alu* elements and human genomic diversity**. *Mol Biol Evol* 2003, **20**:1349-1361.

39. Read LR, Raynard SJ, Ruksc A, Baker MD: **Gene repeat expansion and contraction by spontaneous intrachromosomal homologous recombination in mammalian cells**. *Nucleic Acids Res* 2004, **32**:1184-1196.

40. Saitta SC, Harris SE, Gaeth AP, Driscoll DA, McDonald-McGinn DM, Maisenbacher MK, Yersak JM, Chakraborty PK, Hacker AM, Zackai EH *et al.*: **Aberrant interchromosomal exchanges are the predominant cause of the 22q11.2 deletion**. *Hum Mol Genet* 2004, **13**:417-428.

41. • Gilbert N, Lutz-Prigge S, Moran JV: **Genomic deletions created upon LINE-1 retrotransposition**. *Cell* 2002, **110**:315-325.
It is shown here that in the cultured human cells retrotransposition of LINE1 elements (a family of retroposons of ~6 kb that is very abundant in mammalian genomes) can cause deletions that are much larger (>70 kb) than LINE1 themselves. The authors suggest that retrotransposition of LINE1 could act to reduce genome size.

42. Medstrand P, van de Lagemaat LN, Mager DL: **Retroelement distributions in the human genome: variations associated with age and proximity to genes**. *Genome Res* 2002, **12**:1483-1495.

43. • Gage MJ, Morrow EH: **Experimental evidence for the evolution of numerous, tiny sperm via sperm competition**. *Curr Biol* 2003, **13**:754-757.
It is shown here that sperm competition can lead to a decrease in sperm size and an increase in sperm number. In principle, this process could reduce the size of genome, thereby presenting an additional level of selection (germ cells).

44. • Waltari E, Edwards SV: **Evolutionary dynamics of intron size, genome size, and physiological correlates in archosaurs**. *Am Nat* 2002, **160**:539-552.
Using analysis of phylogenetically independent contrasts and Monte Carlo simulations, the authors shown that the distribution of genome sizes in birds and mammals is underdispersed compared with the Brownian motion model and therefore consistent with strong stabilizing selection. Also, the negative correlation between (body mass-corrected) metabolic rate and genome size is demonstrated in amniotes and archosaurs.

45. Robertson DS: **Feedback theory and Darwinian evolution**. *J Theor Biol* 1991, **152**:469-484.

46. Vinogradov AE: **DNA helix: the importance of being GC-rich**. *Nucleic Acids Res* 2003, **31**:1838-1844.

47. Duret L, Semon M, Piganeau G, Mouchiroud D, Galtier N: **Vanishing GC-rich isochores in mammalian genomes**. *Genetics* 2002, **162**:1837-1847.

48. Arndt PF, Petrov DA, Hwa T: **Distinct changes of genomic biases in nucleotide substitution at the time of mammalian radiation**. *Mol Biol Evol* 2003, **20**:1887-1896.

49. Levine M, Tjian R: **Transcription regulation and animal diversity**. *Nature* 2003, **424**:147-151.

50. Lynch M, Conery JS: **The origins of genome complexity**. *Science* 2003, **302**:1401-1404.

51. Vinogradov AE: **Testing genome complexity**. *Science* 2004, **304**:389-390.

52. • Castillo-Davis CI, Mekhedov SL, Hartl DL, Koonin EV, Kondrashov FA: **Selection for short introns in highly expressed genes**. *Nat Genet* 2002, **31**:415-418.
Using data on gene expression in nematode and human, the authors show that introns in highly expressed genes are substantially shorter than in genes expressed at low levels, which is interpreted as a result of selection for minimization of transcription cost.

53. • Eisenberg E, Levanon EY: **Human housekeeping genes are compact**. *Trends Genet* 2003, **19**:362-365.
Introns, untranslated regions and coding sequences of human housekeeping genes are shown to be shorter than those of more narrowly expressed genes — a feature that is interpreted to be a result of selection for economy.

54. • Urrutia AO, Hurst LD: **The signature of selection mediated by expression on human genes**. *Genome Res* 2003, **13**:2260-2264.
This paper reports that highly expressed human genes encode smaller proteins, have less intronic DNA, and higher codon and amino acid biases — features that are interpreted to be a result of selection for economy. It is also shown that intergenic spacers tend to be shorter for highly expressed genes. The authors control for intergenic spacer length, by using multiple regression analysis, and still find a negative correlation of intronic length with expression level (contrary to what is found in [55•]). The causes of this discrepancy may be that (i) although most of the intronic length covaries with intergenic spacer length, there still remains some room for the effect of selection for economy; (ii) the multiple regression used is not adequate here because the parameters under study might not be normally distributed (a more straightforward method of control — ratio to intergenic spacer length — is used in [55•]).

55. • Vinogradov AE: **Compactness of human housekeeping genes: selection for economy or genomic design?** *Trends Genet* 2004, **20**:248-253.
This paper argues against the explanations presented in [52•–54•], presenting the new data.

56. • Singh ND, Petrov DA: **Rapid sequence turnover at an intergenic locus in *Drosophila***. *Mol Biol Evol* 2004, **21**:670-680.
An exemplar intergenic spacer is studied in the *Drosophila melanogaster* species complex. A rapid DNA sequence turnover is found: the length of the spacer initially increased during the evolution of these species because of several large insertions but subsequently decreased owing to negative indel bias. The authors interpret these findings as evidence for the unconstrained evolution of noncoding DNA owing to random mutations, which is consistent with the 'mutational equilibrium' model.

57. Enard W, Khaitovich P, Klose J, Zollner S, Heissig F, Giavalisco P, Nieselt-Struwe K, Muchmore E, Varki A, Ravid R *et al.*: **Intra- and interspecific variation in primate gene expression patterns**. *Science* 2002, **296**:340-343.

58. Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL: **Sex-dependent gene expression and evolution of the *Drosophila* transcriptome**. *Science* 2003, **300**:1742-1745.

59. Gregory TR: **Genome size is not correlated positively with longevity in fishes (or homeotherms)**. *Exp Gerontol* 2004, **39**:859-860.

60. • Trivers R, Burt A, Palestis BG: **B chromosomes and genome size in flowering plants**. *Genome* 2004, **47**:1-8.
The authors found a positive correlation between the number of B chromosomes, the degree of outbreeding, and genome size in angiosperm plants.