

# Genome size and extinction risk in vertebrates

Alexander E. Vinogradov

*Institute of Cytology, Russian Academy of Sciences, Tikhoretsky Avenue 4, St Petersburg 194064, Russia*  
(aevin@mail.cytspb.rssi.ru)

The hypothesis of 'selfish DNA' is tested for the case of animals using the relation between genome size and conservation status of a given species. In contrast to plants, where the larger genome was previously shown to increase the likelihood of extinction, the picture is more complicated in animals. At the within-families and within-orders levels, the larger genome increases the risk of extinction only in reptiles and birds (which have the smallest genomes among tetrapods). In fishes and amphibians, the effect is caused by the higher taxonomic levels (above order). In several phylogenetic lineages of anamniotes, there is a correlation between a higher fraction of threatened species and a lower number of extant species in a lineage with the larger genome. In mammals, no effect was observed at any taxonomic level. The obtained data support the concept of hierarchical selection. It is also shown that, in plants and reptiles, the probability of being threatened increases from less than 10% to more than 80% with the increase in genome size, which can help in establishing conservation priorities.

**Keywords:** selfish DNA; non-coding DNA; C-value paradox; hierarchical selection; macroevolution; biodiversity

## 1. INTRODUCTION

The role of non-coding DNA, which constitutes the major part of the eukaryotic genome (e.g. *ca.* 99% of the human genome; Venter *et al.* 2001), remains enigmatic. It is still unclear whether the growth of the genome is an adaptation or a consequence of intragenomic activity of transposable elements behaving as autonomous Darwinian units, as was suggested in the long-debated hypothesis of 'selfish DNA' (Doolittle & Sapienza 1980; Orgel & Crick 1980). Were the latter true, the vectors of the intragenomic selection, and selection at the higher levels may differ and the result of intragenomic propagation of selfish genetic units might become maladaptive at the higher selection levels. The threatened plant species (whose populations are on the decline) were shown to have, on average, larger genomes than their more secure relatives, which indicates that the excess of non-coding DNA in the plant genome increases the likelihood of extinction (Vinogradov 2003). This observation, which connected two actively developing fields—genomics and global ecology/species conservation—presented, to my knowledge, the first direct test of (mal)adaptivity of naturally occurring redundant DNA (at least, at the species level). It is important to know whether a similar trend exists in animals.

## 2. MATERIAL AND METHODS

Genome size values were taken from the Animal Genome Size Database (Gregory 2001). The polyploid and poorly determined species (records with 'sp.', 'af.', 'cf.' in the species field) were excluded, and the data of different authors and for infraspecies forms were averaged. The conservation status of these species was determined using the International Union for Conservation of Nature (IUCN) Red List of Threatened Species (<http://www.redlist.org>), which lists species of global conservation concern. Species in the following categories were

considered as threatened (and pooled together): EX (extinct), EW (extinct in the wild), CR (critically endangered), EN (endangered), VU (vulnerable), NT (near threatened) and CD (conservation dependent). The fractions of the total number of known species in different phylogenetic groups, which were assessed for conservation status, were also taken from the Red List database (<http://www.redlist.org/info/tables/table1.html>).

The analysis was made with the raw species points and, to control for phylogenetic conservatism, with within-taxon contrasts (i.e. ratios of genome size of a given species to mean genome size of the taxon to which it belongs, the taxon mean being determined by hierarchical stepwise averaging of genome size for lower taxonomic levels). In plants, genome-size contrasts of a level that is lower than the within-family taxonomic level were found to be unreliable because they were close to measurement errors and within-species variance (Vinogradov 2003). This conclusion is extended here to the animal dataset. The variance components analysis showed that variance of (log-transformed) genome size values was partitioned among taxonomic levels in the following way: within-species (mostly among-authors) 1.7%, among-species within-genera 1.6%, among-genera within-families 2.5%, among-families within-orders 3.4%, and among- and above-orders 90.8%. Thus, the within-genera variance is close to the within-species (among-authors) variance, whereas the majority of the variance resides at the among-orders level and higher. Therefore, only the within-orders and within-families contrasts were used (which, for the whole dataset, control for greater than 90% of phylogeny-related variance). The numbers of extant species (and the general list of families for analysis of palaeontological records) were taken from the Fish-Base database (<http://www.fishbase.org>), the Amphibian Species of the World Database (<http://research.amnh.org/herpetology/amphibia/index.html>) and the EMBL Reptile Database (<http://www.embl-heidelberg.de/~uetz/LivingReptiles.html>). The data on the first appearance in palaeontological records of different fish families were taken from the Fossil Record 2 database (<http://palaeo.gly.bris.ac.uk/frwhole/FR2.html>).

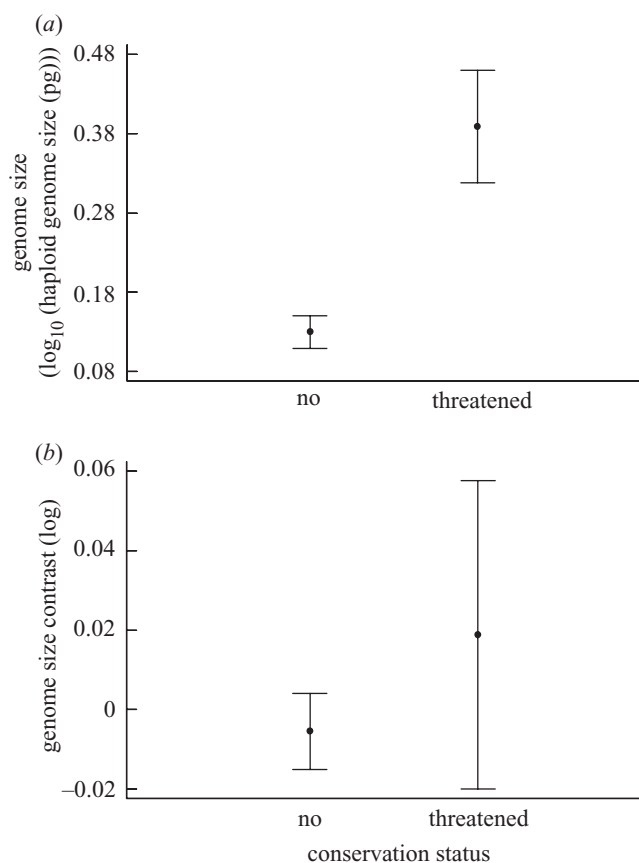


Figure 1. Genome size of fishes with different risks of extinction (means with 95% least significant difference (LSD) intervals). (a) Raw species points (Mann–Whitney,  $p < 10^{-10}$ ); (b) within-orders contrasts (i.e. ratios to orders means; Mann–Whitney,  $p > 0.4$ ).

### 3. RESULTS AND DISCUSSION

#### (a) *Fishes*

In fishes, 76 species were found to be of conservation concern from a total of 928 species for which genome size is known. Analysis of the raw species points showed that the threatened species have, on average, larger genomes (figure 1a). However, contrary to what is found in plants, neither the within-families or within-orders contrasts showed any significant effects, i.e. when genome size values were taken as ratios to families or orders means (figure 1b). It transpired that the effect with the raw species points was owing only to two phylogenetic contrasts: Elasmobranchii (cartilaginous fishes) versus Actinopterygii (ray-finned fishes), and Acipenseriformes (sturgeons) versus other ray-finned fishes. Elasmobranchii and Acipenseriformes have the larger genomes ( $5.6 \pm 0.5$  pg and  $3.1 \pm 0.8$  pg versus  $1.2 \pm 0.1$  pg, respectively) and the higher fractions of threatened species ( $0.24 \pm 0.08$  and  $1.0 \pm 0.1$  versus  $0.04 \pm 0.01$ ) as compared with (other) ray-finned fishes. Neither in the cartilaginous fishes taken separately, nor in the ray-finned fishes with exclusion of Acipenseriformes, did the effect with the raw species points remain valid. Thus, the representation in figure 1a is strongly affected by phylogenetic conservatism.

It is interesting that Elasmobranchii are much older than ray-finned fishes and now contain far fewer extant species. According to the Fossil Record 2 database, the

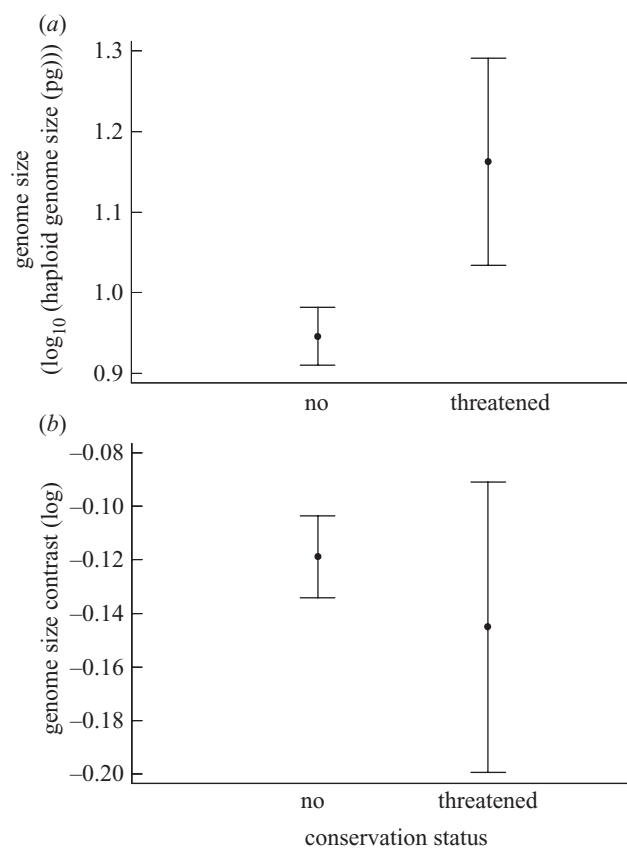


Figure 2. Genome size of amphibians with different risks of extinction (means with 95% LSD intervals). (a) Raw species points (Mann–Whitney,  $p < 0.02$ ); (b) within-orders contrasts (Mann–Whitney,  $p > 0.3$ ).

median upper limit of first appearance in palaeontological records of Elasmobranchii families is 90.4 Myr ago ( $n = 31$ ), whereas for ray-finned fishes it is only 42.1 Myr ago ( $n = 223$ ). According to the FishBase, there are only 939 species of cartilaginous fishes and 25 217 species of ray-finned fishes. Acipenseriformes are also relatively more ancient and show a lower number of extant species than most other ray-finned fishes. The median upper limit of first appearance for Acipenseriformes families is 62 Myr ago, and there are now only 28 species of them (whereas on average an order of ray-finned fishes contains 560 species). Although the lungfishes (which have the largest genomes among fishes) are not listed as threatened, there remain only a few species of them (lungfishes are very ancient: the median upper limit of first appearance of their families is 158 Myr ago). Thus, it seems that, at least at the level of higher taxa, there is now a negative relation of genome size to survival and speciosity in fishes. The older groups with the larger genomes have a higher fraction of threatened species and a lower number of extant species.

#### (b) *Tetrapods*

In amphibians (380 species in the genome size dataset, 28 of which are threatened), there is a negative relationship between genome size and extinction risk for the raw species points but no relationship for the within-families or within-orders contrasts (figure 2). Again, the effect with the raw species points was owing only to the higher taxa:

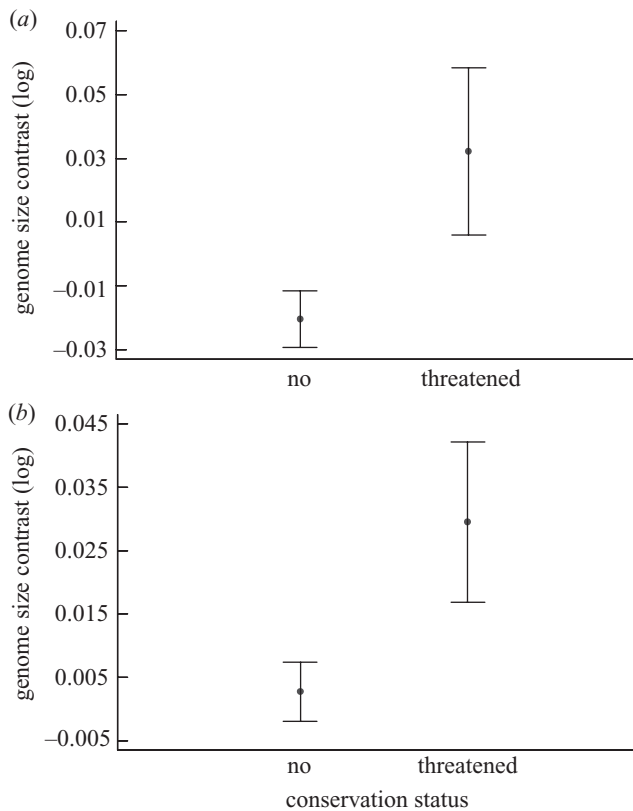


Figure 3. Genome size of reptiles and birds with different risks of extinction (within-orders contrasts, means with 95% LSD intervals). (a) Reptiles (Mann–Whitney,  $p < 0.01$ ); (b) birds (Mann–Whitney,  $p < 0.01$ ).

caudata (salamanders) versus anura (frogs). The former have the larger genomes ( $36.8 \pm 1.2$  pg versus  $4.5 \pm 1.0$  pg), a higher fraction of threatened species ( $0.12 \pm 0.02$  versus  $0.05 \pm 0.02$ ) and a lower number of extant species (502 versus 4837).

In reptiles (288 species in the dataset, 30 of which are threatened) and birds (158 species, 19 of which are threatened), the effect held both for the raw species points and for the within-families and within-orders contrasts (figure 3). Testudines (turtles) have the larger genomes ( $2.9 \pm 0.2$  pg versus  $2.1 \pm 0.1$  pg), a higher fraction of threatened species ( $0.60 \pm 0.15$  versus  $0.02 \pm 0.01$ ) and a lower number of extant species (305 versus 7835) as compared with squamata (lizards, snakes). Crocodiles also have larger genomes ( $3.1 \pm 0.7$  pg), a small number of extant species (23), and seem to have a higher fraction of threatened species ( $0.2 \pm 0.5$ ), although confidence intervals are too broad in the latter case. However, in reptiles the effect held at the level of the lower taxonomic levels as well (figure 3a). In birds, the effect is not owing to flightless birds (which have relatively larger genomes and therefore could show a spurious correlation with extinction risk).

In mammals (316 species, 67 of which are threatened), there was no effect in any type of analysis.

### (c) Plants

The database of threatened plant species used in previous work (UNEP-WCMC; Vinogradov 2003) and the database used here (IUCN) differ because of the implementation of a more recent version of the Red List

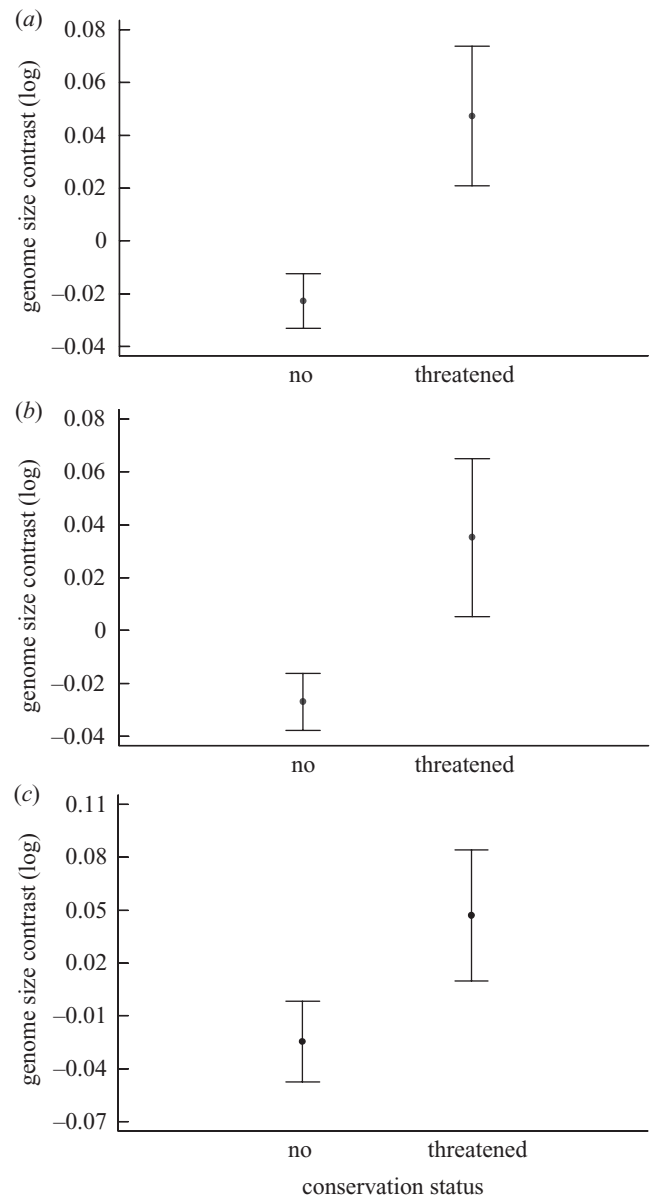


Figure 4. Genome size of plants with different risks of extinction (within-families contrasts, means with 95% LSD intervals). (a) All vascular plants (Mann–Whitney,  $p < 0.001$ ); (b) angiosperms (Mann–Whitney,  $p < 0.01$ ; ANOVA,  $p < 0.01$ ); (c) angiosperms, the effect of life cycle being removed using multifactor ANOVA with factor levels: annuals, biennials, perennials (ANOVA,  $p < 0.01$ ).

in the IUCN database. By contrast, not all plants presented in the UNEP-WCMC database are yet covered by the IUCN database, and it is recommended to use both databases for the finding and retrieval of plants of global conservation concern (<http://www.redlist.org/info/introduction.html>). Therefore, the plant case was rechecked here by combining both databases. The effect still held (figure 4a,b). For angiosperms, there are also data on the life-cycle duration (Bennett & Leitch 2003), which is known to be strongly associated with genome size (Bennett 1998). After controlling for life-cycle duration using multifactor ANOVA (or GLM, general linear model), the effect still held (figure 4c), which indicates that the relation between plant genome size and the risk of extinction is (at least partly) independent of the life cycle.

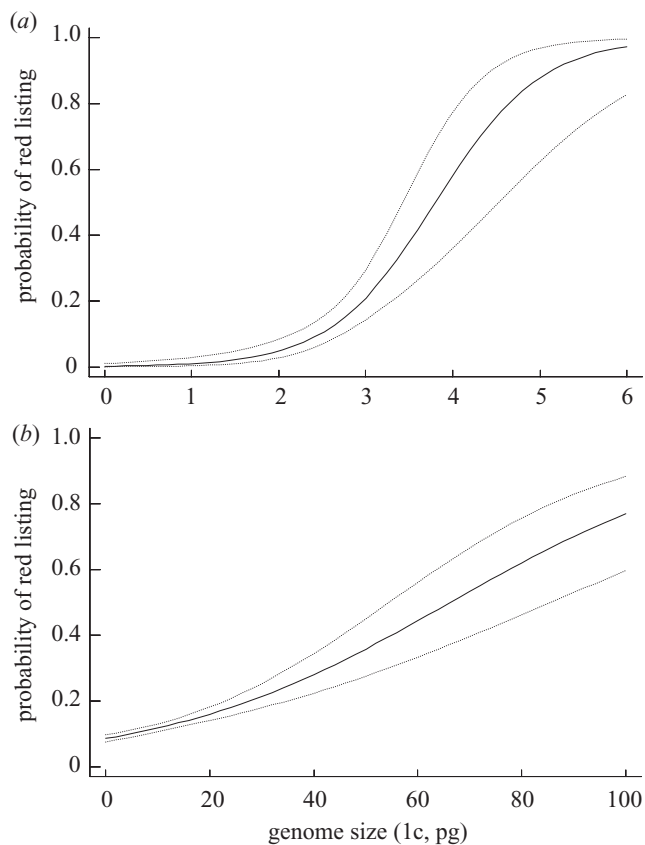


Figure 5. Probability of a species being included in the Red List, determined using logistic regression (with 95% confidence limits). (a) Reptiles. (b) Plants. (In both cases, for logistic regression  $p < 10^{-4}$ .)

#### (d) *Prediction of conservation status*

For conservation effort it may be convenient to use genome size (which can now be determined quite easily and accurately using flow cytometry) as an additional predictor of potential conservation status of a given species. Two examples are shown in figure 5. It can be seen that in reptiles and plants the probability of being included in the Red List, determined using logistic regression, increases from less than 10% to greater than 80% (with confidence intervals up to 20%) with increasing genome size. For birds, the confidence intervals are too large (up to 40%) for accurate prediction (although the logistic-regression model is statistically significant).

#### (e) *Fractions of assessed species*

It is important to estimate what fraction of the total number of species is covered by the current analysis in different taxonomic groups. Assuming the independence of fractions of species for which genome size is known and that were assessed for conservation status, we can take a product of both fractions as a rough estimate of the percentage of species covered by both types of assessment. For cold-blooded animals (fishes, amphibians and reptiles), the percentages of threatened species are about threefold higher in the genome size database than in the total species list. This suggests that the more easily accessed species are more likely to be assessed both for genome size and for conservation status, and hence these two fractions are not independent. Therefore, the

following estimates for cold-blooded animals covered by both types of analysis are related only to the percentages of the total covered species (threatened plus unthreatened), and the percentages of the covered threatened species can be approximately threefold higher. (It is the threatened species that are a minor part of the dataset and therefore determine the statistical significance of the analysis.) For warm-blooded vertebrates (mammals and birds), the percentages of threatened species in the total species list and in the genome size dataset were nearly equal. All known mammals and birds are assessed for conservation status and hence the percentage of those whose genome size is known can be taken as a sought estimate. This makes 6.6% of mammals and 1.6% of birds. In amphibians and reptiles, only 15% of species are assessed for conservation status, which makes percentages of species assessed both for conservation status and for genome size 1.2% and 0.5%, respectively. In fishes, 10% are assessed for conservation status and thereby only 0.4% of species are covered by both assessments. In angiosperm plants, 2–5% are assessed for conservation status, and correspondingly *ca.* 0.04% are covered by both assessments. (The plants and cold-blooded animals designated as ‘not threatened’ should more accurately be regarded as ‘not threatened or not yet assessed’, but the obtained conclusions remained valid because the heterogeneous nature of this group can only decrease the statistical significance of results.)

Thus, the percentages of covered species in reptiles and birds are the lowest among the cold-blooded and the warm-blooded tetrapods, respectively (0.5% in reptiles versus 1.2% in amphibians, and 1.6% in birds versus 6.6% in mammals). However, these smallest fractions (and an even smaller fraction in plants) still revealed a positive link between genome size and the risk of extinction. Therefore, we probably should not expect to find this link in amphibians (at the below-orders level) and mammals in the course of future assessment of genome size and conservation status. Neither is it likely that amphibians and mammals are less subject to the harmful state of the current biosphere as compared with reptiles and birds. In particular, there is a notion that it is amphibians that are now especially vulnerable because they can be very sensitive to pollution and dryness attributed to a global change of climate (Collins & Storer 2003; Rohr & Madison 2003).

## 4. CONCLUSION

Thus, in several phylogenetic groups of vertebrate animals (fishes, amphibians, mammals), the relationship between genome size and the risk of extinction is not revealed (although in fishes and amphibians it can exist at the above-order levels). Probably, the cost of the accumulation of non-coding DNA in their genomes is better balanced with such benefits as, for instance, the lower basal metabolic rate that can be adaptive in ecological niches with a low energy supply (Szarski 1983; Vinogradov 1995, 1997, 1998a; Gregory 2002; Waltari & Edwards 2002; Kozłowski *et al.* 2003; Olmo 2003; Vinogradov & Anatskaya 2004). Although the link between genome size and metabolic rate has not so far been proven in fishes and amphibians (e.g. Gregory 2003), this is probably

because of the strong dependence of the latter parameter on temperature and the adaptation of different species to different temperatures. For instance, it has been shown that the correlation between genome size and duration of embryonic cell cycles (another physiological parameter that is strongly dependent on temperature) in closely related frogs can reach its maximum possible value ( $r = 1.0$ ), but only at the compromise (trade-off) temperature (Vinogradov 1999). A change of temperature causes differences in the duration of the cell cycle changes in each species, in accordance with its geographical habitat range, and at certain temperatures the correlation with genome size becomes non-significant. On the other hand, the link between genome size and metabolic rate has been demonstrated in birds and reptiles (Vinogradov 1997; Gregory 2002; Olmo 2003), where genome enlargement is shown here to increase the risk of extinction. Therefore, it is possible that the accumulation of redundant DNA in their genomes can sometimes be beneficial in the short term but later become maladaptive at the species and lineage levels. It is interesting that birds and reptiles have the smallest genomes among tetrapods, which suggests that selection against the redundant DNA is generally stronger in them. Therefore, it is plausible that even a small amount of redundant DNA in their genomes might become maladaptive in the long term. They also show a separate regression line between genome size and GC content, differing from the amphibian–mammalian line, which indicates a distinctive pattern of evolutionary dynamics of the genome (Vinogradov 1998b).

It was suggested that accumulation of selfish DNA in the genome might become a trap: besides the fact that transposable elements form ‘landing pads’ for other such elements, the longer life cycles and the lower population sizes (because of the larger body sizes resulting from increased cell size) caused by genome enlargement might attenuate purifying selection against larger genomes, thereby simulating a neutralist effect of ‘permissive’ evolution at the organismal level, but threatening survival at the species and lineage levels (Vinogradov 2003). The data obtained here, and in the previous work on plants (Vinogradov 2003), provide factual support for the hotly debated concept of hierarchical selection (reviewed by Gould & Lloyd 1999; Gregory 2004).

The author thanks two anonymous reviewers for helpful comments. 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

Bennett, M. D. 1998 Plant genome values: how much do we know? *Proc. Natl Acad. Sci. USA* **95**, 2011–2016.

- Bennett, M. D. & Leitch, I. J. 2003 Plant DNA C-values database (release 2.0) <http://www.rbghkew.org.uk/cval/homepage.html>.
- Collins, J. P. & Storfer, A. 2003 Global amphibian declines: sorting the hypotheses. *Divers. Distrib.* **9**, 89–98.
- Doolittle, W. F. & Sapienza, C. 1980 Selfish genes, the phenotype paradigm and genome evolution. *Nature* **284**, 601–603.
- Gould, S. J. & Lloyd, E. A. 1999 Individuality and adaptation across levels of selection: how shall we name and generalize the unit of Darwinism? *Proc. Natl Acad. Sci. USA* **96**, 11 904–11 909.
- Gregory, T. R. 2001 Animal genome size database. <http://www.genomesize.com/> (accessed 2003).
- Gregory, T. R. 2002 A bird’s-eye view of the C-value enigma: genome size, cell size, and metabolic rate in the class Aves. *Evolution* **56**, 121–130.
- Gregory, T. R. 2003 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.* **79**, 329–339.
- Gregory, T. R. 2004 Macroevolution, hierarchy theory, and the C-value enigma. *Paleobiology* **30**, 179–202.
- Kozłowski, J., Konarzewski, M. & Gawelczyk, A. T. 2003 Cell size as a link between noncoding DNA and metabolic rate scaling. *Proc. Natl Acad. Sci. USA* **100**, 14 080–14 085.
- Olmo, E. 2003 Reptiles: a group of transition in the evolution of genome size and of the nucleotypic effect. *Cytogenet. Genome Res.* **101**, 166–171.
- Orgel, L. E. & Crick, F. H. 1980 Selfish DNA: the ultimate parasite. *Nature* **284**, 604–607.
- Rohr, J. R. & Madison, D. M. 2003 Dryness increases predation risk in efts: support for an amphibian decline hypothesis. *Oecologia* **135**, 657–664.
- Szarski, H. 1983 Cell size and the concept of wasteful and frugal evolutionary strategies. *J. Theor. Biol.* **105**, 201–209.
- Venter, J. C. (and 273 others) 2001 The sequence of the human genome. *Science* **291**, 1304–1351.
- 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.
- 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.
- Vinogradov, A. E. 1998a Buffering: a possible passive-homeostasis role for redundant DNA. *J. Theor. Biol.* **193**, 197–199.
- Vinogradov, A. E. 1998b Genome size and GC-percent in vertebrates as determined by flow cytometry: the triangular relationship. *Cytometry* **31**, 100–109.
- Vinogradov, A. E. 1999 Genome *in toto*. *Genome* **42**, 361–362.
- Vinogradov, A. E. 2003 Selfish DNA is maladaptive: evidence from the plant Red List. *Trends Genet.* **19**, 609–614.
- Vinogradov, A. E. & Anatskaya, O. V. 2004 Phenological resonance and quantum life-history. *J. Theor. Biol.* **228**, 417–420.
- Waltari, E. & Edwards, S. V. 2002 Evolutionary dynamics of intron size, genome size, and physiological correlates in archosaurs. *Am. Nat.* **160**, 539–552.