Variation across species in the size of the nuclear genome supports the junk-DNA explanation for the C-value paradox

MARK PAGEL,† AND RUFUS A. JOHNSTONE

1 Biological Anthropology, Harvard University, Cambridge, Massachusetts 02138, U.S.A.
2 Department of Plant Sciences, University of Oxford, Oxford OX1 3PS, U.K.

SUMMARY

The amount of DNA in the nuclear genome (the DNA C-value) of eukaryotes varies at least 80000-fold across species, and yet bears little or no relation to organismic complexity or to the number of protein-coding genes. This phenomenon is known as the C-value paradox. One explanation for the C-value paradox attributes the size of the nuclear genome to ‘junk’ (typically non-coding) genetic elements that accumulate until the costs to the organism of replicating excess DNA select against it. Across species, organisms that develop at a slower rate should tolerate more junk DNA. Alternatively, junk DNA may function as a nucleo-skeleton to maintain the volume of the nucleus at a size proportional to the volume of the cytoplasm in the cell. Across species, the DNA C-value is predicted to vary with the nuclear and cytoplasmic volumes of cells. Previous studies have not been able to distinguish between the skeletal-DNA and junk-DNA explanations for the C-value paradox. We report a study of DNA content in 24 salamander species which does. The size of the nuclear genome is correlated with developmental rate even after the effects of nuclear and cytoplasmic volume have been removed. However, genome size is not correlated with cytoplasmic volume after controlling for developmental rate. These results support the view that junk DNA accumulates in the nuclear genome until the costs of replicating it become too great, rather than that it functions as a nucleo-skeleton.

1. INTRODUCTION

The amount of DNA in the nuclear genome of an organism is called the DNA C-value to denote the fact that the haploid genome is fairly constant or ‘characteristic’ within a species. However, there is extraordinary variation in the DNA C-values across species. Eukaryote species have C-values which vary from around 0.009 pg in a yeast (Saccharomyces cerevisiae) to approximately 700 pg in Ameoba dubia (Cavalier-Smith 1985α), an approximately 80000-fold range (1 pg is approximately 10⁶ bases). Surprisingly, however, there is no correspondence between the amount of DNA in the nuclear genome of an organism and its organismic complexity or the number of protein-coding genes, a phenomenon known as the C-value paradox (see Cavalier-Smith 1978, 1985α; Dawkins 1982; Maynard Smith 1989; Li & Graur 1991). For example, genome size in mammals ranges from around 1.5 pg to 5.9 pg, salamanders vary between about 15 pg and 82 pg, and lungfish have up to 142 pg (Olmo 1988) of DNA in their nuclei.

Why amoeba might require so much more DNA than humans is not obvious. Similarly, Orgel & Crick (1980) exhort that ‘it seems totally implausible that the number of radically different genes needed in a salamander is twenty times that in a man’. Several authors suggest that the lack of a relation between genome size, number of protein coding genes, and organismic complexity can be understood by dividing the nuclear DNA into two components: that which is primarily genic or coding DNA, and that which has been called ‘junk’ (Ohno [1972] – including ‘selfish’ (Doolittle & Sapienza 1980; Orgel & Crick 1980) or ‘ignorant’ (Dover 1980) – DNA. The latter refers to genetic elements which typically are non-coding and whose deletion or extensive alteration has a negligible effect on the organism (Orgel et al. 1980). Evidence from renaturation kinetic studies suggests that non-coding (or non-genic) DNA typically comprises 70% or more of the total quantity of DNA of higher organisms (Cavalier-Smith 1985α). Humans, for example, may use only about 9% of their 3.5 pg genome to code for proteins, salamanders are estimated to use only about 1–5% of their DNA to code for proteins, whereas up to 70% of the yeast (S. cerevisiae) genome is genic DNA (Cavalier-Smith 1985α). So an explanation for variation across species in the amount of non-coding DNA can potentially explain much of the C-value paradox.

Forces that can replicate and spread junk genetic elements, such as unequal crossing over and transposition, will increase the size of the genome until the
Table 1. Time until hatching, genome size, cell volume, and nuclear volume in 24 Urodele amphibian species

(Cytoplasmic volume is found as the difference between cell and nuclear volume.)

<table>
<thead>
<tr>
<th>species name</th>
<th>time until hatching/d</th>
<th>genome size (diploid)/pg</th>
<th>cell volume/µm³</th>
<th>nuclear volume/µm³</th>
</tr>
</thead>
<tbody>
<tr>
<td>Desmognathus fuscus</td>
<td>47</td>
<td>30.2</td>
<td>765</td>
<td>122</td>
</tr>
<tr>
<td>Leurognathus marmoratus</td>
<td>83</td>
<td>33.1</td>
<td>417</td>
<td>77</td>
</tr>
<tr>
<td>Eurycea bistriata</td>
<td>28</td>
<td>41.6</td>
<td>1138</td>
<td>141</td>
</tr>
<tr>
<td>Batrachocephalus attenuatus</td>
<td>56</td>
<td>84.0</td>
<td>1233</td>
<td>228</td>
</tr>
<tr>
<td>Eutinina eschscholtzii</td>
<td>112</td>
<td>77.5</td>
<td>1323</td>
<td>281</td>
</tr>
<tr>
<td>Plethodon cinereus</td>
<td>56</td>
<td>45.2</td>
<td>1388</td>
<td>196</td>
</tr>
<tr>
<td>P. glutinosus</td>
<td>60</td>
<td>52.6</td>
<td>1791</td>
<td>326</td>
</tr>
<tr>
<td>Ambystoma maculatum</td>
<td>25</td>
<td>52.3</td>
<td>1192</td>
<td>247</td>
</tr>
<tr>
<td>A. maculatum</td>
<td>43</td>
<td>52.4</td>
<td>559</td>
<td>103</td>
</tr>
<tr>
<td>A. opacum</td>
<td>41</td>
<td>47.7</td>
<td>1611</td>
<td>212</td>
</tr>
<tr>
<td>A. talpoideum</td>
<td>35</td>
<td>62.2</td>
<td>538</td>
<td>90</td>
</tr>
<tr>
<td>A. texanum</td>
<td>71</td>
<td>48.3</td>
<td>1462</td>
<td>287</td>
</tr>
<tr>
<td>Rhynchochloris olympicus</td>
<td>250</td>
<td>124.0</td>
<td>3714</td>
<td>488</td>
</tr>
<tr>
<td>Notophthalmus viridescens</td>
<td>28</td>
<td>72.8</td>
<td>1730</td>
<td>183</td>
</tr>
<tr>
<td>Triturus cristatus</td>
<td>25</td>
<td>46.3</td>
<td>1145</td>
<td>134</td>
</tr>
<tr>
<td>T. marmoratus</td>
<td>28</td>
<td>49.8</td>
<td>1674</td>
<td>208</td>
</tr>
<tr>
<td>T. alpestris</td>
<td>21</td>
<td>48.0</td>
<td>1350</td>
<td>190</td>
</tr>
<tr>
<td>T. vulgaris</td>
<td>20</td>
<td>49.3</td>
<td>1056</td>
<td>127</td>
</tr>
<tr>
<td>Tylototriton verrucosus</td>
<td>20</td>
<td>49.0</td>
<td>524</td>
<td>97</td>
</tr>
<tr>
<td>Amphiuma means</td>
<td>140</td>
<td>149.9</td>
<td>3520</td>
<td>640</td>
</tr>
<tr>
<td>Necturus maculosus</td>
<td>57</td>
<td>165.1</td>
<td>3348</td>
<td>946</td>
</tr>
<tr>
<td>Hyperolius nubilosus</td>
<td>32</td>
<td>38.4</td>
<td>380</td>
<td>49</td>
</tr>
<tr>
<td>H. naevius</td>
<td>32</td>
<td>40.9</td>
<td>681</td>
<td>123</td>
</tr>
</tbody>
</table>

costs paid by the organism for replicating the junk DNA put the organism at a disadvantage. Thus the amount of junk DNA may be determined by a balance between intra-genomic mutation forces that increase genome size and selective forces acting on the organism to reduce genome size (Doolittle & Sapienza 1980; Orgel & Crick 1980). Other things equal, the weaker the selection against junk DNA, the larger the predicted genome size. One cost of excess DNA is that the time required to replicate the genome is increased (see, for example, Bennett 1972). Organisms that have been selected to develop at a slower pace may therefore ‘tolerate’ a greater amount of junk genetic elements, and thus a negative correlation across species between genome size and developmental rate is predicted.

An alternative explanation for the C-value paradox gives the non-genic DNA a functional role in maintaining the volume of the nucleus at a metabolically favourable size relative to the volume of the cytoplasm in the cell (Cavalier-Smith 1985d). Cavalier-Smith (1985d) suggests that larger cells require more surface area on the nuclear membrane to maintain transport of RNA from the nucleus to the cytoplasm at an acceptable rate. Non-genic DNA is assumed to perform a ‘skeletal’ function within the nucleus by, through its sheer bulk, determining nuclear volume, and thus the area of the nuclear membrane (Cavalier-Smith 1985d). The DNA C-value is thus predicted to vary across species positively with cytoplasmic volume. Further, because organisms with large cells tend to be slower growing whereas faster-growing organisms typically have smaller cells, Cavalier-Smith (1985a) also predicts a negative correlation across species between the size of the genome and developmental rate. Thus, from both the junk-DNA and the skeletal-DNA ideas, a negative correlation between developmental rate and the DNA C-value is expected. However, according to the skeletal-DNA idea, this relation occurs only secondarily as a result of the relations amongst developmental rate and nuclear and cytoplasmic volume.

Both the skeletal-DNA and junk-DNA explanations for the C-value paradox are supported by the cellular and life history data. In a variety of taxonomic groups, the size of the genome is positively correlated with cell-cycle length (the time required for a cell to divide, an inverse measure of rate) and negatively correlated with developmental rate (Bennett 1972; Oeldorf et al. 1978; Horner & MacGregor 1983; Sessions & Larson 1987). Further, the DNA C-value has been found in several studies to correlate positively with cytoplasmic and nuclear volume (Horner & MacGregor 1983; Olmo 1983; Cavalier-Smith 1985; Sessions & Larson 1987). However, only the junk DNA explanation predicts that there will be a relation between developmental rate and genome size independent of the cytoplasmic and nuclear volumes, a relation that has not been tested for. It is this difference between the two points of view that we examine in this study.

2. METHODS

We collected data from the literature on haploid genome size (pg), cytoplasmic volume and nuclear volume (Olmo & Morescalchi 1975; Olmo & Morescalchi 1978; Horner & MacGregor 1983; Olmo 1983; Sessions & Larson 1987) and, as an inverse measure of developmental rate, on the time until hatching (Oyama 1929a, b; Bishop 1943; Steward 1969; Horner & MacGregor 1983; Larson 1984; Duellman & Trueb 1986; Das 1987; Hon 1987; Sembitsch et al. 1988).
in 24 species of Urodele amphibians from seven families (Ambystomatidae, Amphiumidae, Cryptobranchidae, Hynobiidae, Plethodontidae, Proteidae, and Salamandridae). Urodeles are a particularly suitable group for studying variation in genome size because they exhibit a wider range of C-values than any other order or class of terrestrial vertebrates. The cytoplasmic and nuclear volumes are measured in erythrocytes which are nucleated in these species. Time until hatching is the number of days between the time that eggs are laid and the emergence of larvae. The cellular and life-history data are present in table 1.

With the exception of a study by Sessions & Larson (1987), previous work on C-value variation has used across-species correlational analyses. Such analyses ignore the phylogenetic relationships amongst species, and potentially bias results by treating species as independent data points in statistical analyses. To avoid these problems we applied a variant of Felsenstein’s (1985) method of pairwise contrasts (Pagel 1992) to a phylogeny of the Urodeles. This method adapts Felsenstein’s approach to a phylogeny for which the lengths of the branches of the phylogenetic tree are not known and may be used when the phylogeny is incompletely specified (see also Harvey & Pagel 1991). The method computes a set of comparisons for each variable between pairs of species that share an immediate common ancestor, between pairs of species and higher nodes that share an immediate common ancestor, and between pairs of the higher nodes of the phylogeny. The pairwise comparisons are defined by the phylogeny, each is used as an independent data point in statistical tests, and statistical trends can be interpreted as would an analysis across species (Felsenstein 1985; Pagel & Harvey 1989; Harvey and Pagel 1991). By using this approach, a significant relation between two variables is evidence that the comparative relationship has evolved independently in each of a number of phylogenetic sister-groups (Grafen 1989; Pagel & Harvey 1989; Harvey & Pagel 1991).

We assembled a phylogeny of the Urodeles from several sources. Duellman’s (1989) scheme provided relationships at the familial level and above. The phylogeny of the Plethodontidae was taken from Sessions & Larson (1987), and the salamandroid phylogeny was taken from Wake & Ozemt (1969). The ambystomatid phylogeny was derived from Kraus (1988). Remaining phylogenetic classifications followed taxonomic practice (Duellman & Trueb 1986). The phylogenetic tree derived from these classifications and the pairwise comparisons as defined by this tree are shown in figure 1.

3. RESULTS

We tested the comparative predictions by calculating the correlations amongst the pairwise contrasts on the four variables (table 2). Genome size correlates

![Figure 2. Plot of genome size comparisons (see text) against time until hatching comparisons. Each point on the plot represents a pairwise comparison calculated between two species, a species and a higher node, or between two higher nodes. The correlation is equal to 0.69, p < 0.01 (two-tailed).](http://rspb.royalsocietypublishing.org/Downloaded from http://rspb.royalsocietypublishing.org)
Table 2. Correlations amongst pairwise comparisons on genome size, cytoplasmic and nuclear volume, and time until hatching in Urodeles

(All correlations are significant at $p < 0.01$ or less (two-tailed $p$-values).)

<table>
<thead>
<tr>
<th></th>
<th>genome size</th>
<th>cytoplasmic volume</th>
<th>nuclear volume</th>
<th>time until hatching</th>
</tr>
</thead>
<tbody>
<tr>
<td>genome size</td>
<td>—</td>
<td>0.56</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>cytoplasmic volume</td>
<td>—</td>
<td>—</td>
<td>0.68</td>
<td>0.69</td>
</tr>
<tr>
<td>nuclear volume</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.53</td>
</tr>
</tbody>
</table>

Figure 3. Plot of residual genome size comparisons (see text) against residual time until hatching comparisons. Each point on the plot represents a pairwise comparison calculated between two species, a species and a higher node, or between two higher nodes, after statistically removing the effects of nuclear and cytoplasmic volume. The correlation is equal to 0.57, $p < 0.025$ (two-tailed).

significantly with time until hatching, and with cytoplasmic and nuclear volume. Figure 2 portrays the relation between the pairwise contrasts on genome size and those for time until hatching, showing the tendency for genome size to be larger in species that take longer to hatch.

The results in Table 2 support both explanations for the C-value paradox. However, only the junk-DNA explanation predicts that there will be a relation between genome size and developmental rate (time until hatching) that is independent of the cellular characteristics. We used partial correlation to test this prediction by controlling statistically for the associations of genome size and time until hatching with nuclear and cytoplasmic volume. That is, we calculated residual values for genome size and for time until hatching that are uncorrelated with either of the potentially confounding variables. The correlation between the time until hatching residuals and the genome size residuals is significant (partial $r = 0.57$, $p < 0.025$; Spearman rank correlation $r_s = 0.45$, $p < 0.033$; figure 3. All $p$-values reported are for two-tailed tests). This shows that there is a significant association between genome size and time until hatching that cannot be attributed either to cytoplasmic or to nuclear volume.

The skeletal-DNA ideas link genome size to cytoplasmic volume via the effect of genome size on nuclear volume. However, the observed correlation between genome size and cytoplasmic volume (table 2) may arise solely because each is associated with developmental rate. This seems to be the case: after controlling for time until hatching, genome size does not correlate significantly with cytoplasmic volume ($r = 0.29$, $p > 0.20$), although perhaps not surprisingly genome size remains correlated with nuclear volume ($r = 0.51$, $p < 0.05$; $r_s = 0.42$, $p = 0.053$). Further evidence against the idea that it is genome size which determines the relation between nuclear volume and cytoplasmic volume comes from examining the partial correlation of nuclear volume with cytoplasmic volume controlling for genome size. If genome size determines nuclear volume then the partial correlation should be non-significant. However, after controlling for genome size, nuclear volume and cytoplasmic volume remain highly correlated (partial $r = 0.77$, $p < 0.001$; $r_s = 0.66$, $p < 0.003$). Thus, the volume of the nucleus may be functionally related to the volume of the cytoplasm, but if so their relation is not determined by the amount of nuclear DNA.

4. DISCUSSION

Our results provide a test that can distinguish between two opposing explanations for the C-value paradox. According to proponents of the junk-DNA explanation (Ohno 1972; Orgel & Crick 1980; Doolittle & Sapienza 1980; Orgel et al. 1980), junk, parasitic or selfish DNA accumulates in the genomes of organisms until the costs of replicating it become too great. Organisms that develop at a slower pace are predicted to be able to tolerate more junk DNA. The skeletal-DNA explanation, however, rests on the belief that organisms make use of the junk DNA to maintain the volume of the nucleus at a metabolically favourable size relative to the volume of cytoplasm in the cell. Species in our study with longer times until hatching (slower developmental rate) have larger genomes, independent of their cytoplasmic and nuclear volumes. However, cytoplasmic volume does not correlate with genome size after controlling for the effects of developmental rate. These results suggest that genome size is associated directly with time until hatching (developmental rate), and not, as the skeletal-DNA idea would suggest, with cytoplasmic volume. Further, on the assumption that most of the variation in genome size across the 24 species of Urodeles in our study results from variation in junk DNA (see evidence for this assumption presented in the Introduction), our results provide direct support for the view that the nuclear genomes of eukaryotes accumulate junk DNA until the costs to the organism of replicating it become too great. The idea that there is a functional relation between the surface area of the nuclear membrane and

the volume of the cytoplasm (Cavalier-Smith 1985b) seems reasonable and is supported by the results of our study. What is clear, however, is that this relation is independent of the size of the genome, because nuclear volume and cytoplasmic volume are highly correlated even after removing the correlated effects of genome size.

The proportion of non-coding against coding DNA varies greatly across organisms, and our results suggest that the variation depends principally on an organism’s tolerance for junk DNA. In this light it is interesting that the bacteria, for which fast replication times may be highly advantageous, are thought to have little if any non-coding DNA (Cavalier-Smith 1985a). Similarly, the mitochondrion is thought to be free of non-coding DNA (Li & Graur 1991). Perhaps this is due to intense competition among these organelles to replicate quickly at meiosis to ensure that they will be represented in the daughter cells. However, even though DNA can merely accumulate as ‘junk’ when the costs to the organism are low, it may benefit the organism in other ways. For example, Bennett (1972) reports that the size of the nuclear genome exerts a profound influence on the minimum development time of a species: Bennett found cell-cycle times of up to 300 h in some plants. One possibility, then, is that when natural selection favours slower development, individuals with larger amounts of junk DNA may be favoured, if the metabolic costs of having extra DNA are not onerous. That is, organisms may use the accumulation of junk DNA as a mechanism for slowing rates of development via increased cell-cycle lengths. Conversely, when selection favours faster development, individuals with mechanisms for deleting junk DNA may be at an advantage. Such mechanisms are predicted to be found more commonly in those lineages that have evolved to develop faster than their ancestral species. Large regions of non-transcribed DNA may benefit longer-lived organisms by acting as inert repositories for retroviruses (Bremnerman 1987). Longer-lived species might be expected to suffer the onslaughters of a greater number of retroviruses in their lifetimes. If retroviruses reverse-transcribe themselves into randomly chosen portions of the genome, then having large regions of non-transcribed DNA could decrease the likelihood that the retrovirus ever manifests itself.

Large regions of junk DNA may also protect organisms against the potentially destructive effects of genetic recombination. Recombination, for all of its apparent advantages, can also break apart useful genes. Having long introns composed of junk DNA might increase the chances of a recombination event falling in a non-transcribed region of a gene, rather than in a coding region. Mitochondrial DNA is again interesting in this light: it does not recombine, and, as mentioned above, it is thought to lack non-coding DNA. In contrast, domesticated animals have undergone strong artificial selection and display very high levels of recombination (Burt & Bell 1987). Support for our view that junk DNA may protect against recombination breaking apart useful genes would be found if domesticated animals also have a large amount of DNA relative to their non-domesticated ancestors or to other non-domesticated animals with similar life-history characteristics. More to the point, we would predict that in, domesticated animals, exons are separated by unusually long introns.

A simplistic view of the C-value paradox would hold that we should expect a single allometric relation between genome size and developmental rate, right across the animal and plant kingdom. However, there are many reasons that such a view is incorrect. Just as a mammal and a bird of similar body size have different amounts of their total body masses given over to bone mass, the DNA C-value of a species may vary amongst organisms with similar life histories but of different phylogenetic origins or different lifestyles. For example, the number of independent sites of replication along the genome may be shared by many species within a specified group, but vary across groups. Variation in the number of such sites may influence the amount of time required to replicate the genome, and thus influence the temporal cost that a particular genome size imposes. Thus, for example, the fact that humans with a much slower developmental rate than salamanders nevertheless have much less DNA should not be taken as evidence against the junk DNA explanation for the C-value paradox.

We thank Peter Holland and Ruth Mace for comments on an earlier draft of this manuscript. M.P. was supported by a grant from the Commission of European Communities, and by SERC grant No. GR/F 98727.

REFERENCES


Received 25 February 1992; accepted 8 April 1992