

# Eukaryotic non-coding DNA is functional: evidence from the differential scaling of cryptomonad genomes

Margaret J. Beaton\* and Thomas Cavalier-Smith†

Canadian Institute for Advanced Research Evolutionary Biology Programme, Department of Botany, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z4

Genic DNA functions are commonplace: coding for proteins and specifying non-messenger RNA structure. Yet most DNA in the biosphere is non-genic, existing in nuclei as non-coding or secondary DNA. Why so much secondary DNA exists and why its amount per genome varies over orders of magnitude (correlating positively with cell volume) are central biological problems. A novel perspective on secondary DNA function comes from natural eukaryote–eukaryote chimaeras (cryptomonads and chlorarachneans) where two phylogenetically distinct nuclei have coevolved within one cell for hundreds of millions of years. By comparing cryptomonad species differing 13-fold in cell volume, we show that nuclear and nucleomorph genome sizes obey fundamentally different scaling laws. Following a more than 125-fold reduction in DNA content, nucleomorph genomes exhibit little variation in size. Furthermore, the present lack of significant amounts of nucleomorph secondary DNA confirms that selection can readily eliminate functionless nuclear DNA, refuting ‘selfish’ and ‘junk’ theories of secondary DNA. Cryptomonad nuclear DNA content varied 12-fold: as in other eukaryotes, larger cells have extra DNA, which is almost certainly secondary DNA positively selected for a volume-related function. The skeletal DNA theory explains why nuclear genome size increases with cell volume and, using new evidence on nucleomorph gene functions, why nucleomorph genomes do not.

**Keywords:** C-value paradox; DNA content; skeletal DNA; cryptomonads; nucleomorphs; nuclei

## 1. INTRODUCTION

Despite numerous theories and attempts to evaluate their merits, there is no general agreement about why most eukaryotic cells contain orders of magnitude more DNA than required to code for their proteins (Cavalier-Smith 1985*a*). Eukaryote genome sizes vary by about 300 000-fold (minimum = 0.0023 pg in the microsporidian *Encephalitozoon intestinalis* (Vivares 1999), maximum = 700 pg in *Amoeba dubia* (Friz 1968) (1 pg = 980 Mb)), with no overall correlation with organismal complexity or estimated gene numbers (Cavalier-Smith 1985*b*). In fact, the bulk of nuclear DNA (over 99% for many taxa) is present as non-coding or secondary DNA. In contrast to its usual lack of correspondence with organismal complexity, strong positive correlations have been described between genome size and both nuclear and cell volumes in unicells and multicellular organisms alike (Bennett 1972; Shuter *et al.* 1983; Cavalier-Smith 1985*c*).

Two radically different classes of explanation for the persistence and variability in amount of secondary DNA have been developed. One suggests that secondary DNA has no function and accumulates purely through mutation

pressure (Ohno 1972; Doolittle & Sapienza 1980; Orgel & Crick 1980; Maynard Smith & Szathmari 1997). Selection against secondary DNA is assumed to be simply too weak to eliminate it, and systematically less effective in larger cells than in smaller ones. As a result, there is a trend towards an overall increase in genome size. Secondary DNA can be regarded as purely neutral ‘junk’, maintained in the genome simply by its physical linkage to genic DNA (Ohno 1972). Another mutation pressure theory referred to the majority of non-coding DNA as genetic parasites or ‘selfish DNA’ and offered a mechanism (e.g. duplicative transposition) for its active accumulation (Doolittle & Sapienza 1980; Orgel & Crick 1980). In general, mutation pressure theories offer no clues as to the basis for the observed relationships between genome size and cellular parameters, although the presence of these correlations is insufficient by itself to preclude their validity.

In marked contrast to the purely mutational explanations, functional theories argue that secondary DNA has a sequence-independent function by virtue of its sheer bulk. Most functional theories assert that cell size is adaptively important (for which there is much evidence; Cavalier-Smith 1985*c*) and that the genome-size–cell-volume relationship is the key to explaining the continued presence of non-coding DNA. The most specific of the functional theories is the skeletal DNA hypothesis, which states that the mass of DNA (in conjunction with its

\*Author for correspondence and present address: Biology Department, Mount Allison University, 63B York Street, Sackville, NB, Canada E4L 1G7 (mbeaton@mta.ca).

†Present address: Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

Table 1. *Cryptomonad cell volumes and nuclear genome sizes*

(Strain code: CCAP, Culture Collection of Algae and Protozoa; CCMP, Culture Collection for Marine Phytoplankton; NEPCC, Northeast Pacific Culture Collection; UTEX, Culture Collection of Algae at the University of Texas at Austin. Dashes indicate data not determined.)

species	strain	cell volume ( $\mu\text{m}^3$ )	G <sub>2</sub> :G <sub>1</sub> ratio	DNA content (pg)
<i>Campylomonas reflexa</i>	CCMP 1177	1137.7	—	8.80
<i>Chroomonas pauciplastida</i>	CCMP 268	145.0	2.0	1.46
<i>Chroomonas placoidea</i>	CCAP 978/8	163.7	1.8	1.74
<i>Chroomonas diplococcus</i>	UTEX 2422	212.9	2.1	4.24
<i>Cryptomonas ozolini</i>	UTEX 2194	401.7	1.9	2.67
<i>Cryptomonas irregularis</i>	CCAP 979/7	456.3	1.8	2.76
<i>Cryptomonas ovata</i>	CCAP 979/61	1072.8	2.0	7.35
<i>Cryptomonas czosnowski</i>	CCAP 979/67	1151.4	—	5.80
<i>Guillardia theta</i>	CCMP 327	141.9	1.9	1.80
<i>Hanusia phi</i>	CCMP 325	205.0	—	2.32
<i>Hemiselmis virescens</i>	UTEX 1011	97.5	1.8	0.88
<i>Plagioselmis prolunga</i>	CCMP 644	88.8	—	0.72
<i>Rhodomonas (Pyrenomonas) salina</i>	NEPCC 076	395.7	1.8	3.60
<i>Rhodomonas (Pyrenomonas) salina</i>	CCMP 1319	294.6	1.9	3.72
<i>Rhodomonas</i> sp	CCMP 768	476.6	—	3.31
<i>Stoeratula major</i>	CCMP 320	922.2	1.9	4.74
unidentified	CCMP 1167	1174.4	2.0	7.57

folding pattern) directly determines the nuclear volume. In order to maintain a balance between the rates of nuclear RNA and cytoplasmic protein synthesis during growth, it is argued that cells of differing volumes must have a constant cytonuclear ratio, so that nuclear volume must coevolve with cell volume, which is itself roughly optimized by selection (Cavalier-Smith 1978, 1985c, 1991). This functional theory applies to both unicells and multicells, but its application to the latter is more complex; for simplicity the extra complications in multicells (see Cavalier-Smith 1985c, 1991) and protists with multiple nuclei or multiple fission (see Cavalier-Smith 1980, 1985c) are not considered here.

Deciding between these two radically different theories is doubly difficult because of the impracticality of manipulating genome sizes sufficiently for experimentally analysing the basis for the observed evolutionary correlations, and our imperfect understanding of the genetic determinants of cell and nuclear volumes. Here we present new data and arguments from two natural evolutionary experiments that offer a novel perspective on the correlation between cell size and nuclear genome size. These experiments were the independent origin several hundred million years (Myr) ago of two types of chimaeric binucleate algal cells by the permanent merger of an algal symbiont with a flagellate host (Cavalier-Smith 1995; Gilson *et al.* 1997). The merger of a green alga (Van de Peer *et al.* 1996; Cavalier-Smith *et al.* 1996; Ishida *et al.* 1999) with a protozoan flagellate host (Cavalier-Smith & Chao 1997) produced one such chimaera that diversified to form the chlorarachnean algae (Ishida *et al.* 1999). The other resulted from the endosymbiosis of a red alga (Douglas *et al.* 1991; Maier *et al.* 1991; Cavalier-Smith *et al.* 1996) in an unknown biflagellate host, establishing cryptomonad ancestors.

In both groups of algal chimaeras, the symbiont's nucleus (referred to as the nucleomorph), although no longer independent, has been retained in the cell.

Possessing a double envelope with characteristic nuclear pores (Greenwood 1974; Hibberd & Norris 1984), three minute linear chromosomes with telomeres, and densely packed genes (Gilson & McFadden 1996; Gilson *et al.* 1997; Zauner *et al.* 1999), nucleomorphs are undoubtedly genuine nuclei which have undergone the greatest genomic reduction in all eukaryotic history (Gilson *et al.* 1997; McFadden *et al.* 1997)—except for some total losses of algal symbiont nuclei during secondary symbiogenesis (Cavalier-Smith 1995). As a result of this reduction, chlorarachnean and cryptomonad nucleomorphs contain only 380–450 kb (Gilson & McFadden 1999) and 450–710 kb of DNA (Rensing *et al.* 1994), respectively. The intracellular coevolution of independently derived nucleus and nucleomorph for about 530 Myr (Cavalier-Smith 1999) allows us to use nucleomorphs as internal controls for understanding the selective forces acting on the main cell nucleus as a function of cell size.

In the present study, we show that the DNA content of the cryptomonad cell nucleus scales with cell size as in other eukaryotes, but the nucleomorph genome does not. We argue that these radically different scaling laws for the two eukaryote genomes coevolving in the same cell demonstrate very strong selection against secondary DNA in nucleomorphs but for it in nuclei. This appears to refute mutation pressure explanations and points to a positively selected function for nuclear secondary DNA. We explain how the skeletal DNA hypothesis can explain the divergent scaling patterns.

## 2. METHODS

We investigated the relationships between cell volume and nuclear and nucleomorph DNA contents in diverse cryptomonad species; cell volume and nuclear DNA content estimates are from the present study and nucleomorph size estimates were taken from Rensing *et al.* (1994). Seventeen strains (from culture collections shown in table 1) were grown at 20 °C under a

16 L:8 D regime. Dimensions (length and width) for each strain were estimated for 20–100 cells treated with Lugol's iodine fixative, and the measurements were incorporated into cell volume calculations using the formula for an oblate spheroid,  $\text{vol} = \pi/6 \times l \times w^2$ , where  $l$  and  $w$  are the mean length and width, respectively.

Cellular DNA contents were determined by flow cytometry for the same strains. Using cultures at densities of approximately  $10^6$  cells  $\text{ml}^{-1}$ , 1–3 ml of cells were washed one to two times in culture media, fixed in ice-cold 70% methanol in phosphate buffered saline (PBS) and held at 4 °C. Prior to measurement, fixed cells were washed twice in PBS before adding RNAase ( $100 \mu\text{g ml}^{-1}$  final concentration) and propidium iodide ( $50 \mu\text{g ml}^{-1}$  final concentration). Fluorescence was measured for 10 000 events per strain on a Becton Dickson Facscan. To confirm that autofluorescence did not interfere with the staining, unstained cells were routinely monitored. Reconstituted chicken red blood cells (2.33 pg DNA per cell) were used as a DNA standard, and as an added check of the measurement, the dinoflagellate *Amphidinium carterae* was also included. Our experience with cryptomonad DNA (especially *Guillardia theta*) separated by CsCl density centrifugation suggests that chloroplast, mitochondrial and nucleomorph DNA are together about 10% of total cellular DNA. As this is comparable to the estimated measurement error, we used total cell DNA values as estimates of nuclear DNA content.

### 3. RESULTS

#### (a) Variation in DNA contents

Two peaks (one major and one minor) were visible in the fluorescence histograms for most of the algal strains, identifying cells in the  $G_1$  and  $G_2$  stages of the cell cycle, respectively. The  $G_2$ : $G_1$  ratios were close to 2.0 (table 1), indicating a doubling of DNA contents, although the average of the ratios (1.91) was slightly less than expected (2.0), probably owing to an overlap of the upper and lower tails of the two distributions. In the present study, cells of the dinoflagellate, *A. carterae*, contained 6.4 pg of DNA, in good agreement with a previous estimate of 6.5 pg (Boucher *et al.* 1991).

We found that cryptomonad nuclear DNA contents vary over 12-fold from 0.72 to 8.8 pg per cell (table 1). The largest cells have 2.5 times as much DNA as a human sperm. Few measurements of DNA contents for cryptomonads have been reported that can be used for comparative purposes. Using flow cytometry, Velduis *et al.* (1997) found a 12-fold range in DNA contents among five cryptomonad strains (reported as chicken red blood cell units per cell). The converted estimate for one strain common to our study, *Storeatula major*, was 5.3 pg, about 10% higher than we found. Furthermore, Boucher *et al.* (1991) reported flow cytometric results for *Cryptomonas maculata* of 1.5 pg, which falls within our range. However, in contrast to flow cytometry studies, Hansmann & Eschbach (1990) reported a nuclear DNA content for *Rhodomonas (Pyrenomonas) salina* of 1.1 pg using a diphenylamine method following biochemical DNA extraction. This estimate is several-fold less than ours. Clearly, further study is required to confirm the absolute quantities of DNA per cell, but the relative amounts for strains should be valid. Whether the biochemical measurement for *R. salina* (Hansmann & Eschbach 1990) slightly

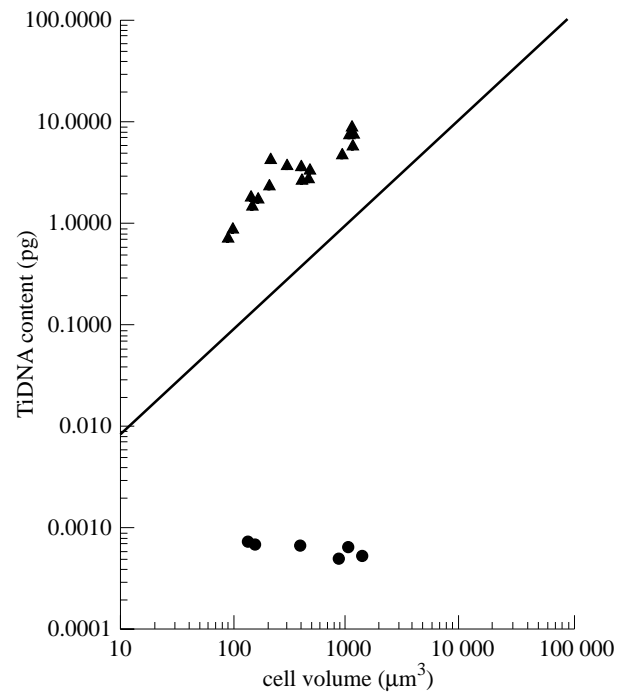


Figure 1. Relationship between cell volume and nuclear (triangles) and nucleomorph (circles) DNA content. Nucleomorph genome sizes estimated from Rensing *et al.* (1994) (490 kb, 655 kb, 638 kb, 515 kb, 710 kb, and 660 kb for *Chroomonas paramecium*, *Chroomonas* sp., *Cryptomonas ovata*, *Guillardia theta*, *Rhodomonas mariana*, and *Rhodomonas salina*, respectively); other data from table 1. The line represents a regression established earlier (Shuter *et al.* 1983) between cell volume and nuclear DNA contents in other unicellular eukaryotes (mainly dinoflagellates and diatoms). DNA conversion, 1 pg = 980 Mb.

underestimates the DNA content or flow cytometry measurements are slightly inflated (or both) is irrelevant to the central evolutionary arguments of the present paper.

#### (b) Different scaling laws for nuclei and nucleomorphs

As expected, DNA content correlated positively with cell volume ( $r = 0.9$ ,  $p < 0.001$ ), with a slope of 0.74 (see figure 1). This agrees reasonably with an earlier analysis for a wide taxonomic range of unicells (slopes for eukaryotes and prokaryotes were 1.0 and 0.28, respectively; Shuter *et al.* 1993), but we are unsure why the data systematically lie above the regression for other protists. This displacement implies that cryptomonads have more nuclear DNA per unit cell volume than previously studied protists, although the difference may be exaggerated by cell shrinkage during fixation or the imperfect resolution of  $G_1$  and  $G_2$  cell distributions. Similar discrepancies have been noted for a variety of data sets where DNA content and cell volume estimates have been obtained by different methods (as illustrated in Shuter *et al.* (1983) and Boucher *et al.* (1991)). What matters for the present discussion, however, is that cryptomonad nuclei show a marked increase in DNA content with cell size, as in all eukaryote groups previously studied. Furthermore, as gene numbers are expected to be approximately constant within the group, differences in nuclear content can be attributed to

secondary DNA. But in great contrast, nucleomorph genome size is essentially independent of cell volume (slope close to zero) (figure 1). Therefore, nucleomorph genomes follow a very different scaling law from their nuclear counterparts, demonstrating profoundly different evolutionary forces acting on the genome size of two phylogenetically distinct nuclei within a single cell.

#### 4. DISCUSSION

A comprehensive explanation for nuclear secondary DNA must explain why it increases with cell size and, equally importantly, why the corresponding nucleomorph DNA does not. The amount of nuclear non-coding DNA cannot be definitively determined from our results, as neither the ploidy level nor the exact number of genes is known. However, the strains studied may be mostly haploid, as sex with alternation of haploid and diploid generations is known only for one of the species not included here (Hill & Wetherbee 1986). If we assume that their nuclei have 15 000 protein genes with mean coding length 1 kb, then the coding part of their genome would be 15 Mb. As their cell DNA contents vary from 705 to 8620 Mb (table 1), 97.8–99.8% is likely to be non-coding.

The precise degree of genomic reduction by cryptomonad nucleomorphs compared to their red algal ancestors is also difficult to evaluate. The smallest nuclear genome reported for a single-celled red alga (*Cyanidioschyzon*) is about 0.01 pg (Maleszka 1993), but this organism is restricted to acidic hot springs in which potential flagellate hosts could not grow. Possibly a better indicator of the likely ancestral size is the unicellular mesophilic red alga, *Porphyridium aeruginosum*, with a genome of 0.13 pg (Barnes *et al.* 1982), making a reasonable minimal estimate for the genome size of the ancestral red algal endosymbiont about 0.1 pg (*ca.* 98 Mb). If we assume that the ancestral red algal genome had 10 000 genes (less than cryptomonads due to the absence of flagellar apparatus and extra membranes) of average length 1 kb, then 10 Mb of the 98 Mb (*i.e.* slightly over 10%) ancestral genome had a protein-coding function.

##### (a) *Drastic genome reduction in nucleomorphs*

Thus, cryptomonad nucleomorphs must have been miniaturized by at least 125-fold. For *Guillardia theta*, only about 9% of nucleomorph DNA is non-coding: introns are virtually absent and genes are separated by only about 75 bases (Zauner *et al.* 1999). Furthermore, while the protein-coding part of this genome has been reduced 21-fold compared to the ancestral red algal genome, secondary DNA has been even more effectively discarded (1950-fold; see table 2). The successful elimination of secondary DNA from nucleomorphs seems to prove that most of it could have been eliminated from the nucleus also, if it really had no function.

Although it is widely assumed that mutational explanations, specifically the selfish DNA hypothesis, can explain the existence of so much non-coding DNA (Maynard Smith & Szathmary 1997), comparative evidence strongly indicates that only a small fraction of secondary DNA in nuclei (*e.g.* most transposons) is simply genetically parasitic. A significant fraction of non-coding DNA, notably introns (Cavalier-Smith 1978, 1985*d*) and dis-

persed repetitive sequences (*e.g.* Alu repeats, L1 elements), is likely to be 'selfish' in origin (Cavalier-Smith 1985*e*), but is today predominantly not actively spreading genetic parasites. As ex-selfish DNA, their maintenance and persistence in the genome must be explained in the same way as other non-coding DNA. Most secondary DNA must be either neutral junk or functional.

Furthermore, the miniaturized nucleomorph genome compared with nuclei, and the absence of any evidence for larger nucleomorphs in bigger cells, show that mutation pressure does not inexorably increase the size of eukaryotic genomes. This appears to refute the mutation pressure theories, which would expect non-coding DNA to increase equally in both nuclei and nucleomorphs. Effective pressure for genomic reduction is also emphasized by the very short, 19–20 nucleotide splicing introns present in chlorarachnean nucleomorph chromosomes (Gilson & McFadden 1996), showing that selection can readily shorten but seldom eliminate introns. Moreover, the minuteness of nucleomorph genomes cannot be explained by intracellular competition for rapid replication (as has been invoked for mitochondria and chloroplasts) as there is typically only a single copy per cell. Thus, nuclear secondary DNA must be positively selected for rather than selectively neutral junk.

What exactly is the cost of extra non-coding DNA to a cell? How mechanistically does it lower net reproductive rates? We suggest that it is the cost in energy and scarce nutrients, such as phosphate and fixed nitrogen, needed to synthesize extra DNA and histones that would lower cell reproductive rates compared with a competitor with less DNA. If the selective disadvantage is sufficient to remove quite small amounts of non-coding DNA from the nucleomorph, it should be even more effective in removing the much larger, and thus more costly, amounts from the nucleus if that were non-functional.

The presence of a special genome-reducing mechanism in the nucleomorph seems unlikely. The mechanism of reduction of all genomes is simply deletion mutations followed by selection for smaller variants, and there is no reason to think that deletions occur in a higher frequency in the nucleomorph than in the nucleus. Because the nucleus has a much higher fraction of non-coding DNA than the nucleomorph, the fraction of its deletions that are viable (*i.e.* not eliminating essential genes) should be higher. As the residual amount of secondary DNA in the nucleomorph is so small, viable deletions must be shorter, on average, for continued genome reduction, and therefore yield less cost saving than those in the nucleus. Thus in the absence of a function, secondary DNA should be even more efficiently eliminated in the nucleus than the nucleomorph.

##### (b) *Large nuclear genomes: mutational and selective causes*

We argue that the skeletal DNA hypothesis (Cavalier-Smith 1985*c*, 1991) simply explains the opposing evolutionary forces acting on the two genomes. This theory rests on the fact that cells typically exhibit balanced growth, such that all components double between divisions, and evidence from cell-cycle experiments supports this tenet. Therefore, for a cell to evolve a larger volume there must be a corresponding increase in ribosomes and

Table 2. Estimated reduction of coding and non-coding DNA of the nucleomorph genome of *Guillardia theta* in relation to genome size estimates

(Contributions of coding and non-coding DNA to the nucleomorph genome are taken from Zauner *et al.* (1999) and remaining estimates are described in the text.)

	<i>Guillardia theta</i>			reduction of <i>Nm</i>
	nucleus	nucleomorph	ancestral red alga	
genome size (Mb)	1764	0.515	98	195-fold
protein-coding genes (Mb)	15	0.47	10	21-fold
end-product (Mb)	14.25	0.023	9.5	413-fold
housekeeping (Mb)	0.75	0.45	0.5	10%
secondary DNA (Mb)	1749	0.045	88	1956-fold

in ribosomal RNA (rRNA) and messenger RNA (mRNA). Consequently, the nucleus must accommodate more RNA polymerases, RNA-processing enzymes and nascent RNA, and so be correspondingly larger. Thus larger cycling cells must necessarily have larger nuclei. The skeletal DNA theory adds two key ideas to this established principle of cell biology. First, the mechanistic thesis that nuclear volume is determined by the total mass of DNA that it contains and the manner in which the DNA is folded, for which there is reasonable experimental evidence (Cavalier-Smith 1991). For example, in frog eggs, DNA acts as the sequence-independent nucleating agent for nuclear assembly (Forbes *et al.* 1983). Second, the evolutionary thesis that the simplest and most usual way to increase or decrease nuclear volume is by changing genome size rather than by altering the DNA folding pattern or its attachment to the nuclear envelope, both of which appear relatively constant among related taxa. If both points are correct, then an evolutionary correlation between genome size and cell size necessarily follows. Cell volume itself is assumed to be controlled genically, not by DNA content (Cavalier-Smith 1985c). Both genic and secondary DNA will function as a nuclear skeleton, but as the copy number of protein-coding genes generally needs to be no greater in larger cells, the fraction of secondary (i.e. purely skeletal) DNA rises in larger cells. Furthermore, the theory expects nuclear genome size to scale with a slope marginally less than one, if larger cells are not to have longer cell cycles (actually they do somewhat (Cavalier-Smith 1985c, 1985f), which slightly reduces the predicted scaling factor).

As stressed earlier (Cavalier-Smith 1978), the larger amounts of skeletal DNA in bigger nuclei provide a secure and extensive habitat for the multiplication of selfish genetic elements, which spread easily in sexual eukaryotes (Maynard Smith & Szathmary 1997). However, the cryptomonad and chlorarachnean binucleate chimaeras compellingly argue that the abundance of selfish elements is a consequence of large genomes, not their decisive determinant (Cavalier-Smith 1978, 1985e). There is no contradiction in recognizing an important role for mutational and selfish processes in many qualitative aspects of

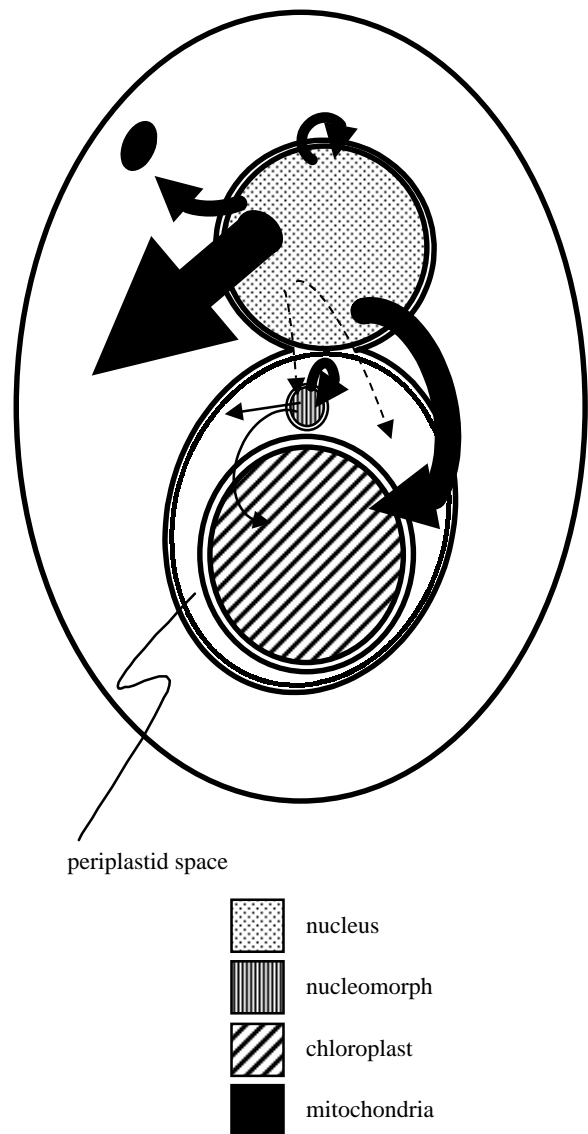


Figure 2. Relative contributions of nuclear and nucleomorph genes to cellular activity. The width of the arrow indicates the relative proportion of proteins transported from the nucleus and nucleomorph to the four cell compartments. For example, RNA from the nucleus is transported to, and active in, the cytoplasm. Broken arrows represent possible but unconfirmed protein transport routes.

genome evolution (Cavalier-Smith 1993), yet accepting a decisive role for selection for optimal nuclear size in governing the overall scale of the genome within which these molecular biases operate (Cavalier-Smith 1978). The junk DNA theory's assumption that duplications and insertions predominate over deletions has also been strongly contradicted by recent evidence for the reverse in *Drosophila*, where mutation pressure favours smaller rather than larger genomes (Petrov *et al.* 1998).

(c) **Why is nucleomorph genome size independent of cell size?**

The answer, we suggest, lies in the very different functional spectrum of nucleomorph genes compared with the nucleus (figure 2). The scaling for nucleomorph genomes expected by the skeletal DNA hypothesis will depend on how the demand for their gene products quantitatively

increases with cell size. In both cryptomonad and chlorarachnean nucleomorphs most identified genes are house-keeping genes, e.g. those involved in DNA replication, transcription, translation, and RNA and protein processing within the nucleomorph and its surrounding periplastid space (the relict cytosol of the former symbiotic algae: Cavalier-Smith 1995; Gilson *et al.* 1997). Missing from these genomes are genes encoding proteins for primary and secondary metabolism or cytosolic structures. Therefore, the proportion of the nucleomorph genome encoding end-products required in higher amounts with larger cell size is very low (figure 2). So far, only a handful of such end-product genes have been identified. They include a very few plastid proteins (protease clpP in the chlorarachnean (Gilson & McFadden 1996), and FtsZ and rubredoxin in the cryptomonad nucleomorph (Zauner *et al.* 1999)) and tubulin genes in cryptomonads (Zauner *et al.* 1999). With most of the *G. theta* nucleomorph genome sequenced (S. E. Douglas, U.-G. Maier, T. Cavalier-Smith and M. J. Beaton, unpublished data), we estimate that end-product genes are less than 5%, while in nuclei they probably exceed 95%, of total gene numbers.

Assuming balanced growth, doubling the cell volume without increasing cell-cycle length (i.e. scale with a slope close to zero), requires a 99% increase in the nuclear volume (i.e. scale with a slope somewhat less than one, depending on how much cryptomonad cell cycles increase in length in larger cells). However, given the radically different nature of its gene products, only a marginal volume increase would be required of the nucleomorph (thus scaling with a slope scarcely above zero). Thus the strongly divergent scaling of nuclear and nucleomorph genomes with cell size is predicted by the skeletal DNA theory of the C-value paradox, but not by mutation pressure theories (whether selfish or junk DNA), which cannot account for the differing scaling laws. This provides the strongest evidence yet against the neutral and selfish DNA theories and in favour of a positive, but sequence independent, function for secondary nuclear DNA.

The present discussion has assumed that the DNA packing ratio (i.e. the DNA mass per unit nuclear volume) is approximately constant for both eukaryotic genomes. This assumption needs testing by direct volumetric measurements for cryptomonads with different cell sizes. Furthermore, a survey of packing ratios for a variety of protists would offer insights into its conservation across major phylogenetic divergences.

## 5. CONCLUSION

The skeletal DNA theory may not be the final answer to the question why nuclei possess so much secondary DNA. But any alternative explanation for nuclear genome size evolution will be acceptable only if it can quantitatively explain the scaling laws discussed here at least as well as the skeletal hypothesis. Certainly, these alternatives must reasonably account for very effective divergent selection for small nucleomorph genomes and for far larger nuclear genomes in the same cell.

We have argued that the majority of the secondary DNA present in the nucleus of the ancestral red alga has

been successfully eliminated from cryptomonad nucleomorphs. But we do not claim that selection has eradicated all their non-coding DNA, because a consequence of ever-shrinking non-coding segments is the increasing probability that random deletions will more often remove segments of coding DNA, so fewer will be viable. Eventually a mutation–selection equilibrium will be attained at which a low rate of selective loss is balanced by occasional new duplications or insertions. The short intergenic spacers remaining in the nucleomorph (Zauner *et al.* 1999) are candidate sites for such ongoing mutational gains and losses. Whether the small variation in nucleomorph genome size is caused by variable amounts of residual non-coding DNA or by variation in gene content (or a combination of both) has yet to be investigated. Comparative studies of nucleomorph genomes could help elucidate the interplay between mutational and selective forces in the evolution of eukaryote genomes.

We thank Helen Merkens and Dr Xiaonan Wu for technical assistance, Natural Sciences and Engineering Research Council (Canada) for a research grant, the Canadian Institute for Advanced Research for Fellowship support for T.C.-S., and anonymous reviewers for helpful comments.

## REFERENCES

- Barnes, K. L., Craigie, R. A., Cattini, P. & Cavalier-Smith, T. 1982 Chromatin from the unicellular red alga *Porphyridium* has a nucleosome structure. *J. Cell Sci.* **57**, 151–160.
- Bennett, M. D. 1972 Nuclear DNA content and minimum generation time in plants. *Proc. R. Soc. Lond.* **B178**, 259–275.
- Boucher, N., Vaulot, D. & Partensky, F. 1991 Flow cytometric determination of phytoplankton DNA in cultures and oceanic populations. *Mar. Ecol. Prog. Ser.* **71**, 75–84.
- 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.
- Cavalier-Smith, T. 1980 *r*- and *K*-tactics in the evolution of protist developmental systems: cell and genome size, phenotype diversifying selection, and cell cycle patterns. *BioSystems* **12**, 43–59.
- Cavalier-Smith, T. (ed.) 1985*a* Introduction: the evolutionary significance of genome size. In *The evolution of genome size*, pp. 1–36. Chichester, UK: Wiley.
- Cavalier-Smith, T. (ed.) 1985*b* Eukaryote gene numbers, non-coding DNA and genome size. In *The evolution of genome size*, pp. 69–103. Chichester, UK: Wiley.
- Cavalier-Smith, T. (ed.) 1985*c* Cell volume and the evolution of eukaryote genome size. In *The evolution of genome size*, pp. 105–184. Chichester, UK: Wiley.
- Cavalier-Smith, T. (ed.) 1985*d* Selfish DNA and the origin of introns. *Nature* **383**, 283–284.
- Cavalier-Smith, T. (ed.) 1985*e* Selfish DNA, intragenomic selection and genome size. In *The evolution of genome size*, pp. 253–265. Chichester, UK: Wiley.
- Cavalier-Smith, T. (ed.) 1985*f* DNA replication and the evolution of genome size. In *The evolution of genome size*, pp. 211–251. Chichester, UK: Wiley.
- Cavalier-Smith, T. 1991 Coevolution of vertebrate genome, cell, and nuclear sizes. In *Symposium on the evolution of terrestrial vertebrates* (ed. G. Ghiara, F. Angelini, E. Olmo & L. Varano), pp. 51–86. Selected Symposia and Monographs, U.Z.I., 4. Modena, Italy: Mucchi.
- Cavalier-Smith, T. 1993 Evolution of the eukaryotic genome. In *The eukaryotic genome* (ed. P. Broda, S. G. Oliver & P. Sims), pp. 333–385. Cambridge University Press.

- Cavalier-Smith, T. 1995 Membrane heredity, symbiogenesis and the multiple origins of algae. In *Biodiversity and evolution* (ed. R. Arai, M. Kato & Y. Doi), pp. 75–114. Tokyo: The National Science Museum Foundation.
- Cavalier-Smith, T. 1999 Flagellate megaevolution: the basis for eukaryote diversification. In *The flagellates* (ed. B. S. C. Leadbeater & J. C. Green). London: Taylor & Francis. (In the press.)
- Cavalier-Smith, T. & Chao, E. E. 1997 Sarcomonad ribosomal RNA sequences, rhizopod phylogeny, and the origin of euglyphid amoebae. *Archiv Protistenk.* **147**, 227–236.
- Cavalier-Smith, T., Couch, J. A., Thorsteinsen, K. E., Gilson, P., Deane, J. A., Hill, D. R. A. & McFadden, G. I. 1996 Cryptomonad nuclear and nucleomorph 18S rRNA phylogeny. *Eur. J. Phycol.* **31**, 315–328.
- Doolittle, W. F. & Sapienza, C. 1980 Selfish genes, phenotype paradigm and genome evolution. *Nature* **284**, 617–618.
- Douglas, S. E., Murphy, C. A., Spencer, D. F. & Gray, M. W. 1991 Cryptomonad algae are evolutionary chimaeras of two phylogenetically distinct unicellular eukaryotes. *Nature* **350**, 148–151.
- Forbes, D. J., Kirschner, M. W. & Newport, J. W. 1983 Spontaneous formation of nuclear-like structure around bacteriophage DNA microinjected into *Xenopus* eggs. *Cell* **34**, 13–23.
- Friz, C. T. 1968 The biochemical composition of the free-living amoebae *Chaos chaos*, *Amoeba dubia*, and *Amoeba proteus*. *Comp. Biochem. Physiol.* **26**, 81–91.
- Gilson, P. R. & McFadden, G. I. 1996 The miniaturized nuclear genome of a eukaryotic endosymbiont contains genes that overlap, genes that are cotranscribed, and the smallest known spliceosomal introns. *Proc. Natl Acad. Sci. USA* **93**, 7737–7742.
- Gilson, P. R. & McFadden, G. I. 1999 Molecular, morphological and phylogenetic characterization of six chlorarachniophyte strains. *Phycol. Res.* **47**, 7–20.
- Gilson, P. R., Maier, U.-G. & McFadden, G. I. 1997 Size isn't everything: lessons in genetic miniaturisation from nucleomorphs. *Curr. Opin. Genet. Dev.* **7**, 800–806.
- Greenwood, A. D. 1974 The Cryptophyta in relation to phylogeny and photosynthesis. In *Eighth International Congress of Electron Microscopy* (ed. J. V. Sanders & D. J. Goodchild), pp. 566–567. Canberra, Australia: Australian Academy of Science.
- Hansmann, P. & Eschbach, E. 1990 Isolation and preliminary characterization of the nucleus and the nucleomorph of a cryptomonad, *Pyrenomonas salina*. *Eur. J. Cell Biol.* **52**, 373–378.
- Hibberd, D. J. & Norris, R. E. 1984 Cytology and ultrastructure of *Chlorarachnion reptans* (Chlorarachniophyta divisio nova, Chlorarachniophyceae classis nova). *J. Phycol.* **20**, 310–330.
- Hill, D. R. A. & Wetherbee, R. 1986 *Proteomonas sulcata* gen. et sp. nov. (Cryptophyceae), a cryptomonad with two morphologically distinct and alternating forms. *Phycologia* **25**, 521–543.
- Ishida, K., Green, B. R. & Cavalier-Smith, T. 1999 Diversification of a chimaeric algal group, the chlorarachniophytes: phylogeny of nuclear and nucleomorph small subunit rRNA genes. *Mol. Biol. Evol.* **16**, 321–331.
- McFadden, G. I., Gilson, P. R., Douglas, S. E., Hofmann, C. J. B. & Maier, U.-G. 1997 Bonsai genomics: sequencing the smallest eukaryotic genomes. *Trends Genet.* **13**, 46–49.
- Maier, U.-G., Hoffmann, C. J. B., Eschbach, S., Wolters, J. & Igloi, G. L. 1991 Demonstration of nucleomorph encoded eukaryotic small subunit ribosomal RNA in cryptomonads. *Mol. Gen. Genet.* **230**, 155–160.
- Maleszka, R. 1993 Electrophoretic analysis of the nuclear and organellar genomes in the ultra-small alga *Cyanidioschyzon merolae*. *Curr. Genet.* **24**, 548–550.
- Maynard Smith, J. & Szathmary, E. 1997 *The major transitions in evolution*. Oxford University Press.
- Ohno, S. 1972 So much 'junk' DNA in the genome. In *Evolution of genetic systems*, vol. 23 (ed. H. H. Smith), pp. 366–370. Brookhaven Symposia in Biology. New York: Gordon & Breach.
- Orgel, L. E. & Crick, F. H. C. 1980 Selfish DNA: the ultimate parasite. *Nature* **284**, 604–607.
- Petrov, D. A., Chao, Y.-C., Stephenson, E. C. & Hartl, D. L. 1998 Pseudogene evolution in *Drosophila* suggests a high rate of DNA loss. *Mol. Biol. Evol.* **15**, 1562–1567.
- Rensing, S. A., Goddemeier, M., Hoffmann, C. J. B. & Maier, U.-G. 1994 The presence of a nucleomorph *hsp70* gene is a common feature of Cryptophyta and Chlorarachniophyta. *Curr. Genet.* **26**, 451–455.
- Shuter, B. J., Thomas, J. E., Taylor, W. D. & Zimmerman, A. M. 1983 Phenotypic correlates of genomic DNA content in unicellular eukaryotes and other cells. *Am. Nat.* **122**, 26–44.
- Van de Peer, Y., Rensing, S., Maier, U.-G., & De Wachter, R. 1996 Substitution rate calibration of small subunit RNA identifies chlorarachniophyte endosymbionts as remnants of green algae. *Proc. Natl Acad. Sci. USA* **93**, 7744–7748.
- Velduis, M. J., Cucci, T. L. & Sieracki, M. E. 1997 Cellular DNA content of marine phytoplankton using two new fluorochromes: taxonomic and ecological implications. *J. Phycol.* **33**, 527–541.
- Vivares, C. P. 1999 On the genome of Microsporidia. *J. Euk. Microbiol.* **46**(Suppl.), 16A.
- Zauner, S., Fraunholz, M., Wastl, J., Penny, S., Beaton, M. J., Cavalier-Smith, T., Maier, U.-G. & Douglas, S. 1999 Chloroplast protein and centrosomal genes, a novel tRNA intron, and odd telomeres in an unusually compact eukaryotic genome—the cryptomonad nucleomorph. *Proc. Natl Acad. Sci. USA*. (Submitted.)

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.