ELSEVIER

Causes and effects of nuclear genome reduction

Patrick J Keeling and Claudio H Slamovits

Eukaryotic nuclear genomes are generally considered to be large and gene-sparse, but extreme reduction has taken place several times, resulting in small genomes with a high gene-density. This process involves loosing genes, compacting those that remain, or often both. Recently sequenced nuclear genomes include several that have converged to similar gene-densities by many means: variation in numbers and lengths of genes, intergenic regions and introns all contribute, but not equally in any given genome. Genomes of microsporidia and nucleomorphs have taken compaction much further, and in these hyper-compacted genomes there is evidence that some basic processes such as gene expression might be affected by genome form. In these genomes, normally weak forces might become more significant drivers of genome evolution.

Addresses
Canadian Institute for Advanced Research, Botany Department, University of British Columbia, 3529–6270 University Boulevard, Vancouver, BC, V6T 1Z4, Canada

Corresponding author: Keeling, Patrick J (pkeeling@interchange.ubc.ca)

Current Opinion in Genetics & Development 2005, **15**:601–608

This review comes from a themed issue on
Genomes and evolution
Edited by Stephen J O'Brien and Claire M Fraser

Available online 26th September 2005

0959-437X/$ – see front matter

DOI 10.1016/j.gde.2005.09.003

Introduction

At the bottom of the rabbit hole, Alice found a bottle labeled, ''Drink Me''. When she did, Alice shrank to a perfectly functioning, ten-inch miniature of herself. In reality, shrinking can be more difficult than simply drinking a potion, because the component parts of many systems are not themselves shrinkable, and so the system fails to function properly. In the world of eukaryotic nuclear genomes this is probably true, despite the fact that they vary in size by factors of hundreds of thousands (Figure 1), much more than all of Alice's many transformations combined.

Variations in genome size have been a persistent puzzle, mostly because genome size does not correspond to organismal complexity, often referred to as the 'C-value paradox'. Many other characteristics have been tied to genome size, including metabolic rates, body size, effective population size, and cell size or nucleus size [1–3,4•], the latter being the characteristic that most uniformly correlates with genome size. As with all complex characteristics, however, genome size is probably controlled by a variety of factors of varying importance in different organisms. These correlations are probably important across a broad spectrum of eukaryotes, but do not explain everything. Moreover, some genomes seem to depart from any otherwise stable trends, and these are often at the extremes of genome size [4•].

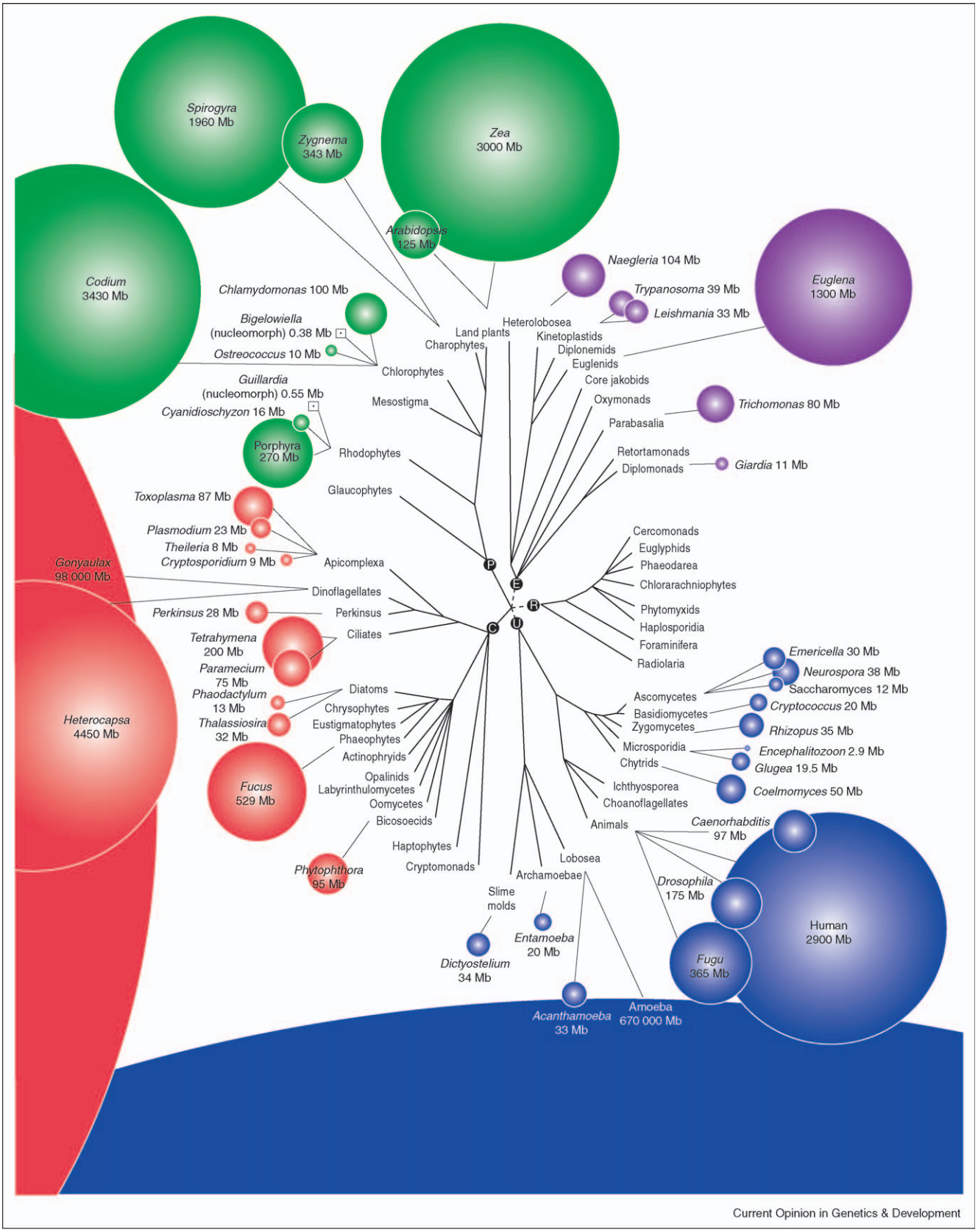
Here, we review the smallest eukaryotic genomes, illustrate some of the recent findings on how compaction impacts not only genome form but also function, and, along the way, point out some interesting characteristics of cells that have enabled compaction to take place. Some of the more familiar 'compact' genomes, such as that of yeast, are in reality only mildly compacted. Mildly compacted genomes (with gene densities of approximately 2 kb/gene) are not particularly uncommon and probably follow most of the same rules as larger genomes. A few genomes, however, are hyper-compacted (with densities closer to 1 kb/gene) and it is emerging that, at some point, such genomes might depart from some rules that govern other genomes — a point when compaction begins to affect basic processes such as replication, expression or recombination. At this point, processes and pressures that are common to all genomes but are generally relatively insignificant might become more important in genome evolution.

Different ways to shrink a genome: reduction versus compaction

Before examining a small genome, it is helpful to consider a simplistic division of ways that a genome can shrink: it can loose genes (elimination) or it can pack genes into a smaller space (compaction) [5]. Gene elimination results in a genome with a reduced coding capacity and a simplified proteome, whereas compaction by itself does not change the coding capacity but results in a higher gene density. It is relatively easy to explain gene elimination in many instances because the organisms are endosymbionts or intracellular parasites. These genomes have shed a large number of genes because they rely on a host cell for nutrients and biosynthesis of small molecules. Why compaction takes place is less obvious, and its effects on the genome are also more subtle, although not insignificant.

Returning to Alice, depending on what she ate (e.g. eating mushrooms versus eating cake), the effect on the way she shrank was different: sometimes she simply got smaller

Figure 1


Spirogyra
1960 Mb
Zygnema
343 Mb
Zea
3000 Mb
Arabidopsis
125 Mb
Codium
3430 Mb
Chlamydomonas 100 Mb
Bigelowiella
(nucleomorph) 0.38 Mb
Ostreococcus 10 Mb
Guillardia
(nucleomorph) 0.55 Mb
Cyanidioschyzon 16 Mb
Porphyra
270 Mb
Toxoplasma 87 Mb
Plasmodium 23 Mb
Theileria 8 Mb
Cryptosporidium 9 Mb
Gonyaulax
98 000 Mb
Perkinsus 28 Mb
Tetrahymena
200 Mb
Paramecium
75 Mb
Phaodactylum
13 Mb
Thalassiosira
32 Mb
Heterocapsa
4450 Mb
Fucus
529 Mb
Phytophthora
95 Mb
Naegleria 104 Mb
Trypanosoma 39 Mb
Leishmania 33 Mb
Euglena
1300 Mb
Heterolobosea
Kinetoplastids
Diplonemids
Euglenids
Land plants
Charophytes
Chlorophytes
Mesostigma
Rhodophytes
Glaucophytes
Apicomplexa
Dinoflaglellates
Perkinsus
Ciliates
Diatoms
Chrysophytes
Eustigmatophytes
Phaeophytes
Actinophryids
Opalinids
Labyrinthulomycetes
Oomycetes
Bicosoecids
Haptophytes
Cryptomonads
Core jakobids
Oxymonads
Parabasalia
Trichomonas 80 Mb
Retortamonads
Diplomonads
Giardia 11 Mb
Cercomonads
Euglyphids
Phaeodarea
Chlorarachniophytes
Phytomyxids
Haplosporidia
Foraminifera
Radiolaria
Emericella 30 Mb
Neurospora 38 Mb
Saccharomyces 12 Mb
Ascomycetes
Basidiomycetes
Zygomycetes
Cryptococcus 20 Mb
Rhizopus 35 Mb
Microsporidia
Chytrids
Encephalitozoon 2.9 Mb
Glugea 19.5 Mb
Ichthyosporea
Choanoflagellates
Coelmomyces 50 Mb
Animals
Caenorhabditis
97 Mb
Drosophila
175 Mb
Human
2900 Mb
Fugu
365 Mb
Lobosea
Archamoebae
Slime
molds
Entamoeba
20 Mb
Dictyostelium
34 Mb
Acanthamoeba
33 Mb
Amoeba
670 000 Mb
P
E
R
C
U
Current Opinion in Genetics & Development

and sometimes parts of her shrank more than others. Examining different groups of eukaryotes with compacted genomes we also see different combinations of elimination versus compaction.

We are fortunate now to have complete genome sequences for several eukaryotes with mildly compacted genomes, and also to have more than one closely related species in a couple of cases (Table 1). Focusing on protists, the best resource of complete genome sequences is in the Apicomplexa, a diverse phylum of intracellular parasites. In addition to *Plasmodium* [6,7], there are now complete genomes for two species of *Cryptosporidium* [8•,9•] and two species of *Theileria* [10•,11•], all of which have mildly compacted genomes with densities in the order of twice that seen in *Plasmodium*. Each has achieved this compaction by different means. All have eliminated genes to about the same extent, but *Theileria* has reduced its intergenic spaces more than *Cryptosporidium*, whereas *Cryptosporidium* has reduced its intron content considerably more than *Theileria* (Table 1). Comparing the genomes of congeneric relatives has already revealed a high degree of conservation in gene order in *Theileria*, and many features relating to the expansion of species-specific gene families in both genera [8•–11•]. In other groups, even just considering the simplistic characteristics of coding capacity, gene density and intron numbers, we see similar variations in modes of reduction: *Dictyostelium* has retained many genes and introns, but reduced its intergenic regions [12], whereas *Entamoeba* has reduced all three characteristics [13], and *Cyanidioschyzon* has reduced gene number and lost nearly all of its introns [14]. Interestingly, even though many of these genomes have reduced by eliminating genes, there is also evidence for acquiring new genes by horizontal gene transfer in several parasites with reduced genomes [13,15–20].

One of the more interesting compacted genomes is that of the ciliate *Paramecium* [21••]. This genome is relatively large and contains an estimated 30 000 genes. It has clearly eliminated a small amount, and yet it has compacted this large repertoire of genes significantly. It has retained many introns but shrunk them near to the lower limit known, and, most interestingly, has drastically reduced intergenic lengths (Table 1). This genome shows, perhaps better than does any other eukaryote to date, that elimination and compaction do not necessarily go together. We do not know why, but one possible factor is the separation of germ and somatic nuclei. In the development of somatic chromosomes from germ-line chromosomes, a substantial amount of sequence is discarded: nearly all non-genic, including most or all of the transposable elements in the genome. The somatic genome, therefore, has a higher gene density: perhaps this separation of selective pressures shields the somatic genome from negative selection on some otherwise deleterious effects of compaction.

Extreme compaction: microsporidia and nucleomorphs

Even the most gene-dense of the genomes described above do not fall below 2 kb/gene overall, suggesting that beyond this density functional difficulties might arise. It is easy to imagine what kinds of problem might be faced: for instance, regulatory regions interfering with one another, or transcriptional fronts colliding [22]. However, in three different cases, nuclear gene densities have hyper-compacted: in microsporidia and the nucleomorphs of cryptomonads and chlorarachniophytes.

Microsporidia are a group of diverse obligate intracellular parasites now known to be closely related to fungi. This has been a contentious issue because they were formerly thought to be early-diverging eukaryotes, until conflicting phylogenetic data suggested that they were fungi [23]. The reason behind this incongruence has now been shown in a genome-wide phylogenetic analysis, which demonstrated a strong correlation between how highly divergent a gene was and its tendency to show microsporidia branching early in eukaryotes, a classic phylogenetic artefact [24•]. Microsporidia are highly adapted to their parasitic way of life: they have a highly advanced and sophisticated infection mechanism (Figure 2), but they are also degenerate in many ways. Known microsporidian genomes range in length from 19.5 Mb to a mere 2.3 Mb [25–28], and the complete 2.9 Mb sequence of the *Encephalitozoon cuniculi* genome has been determined and shown to be a model of both elimination and compaction [29,30]. The genome contains just fewer than 2000 protein-coding genes, a great proportion of which are involved in replication and gene expression. Many biosynthetic pathways have been lost — particularly those involving the biosynthesis of small molecules such as nucleotides or amino acids — which, not surprisingly, suggests a heavy reliance on the host cell for nutrients. In terms of compaction, the average intergenic size is only 129 bp; by comparison, this is approximately a quarter of the size of that found in the mildly compacted genome of *Saccharomyces*. Furthermore, there are few short repeats, only one repeated block of the genome, and no evidence

(Figure 1 Legend) The tree of eukaryotes, showing some variations in genome size. The tree is a composite based on a variety of available data according to the study by Keeling *et al*. [48]. There are five hypothetical supergroups, indicated by black circles at their bases: Excavates (E), Unikonts (U), Plants (P), Chromalveolates (C) and Rhizaria (R). Genome sizes are derived either from complete sequences or from estimates based on methods such as CHEF (contour-clamped homogeneous electric-field) gel electrophoresis or quantifying DNA content. Some estimates — in particular, those of larger genomes — are probably erroneous (for example, as a result of polyploidy) but it is nevertheless clear that the range of genome sizes in eukaryotes is vast. Genome sizes are taken from too many sources to list them all, but useful compilations can be found in the studies by Lynch and Conery [2] and Kapraun [3], and at www.cbs.dtu.dk/databases/DOGS.

Table 1

Some characteristics of model genomes compared with compact genomes[*].

Organism	Genome size (Mb)	Number of genes	Gene density (kb/gene)	Mean intergenic distance (bp)	% Intron-containing [¶]	Mean intron size (bp)
Homo sapiens	2851	22 287	127.90	$\sim 10^4$	85	3365
Arabidopsis thaliana	125	25 498	4.90	2900	79	170
Saccharomyces cerevisiae	12.50	5770	2.09	500	5	287
Apicomplexa						
Plasmodium falciparum	12.03	5268	4.34	1694	54	179
Theileria parva	8.31	4035	2.20	405	74	94
Theileria annulata	8.35	3792	2.05	369	71	69
Cryptosporidium parvum	9.11	3952	2.30	566	5	[‡]
Cryptosporidium hominis	9.16	3994	2.29	716	5–20	[‡]
Red Algae						
Cyanidioschyzon merolae	16.52	5331	3.10	[‡]	0.5	[‡]
Amoebozoa						
Dictyostelium discoideum	33.82	12 500	2.72	[‡]	69	146
Entamoeba histolytica	23.75	9938	2.39	[‡]	25	[‡]
Ciliate						
Paramecium tetraurelia [∞]	~100	~30 000	2.14	202	84	25
Microsporidia						
Encephalitozoon cuniculi	2.51	1997	1.25[§]	129	0.6	[‡]
Nucleomorphs						
Guillardia theta	0.551	486	1.13[§]	70	1.8	46
Bigelowiella natans	0.373	~308	1.21[§]	113	80	19

[*] Values for several of these characteristics vary among published reports, in part because they change as annotations improve, and in part because they are measured in different ways. Accordingly, the values should be taken as an approximation that demonstrates the general trends in the genome. In general, the values reported here are taken from the most current source or from the report describing the genome

[‡] Value not available.

[§] Overall gene density is based on genome size divided by protein-coding gene number. However in the smallest genomes, essential non-coding regions (e.g. telomeres) and small RNA genes (e.g. tRNAs) have a disproportionate affect on density values compared with their affect on large genomes. The effective gene density in the chromosomal cores can be much higher (e.g. 20% higher in *E. cuniculi*), so this value is useful for comparison, but a more interesting value with respect to the effects of compaction is the mean intergenic length.

[¶] The number of genes containing introns is most commonly reported, but in terms of compaction this value is only partly informative since the number of introns in each of these genes is also relevant, or the total number of genes across the genome.

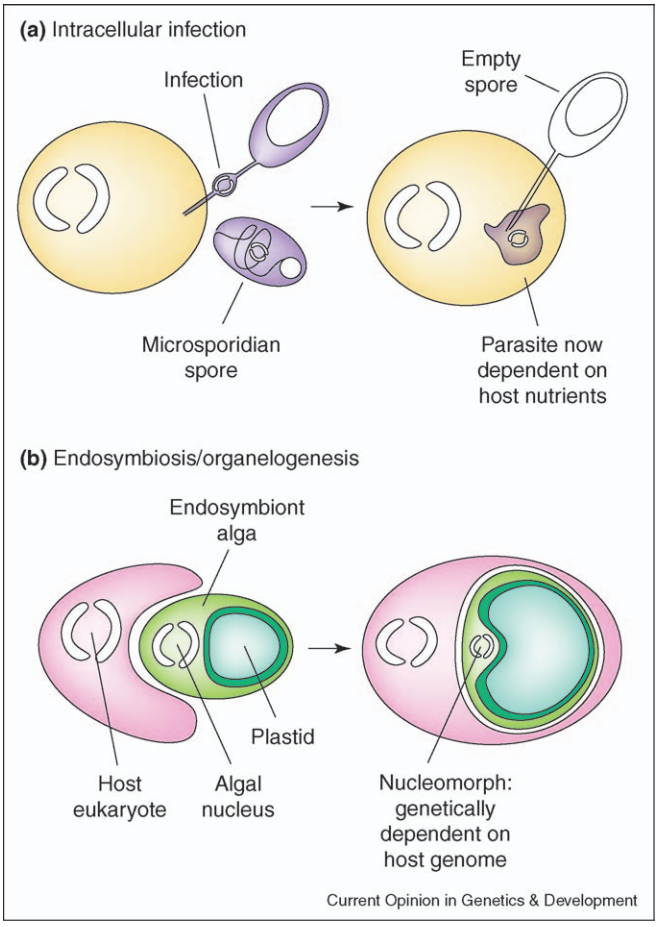
[∞] Based on partial genome.

of transposons. There are only 13 introns, located within 12 genes, known in the entire genome. Overall, these characteristics result in a gene density of 1.25 bp/gene [30].

Nucleomorphs are relict nuclei of endosymbiotic algae found in cryptomonads and chlorarachniophytes (Figure 2) [31,32]; other groups have endosymbiotic algae that retain nuclei [33], but virtually nothing is known about their genome form or content. The apparent function of nucleomorph genomes is to supply a few proteins to the plastid with which they are associated. In most secondary plastids, these protein genes have all moved to the host nucleus, rendering the endosymbiont genome obsolete [34,35]. In cryptomonads and chlorarachniophytes, however, the endosymbionts have retained a small number of essential genes — encoding plastid proteins and, perhaps, also proteins involving in targeting — that cannot be lost until these genes also move to the nucleus. Cryptomonad and chlorarachniophyte nucleomorphs evolved independently from red and green algae, respectively, but they share many features of overall structure in common [31]. Their genomes are all composed of three linear chromosomes with gene-dense cores and rRNA operons as subtelomeric repeats. They are by far the smallest nuclear genomes known: the genomes of cryptomonads range from 450 to 710 kb [36], and those of chlorarachniophytes range from 373 to only 455 kb [31,37].

Complete genome sequences are known from a representative of both groups: the cryptomonad *Guillardia theta* [38] and the chlorarachniophyte *Bigelowiella natans* ([39] and PR Gilson, V Su, CH Slamovits, ME Reith, PJ Keeling and GI McFadden, unpublished). These genomes are similar to microsporidia in terms of compaction, but have eliminated far more genes because of their even greater dependence on their host. The potential for reduction in nucleomorphs is higher than that in microsporidia for an intriguing reason: not only are they dependent upon their hosts for energy and small molecules but, because secondary algae have transferred many genes to

Figure 2



Intracellular life and small genomes. **(a)** Spores are the only stage of microsporidian obligate intracellular parasites that survive outside another cell. To the left, two spores (purple) are shown: one is in its dormant state (below); and one (above) is infecting a host (yellow). Infection takes place using a projectile tube that injects the parasite directly into the host cytoplasm, as shown to the right. Because all growth and development occurs inside this host cell, the parasite become deeply dependent upon the host for energy and nutrients, enabling them to lose a great number of genes, and their genomes to shrink (although this does not explain why they compact). **(b)** Nucleomorphs arose when an alga (green) was eaten by another eukaryote (pink), as shown on the left. Usually, this meal would be digested, but on several occasions the alga was retained and the two cells integrated in a process called secondary endosymbiosis, forming a new algal lineage. Sometimes secondary endosymbionts lose their nucleus altogether, but in two instances the nucleus of the endosymbiont was not lost and, instead, degenerated substantially to what we now call a nucleomorph (shown on right). The endosymbiont is dependent on the host for energy and nutrients, and so the degeneration of nucleomorph genomes is similar in many ways to that seen in microsporidian genomes. In addition, however, many nucleomorph genes appear to have moved to the host nucleus, and their products are targeted back to the endosymbiont, enabling nucleomorphs to degenerate even further than a parasite might. Overall, endosymbionts and intracellular parasites and the relationship of each to their host differ in many ways, but there are also many similarities and parallels.

the host and developed a mechanism to target the protein products back to the plastid [34], they have also become dependent on their host for most of their proteins. Accordingly, nucleomorphs might encode fewer proteins than are needed for the most basic functions. Indeed, the complete *G. theta* nucleomorph genome is missing genes for several essential proteins, such as DNA polymerases [38]. It has long been speculated that the genes for such proteins have moved to the host nucleus, but, until recently, no such gene was found. Now, however, *G. theta* nuclear genes for several putatively endosymbiont-targeted proteins have been found [40••]. These genes are one of the 'holy grails' of nucleomorph research because they hold many keys to important questions about protein-trafficking. Indeed, expressing these genes as green fluorescent protein (GFP)-fusions in a genetically tractable diatom has already proposed an intriguing mechanism for traversing the long-mysterious third membrane of complex plastids: it is suggested that a plastid outer-membrane complex is duplicated and present in both membranes, although the two complexes are subtly different [40••].

Genome-wide effects of compaction

Aside from altering the form of the genome, one of the first features of a functional nature to be noted in the hyper-compacted genomes of both microsporidia and nucleomorphs was that the proteins encoded in them tend to be smaller than their homologues in related genomes [30,38]. In *E. cuniculi* it was hypothesized that these smaller proteins have not arisen directly as a result of compaction, however, because a shrinking proteome could result in simpler interaction networks, which would, in turn, facilitate the loss of protein domains responsible for these interactions [30].

Perhaps a stronger link to compaction is seen in conservation of the overall gene order of the genome in microsporidia. Comparisons between the distantly related microsporidia *E. cuniculi* and *Antonospora locustae* revealed a relatively high number of conserved gene-pairs and showed that the intergenic regions between these conserved pairs were markedly shorter in both genomes [41••]. It was hypothesized that the short intergenic regions slowed genomic rearrangements simply by reducing the number of possible breakpoints [41••]. In the mildly compacted genome of yeasts, conservation of synteny has been linked to short intergenic regions, but, overall, other factors such as co-expression appear to be more important [42]. The extreme conditions in microsporidian genomes might, therefore, place new prominence on otherwise less significant forces. Currently, there is no comparative genomic data from nucleomorphs, but they offer a very interesting system to study synteny over time, because the host nuclear genome has been evolving in parallel with the nucleomorph. If one could demonstrate that two host genomes shared significantly less synteny than the nucleomorph genomes from the same species, it would prove that the nucleomorph genome was evolving more slowly than that of the host, because both host and nucleomorph genomes would have diverged at exactly the same time. On a more practical level, the conserved synteny in microsporidia is also an ideal guide for gene discovery, particularly as the divergent nature of their genes makes identifying them a challenge. Recently, this conservation was used to identify, in the absence of sequence similarity, *A. locustae* homologues of two *E. cuniculi* proteins that are crucial for infection [43•]. Functional studies confirm this identification and highlight the practical value of understanding the dynamics of a genome.

Although these characteristics are each intriguing aspects of reduction, they do not obviously explain the rarity of hyper-reduction in nuclear genomes. One potential explanation has emerged not from the genomes themselves but from examining gene expression within them. Expressed sequence tag (EST) projects from microsporidia and both nucleomorph genomes have revealed a high frequency of overlapping transcription in all three systems [44••]. In most eukaryotes, transcripts from one gene do not typically overlap with those of adjacent genes, and if they are engineered to do so they often have disastrous effects on expression of one or both genes [22,45,46]. In microsporidia, however, about 15% of transcripts appear to initiate within the upstream gene, terminate within or beyond the downstream gene, or both [44••,47]. In nucleomorphs, little initiation within upstream genes was observed, but most transcripts read through into downstream genes (97% in *G. theta* and 82% in *B. natans*), sometimes encoding all or parts of as many as four genes [44••]. It is possible that one of the forces containing mildly compacted genomes to about 2 kb/gene is the potential for adjacent genes to interfere with one another's expression, and that hyper-compact genomes have overcome this constraint, but this hypothesis needs a great deal of work for confirmation.

Conclusions and more questions

Small genomes tend to be the first to get sequenced, so even in these early days of comparative nuclear genomics there are already several reduced genomes for further analysis and comparison, but this is only the beginning. New genomes of potential interest to reduction and compaction are on the horizon, including those of the reduced green alga *Ostreococcus*, the ciliate *Tetrahymena*, and additional Apicomplexa and microsporidia, some with much larger genomes than those presently analysed. The number of genomes with densities around 2 kb/gene is interesting, but whether it represents some functional line that is difficult to cross needs not only more genomic data to test the robustness of the observation but also experimental data that might explain what are the constraints on compaction. Moreover, there appear to be correlations between compaction, synteny and transcriptional properties in hyper-compacted genomes, but, again, the full strength of comparative analyses have not yet been brought to bear on these questions.

Acknowledgements

We thank GI McFadden, PR Gilson and V Su for allowing us to discuss unpublished data from the collaborative *B. natans* nucleomorph genome project, F Delbac and U-G Maier for giving us access to in-press manuscripts, and B Williams for critical comments. Work on microsporidian genomics is supported by the Canadian Institutes for Health Research. PJK is a Fellow of the Canadian Institute for Advanced Research and a New Investigator of the Michael Smith Foundation for Health Research and the Canadian Institutes for Health Research.

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. Vinogradov AE: **Evolution of genome size: multilevel selection, mutation bias or dynamical chaos?** *Curr Opin Genet Dev* 2004, **14**:620-626.

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

3. Kapraun DF: **Nuclear DNA content estimates in multicellular green, red and brown algae: phylogenetic considerations**. *Ann Bot (Lond)* 2005, **95**:7-44.

4. • Cavalier-Smith T: **Economy, speed and size matter: evolutionary forces driving nuclear genome miniaturization and expansion**. *Ann Bot (Lond)* 2005, **95**:147-175.
A current review that covers a range of questions surrounding eukaryotic genome size by synthesizing data from the molecular, cell and population levels.

5. Keeling PJ: **Reduction and compaction in the genome of the apicomplexan parasite, *Cryptosporidium parvum***. *Dev Cell* 2004, **6**:614-616.

6. Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S *et al.*: **Genome sequence of the human malaria parasite *Plasmodium falciparum***. *Nature* 2002, **419**:498-511.

7. Carlton JM, Angiuoli SV, Suh BB, Kooij TW, Pertea M, Silva JC, Ermolaeva MD, Allen JE, Selengut JD, Koo HL *et al.*: **Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii***. *Nature* 2002, **419**:512-519.

8. • Abrahamsen MS, Templeton TJ, Enomoto S, Abrahante JE, Zhu G, Lancto CA, Deng M, Liu C, Widmer G, Tzipori S *et al.*: **Complete genome sequence of the apicomplexan, *Cryptosporidium parvum***. *Science* 2004, **304**:441-445.
See annotation [9•].

9. • Xu P, Widmer G, Wang Y, Ozaki LS, Alves JM, Serrano MG, Puiu D, Manque P, Akiyoshi D, Mackey AJ *et al.*: **The genome of *Cryptosporidium hominis***. *Nature* 2004, **431**:1107-1112.
This study and that by Abrahamsen *et al*. [8•] report the complete genomes of *C. parvum* and *C. hominis*, respectively, together providing a comparison between the modes of reduction in two closely related, mildly compacted genomes. In these *Cryptosporidium* species, unlike in *Theileria* [11•], there is almost no variation in gene content.

10. • Pain A, Renauld H, Berriman M, Murphy L, Yeats CA, Weir W, Kerhornou A, Aslett M, Bishop R, Bouchier C *et al.*: **Genome of the host-cell transforming parasite Theileria annulata compared with *T. parva***. *Science* 2005, **309**:131-133.
See annotation [11•].

11. • Gardner MJ, Bishop R, Shah T, de Villiers EP, Carlton JM, Hall N, Ren Q, Paulsen IT, Pain A, Berriman M *et al.*: **Genome sequence of *Theileria parva*, a bovine pathogen that transforms lymphocytes**. *Science* 2005, **309**:134-137.
As above, this study and that by Gardner *et al*. [10•] describe two complete sequences of closely related species with mildly compacted genomes. These reports once again show different modes of reduction in addition to interesting species-specific amplifications of gene families, and a high degree of gene-order conservation.

12. Eichinger L, Pachebat JA, Glockner G, Rajandream MA, Sucgang R, Berriman M, Song J, Olsen R, Szafranski K, Xu Q *et al.*: **The genome of the social amoeba *Dictyostelium discoideum***. *Nature* 2005, **435**:43-57.

13. Loftus B, Anderson I, Davies R, Alsmark UC, Samuelson J, Amedeo P, Roncaglia P, Berriman M, Hirt RP, Mann BJ *et al.*: **The genome of the protist parasite *Entamoeba histolytica***. *Nature* 2005, **433**:865-868.

14. Matsuzaki M, Misumi O, Shin IT, Maruyama S, Takahara M, Miyagishima SY, Mori T, Nishida K, Yagisawa F, Yoshida Y *et al.*: **Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D**. *Nature* 2004, **428**:653-657.

15. Fast NM, Law JS, Williams BA, Keeling PJ: **Bacterial catalase in the microsporidian *Nosema locustae*: implications for microsporidian metabolism and genome evolution**. *Eukaryot Cell* 2003, **2**:1069-1075.

16. Huang J, Mullapudi N, Sicheritz-Ponten T, Kissinger JC: **A first glimpse into the pattern and scale of gene transfer in Apicomplexa**. *Int J Parasitol* 2004, **34**:265-274.

17. Huang J, Mullapudi N, Lancto CA, Scott M, Abrahamsen MS, Kissinger JC: **Phylogenomic evidence supports past endosymbiosis, intracellular and horizontal gene transfer in *Cryptosporidium parvum***. *Genome Biol* 2004, **5**:R88.

18. Richards TA, Hirt RP, Williams BA, Embley TM: **Horizontal gene transfer and the evolution of parasitic protozoa**. *Protist* 2003, **154**:17-32.

19. Striepen B, Pruijssers AJ, Huang J, Li C, Gubbels MJ, Umejiego NN, Hedstrom L, Kissinger JC: **Gene transfer in the evolution of parasite nucleotide biosynthesis**. *Proc Natl Acad Sci USA* 2004, **101**:3154-3159.

20. Striepen B, White MW, Li C, Guerini MN, Malik SB, Logsdon JM Jr, Liu C, Abrahamsen MS: **Genetic complementation in apicomplexan parasites**. *Proc Natl Acad Sci USA* 2002, **99**:6304-6309.

21. •• Zagulski M, Nowak JK, Le Mouel A, Nowacki M, Migdalski A, Gromadka R, Noel B, Blanc I, Dessen P, Wincker P *et al.*: **High coding density on the largest *Paramecium tetraurelia* somatic chromosome**. *Curr Biol* 2004, **14**:1397-1404.
Paramecium has a large genome with many genes but, despite this, the somatic chromosome described here is highly compacted with very short intergenic regions. This shows that compaction can take place without reducing the proteomic complexity of the organism. In the case of ciliates, this might be related to the generation of a somatic nucleus from a germ line nucleus, which includes the elimination of a great deal of non-coding DNA from the genome.

22. Shearwin KE, Callen BP, Egan JB: **Transcriptional interference — a crash course**. *Trends Genet* 2005, **21**:339-345.

23. Keeling PJ, Fast NM: **Microsporidia: biology and evolution of highly reduced intracellular parasites**. *Annu Rev Microbiol* 2002, **56**:93-116.

24. • Thomarat F, Vivarès CP, Gouy M: **Phylogenetic analysis of the complete genome sequence of *Encephalitozoon cuniculi* supports the fungal origin of microsporidia and reveals a high frequency of fast-evolving genes**. *J Mol Evol* 2004, **59**:780-791.
The debate over the phylogenetic position of microsporidia has endured for many years because different genes often yield different trees. Here, a large proportion of the protein-coding genes from the *E. cuniculi* genome were analysed and a strong correlation was found between the level of divergence and the phylogeny, which should put this debate to rest. Divergent genes tend to show microsporidia branching early whereas more-conserved genes show them allied with fungi.

25. Streett DA: **Analysis of *Nosema locustae* (Microsporidia: Nosematidae) chromosomal DNA with pulsed-field gel electrophoresis**. *J Invert Pathol* 1994, **63**:301-303.

26. Biderre C, Pagès M, Méténier G, David D, Bata J, Prensier G, Vivarès CP: **On small genomes in eukaryotic organisms: molecular karyotypes of two microsporidian species (Protozoa) parasites of vertebrates**. *C R Acad Sci III* 1994, **317**:399-404.

27. Biderre C, Pagès M, Méténier G, Canning EU, Vivarès CP: **Evidence for the smallest nuclear genome (2.9 Mb) in the microsporidium *Encephalitozoon cuniculi***. *Mol Biochem Parasitol* 1995, **74**:229-231.

28. Peyretaillade E, Biderre C, Peyret P, Duffieux F, Méténier G, Gouy M, Michot B, Vivarès CP: **Microsporidian *Encephalitozoon cuniculi*, a unicellular eukaryote with an unusual chromosomal dispersion of ribosomal genes and a LSU rRNA reduced to the universal core**. *Nucleic Acids Res* 1998, **26**:3513-3520.

29. Peyret P, Katinka MD, Duprat S, Duffieux F, Barbe V, Barbazanges M, Weissenbach J, Saurin W, Vivarès CP: **Sequence and analysis of chromosome I of the amitochondriate intracellular parasite *Encephalitozoon cuniculi* (Microspora)**. *Genome Res* 2001, **11**:198-207.

30. Katinka MD, Duprat S, Cornillot E, Méténier G, Thomarat F, Prenier G, Barbe V, Peyretaillade E, Brottier P, Wincker P *et al.*: **Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi***. *Nature* 2001, **414**:450-453.

31. Gilson PR, McFadden GI: **Jam packed genomes–a preliminary, comparative analysis of nucleomorphs**. *Genetica* 2002, **115**:13-28.

32. Cavalier-Smith T: **Nucleomorphs: enslaved algal nuclei**. *Curr Opin Microbiol* 2002, **5**:612-619.

33. Dodge JD: **Observations on the fine structure of the eyespot and associated organelles in the dinoflagellate *Glenodinium foliaceum***. *J Cell Sci* 1969, **5**:479-493.

34. McFadden GI: **Primary and secondary endosymbiosis and the origin of plastids**. *J Phycol* 2001, **37**:951-959.

35. Keeling PJ: **The diversity and evolutionary history of plastids and their hosts**. *Am J Bot* 2004, **91**:1481-1493.

36. Rensing SA, Goddemeier M, Hofmann CJ, Maier UG: **The presence of a nucleomorph *hsp70* gene is a common feature of Cryptophyta and Chlorarachniophyta**. *Curr Genet* 1994, **26**:451-455.

37. Gilson PR, McFadden GI: **Molecular, morphological and phylogenetic characterization of six chlorarachniophyte strains**. *Phycol Res* 1999, **47**:7-19.

38. Douglas S, Zauner S, Fraunholz M, Beaton M, Penny S, Deng LT, Wu X, Reith M, Cavalier-Smith T, Maier UG: **The highly reduced genome of an enslaved algal nucleus**. *Nature* 2001, **410**:1091-1096.

39. Gilson PR, McFadden GI: **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* 1996, **93**:7737-7742.

40. •• Gould SB, Sommer MS, Hadfi K, Zauner S, Kroth PG, Maier U-G: **Protein targeting into the complex plastid of cryptophytes**. *J Mol Evol*, in press.
It has long been known that cryptomonad and chlorarachniophyte algae must have nuclear-encoded genes for proteins that are targeted to the cytosol of the eukaryotic endosymbiont. These proteins are crucially important to our understanding of targeting to secondary plastids: because they only travel part way to the plastid, they can be seen as a 'natural dissection' of the targeting pathway. This report identifies and characterises the first such genes and analyses their localization in a diatom genetic system.

41. •• Slamovits CH, Fast NM, Law JS, Keeling PJ: **Genome compaction and stability in microsporidian intracellular parasites**. *Curr Biol* 2004, **14**:891-896.
Comparing the frequency of conserved gene pairs in two distantly related microsporidia, the authors revealed a high level of conservation, and this conservation was more likely in cases in which intergenic regions were short. The study concluded that the compaction of the genome probably slowed down the rate of naturally occurring reorganizations by limiting the number of non-deleterious breakpoints.

42. Hurst LD, Williams EJ, Pal C: **Natural selection promotes the conservation of linkage of co-expressed genes**. *Trends Genet* 2002, **18**:604-606.

43. • Polonais V, Prensier G, Méténier G, Vivarès CP, Delbac F: **Microsporidian polar tube proteins: highly divergent but closely linked genes encode PTP1 and PTP2 in members of the evolutionarily distant *Antonospora* and *Encephalitozoon* groups**. *Fungal Genet Biol* 2005, **43**:491-803.
Functionally annotating genes that are poorly conserved at the primary sequence level is a huge challenge, even in small genomes. Here, the authors demonstrated that synteny can be used to identify homologues of important genes when no detectable sequence conservation exists. Functional studies were also used to confirm the identity of the genes, suggesting that synteny will be an important tool for translating functional annotation from one microsporidian to another.

44. •• Williams BAP, Slamovits CH, Patron NJ, Fast NM, Keeling PJ: **A high frequency of overlapping gene expression in compacted eukaryotic genomes**. *Proc Natl Acad Sci USA* 2005, **102**:10936-10941.
Expressed sequence tag (EST) data from a microsporidian and two nucleomorphs showed that cDNAs in all three systems frequently include sequence of more than one gene. Such overlapping transcription can be a problem in other nuclear genomes, and this transcriptional interference is one possible barrier to further reduction in many mildly compacted genomes. In these hyper-compacted genomes, however, overlapping transcription might be tolerated at very high levels: nearly all *G. theta* nucleomorph transcripts extend into downstream genes.

45. Prescott EM, Proudfoot NJ, Furger A, Dye MJ, Greger IH: **Transcriptional collision between convergent genes in budding yeast. Integrating mRNA processing with transcription Poly(A) signals control both transcriptional termination and initiation between the tandem *GAL10* and *GAL7* genes of *Saccharomyces cerevisiae***. *Proc Natl Acad Sci USA* 2002, **99**:8796-8801.

46. Springer C, Valerius O, Strittmatter A, Braus GH: **The adjacent yeast genes *ARO4* and *HIS7* carry no intergenic region**. *J Biol Chem* 1997, **272**:26318-26324.

47. Slamovits CH, Keeling PJ: **Class II photolyase in a microsporidian intracellular parasite**. *J Mol Biol* 2004, **341**:713-721.

48. Keeling PJ, Burger G, Durnford DG, Lang BF, Lee RW, Pearlman RE, Roger AJ, Gray MW: **The tree of eukaryotes**. *Trends Ecol Evol* 2005, in press.