

Minireview

## Nucleomorph genomes: much ado about practically nothing

Paul R Gilson

Address: Centre for Cellular and Molecular Biology, School of Biological and Chemical Sciences, Deakin University, Victoria 3125, Australia.  
E-mail: prgilson@deakin.edu.au

Published: 30 July 2001

*Genome Biology* 2001, **2(8)**:reviews1022.1-1022.5

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2001/2/8/reviews/1022>

© BioMed Central Ltd (Print ISSN 1465-6906; Online ISSN 1465-6914)

### Abstract

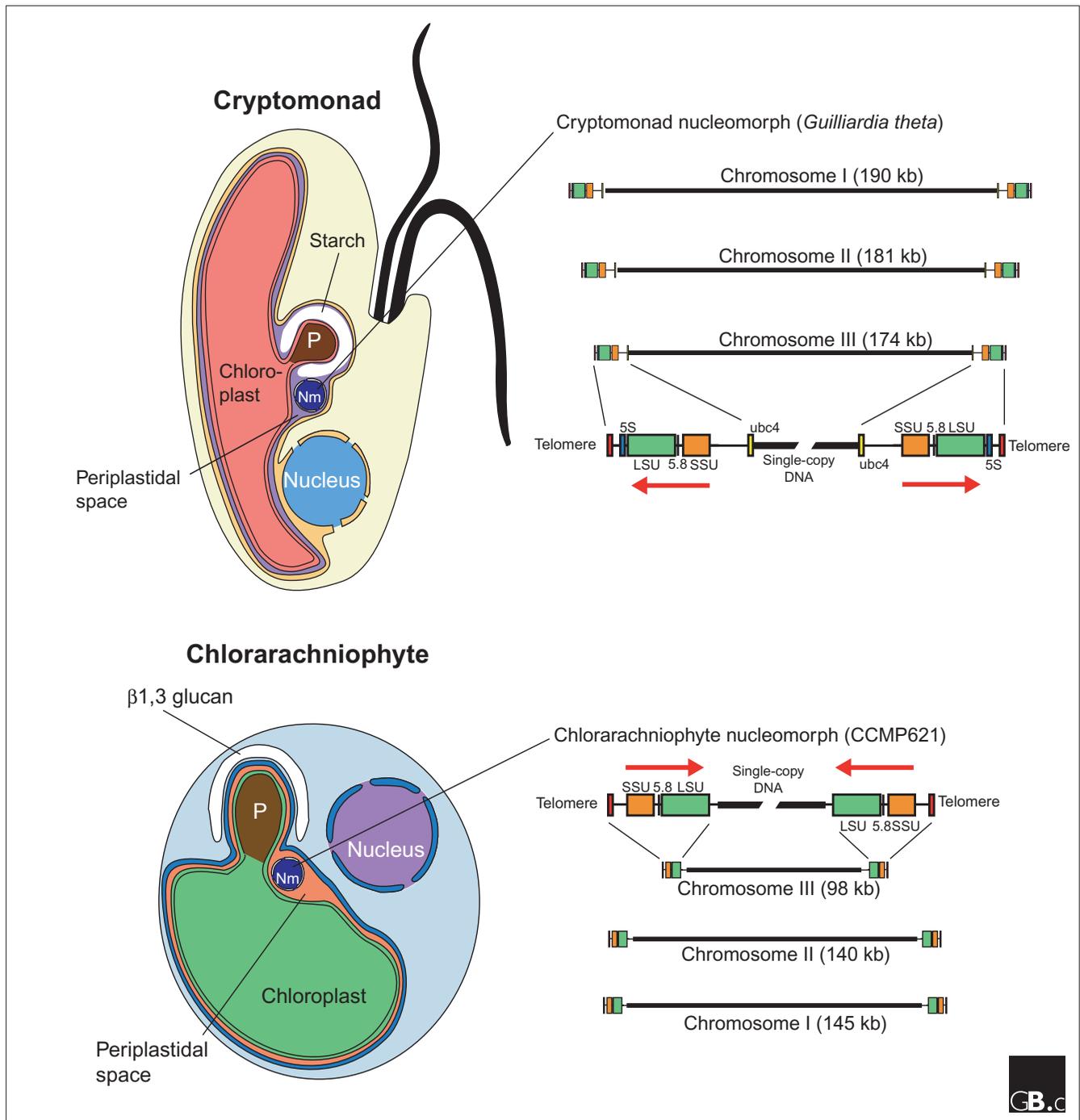
The DNA sequence of one of the smallest eukaryotic genomes has recently been finished - that of the reduced nucleus, or nucleomorph, of an algal endosymbiont that resides within a cryptomonad host cell. Its sequence promises insights into chloroplast acquisition, the constraints on genome size and the basic workings of eukaryotic cells.

It may surprise many that the sixth eukaryotic genome to be sequenced comes not from some familiar model organism but from a nucleus whose DNA content is less than that of most bacteria [1]. This nucleus comes from *Guillardia theta*, a member of a group of biflagellate protists called cryptomonads (Figure 1). The tiny nucleus, called a nucleomorph, is not the main cellular nucleus but is the residual nucleus of a red algal endosymbiont that resides inside the endoplasmic reticulum (ER) of its host cell (Figure 1) [2]. Another group of protists, called chlorarachniophytes, also have nucleomorphs, and these are the remnant nuclei of green algal endosymbionts (Figure 1) [3,4]. Both protist groups are believed to have acquired their endosymbionts about 600 million years ago [5] when, instead of digesting their algal-prey meals as they would normally do, the host cells retained the algae, to feed the heterotrophic host cell with carbohydrates made by the algal chloroplast. Over time the algae and hosts became interdependent and the algae lost their autonomy, becoming highly streamlined for photosynthesis. Most of the trappings normally found in a eukaryotic cell, such as mitochondria, endoplasmic reticulum and flagella, have been entirely lost from the endosymbionts. All that is left is a prominent double-membrane-bound chloroplast surrounded by a small volume of cytoplasm (called the periplastidal space), in which resides the diminutive nucleomorph (Figure 1). A relic plasma membrane still envelops the alga that is in turn surrounded by a modified food vacuole in chlorarachniophytes or by the host cell's ER in cryptomonads [6]. The chloroplast is, in effect, bound by four membranes (Figure 1).

The primary chloroplasts of red algae, glaucophytes, green algae and plants are bound by two membranes and are derived from cyanobacterial endosymbionts, but chloroplasts bound by three or four membranes are also common [7]. These so-called 'multi-membraned' or 'complex' plastids are derived from engulfed algae but, unlike the chloroplasts of cryptomonads and chlorarachniophytes, they have completely lost their nuclei: presumably, the nuclei went through a nucleomorph-like stage as redundant genes were lost and essential genes were transferred to the nucleus of the host cell. Algae that contain complex chloroplasts - for example, diatoms, dinoflagellates and coccolithophorids - are amongst the most prolific photosynthesizers on earth, and study of nucleomorphs may offer important insights into how these algae acquired their chloroplasts and then went on ultimately to lose their nucleomorphs. Projects to sequence the nucleomorph genomes of cryptomonads and chlorarachniophytes therefore began several years ago [8], and the cryptomonad nucleomorph genome sequence was recently completed [1].

### A lean, mean, gene machine

At a scant 551 kilobases (kb), the nucleomorph genome of *G. theta* is one of the smallest and most gene-dense genomes known (1 gene per 977 base-pairs) [1]. Only the partially sequenced nucleomorph genome of an undescribed chlorarachniophyte species (CCMP621) is smaller, at 380 kb, and it is almost as gene-dense (1 gene per 1,141 base-pairs [9]). Studies in higher plants have shown that the vast



**Figure 1**  
 Diagrams of cryptomonad and chlorarachniophyte cells and their nucleomorph (Nm) genomes. Cryptomonad and chlorarachniophyte cells (left) contain red and green algal endosymbionts, respectively, the reduced nuclei (nucleomorphs) of which encode three repeat-capped chromosomes (right). In cryptomonads, starch is stored within the endosymbiont and often sheaths the chloroplast's own pyrenoid (P), a specialized region of the chloroplast. In chlorarachniophytes, carbohydrate storage has moved from a chloroplast starch-based system to one based on  $\beta$ 1,3 glucan within the host cell.

majority of proteins required by the chloroplast are nucleus-encoded and are post-translationally targeted to the organelle after synthesis in the cytoplasm [10]. It follows, then, that the nucleomorph, once the nucleus of an alga,

should be enriched for chloroplast genes. But sequencing of the *G. theta* nucleomorph has dispelled this assumption, because the genome encodes a mere 30 chloroplast proteins, even fewer than are in the chloroplast genome [1,11]. Most

chloroplast genes appear to have been transferred to the nucleus of the host cell, and their proteins are targeted back to the chloroplast via the ER [12].

Most of *G. theta*'s nucleomorph is occupied by a surprisingly diverse set of 'genetic housekeeping' genes [1]. These genes perform functions such as transcription, transcript processing (splicing, capping and polyadenylation), translation, protein folding and degradation, cell-cycle control, mitosis and some aspects of DNA packaging, synthesis and repair [1]. Some genetic housekeeping proteins expected to be present are in fact absent from the nucleomorph gene list, for example DNA polymerases and some ribosomal proteins. These are probably provided by the host cell, so although the endosymbiont is 'conceptually equivalent to a cell', it is not capable of living autonomously [1].

Of the 30 chloroplast proteins made by the nucleomorph, only two appear to be part of the photosynthetic apparatus [1]. Perhaps the tiny nucleomorph/periplastic compartment cannot synthesize chloroplast proteins in sufficient quantities to equip the organelle, or the photosynthetic genes are not tolerant of the elevated mutational environment in the nucleomorph. Preliminary data from the unfinished chlorarachniophyte nucleomorph genome indicate that it too encodes very few chloroplast genes [13]. Apart from chloroplast genes, very few other genes are present in the *G. theta* nucleomorph that encode functions directly useful to the rest of the cell [1]. These so-called 'end product' functions [14] include starch metabolism and ion transport as well as chloroplast functions [1].

In red algae starch is stored in the cytoplasm, and in the cryptomonad endosymbiont it is stored in an equivalent compartment, the periplastidal space (Figure 1). In chlorarachniophytes, carbohydrate storage has been transferred from the chloroplast to the host cell and changed from starch to a  $\beta$ 1,3 glucan (Figure 1) [15]. It will be interesting to determine whether the transfer of carbohydrate metabolism from endosymbiont to host and the concomitant loss of genes for carbohydrate metabolism in chlorarachniophytes contributes to their nucleomorph genomes being smaller than those of cryptomonads. The function of nucleomorphs appears to be to produce a relatively small number of proteins vital for chloroplast and other cellular functions. But to produce these essential proteins nucleomorphs need a simplified gene-expression, DNA-replication and mitotic apparatus encoded by hundreds of additional genetic housekeeping genes.

### Converging genomes

Interestingly, the genomes of both cryptomonad and chlorarachniophyte nucleomorphs contain just three tiny chromosomes [13]. Douglas *et al.* [1] provide the first credible explanation for why both genomes, despite traveling independent evolutionary pathways, have arrived at similar endpoints.

To start with, both genomes are similar in size because they contain a common core of genetic housekeeping genes. Some of these encode histones ([1] and my unpublished observations), so it is likely that the DNA is wound around histone cores. This would make the length of each chromosome about the width of a nucleomorph (1.5  $\mu$ m), which means that each chromosome is just about the right size to be segregated by a mitotic spindle during mitosis. If there were fewer than three chromosomes they would be too long to be segregated within the confines of the nucleomorph, unless they were condensed into higher-order structures (which they appear not to be) [16]; if there were more than three chromosomes, each might be too small to be viable [17].

Nucleomorph chromosomes probably have centromeres, given that they encode centromere proteins [1]. It has been suggested that gene-free regions in the centres of each cryptomonad chromosome might comprise centromere DNA [1], although these regions lack any of the common repetitive motifs typical of centromeres [18]. In the *G. theta* nucleomorph all three tubulins ( $\alpha$ ,  $\beta$  and  $\gamma$ ) are present, as well as centrosomal proteins, suggesting that the chromosomes may be segregated by a rudimentary spindle during mitosis [1,12].

The ends of nucleomorph chromosomes are also remarkably similar in both cryptomonads and chlorarachniophytes [5,19]. Each chromosome is capped with inverted repeats each comprising a single ribosomal DNA gene unit (rDNA; made up of 18S, 5.8S and 23S genes) linked to a telomere (Figure 1). The rDNA is in the opposite orientation relative to the telomeres in each type of nucleomorph. Presumably, recombination between chromosome ends and resultant gene convergence keeps the repeats of each chromosome identical to those of the other chromosomes [20]. Very few open reading frames reside in the inverted repeats, and it has been suggested that DNA replication origins may be located here as well [1].

### Compact genomes and pygmy introns

Perhaps one of the most striking features of nucleomorphs is the extraordinary compactness of their coding sequences. Spacers between genes are extremely short, or in many cases entirely absent [1,19]. The coding sequences of some genes, encoded on opposite strands, overlap by as much as 76 base-pairs in the cryptomonad [1]. Such is the proximity of many genes that they must use cryptic regulatory sequences within the coding regions of neighboring genes. Chlorarachniophyte nucleomorph genes have a surfeit of pygmy-sized introns, ranging from 18 to 20 nucleotides in length, with 19 being the most common [19]. These introns still have standard GT-AG sequences at their borders, as do the introns in the cryptomonad nucleomorph. Cryptomonad introns, at 42-55 nucleotides, are somewhat larger than chlorarachniophyte introns, and the intron density is strikingly different between the two types of nucleomorph genome [1].

Chlorarachniophyte intron density, at 3.3 per kilobase of coding sequence, is only slightly less than average for eukaryotes (3.7 introns per kilobase) [21]. Genes of the cryptomonad nucleomorph, on the other hand, are intron-poor. The entire genome contains only 17 introns [22], which is fewer than the number of spliceosomal components required to remove them. The difference in intron density between the two types of nucleomorph probably reflects the intron density of their respective algal ancestors, indicating that once acquired introns are not easily lost, despite the efficacy with which other nucleomorph DNA has been removed during evolution [1,13].

### Big lessons from the smallest genomes

Compared to the genomes of nucleomorphs, 91% of which are coding sequences [23], the genomes of most other eukaryotes, such as humans (1.5% coding sequence [24]), are positively flabby. The question of why eukaryotic genomes vary in size over many orders of magnitude (over 200,000-fold) is known as the C-value paradox (or C-enigma), and has not been settled yet (see [25] for recent discussion). One rule holds true, however: the genome size of an organism is proportional to its cell size, so big cells have big nuclei and hence big genomes. But this rule breaks down when applied to nucleomorphs. The genomes of these nuclei do not scale with cell size, although their host cell's nuclei do [14,23]. The factors that determine genome size in most cells have therefore apparently been suspended for nucleomorphs, and it has been proposed instead that there is strong selection for the removal of non-coding DNA [14,23]. Recently, however, it has been suggested (D. Petrov, personal communication) that a lack of transposable element activity may have reduced the amount of non-coding DNA in nucleomorphs, rather than there having been selection-driven DNA removal.

Transposable elements comprise 45% of the human genome, and much of the remaining non-coding DNA may be derived from the decaying sequences of ancient transposable elements [24]. Transposable elements have been discovered in nearly all eukaryotic species in which their presence has been assessed, and they are probably ubiquitous [26]. Interestingly, the only nuclei in which they have not been identified are nucleomorphs [19]. One possible explanation is that nucleomorphs are reproductively isolated from each other, because sexual stages between the host cells that would allow nucleomorphs/endosymbionts from different cells to occupy a common cytoplasm and to exchange DNA might be quite rare (and such sexual stages have yet to be well characterized) [27,28]. Even if host cell sex does occur, the extra membranes surrounding nucleomorphs are likely to form a barrier to DNA exchange. Transposable elements do not flourish in the absence of DNA exchange, and an entire class of transposable element that is present in sexual relatives is absent from asexual rotifers [26]. Without DNA exchange to

introduce new transposable elements into nucleomorphs, the old ones will have been gradually inactivated by mutation. Without the DNA bulk-generating activity that results from transposable-element proliferation, small-scale deletions, which occur more frequently than insertions [22], will have removed all non-essential DNA sequences, including transposable elements, and nucleomorph genomes will therefore have become extremely compact. Another consequence of the absence of sexual DNA exchange is that those genes that do remain have often become highly divergent and are in some cases truncated [1].

It is probable that a lack of sex ultimately sealed the fate of nucleomorphs in the algae that have lost them. Nucleomorph genes encoding chloroplast or other end-product functions useful to the rest of the cell must have continuously accumulated errors that could not be corrected because recombination with fitter versions from other nucleomorphs was not possible. Nucleomorph genes may have eventually been replaced by fitter copies previously transferred to the host cell nucleus. These new proteins would have needed to acquire correct gene-regulatory information and protein-targeting addresses to be able to return to the endosymbiont from the host cell. It is possible that some nucleomorph genes were even replaced by duplicated versions of host-cell genes. Eventually, all essential end-product-function nucleomorph-encoded proteins were replaced by host-cell nuclear versions, until the nucleomorph's genetic housekeeping proteins had nothing left to make. Once this occurred, the nucleomorph would have rapidly been lost.

It is unknown whether a similar fate awaits cryptomonad and chlorarachniophyte nucleomorphs or whether these genomes have now stabilized and will not lose many more genes, rather like the genomes of chloroplasts and mitochondria. Comparative analyses of chloroplast and mitochondrial genomes have revealed a core of genes commonly encoded by organelles and rarely encoded by nuclei. It is thought that because the concentrations of these proteins have to be continuously and rapidly adjusted they are made 'on site' and are therefore encoded by the organelle [29]. Comparative analysis of chlorarachniophyte and cryptomonad nucleomorph genome sequences may yet reveal a small core of genes with essential functions, which may be unwilling to take up residence in the host cell, so stalling the nucleomorph's demise. The insights gained from the sequence of *G. theta*'s nucleomorph genome lead to great expectations of nucleomorph comparative genomics.

### References

1. Douglas S, Zauner S, Fraunholz M, Beaton M, Penny S, Deng L-T, Wu X, Reith M, Cavalier-Smith T, Maier U-G: **The highly reduced genome of an enslaved algal nucleus.** *Nature* 2001, **410**:1091-1096.
2. Douglas SE, Murphy CA, Spencer DF, Gray MW: **Cryptomonad algae are evolutionary chimaeras of two phylogenetically distinct unicellular eukaryotes.** *Nature* 1991, **350**:148-151.

3. McFadden GI, Gilson PR, Hofmann CJ, Adcock GJ, Maier U-G: **Evidence that an amoeba acquired a chloroplast by retaining part of an engulfed eukaryotic alga.** *Proc Natl Acad Sci USA* 1994, **91**:3690-3694.
4. Van de Peer Y, Rensing SA, Maier U-G, de Wachter R: **Substitution rate calibration of small subunit rRNA identifies chlorarachniophyte endosymbionts as remnants of green algae.** *Proc Natl Acad Sci USA* 1996, **93**:7732-7736.
5. Zauner S, Fraunholz M, Wastl J, Penny S, Beaton M, Cavalier-Smith T, Maier U-G, Douglas SE: **Chloroplast protein and centrosomal genes, a tRNA intron, and odd telomeres in an unusually compact eukaryotic genome, the cryptomonad nucleomorph.** *Proc Natl Acad Sci USA* 2000, **97**:200-205.
6. Cavalier-Smith T: **Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryotic family tree.** *J Euk Microbiol* 1999, **47**:347-366.
7. Douglas S: **Plastid evolution: origins, diversity, trends.** *Curr Opin Genet Dev* 1998, **8**:655-661.
8. McFadden GI, Gilson PR, Douglas SE, Cavalier-Smith T, Hofmann CJ, Maier UG: **Bonsai genomics: sequencing the smallest eukaryotic genomes.** *Trends Genet* 1997, **13**:46-49.
9. Gilson PR, McFadden GI: **Jam packed genomes - a comparative analysis of nucleomorphs.** *Genetica* 2001, in press.
10. Abdallah F, Salamini F, Leister D: **A prediction of the size and evolutionary origin of the proteome of chloroplasts of *Arabidopsis*.** *Trends Plant Sci* 2000, **5**:141-142.
11. Douglas SE, Penny SL: **The plastid genome of the cryptophyte alga, *Guillardia theta*: complete sequence and conserved syntenic groups confirm its common ancestry with red algae.** *J Mol Evol* 1999, **48**:236-244.
12. Keeling PJ, Deane JA, Hink-Schauer C, Douglas SE, Maier U-G, McFadden GI: **The secondary endosymbiont of the cryptomonad *Guillardia theta* contains alpha-, beta-, and gamma-tubulin genes.** *Mol Biol Evol* 1999, **16**:1308-1313.
13. Gilson PR, Maier U-G, McFadden GI: **Size isn't everything - lessons in genetic miniaturisation from nucleomorphs.** *Curr Opin Genet Dev* 1997, **7**:800-806.
14. Cavalier-Smith T, Beaton MJ: **The skeletal function of non-genic nuclear DNA: new evidence from ancient cell chimeras.** *Genetica* 1999, **106**:3-13.
15. McFadden GI, Gilson PR, Sims IM: **Preliminary characterization of carbohydrate stores from chlorarachniophytes (division: Chlorarachniophyta).** *Phycol Res* 1997, **45**:145-151.
16. Gilson PR, McFadden GI: **A grin without a cat.** *Nature* 2001, **410**:1040-1041.
17. Murray A, Szostak J: **Chromosome segregation in mitosis and meiosis.** *Annu Rev Cell Biol* 1985, **1**:289-315.
18. Heslop-Harrison JS, Murata M, Ogura Y, Schwarzacher T, Motoyoshi F: **Polymorphisms and genomic organization of repetitive DNA from centromeric regions of *Arabidopsis* chromosomes.** *Plant Cell* 1999, **11**:31-42.
19. Gilson PR, McFadden GI: **The miniaturized nuclear genome of a eukaryotic endosymbiont contains genes that overlap, genes that are cotranscribed, and smallest known spliceosomal introns.** *Proc Natl Acad Sci USA* 1996, **93**:7737-7742.
20. Gilson PR, McFadden GI: **The chlorarachniophyte: a cell with two different nuclei and two different telomeres.** *Chromosoma* 1995, **103**:635-641.
21. Deutsch M, Long M: **Intron-exon structures of eukaryotic model organisms.** *Nucleic Acids Res* 1999, **27**:3219-3228.
22. Duret L: **Why do genes have introns? Recombination might add a new piece to the puzzle.** *Trends Genet* 2001, **17**:172-175.
23. Beaton MJ, Cavalier-Smith T: **Eukaryotic non-coding DNA is functional: evidence from the differential scaling of cryptomonad genomes.** *Proc R Soc Lond B Biol Sci* 1999, **266**:2053-2059.
24. International Human Genome Sequencing Consortium: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409**:860-921.
25. Gregory TR: **Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma.** *Biol Rev* 2001, **76**:65-101.
26. Arkhipova I, Meselson M: **Transposable elements in sexual and asexual taxa.** *Proc Natl Acad Sci USA* 2000, **97**:14473-14477.
27. Grell KG: **Indications of sexual reproduction in the plasmodial protist *Chlorarachnion reptans* Geitler.** *Z Naturforsch* 1990, **45c**:112-114.
28. Hill DRA, Wetherbee R: ***Proteomonas sulcata* gen. et sp. nov. (Cryptophyceae), a cryptomonad with two morphologically distinct and alternating forms.** *Phycologia* 1986, **25**:521-543.
29. Race HL, Herrmann RG, Martin W: **Why have organelles retained genomes?** *Trends Genet* 1999, **15**:364-370.