subsequent direct sequence assays were performed from each purified DNA extract. The mtDNA sequences derived from each bone were therefore multiply verified through independent samples, extractions, amplifications and sequence determinations. In all cases for each bone in the panel, the replicated mtDNA sequences were consistent across all samples and extractions. The veracity and integrity of the aurochs sequence haplogroup is strongly supported by the congruity of the bones analysed in Dublin (CHWF, TP65, CPC98 and NORF) with those previously examined in Oxford (D740 and D812).

Analysis of mitochondrial sequence data

We aligned sequences by eye. Data from previous studies[5–7,17] were also included and novel sequences were submitted to GenBank (accession numbers AF336383–AF3366748). We constructed a neighbour-joining phylogeny using uncorrected distances with gapped positions excluded and the Neighbour program in the Phylip package[19]. Arlequin 2.0 (ref. 20) was used to estimate nucleotide diversity values for each population, to estimate parameters and goodness of fit from mismatch distributions and to compute Fu's *Fs* test of selective neutrality[10]. Phylogenetic analysis was also performed using reduced median networks, constructed manually according to ref. 21. Sites prone to hypermutablity were identified using a network of zebu and taurine samples examined in ref. 6. For the reduced median networks shown, most of the reductions (71%) were at these sites. The mutation rate for the 240-bp region of the cattle D-loop was calculated using the estimate of the transition/transversion ratio for the sequence data presented here (61/1). This agrees well with previous estimates on smaller data sets 57/1 (ref. 6) and 41/1 (ref. 7), and as the new rate incorporates data presented in those papers it is used in preference to either. The *Bison–Bos* divergence is considered to be around 1 Myr. The three transversions observed in the 240-bp region between the *Bison–Bos* groups constitute the equivalent of 183 transitions. The one-lineage rate was estimated as 38% per Myr or one substitution per 10,928 yr. The central 95% credible region for the expansion time was calculated for each of the four main cattle haplogroups using the program CRED[22].

Received 10 August 2000; accepted 19 January 2001.

1. Smith, B. D. *The Emergence of Agriculture* (Scientific American Library, New York, 1995).
2. Harris, D. R. in *The Origins and Spread of Agriculture and Pastoralism in Eurasia* (ed. Harris, D. R.) (UCL, London, 1996).
3. Epstein, H. *The Origin of the Domestic Animals of Africa* (Africana, New York, 1971).
4. Bökönyi, S. *The History of Domestic Animals in Central and Eastern Europe* (Akadémiai Kiadó, Budapest, 1974).
5. Loftus, R. T., MacHugh, D. E., Bradley, D. G., Sharp, P. M. & Cunningham, P. Evidence for two independent domestications of cattle. *Proc. Natl Acad. Sci. USA* **91,** 2757–2761 (1994).
6. Bradley, D. G., MacHugh, D. E., Cunningham, P. & Loftus, R. T. Mitochondrial diversity and the origins of African and European cattle. *Proc. Natl Acad. Sci. USA* **93,** 5131–5135 (1996).
7. Bailey, J. F. *et al*. Ancient DNA suggests a recent expansion of European cattle from a diverse wild progenitor species. *Proc. R. Soc. Lond. B* **263,** 1467–1473 (1996).
8. Richards, M. B., Macaulay, V. A., Bandelt, H.-J. & Sykes, B. C. Phylogeography of mitochondrial DNA in western Europe. *Ann. Hum. Genet.* **62,** 241–260 (1998).
9. Rogers, A. R. & Harpending, H. Population growth makes waves in the distribution of pairwise genetic differences. *Mol. Biol. Evol.* **9,** 552–569 (1992).
10. Fu, Y.-X. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147,** 915–925 (1997).
11. Excoffier, L. & Schneider, S. Why hunter-gatherer populations do not show signs of Pleistocene demographic expansions. *Proc. Natl Acad. Sci. USA* **96,** 10597–10602 (1999).
12. Cymbron, T., Loftus, R. T., Malheiro, M. I. & Bradley, D. G. Mitochondrial sequence variation suggests an African influence in Portuguese cattle. *Proc. R. Soc. Lond. B* **266,** 597–603 (1999).
13. Troy, C. S. *Mitochondrial DNA Phylogeny and Biogeography of Domestic Cattle.* Ph.D Thesis, Univ.-Dublin (1998).
14. Macaulay, V. *et al*. The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am. J. Hum. Genet.* **64,** 232–249 (1999).
15. Wendorf, F. & Schild, R. Nabta Playa and its role in northeastern African prehistory. *J. Anthropol. Archaeol.* **17,** 97–123 (1998).
16. Yang, D. Y., Eng, B., Waye, J. S., Dudar, J. C. & Saunders, S. R. Technical note: improved DNA extraction from ancient bones using silica-based spin columns. *Am. J. Phys. Anthropol.* **105,** 539–543 (1998).
17. MacHugh, D. E. *et al*. Early mediaeval cattle remains from a Scandinavian settlement in Dublin: genetic analysis and comparison with extant breeds. *Phil. Trans. R. Soc. Lond. B* **354,** 99–109 (1999).
18. Anderson, S. *et al*. Complete sequence of bovine mitochondrial DNA. *J. Mol. Biol.* **156,** 683–717 (1982).
19. Felsenstein, J. PHYLIP: Phylogeny Inference Package. (University of Washington, Seattle, 1993).
20. Schneider, S., Roessli, D. & Excoffier, L. *ARLEQUIN Ver. 2.0: a Software for Population Genetics Data Analysis* (Department of Anthropology, University of Geneva, Switzerland, 1999).
21. Bandelt, H. J., Forster, P., Sykes, B. C. & Richards, M. B. Mitochondrial portraits of human populations using median networks. *Genetics* **141,** 743–753 (1995).
22. Macaulay, V. A. CRED: Credible Regions for Coalescence Times. (Univ. Oxford, 1998).

Acknowledgements

We thank C. Hawkes, O. Ertugrul, A. H. Al Haboby, A. H. Harba, M. A. A. El-Barody, E. Thompson, T. Goodchild, H. Halila, A. Swaid, G. Guneren, B. Tekbas, M. Bruford, B. Sauveroche, G. Kana, D. Achu-Kwi, M. Diallo, L. Gnaho, K. Papadopolous, A. G. Georgoudis, C. Gaillard, O. Hanotte, E. Rege, the Nordic GeneBank and C. Hawes for assistance or provision of samples. We also thank the Highland Cattle Society and the Black and Belted Galloway Societies for sample information. This work was partly funded by a European Commission contract. D.E.M. was supported by a Wellcome Trust Fellowship in Bioarchaeology. J.F.B. is a European Commission Marie Curie Fellow. Radiocarbon dating at the Oxford Radiocarbon Accelerator Unit was funded by NERC.

Correspondence and requests for materials should be addressed to D.G.B. (e-mail: dbradley@tcd.ie).

The highly reduced genome of an enslaved algal nucleus

Susan Douglas*, Stefan Zauner†, Martin Fraunholz†‡, Margaret Beaton‡§, Susanne Penny*, Lang-Tuo Deng§, Xiaonan Wu§, Michael Reith*, Thomas Cavalier-Smith‡§ & Uwe-G Maier†

* *National Research Council of Canada Institute for Marine Biosciences and Program in Evolutionary Biology, Canadian Institute of Advanced Research, 1411 Oxford Street, Halifax, Nova Scotia B3H 3ZI, Canada*

† *Cell Biology and Applied Botany, Philipps-University Marburg, Karl-von-Frisch-Strasse, D-35032 Marburg, Germany*

§ *Program in Evolutionary Biology, Canadian Institute of Advanced Research, Department of Botany, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada*

Chromophyte algae differ fundamentally from plants in possessing chloroplasts that contain chlorophyll *c* and that have a more complex bounding-membrane topology[1]. Although chromophytes are known to be evolutionary chimaeras of a red alga and a non-photosynthetic host[1], which gave rise to their exceptional membrane complexity, their cell biology is poorly understood. Cryptomonads are the only chromophytes that still retain the enslaved red algal nucleus as a minute nucleomorph[2–4]. Here we report complete sequences for all three nucleomorph chromosomes from the cryptomonad *Guillardia theta*. This tiny 551-kilobase eukaryotic genome is the most gene-dense known, with only 17 diminutive spliceosomal introns and 44 overlapping genes. Marked evolutionary compaction hundreds of millions of years ago[1,4,5] eliminated nearly all the nucleomorph genes for metabolic functions, but left 30 for chloroplast-located proteins. To allow expression of these proteins, nucleomorphs retain hundreds of genetic-housekeeping genes[5]. Nucleomorph DNA replication and periplastid protein synthesis require the import of many nuclear gene products across endoplasmic reticulum and periplastid membranes. The chromosomes have centromeres, but possibly only one loop domain, offering a means for studying eukaryotic chromosome replication, segregation and evolution.

Soon after the symbiogenetic origin of chloroplasts from cyanobacteria[1] to form the common ancestor of green plants, red and glaucophyte algae (kingdom Plantae[6,7]), even more complex eukaryotic cells arose by secondary symbiogenesis[1,3,4] (Fig. 1). Such chimaeric integration of two evolutionarily distant eukaryotic cells occurred independently in the common ancestor of cryptomonads and other chromophytes, in which the endosymbiont was a red alga, and in chlorarachneans, which acquired a green alga[1,3,4]. In both cryptomonads and chlorarachneans, a flagellate host contributed the nucleus, endomembranes and mitochondria to the chimaera, whereas the photosynthetic endosymbiont provided its chloroplast, plasma membrane (the periplastid membrane[1,3,4]) and a second nucleus (the nucleomorph), which became miniaturized[3–5]. The nucleomorph of both groups kept its envelope, nuclear pores[8] and three minute chromosomes[9]. In the ancestor of cryptomonads and chromobiotes (treated as kingdom Chromista[6,8]) but not chlorarachneans, the former food vacuole membrane originally enclosing the enslaved endosymbiont apparently fused with the nuclear envelope[10], placing it in the rough endoplasmic reticulum[8,10] (RER; Fig. 1). Cryptomonad cells depend on four genomes, each encoding distinct protein synthesis machineries in discrete

‡ Present addresses: Goddard Laboratories, University of Philadelphia, 415 South University Avenue, Philadelphia, Pennsylvannia 19104-6018, USA (M.F.); Biology Department, 63B York Street, Mount Allison University, Sackville, NB, Canada E4L 1G7 (M.B.); Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK (T.C.-S.).

© 2001 Macmillan Magazines Ltd

compartments, between which proteins are translocated. Until now, understanding how these genomes cooperate has been limited by the availability of only partial sequences for nucleomorph genomes[11,12].

We report the first complete genome sequence of a nucleomorph (551,264 base pairs (bp)), which proves conclusively that it is a vestigial nucleus[13] and that cryptomonads comprise one eukaryotic cell nested in another. Gene density is extremely high (1 gene per 977 bp) and non-coding regions are ultrashort, with only one pseudogene φ*rpl24* (Fig. 2). The six chromosome ends are identical repeats, comprising telomeres ($([AG]_7AAG_6A)_{11}$, 5S and 28S/5.8S/18S ribosomal RNA genes and five open reading frames (ORFs); the ubiquitin conjugation enzyme gene is repeated at five ends, and the TATA-box binding protein gene (*tfIId*) at three ends. Except in these repeats and five central 200–500-bp regions (possibly centromeric), intergenic spacers have few, if any, nucleotides. Forty-four genes overlap by up to seventy-six nucleotides (Fig. 2). One gene has three copies, one on chromosome 1 (ORF 160a) and two on chromosome 3 (ORF 160b,c); the only other repeated genes are in the termini. Coding regions of some genes are shorter than their homologues in other organisms.

Only 17 protein-coding genes contain spliceosomal introns (42–52 bp long; see Supplementary Information), all located in the 5′ region, as in yeast[14], many immediately after the initiator AUG. Eleven are in ribosomal protein genes, as in yeast where their splicing negatively regulates messenger RNA levels[14]. Like the even shorter pygmy introns of chlorarachnean nucleomorphs[12], they have standard GT/AG boundaries. Twelve transfer RNA genes have protein-spliced introns. The marked contrast between the effective elimination of non-coding DNA from cryptomonad nucleomorphs and the accumulation of vast amounts of non-coding DNA in coexisting cryptomonad nuclei indicates that nuclear non-coding DNA in general is functional and positively selected[5], and is not purely selfish or junk. Chromosomal A+T content varies, suggesting that there have been three different mutational and selective pressures on base composition: on the terminal repeats (45.5% G+C); on housekeeping genes with a very low G+C content (23%); and on transfer RNA genes and genes for plastid proteins with intermediate G+C content (35%).

The function of 219 of the 464 putative protein-coding genes is unknown, but 31 have convincing database matches—11 to cyanobacteria and 20 to eukaryotes. Retention of the latter, even in this exceptionally compacted genome, shows that they must have important functions in all eukaryotes. The other 245 genes have homologues of known function, mostly for chromosome reproduction or gene expression, with very few for cytoplasmic functions such as protein assembly and degradation, signal transduction/regulation, cell-cycle control and membrane transport (Fig. 3). Only one gene for metabolism (carotenoid synthesis) was found. There are 47 different genes for non-mRNAs (rRNA, tRNA, small nuclear RNA, small nucleolar RNA).

Three conclusions can be drawn about intergenome cooperation in these complex eukaryote/eukaryote chimaeras. First, most identified genes ($>$ 250/302) are needed simply for self-perpetuation of the nucleomorph and its periplastid ribosomes. End products directly useful to the rest of the cell are encoded by only a few genes including 30 chloroplast proteins, 3 transporters (Sut1 for sulphate; Kea1 for potassium; Rli1, an ABC transporter), an anabolic enzyme (Ggt), and a few regulatory enzymes. Second, even fewer cryptomonad plastid proteins are encoded by the nucleomorph than by the plastid genome[15], so at least a thousand more[1] must be imported into the plastid across four membranes, not two membranes as for nucleomorph-encoded ones. Third, as certain well-conserved genes essential for nucleomorph functions are absent from the nucleomorph genome, these functions must be provided by nuclear genes and imported into the periplastid compartment.

It was not known previously that nuclear gene products not destined for the chloroplast were imported into the periplastid space[4]. As in mitochondria and chloroplasts, DNA polymerase genes are absent. DNA polymerases must be nuclear encoded and imported across the RER and periplastid membranes, and onwards into the nucleomorphs through their nuclear pores. As we identified so few genes for transporters, most must be encoded by nuclear genes or nucleomorph ORFs. As chromobiotes lost the nucleomorph, they are likely to have homologous transport proteins in their nuclei. If chromists are sisters of alveolate protozoa, sharing a photosynthetic common ancestor[1,4,16], some may be also present in Sporozoa, such as malaria parasites, and important for their periplastid membranes and as potential drug targets.

Under 10% of the genes encode end-product functions[5] that are useful to the rest of the cell. Originally three end-product functions

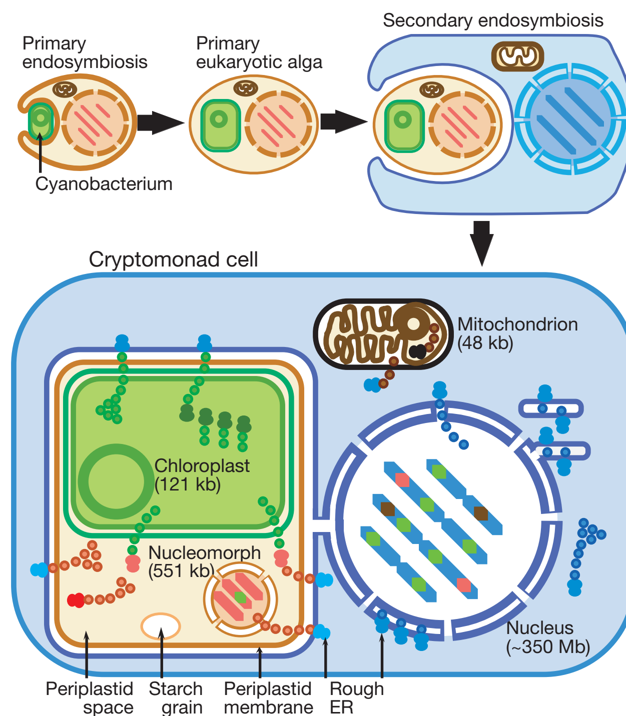

Figure 1 Secondary symbiogenetic origin and membrane topology of cryptomonads. After the primary endosymbiotic incorporation of a cyanobacterium to form the first chloroplast (green), many of its genes were transferred into the host nucleus. After this ancestral plant diversified to form green plants, red algae and glaucophytes, more complex algae were formed by independent secondary symbioses involving green or red algae, after which many or all plastid protein genes were transferred from the algal to the secondary host's nucleus. Shown here is the symbiosis of a red alga to form cryptomonads, where (as in all chromists) the food vacuole membrane fused with the RER. This fusion did not occur in alveolates, chlorarachneans or euglenoids. Former cyanobacterial genes now inserted in nucleomorph or nuclear chromosomes are shown in green, and former red algal genes now in the host nucleus in red. In cryptomonads, the chloroplast and nucleomorph (former red algal nucleus) are topologically in the periplastid space (starch- and ribosome-containing residual cytoplasm of the former red algal cell, yellow) in the periplastid membrane (former red algal plasma membrane), which is located in the lumen of the host's rough endoplasmic reticulum (RER). Chloroplast proteins are coded by three genomes (chloroplast, nucleomorph and nucleus) and mitochondrial proteins by two genomes (mitochondrion and nucleus). Nucleomorph and periplastid proteins are coded by two genomes (nucleus and nucleomorph). Coloured dots indicate protein translocation pathways in the periplastid complex: nuclear- or nucleomorph-encoded proteins targeted to the chloroplast are green; nuclear-encoded proteins imported into the periplastid space, and both nuclear- and nucleomorph-encoded proteins imported into the nucleomorph, are red. Mitochondrial proteins (brown) are encoded by both nuclear and mitochondrial genomes.

© 2001 Macmillan Magazines Ltd

Figure 2 The cryptomonad nucleomorph chromosomes showing gene locations. Colours show the functional categories of the genes specified in Fig. 3. The number of overlapping nucleotides between adjacent genes is indicated beside the brackets. Each chromosome is displayed as if broken into two fragments near its midpoint. Labels on the left of each chromosome indicate genes transcribed towards its left end (top left); labels on the right show those transcribed towards the right end (bottom right).

© 2001 Macmillan Magazines Ltd

were envisaged: plastid proteins not made in the chloroplast itself or in the host cytosol; periplastid space functions in starch metabolism; and periplastid membrane transporters mediating metabolite exchange between host and symbiont. We identified 30 genes encoding proteins for chloroplast function, including two known proteins[11], but very few for periplastid functions.

Sequence comparisons show that two nucleomorph genes for plastid proteins are eukaryotic inventions (Iap100, Met), one (CbbX) is α-proteobacterial[17], and the others are cyanobacterial. Only two encode molecules directly needed for photosynthesis: an electron transfer molecule (rubredoxin)[11] and a carotene-binding protein (Hlip). Some genes are needed for chloroplast division (FtsZ[11]) or gene expression (ribosomal protein Rps15, products for DNA and RNA metabolism, and a factor for RuBisCo expression, CbbX). Others encode components of the chloroplast protein-import machinery (Iap100, Tic22), translocation into the thylakoid lumen (Tha4, SecE), chaperonins (Cpn60, Hcf136) and a protease (ClpP).

All are essential for chloroplast function, explaining why the nucleomorph has persisted for hundreds of millions of years. Compared with their cyanobacterial homologues, the nucleomorph-encoded plastid proteins have amino-terminal extensions (transit peptides; see Supplementary Information) to specify import into the chloroplast, although only those of Hcf136, Cpn60, Tic22 and GyrB were recognized by the search tool ChloroP (http://www.cbs.dtu.dk/services/ChloroP/). Nevertheless, *in vitro* the N-terminal extension of rubredoxin acts as a transit peptide with pea or cryptomonad plastids and is removed after import[18].

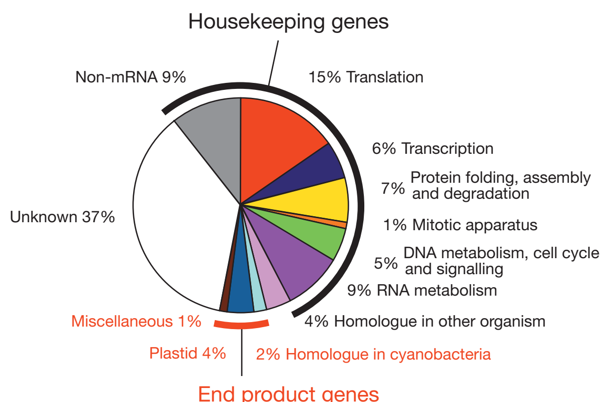

Figure 3 The fraction of genes in each functional category and their classification as end-product or genetic housekeeping genes. The genes in each category are listed in the Supplementary Information.

Acquiring solar-powered carbohydrate synthesis is the main advantage of enslaving a photosynthetic cell. As glucose produced by degrading periplastid starch is used mainly in the host cytosol, starch metabolism must be responsive to its needs. If the phosphatidylinositol-4-OH kinase (PI(4)K) is associated with the periplastid membrane, it may mediate such signalling through the second messenger inositol trisphosphate. In response, nucleomorph-encoded GTP-binding proteins (GbIp, GTPbp) may regulate periplastid starch metabolism and protein synthesis, as may a protein kinase similar to glycogen synthase kinase; this AMP-activated kinase (AakB) might control starch degradation (like its homologue in animal glycogen degradation). The nucleomorph has four other kinase genes (gs, mps1, snf1, snf2). Protein phosphatase, Pp1, may function complementarily in starch regulation and/or kinase-regulated cell-cycle controls. The presence of a geranylgeranyl-transferase (Ggt) lacking a transit peptide indicates that at least one intermediate stage in carotenoid synthesis may occur in the periplastid space. As plastids typically make carotenoids using isoprenoid precursors from the host, intermediates presumably traverse the periplastid space.

The red algal symbiont lost mitochondria, peroxisomes, lysosomes, Golgi membranes and most metabolic enzymes. Possibly it also lost all ER lumenal proteins and protein glycosylation, perhaps facilitating an unprecedented simplification of eukaryotic membrane-protein targeting. This red algal vestige may differ from true eukaryotic and bacterial cells by lacking co-translational protein insertion into membranes by the signal recognition particle (SRP): the nucleomorph outer membrane, phylogenetically equivalent to RER[8], uniquely lacks obvious surface ribosomes[13,19], and we found no 7S SRP RNA or other genes for SRPs. Perhaps nucleomorph envelope and periplastid membrane proteins are inserted post-translationally, as in mitochondria and plastids[1] (which analogously both lost the ancestral eubacterial SRP RNA), using the nucleomorph-encoded Hsp70 chaperone.

Genes for DNA polymerase are absent, but there are genes for its sliding clamp (Pcna), a replication co-factor (RfC) for primer extension, and a RecA-like protein (Rad51). Replicon origins might lie in the terminal repeats proximal to the 18S rRNA genes, where there is more non-coding DNA than elsewhere; if there are no others, each chromosome will consist of just two 80–100-kilobase (kb) replicons. As eukaryotic replicons average 100 kb (ref. 20), no origins need lie in unique chromosome regions where intergenic spacers are almost all shorter than the minimal 250 bases needed for origins.

Genes for three core histones (H2b, H3, H4) are present. H2b is

Table 1 Codon usage, introns and chromosomal location of tRNA genes

Codon	Chr*	%	Codon	Chr*	%	Codon	Chr*	%	Codon	Chr*	%
TTC F	2†	21.4	TCC S		5.6	TAC Y	1†	20.2	TGC C	3†	24.6
TTT F		78.6	TCT S	2†	34.1	TAT Y		79.8	TGT C	2†	75.4
TTA L	3†	50.8	TCA S		34.1	TAA Z		60.0	TGA Z		20.0
TTG L	2†	17.6	TCG S	1	5.6	TAG Z		20.0	TGG W	3†	100
CTC L		1.9	CCC P		12.6	CAC H	2	18.4	CGC R		0.8
CTT L	3	14.9	CCT P	3	32.3	CAT H		81.6	CGT R	3	2.4
CTA L	1	12.2	CCA P		48.3	CAA Q	1	78.2	CGA R	2	8.1
CTG L		2.7	CCG P	1	6.9	CAG Q	2	21.8	CGG R		2.4
ATC I		12.9	ACC T		10.6	AAC N	3	23.6	AGC S	2	6.0
ATT I	1	48.6	ACT T	2	34.1	AAT N		76.4	AGT S		14.5
ATA I	2	38.5	ACA T	3	49.2	AAA K	2	87.1	AGA R	1†	77.4
ATG M_i	3	100	ACG T	3	6.2	AAG K	2	12.9	AGG R	3†	8.9
ATG M_e	3†	100									
GTC V		7.5	GCC A		9.1	GAC D	1	18.6	GGC G	3	9.2
GTT V	2†	48.4	GCT A		32.6	GAT D		81.4	GGT G		36.2
GTA V	1	34.4	GCA A	3	50.8	GAA E		84.1	GGA G	3	48.5
GTG V		9.7	GCG A		7.6	GAG E		15.9	GGG G		6.1

M_e, elongator methionine; M_i, initiatior methionine.
* Chromosome on which tRNA gene is located.
† tRNA gene contains non-spliceosomal intron.

© 2001 Macmillan Magazines Ltd

exceptionally divergent with its normally acetylated basic N-terminal tail mostly deleted, so nucleosomes may be simplified compared with those of typical nuclei. Histones probably undergo acetylation, as genes for a histone acetyltransferase (Hat) and deacetylase (Hda) are found, and all H3 and H4 acetylation sites are conserved. H3 and H4 genes are adjacent, divergently transcribed and relatively well-conserved, but have evolved more rapidly than nuclear homologues (like most nucleomorph genes). Although H2a and H1 are undetectable, neither is likely to be dispensable, as a string of nucleosomes, unlike a 30-nm filament requiring H1, would be too long for segregation without them.

One gene each for α-, β-, and γ-tubulin is present[11], and in addition to γ-tubulin three other centrosomal proteins[21] are nucleomorph-encoded—Ranbpm[11]; Hsp82 (in the cytosolic Hsp90 family) and Hsp 70 (cytosolic family). As centrosome activation involves the cyclinB–cdc2 complex[22] (both nucleomorph encoded), mitotic control is present. Nuclear localization signals on Hsp82 and Hsp70 suggest that the centrosome is inside the nucleomorph, consistent with the intact envelope during division[19] and the intranuclear red algal spindles. Presence of a centromere-specific H3-like histone (Cenp-A) suggests that nucleomorph chromosomes have centromeres, confirming that nucleomorphs have a relict mitotic apparatus[11] despite the lack of ultrastructural evidence for a spindle[19].

We cannot be certain of centromere positions. Only the terminal repeats have significant non-coding DNA regions repeated on all chromosomes, but we doubt that they are centromeric because two per chromosome would upset segregation. Assuming nucleomorph centromere sizes to be similar to *Saccharomyces cerevisiae*[23], chromosome 1 has two non-genic regions long enough to be centromeric (520 nucleotides between ORFs 147 and 471; 540 nucleotides between ORFs 244 and 446); chromosome 2 has one (~200 nucleotides between ORFs 80a and 231); chromosome 3 has two (374 nucleotides between ORFs 141 and 180; ~300 nucleotides between ORF 62 and Ebi). As these candidate centromere regions are near chromosome midpoints and other intergenic regions seem too short, all three chromosomes may be metacentric.

Eukaryotic chromosomes are linear arrays of looped domains attached to a nuclear skeleton by their replicon origins[24]. Each nucleomorph chromosome may constitute a single loop domain, being similar in size. If nucleomorph histones compact the DNA into typical 30-nm chromatin threads, the longest chromosome would be 1.5 μm in the absence of higher-order condensation; if metacentric, each chromosome arm would be 0.75 μm. As nucleomorphs are about 1.5-μm across and longer when dividing[19], extra folding would be unneccessary for mitotic segregation, unlike for other eukaryotic chromosomes. Compacted chromosomes cannot be seen during nucleomorph division[19]. Analogous folding constraints may explain why cryptomonad and chlorarachnean nucleomorphs retained three short chromosomes[3,9,12] instead of aggregating them into a larger one, which would be feasible only with higher-order mitotic compaction.

Genes for key cell-cycle control proteins include a replication-licensing protein (Mcm2), which mediates the G1/S-phase transition[22], and a cyclin-dependent Cdc2 kinase and its cyclin B, which are involved in the G2/M-phase checkpoint. The persistence of a nucleomorph-specific cell-cycle kinase system, even in this simplified cell vestige, emphasizes that it remains conceptually equivalent to a cell, despite its long integration and loss of metabolism and most genes. The nucleomorph encodes ubiquitin (fused to two ribosomal proteins), the ubiquitin-fusion-degradation enzyme (Ufd), and three E2 enzyme subunits (Ubc2, Ubc4, UbcE2). The cell-cycle regulatory role of proteasomes (degrading cyclin B to ensure exit from mitosis, and also controlling exit from S phase[25]) may be the key reason why such complex multiprotein assemblies as proteasomes and their associated ubiquitin pathway are conserved in the periplastid space, otherwise so reduced in normal cytosolic functions. Supporting this is a gene for an AAA-ATPase (Cdc48; plus a partially related gene *cdc48a*) implicated in ubiquitin-dependent mitotic events, including membrane fusion[26]. There are 21 nucleomorph genes for different subunits of the 20S core proteasome and its 19S cap. Two 20S proteasome subunits (α4 and β2) and ten 19S cap subunits are absent, and must be imported like all 11S activator subunits.

Elements of the nuclear pore-complex export/import machinery, the key nucleolar structural protein fibrillarin (Nop1), and elaborate RNA processing machinery confirm that the nucleomorph is a miniaturized nucleus. The surmise[13] that nucleomorph envelope pores are genuine nuclear pores is confirmed by well-conserved importin genes (*impA*; *imb1*) and a protein interacting with nuclear pores and needed for transport (CRM). Neither transcription nor RNA processing was radically simplified by symbiogenesis. All three RNA polymerases and several transcription factors are nucleomorph encoded, including TfIID for promoter recognition, TfIIB for binding polymerase II, and two components (Rad3; Rad25) of TfIIH. Nucleomorph messengers are probably capped and polyadenylated, for mRNA-capping enzyme (Mce), cap-binding protein (Cbp), and poly(A)-binding protein (Pab) genes exist. Many spliceosomal splicing components (including U6 snRNA and snRNPE, as in chlorarachneans[12]) and some for tRNA intron removal are present, as are 5 RNAs and 17 proteins of the nucleolar snoRNP machinery for processing rRNA (cleavage, methylation and pseudouridylation). The gene for pseudouridylate synthase is of the eukaryotic type (*cbf5*) with ancillary centromere–microtubule binding properties, but lacks several normally conserved features. The thesis that snoRNAs are ancestral for all cells but were lost in eubacteria by 'streamlining'[27] is rendered implausible by their persistence in this ultra-streamlined genome.

Thirty-seven genes encode large subunit periplastid ribosomal proteins and 28 small subunit ones, so about 14 are imported or coded by unidentified ORFs. The nucleomorph encodes 37 tRNAs (12 with introns; Table 1) and standard protein synthesis initiation and elongation factors, but only one amino-acyl-tRNA synthetase (seryl). As we found no gene for glutamyl-tRNA, it must be imported unless its codons are recognized by another tRNA with 'super-wobble'.

Cryptomonad and chlorarachnean nucleomorphs are natural experiments in genome miniaturization and cell simplification that can test basic ideas about genome and cell functions. If domain organization, folding, and mitotic segregation of nucleomorph chromosomes are indeed exceptionally simplified, they could become important models for understanding the more complex chromosomes of typical nuclei. □

Methods

Nucleomorph DNA

As nucleomorph DNA is only about 0.1% of cellular DNA, we could not obtain pure enough samples or sufficient amounts of it for a single mechanically fragmented library. We prepared several libraries slightly contaminated by DNA from the other three genomes: random libraries from the nucleomorph DNA band on CsCl density gradients[28] (with major mitochondrial DNA contamination) and others enriched in one individual nucleomorph chromosome extracted from bands excised from pulsed field gels[9]. DNA was partially digested by *Sau*3AI and cloned into phage λ as described[11], or completely digested with one of *Eco*RI, *Hind*III, *Pst*I, *Bgl*II, *Xba*I, *Cla*I, *Spe*I, *Bcl*I or *Bam*HI and cloned into plasmids (pUC18).

Sequencing and gene identification

Sequencing, editing and contig assembly were done as described[11]. Clones were sequenced fully on both strands by primer walking, and gaps between contigs filled by polymerase chain reaction of genomic DNA. Preliminary gene identification and analysis used MAGPIE[29]. tRNA and snoRNA genes were detected using the search tools tRNAscan (http://www.genetics.wustl.edu/eddy/ tRNAscan-SE) and snoscan[30]. Further gene analyses used a Perl script to identify all possible ORFs and search them against GenBank (BLAST results available at http:// reith.imb.nrc.ca/nucleomorph/nucleomorph.html). Spliceosomal introns were first found manually, and then using a Perl script that searched for conserved splice junctions and extension of the ORF by removing the putative intron.

© 2001 Macmillan Magazines Ltd

Received 5 September 2000; accepted 14 February 2001.

1. Cavalier-Smith, T. Membrane heredity and early chloroplast evolution. *Trends Plant Sci.* **5,** 174–182 (2000).
2. Douglas, S. E., Murphy, C. A., Spencer, D. F. & Gray, M. W. Cryptomonad algae are evolutionary chimaeras of two phylogenetically distinct unicellular eukaryotes. *Nature* **350,** 148–151 (1991).
3. Maier, U.-G., Douglas, S. & Cavalier-Smith, T. The nucleomorph genomes of cryptophytes and chlorarachniophytes. *Protist* **151,** 103–109 (2000).
4. Cavalier-Smith, T. Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. *J. Euk. Microbiol.* **46,** 347–366 (1999).
5. Cavalier-Smith, T. & Beaton, M. J. The skeletal function of non-genic nuclear DNA: new evidence from ancient cell chimaeras. *Genetica* **106,** 3–13 (1999).
6. Cavalier-Smith, T. A revised six-kingdom system of life. *Biol. Rev.* **73,** 203–266 (1998).
7. Moreira, D., Le Guyader, H. & Philippe, H. The origin of red algae: implications for the evolution of chloroplasts. *Nature* **405,** 69–72 (2000).
8. Cavalier-Smith, T. in *Progress in Phycological Research* Vol. 4 (eds Round, F. E. & Chapman, D. J.) 309–347 (Biopress, Bristol, 1986).
9. Eschbach, S., Hofmann, C. J. B., Maier, U.-G., Sitte, P. & Hansmann, P. A eukaryotic genome of 660 kb: electrophoretic karyotype of nucleomorph and cell nucleus of the cryptomonad alga, *Pyrenomonas salina*. *Nucleic Acids Res.* **19,** 1779–1781 (1991).
10. Whatley, J. M., John, P. & Whatley, F. R. From extracellular to intracellular: the establishment of mitochondria and chloroplasts. *Proc. Roy. Soc. Lond. B.* **204,** 165–187 (1979).
11. Zauner, S. *et al.* 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* **97,** 200–205 (2000).
12. Gilson, P. R. & McFadden, G. I. 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 (1996).
13. Ludwig, M. & Gibbs, S. P. Are the nucleomorphs of cryptomonads and *Chlorarachnion* the vestigial nuclei of eukaryotic endosymbionts? *Ann. NY Acad. Sci.* **503,** 198–211 (1987).
14. Spingola, M., Grate, L., Haussler, D. & Ares, M. Genome-wide bioinformatic and molecular analysis of introns in *Saccharomyces cerevisiae*. *RNA* **5,** 221–234 (1999).
15. Douglas, S. E. & Penny, S. L. The plastid genome of the cryptophyte alga, *Guillardia theta*: complete sequence and conserved synteny groups confirm its common ancestry with red algae. *J. Mol. Evol.* **48,** 236–244 (1999).
16. McFadden, G. I. Mergers and acquisitions: malaria and the great chloroplast heist. *Genome Biol.* **1,** 1026.1–1026.4 (2000).
17. Maier, U. G., Fraunholz, M., Zauner, S., Penny, S. & Douglas, S. A nucleomorph-encoded CbbX and the phylogeny of RuBisCo regulators. *Mol. Biol. Evol.* **17,** 576–583 (2000).
18. Wastl, J. & Maier, U. -G. Transport of proteins into cryptomonads complex plastids. *J. Biol. Chem.* **275,** 23194–23198 (2000).
19. Meyer, S. The taxonomic implications of the ultrastructure and cell division of a stigma-containing *Chroomonas* sp. (Cryptophyceae) from Rocky Bay, Natal, South Africa. *S. Afr. J. Bot.* **53,** 129–139 (1987).
20. Berezney, R., Dubey, D. D. & Huberman, J. A. Heterogeneity of eukaryotic replicons, replicon clusters, and replication foci. *Chromosoma* **108,** 471–484 (2000).
21. Lange, B. H. M., Bachi, A., Wilm, M. & Gonzalez, C. Hsp90 is a core centrosomal component and is required at different stages in the centrosome cycle in *Drosophila* and vertebrates. *EMBO J.* **19,** 1252–1262 (2000).
22. De Souza, C. P., Ellem, K. A. & Gabrielli, B. G. Centrosomal and cytoplasmic Cdc2/cyclin B1 activation precedes nuclear mitotic events. *Exp. Cell Res.* **257,** 11–21 (2000).
23. Hegemann, J. H. & Fleig, U. N. The centromere of budding yeast. *BioEssays* **15,** 451–460 (1993).
24. Paul, A. L. & Ferl, R. J. Higher-order chromatin structure: looping long molecules. *Plant Mol. Biol.* **41,** 713–720 (1999).
25. Kawahara, H. *et al.* Inhibiting proteasome activity causes overreplication of DNA and blocks entry into mitosis in sea urchin embryos. *J. Cell Sci.* **113,** 2659–2670 (2000).
26. Meyer, H. H., Shorter, J. G., Seemann, J., Pappin, D. & Warren, G. A complex of mammalian ufd1 and npl4 links the AAA-ATPase, p97, to ubiquitin and nuclear transport pathways. *EMBO J.* **19,** 2181–2192 (2000).
27. Poole, A., Jeffares, D. & Penny, D. Early evolution: prokaryotes, the new kids on the block. *BioEssays* **21,** 880–889 (1999).
28. Douglas, S. E. Physical mapping of the plastid genome from the chlorophyll c-containing alga, *Cryptomonas* ϕ. *Curr. Gen.* **14,** 591–598 (1988).
29. Gaasterland, T. & Sensen, C. W. Fully automated genome analysis that reflects user needs and preferences: A detailed introduction to the MAGPIE system architecture. *Biochimie* **78,** 302–310 (1996).
30. Lowe, T. M. & Eddy, S. R. A computational screen for methylation guide snoRNAs in yeast. *Science* **283,** 1168–1171 (1999).

Supplementary information is available on *Nature*'s World-Wide Web site (http://www.nature.com) or as paper copy from the London editorial office of *Nature*.

Acknowledgements

We are grateful to M. Johannsen for technical assistance; P. Gordon for assistance with MAGPIE; T. Lowe for helping us identify tRNAs and snoRNAs; and R. Redfield for preparing Fig. 1 and valuable comments on the manuscript. Our research was supported by the Natural Sciences and Engineering Council (NSERC), Canada, the Deutsche Forshungsgemeinschaft, Germany, and the NERC (UK). We thank the Canadian Institute for Advanced Research and NERC for fellowships for T.C.-S. and NSERC for a fellowship for M.J.B.

Correspondence and requests for materials should be addressed to T.C.-S. (e-mail: tom.cavalier-smith@zoo.ox.ac.uk). Sequences have been deposited in GenBank under accession numbers AF165818, AJ010592 and AF083031.

Nonlinear effects of large-scale climatic variability on wild and domestic herbivores

Atle Mysterud*, Nils Chr. Stenseth*, Nigel G. Yoccoz*†, Rolf Langvatn‡ & Geir Steinheim§

* *Department of Biology, Division of Zoology, University of Oslo, PO Box 1050 Blindern, N-0316 Oslo, Norway*
† *Department of Arctic Ecology, Norwegian Institute for Nature Research (NINA), Polar Environmental Centre, N-9296 Tromsø, Norway*
‡ *University Courses on Svalbard (UNIS), N-9170 Longyearbyen, Spitsbergen, Norway*
§ *Department of Animal Science, Agricultural University of Norway, PO Box 5025, N-1432 Ås, Norway*

Large-scale climatic fluctuations, such as the North Atlantic Oscillation (NAO)[1,2], have been shown to affect many ecological processes[3–6]. Such effects have been typically assumed to be linear. Only one study has reported a nonlinear relation[7]; however, that nonlinear relation was monotonic (that is, no reversal). Here we show that there is a strong nonlinear and non-monotonic (that is, reversed) effect of the NAO on body weight during the subsequent autumn for 23,838 individual wild red deer (*Cervus elaphus*) and 139,485 individual domestic sheep (*Ovis aries*) sampled over several decades on the west coast of Norway. These relationships are, at least in part, explained by comparable nonlinear and non-monotonic relations between the NAO and local climatic variables (temperature, precipitation and snow depth). The similar patterns observed for red deer and sheep, the latter of which live indoors during winter and so experience a stable energy supply in winter, suggest that the (winter) climatic variability (for which the index is a proxy) must influence the summer foraging conditions directly or indirectly.

Local weather patterns[8] and large-scale climatic variability[6,9–11] affect demography and population dynamics of temperate ungulates. The NAO[1] is a large-scale fluctuation in atmospheric mass between the subtropical North Atlantic region (centred on the Azores) and the subpolar North Atlantic region (centred on Iceland)[12]. The NAO is positively correlated with temperature and precipitation at the west coast of Norway during winter[1]. Because temperatures in this region often are around 0 °C during winter, precipitation (as rain) is positively correlated and snow depth is negatively correlated with the NAO at low elevation (below 400 m), and precipitation (as snow) is positively correlated with the NAO at high elevation[11]. Similar differential impacts of global climatic warming at low-altitude and high-altitude habitats have been reported from the Rocky Mountains[13].

Using wild red deer (*Cervus elaphus*) and domestic sheep (*Ovis aries*) lambs as examples of wild and domestic herbivores, we have explored comparatively the effect of the NAO (being a proxy for the interannual variation in winter temperature and precipitation along the west coast of Norway[1,11]) on individual performance (measured as body weight). As domestic sheep in Norway are free-ranging only from May/June to September/October, we can separate, through a comparison of the two species, the direct effects (costs of thermoregulation and movement in snow) and indirect effects (delayed snow melt can affect foraging conditions during summer) of winter climate.

In our study region, sheep and red deer in Norway are sympatric and have very similar diet composition and habitat use[14]. During winter, however, sheep are fed indoors, and as a result experience a stable interannual energy supply during that season. Thus, any relation between growth of lambs (during summer) and winter

© 2001 Macmillan Magazines Ltd