moving along the gene in one base frames, thus, allowing a mutation to be detected by a difference in the pattern of hybridization. There are many theoretical problems but, nevertheless, in July 1996 a chip for HIV enzyme sequence analysis was released together with an instrument for its assay[21]. There has been a recent description of an alternative method of synthesis and use of such chips[22] and the field has recently been reviewed in the columns of this journal[23].

## Outlook

How is the scientist, faced with a mutation detection problem, to decide among these methods? One almost needs to generate an algorithm from the characteristics in Box 1 and Table 1 depending on the project, the gene, the size, the background of workers, and so on. It is not easy, and the chosen method takes some months to be established with confidence in the laboratory. To alleviate some of these problems, an Australian 'Genome Centre' (Web site: http://www.agrf.org.au) will be initiated in 1997 that will offer mutation detection (together with linkage and sequencing) on a subscription basis. We hope that this will improve the predicament of those detecting mutations and possibly be a model for mutation detection in the future.

## References


1 Cotton, R.G.H. (1993) *Mutat. Res.* 285, 125–144
2 Grompe, M. (1993) *Nat. Genet.* 5, 111–117
3 Dracopoli, N. *et al.* (1994) *Current Protocols in Human Genetics*, Wiley & Sons
4 Landegren, U. (1996) *Laboratory Protocols for Mutation Detection*, Oxford University Press
5 Cotton, R.G.H. (1996) *Mutation Detection*, Oxford University Press
6 Dianzani, I., Camaschella, C., Ponzone, A. and Cotton, R.G.H. (1993) *Trends Genet.* 9, 403–405
7 Forrest, S., Cotton, R.G.H., Landegren, U. and Southern, E. (1995) *Nat. Genet.* 10, 375–376
8 Whitall, R. *et al.* (1995) *J. Med. Genet.* 32, 509–515
9 Lyamichev, V., Brow, M.D. and Dahlberg, J.E. (1993) *Science* 260, 778–783
10 Brow, M.A.D. *et al. J. Clin. Microbiol.* (in press)
11 Kumar, R. *et al.* (1995) *Carcinogenesis* 16, 2667–2673
12 Wu, Y. *et al.* (1996) *Hum. Mutat.* 8, 160–167
13 Underhill, P.A. *et al.* (1996) *Proc. Natl. Acad. Sci. U. S. A.* 93, 196–200
14 Ophoff, R.A. *et al.* (1996) *Cell* 87, 1–20
15 Nash, K.A. and Inderlied, C.B. (1996) *Antimicrob. Agents Chemother.* 40, 1748–1750
16 Verpy, E. *et al.* (1996) *Am. J. Hum. Genet.* 58, 308–319
17 Youil, R., Kemper, B. and Cotton, R.G.H. (1996) *Genomics* 32, 431–435
18 Smith, J. and Modrich, P. (1996) *Proc. Natl. Acad. Sci. U. S. A.* 93, 4374–4379
19 Wagner, R., Debbie, P. and Radman, M. (1995) *Nucleic Acids Res.* 23, 3944–3948
20 Lispshutz, R.J. *et al.* (1995) *BioTechniques* 18, 442–447
21 Kreiner, T. (1996) *Am. Lab.* 28, 39–43
22 Yershov, G. *et al.* (1996) *Proc. Natl. Acad. Sci. U. S. A.* 93, 4913–4918
23 Southern, E. (1996) *Trends Genet.* 12, 110–115


# Bonsai genomics: sequencing the smallest eukaryotic genomes


GEOFFREY I. McFADDEN*, PAUL R. GILSON*, SUSAN E. DOUGLAS§, THOMAS CAVALIER-SMITH‡, CLAUDIA J.B. HOFMANN¶ AND UWE-G. MAIER¶

mc_fadden@botany.unimelb.edu.au
susan.douglas@nrc.ca
tom@tcs.botany.ubc.ca
uwe.maier@btr0x1.hrz.uni-bayreuth.de

*PLANT CELL BIOLOGY RESEARCH CENTRE, SCHOOL OF BOTANY, UNIVERSITY OF MELBOURNE, PARKVILLE, VICTORIA 3052, AUSTRALIA; §NATIONAL RESEARCH COUNCIL OF CANADA, INSTITUTE FOR MARINE BIOSCIENCE, 1411 OXFORD STREET, HALIFAX, NS, CANADA B3H 3Z1; ‡EVOLUTIONARY BIOLOGY PROGRAM, CANADIAN INSTITUTE FOR ADVANCED RESEARCH, DEPARTMENT OF BOTANY, UNIVERSITY OF BRITISH COLUMBIA, VANCOUVER, BC, CANADA V6T 1Z4; ¶UNIVERSITAET BAYREUTH, LEHRSTUHL FUR PFLANZENSYSTEMATIK, UNIVERSITAETSSTR. 30, D-95440 BAYREUTH, BAVARIA, GERMANY.


Several projects characterizing the complete sequence of whole genomes are under way. In addition to several mitochondrial and plastidial genomes, the complete genome sequence of two prokaryotes and one eukaryote (budding yeast) have now been determined[1–3]. In the near future, the sequences of several other prokaryote genomes will be completed and substantial progress with nuclear genomes from higher plants, algae and animals will have been made. Studies of complete genomes from prokaryotic and eukaryotic organisms help us to understand cellular functions and to predict the function of uncharacterized gene products. We also gain insight into the minimal number of subunits in protein complexes, and the repertoire of essential 'housekeeping genes'. The complete sequencing of genomes will also facilitate more reliable phylogeny reconstruction, particularly for eukaryotic groups.

## A new genome project

A new genome project, initiated by groups from Canada, Australia and Germany, focuses on the smallest eukaryotic genomes investigated so far. These genomes, known as nucleomorphs, are restricted to two groups of algae, the cryptophytes and chlorarachniophytes (Fig. 1). Nucleomorphs are remnant nuclei of former free-living eukaryotic algae that have been engulfed by another eukaryotic cell and put to work as solar-powered food factories (secondary endosymbiosis, Fig. 2)[4–6]. The resulting chimaeric cell contains four genomes: the plastid, the mitochondrion, the host nucleus and the nucleomorph (endosymbiont nucleus). Coevolution between the two partner cells has resulted in the nucleus of the newest host assuming master control of plastid gene expression and the endosymbiont



 0168-9525/97/$17.00
PII: S0168-9525(97)01010-X

nucleus has undergone drastic reduction. Indeed, for some lineages (heterokont and haptophyte algae) only the plastid compartment and endosymbiont plasma membrane have been retained, and the endosymbiont nucleus has disappeared. In cryptophytes and chlorarachniophytes, however, the reduced nucleus, and a modicum of cytoplasm from the endosymbiont, persist. These represent a minimal eukaryotic organizational unit, probably semi-autonomous, like mitochondria and plastids, but eukaryotic in essence.

### The nucleomorph

So far, all nucleomorphs examined contain three small linear chromosomes. The total genome length in chlorarachniophytes is 380 kb (the smallest eukaryotic genome), but is marginally greater in cryptophytes averaging 600 kb (Refs 6, 7). Despite these superficial similarities in karyotype and genome size, gene sequence phylogenies demonstrate that cryptophytes and chlorarachniophytes evolved independently from separate secondary endosymbiotic events[8,9]. The similarities in gross structure (in both types of nucleomorph the three chromosomes carry rRNA genes and chromosome size ranges from only 95–230 kb) could indicate convergence at a minimal operational chromosome size and number. As mentioned above, many algae are believed to have evolved by secondary endosymbiosis, but retain only the plastid from the eukaryotic endosymbiont. Retention of the endosymbiont's nucleus and cytoplasm in cryptophytes and chlorarachniophytes raises the question of what role the nucleomorph, itself plays.

The most likely function is maintenance of the plastid, either by supplying essential proteins or by facilitating the passage of proteins originating from the secondary host nucleocytoplasmic compartment. Sequence characterization is the optimal way to tease out nucleomorph function. Preliminary work has already shown that a diverse cadre of housekeeping genes are encoded on the nucleomorph chromosomes of cryptophytes and chlorarachniophytes[10,11]. Because these encode transcription, mRNA processing, translation, protein degradation, and signal transduction-related functions, the nucleomorph apparently contains the machinery for expressing its information content.

### Importance of studying a reduced nucleus for a genome project

Given that they are functionally important, the nucleomorphs of cryptophytes and chlorarachniophytes can provide useful data for studying processes of coordinated gene expression and also for phylogenetic reconstruction. Our research is interdisciplinary and seeks to answer various questions of phylogeny and function. How many genes are really located on these genomes? What are the structural requirements for maintenance of small chromosomes? Are typical components of chromosome replication and division (telomeres, centromeres, replication origins) still identifiable? Are transcriptional regulation mechanisms required in these minimal genomes, and are there processing mechanisms for mRNAs? Are intron locations conserved? Have introns undergone reduction and/or removal during genome streamlining? Does RNA editing occur? What is the nature of the nucleomorph-encoded genes? Does the nucleomorph encode proteins transported into the plastid? Answers to some of these questions are already emerging.

### Genetically different sources contribute to multiprotein complexes

Like other diminutive genomes, nucleomorph DNA is rich in genes. In *Chlorarachnion*, noncoding regions between genes are less than 100 bp; in cryptomonads the average intergenic region is somewhat larger. Furthermore, no extended noncoding regions have been detected in either nucleomorph, making a sequencing project particularly attractive in a eukaryote realm. Despite this compactness, there is only room for approximately 200–300 genes in a nucleomorph. This cohort seems insufficient to maintain functional and structural control of both the plastid and the endosymbiont's cytoplasm. The missing functions appear to be provided by the nucleus of cryptomonads or chlorarachniophytes (G.I. McFadden and U-G. Maier, unpublished). It is, thus, apparent that gene expression in three genomes (nucleomorph, host nucleus and plastid) must be coordinated. For instance, in cryptomonads the protein degradation label, ubiquitin, is encoded by the host nucleus, but the ubiquitin conjugation enzyme is encoded by a nucleomorph gene (U-G. Maier, unpublished). Thus, in addition to plastid and mitochondrial protein complexes, nucleomorph-containing algae represent an additional system where subunits of protein complexes are provided by genomes from phylogenetically distant origins. As far as we know, this is the only system in which two different eukaryotic genomes act in concert with a prokaryotic genome to achieve a common goal: photosynthesis.

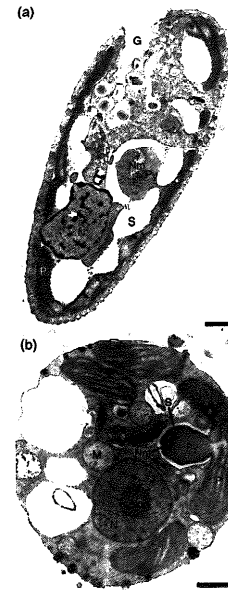


FIGURE 1. Electron micrographs of (a) *Rhodomonas salina* (cryptophytes), (b) *Chlorarachnion reptans* (chlorarachniophytes). Abbreviations: E, ejectosome; G, gullet; M, mitochondrion; N, nucleus; Nm, nucleomorph; No, nucleolus; P, plastid; Py, pyrenoid or plastid protein body; S, polysaccharide grain.

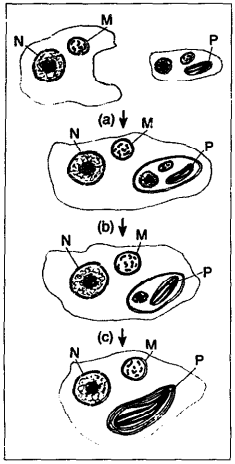


FIGURE 2. The evolution of algae by secondary endosymbiosis. (a) A phototrophic eukaryote is engulfed by another phagotrophic eukaryote. (b) Continued symbiosis leads to reduction of the symbiont, including partial reduction of the nucleus to a nucleomorph. (c) Further reduction leads to an alga whose plastid is surrounded by four membranes. Abbreviations: M, mitochondrion; N, nucleus; P, plastid.

## Structural and functional aspects of the nucleomorph

Nucleomorph chromosomes are linear[12,13] and, in chlorarachniophytes at least, there are small telomeres at the ends[13]. Although a eukaryotic chromosomal architecture is evident, it is intriguing that nucleomorphs divide in the absence of a mitotic spindle and without detectable condensation of chromosomes[14]. So, will we find centromeres on these tiny chromosomes or will they have a novel chromosome segregation mechanism with a unique primary structure at the DNA level? And how many replication origins can we anticipate on chromosomes that are barely more than twice the size of lambda phage? Do such small chromosomes require histones? Why are the chromosomes in all the different nucleomorphs examined to date never less than 95 kb? Is this a minimal size for viability?

Gene regulation systems are just beginning to be identified in nucleomorph genomes. Nucleomorph cDNAs have been isolated[10,11]. These cDNAs are polyadenylated, apparently using a standard (ATTAAA) polyadenylation motif[11]. In addition, components of the transcription machinery, such as a TATA-box-binding protein have been isolated (U-G. Maier, unpublished). In *Chlorarachnion*, the nucleomorph genes contain an unexpected surfeit of introns, and several components of the intron removal machinery (spliceosomes) have been identified[11]. Transcription has also been shown to include multiple open reading frames within single mRNAs suggesting that the squeeze on gene spacing has lead to fusion of transcription to create operon-like multicistronic messages[11,15].

And what about other genes? Interestingly, the nucleomorph seems to be a gold mine for the identification of housekeeping genes. The sequences of these genes are helping to identify important functional domains, particularly where very few examples are known from other systems. Perhaps more importantly, a number of ORFs of unknown function revealed by other eukaryotic sequencing projects are also being found in nucleomorphs (G.I. McFadden, unpublished). Because the nucleomorph genome encodes predominantly core components of eukaryotic metabolism, it is highly likely that these ORFs represent genes of fundamental importance. In this respect, nucleomorphs could serve as indicator genomes for yeast or *Caenorhabditis elegans* by identifying some genes universal to all eukaryotes, similar to the way the relatively small pufferfish genome serves as an index for the vastly larger human genome[16].

## The plastid

Several nucleomorph genes encode plastid-located proteins (G.I. McFadden and U-G. Maier, unpublished). In fact, the *raison d'être* for the nucleomorph is thought to be provision of genes necessary for plastid function. However, to elucidate the teamwork between the plastid, nucleomorph and nuclear genomes in the maintenance of a functional plastid compartment, it is also necessary to obtain as much information about the plastid as possible. Knowledge of the gene complement of the plastid will be useful in predicting which genes are likely to be encoded in the nucleomorph. In addition, phylogenetic analysis of plastid genes from cryptophytes will also elucidate the closest relative of the photosynthetic eukaryote (and, thus, of the nucleomorph genome) that gave rise to the plastid of the cryptomonad cell by secondary endosymbiosis. Therefore, determination of the complete sequence of the plastid genome of the cryptophyte *Guillardia theta* is included in our genome project.

To date, sequence analysis of this relatively small (118 kb) plastid genome has revealed a variety of genes not reported from the extensively studied land plant plastid genomes. These genes encode proteins involved in such functions as electron transport, photosynthesis, DNA binding, protein secretion, transcriptional regulation, protection from heat shock effects, and biosynthesis of amino acids, chlorophyll, carotenoids, phycobilins and fatty acids. In addition, genes for tRNAs and rRNAs are present as well as a large number of ycfs (homologous ORFs found in several plastid genomes). The cryptomonad plastid genome is characterized by a small rRNA-encoding inverted repeat (4.9 kb) as well as very tight packing of its genes, with organization of related genes into large transcriptionally linked clusters. Numerous gene clusters are conserved between cyanobacteria and the plastid genomes of cryptophytes, rhodophytes and chromophytes, indicating their common ancestry.

## Conclusions

Genome projects for eubacteria, archaebacteria, plants, animals and fungi provide an overview of coding functions in whole genomes and elucidation of fundamental biological phenomena. These include the deduction of still unknown metabolic pathways and a better comprehension of general regulatory systems involved in gene expression. However, the evolutionary implications of comparative genome analyses remain understated.

It is not expected that the nucleomorph genome projects will lead to a complete understanding of how a

cell or an organism works. However, determination of the entire sequence of these miniaturized eukaryotic genomes is a unique opportunity to study gene expression strategies, and to understand what comprises the minimal chromosome. Additionally, the nucleomorph genes serve as indicators of core eukaryotic metabolism without the distractions of thousands of genes encoding organism-specific functions.

The algal nucleomorph–nucleus system is unique with respect to eukaryotic protein complexes. Nowhere else are four genomes of diverse evolutionary origin amalgamated into one cell. Coordinated activity ensures that components of three separate genetic compartments combine in the plastid to power the conglomerate. Unravelling the mechanisms of this process presents a stimulating challenge.

To us, the nucleomorph is analogous to dwarf plants produced by the venerable oriental art of bonsai. Trimmed and shaped over great spans of time, bonsai trees are perfect miniatures with all the pleasing qualities of form packed into an easily appreciable size. Endosymbiosis has created a bonsai nucleus for us in the form of the nucleomorph. Genome sequencing now provides us with the means to begin to appreciate a little piece of Nature's art.

**Acknowledgements**

The nucleomorph–plastid projects are supported by the Natural Sciences and Engineering Research Council (Canada), the Australian Research Council (Australia), and the Deutsche Forschungsgemeinschaft (Germany).

**References**

1 Fleischmann, R.D. *et al.* (1995) *Science* 269, 496–512
2 Fraser, C.M. *et al.* (1995) *Science* 270, 397–403
3 Johnston, M. (1996) *Curr. Biol.* 6, 500–503
4 Douglas, S.E., Murphy, C.A., Spencer, D.F. and Gray, M.W. (1991) *Nature* 350, 148–151
5 Maier, U-G. *et al.* (1991) *Mol. Gen. Genet.* 230, 155–160
6 McFadden, G.I. *et al.* (1994) *Proc. Natl. Acad. Sci. U. S. A.* 91, 3690–3694
7 Rensing, S.A., Goddemeier, M., Hofmann, C.J.B. and Maier, U-G. (1994) *Curr. Genet.* 26, 451–455
8 Van de Peer, Y., Rensing, S.A., Maier, U-G. and De Wachter, R. (1996) *Proc. Natl. Acad. Sci. U. S. A.* 93, 7732–7736
9 Cavalier-Smith, T. *et al.* (1996) *Eur. J. Phycol.* 31, 315–328
10 Hofmann, C.J.B. *et al.* (1994) *Mol. Gen. Genet.* 243, 600–604
11 Gilson, P.R. and McFadden, G.I. (1996) *Proc. Natl. Acad. Sci. U. S. A.* 93, 7737–7742
12 Eschbach, S. *et al.* (1991) *Nucleic Acids Res.* 19, 1779–1781
13 Gilson, P.R. and McFadden, G.I. (1995) *Chromosoma* 103, 635–641
14 McKerracher, L. and Gibbs, S.P. (1982) *Can. J. Bot.* 60, 2440–2452
15 Palmer, J.D. and Delwiche, C.W. (1996) *Proc. Natl. Acad. Sci. U. S. A.* 93, 7432–7435
16 Elgar, G. *et al.* (1996) *Trends Genet.* 12, 145–151

# MEETING REPORTS

## *Saccharomyces sapiens*

YEAST GENETICS AND HUMAN DISEASE, BALTIMORE, MD, USA, 14–17 NOVEMBER 1996.

Six months after the complete yeast (*Saccharomyces cerevisiae*) genome sequence was released, and with the human EST (expressed sequence tag) database swelling past the 400 000 mark, this meeting marked a historic intersection between the two fields of research. The scene was set by the keynote speakers who echoed Mendeleev, proclaiming that what we have in the sequence is a periodic table for eukaryotes – the challenge is how to use it. The human genome sequence could be completed as early as 2005, and so lessons learned in yeast(s) will be fundamental in deciding how best to use genomic information in the future.

A brief tour of the yeast genome showed that, of the 5885 known (or predicted) genes, only about 40% have any genetic or biochemical data associated with them. The progression from systematic sequencing to systematic gene knockouts is already under way. Some 266 of the 267 predicted ORFs on chromosome VIII have been knocked out, and less than a fifth of them are essential genes (Mark Johnston, Washington Univ., USA). A quasi-random gene disruption scheme is being employed by Pat Brown (Stanford Univ., USA). Yeast cells are mutagenized by inducing transposition of the naturally occurring Ty1 elements. This population of cells can then be selected under various conditions, and the location of transposons mapped by PCR. This approach can be used to associate genes with selected phenotypes, and to define ORFs which were not large enough to be spotted from the sequence data alone.

An alternative, more systematic approach involves tagging individual deletants with a unique oligonucleotide sequence (Ron Davis, Stanford Univ., USA)[1]. To make these tagged gene deletions, a yeast selectable marker is amplified by PCR, adding homologous ends for recombination, thereby allowing replacement of a specific segment in the genome. At the same time, a known 20mer sequence is added which becomes a unique molecular 'bar code' for that particular gene deletion. By hybridization to an array of oligonucleotides (such as on a DNA chip – see below), these barcodes can be read out, revealing which gene deletions are present in the population of cells examined. This allows large collections of deletants to be analysed in parallel.

DNA chips (David Lockhart, Affymetrix, USA) were mentioned by a number of speakers. These are essentially an ordered array of up to 65000 oligonucleotides immobilized on derivatized glass slides. By hybridizing fluorescent labelled DNA or RNA probes to the chips, it is




PII: S0168-9525(97)01014-7