Reference: *Biol. Bull.* **196:** 400–403. (June 1999)

A Genomics Approach to Mitochondrial Evolution

MICHAEL W. GRAY[1], GERTRAUD BURGER[2], ROBERT CEDERGREN[2], G. BRIAN GOLDING[3], CLAUDE LEMIEUX[4], DAVID SANKOFF[5], MONIQUE TURMEL[4], AND B. FRANZ LANG[2]

Program in Evolutionary Biology, Canadian Institute for Advanced Research, [1]Department of Biochemistry, Dalhousie University, Halifax, Nova Scotia, Canada; [2]Département de biochimie et [5]Centre de recherches mathématiques, Université de Montréal, Montréal, Québec, Canada; [3]Department of Biology, McMaster University, Hamilton, Ontario, Canada; and [4]Département de biochimie, Université Laval, Québec, Québec, Canada

The pathways of mitochondrial DNA (mtDNA) evolution are obscured by the extraordinary diversity in size, gene content, gene arrangement, and mode of expression of characterized mitochondrial genomes (1–3). For this reason, a comparison of mtDNAs in animals (4), plants (5), and fungi (6, 7) leads to few conclusions about the evolutionary pressures and mechanisms by which the mitochondrial genome has become so different in these multicellular eukaryotic lineages, or about the form and gene content of the mtDNA ancestor (proto-mitochondrial genome). Not only do the mitochondrial genomes in multicellular eukaryotes, especially animals and fungi, appear to be highly derived, but these lineages represent relatively recently evolved, terminal branches in the eukaryotic phylogenetic tree. Hence, the multicellular eukaryotes are likely to give us a highly biased picture of the mitochondrial genome and its evolution. To provide a phylogenetically more representative sampling of mitochondrial genome structure, it is necessary to examine mtDNA in a diversity of protists (primarily unicellular eukaryotes), a vast assemblage that comprises most of the evolutionary diversity of the eukaryotic lineage (8). Among the questions that a comprehensive and systematic comparative study of protist mtDNAs will better address are the following: Did the mitochondrial genome arise only once (monophyletic origin) or more than once (polyphyletic origin)? How was (were) the ancestral proto-mitochondrial genome (or genomes) organized, and what genes did it (they) contain? What are the pathways and mechanisms of mtDNA evolution that account for the current observed structural variation?

This paper was originally presented at a workshop titled *Evolution: A Molecular Point of View.* The workshop, which was held at the Marine Biological Laboratory, Woods Hole, Massachusetts, from 24–26 October 1997, was sponsored by the Center for Advanced Studies in the Space Life Sciences at MBL and funded by the National Aeronautics and Space Administration under Cooperative Agreement NCC 2-896.

A program to sequence protist mitochondrial DNAs

In 1992, the Organelle Genome Megasequencing Program (OGMP) was initiated as a consortium of seven investigators from four eastern Canadian universities, having as one of its goals the complete sequencing of mitochondrial genomes from a wide range of protists. The experimental arm of the OGMP is a centralized sequencing facility (Sequencing Unit) located at the University of Montreal. The Unit comprises a Molecular Biology Division, within which sequences are determined and analyzed, and an Informatics Division, which develops informatics tools to aid in all aspects of sequence determination and analysis. To date, the OGMP has determined and analyzed 11 complete protist mtDNA sequences, with 3 additional mtDNA sequencing projects nearing completion. The accumulating data (representing >750,000 bp of finished sequence so far) are providing new insights into the questions raised above.

Gene content and genome organization in protist mitochondrial DNAs

To date, detailed descriptions of three protist mtDNAs sequenced by the OGMP have been published, representing a chlorophyte alga, *Prototheca wickerhamii* (9, 10); an amoeboid protozoon, *Acanthamoeba castellanii* (11); and a jakobid flagellate, *Reclinomonas americana* (12). For the other protist mtDNA sequencing projects, individual de-

scriptions summarizing salient features of gene content, organization, and expression are available through the OGMP website (see below). Sequence data generated for protist mtDNAs by the OGMP and other groups have been compiled and reviewed in detail elsewhere (13). Among the findings and conclusions to emerge from this work are the following:

1. By and large, protist mtDNAs are A + T-rich genomes (usually >70%) with a median size of around 40 kbp. Typically, they are compactly organized, gene-rich, circular-mapping genomes with few, or no, large noncoding regions. Intergenic spacers usually account for less than 10% of the mtDNA, with coding regions often overlapping.
2. In general, protist mtDNAs encode substantially more genes than do animal and fungal mtDNAs. Additional genes that are widely found in protist mtDNAs include ones specifying extra subunits of respiratory complexes I (*nad*) and V (*atp*) and both large subunit and small subunit ribosomal proteins (*rpl* and *rps,* respectively). Animal and fungal mtDNAs completely, or almost completely (14), lack genes for ribosomal proteins. Animal mitochondrial genomes, in particular, appear to be exquisitely condensed versions of a much larger, protist-like progenitor.
3. In overall gene content, protist mtDNAs more closely resemble the much larger mitochondrial genomes of land plants. Genes shared between plant and protist mtDNAs include not only those that are absent from animal and fungal mtDNAs (noted above), but occasionally also genes encoding subunits of complex II (*sdh;* ref. 15) and 5S rRNA (*rrn5;* ref. 16). In spite of their larger size, plant mtDNAs do not encode substantially more genes than protist mtDNAs (13). In marked contrast to their animal counterparts, land plant mtDNAs appear to have expanded considerably in size, mostly through the acquisition of large amounts of noncoding spacer DNA.
4. Typically, the ribosomal and transfer RNAs encoded by protist mtDNAs are strikingly eubacteria-like in their secondary structures, lacking the anomalous features found in their fungal and particularly animal mitochondrial counterparts. Although a few protist mitochondrial genomes encode a minimal set of tRNA species that is sufficient to recognize all codons found in protein-coding genes, most do not (13). In these cases, the import of nucleus-encoded, cytosolic tRNAs into the mitochondria is assumed to make up the deficit.
5. At least half of the protist mtDNAs examined appear to use the standard genetic code, as does the plant mitochondrial genome. In the remaining protist mtDNAs, the codon UGA is used to specify tryptophan rather than termination. Phylogenetic considerations strongly indicate that this codon change has occurred independently in several branches of the eukaryotic tree during mitochondrial genome evolution.
6. Introns, either group I or group II, are conspicuously rare in protist mitochondrial genomes. Several protist mtDNAs lack introns entirely. Where introns are present, the evidence suggests that a number of them have been acquired by means of horizontal transfer events, rather than being inherited vertically from the proto-mitochondrial genome.
7. Aspects of protist mitochondrial gene content, primary and secondary structure, gene organization, and gene expression also serve to emphasize the underlying eubacterial character of the mitochondrial genome, reinforcing the theory that mitochondria originated in evolution from eubacteria-like symbionts. Mitochondrion-specific patterns of gene organization strongly support a single (monophyletic) origin of the mitochondrial genome.
8. Although some of the protist mtDNAs initially examined (*e.g.,* those of the green alga *Chlamydomonas reinhardtii* and the apicomplexan *Plasmodium falciparum,* the human malaria parasite) proved to have a very unusual—indeed bizarre—structure, subsequent work by the OGMP and other groups has indicated that these mtDNAs are likely to be highly derived. In fact, the most ancestral (least derived) mitochondrial genomes are to be found among the protists, particularly in free-living species whose mtDNA has not undergone the dramatic re-tailoring suffered by its counterpart in many parasites.
9. The most ancestral mitochondrial genome so far described, that of *Reclinomonas americana* (12), resembles nothing so much as a miniaturized eubacterial genome. In addition to encoding 5S and RNase P RNAs that are eubacteria-like in primary and secondary structure, *R. americana* mtDNA contains a set of genes never before found in mtDNA. Among the most intriguing are genes (*rpoA-D)* encoding a eubacteria-like ($\alpha_2\beta\beta'\sigma$) RNA polymerase. This finding raises intriguing questions about the evolution of the mitochondrial transcription system (17, 18), because a single-subunit, nucleus-encoded, bacteriophage T3/T7-like enzyme functions as the mitochondrial RNA polymerase in other eukaryotes (see, *e.g.,* ref. 19).

Future prospects and on-line access to the OGMP

Sequencing of protist mtDNAs is continuing in the OGMP, with the current emphasis on identification of fur-

ther examples of ancestral mitochondrial genomes from organisms branching deep within the eukaryotic lineage. The new insights obtained will undoubtedly influence our ideas about the origin of the eukaryotic cell, and the crucial role that the acquisition of mitochondria has played in this process.

Information about the OGMP and about individual mtDNA sequencing projects and informatics tools may be obtained by consulting World Wide Web sites for the following organizations (URLs given below): the OGMP, the allied Protist Image Database (PID), the Organelle Genome Database (GOBASE) (20), and the Fungal Mitochondrial Genome Project (FMGP).

OGMP	http://megasun.bch.umontreal.ca/ogmp
PID	http://megasun.bch.umontreal.ca/protists
GOBASE	http://megasun.bch.umontreal.ca/gobase/
FMGP	http://megasun.bch.umontreal.ca/People/lang/FMGP/

Acknowledgments

Work described in this paper was supported by a Special Project grant (SP-34) from the Medical Research Council of Canada, a grant from the Canadian Genome Analysis and Technology Program (GO-12323), and by generous donations of equipment from Sun Microsystems Inc. and LI-COR Inc. Salary and interaction support from the Canadian Institute for Advanced Research (Program in Evolutionary Biology) are also gratefully acknowledged.

Literature Cited

1. **Gray, M. W. 1989.** Origin and evolution of mitochondrial DNA. *Annu. Rev. Cell Biol.* **5:** 25–50.
2. **Gray, M. W. 1992.** The endosymbiont hypothesis revisited. *Int. Rev. Cytol.* **141:** 233–357.
3. **Gillham, N. W. 1994.** *Organelle Genes and Genomes,* Oxford University Press, New York.
4. **Wolstenholme, D. R. 1992.** Animal mitochondrial DNA: structure and evolution. *Int. Rev. Cytol.* **141:** 173–216.
5. **Hanson, M. R., and O. Folkerts. 1992.** Structure and function of the higher plant mitochondrial genome. *Int. Rev. Cytol.* **141:** 129–172.
6. **Clark-Walker, G. D. 1992.** Evolution of mitochondrial genomes in fungi. *Int. Rev. Cytol.* **141:** 89–127.
7. **Paquin, B., M.-J. Laforest, L. Forget, I. Roewer, Z. Wang, J. Longcore, and B. F. Lang. 1997.** The fungal mitochondrial genome project: evolution of fungal mitochondrial genomes and their gene expression. *Curr. Genet.* **31:** 380–395.
8. **Patterson, D. J., and M. L. Sogin. 1992.** Eukaryote origins and protistan diversity. Pp. 13–46 in *The Origin and Evolution of the Cell.* H. Hartman and K. Matsuno, eds. World Scientific, Singapore.
9. **Wolff, G., G. Burger, B. F. Lang, and U. Kück. 1993.** Mitochondrial genes in the colourless alga *Prototheca wickerhamii* resemble plant genes in their exons but fungal genes in their introns. *Nucleic Acids Res.* **21:** 719–726.
10. **Wolff, G., I. Plante, B. F. Lang, U. Kück, and G. Burger. 1994.** Complete sequence of the mitochondrial DNA of the chlorophyte alga *Prototheca wickerhamii.* Gene content and genome organization. *J. Mol. Biol.* **237:** 75–86.
11. **Burger, G., I. Plante, K. M. Lonergan, and M. W. Gray. 1995.** The mitochondrial DNA of the amoeboid protozoon, *Acanthamoeba castellanii*: complete sequence, gene content and genome organization. *J. Mol. Biol.* **245:** 522–537.
12. **Lang, B. F., G. Burger, C. J. O'Kelly, R. Cedergren, G. B. Golding, C. Lemieux, D. Sankoff, M. Turmel, and M. W. Gray. 1997.** An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. *Nature* **387:** 493–497.
13. **Gray, M. W., B. F. Lang, R. Cedergren, G. B. Golding, C. Lemieux, D. Sankoff, M. Turmel, N. Brossard, E. Delage, T. G. Littlejohn, I. Plante, P. Rioux, D. Saint-Louis, Y. Zhu, and G. Burger. 1998.** Genome structure and gene content in protist mitochondrial DNAs. *Nucleic Acids Res.* **26:** 865–878.
14. **Paquin, B., and B. F. Lang. 1996.** The mitochondrial DNA of *Allomyces macrogynus:* the complete genomic sequence from an ancestral fungus. *J. Mol. Biol.* **255:** 688–701.
15. **Burger, G., B. F. Lang, M. Reith, and M. W. Gray. 1996.** Genes encoding the same three subunits of respiratory complex II are present in the mitochondrial DNA of two phylogenetically distant eukaryotes. *Proc. Natl. Acad. Sci. USA* **93:** 2328–2332.
16. **Lang, B. F., L. J. Goff, and M. W. Gray. 1996.** A 5 S rRNA gene is present in the mitochondrial genome of the protist *Reclinomonas americana* but is absent from red algal mitochondrial DNA. *J. Mol. Biol.* **261:** 607–613.
17. **Cermakian, N., T. M. Ikeda, P. Miramontes, B. F. Lang, M. W. Gray, and R. Cedergren. 1997.** On the evolution of the single-subunit RNA polymerases. *J. Mol. Evol.* **45:** 671–681.
18. **Gray, M. W., and B. F. Lang. 1998.** Transcription in chloroplasts and mitochondria: a tale of two polymerases. *Trends Microbiol.* **6:** 1–3.
19. **Cermakian, N., T. M. Ikeda, R. Cedergren, and M. W. Gray. 1996.** Sequences homologous to yeast mitochondrial and bacteriophage T3 and T7 RNA polymerases are widespread throughout the eukaryotic lineage. *Nucleic Acids Res.* **24:** 648–654.
20. **Korab-Laskowska, M., P. Rioux, N. Brossard, T. G. Littlejohn, M. W. Gray, B. F. Lang, and G. Burger. 1998.** The Organelle Genome Database Project (GOBASE). *Nucleic Acids Res.* **26:** 138–144.

Discussion

MARGULIS: Can we argue that you have nearly a *Paracoccus* or other gram negative bacterial genome here, and therefore *Reclinomonas* acquired the oxygen-respiring symbiont very recently?

GRAY: In other mitochondrial genomes, the genes are a subset of the ones in *Reclinomonas* mitochondrial DNA. Exactly the same set of small subunit ribosomal protein genes is present in *Reclinomonas* mitochondrial DNA as in the mitochondrial genome of the liverwort *Marchantia polymorpha.* In all other mitochondrial DNAs, the ribosomal protein genes are a subset of the *Reclinomonas* ones.

MARGULIS: Is there no possibility of molecular convergence?

GRAY: I would say that the number of characters that we are starting to accumulate makes it very difficult to see how independent reduction of much larger genomes could converge on these multiple specific patterns, which include not only gene content but also specific losses of ribosomal protein genes within a large operon. It depends on your viewpoint. How easy or difficult is it to converge on this kind of pattern? I think it is probably difficult if you assume that various gene losses are independent events—that all of those genes didn't move at the same time. Even if they did, the same set would have had to have been selected for retention each time.

MARGULIS: Is it a question of relentless selection?

GRAY: This is looking at it with the idea that there is probably no good reason why those particular genes were lost.

CAVALIER-SMITH: The *Reclinomonas* ribosomal RNA sequence isn't deep, unlike the *Naegleria* one, but the depth of the *Naegleria* one on the ribosomal RNA tree is not supported by any currently available protein data. It could well be that at least all of the mitochondria-containing eukaryotes are simply a star phylogeny. If we look at it in that way, and also realize that the divergence of the plastid-containing groups is part of that star-like phylogeny, then we could say that possibly the serial endosymbiosis theory is wrong, and that there was, instead, a simultaneous symbiotic origin of mitochondria chloroplasts and, as I have argued, peroxisomes.

GRAY: The issue is, if *Reclinomonas* and the jakobids are deep, where is the crown radiation relative to the origin of the eukaryotes? Are these two points really far apart in time, or are they relatively close together? It could well be that all of these symbiotic events happened, if not simultaneously, then very close together in time.

KATZ: It is hard to evaluate the trees without any knowledge of the alignment and phylogenetic algorithms that you are using. What algorithms were you using for the trees that were uniting the *Reclinomonas* with the land plants? Were those relationships robust if you tried other phylogenetic methods?

GRAY: The tree was generated from a sequence alignment based on secondary structure, so that conserved regions were selected from complete ribosomal RNA sequences. The data set comprises about 290 eubacterial or organellar sequences, so what I showed was simply the alpha-proteobacterial branch of a much larger assemblage. The tree is a neighbor-joining one, and this particular example wasn't bootstrapped. I don't place any great significance on a mitochondrial small subunit ribosomal RNA tree because we've had this artifact (plants separate from other eukaryotes) for many years now. Because the tree represents the first time that we have found an organism (*i.e., Reclinomonas*) that goes together with land plants, I offer it simply as an illustration of what I think is happening here.

LANDWEBER: Do maximum likelihood and parsimony give you the same result?

GRAY: We haven't applied these methods to this particular data set.

JEFF PALMER: I'd like to respond to something Tom (Cavalier-Smith) said about possible simultaneous origin of chloroplasts and mitochondria. If that were the case, then a number of eukaryotes would have lost their chloroplasts entirely, which is reasonable; but it becomes less and less likely when we look at fully sequenced genomes such as those of yeast and *C. elegans.* If chloroplast loss had occurred in these organisms, and there had been substantial transfer of chloroplast-derived genes to the nucleus in the common ancestor of these groups, you should find nuclear genes of chloroplast origin in yeast and *C. elegans,* and there don't seem to be any.

CAVALIER-SMITH: In my presentation, I had chloroplasts originating a bit after the primary divergence of plants from the ancestor of animals and fungi, but I still wonder whether we can really be sure about what things might be retained after chloroplast loss.