

An Ancient Fission of Mitochondrial *cox1*

Ryan M.R. Gawryluk and Michael W. Gray*

Centre for Comparative Genomics and Evolutionary Bioinformatics, Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, Canada

*Corresponding author: E-mail: m.w.gray@dal.ca.

Associate editor: Martin Embley

Abstract

Many genes inherited from the α -proteobacterial ancestor of mitochondria have undergone evolutionary transfer to the nuclear genome in eukaryotes. In some rare cases, genes have been functionally transferred in pieces, resulting in split proteins that presumably interact in trans within mitochondria, fulfilling the same role as the ancestral, intact protein. We describe a nucleus-encoded mitochondrial protein (here named Cox1-c) in the amoeboid protist *Acanthamoeba castellanii* that is homologous to the C-terminal portion of conventional mitochondrial Cox1, whereas the corresponding portion of the mitochondrion-encoded *A. castellanii* Cox1 is absent. Bioinformatics searches retrieved nucleus-encoded Cox1-c homologs in most major eukaryotic supergroups; in these cases, also, the mitochondrion-encoded Cox1 lacks the corresponding C-terminal motif. These data constitute the first report of functional relocation of a portion of *cox1* to the nucleus. This transfer event was likely ancient, with the resulting nuclear *cox1-c* being differentially activated across the eukaryotic domain.

Key words: mitochondria, gene transfer, gene fission, eukaryotic evolution.

Mitochondria are biochemically and ultrastructurally diverse eukaryotic organelles whose complete protein repertoire (proteome) ranges from several hundred to >1,000 polypeptides (Richly et al. 2003; Sickmann et al. 2003; Heazlewood et al. 2004; Smith et al. 2007). Although mitochondria have retained a relict genome (mitochondrial DNA [mtDNA]), the majority (>95%) of proteins that function in mitochondria are encoded in the nucleus (nuclear DNA [nuDNA]) and imported posttranslationally (Neupert and Hermann 2007). Many genes formerly encoded by mtDNA have been functionally transferred to the nucleus via endosymbiotic gene transfer (EGT; see Gray et al. 1999). Functional mitochondrion-to-nucleus EGT no longer occurs in animals but is still an active process in green algae (Pérez-Martínez et al. 2001) and land plants (Adams et al. 2000; Adams and Palmer 2003).

In most cases, the integrity of a gene is preserved in the course of gene transfer (i.e., the gene remains a single, continuous entity). However, several cases have been reported of fission in genes coding for mitochondrial proteins, resulting in either both halves encoded in the nuDNA (Gawryluk and Gray 2009; Pérez-Martínez et al. 2001) or one part encoded in the mtDNA and the other in the nuDNA (Adams et al. 2001; Funes et al. 2002). A notable example of mitochondrial gene fission is *cox2* in chlorophyte algae. Whereas both halves (Cox2a and Cox2b) of the typical Cox2 protein are encoded in the nuDNA in *Chlamydomonas* (Pérez-Martínez et al. 2001), Cox2a has been retained in the mtDNA of *Scenedesmus* (Nedelcu et al. 2000), whereas *Scenedesmus* Cox2b is nucleus-encoded (Funes et al. 2002): a clear scenario for the separate functional transfer of two halves of a mitochondrial gene to the nucleus.

From an evolutionary perspective, most documented examples of fissioned mitochondrial genes are relatively recent, that is, the events leading to them have occurred

within well-recognized lineages, such as chlorophyte algae (Pérez-Martínez et al. 2001), euglenozoons (Gawryluk and Gray 2009), or eudicots (Adams et al. 2001), and are therefore phylogenetically meaningful (Gawryluk and Gray 2009) and relatively straightforward to interpret (but cf. Funes et al. 2002; Waller and Keeling 2006).

We describe here a previously undetected split in the C-terminal region of the cytochrome *c* oxidase subunit 1 gene (*cox1*), a fission that has a punctate distribution within eukaryotes. We suggest that this distribution coupled with the phylogenetic breadth covered by this unique molecular event (which is evident in at least four of six currently proposed [Keeling et al. 2005] eukaryotic supergroups) reflects an ancient partial gene transfer that was differentially activated in diverse eukaryotic lineages.

During an ongoing investigation of the mitochondrial proteome of the amoeboid protozoon *Acanthamoeba castellanii*, a high-confidence peptide that matched an inferred 24.8-kDa nucleus-encoded protein (based on a translated expressed sequence tag [EST]) was identified by tandem mass spectrometry. Blastp (Altschul et al. 1997) analyses demonstrated that a short (~25 amino acid) region of this protein exhibits strong similarity to a conserved motif present within the C-terminal region of Cox1 proteins from α -proteobacteria and minimally divergent mtDNAs (e.g., *Reclinomonas* and *Marchantia*; see fig. 1), whereas the remainder of the inferred protein sequence shows no strong similarity to proteins in available sequence databases. Multiple protein alignments established that the mtDNA-encoded “fused” Cox1/2 of *Acanthamoeba* (Burger et al. 1995; Lonergan and Gray 1996) lacks this C-terminal portion, which we consider to be an ancestral part of a covalently continuous (i.e., conventional) Cox1 due to its presence in this form in α -proteobacteria and *Reclinomonas*

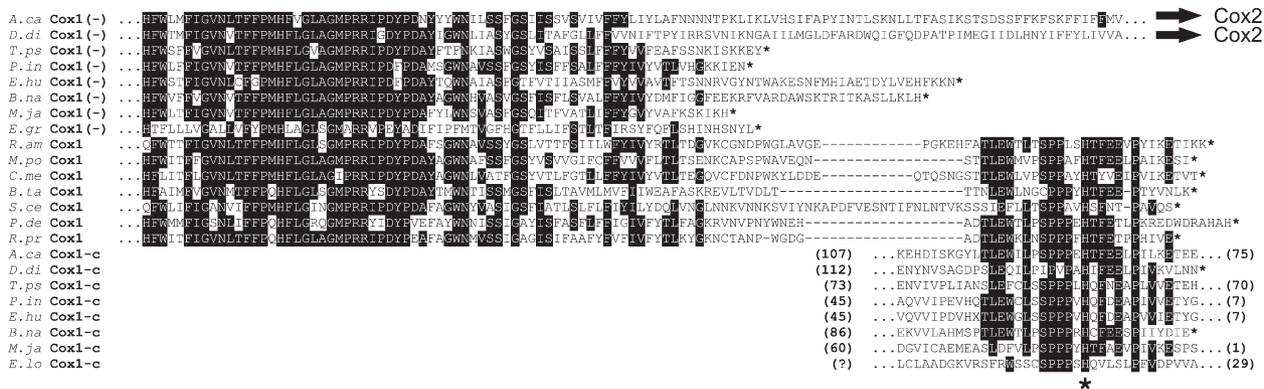


Fig. 1. Partial alignment of Cox1, Cox1(-), and Cox1-c protein sequences. The C-terminal region of Cox1 was aligned with inferred Cox1-c protein sequences. In this alignment, “...” denotes that the sequence continues, whereas “*” represents the stop codon. The large * demarcates H503 of bovine Cox1, which is 100% conserved in Cox1-c. Numbers adjacent to Cox1-c sequences indicate the number of N-terminal and C-terminal amino acids omitted from the alignment. Species abbreviations: *A.ca*, *Acanthamoeba castellanii*; *D.di*, *Dictyostelium discoideum*; *T.ps*, *Thalassiosira pseudonana*; *P.in*, *Phytophthora infestans*; *E.hu*, *Emiliania huxleyi*; *B.na*, *Bigelowiella natans*; *M.ja*, *Malawimonas jakobiformis*; *E.gr*, *Euglena gracilis*; *R.am*, *Reclinomonas americana*; *M.po*, *Marchantia polymorpha*; *C.me*, *Cyanidioschyzon merolae*; *B.ta*, *Bos taurus*; *S.ce*, *Saccharomyces cerevisiae*; *P.de*, *Paracoccus denitrificans*; *R.pr*, *Rickettsia prowazekii*; *E.lo*, *Euglena longa*. GenBank accession numbers are provided in [supplementary table 1](#) (Supplementary Material online).

mitochondria (Lang et al. 1997). We refer to the nuDNA-encoded *Acanthamoeba* protein homologous to the C-terminal portion of full-length Cox1 as “Cox1-c” and the truncated mtDNA-encoded Cox1 homolog as “Cox1(-).”

The *Acanthamoeba* Cox1-c sequence was used to query the proteome of a related amoebozoan, *Dictyostelium discoideum* (Eichinger et al. 2005) via Blastp. This search identified a homolog that is similar to Cox1-c in the region of the conserved motif of conventional Cox1, although the degree of similarity is not as great as in the case of the *Acanthamoeba* Cox1-c. The *Dictyostelium* Cox1-c homolog had already been identified as the largest nuDNA-encoded constituent of the *Dictyostelium* cytochrome oxidase complex and was previously named CoxIV (Bisson and Schiavo 1986). However, the authors failed to recognize the similarity to the C-terminus of conventional Cox1 (likely because the motif is not as well conserved in *Dictyostelium*) and suggested instead that this protein might be related to subunit VI of yeast and subunit Va of human cytochrome oxidase (Rizzuto et al. 1991). Together, these results suggest that the novel *Acanthamoeba* Cox1-c protein is a bona fide component of the cytochrome oxidase complex and that *Dictyostelium* CoxIV is homologous to the C-terminal portion of conventional Cox1.

We generated a phylogenetically broad alignment in order to examine the C-terminal portion of mtDNA-encoded Cox1 across the diversity of eukaryotes (fig. 1). The alignment revealed that the ancestral C-terminal motif of Cox1 is missing from the mtDNA of a wide variety of eukaryotes (i.e., they encode Cox1(-) rather than Cox1), including species from four of the six currently proposed eukaryotic supergroups. In contrast to *Acanthamoeba* and *Dictyostelium*, where the continuous nature of the Cox1/2 open reading frame (Burger et al. 1995) obscures the actual C-terminus of mature Cox1(-), other inferred Cox1(-) sequences are clearly truncated.

We hypothesized that separate, nuDNA-encoded homologs of Cox1-c might be present in those eukaryotes with an mtDNA-encoded Cox1(-). Searches of EST, nuclear genomic, and inferred protein sequence databases allowed us to identify nuDNA-encoded homologs of Cox1-c in many of the species lacking the motif in mtDNA-encoded Cox1(-)

Table 1. The Structure of Cox1 across Eukaryotes.

Eukaryotic Supergroup	Organism	mtDNA Cox1	Cox1-c	Comments
Excavata	<i>Reclinomonas</i>	Cox1	?	
	<i>Naegleria</i>	Cox1	No	
	<i>Euglena</i>	Cox1(-)	Yes	
	<i>Trypanosoma</i>	Cox1(-)	?	
	<i>Malawimonas</i>	Cox1(-)	Yes	
	<i>Acanthamoeba</i>	Cox1(-)	Yes	Continuous Cox1/2
Amoebozoa	<i>Dictyostelium</i>	Cox1(-)	Yes	Continuous Cox1/2
	<i>Physarum</i>	Cox1(-)	Yes	
Rhizaria	<i>Hyperamoeba</i>	?	Yes	
	<i>Bigelowiella</i>	Cox1(-)	Yes	
Chromalveolata	<i>Tetrahymena</i>	Cox1(-)	?	
	<i>Thalassiosira</i>	Cox1(-)	Yes	
	<i>Guillardia</i>	Cox1	No	
	<i>Cafeteria</i>	Cox1(-)	?	
	<i>Phytophthora</i>	Cox1(-)	Yes	
	<i>Emiliania</i>	Cox1(-)	Yes	
Opisthokonta	<i>Plasmodium</i>	Cox1(-)	?	
	<i>Bos</i>	Cox1	No	
Plantae	<i>Monosiga</i>	Cox1	No	
	<i>Saccharomyces</i>	Cox1	No	
	<i>Arabidopsis</i>	Cox1	No	
Plantae	<i>Marchantia</i>	Cox1	No	
	<i>Chlamydomonas</i>	Cox1	No	
	<i>Porphyra</i>	Cox1	No	

NOTE.—Organisms are categorized according to eukaryotic supergroup membership (Keeling et al. 2005). “Cox1” refers to intact (conventional) Cox1 in the mtDNA, whereas “Cox1(-)” specifies truncated mtDNA-encoded Cox1. “Cox1-c” is a nuDNA-encoded protein homologous to the C-terminal portion of Cox1. GenBank accession numbers are provided in [supplementary table 1](#) (Supplementary Material online).

(fig. 1). Cox1-c homologs exist in the supergroups Amoebozoa, Chromalveolata, Rhizaria, and Excavata (table 1), all of which also encode a Cox1(-) in their mtDNA. Although not all Cox1-c homologs are as similar to the C-terminus of conventional Cox1 as is *Acanthamoeba* Cox1-c, a consensus motif S-P-P-P-X-H (the H 100% conserved) aids in the identification of bona fide Cox1-c homologs (fig. 1, * denotes conserved H). As expected, the mitochondrial targeting peptide program TargetP (Emanuelsson et al. 2000) predicts mitochondrial localization for Cox1-c homologs having complete N-terminal sequence, at a probability of ≥ 0.65 .

The retention of Cox1-c coding sequence in the nuDNA of many eukaryotes after it was lost from the mtDNA argues strongly for the functional importance of this region of Cox1. None of the residues required for coordination of metal clusters is located within the extreme C-terminus of Cox1; however, crystal structures of the bovine enzyme in reduced and oxidized states demonstrate that H503 (which corresponds to the 100% conserved H residue of Cox1-c) is part of an important controlling site for entrance of protons into the dioxygen reduction pathway (Muramoto et al. 2007). Because proton pumping and dioxygen reduction are fundamental functions of cytochrome oxidase, these data suggest that Cox1-c (and the C-terminal portion of full-length Cox1) serves a critical function in the cytochrome oxidase complex, thereby explaining why this part of the protein is not easily lost.

Interestingly, full-length Cox1 proteins are interspersed with split Cox1 proteins within eukaryotic supergroups (table 1). For instance, within Excavata, Cox1 is full length in the jakobids (*Reclinomonas*) and heteroloboseans (*Naegleria*) but split in euglenozoans (*Euglena*; *Trypanosoma* and *Diplonema* encode Cox1(-), but corresponding nuDNA-encoded proteins could not be identified). Within Chromalveolata, haptophytes such as *Emiliania* have Cox1(-) proteins and nuDNA-encoded Cox1-c, whereas related cryptophytes such as *Guillardia* have full-length mtDNA-encoded Cox1 proteins. Thus far, the only supergroups in which the described Cox1 fission has not been detected are Opisthokonta (including fungi, animals, and their unicellular ancestors) and Plantae (comprising red algae, green algae, and land plants). Conversely, all amoebozoans examined so far have a nuDNA-encoded Cox1-c and mtDNA-encoded Cox1(-). From the perspective of parsimony, the widespread yet punctate distribution of the described *cox1* split suggests that sequence encoding the small C-terminal domain of Cox1 was transferred to the nuclear genome early in eukaryotic evolution and differentially lost/activated in different lineages. Alternatively, it is possible that the sequence encoding the C-terminal region of Cox1 (or *cox1* in its entirety) was transferred to the nucleus many times independently. Although this scenario appears less likely, our finding could reflect many parallel failed attempts at functional relocation of *cox1* to the nucleus (perhaps due to the high hydrophobicity of the intact Cox1 protein), with only a sequence encoding this

small, hydrophilic C-terminal end acquiring the necessary expression and targeting signals.

Supplementary Material

Supplementary table 1 is available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

R.M.R.G. was supported by a Canada Graduate Scholarship Doctoral from the Natural Sciences and Engineering Research Council and a Predoctoral Scholarship from the Killam Trusts. This work was supported by operating grant MOP-4124 from the Canadian Institutes of Health Research.

References

- Adams KL, Daley DO, Qiu YL, Whelan J, Palmer JD. 2000. Repeated, recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants. *Nature*. 408:354–357.
- Adams KL, Ong HC, Palmer JD. 2001. Mitochondrial gene transfer in pieces: fission of the ribosomal protein gene *rpl2* and partial or complete gene transfer to the nucleus. *Mol Biol Evol*. 18: 2289–2297.
- Adams KL, Palmer JD. 2003. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Mol Phylogenet Evol*. 29:380–395.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 25:3389–3402.
- Bisson R, Schiavo G. 1986. Two different forms of cytochrome c oxidase can be purified from the slime mold *Dictyostelium discoideum*. *J Biol Chem*. 261:4373–4376.
- Burger G, Plante I, Lonergan KM, Gray MW. 1995. The mitochondrial DNA of the amoeboid protozoan, *Acanthamoeba castellanii*: complete sequence, gene content and genome organization. *J Mol Biol*. 245:522–537.
- Eichinger L, Pachebat JA, Glockner G, et al. (97 co-authors). 2005. The genome of the social amoeba *Dictyostelium discoideum*. *Nature*. 435:43–57.
- Emanuelsson O, Nielsen H, Brunak S, von Heijne G. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol*. 300:1005–1016.
- Funes S, Davidson E, Reyes-Prieto A, Magallón S, Herion P, King MP, González-Halphen D. 2002. A green algal apicoplast ancestor. *Science*. 298:2155.
- Gawryluk RM, Gray MW. 2009. A split and rearranged nuclear gene encoding the iron-sulfur subunit of mitochondrial succinate dehydrogenase in Euglenozoa. *BMC Res Notes*. 2:16.
- Gray MW, Burger G, Lang BF. 1999. Mitochondrial evolution. *Science*. 283:1476–1481.
- Heazlewood JL, Tonti-Filippini JS, Gout AM, Day DA, Whelan J, Millar AH. 2004. Experimental analysis of the *Arabidopsis* mitochondrial proteome highlights signaling and regulatory components, provides assessment of targeting prediction programs, and indicates plant-specific mitochondrial proteins. *Plant Cell*. 16:241–256.
- Keeling PJ, Burger G, Durnford DG, Lang BF, Lee RW, Pearlman RE, Roger AJ, Gray MW. 2005. The tree of eukaryotes. *Trends Ecol Evol*. 20:670–676.
- Lang BF, Burger G, O’Kelly CJ, Cedergren R, Golding GB, Lemieux C, Sankoff D, Turmel M, Gray MW. 1997. An ancestral

- mitochondrial DNA resembling a eubacterial genome in miniature. *Nature*. 387:493–497.
- Loneragan KM, Gray MW. 1996. Expression of a continuous open reading frame encoding subunits 1 and 2 of cytochrome *c* oxidase in the mitochondrial DNA of *Acanthamoeba castellanii*. *J Mol Biol*. 257:1019–1030.
- Muramoto K, Hirata K, Shinzawa-Itoh K, Yoko-o S, Yamashita E, Aoyama H, Tsukihara T, Yoshikawa S. 2007. A histidine residue acting as a controlling site for dioxygen reduction and proton pumping by cytochrome *c* oxidase. *Proc Natl Acad Sci USA*. 104:7881–7886.
- Nedelcu AM, Lee RW, Lemieux C, Gray MW, Burger G. 2000. The complete mitochondrial DNA sequence of *Scenedesmus obliquus* reflects an intermediate stage in the evolution of the green algal mitochondrial genome. *Genome Res*. 10:819–831.
- Neupert W, Herrmann JM. 2007. Translocation of proteins into mitochondria. *Annu Rev Biochem*. 76:723–749.
- Pérez-Martínez X, Antaramian A, Vázquez-Acevedo M, Funes S, Tolkunova E, d'Alayer J, Claros MG, Davidson E, King MP, González-Halphen D. 2001. Subunit II of cytochrome *c* oxidase in chlamydomonad algae is a heterodimer encoded by two independent nuclear genes. *J Biol Chem*. 276:11302–11309.
- Richly E, Chinnery PF, Leister D. 2003. Evolutionary diversification of mitochondrial proteomes: implications for human disease. *Trends Genet*. 19:356–362.
- Rizzuto R, Sandonà D, Capaldi RA, Bisson R. 1991. Nucleotide sequence of a cDNA coding for the mitochondrial precursor protein of cytochrome *c* oxidase subunit IV from the slime mold *Dictyostelium discoideum*. *Biochim Biophys Acta*. 1090:125–128.
- Sickmann A, Reinders J, Wagner Y, et al. (13 co-authors). 2003. The proteome of *Saccharomyces cerevisiae* mitochondria. *Proc Natl Acad Sci USA*. 100:13207–13212.
- Smith DG, Gawryluk RM, Spencer DF, Pearlman RE, Siu KW, Gray MW. 2007. Exploring the mitochondrial proteome of the ciliate protozoon *Tetrahymena thermophila*: direct analysis by tandem mass spectrometry. *J Mol Biol*. 374:837–863.
- Waller RF, Keeling PJ. 2006. Alveolate and chlorophycean mitochondrial *cox2* genes split twice independently. *Gene*. 383:33–37.