Late Mitochondrial Acquisition, Really?

Mauro Degli Esposti[1,2,*]

[1]Italian Institute of Technology, Genoa, Italy

[2]Centre for Genomic Sciences, UNAM Cuernavaca, Mexico

*Corresponding author: E-mail: mauro.degliesposti@iit.it, mauro1italia@gmail.com.

Abstract

This article provides a timely critique of a recent Nature paper by Pittis and Gabaldón that has suggested a late origin of mitochondria in eukaryote evolution. It shows that the inferred ancestry of many mitochondrial proteins has been incorrectly assigned by Pittis and Gabaldón to bacteria other than the aerobic proteobacteria from which the ancestor of mitochondria originates, thereby questioning the validity of their suggestion that mitochondrial acquisition may be a late event in eukaryote evolution. The analysis and approach presented here may guide future studies to resolve the true ancestry of mitochondria.

**Key words:** eukaryogenesis, mitochondria, bacteria.

Introduction

Current views of eukaryotic cell evolution disagree on whether the endosymbiotic acquisition of mitochondria occurred early (Lane et al. 2013) or late (Ettema 2016; Pittis and Gabaldón 2016). A recent article in *Nature* concludes that mitochondrial endosymbiosis was a late event (Pittis and Gabaldón 2016). However, this conclusion is based upon phylogenetic inferences that challenge the accumulated knowledge of proteobacterial evolution. An early branching separated the anaerobic δ and ε proteobacteria from the predominantly aerobic or facultatively anaerobic α, β, and γ proteobacteria, which share ubiquinone (Q, fig. 1a and b) with eukaryotes (Aussel et al. 2014). In contrast, Pittis and Gabaldón (2016) report that only 13% of the eukaryotic protein families analyzed would be phylogenetically related to α, β, and γ proteobacteria. Remarkably, more eukaryotic families would be phylogenetically related to δ and ε proteobacteria (57) than to β (1) or γ (33) proteobacteria (Pittis and Gabaldón 2016), in open contrast with the established evolutionary sequence of prokaryotes and mitochondria (fig. 1a and b).

Results and Discussion

Phylogenetic analysis of a protein family with inferred ancestry to ε proteobacteria, D-fructose-1,6-bisphosphate 1-phosphohydrolase, shows that such an inference is correct only for a chloroplast form of the protein, while the major bifunctional form displays an ancestry with β and γ proteobacteria

(supplementary fig. S1, Supplementary Material online), in agreement with earlier results (Martin et al. 1996). Another contradiction with the established evolution of proteobacteria is the finding that the stem length of protein families with γ proteobacteria as their sister group is systematically longer than that of protein families with α proteobacteria as their sister group (Pittis and Gabaldón 2016). In my previous phylogenetic studies, proteins from α taxa have always shown deeper branches than those of γ taxa (Degli Esposti 2014; Degli Esposti et al. 2015). I have thus undertaken further phylogenetic analysis of the protein families listed as having different proteobacterial and bacterial ancestry (Pittis and Gabaldón 2016), which have well-known biochemical properties and reside in the proteome of mitochondria and related organelles (see Materials and Methods for details). While there appears to be concordance in the percentage of proteins with inferred ancestry to γ proteobacteria, there is a remarkable difference in the percentage of proteins inferred to have bacterial ancestry, as shown in figure 1c. Most of these proteins turn out to have an underlying ancestry to α, β, and γ proteobacteria—hereafter indicated as "aerobic proteobacteria"—once common instances of lateral gene transfer (LGT) from δ proteobacteria and other bacterial phyla are taken into account (Thiergart et al. 2012; Rochette et al. 2014). Accurate phylogenetic trees of key bioenergetic proteins such as COX1 (Degli Esposti 2014), cytochrome c_1 (fig. 1b), and cytochrome b (fig. 2) show that the mitochondrial (eukaryotic) orthologs are embedded
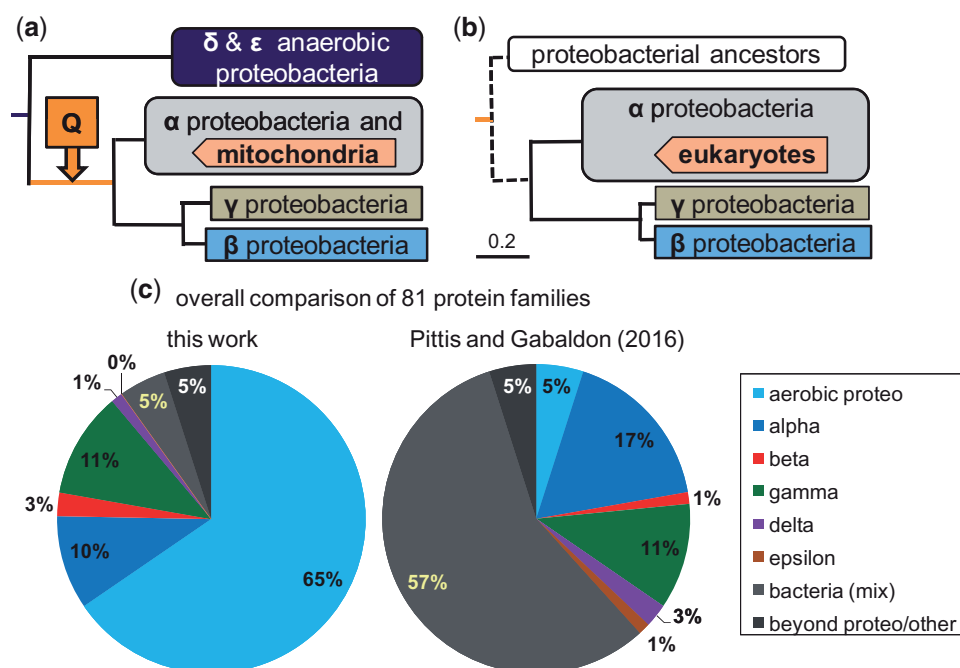
Fig. 1.—The drawings represent phylogenetic trees of proteobacterial and eukaryotic proteins. The accepted pattern for the evolution of proteobacteria and mitochondria is represented in (*a*), showing the appearance of the biosynthetic pathways for Q after the separation of the α, β, and γ proteobacteria (Aussel et al. 2014; Degli Esposti 2014; Gray 2015). (*b*) The phylogenetic pattern of aerobic proteobacteria is well matched by the NJ tree obtained with previously reported methods (Degli Esposti et al. 2015) for cytochrome c_1, a nuclear encoded subunit of mitochondrial bc_1 complex that coevolved with its Q substrate. All mitochondrial proteins of eukaryotes cluster in a branch that is embedded within the clade of their α proteobacterial homologs, as reported for COX subunits (Degli Esposti 2014). Bar indicates the fraction of amino acid changes estimated per unit of branch length (Degli Esposti et al. 2015). (*c*) The prokaryotic ancestry was deduced from the phylogenetic trees of 81 proteins that had been previously reported by Pittis and Gabaldón (2016). These proteins have been chosen and analyzed as described in the Materials and Methods section and their list is available upon request. Proteins with deduced ancestry to combinations of α, β, and γ proteobacteria as in (*b*) are abbreviated as "aerobic proteo"; this definition has been equally applied to the proteins reported by Pittis and Gabaldón (2016) that show any combination of aerobic proteobacteria in their ancestry. Proteins showing sister groups with orthologs from mixed proteobacterial and bacterial taxa are included under the label of "bacterial (mix)," corresponding to the broad definition of bacterial ancestry used by Pittis and Gabaldón (2016). Proteins showing clear ancestry only to bacteria are instead labeled "beyond proteo/others." The percentage values of the various categories did not significantly change by increasing the numbers of proteins compared. Of note, repeating the same phylogenetic analysis without the eukaryotic groups has shown α proteobacterial ancestry for 78% of the proteins, in agreement with earlier estimates of the intrinsically mosaic nature of proteobacterial genomes (Esser et al. 2007).

in a clade that includes all their homologs of α proteobacteria, in agreement with the established pattern of proteobacterial and mitochondrial evolution (fig. 1a and b). In turn, this clade is sister to a clade containing the homologs of β and γ proteobacteria (Degli Esposti 2014), thereby defining a general ancestry to aerobic proteobacteria for eukaryotic bioenergetic proteins coded by either mitochondrial, such as cytochrome *b* (fig. 2), or nuclear DNA, such as cytochrome c_1 (fig. 1b). Therefore, the true ancestry of mitochondrial bioenergetic proteins would be aerobic proteobacteria instead of the narrow taxonomic subdivisions followed by Pittis and Gabaldón (2016) and previous works—see Thiergart et al. (2012) and Degli Esposti (2014) for a review. I now found that this concept applies also to many proteins analyzed by Pittis and Gabaldón (2016) that reside in mitochondria (fig. 1c).

Overall, the proteins analyzed here show a common ancestry with aerobic proteobacteria, which exhibits different levels of resolution depending upon their phylogenetic signal and LGT interferences (Rochette et al. 2014). As an example of this situation for proteins inferred to pre-date mitochondrial endosymbiosis (Pittis and Gabaldón 2016), LGT interference is clearly observed in the phylogeny of nitrite reductase, a multidomain enzyme of fungi and heterokonts which is engaged in bioenergy production during anaerobiosis (Degli Esposti 2014). The monophyletic group of eukaryotic nitrite reductase have sister proteins from a few Planctomycetes and a single δ proteobacterium, forming a clade that is sister to a group containing homologs from α and γ proteobacteria, together with a single protein from the Bacteroidetes, *Flexithrix dorotheae* (supplementary fig. S2, Supplementary Material online). If one excludes this case and the similar one for the δ proteobacterial protein, the overall phylogeny of eukaryotic nitrite reductase conforms to a basal ancestry of aerobic proteobacteria intermixed with Planctomycetes. An equivalent situation has been encountered for cytochrome *b*, the
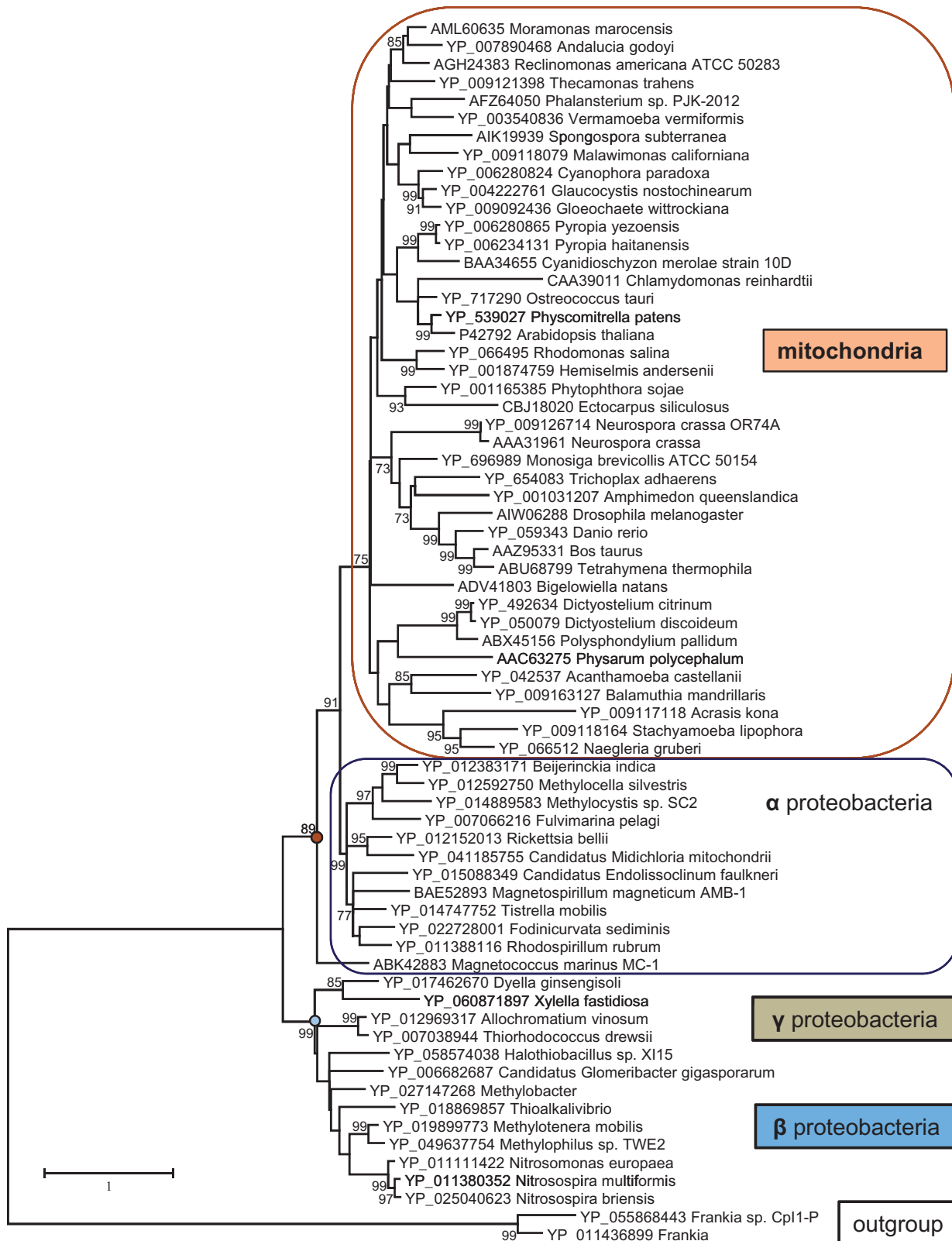
Fig. 2.—Phylogenetic tree of cytochrome *b* of the *bc₁* complex. The sequence of the cytochrome *b* encoded in the mitochondrial genome of the Jakobida *Moramonas* (Strassert et al. 2016) has been first used as a query in a DELTABLAST search extended to 1,000 proteins from all deep branching lineages of eukaryotes, as described in the Materials and Methods section. After accurate alignment (cf. Degli Esposti 2014) of a selection of 80 sequences

(continued)

mitochondrial homologs of which have a solid ancestry in aerobic proteobacteria because they are part of the bc_1 operon (Degli Esposti 2014), as also indicated by the trees shown here (figs. 1b and 2).

The most striking difference between the present results and those reported earlier (Pittis and Gabaldón 2016) consists in the much reduced proportion of proteins with inferred bacterial ancestry, that is, those that would have a mixture of orthologs from proteobacteria and other bacterial lineages as sister group. Multiple reasons could account for this, but methodological differences in the building and analysis of the phylogenetic trees appear to be minor, at least for the bioenergetic proteins examined here (figs. 1 and 2) and earlier (Degli Esposti 2014). A major reason, I surmise, could be the odd topology of fast evolving proteins in parasitic eukaryotic lineages such as Apicomplexa and Kinetoplastida, which often do not cluster together with other eukaryotic orthologs in maximum-likelihood trees, as documented here for cytochrome b (fig. 2). The addition of deep branching eukaryotic lineages not considered before (Rochette et al. 2014; Derelle et al. 2015; Pittis and Gabaldón 2016) often helps clustering these fast evolving proteins into a monophyletic clade, as predicted (Ku et al. 2015). Another reason is the presence of multiple isoforms for the eukaryotic orthologs of bacterial proteins, as in the case shown here in supplementary fig. S1, Supplementary Material online. Ultimately, the known limits of phylogenetic analysis based upon immediate sister topology (Rochette et al. 2014; Ku et al. 2015) may have produced fundamental differences between the results of Pittis and Gabaldón (2016) and those presented here (fig. 1).

In conclusion, many eukaryotic proteins inferred to have bacterial ancestry by Pittis and Gabaldón (2016) turn out to actually have an underlying ancestry with aerobic proteobacteria, as in the case of mitochondrial proteins (fig. 1). Therefore, several results reported by Pittis and Gabaldón (2016) appear to derive from incorrect inferences, seriously weakening their statistical analysis of stem length that has led to the conclusion that mitochondrial acquisition would be a late event in the evolution of eukaryotes. The symbiosis with proto-mitochondria is unlikely to have occurred in multiple steps either (Ku et al. 2015), contrary to what implied from the work of Pittis and Gabaldón (2016) (Ettema 2016).

Materials and Methods

A systematic approach has been followed to analyze the phylogenesis of various proteins showing orthologs in both prokaryotes and eukaryotes as considered by Pittis and Gabaldón (2016). First, proteins were collected from the published proteomes of mitochondria, hydrogenosomes, and mitosomes (Atteia et al. 2009; Barberà et al. 2010; Jedelský et al. 2011; Horváthová et al. 2012; Gawryluk et al. 2014; Calvo et al. 2016) and screened for their known structural and functional properties. Those with unknown function were discarded, together with the proteins that are involved in the interaction with DNA or RNA, which may show interference with archean orthologs having equivalent function (Rochette et al. 2014; Ku et al. 2015). The remaining proteins formed the MitoBac database, currently containing approximately 250 proteins with defined function in mitochondria and related organelles. The database includes more than 100 proteins previously analyzed by Pittis and Gabaldón (2016) and a dozen proteins used in earlier studies on the phylogenesis of eukaryotes (Derelle et al. 2015). The phylogenetic pattern of these proteins was deduced by following the robust phylogeny of mitochondria-encoded bioenergetic proteins that are common to all organisms with sequenced mtDNA (Kannan et al. 2014), namely cytochrome b (fig. 2) and COX1 (Degli Esposti 2014), as well as of nuclear encoded cytochrome c_1 (fig. 1b). The branching order of the major clades in the trees for these proteins was very similar using different methods (cf. fig. 2) and therefore could be used as a reference to verify the presence of LGT or other distorsions in the neighbor-joining (NJ) trees obtained from DELTABLAST searches (Boratyn et al. 2012). Such searches were initially applied to complex I subunits of the Nuo operon, which has been transmitted almost in its entirety to the mtDNA of protists (only the NuoE and NuoF subunits are regularly encoded in nuclear DNA—Kannan et al. 2014; Gray 2015). Except for the short NuoK subunit, the tree topology of the Nuo subunits was shown to essentially coincide with that of cytochrome b or c_1, once sporadic cases of LGT by δ proteobacteria (cf. Thiergart et al. 2012) or the insertion of orthologs from Bacteroidetes and Planctomycetes in the sister group of γ and β proteobacteria (supplementary fig. S2, Supplementary Material online) were taken into account, as previously described by Rochette et al. (2014). Equivalent DELTABLAST searches (Boratyn et al.

Fɪɢ. 2.—Continued

that would match the taxonomic distribution and branching order of the NJ tree obtained with all the results of the DELTABLAST (Degli Esposti et al. 2015), a maximum-likelihood (ML) tree was generated using the program MEGA 6 (Tamura et al. 2013). ProtTest was then used to select the best fit model based following the Akaike information criterion (Abascal et al. 2005). Branch support was evaluated with 500 bootstrap pseudoreplicates and is annotated in percentage values within the tree. Sequences from α proteobacteria plus mitochondria form a sister clade to their homologs in γ and β proteobacteria as in the trees obtained for cytochrome c_1 (fig. 1b) and COX1 (Degli Esposti 2014). Cytochrome b_6-like sequences from Actinomycetes have been used as the outgroup to root the tree. The cytochrome b sequences from Apicomplexa, for example, *Plasmodium berghei*, had to be removed from the initial selection for this tree because their large variation prevented clustering with other mitochondrial sequences in a single monophyletic clade as obtained with NJ trees.

2012) were then applied to other proteins of the MitoBac database using orthologs in *Methylocystis* species as queries. The taxonomic breath of such searches was extended to 1,000 sequences and included, in addition to the eukaryotic groups used earlier (Rochette et al. 2014; Derelle et al. 2015; Ku et al. 2015; Pittis and Gabaldón 2016), the deep branching lineages of Glaucophyta, Cryptomonads, Hypotrichous ciliates, Apusozoa, and Heterolobosea (Burki 2014). Complementary phylogenetic analysis was carried out with the programs MEGA5.2 and MEGA6 (Tamura et al. 2013) using selections of proteins that could match the general branching order found in the NJ trees that were obtained with all the results of the DeltaBLAST searches, as previously described (Degli Esposti et al. 2015); see figure 2 as an example. Parallel searches and analyses were then undertaken without eukaryotic lineages to verify bacterial mosaicism (Esser et al. 2007) in the proteome of *Methylocystis*, which was chosen as the α proteobacterial taxon of reference because of its possible closeness to proto-mitochondria (Degli Esposti 2014).

Acknowledgments

Supplementary Material

Supplementary figures S1 and S2 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Literature Cited

Abascal F., Zardoya R., Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21:2104–2105.

Atteia A., et al. 2009. A proteomic survey of *Chlamydomonas reinhardtii* mitochondria sheds new light on the metabolic plasticity of the organelle and on the nature of the alpha-proteobacterial mitochondrial ancestor. Mol Biol Evol. 26:1533–1548.

Aussel L., et al. 2014. Biosynthesis and physiology of coenzyme Q in bacteria. Biochim Biophys Acta. 1837:1004–1011.

Barberà MJ., et al. 2010. *Sawyeria marylandensis* (Heterolobosea) has a hydrogenosome with novel metabolic properties. Eukaryot Cell. 9:1913–1924.

Boratyn GM, et al. 2012. Domain enhanced lookup time accelerated BLAST. Biol Direct. 7:12.

Burki F. 2014. The eukaryotic tree of life from a global phylogenomic perspective. Cold Spring Harb Perspect Biol. 6:a016147.

Calvo SE, Clauser KR, Mootha VK. 2016. MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. Nucleic Acids Res. 44:D1251–D1257.

Degli Esposti M. 2014. Bioenergetic evolution in proteobacteria and mitochondria. Genome Biol Evol. 6:3238–3251.

Degli Esposti M., et al. 2015. Molecular evolution of cytochrome bd oxidases across proteobacterial genomes. Genome Biol Evol. 7:801–820.

Derelle R., et al. 2015. Bacterial proteins pinpoint a single eukaryotic root. Proc Natl Acad Sci U S A. 112:E693–E699.

Esser C, Martin W, Dagan T. 2007. The origin of mitochondria in light of a fluid prokaryotic chromosome model. Biol Lett. 3:180–184.

Ettema TJ. 2016. Evolution: mitochondria in the second act. Nature 531:39–40.

Gawryluk RM, Chisholm KA, Pinto DM, Gray MW. 2014. Compositional complexity of themitochondrial proteome of a unicellular eukaryote (*Acanthamoeba castellanii*, supergroup Amoebozoa) rivals that of animals, fungi, and plants. J Proteomics. 109:400–416.

Gray MW. 2015. Mosaic nature of the mitochondrial proteome: implications for the origin and evolution of mitochondria. Proc Natl Acad Sci U S A. 112:10133–10138.

Horváthová L., et al. 2012. Transcriptomic identification of iron-regulated and iron-independent gene copies within the heavily duplicated *Trichomonas vaginalis* genome. Genome Biol Evol. 4:1017–1029.

Jedelský PL., et al. 2011. The minimal proteome in the reduced mitochondrion of the parasitic protist *Giardia intestinalis*. PLoS One 6:e17285.

Kannan S, Rogozin IB, Koonin EV. 2014. MitoCOGs: clusters of orthologous genes from mitochondria and implications for the evolution of eukaryotes. BMC Evol Biol. 14:237.

Ku C., et al. 2015. Endosymbiotic origin and differential loss of eukaryotic genes. Nature 524:427–432.

Lane N, Martin WF, Raven JA, Allen JF. 2013. Energy, genes and evolution: introduction to an evolutionary synthesis. Philos Trans R Soc Lond B Biol Sci. 368:20120253.

Martin WF, Mustafa AZ, Henze K, Schnarrenberger C. 1996. Higher-plant chloroplast and cytosolic fructose-1,6-bisphosphatase isoenzymes: origins via duplication rather than prokaryote-eukaryote divergence. Plant Mol Biol. 32:485–491.

Pittis AA, Gabaldón T. 2016. Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. Nature 531:101–104.

Rochette NC, Brochier-Armanet C, Gouy M. 2014. Phylogenomic test of the hypotheses for the evolutionary origin of eukaryotes. Mol Biol Evol. 31:832–845.

Strassert JF., et al. 2016. *Moramonas marocensis* gen. nov., sp. nov.: a jakobid flagellate isolated from desert soil with a bacteria-like, but bloated mitochondrial genome. Open Biol. 6:150239.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 30:2725–2729.

Thiergart T, Landan G, Schenk M, Dagan T, Martin WF. 2012. An evolutionary network of genes present in the eukaryote common ancestor polls genomes on eukaryotic and mitochondrial origin. Genome Biol Evol. 4:466–485.

Associate editor: Bill Martin