

Multiple origins of eukaryotic *cox15* suggest a horizontal gene transfer from bacteria to jakobid mitochondrial DNA

Journal:	<i>Molecular Biology and Evolution</i>
Manuscript ID:	MBE-14-1252.R2
Manuscript Type:	Article
Date Submitted by the Author:	08-Sep-2015
Complete List of Authors:	He, Ding; Evolutionary Biology Centre, Uppsala University, Program in Systematic Biology Fu, Cheng-Jie; Evolutionary Biology Centre, Uppsala University, Program in Systematic Biology Baldauf, Sandra; Evolutionary Biology Centre, Uppsala University, Program in Systematic Biology
Key Words:	α -proteobacterium, mitochondrial evolution, heme A synthase (HAS), endosymbiotic gene transfer (EGT), horizontal gene transfer (HGT), last eukaryote common ancestor (LECA)

SCHOLARONE™
Manuscripts

Manuscript type: Article (Discoveries)

Multiple origins of eukaryotic *cox15* suggest a horizontal gene transfer from bacteria to jakobid mitochondrial DNA

Ding He, Cheng-Jie Fu* and Sandra L. Baldauf

Program in Systematic Biology, Department of Organismal Biology,

Uppsala University, 75236 Uppsala, Sweden

*Author for Correspondence:

Cheng-Jie Fu

Norbyvägen 18D, Evolutionary Biology Centre, 75236 Uppsala, Sweden

Tel: +46 (0)18-471-2777;

Fax: +46 (0)18-471-6452;

Email: chengjie.fu@ebc.uu.se; chengjie.fu@gmail.com

Abstract

The most gene-rich and bacterial-like mitochondrial genomes (mtDNAs) known are those of Jakobida (Excavata). Of these, the most extreme example to date is the *Andalucia godoyi* mtDNA, including a *cox15* gene encoding the respiratory enzyme heme A synthase (HAS), which is nuclear-encoded in nearly all other mitochondriate eukaryotes. Thus *cox15* in eukaryotes appears to be a classic example of mitochondrion-to-nucleus (endosymbiotic) gene transfer, with *A. godoyi* uniquely retaining the ancestral state. However, our analyses reveal two highly distinct HAS types (encoded by *cox15-1* and *cox15-2* genes) and identify *A. godoyi* mitochondrial *cox15*-encoded HAS as type-1 and all other eukaryotic *cox15*-encoded HAS as type-2. Molecular phylogeny places the two HAS types in widely separated clades with eukaryotic type-2 HAS clustering with the bulk of α -proteobacteria (>670 sequences), while *A. godoyi* type-1 HAS clusters with an eclectic set of bacteria and archaea including two α -proteobacteria missing from the type-2 clade. This wide phylogenetic separation of the two HAS types is reinforced by unique features of their predicted protein structures. Meanwhile, RNA-seq and genomic analyses fail to detect either *cox15* type in the nuclear genome of any jakobid including *A. godoyi*. This suggests that not only is *cox15-1* a relatively recent acquisition unique to the *Andalucia* lineage, but the jakobid last common ancestor probably lacked both *cox15* types. These results indicate that uptake of foreign genes by mtDNA is more taxonomically widespread than previously thought. They also caution against the assumption that all α -proteobacterial-like features of eukaryotes are ancient remnants of endosymbiosis.

Key words

α -proteobacterium, mitochondrial evolution, heme A synthase (HAS), endosymbiotic gene transfer (EGT), horizontal gene transfer (HGT), last eukaryote common ancestor (LECA)

Introduction

Despite continued debate, there is still abundant evidence to support the theory that mitochondria arose by endosymbiosis of a bacterium (Zimorski et al. 2014). The best candidate source for this bacterium is also still α -proteobacteria (Gray 2012; Müller et al. 2012; Esposti 2014). This is primarily because most of the few remaining genes in mitochondrial genomes (mtDNAs) show α -proteobacterial affinity (Andersson et al. 2003; Esser et al. 2004; Fitzpatrick et al. 2006; Williams et al. 2007; Rodríguez-Ezpeleta and Embley 2012), as do 10-20% of the nuclear genes that encode 95-99% of the mitochondrial proteome (Szkłarczyk and Huynen 2010; Thiergart et al. 2012; Gray 2015). Thus α -proteobacterial-like nuclear genes encoding mitochondrial proteins are presumed to have originated by transfer of genetic material from the mitochondrion to the nucleus (endosymbiotic gene transfer or EGT) (Timmis et al. 2004). Although most functional EGT appears to have occurred before the last eukaryote common ancestor (LECA), this is clearly still an active evolutionary process (Adams and Palmer 2003; Bock and Timmis 2008; Kleine et al. 2009; Wang et al. 2012).

Unique among eukaryotes, Jakobida (Excavata) have strikingly gene-rich mtDNAs with bacterial-like organization. These include relicts of nearly canonical bacterial operons, a multi-subunit bacterial-type RNA polymerase and putative Shine-Dalgarno motifs preceding many genes (Lang et al. 1997). The most extreme

example to date of these gene-rich mtDNAs is that of *Andalucia godoyi* (Burger et al. 2013). This mtDNA carries 66 protein-coding genes including a *cox15* gene encoding heme A synthase (HAS), which is nuclear encoded in all other eukaryotes. HAS is a key enzyme in aerobic respiration catalyzing the terminal step in heme A synthesis (Glerum et al. 1997; Antonicka et al. 2003; Hederstedt 2012).

Heme A functions as a key prosthetic group in subunit I of the cytochrome c oxidase complex (CcO, COX or Complex IV), the terminal acceptor complex in the mitochondrial electron transport chain (Kim et al. 2012). Heme A is synthesized in the mitochondrion from heme B (protoheme) in two consecutive steps, the first catalyzed by heme O synthase (HOS, encoded by *cox10*) and the second by HAS (Barros et al. 2001). Within the mitochondrion, HAS is found embedded in the inner mitochondrial membrane where it is closely associated with CcO assembly intermediates (Bareth et al. 2013). Structurally, HAS is dominated by an eight helix, transmembrane pore-forming domain built from two tandem copies of a four-helix unit (Hederstedt 2012).

Since nuclear *cox15* is nearly universal among eukaryotes and shows strong affinity for its α -proteobacterial homolog (*ctaA*; Hannappel et al. 2012), it appears to be a classic example of EGT with *A. godoyi* mtDNA uniquely retaining the ancestral state (Burger et al 2013). However, we find strong evidence for two very different types of HAS, and these are only very distantly related to each other. Moreover, *A. godoyi* mitochondrial *cox15* appears to encode one type of HAS while all other eukaryotic nuclear *cox15* genes encode the other HAS type. This distinction is supported by molecular phylogenetic and protein structural analyses. We also find evidence suggesting that the eukaryotic nuclear type *cox15* was lost early in jakobid evolution. Together these data indicate an endosymbiotic origin for eukaryotic

nuclear *cox15* but a relatively recent, non-endosymbiotic origin for *A. godoyi* mitochondrial *cox15*.

Results

Multiple origins of mitochondrial heme A synthase (*cox15*)

All detectable homologs of HAS were retrieved using a wide variety of *cox15* deduced amino acid sequences as queries. Sequences were found in nearly all aerobic α -proteobacteria, a taxonomically broad scattering of Archaea and other Bacteria, and all examined mitochondriate eukaryotes except rhodophyte and glaucophyte algae (Supplementary table S1 and Supplementary data file). Phylogenetic analyses of the full set of sequences identify two highly distinct clades. These clades are separated from each other by a long unbroken branch that receives 100% maximum-likelihood bootstrap (mlBP) support and 1.0 Bayesian inference posterior probability (biPP) (Figure 1). To simplify nomenclature, the proteins in these two clades are referred to from here on as type-1 and type-2 HAS and their respective encoding genes as *cox15-1* and *cox15-2*.

The type-1 HAS clade includes an assortment of aerobic Archaea, a small but taxonomically diverse sampling of Bacteria including two α -proteobacteria and a single eukaryotic sequence, that of *A. godoyi* mitochondrial *cox15* (Figure 1). The type-2 cluster includes all eukaryotic nuclear *cox15* encoded sequences along with the vast majority of α -proteobacteria, a smattering of other proteobacteria and a single clade of Bacteroidetes (Figure 1). The eukaryotic type-2 clade includes representatives of a wide taxonomic sampling of mitochondriate eukaryotes, specifically Amorphea (Holozoa, Fungi, Amoebozoa), Diaphoretickes (Viridiplantae, Stramenopila, Alveolata, Rhizaria, Cryptophyta, Haptophyta) and Discoba

(Discicristata, Tsukubamonadida) (Adl et al. 2012; He et al. 2014) (Figure 1, Supplementary figure S1 and Supplementary table S1). Thus all detectable HAS sequences form two distinct clades separated by a large evolutionary distance with all examined eukaryotic nuclear sequences identified as *cox15-2* and the *A. godoyi* sequence as a *cox15-1*. Moreover, all eukaryotic *cox15-2* encoded HAS sequences form a well-supported sub-clade within the type-2 subtree (76-84% mIBP, 0.99 biPP, Figure 1), while *A. godoyi cox15-1* encoded HAS is grouped with a subset of bacteria within the type-1 subtree, also with strong support (Hillis and Bull 1993; Alfaro et al. 2003) (73% mIBP, 0.99 biPP, Figure 1).

Unique structural features of the two HAS types

The two HAS types show distinct structural features including characteristic functional residues and type-specific loop regions (Figure 1 and Figure 2). The most distinctive difference between the HAS types is a pair of cysteine residues (C1 and C2) that are universal among type-1 HAS but missing from all type-2 sequences (Supplementary figure S2). Site-directed mutagenesis in the gram-positive bacterium *Bacillus subtilis*, which uses a type-1 HAS, indicate that these residues are critical for enzyme function in this organism (Mogi 2009; Hederstedt 2012). Nonetheless, the cysteine-depleted type-2 HAS, where found also appears to be sufficient for heme A synthesis, e.g., the *cox15* gene of the protistan parasite *Trypanosoma cruzi* (Excavata) restores heme A synthesis to a *cox15* knockout of its distant relative the yeast *Saccharomyces cerevisiae* (Buchensky et al. 2010).

The two HAS types also show markedly different length variation, particularly in the hydrophilic loop region between the first two α -helices ($\alpha 1$ and $\alpha 2$) (Supplementary figure S3). This region is nearly invariant in length in type-2 HAS but

highly variable in type-1 ([Supplementary figure S2](#)) and appears to be important for protein-protein interactions (Hannappel et al. 2012). Thus, insertions or deletions in this loop region of the protein structure may relate to differences in the complement of cytochrome c oxidase biogenesis co-factors with which the two different HAS types interact (Khalimonchuk and Rödel 2005; Bareth et al. 2013).

Absence of *cox15* in jakobid nuclear DNA

The *cox15* composition of jakobid nuclear genomes was investigated using PCR and transcriptomics. PCR assays were conducted on total DNA from three aerobic jakobid species - *A. godoyi*, *Seculamonas ecuadoriensis* and *Jakoba libera*, which together cover the breadth of known jakobid taxonomic diversity (Lara et al. 2006; Burger et al. 2013). Assays were conducted using several sets of degenerate PCR primers for both *cox15* types ([Supplementary table S2](#)). These assays readily identified *A. godoyi cox15-1*, and direct sequencing of the PCR product confirms that the amplicon is *cox15-1* and shows no evidence of a mixed signal in the sequence. However, no evidence was found for a *cox15-2* gene in any of the jakobids, nor was evidence found for a *cox15-1* gene in any jakobid other than *A. godoyi*. RT-PCR did confirm active transcription of *A. godoyi cox15-1* ([Supplementary figure S4](#)) as recently reported elsewhere (Valach et al. 2014).

Since PCR detection may fail for reasons other than absence of the target sequence, we also performed high-throughput RNA-seq analysis of *A. godoyi* and *S. ecuadoriensis* mRNA. RNA-seq has proven highly effective for detecting even low-level expressed genes (Haas et al, 2012). The final assembled transcriptomes consist of 20,146 transcripts for *A. godoyi* (18,541 predicted genes with average length of 1,101 bp) and 32,071 transcripts for *S. ecuadoriensis* (26,835 genes with

average length of 925 bp) (Fu C-J, He D, Baldauf SL, to be published elsewhere). Both assemblies are estimated to be over 80% complete based on their coverage of the 248 ultra-conserved core eukaryotic genes (CEGMA; Parra et al. 2007). However, no transcript corresponding to a possible homolog of either *cox15-1* or *cox15-2* was detected in either of the two jakobid transcriptomes with BLAST or HMM searches using an extremely permissive e-value cut-off of 1000 (10^3) ([Supplementary data file](#)). Likewise, no *cox15* related sequences were detected in the limited available EST data for other jakobids including *J. libera*, *J. bahamiensis*, *Reclinomonas americana* and *Histiona aroides* (TBestDB, O'Brien et al. 2007). Thus, the known diversity of jakobids appears to lack both *cox15* types, with the exception of *A. godoyi*, which, unique among eukaryotes, possesses *cox15-1* and carries it exclusively in mtDNA.

Nonetheless, transcripts for other conserved factors required for CcO complex assembly, such as *cox10* and *shy1* (Bareth et al. 2013) are easily detected in both the *A. godoyi* and *S. ecuadoriensis* transcriptomes. Furthermore, the *cox10* protein product HOS has been shown to form a 1:1 complex with HAS in the bacteria *Rhodobacter sphaeroides* and *B. subtilis* (Brown et al. 2004). Although HOS and HAS are thought to be unlinked *in vivo* and differentially regulated in *S. cerevisiae* (Barros and Tzagoloff 2002), the mRNA level of *cox15* in *S. cerevisiae* is in fact nearly 10-fold in excess over that of *cox10* (Wang et al. 2009). Thus, the fact that *cox10* is readily detected in the two jakobid transcripts further suggests that the absence of a *cox15* transcript in these data is not likely to be due to a low level of transcription ([Supplementary data file](#)).

Genomic context of *A. godoyi* mitochondrial *cox15*

Since *A. godoyi* *cox15-1* appears to represent a relatively recent acquisition, we inspected its genomic context relative to that of other jakobid mtDNAs. *A. godoyi* *cox15-1* is embedded within a conserved tRNA cluster but with unique flanking tRNA genes. These include a *trnT*, which is not found in other jakobid mtDNAs (Burger et al. 2013) (Figure 3). Moreover, unlike other genes in this cluster, *A. godoyi* *cox15-1* occurs on the anti-sense DNA strand as well as in the opposite orientation from almost all protein-coding sequences in the flanking regions (Figure 3). Nonetheless, analyses of codon usage bias and position-specific GC content show that the patterns for *cox15-1* are consistent with those of the other protein-coding sequences in *A. godoyi* mtDNA (Fu et al. 2014). Although *cox15-1* lacks a putative Shine-Dalgarno-like ribosome binding motif, these are also lacking in 14 other predicted protein-coding genes in *A. godoyi* mtDNA (Burger et al. 2013).

DISCUSSION

The origin of eukaryotic cox15 genes

We find that the CcO biogenesis factor HAS and its encoding gene *cox15* exist in two different forms with very distinct structural features (Figure 2) and phylogenetic histories (Figure 1). Among eukaryotes, the *cox15-2* gene is nearly universal and always nuclear-encoded, and its prokaryotic closest relatives include nearly all detected α -proteobacteria sequences. In contrast, *cox15-1* is found in eukaryotes exclusively in *A. godoyi* mtDNA. Furthermore, molecular phylogeny places *A. godoyi* *cox15-1* far outside of the eukaryotes *cox15-2* clade, together with a taxonomic scattering of bacterial and archaeal sequences type-2 clade (Figure 1). Together these data lead to three alternative evolutionary scenarios for *cox15* evolution in eukaryotes, depending primarily on the *cox15* complement of LECA, i.e.,

whether LECA had A) only *cox15-1*, B) both *cox15* types, or C) only *cox15-2* (Figure 4).

A) If LECA had only *cox15-1*, *A. godoyi* would be the only eukaryote to retain the original endosymbiotically derived *cox15* (Figure 4A). This would mean that *cox15-2* was acquired by eukaryotes by HGT from bacteria, and also spread across eukaryotes by a minimum of one additional HGT event. The latter is required in order to accommodate the strongly supported monophyly of eukaryotic *cox15-2* and its disjunct distribution in eukaryotes (Figure 1) – *i.e.* the absence of *cox15-2* in jakobids but presence in at least two major lineages branching separately from Jakobida, Discicristata + Tsukubamonadida and Amorphea + Diaphoretickes (Adl et al. 2012) (Figure 4A). It should be noted that this assumes a neozoan-excavate root for eukaryotes (He et al. 2014); any other root requires additional HGT events.

B) If LECA had both *cox15* types, this would at least eliminate the need for HGT (Figure 4B). However, this scenario would also require co-existence and selective pressure to maintain two HAS types in the same cell and for an extended period of time – at a minimum, from the advent of mitochondria through the divergence of Discoba from its last common ancestor with Amorphea + Diaphoretickes and further within Discoba until the divergence of the unique ancestor of *Andalucia* from the remaining Jakobida (Figure 4B). However we find only a single example of co-occurrence of both *cox15* types in the many hundreds of prokaryotes possessing these genes. Both type-1 and type-2 sequences are found in the ζ -proteobacterium SCGC AB-604-B04 (WP_018280647 and WP_026194051, respectively). However the former sequence lacks the first two of the four otherwise universal cysteine residues that we find to be a hallmark of type-1 HAS and which appear to be critical for its function (Mogi 2009; Hederstedt 2012) (Figure 2).

C) Thus the simplest scenario for *cox15* evolution in eukaryotes is Scenario C - a relatively recent acquisition of *cox15-1* by *A. godoyi* mtDNA (Figure 4C). The facts that eukaryotic *cox15-2* is nearly universal among eukaryotes, shows a single common origin with strong support and is closely related to the homologous gene in the vast majority of α -proteobacteria (Figure 1 and Figure 2), together strongly argue for *cox15-2* as the ancestral eukaryotic protein likely inherited from the mitochondrion. Meanwhile, *cox15-1* is found in no eukaryote other than *A. godoyi*, including a large taxonomic diversity of jakobid mtDNAs (Burger et al. 2013) and a smaller but still taxonomically broad sampling of jakobid nuclear DNAs (this study). Thus, the simplest and most likely explanation of *cox15* distribution in eukaryotes is an endosymbiotic origin of *cox15-2* followed by a single relatively recent acquisition of a bacterial *cox15-1* by *A. godoyi* mtDNA (Figure 4C).

It should be noted that within the *cox15-1* clade, the *A. godoyi* sequence groups strongly with two α -proteobacteria, *Geminicoccus roseus* and *Tistrella mobilis* plus several unclassified environmental sequences (73% mIBP, 0.99 biPP; Figure 1). However, the two named α -proteobacterial species in the type-1 clade are not closely related to each other (Ferla et al. 2013), and all other examined close relatives of both species (Rhodospirillales) are found exclusively in the type-2 clade. Thus, it is highly unlikely that *A. godoyi cox15-1* entered eukaryotes via mitochondrial endosymbiosis. In fact, both *cox15* types show evidence of HGT in bacteria, as β -, and γ -proteobacteria are also split between the type-1 and type-2 clades along with scattered representatives of additional bacterial groups. It should be noted that rooting the HAS tree with Archaea (Figure 1), which have a presumably ancestral-like HAS with only a single copy of the four-helix transmembrane domain (Lewin and Hederstedt 2006) only strengthens support for scenario C (Figure 4C).

Analogous HAS activity in jakobids?

The apparent total absence of *cox15* in most jakobids is potentially very problematic. Cytochrome c oxidase is a multi-protein complex essential for mitochondrial electron transport, and it is also the only respiratory complex that requires heme A as a co-factor (Kim et al. 2012). Therefore all aerobic eukaryotes should possess HAS activity. However, our BLAST/HMM searches indicate that *cox15* is also missing from the completely sequenced genomes of one glaucophyte and five rhodophyte algae ([Supplementary table S1](#) and [Supplementary data file](#)). Nonetheless, heme A presence has been demonstrated in at least one of these algae, *Galdieria sulphuraria* (Weinstein and Beale 1984). This suggests the presence of an analog of HAS in this genome. It should be noted that HAS does not show a generally high level of sequence conservation and only a few residues appear to be widely conserved and essential for activity ([Figure 2](#), [Supplemental Figures S2](#) and [S3](#)), which is not unusual for a membrane-spanning protein (Sharma and Wikström 2014). Therefore it may not be altogether surprising that HAS activity could evolve multiple times, either *de novo* or by minor modification of an existing membrane protein.

Although some bacteria can functionally replace heme A with other heme groups in heme-copper oxidases involved in aerobic respiration (Sone and Fujiwara 1991; Contreras-Zentella et al. 2003), this is still expected to reduce enzyme activity (Hederstedt 2012). Furthermore, inspection of CcO subunit 1 (COX1) sequences from across eukaryotes indicates that all the key metal-binding sites are highly conserved in jakobids, as well as in rhodophyte and glaucophyte algae ([Supplementary figure S5](#)). This may suggest that these COX1 subunits are binding

a normal heme A. Evidence of an analogous enzyme substituting for an essential organellar function has been recently demonstrated for mitochondria-related organelles (MROs). In this case, the Iron–sulfur cluster (ISC) synthesizing system, the only essential function of yeast mitochondria and probably at least some MROs (Lill and Mühlenhoff 2008), can be replaced by acquisition of an analogous system via HGT (Maralikova et al. 2010; Nývltová et al. 2013; Stairs et al. 2014).

The case for mitochondrial HGT

Although the idea of a protein-coding gene of foreign origin being inserted and functionally integrated into a mitochondrial genome may seem problematic, DNA uptake by plant mitochondria has been known for some time. This includes nuclear and chloroplast sequences from the same cell as well as mtDNA sequences from distant plant relatives (e.g., Bergthorsson et al. 2003; Hao et al. 2010; Wang et al. 2012; Rice et al. 2013). Isolated plant mitochondria have also been shown to actively take up foreign DNA in culture (Koulintchenko et al. 2003; Milesheina et al 2011), and mammalian and yeast mitochondria have also been shown to be naturally competent (Koulintchenko et al. 2006; Wu et al. 2015). In fact cancer cells depleted of mtDNA are able to acquire mtDNA from neighboring normal cells, thereby restoring oxidative phosphorylation, although the mechanism of DNA uptake is still unclear (Tan et al. 2015; Maiuri and Kroemer 2015).

However, other than plants, only two cases of protein-coding gene acquisition by mtDNA have been reported. These are a DNA adenine methylase (DAM) gene in the mtDNA of the haptophyte *Emiliania huxleyi* (Sánchez-Puerta et al. 2004) and a *MutS* gene potentially involved in self-sufficient DNA mismatch repair found in octocoral mtDNAs (Bilewitch and Degnan 2011). In the case of the octocoral *mtMutS*

gene, the origin is unclear as the gene shows only very weak similarity to a sequence from a single ϵ -proteobacterium or possibly a DNA virus (Bilewitch and Degnan 2011). There is also no clear indication of the origin of the *Emiliana* gene. In addition, in both cases, the functional significance of these novel genes also remains unclear.

The physical mechanism of prokaryote-to-eukaryote HGT is intriguing and still largely unknown (Bruto et al. 2013; Huang 2013), although heterotrophic protists should have ample opportunity to acquire genes from their bacterial food (Doolittle 1998) or frequent endosymbionts (Nowack and Melkonian 2010). While the growing list of examples identified to date has involved HGT of bacterial genes to eukaryotic nuclei (Keeling and Palmer 2008; Andersson 2009; Schönknecht et al. 2014), there is no theoretical reason why mtDNA should be exempt. Thus it is perhaps not entirely surprising that some mtDNAs should be able to acquire and express foreign genes. These results further caution against assumptions that all bacterial-like features of mitochondria are endosymbiotically derived, even those that appear to trace to α -proteobacteria (Richards and Archibald 2011; He et al. 2014).

Materials and Methods

Initial sequence vetting

The deduced amino acid sequence of *A. godoyi* mitochondrial *cox15* (GenBank accession number YP_007890502) and heme A synthase (HAS) homologs from five experimentally determined mitochondrial proteomes ([Supplementary table S1](#)) were used to query the NCBI, JGI Genome Portal (Nordberg et al. 2014) and Integrated Microbial Genomes (Markowitz et al. 2012) databases for homologs of selected eukaryotic representatives at the “Phylum” level ([Supplementary table S1](#)) using BLASTp (e-value < 1e-20) or tBLASTn (e-value < 1e-5). The retrieved eukaryotic

HAS sequences were then used to query the same databases for bacterial and archaeal homologs using the same procedure. The major α -proteobacterial lineages Rickettsiales, Rhodospirillales, Rhodobacterales, Polymorphum, SAR11, SAR116, Caulobacterales, Parvularculales, Rhizobiales, Sphingomonadales, Geminicoccus and Micavibrio were queried individually.

Hidden Markov model (HMM)-based homology search

More sensitive homology searches were conducted using the HMMER package (Finn et al. 2011). To maximize sensitivity, we built separate HMM profiles from alignments of each HAS type using hmmbuild with default settings. The two resulting HAS type-specific HMMs were then used to query the deduced amino acids sequences of transcriptomic assemblies of *A. godoyi* and *S. ecuadoriensis*, and complete genome assemblies of five rhodophytes (*Pyropia yezoensis*, *Porphyridium purpureum*, *Cyanidioschyzon merolae*, *Galdieria sulphuraria*, and *Chondrus crispus*) and one glaucophyte (*Cyanophora paradoxa*). These hmmsearches employed a less stringent e-value threshold of 10^3 (default is 10) for both whole sequence (-E) and domain (-domE) criteria ([Supplementary data file](#)). All hits were aligned with the respective HMM-source alignment using hmmsalign to verify their homology.

Phylogenetic analyses

Sequences were aligned using MAFFT 7.130b with default settings (Katoh and Standley 2013). Regions of ambiguous alignment were masked using either GBlocks (Talavera and Castresana 2007) for control trees or SeqFIRE for the final alignment (Ajawatanawong et al. 2012), both with default settings. Control phylogenies to identify obvious in-paralogs and redundant sequences were constructed using

PhyML with the LG amino-acid substitution model (Le and Gascuel 2008), four discrete gamma rate categories and empirical amino acid frequencies. HAS sequences from all examined Alveolata (eukaryotes), Rickettsiales and SAR11 (α -proteobacteria) lineages show sequence signatures consistent with type-2 HAS, but these sequences were excluded from the final set of phylogenetic analyses to avoid potential artifact caused by the extreme AT-bias and/or fast evolutionary rates of these sequences (Rodríguez-Ezpeleta et al. 2007; Viklund et al. 2012).

After removing in-paralogs and redundant sequences, trees were constructed using maximum likelihood with bootstrapping using RAxML 8.0.0 and the same model settings as above. Analyses consisted of at least 100 rapid bootstrap replicates as determined automatically with the built-in majority rule criterion (Stamatakis 2014). Bayesian inference analysis utilized two independent chains in PhyloBayes MPI 1.5a with the model setting of empirical profile mixture C20 + GTR + 4 discrete gamma rate categories (Lartillot et al. 2013). Chain convergence was diagnosed based on a comparison of the frequency of all bipartitions after discarding ~50% of the cycles as burn-in (35000 trees were discarded from both chains, with 23903 trees remaining in one chain and 35073 trees in the other, maxdiff = 0.26).

Protein structural analysis

Transmembrane helix structures were predicted for all deduced protein sequences using TMHMM Server v. 2.0 (cbs.dtu.dk/services/TMHMM/) and drawn with the online tool TOPO2 (sacs.ucsf.edu/TOPO2/). The predicted structure for the *A. godoyi* mitochondrial *cox15* encoded HAS was modeled onto the structure of the cytochrome b6 homodimer, its closest relative with a determined structure (Top HMM hit, sequence identities = 17%). Cytochrome b6 is structurally similar to the core of

mitochondrial cytochrome bc1 (Complex III). These complexes perform analogous catalytic reactions and show strong similarity in their respective heme-binding transmembrane domains (Baniulis et al. 2008). The crystal structure of cyanobacteria *Nostoc* sp. PCC 7120 Cyt b6 was retrieved from the RCSB Protein Data Bank (PDB ID: 4ogq_A), trimmed to the transmembrane domains, and residues were then modified to produce the type-1 HAS with histidines H1-H4 using Swiss-PdbViewer v4.1.0 (Guex and Peitsch 1997). *A. godoyi* HAS residues (H⁷⁶, H¹⁴³) corresponding to the first pair of highly conserved histidine residues in nearly all HAS homologs (H1, H2) were then superimposed onto the four transmembrane helix core. Positions of residues (C⁴⁰, C⁴⁶) corresponding to a pair of nearly invariant cysteine residues in type-1 HAS (C1, C2) are located within the flexible loop region between α 1 and α 2. Multiple sequence alignment of mitochondrial COX1 was conducted with Expresso (Tommaso et al. 2011), using bovine heart CcO protein structure as the template (PDB ID: 1occ_A).

Indel analysis

Indels were extracted from an unedited HAS alignment including a broad taxonomic sampling of 36 eukaryotes, 28 α -proteobacteria, 36 non- α -proteobacteria, 20 archaea and seven metagenomic sequences. Protein indels were identified using the combined indel/conserved block option in SeqFIRE (Ajawatanawong et al. 2012). Conserved regions were first identified using a 40% gaps cut-off, 75% amino acid conservation threshold and three-residue minimum conserved and maximum non-conserved block sizes. Indel regions were then refined using the partial/twilight treatment to accommodate incomplete and divergent sequences using a minimum inter-indel space of three. This procedure identified an indel signature that retains the

1
2
3 loop regions between C2 and H1 ([Supplementary figure S2](#)) and rejects a superficial
4
5 similarity predicted to be uniquely shared by *A. godoyi* and *T. mobilis* based on more
6
7 limited taxon sampling (Burger et al. 2013).
8
9

10 11 **Cell culture and nucleic acid extraction**

12 *Andalucia godoyi* (strain “And28”) and *Seculamonas ecuadoriensis* (ATCC No.
13
14 50688) cells were grown in Page's modified Neff's amoeba saline (Page 1988)
15
16 supplemented with LB (1:300) in 2000 ml flasks and shaken on a rotary shaker (80
17
18 cycles/min) at room temperature. Cells were fed with pre-cultured live *Klebsiella sp.*
19
20 as food and harvested in 50 ml corning tubes after 72 hours at a cell density of
21
22 approximately 2×10^5 /ml. Total DNA was extracted as described in (Lara et al. 2006).
23
24 Total RNA was extracted using TRI Reagent LS (Sigma-Aldrich) according to the
25
26 manufacturer's protocol.
27
28
29
30
31
32
33

34 **RNA sequencing and data processing**

35
36 Stranded mRNA transcriptomic libraries of *A. godoyi* and *S. ecuadoriensis* were
37
38 constructed using Illumina's Truseq RNA sample prep kit and sequenced on high-
39
40 throughput illumina platforms (Miseq and HiSeq2000, respectively; Macrogen, South
41
42 Korea). The paired-end raw reads were treated to remove adapter sequences and
43
44 low quality bases with Trimmomatic v.0.32 (Bolger et al. 2014) with default settings.
45
46 *De novo* transcriptome assemblies were generated using Trinity following the
47
48 protocol documented in (Grabherr et al. 2011) with the --jaccard_clip option set to
49
50 yes and all other parameters set to default. The transcriptomic sequencing data of
51
52 *Tsukubamonas globosa* (454 Titanium platform, Roche) was retrieved from GenBank
53
54
55
56
57
58
59
60

(Accession No. DRR014073), and transcripts were assembled using Newbler (v2.6, Roche).

cDNA synthesis and RT-PCR

For cDNA synthesis, 0.5 µg *A. godoyi* total RNA was first treated with DNase (Thermo). The reverse transcription reaction was performed at 40 °C for 30 min with the hexanucleotide random primer mix included in the Phusion RT-PCR Kit (Thermo). PCR amplification of cDNA products was performed using the Phusion High-Fidelity DNA Polymerase (Thermo) using two sets of primer pairs (Ag_cox15) (Supplementary table S2). PCR reaction conditions were: 1 min at 98 °C, 30 cycles of 5 sec at 98 °C, 10 sec at 65 °C and 20 sec at 72 °C; and 72 °C for 5 min. PCR amplicons were cleaned with ExoSap-IT (GE Healthcare) and sent for direct sequencing (Macrogen, South Korea). The *cox1* gene was used as a reference gene for semi-quantitative measures of relative transcriptional levels, using a set of primer pairs (Ag_cox1) (Supplementary table S2).

Degenerate PCR assay for *cox15* genes

PCR amplification using degenerate primers on jakobid (*A. godoyi*, *S. ecuadoriensis* and *J. libera*) total DNA was performed using the Phusion High-Fidelity DNA Polymerase (Thermo). Degenerate primer sets were designed based on separate multiple sequence alignments of each *cox15* type (Supplementary table S2). A touchdown PCR approach was used with annealing temperature stepdowns of 0.5 °C per cycle from 70 °C to 63 °C. Cycling conditions were as follow: 30 sec at 98 °C, 14 cycles of 5 sec at 98 °C and 30 sec at 72 °C, followed by 25 cycles of 5 sec at 98 °C, 10 sec at 63 °C and 30 sec at 72 °C, followed by a final polishing step at 72 °C for 5

min. Total DNA of *Acrasis kona* (Fu et al. 2014) was used for the control reaction [cox15-1(-), cox15-2(+)]. PCR products were analyzed by 1.5% agarose gel electrophoresis, and all even-approximately correctly sized fragments were isolated and cleaned with ExoSap-IT (GE Healthcare) and sent for direct sequencing (Macrogen, South Korea).

Supplementary Material

Supplemental Information includes [Supplementary tables S1-S2](#), [Supplementary figures S1-S5](#) and [Supplementary data file](#).

Acknowledgements:

We thank Alastair Simpson (Dalhousie University) for kindly providing the *Andalucia godoyi* culture and *Jakoba libera* DNA, Anders Larsson (Uppsala University) for help setting up the online BLAST server. We thank Cyberinfrastructure for Phylogenetic Research (CIPRES, phylo.org) and the Uppsala Multidisciplinary Center for Advanced Computational Science (Uppmax, uppmax.uu.se) for the use of computational resources and the Joint Genome Institute (jgi.doe.gov) and iMicrobe project (data.imicrobe.us) for the use of unpublished sequence data. C.-J. F., D.H. conceived and designed the study, with guidance from S.L.B. D.H. performed phylogenetic and indel analysis. C.-J. F. conducted all experiments, performed transcriptome assembly and protein structural analysis. C.-J. F., D.H., and S.L.B. wrote the paper. This work was supported by grants from the Swedish Research Council (Vetenskapsrådet) to S.L.B. C.-J.F. and S.L.B. are members of the Uppsala Centre for Evolution and Genomics (UCEG) research environment at Uppsala University.

Reference:

Adams, K. L., and J. D. Palmer. 2003. Evolution of mitochondrial gene content: Gene loss and transfer to the nucleus. *Molecular phylogenetics and evolution* **29**:380-395.

Adl, S. M., A. G. B. Simpson, C. E. Lane, J. Lukeš, D. Bass, S. S. Bowser, M. W. Brown, F. Burki, M. Dunthorn, V. Hampl, A. Heiss, M. Hoppenrath, E. Lara, L. L. Gall, D. H. Lynn, H. McManus, E. A. D. Mitchell, S. E. Mozley-Stanridge, L. W. Parfrey, J. Pawlowski, S. Rueckert, L. Shadwick, C. L. Schoch, A. Smirnov, and F. W. Spiegel. 2012. The revised classification of eukaryotes. *Journal of Eukaryotic Microbiology* **59**:429-493.

Ajawatanawong, P., G. C. Atkinson, N. S. Watson-Haigh, B. MacKenzie, and S. L. Baldauf. 2012. SeqFIRE: a web application for automated extraction of indel regions and conserved blocks from protein multiple sequence alignments. *Nucleic acids research* **40**:W340-W347.

Alfaro, M. E., S. Zoller, and F. Lutzoni. 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. *Molecular biology and evolution* **20**:255-266.

Andersson, J. O. 2009. Gene transfer and diversification of microbial eukaryotes. *Annu Rev Microbiol* **63**:177-193.

Andersson, S. G. E., O. Karlberg, B. Canbäck, C. G. Kurland, F. R. Whatley, M. Van Der Giezen, W. Martin, A. G. M. Tielens, J. F. Allen, and J. A. Raven. 2003. On the origin of mitochondria: A genomics perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences* **358**:165-179.

Antonicka, H., A. Mattman, C. G. Carlson, D. M. Glerum, K. C. Hoffbuhr, S. C. Leary, N. G. Kennaway, and E. A. Shoubridge. 2003. Mutations in COX15 produce a defect in the mitochondrial heme biosynthetic pathway, causing early-onset fatal hypertrophic cardiomyopathy. *Am J Hum Genet* **72**:101-114.

Baniulis, D., E. Yamashita, H. Zhang, S. S. Hasan, and W. A. Cramer. 2008. Structure-Function of the Cytochrome b(6)f Complex. *Photochemistry and Photobiology* **84**:1349-1358.

Bareth, B., S. Dennerlein, D. U. Mick, M. Nikolov, H. Urlaub, and P. Rehling. 2013.

- The heme a synthase Cox15 associates with cytochrome c oxidase assembly intermediates during Cox1 maturation. *Molecular and cellular biology* **33**:4128-4137.
- Barros, M. H., C. G. Carlson, D. M. Glerum, and A. Tzagoloff. 2001. Involvement of mitochondrial ferredoxin and Cox15p in hydroxylation of heme O. *FEBS Lett* **492**:133-138.
- Barros, M. H., and A. Tzagoloff. 2002. Regulation of the heme A biosynthetic pathway in *Saccharomyces cerevisiae*. *FEBS Letters* **516**:119-123.
- Bergthorsson, U., K. L. Adams, B. Thomason, and J. D. Palmer. 2003. Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature* **424**:197-201.
- Bilewitch, J. P., and S. M. Degnan. 2011. A unique horizontal gene transfer event has provided the octocoral mitochondrial genome with an active mismatch repair gene that has potential for an unusual self-contained function. *BMC evolutionary biology* **11**:228.
- Bock, R., and J. N. Timmis. 2008. Reconstructing evolution: gene transfer from plastids to the nucleus. *Bioessays* **30**:556-566.
- Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**:2114-2120.
- Brown, B. M., Z. H. Wang, K. R. Brown, J. A. Cricco, and E. L. Hegg. 2004. Heme O synthase and heme A synthase from *Bacillus subtilis* and *Rhodobacter sphaeroides* interact in *Escherichia coli*. *Biochemistry* **43**:13541-13548.
- Bruto, M., C. Prigent-Combaret, P. Luis, G. Hoff, Y. Moënné-Loccoz, and D. Muller. 2013. Horizontal Acquisition of Prokaryotic Genes for Eukaryote Functioning and Niche Adaptation. Pp. 165-179 in P. Pontarotti, ed. *Evolutionary Biology: Exobiology and Evolutionary Mechanisms*. Springer.
- Buchensky, C., P. Almiron, B. S. Mantilla, A. M. Silber, and J. A. Cricco. 2010. The *Trypanosoma cruzi* proteins TcCox10 and TcCox15 catalyze the formation of heme A in the yeast *Saccharomyces cerevisiae*. *Fems Microbiology Letters* **312**:133-141.
- Burger, G., M. W. Gray, L. Forget, and B. F. Lang. 2013. Strikingly bacteria-like and gene-rich mitochondrial genomes throughout jakobid protists. *Genome biology and evolution* **5**:418-438.
- Contreras-Zentella, M., G. Mendoza, J. Membrillo-Hernandez, and J. E. Escamilla.

2003. A novel double heme substitution produces a functional bo(3) variant of the quinol oxidase aa(3) of *Bacillus cereus* - Purification and partial characterization. *Journal of Biological Chemistry* **278**:31473-31478.

Doolittle, W. F. 1998. You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trends in Genetics* **14**:307-311.

Esposti, M. D. 2014. Bioenergetic evolution in proteobacteria and mitochondria. *Genome Biol Evol* **6**:3238-3251.

Esser, C., N. Ahmadinejad, C. Wiegand, C. Rotte, F. Sebastiani, G. Gelius-Dietrich, K. Henze, E. Kretschmann, E. Richly, D. Leister, D. Bryant, M. A. Steel, P. J. Lockhart, D. Penny, and W. Martin. 2004. A genome phylogeny for mitochondria among α -proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. *Molecular biology and evolution* **21**:1643-1660.

Ferla, M. P., J. C. Thrash, S. J. Giovannoni, and W. M. Patrick. 2013. New rRNA gene-based phylogenies of the Alphaproteobacteria provide perspective on major groups, mitochondrial ancestry and phylogenetic instability. *PLoS ONE* **8**:e83383.

Finn, R. D., J. Clements, and S. R. Eddy. 2011. HMMER web server: interactive sequence similarity searching. *Nucleic acids research* **39**:W29-W37.

Fitzpatrick, D. A., C. J. Creevey, and J. O. McInerney. 2006. Genome phylogenies indicate a meaningful α -proteobacterial phylogeny and support a grouping of the mitochondria with the Rickettsiales. *Molecular biology and evolution* **23**:74-85.

Fu, C.-J., S. Sheikh, W. Miao, S. G. Andersson, and S. L. Baldauf. 2014. Missing Genes, Multiple ORFs, and C-to-U Type RNA Editing in *Acrasis kona* (Heterolobosea, Excavata) Mitochondrial DNA. *Genome biology and evolution* **6**:2240-2257.

Glerum, D. M., I. Muroff, C. Jin, and A. Tzagoloff. 1997. COX15 codes for a mitochondrial protein essential for the assembly of yeast cytochrome oxidase. *Journal of Biological Chemistry* **272**:19088-19094.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. D. Zeng, Z. H. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K.

- Lindblad-Toh, N. Friedman, and A. Regev. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**:644-U130.
- Gray, M. W. 2012. Mitochondrial evolution. *Cold Spring Harbor perspectives in biology* **4**:a011403.
- Gray, M. W. 2015. Mosaic nature of the mitochondrial proteome: Implications for the origin and evolution of mitochondria. *Proceedings of the National Academy of Sciences of the United States of America* **112**:10133-10138.
- Guex, N., and M. C. Peitsch. 1997. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* **18**:2714-2723.
- Haas, B. J., M. Chin, C. Nusbaum, B. W. Birren, and J. Livny. 2012. How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes? *Bmc Genomics* **13**.
- Hannappel, A., F. A. Bundschuh, and B. Ludwig. 2012. Role of Surf1 in heme recruitment for bacterial COX biogenesis. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* **1817**:928-937.
- Hao, W., A. O. Richardson, Y. Zheng, and J. D. Palmer. 2010. Gorgeous mosaic of mitochondrial genes created by horizontal transfer and gene conversion. *Proceedings of the National Academy of Sciences of the United States of America* **107**:21576-21581.
- He, D., O. Fiz-Palacios, C.-J. Fu, J. Fehling, C.-C. Tsai, and S. L. Baldauf. 2014. An Alternative Root for the Eukaryote Tree of Life. *Current Biology* **24**:465-470.
- Hederstedt, L. 2012. Heme A biosynthesis. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* **1817**:920-927.
- Hillis, D. M., and J. J. Bull. 1993. An Empirical-Test of Bootstrapping as a Method for Assessing Confidence in Phylogenetic Analysis. *Systematic biology* **42**:182-192.
- Huang, J. L. 2013. Horizontal gene transfer in eukaryotes: The weak-link model. *Bioessays* **35**:868-875.
- Katoh, K., and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* **30**:772-780.
- Keeling, P. J., and J. D. Palmer. 2008. Horizontal gene transfer in eukaryotic evolution. *Nature Reviews Genetics* **9**:605-618.

Khalimonchuk, O., and G. Rödel. 2005. Biogenesis of cytochrome c oxidase. *Mitochondrion* **5**:363-388.

Kim, H. J., O. Khalimonchuk, P. M. Smith, and D. R. Winge. 2012. Structure, function, and assembly of heme centers in mitochondrial respiratory complexes. *Biochimica Et Biophysica Acta-Molecular Cell Research* **1823**:1604-1616.

Kleine, T., U. G. Maier, and D. Leister. 2009. DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. *Annu Rev Plant Biol* **60**:115-138.

Koulintchenko, M., Y. Konstantinov, and A. Dietrich. 2003. Plant mitochondria actively import DNA via the permeability transition pore complex. *EMBO Journal* **22**:1245-1254.

Koulintchenko, M., R. J. Temperley, P. A. Mason, A. Dietrich, and R. N. Lightowers. 2006. Natural competence of mammalian mitochondria allows the molecular investigation of mitochondrial gene expression. *Human Molecular Genetics* **15**:143-154.

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.

Lara, E., A. Chatzinotas, and A. G. Simpson. 2006. Andalusia (n. gen.) - The deepest branch within Jakobids (Jakobida; Excavata), based on morphological and molecular study of a new flagellate from soil. *Journal of Eukaryotic Microbiology* **53**:112-120.

Lartillot, N., N. Rodrigue, D. Stubbs, and J. Richer. 2013. PhyloBayes MPI: Phylogenetic Reconstruction with Infinite Mixtures of Profiles in a Parallel Environment. *Systematic biology* **62**:611-615.

Le, S. Q., and O. Gascuel. 2008. An improved general amino acid replacement matrix. *Molecular biology and evolution* **25**:1307-1320.

Lewin, A., and L. Hederstedt. 2006. Compact archaeal variant of heme A synthase. *FEBS Letters* **580**:5351-5356.

Lill, R., and U. Mühlenhoff. 2008. Maturation of iron-sulfur proteins in eukaryotes: Mechanisms, connected processes, and diseases. *Annual Review of Biochemistry* **77**:669-700.

Maiuri, M. C., and G. Kroemer. 2015. Essential role for oxidative phosphorylation in

- cancer progression. *Cell Metabolism* **21**:11-12.
- Maralikova, B., V. Ali, K. Nakada-Tsukui, T. Nozaki, M. van der Giezen, K. Henze, and J. Tovar. 2010. Bacterial-type oxygen detoxification and iron-sulfur cluster assembly in amoebal relict mitochondria. *Cellular Microbiology* **12**:331-342.
- Markowitz, V. M., I.-M. A. Chen, K. Palaniappan, K. Chu, E. Szeto, Y. Grechkin, A. Ratner, B. Jacob, J. Huang, and P. Williams. 2012. IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic acids research* **40**:D115-D122.
- Mileshina, D., M. Koulintchenko, Y. Konstantinov, and A. Dietrich. 2011. Transfection of plant mitochondria and in organello gene integration. *Nucleic acids research* **39**:e115.
- Mogi, T. 2009. Probing structure of heme A synthase from *Bacillus subtilis* by site-directed mutagenesis. *Journal of biochemistry* **145**:625-633.
- Müller, M., M. Mentel, J. J. Van Hellemond, K. Henze, C. Woehle, S. B. Gould, R. Y. Yu, M. Van Der Giezen, A. G. M. Tielens, and W. F. Martin. 2012. Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. *Microbiology and Molecular Biology Reviews* **76**:444-495.
- Nordberg, H., M. Cantor, S. Dusheyko, S. Hua, A. Poliakov, I. Shabalov, T. Smirnova, I. V. Grigoriev, and I. Dubchak. 2014. The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. *Nucleic acids research* **42**:D26-D31.
- Nowack, E. C. M., and M. Melkonian. 2010. Endosymbiotic associations within protists. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**:699-712.
- Nývltová, E., R. Šuták, K. Harant, M. Šedinová, I. Hrdý, J. Pačes, Č. Vlček, and J. Tachezy. 2013. NIF-type iron-sulfur cluster assembly system is duplicated and distributed in the mitochondria and cytosol of *Mastigamoeba balamuthi*. *Proceedings of the National Academy of Sciences of the United States of America* **110**:7371-7376.
- O'Brien, E. A., L. B. Koski, Y. Zhang, L. S. Yang, E. Wang, M. W. Gray, G. Burger, and B. F. Lang. 2007. TBestDB: a taxonomically broad database of expressed sequence tags (ESTs). *Nucleic acids research* **35**:D445-D451.
- Page, F. C. 1988. A new key to freshwater and soil gymnamoebae: with instructions for culture. Freshwater Biological Association Ambleside, Cumbria.

Parra, G., K. Bradnam, and I. Korf. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**:1061-1067.

Rice, D. W., A. J. Alverson, A. O. Richardson, G. J. Young, M. V. Sanchez-Puerta, J. Munzinger, K. Barry, J. L. Boore, Y. Zhang, C. W. DePamphilis, E. B. Knox, and J. D. Palmer. 2013. Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. *Science* **342**:1468-1473.

Richards, T. A., and J. M. Archibald. 2011. Cell evolution: Gene transfer agents and the origin of mitochondria. *Current Biology* **21**:R112-R114.

Rodríguez-Ezpeleta, N., H. Brinkmann, B. Roure, N. Lartillot, B. F. Lang, and H. Philippe. 2007. Detecting and overcoming systematic errors in genome-scale phylogenies. *Systematic biology* **56**:389-399.

Rodríguez-Ezpeleta, N., and T. M. Embley. 2012. The SAR11 group of alpha-proteobacteria is not related to the origin of mitochondria. *PLoS ONE* **7**:e30520.

Sánchez-Puerta, M. V., T. R. Bachvaroff, and C. F. Delwiche. 2004. The complete mitochondrial genome sequence of the haptophyte *Emiliana huxleyi* and its relation to heterokonts. *DNA Research* **11**:1-10.

Schönknecht, G., A. P. M. Weber, and M. J. Lercher. 2014. Horizontal gene acquisitions by eukaryotes as drivers of adaptive evolution. *Bioessays* **36**:9-20.

Sharma, V., and M. Wikström. 2014. A structural and functional perspective on the evolution of the heme-copper oxidases. *FEBS Letters* **588**:3787-3792.

Sone, N., and Y. Fujiwara. 1991. Heme-O Can Replace Heme-a in the Active-Site of Cytochrome-C-Oxidase from Thermophilic Bacterium *Ps3*. *FEBS Letters* **288**:154-158.

Stairs, C. W., L. Eme, M. W. Brown, C. Mutsaers, E. Susko, G. Dellaire, D. M. Soanes, M. van der Giezen, and A. J. Roger. 2014. A *SUF* Fe-S Cluster Biogenesis System in the Mitochondrion-Related Organelles of the Anaerobic Protist *Pygusua*. *Current Biology* **24**:1176-1186.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**:1312-1313.

Szklarczyk, R., and M. A. Huynen. 2010. Mosaic origin of the mitochondrial proteome. *Proteomics* **10**:4012-4024.

Talavera, G., and J. Castresana. 2007. Improvement of phylogenies after removing

- divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic biology* **56**:564-577.
- Tan, A. S., J. W. Baty, L. F. Dong, A. Bezawork-Geleta, B. Endaya, J. Goodwin, M. Bajzikova, J. Kovarova, M. Peterka, B. Yan, E. A. Pesdar, M. Sobol, A. Filimonenko, S. Stuart, M. Vondrusova, K. Kluckova, K. Sachaphibulkij, J. Rohlena, P. Hozak, J. Truksa, D. Eccles, L. M. Haupt, L. R. Griffiths, J. Neuzil, and M. V. Berridge. 2015. Mitochondrial genome acquisition restores respiratory function and tumorigenic potential of cancer cells without mitochondrial DNA. *Cell Metabolism* **21**:81-94.
- Thiergart, T., G. Landan, M. Schenk, T. Dagan, and W. F. Martin. 2012. An evolutionary network of genes present in the eukaryote common ancestor polls genomes on eukaryotic and mitochondrial origin. *Genome biology and evolution* **4**:466-485.
- Timmis, J. N., M. A. Ayliffe, C. Y. Huang, and W. Martin. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nature Reviews Genetics* **5**:123-135.
- Tommaso, P., S. Moretti, I. Xenarios, M. Orobittg, A. Montanyola, J. M. Chang, J. F. Taly, and C. Notredame. 2011. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic acids research* **39**:W13-W17.
- Valach, M., S. Moreira, G. N. Kiethega, and G. Burger. 2014. Trans-splicing and RNA editing of LSU rRNA in *Diplonema* mitochondria. *Nucleic acids research* **42**:2660-2672.
- Viklund, J., T. J. Ettema, and S. G. Andersson. 2012. Independent Genome Reduction and Phylogenetic Reclassification of the Oceanic SAR11 Clade. *Molecular biology and evolution* **29**:599-615.
- Wang, D., A. H. Lloyd, and J. N. Timmis. 2012. Environmental stress increases the entry of cytoplasmic organellar DNA into the nucleus in plants. *Proceedings of the National Academy of Sciences of the United States of America* **109**:2444-2448.
- Wang, Z., Y. Wang, and E. L. Hegg. 2009. Regulation of the heme A biosynthetic pathway: differential regulation of heme A synthase and heme O synthase in *Saccharomyces cerevisiae*. *Journal of Biological Chemistry* **284**:839-847.
- Weinstein, J. D., and S. I. Beale. 1984. Biosynthesis of Protoheme and Heme-a

Precursors Solely from Glutamate in the Unicellular Red Alga *Cyanidium-Caldarium*. *Plant Physiology* **74**:146-151.

Williams, K. P., B. W. Sobral, and A. W. Dickerman. 2007. A robust species tree for the Alphaproteobacteria. *Journal of Bacteriology* **189**:4578-4586.

Wu, B., A. Buljic, and W. Hao. 2015. Extensive Horizontal Transfer and Homologous Recombination Generate Highly Chimeric Mitochondrial Genomes in Yeast. *Molecular biology and evolution* DOI:10.1093/molbev/msv1127.

Zimorski, V., C. Ku, W. F. Martin, and S. B. Gould. 2014. Endosymbiotic theory for organelle origins. *Current Opinion in Microbiology* **22**:38-48.

Figure Legends:

Figure 1. Molecular phylogeny and sequence signatures of heme A synthase (HAS). The tree shown was derived based on 253 aligned amino-acid positions of deduced HAS sequences using RAxML 8.0.0 with the LG+4Γ+F model. Branch support values are only shown for major clades, where significant (mlBP ≥ 70% and biPP ≥ 0.95). The 84% mlBP value in parentheses indicates support for a monophyletic eukaryote type-2 HAS clade (*cox15-2*) in analyses excluding the distantly related type-1 HAS (*cox15-1*) clade. Variations in eight key amino acid residues are indicated according to the key at the top left on schematic representations of predicted protein structures, with transmembrane helices indicated by gray rectangles. Presence/absence of each of the eight residues are indicated separately to the right of eleven taxa where the residues show patchy distribution, with filled or empty circles indicating presence or absence, respectively. Branches are drawn to scale as indicated by the scale bar at the bottom.

Figure 2. Sequence and predicted structure comparisons for the two HAS types.

Top: A predicted structure of *A. godoyi* mitochondrial *cox15* encoded HAS (YP_007890502) is shown (left), followed by an extraction of the basic four transmembrane helix unit (middle), and (right) a schematic of the same with filled circles indicating functionally important histidine residues in purple (H1, H2) and type-specific cysteine/hydrophobic residues in blue/yellow. Bottom: The variable length loop region between a1 and a2 (dotted magenta circles in schematics above) is shown for a representative set of taxa. The first two of the four nearly invariant cysteine residues (C1, C2) in type-1 HAS and the replacing hydrophobic amino acids in type-2 HAS are highlighted in yellow and light blue, respectively. The histidine residue (H1) found in nearly all examined HAS sequences is highlighted in red. Sequences are grouped by higher taxonomy as indicated to the left, with the exception of *A. godoyi* (red) and *T. mobilis* (YP_006372228) plus *G. roseus* (WP_027133088) (blue purple). A sequence alignment of the full taxon set can be found in [Supplementary figure S2](#).

Figure 3. Comparison of jakobid mtDNA loci corresponding to the *A. godoyi cox15-1* insertion site. Top: The gene content of a conserved tRNA cluster in jakobid mtDNAs is shown with tRNA genes (solid black lines) and protein-coding genes (color frame) indicated. The *A. godoyi cox15-1* is embedded in this cluster, where it occurs on the anti-sense DNA strand as indicated with yellow arrows. Bottom: The overall gene synteny among jakobid mtDNAs is shown, with a circle to indicate the regions surrounding the tRNA cluster enlarged at the top of the figure. Genes are color-coded by functional category as indicated in the box at the far bottom of the figure (adapted from Fu et al. 2014). The phylogenetic relationships among the five taxa are

indicated immediately below their respective synteny maps as follows, with GenBank accession numbers in parentheses: *A. god* (*Andalucia godoyi*, KC353352), *H. aro* (*Histiona aroides*, KC353353), *J. bah* (*Jakoba bahamiensis*, KC353354), *R. ame* (*Reclinomonas americana*, KC353356) and *S. ecu* (*Seculamonas ecuadoriensis*, KC353359).

Figure 4. Three possible scenarios for the origin and evolution of two *cox15* genes in eukaryotes. The different scenarios are shown in parts A-C and depend on the gene content of LECA, *i.e.* whether LECA carried (A) only *cox15-1*, (B) both *cox15-1* and *cox15-2*, or (C) only *cox15-2*. Schematic representations of the two distinct *cox15* types and gene loss are indicated according to the key at the bottom. Horizontal gene transfer (HGT) and endosymbiotic gene transfers (EGT) are indicated by red straight arrow and black curved arrow, respectively. LECA: last eukaryotic common ancestor.

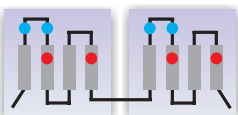
- Cysteine
- Hydrophobic amino acids other than Cysteine
- Histidine

Eukaryotes

76(84)

0.99

Type-2 HAS
(cox15-2)



α -proteobacteria

ζ -proteobacteria

α -proteobacteria

γ -proteobacteria

β -proteobacteria

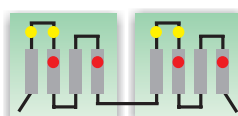
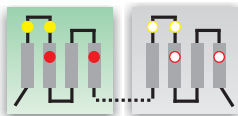
Bacteria

Bacteroidetes

α -proteobacteria

100

1.0



Type-1 HAS
(cox15-1)

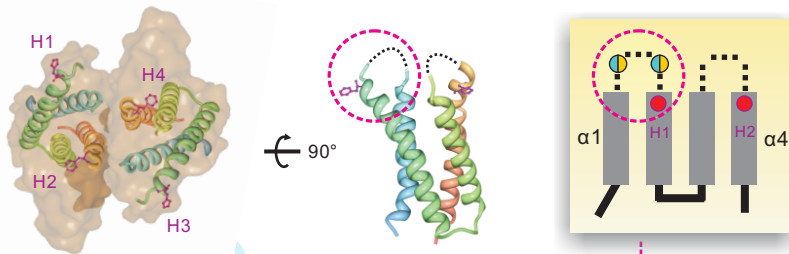
1.0

Archaea

Bacteria

Eukaryote

1.0



Eukaryotes	GGLTRLTRSGLSM	-TDWK-	QGGL--	PQPTDDE---	WQREFERYKRFPEWQQ---	-RQS---	MTVEDFKFIYFWEYGR	RMLGRV						
	GGITRLTSGLSM	-VEWKP-	TGILL-	PPTKASCLE	GWEEFNKYSFPEYQK-	-HNDE---	-LSLAEFKWIWMFWEYGR	RMAGRA						
	GGITRLMTKSGLSM	-TDWK-	VQGSLL-	PPSEAE---	-WEAEFDRYKAFPEWQQ---	-RRS---	MTLDEFKHIFWEYGR	RMMLGRA						
	GGITRLTSGLSM	-TDWRP-	QGKR-	WPRNDAE---	-WEAEFDRWKQFPEYQR-	-LYKGS-	PFSDMEFKGIFWEFWEYGR	RMGMRG						
	GGITRLTSGLSL	-TEWKPP-	AGAW-	PPMTPED---	-WQREFDLYTQSPFVWKQRDPNT-	-FSDMEF---	-EWA	RLQLGRL						
	GGITRLTSGLSM	-VWKWP-	HSEI-	PPMTEEE---	-WMAEFEEFKTYPEWIMKRQYEN-	-FTLQDFKFIYFWEYGR		RMFGRL						
	GGITRLTSGLSM	-VEWKP-	VGAII-	PPLSQEE---	-WEEEFNRYKQYPEYQ---	-LNKG---	MTVEEFKFIYFWEYGR	RLMGRLL						
	GGITRLTSGLSMQTDWS-	FQGLL-	PPLNEE---	-WKQEFKQYKQFPEFKRNPND-	-FSLRDFKFIYFWEYGR			RMYGFR						
	GGITRLTSGLSL	-VEWKP-	TGII-	PPLNNAR---	-WEEEFKQYKQFPEYKL-	-LNNY---	MTLPEFKYIYFWEYGR	RIWGRA						
	GGITRLTSGLSM	-VDWRP-	TGWL-	PPLNEAQ---	-WQAIFDLYKQSPQFKE-	-FNFD-	MDLDGGRGFIWLEFLR	RLWGRLL						
α-proteobac	GGATRLTSGLSL	-MEWAP-	SGTL-	PPLSQAE---	-WQRLYDLRTIYQYEL-	-VNGQ---	-FGLEGFKQIFWLEWT	RLWGRLL						
	GGVTRLTNSGLSM	-VEWKP-	TGWL-	PPLSEQA---	-WMDEFEKYKAFPEYQK-	-INKG---	MSLDEFKGIFAFEYGR	RVLGRI						
	GGATRLTSGLSL	-VEWKP-	VGI-	PPFTEAA---	-WLEEFAYKQFPEYQL-	-INHG---	MTLSEFKFIYFWEYGR	RLLGRL						
	GGATRLTSGLSM	-MTWDP-	LMGWI-	PPLNRAQ---	-WLDSFEHYRKIPFEY-	-LNP-	MDMEGYKGIIFLLEYGR	RVWGRLL						
	GGVTRLTSGLSL	-VEWQP-	LVGTV-	PPLSQND---	-WDLFEKYHQTQYKK-	-VNLG---	MSLEEFKTIFFWEYGR	RLVGRV						
	GGVTRLTSGLSL	-VEWKP-	VGVVL-	PPLNNEE---	-WLQFEFKYQYYPEYQK-	-VNTG---	MDLAGFKQIFWEYGR	RLLGRL						
	GGVTRLTSGLSM	-VEWRP-	MGIW-	PPIGEAA---	-WQEVFAKYQRFPEYQK-	-INQG---	MGLEDKFKIYFWEYGR	RLVGRLL						
	GGITRLTSGLSL	-SNYKL-	TGTI-	PPLNDE---	-WQEAFLYKQYPEYQK-	-LNYH---	-FELEDKFIYFWEYGR	RVIGRL						
	GGITRLTNSGLSM	-TDWHL-	TDTF-	PPLTDEK---	-WMAAFEEYKQFPEYQK-	-INIHD-	-FTVEDFKYIYFWEYGR	RFIGRI						
	GGITRLTNSGLSM	-TDWHL-	TDTF-	PPLTTEEK---	-WQQAFFEEYKQFPEYQL-	-INIHD-	-FDLSDYKFIYFWEYGR	RFIGRI						
Bacteroidetes	GGITRLTQSGLSM	-VKWEP-	KGAI-	PPLTEEA---	-WQEEFRAYQSSPEYEV-	-YNTH---	-FGLSEFKQIYFWEYGR	RLLGRI						
	GVYTAATGSGLAC	QAQWPL	SGDQL-	I	PAIT-		-INP-	-DFI	EWFR	RAWAMV				
	GLTYATKAGALTC	EQRWPL	CDGWMGL	FPA			-NLV-	-SFI	EWFS	RLAML				
	GAYTSAIGAGLSC	-PDWPT	TCYGTW-	VPLTQPEIIA			-NSP	YSALQ-		I	FAEWA	RLGAMT		
	GAFTRAYGAGMG	GPDWPT	TCNGEI	-VPFTS			-DTA-			T	LLEYF	RVAAGL		
	GAFTRAYGAGMG	GPDWPT	TCNGEI	-VPLTA			-NTA-			T	LLEYF	RVAAGL		
	GSYVSARGAGLAC	-PDWPL	CP-	FFG			-DEL-			V	LEEF	RAFAIV		
	GACITRLTSGLSL	-PDWPL	CYGLW-	LPTYSKI		HLLP-		EIH	YGFSO-		V	FFWFS	RFIAGI	
	GAVIRATNSGLSC	-PDWPH	YGFVW-	VPTPAKL		-ATVP-		GID	YTFQT-		V	MYEWW	RFIAGI	
	GGVVRITNSGLSC	-PDWPT	CYGYW-	LPLPSLL		-EQVP-		-NMG	YEEYQ-		V	MLEW	RLIGGV	
Archaea	GASITLTSGLSL	-PDWPL	CYGLW-	IPLSKLL		-LLIS-		-DLN	YEEYQ-		V	MLEW	RLNAAI	
	GAWWRLTSGMS	-PDWPL	CYGYI	-FPTNKI		-VNI-		-NIE	YSYFQ-		I	FLEW	RLNAAI	
	GAWWRLTSGMS	-PDWPL	CYGYI	-FPTNKI		-ENI-		-NIE	YSYFQ-		I	FLEW	RLNAAI	
	GSVARTLTSGLSC	-PDWPL	CYGLW-	FNQEKI		-STIS-		-DVN	YEEYQ-		I	MLEW	RFNAAI	
	GGILIRIYDAGES	-PDWPL	CYGTFG	-FDISEE	-QEA	WYIENP	DEVDSRGS	GHR	YTTFFQ-		I	FTEWL	RLIAGL	
	GGILIRIYDAGES	-PDWPL	CYGTFG	-FDISEE	-QEA	WYIENP	DEVDSRGS	GHR	YTTFFQ-		I	FTEWL	RLIAGL	
	GAAATRVNAGLAC	-PDWPL	CYGOW-	YSQOM			-NLQ-			I	FLEW	RLDASL		
	GPLYRLTSGLAC	-PDWPF	CFGKI	FPDF			-DFN-			I	FMEVG	RRYSYG		
	GGTVRVITGAGMG	GPDWPL	LVNGQV	-IPEW-			-DVL			A	WTEFI	RLVVALA		
	GAIVRVITSGLG	-PTWPT	CFPDSM	TPRADAL		-VP-		-WQH-		A	IFEGNR	ITLGL		
Bacteria	GAAVRITLTSGLG	-PNWPM	CTDS-	LTNTPM-			-GIH-			A	IFEGNR	ITLGVV		
	GSFTRLTSGLGL	-PDWPG	CYGTS	-NPFHARD	IIHAAQ	AMP-		-SGP	VTWMK-		-GF	IEMT	RYFALA	
	GGFTRLTSGLGL	-PDWPG	CYGTS	-SPFIIA	AAHIIHAAQ	AMP-		-TGP	VSMTK-		-AW	EMI	RYFAMA	
	GAYTRLSADAGLGL	-PDWPG	CYGHFS	-VPHHED	-VLRANI	INFP-		-ERE	IEHEK-		-AW	EMI	RYFAGT	
	GAYTRLSADAGLGL	-PDWPG	CYGHFA	-VPSASHD	-LANIES	RFP-		-DAQ	IEPEK-		-AW	EMI	RYIAGT	
	β-, γ-proteobac													

C1

C2

Type-specific region

H1

