SUPPORTING INFORMATION APPENDIX

Mitochondrial Genomes are Retained by Selective Constraints on Protein Targeting

Patrik Björkholm#, Ajith Harish#, Erik Hagström, Andreas M. Ernst, Siv G.E. Andersson*

Supporting methods:	**pages 1-7**
Supporting figures S1-S4:	**pages 9-11**
Supporting tables S1-S8:	**pages 12-33**
Supporting references	**page 34**

Materials and methods

Mitochondrial respiratory chain dataset

Proteins in the human mitochondrial respiratory chain were identified and downloaded from the HUGO nomenclature Committee webpage (1). Proteins in the yeast mitochondrial respiratory chain were downloaded from NCBI Organelle Genome Resource (http://www.ncbi.nlm.nih.gov/genomes/GenomesHome.cgi?taxid=2759&hopt=html).

Mitochondrial structure dataset

All membrane structures from the OPM web-resource that were annotated as mitochondrial (inner membrane) and that contained trans-membrane components were downloaded as Fasta sequences (2). Only proteins showing to have a transmembrane region were kept.

Human and yeast mitochondrial GFP-dataset

Yeast and human proteins localized to mitochondria with the aid of GFP were downloaded from Mitominer (http://mitominer.mrc-mbu.cam.ac.uk/release-3.1/begin.do) (9). We only kept proteins for which hydrophobic segments were identified and annotated or published as mitochondrial proteins in Uniprot.

Chloroplast photosynthetic dataset

Photosynthetic proteins were identified using the KEGG resource (3) and linked to Uniprot where their sequences were downloaded. If multiple identifiers were available the first reviewed protein in the list was selected (4).

Alphaproteobacterial and cyanobacterial datasets

Respiratory chain proteins from *Rickettsia prowazekii* (strain Madrid E) were downloaded from Uniprot. Bacterial homologs of two genes from the respiratory chain complex (*cox1* and *cytb*) and two photosynthetic genes (*psaA* and *psbB*) were identified and downloaded using uniprot families (4). Cox1 and CytB protein sequences were downloaded from the following species: *Paracoccus denitrificans, Bradyrhizobium diazoefficiens, Brucella suis, Magnetospirillum magneticum, Rickettsia prowazeki, Chlamydomonas reinhardtii,* *Caenorhabditis elegans, Danio rerio, Drosophila melanogaster, Gallus Gallus, Canis familiaris, Bos Taurus* and *Arabidopsis thaliana*. PsaA and PsbB protein sequences were downloaded from the following species: *Acaryochloris marina, Nostoc sp., Synechococcus elongatus, Anabaena variabilis, Gloeobacter violaceus, Oryza sativa, Populus alba, Vitis vinifera, Conocephalum conicum* and *Zea mays*.

Prediction of the hydrophobicity of transmembrane domains

Transmembrane domains were predicted and their hydrophobicity was calculated using DGpred (6), which provides a score that is based on hydrophobicity for each transmembrane domain. This score represent the free energy of membrane insertion for that specific TMD (kcal/mol) and normally ranges between +3 and -3. The boundary for SRP-recognition of a transmembrane domain has been approximated to zero and increases in efficiency with lower dG-value (5). Transmembrane domains were categorized as marginally hydrophobic if they had a dG-value higher then 0 (3 ➔ 0), while transmembrane domains that had a score between 0 or lower (0 ➔ -∞) were categorized as hydrophobic.

Construct design, cloning, sequences

Amino acid sequences of human ATP synthase subunit a (ATP6, Uniprot P00846), human cytochrome c oxidase subunit 1 (COX1, Uniprot P00395), human apocytchrome b (CYB, Uniprot P00156) and human Erlin-1 (ERLN1, Uniprot O75477) were downloaded from Uniprot (http://www.uniprot.org/). At the amino-terminus, a MDYKDDDDK- (initiator Met and FLAG tag) and AANDIL -sequence (linker) were added. The sequences were then flanked with 5´-EcoRI and 3`-NotI restrictions sites. The synthetic genes were obtained from Life Technologies, and sub-cloned via the EcoRI and NotI sites to a CMV-promotor-containing pSNAPf vector (NEB; subcloning led to the removal of SNAPf). OMP25-GFP was a kind gift from Prof. Dr. James Rothman (New Haven, U.S.A.).

Sequences (red: initiator Met; green: FLAG tag; yellow: linker)

COX1

>sp|P00395|COX1_HUMAN Cytochrome c oxidase subunit 1 OS=Homo sapiens GN=MT-CO1 PE=1 SV=1
MDYKDDDDKAANDILMFADRWLFSTNHKDIGTLYLLFGAWAGVLGTALSLLIRAELGQPGNLLGNDHIYNVIVTA
HAFVMIFFMVMPIMIGGFGNWLVPLMIGAPDMAFPRMNNMSFWLLPPSLLLLLASAMVEAGAGTGWTVYPPLAGN
YSHPGASVDLTIFSLHLAGVSSILGAINFITTIINMKPPAMTQYQTPLFVWSVLITAVLLLLSLPVLAAGITMLL
TDRNLNTTFFDPAGGGDPILYQHLFWFFGHPEVYILILPGFGMISHIVTYYSGKKEPFGYMGMVWAMMSIGFLGF
IVWAHHMFTVGMDVDTRAYFTSATMIIAIPTGVKVFSWLATLHGSNMKWSAAVLWALGFIFLFTVGGLTGIVLAN
SSLDIVLHDTYYVVAHFHYVLSMGAVFAIMGGFIHWFPLFSGYTLDQTYAKIHFTIMFIGVNLTFFPQHFLGLSG
MPRRYSDYPDAYTTWNILSSVGSFISLTAVMLMIFMIWEAFASKRKVLMVEEPSMNLEWLYGCPPPYHTFEEPVY
MKS

Reverse translated:

gaattcatggactacaaggacgacgacgacaaggccgccaacgacatcctgatgttcgccgacagatggctgttc
agcaccaaccacaaggacatcggcaccctgtacctgctgttcggcgcctgggccggcgtgctgggcaccgccctg
agcctgctgatcagagccgagctgggccagcccggcaacctgctgggcaacgaccacatctacaacgtgatcgtg
accgcccacgccttcgtgatgatcttcttcatggtgatgcccatcatgatcggcggcttcggcaactggctggtg
cccctgatgatcggcgcccccgacatggccttccccagaatgaacaacatgagcttctggctgctgcccccccagc
ctgctgctgctgctggccagcgccatggtggaggccggcgccggcaccggctggaccgtgtacccccccctggcc
ggcaactacagccacccccggcgccagcgtggacctgaccatcttcagcctgcacctggccggcgtgagcagcatc
ctgggcgccatcaacttcatcaccaccatcatcaacatgaagcccccccgccatgacccagtaccagaccccccctg
ttcgtgtggagcgtgctgatcaccgccgtgctgctgctgctgagcctgcccgtgctggccgccggcatcaccatg
ctgctgaccgacagaaacctgaacaccaccttcttcgaccccgccggcggcggcgaccccatcctgtaccagcac
ctgttctggttcttcggccaccccgaggtgtacatcctgatcctgcccggcttcggcatgatcagccacatcgtg
acctactacagcggcaagaaggagcccttcggctacatgggcatggtgtgggccatgatgagcatcggcttcctg
ggcttcatcgtgtgggcccaccacatgttcaccgtgggcatggacgtggacaccagagcctacttcaccagcgcc
accatgatcatcgccatccccaccggcgtgaaggtgttcagctggctggccaccctgcacggcagcaacatgaag
tggagcgccgccgtgctgtgggccctgggcttcatcttcctgttcaccgtgggcggcctgaccggcatcgtgctg
gccaacagcagcctggacatcgtgctgcacgacacctactacgtggtggcccacttccactacgtgctgagcatg
ggcgccgtgttcgccatcatgggcggcttcatccactggttcccccctgttcagcggctacaccctggaccagacc
tacgccaagatccacttcaccatcatgttcatcggc
gtgaacctgaccttcttccccccagcacttcctgggcctgagcggcatgcccagaagatacagcgactaccccgac
gcctacaccacctggaacatcctgagcagcgtgggcagcttcatcagcctgaccgccgtgatgctgatgatcttc
atgatctgggaggccttcgccagcaagagaaaggtgctgatggtggaggagcccagcatgaacctggagtggctg
tacggctgccccccccccctaccacaccttcgaggagcccgtgtacatgaagagcTAGTAGgcggccgc

CYB

>sp|P00156|CYB_HUMAN Cytochrome b OS=Homo sapiens GN=MT-CYB PE=1 SV=2
MDYKDDDDKAANDILMTPMRKTNPLMKLINHSFIDLPTPSNISAWWNFGSLLGACLILQITTGLFLAMHYSPDAS
TAFSSIAHITRDVNYGWIIRYLHANGASMFFICLFLHIGRGLYYGSFLYSETWNIGIILLLATMATAFMGYVLPW
GQMSFWGATVITNLLSAIPYIGTDLVQWIWGGYSVDSPTLTRFFTFHFILPFIIAALATLHLLFLHETGSNNPLG
ITSHSDKITFHPYYTIKDALGLLLFLLSLMTLTLFSPDLLGDPDNYTLANPLNTPPHIKPEWYFLFAYTILRSVP
NKLGGVLALLLSILILAMIPILHMSKQQSMMFRPLSQSLYWLLAADLLILTWIGGQPVSYPFTIIGQVASVLYFT
TILILMPTISLIENKMLKWA

Reverse translated:

gaattcatggactacaaggacgacgacgacaaggccgccaacgacatcctgatgacccccatgagaaagaccaac
cccctgatgaagctgatcaaccacagcttcatcgacctgcccacccccagcaacatcagcgcctggtggaacttc
ggcagcctgctgggcgcctgcctgatcctgcagatcaccaccggcctgttcctggccatgcactacagccccgac
gccagcaccgccttcagcagcatcgcccacatcaccagagacgtgaactacggctggatcatcagatacctgcac
gccaacggcgccagcatgttcttcatctgcctgttcctgcacatcggcagaggcctgtactacggcagcttcctg
tacagcgagacctggaacatcggcatcatcctgctgctggccaccatggccaccgccttcatgggctacgtgctg
ccctggggccagatgagcttctggggcgccaccgtgatcaccaacctgctgagcgccatcccctacatcggcacc
gacctggtgcagtggatctggggcggctacagcgtggacagccccaccctgaccagattcttcaccttccacttc
atcctgcccttcatcatcgccgccctggccaccctgcacctgctgttcctgcacgagaccggcagcaacaacccc
ctgggcatcaccagccacagcgacaagatcaccttccacccctactacaccatcaaggacgccctgggcctgctg
ctgttcctgctgagcctgatgaccctgaccctgttcagccccgacctgctgggcgaccccgacaactacaccctg
gccaacccccctgaacacccccccccacatcaagcccgagtggtacttcctgttcgcctacaccatcctgagaagc
gtgcccaacaagctgggcggcgtgctggccctgctgctgagcatcctgatcctggccatgatccccatcctgcac
atgagcaagcagcagagcatgatgttcagacccctgagccagagcctgtactggctgctggccgccgacctgctg
atcctgacctggatcggcggccagcccgtgagctaccccttcaccatcatcggccaggtggccagcgtgctgtac

ttcaccaccatcctgatcctgatgcccaccatcagcctgatcgagaacaagatgctgaagtgggccTAGTAGgcg
gccgc

ATP6

>sp|P00846|ATP6_HUMAN ATP synthase subunit a OS=Homo sapiens GN=MT-ATP6
PE=1 SV=1
MDYKDDDDKAANDILMNENLFASFIAPTILGLPAAVLIILFPPLLIPTSKYLINNRLITTQQWLIKLTSKQMMTM
HNTKGRTWSLMLVSLIIFIATTNLLGLLPHSFTPTTQLSMNLAMAIPLWAGTVIMGFRSKIKNALAHFLPQGTPT
PLIPMLVIIETISLLIQPMALAVRLTANITAGHLLMHLIGSATLAMSTINLPSTLIIFTILILLTILEIAVALIQ
AYVFTLLVSLYLHDNT

Reverse translated:

gaattcatggactacaaggacgacgacgacaaggccgccaacgacatcctgatgaacgagaacctgttcgccagc
ttcatcgcccccaccatcctgggcctgcccgccgccgtgctgatcatcctgttcccccccctgctgatccccacc
agcaagtacctgatcaacaacagactgatcaccacccagcagtggctgatcaagctgaccagcaagcagatgatg
accatgcacaacaccaagggcagaacctggagcctgatgctggtgagcctgatcatcttcatcgccaccaccaac
ctgctgggcctgctgccccacagcttcacccccaccacccagctgagcatgaacctggccatggccatcccctg
tgggccggcaccgtgatcatgggcttcagaagcaagatcaagaacgccctggcccacttcctgccccagggcacc
cccaccccctgatccccatgctggtgatcatcgagaccatcagcctgctgatccagcccatggccctggccgtg
agactgaccgccaacatcaccgccggccacctgctgatgcacctgatcggcagcgccaccctggccatgagcacc
atcaacctgcccagcaccctgatcatcttcaccatcctgatcctgctgaccatcctggagatcgccgtggccctg
atccaggcctacgtgttcaccctgctggtgagcctgtacctgcacgacaacaccTAGTAGgcggccgc

ERLN1

>sp|O75477|ERLN1_HUMAN Erlin-1 OS=Homo sapiens GN=ERLIN1 PE=1 SV=1
MDYKDDDDKAANDILMTQARVLVAAVVGLVAVLLYASIHKIEEGHLAVYYRGGALLTSPSGPGYHIMLPFITTFR
SVQTTLQTDEVKNVPCGTSGGVMIYIDRIEVVNMLAPYAVFDIVRNYTADYDKTLIFNKIHHELNQFCSAHTLQE
VYIELFDQIDENLKQALQKDLNLMAPGLTIQAVRVTKPKIPEAIRRNFELMEAEKTKLLIAAQKQKVVEKEAETE
RKKAVIEAEKIAQVAKIRFQQKVMEKETEKRISEIEDAAFLAREKAKADAEYYAAHKYATSNKHKLTPEYLELKK
YQAIASNSKIYFGSNIPNMFVDSSCALKYSDIRTGRESSLPSKEALEPSGENVIQNKESTG

Reverse translated:

gaattcatggactacaaggacgacgacgacaaggccgccaacgacatcctgatgacccaggccagagtgctggtg
gccgccgtggtgggcctggtggccgtgctgctgtacgccagcatccacaagatcgaggagggccacctggccgtg
tactacagaggcggcgccctgctgaccagccccagcggccccggctaccacatcatgctgcccttcatcaccacc
ttcagaagcgtgcagaccaccctgcagaccgacgaggtgaagaacgtgccctgcggcaccagcggcggcgtgatg
atctacatcgacagaatcgaggtggtgaacatgctggcccctacgccgtgttcgacatcgtgagaaactacacc
gccgactacgacaagaccctgatcttcaacaagatccaccacgagctgaaccagttctgcagcgcccacaccctg
caggaggtgtacatcgagctgttcgaccagatcgacgagaacctgaagcaggccctgcagaaggacctgaacctg
atggcccccggcctgaccatccaggccgtgagagtgaccaagcccaagatccccgaggccatcagaagaaacttc
gagctgatggaggccgagaagaccaagctgctgatcgccgcccagaagcagaaggtggtggagaaggaggccgag
accgagagaaagaaggccgtgatcgaggccgagaagatcgcccaggtggccaagatcagattccagcagaaggtg
atggagaaggagaccgagaagagaatcagcgagatcgaggacgccgccttcctggccagagagaaggccaaggcc
gacgccgagtactacgccgcccacaagtacgccaccagcaacaagcacaagctgacccccgagtacctggagctg
aagaagtaccaggccatcgccagcaacagcaagatctacttcggcagcaacatccccaacatgttcgtggacagc
agctgcgccctgaagtacagcgacatcagaaccggcagagagagcagcctgcccagcaaggaggccctggagccc
agcggcgagaacgtgatccagaacaaggagagcaccggcTAGTAGgcggccgc

Cell culture and transfection

HeLa cells were cultured in DMEM (Gibco) with 10% FCS at 5% CO_2 and 37°C.

Transfection was performed using FuGENE HD transfection reagent (Promega) following the instructions of the manufacturer (300 ng plasmid/1 µl Fugene HD / 35 mm glass bottom dish).

Confocal microscopy

Cells were plated on 35mm glass bottom dishes (MatTek Corp.) and transfected using FuGENE HD. After 24 h, the cells were washed three times with PBS (pH 7.4) and fixed with 3.6% PFA (Sigma-Aldrich) and 0.1% glutaraldehyde for 20 min on ice. For endomembrane staining, the cells were permeabilized with 0.01% saponin in block buffer (PBS containing 5% goat serum) for 45 min at RT. Immunofluorescence staining was performed by dilution of primary antibodies in block buffer as follows: rabbit anti-FLAG antibody 1:250 (Sigma-Aldrich, F7425), mouse anti-FLAG antibody 1:250 (Sigma-Aldrich, F3040), mouse anti-calnexin 1:500 (Abcam, AF18), rabbit anti-TOM20 1:500 (Santa Cruz, FL-145), and incubated for 1h at RT. After 3 washes for 5 min in PBS, the cells were incubated with secondary antibodies as follows: goat anti-mouse Alexa488 1:1000 (Life Technologies, A11029), goat anti-mouse Alexa647 1:1000 (Life Technologies, A21236), goat anti-rabbit Alexa488 1:1000 (Life Technologies, A11034), goat anti-rabbit Alexa647 1:1000 (Life Technologies, A21245), and incubated for 1h at RT, and subsequently washed 3 times for 5 min in PBS at RT. The samples were then subjected to confocal microscopy using a Zeiss LSM 510 Meta and a 63x, 1.3 NA Plan-NEOFLUAR oil objective. Image processing was performed using Zen 2011 and ImageJ software (http://fiji.sc/Fiji).

Selection of taxa/proteomes for SF-domain analyses

To study phyletic distribution of protein domains and to reconstruct ancestral domain content a reliable phylogeny is essential. We used a published multi-gene phylogeny inferred to propose the root of the eukaryote tree of life (7), which includes major eukaryotic and bacterial taxa. Among the taxa included in the phylogeny we limited our analyses to 43 eukaryotes and 13 bacteria (Table-S6) for which complete genome sequences were available at the time of this study.

Mitochondrial protein sequences were obtained from the MitoCOG collection (8), (ftp://ftp.ncbi.nih.gov/pub/koonin/MitoCOGs/) and specific protein collections for the human mitochondrial ribosome and OXPHOS proteins were obtained from the HUGO Gene Nomenclature Committee at http://www.genenames.org/genefamilies/MRP and http://www.genenames.org/genefamilies/mitocomplex respectively.

Finally biochemically verified mitochondrial proteins downloaded from the MitoMiner database (9). Experimental evidence for mitochondrial localization is obtained from 58 large-scale proteomic analyses and/or from GFP tagged localization. The MitoMiner 'reference set' is a compilation of about 13,000 proteins from 12 model organisms (Table-SX). To be reliably identified as a protein localized to mitochondria, the criteria for a protein to be in the reference set is that mitochondrial localization is required to be supported by at least three independent mass spectrometry experiments or one GFP tagged localization. Although a larger set of proteins (26,000) is available from MitoMiner, we used the reference set to exclude false positives, which is common in proteomic analyses. However, using stringent filters is prone to false negatives. For example, known mitochondrial proteins such as the F1F0 ATP synthase complex subunits were excluded when we applied stringent filters. It is also common with proteomics studies that membrane associated proteins are missed due to difficulties in extracting membrane proteins.

Therefore, the most satisfactory option to compile a nominal 'mitochondrial proteome' was to use a combination of stringently filtered MitoMiner reference set and a collection of mitochondrial genome encoded proteins from the MitoCOG databse.

SCOP superfamily annotations

Structural Classification of Proteins Superfamilies (SCOP-SF): Proteins domain annotations at the level of SCOP superfamily for sequenced genomes were either downloaded from a local installation of SUPERFAMILY (ver 1.75) database. For genomes for which SUPERFAMILY annotations were unavailable, proteins sequences were downloaded from corresponding genome project repositories and SF-domain assignments were carried out as described in (10). A library of hidden Markov models (HMM) representing SCOP Superfamilies and PERL scripts required for SF domain assignments of genome sequences were obtained from the SUPERFAMILY database (ftp.supfam.org). Protein sequences were scored against HMM profiles using the hmmscan program from the HMMER 3.0 software package (11). An *E*-value threshold of 10^{-4} was used to select assignments. Function annotation of SF domains (12) was also obtained from the SUPERFAMILY databse.

Ancestral domain content and domain flux analyses

SF-domains corresponding to the mitochondrial proteome were extracted from each genome and saved as a matrix of characters representing SF-domain occurrence (presence/absence). The most parsimonious scenarios of mitochondrial proteome evolution and the domain content of the mitochondrial ancestor were reconstructed by using generalized parsimony based ancestral-state reconstruction as implemented in PAUP* (13). SF-domain presence/absence was coded as binary character-states and a rooted eukaryotic and bacterial phylogeny (7) was used for character mapping. Based on the understanding that genesis of a structural-domain is very difficult and rare during evolution (14), a gain penalty of 2 and a loss penalty of 1 was applied to the occurrence matrix.

Fig S1

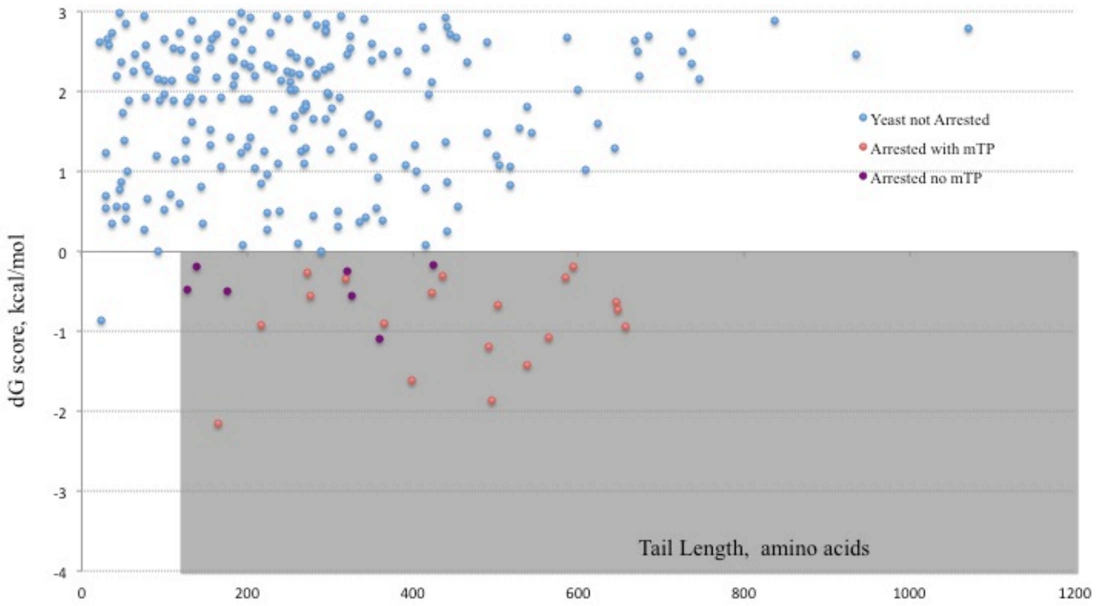

Figure S1. Biophysical characteristics of mitochondrial membrane proteins encoded by the nuclear genome. The figure shows the insertion free energy (deltaG) and the length of the TMD and the following C-terminal tail for human and yeast proteins localized to mitochondria with the aid of green fluorescent tags. The insertion free energy (kcal/mol) has been estimated for either the first TMD with a calculated deltaG value below zero or if no such segment was found, the most hydrophobic segment in the protein. Blue dots represent proteins that are not targeted by SRP. Red dots correspond to proteins that are putative targets for SRP with a mitochondrial target peptide (mTP) and the purple are proteins without an mTP. Characteristics of proteins that are putative targets for recognition by SRP are indicated with the green area.

Figure S2

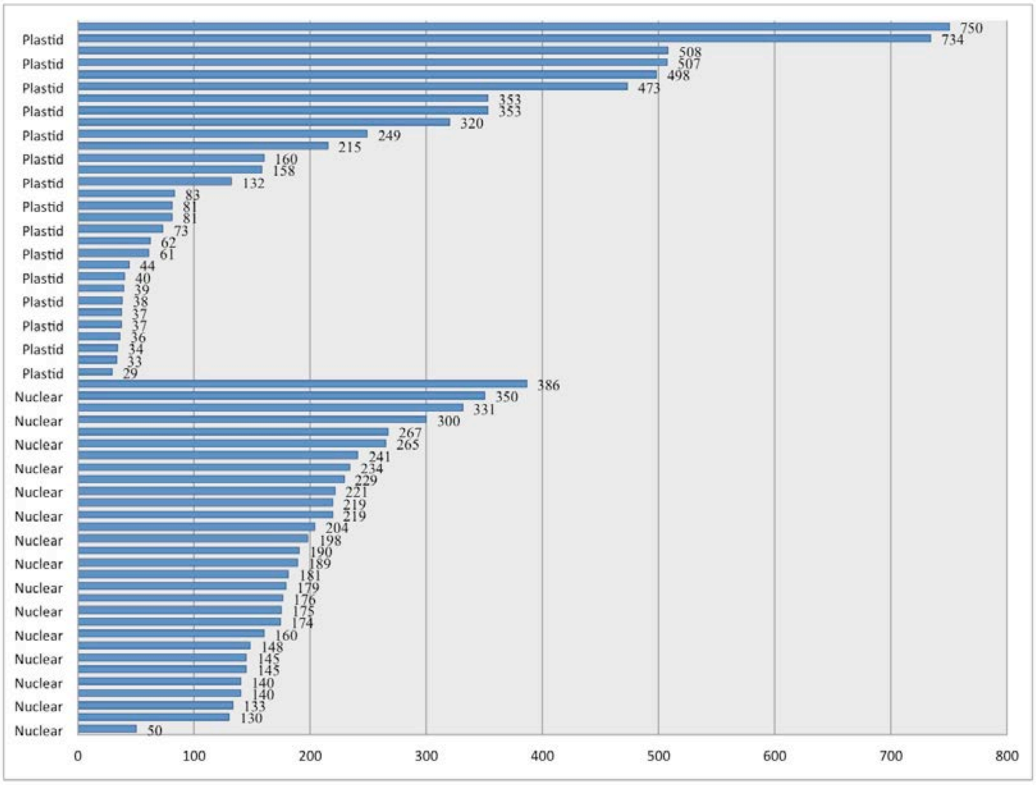

Figure S2. Length of chloroplast proteins. Proteins involved in the photosynthetic apparatus of *Arabidopsis thaliana* are sorted according to their amino acid length for both plastid and nuclear encoded proteins.

Figure S3

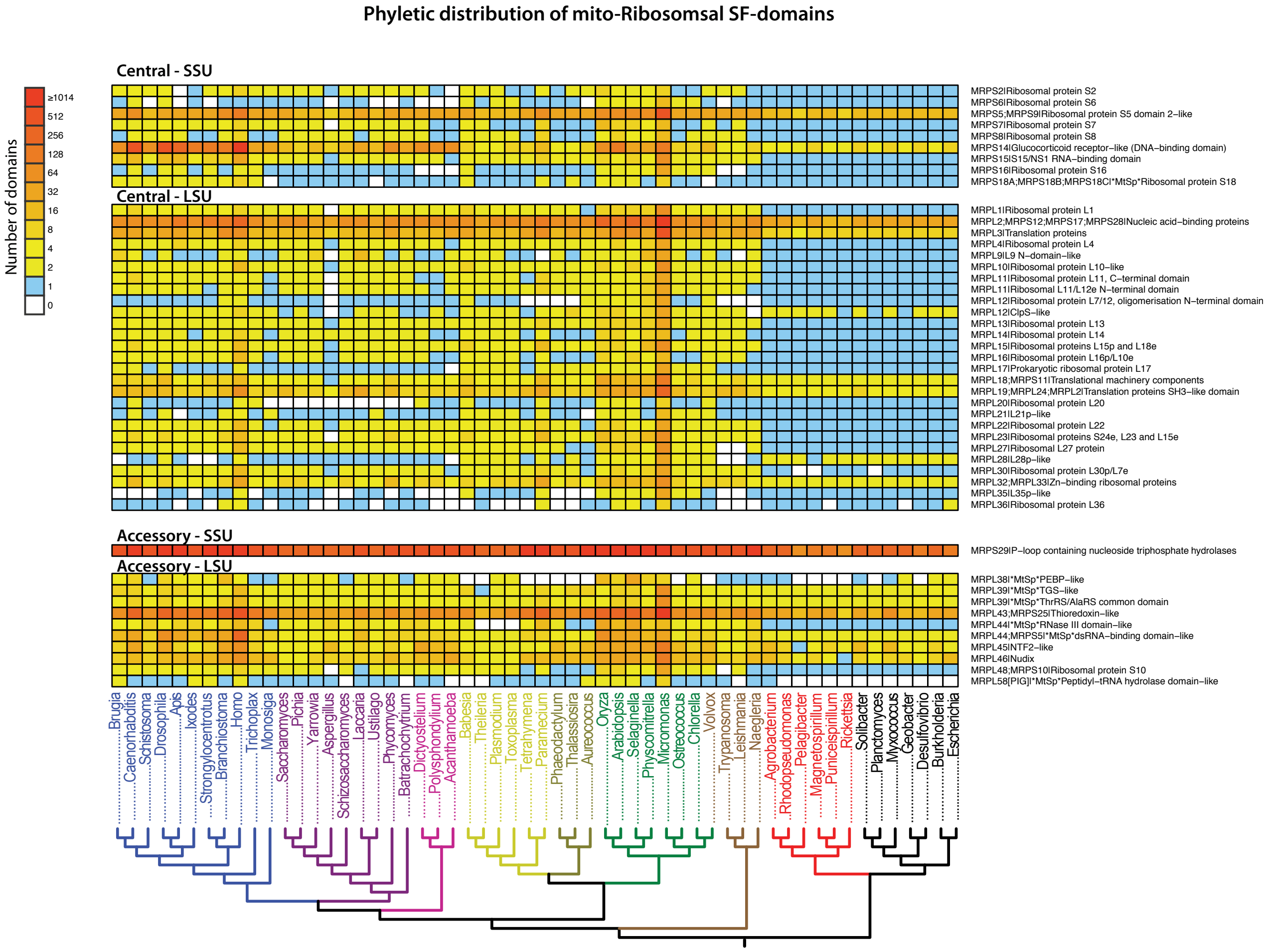

Figure S3. Phyletic distribution of SF-domains of mitochondrial ribosomal proteins in eukaryotes and bacteria. The number of central and accessory SF-domains identified in each species of eukaryotes and bacteria are compared along the phylogeny. Columns correspond to the species in the phylogenetic tree and rows are individual mito-ribosomal SF-domains. The colors indicate the number of domains in each species if present and boxes are white if a SF-domain is absent.

Figure S4

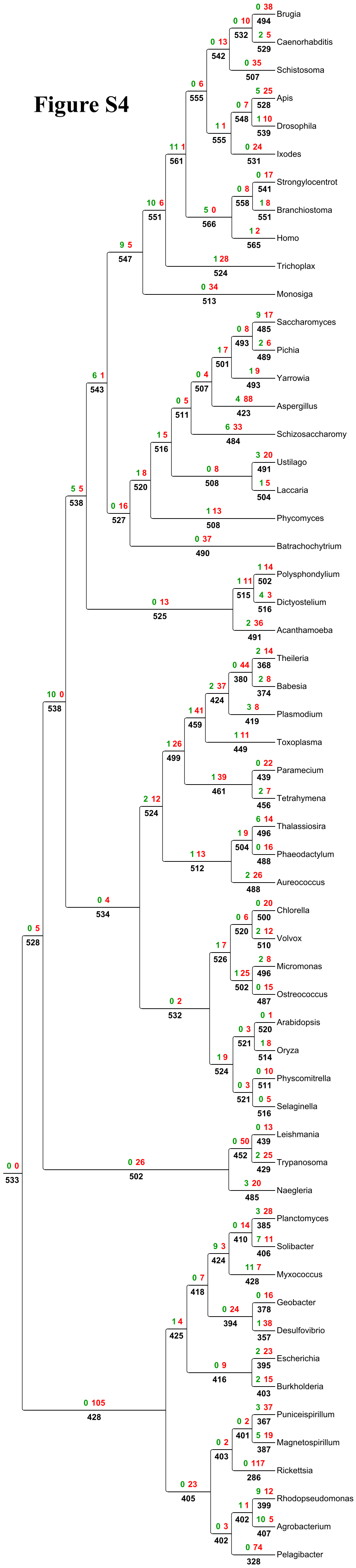

Figure S4. Flux of SF-domains during evolution in the "mitochondrial proteome". Parsimonious reconstruction of flux SF-domains that are associated with mitochondria were mapped onto the eukaryote phylogeny. The number of superfamilies inferred to have been gained (green), lost (red) and the total number present at each node or branch (black) are indicated.

Table S1. Mitochondrial proteins involved in the respiratory chain complex in humans. The proteins are listed according to the complexes in which they occur. Encoded = the genome in which the gene is located (mitochondrial or nuclear); Uniprot ID = the Uniprot identifier of the protein; TMD start = the location of the first hydrophobic segment or the location of the most hydrophobic marginally hydrophobic segment if no hydrophobic segment is present; TMD = length of the transmembrane domain; Hydro = hydrophobicity (free insertion energy (dG, kcal/mol)) of the selected transmembrane domain; Tail = length of the selected transmembrane domain and the C-terminal tail; SRP Target = the prediction of whether the protein would be a target by SRP.

Complex	**Encoded**	**Uniprot ID**	**TMD start**	**TMD**	**Hydro**	**Tail**	**SRP Target**
Complex I	Mito	P03886	3	21	-2,422	295	Yes
Complex I	Mito	P03891	200	22	-0,887	126	Yes
Complex I	Mito	P03897	3	22	-3,46	91	No
Complex I	Mito	P03905	94	23	-0,372	343	Yes
Complex I	Mito	P03901	1	23	-1,045	75	No
Complex I	Mito	P03915	80	22	-0,156	502	Yes
Complex I	Mito	P03923	1	20	-0,846	154	Yes
Complex I	Nuclear	O15239	5	22	-0,056	44	No
Complex I	Nuclear	O43678	n/a	n/a	n/a	n/a	No
Complex I	Nuclear	O95167	17	19	1,374	49	No
Complex I	Nuclear	Q16718	n/a	n/a	n/a	n/a	No
Complex I	Nuclear	P56556	n/a	n/a	n/a	n/a	No
Complex I	Nuclear	O95182	n/a	n/a	n/a	n/a	No
Complex I	Nuclear	P51970	n/a	n/a	n/a	n/a	No
Complex I	Nuclear	Q16795	280	23	1,035	75	No
Complex I	Nuclear	O95299	n/a	n/a	n/a	n/a	No
Complex I	Nuclear	Q86Y39	59	20	1,027	63	No
Complex I	Nuclear	Q9UI09	n/a	n/a	n/a	n/a	No
Complex I	Nuclear	Q9P0J0	27	22	1,475	96	No
Complex I	Nuclear	O14561	n/a	n/a	n/a	n/a	No
Complex I	Nuclear	O75438	12	19	2,626	28	No
Complex I	Nuclear	O95178	55	20	0,936	31	No
Complex I	Nuclear	O43676	70	19	0,967	10	No
Complex I	Nuclear	O95168	88	20	0,351	22	No
Complex I	Nuclear	O43674	70	23	-1,483	97	No
Complex I	Nuclear	O95139	69	20	1,259	40	No
Complex I	Nuclear	P17568	n/a	n/a	n/a	n/a	No
Complex I	Nuclear	O95169	128	23	-1,106	36	No
Complex I	Nuclear	Q9Y6M9	n/a	n/a	n/a	n/a	No
Complex I	Nuclear	O96000	n/a	n/a	n/a	n/a	No
Complex I	Nuclear	Q9NX14	87	23	-2,342	44	No
Complex I	Nuclear	O43677	41	19	-0,043	17	No
Complex I	Nuclear	O95298	59	19	0,354	42	No
Complex I	Nuclear	P28331	n/a	n/a	n/a	n/a	No
Complex I	Nuclear	O75306	184	21	2,463	259	No

Complex I	Nuclear	O75489	7	20	1,782	238	No
Complex I	Nuclear	O43181	13	23	1,92	140	No
Complex I	Nuclear	O43920	n/a	n/a	n/a	n/a	No
Complex I	Nuclear	O75380	n/a	n/a	n/a	n/a	No
Complex I	Nuclear	O75251	79	20	2,575	115	No
Complex I	Nuclear	O00217	55	23	2,223	133	No
Complex I	Nuclear	P49821	n/a	n/a	n/a	n/a	No
Complex I	Nuclear	P19404	n/a	n/a	n/a	n/a	No
Complex I	Nuclear	P56181	n/a	n/a	n/a	n/a	No
Complex II	Nuclear	P31040	451	21	2,76	193	No
Complex II	Nuclear	P21912	180	22	2,152	79	No
Complex II	Nuclear	Q99643	148	22	0,485	0	No
Complex II	Nuclear	O14521	1	23	-0,111	136	Yes
Complex III	Mito	P00156	29	23	-1,019	329	Yes
Complex III	Nuclear	P08574	287	22	-0,187	17	No
Complex III	Nuclear	Q9UDW1	14	23	-0,677	27	No
Complex III	Nuclear	O14957	20	23	-0,074	14	No
Complex III	Nuclear	P14927	n/a	n/a	n/a	n/a	No
Complex III	Nuclear	P31930	n/a	n/a	n/a	n/a	No
Complex III	Nuclear	P22695	213	19	2,829	222	No
Complex III	Nuclear	P47985	210	19	2,574	46	No
Complex III	Nuclear	P07919	n/a	n/a	n/a	v	No
Complex III	Nuclear	O14949	46	19	0,986	18	No
Complex IV	Nuclear	P13073	100	22	-0,394	48	No
Complex IV	Nuclear	Q96KJ9	102	23	-2,249	47	No
Complex IV	Nuclear	P20674	n/a	n/a	n/a	n/a	No
Complex IV	Nuclear	P10606	n/a	n/a	n/a	n/a	No
Complex IV	Nuclear	P12074	37	23	0,665	50	No
Complex IV	Nuclear	Q02221	25	21	-0,671	52	No
Complex IV	Nuclear	P14854	n/a	n/a	n/a	n/a	No
Complex IV	Nuclear	Q6YFQ2	n/a	n/a	n/a	n/a	No
Complex IV	Nuclear	P09669	18	23	0,179	35	No
Complex IV	Nuclear	P24310	51	23	0,396	6	No
Complex IV	Nuclear	P14406	52	23	0,424	9	No
Complex IV	Nuclear	P24311	40	19	1,126	22	No
Complex IV	Nuclear	Q8TF08	41	19	1,138	22	No
Complex IV	Nuclear	P15954	37	21	1,51	6	No
Complex IV	Nuclear	P10176	40	20	0,218	10	No
Complex IV	Nuclear	Q7Z4L0	39	23	0,821	11	No
Complex IV	Mito	P00395	17	23	-1,009	474	Yes
Complex IV	Mito	P00403	27	22	-4,536	179	Yes
Complex IV	Mito	P00414	79	23	-1,361	160	Yes
Complex V	Nuclear	P25705	n/a	n/a	n/a	n/a	No
Complex V	Nuclear	P06576	n/a	n/a	n/a	n/a	No
Complex V	Nuclear	P36542	n/a	n/a	n/a	n/a	No

Complex V	Nuclear	P30049	62	20	2,639	87	No
Complex V	Nuclear	P56381	n/a	n/a	n/a	n/a	No
Complex V	Nuclear	P24539	92	23	1,678	142	No
Complex V	Nuclear	P05496	115	22	-1,862	0	No
Complex V	Nuclear	Q06055	120	22	-1,862	0	No
Complex V	Nuclear	P48201	121	22	-1,862	0	No
Complex V	Nuclear	O75947	n/a	n/a	n/a	n/a	No
Complex V	Nuclear	P56385	11	23	1,507	36	No
Complex V	Nuclear	P18859	n/a	n/a	n/a	n/a	No
Complex V	Nuclear	P56134	63	20	0,868	12	No
Complex V	Nuclear	O75964	68	22	2,079	14	No
Complex V	Nuclear	Q7Z4Y8	68	23	2,772	10	No
Complex V	Nuclear	P48047	n/a	n/a	n/a	n/a	No
Complex V	Nuclear	Q9UII2	1	23	2,802	83	No
Complex V	Mito	P00846	9	23	-1,173	217	Yes
Complex V	Mito	P03928	7	23	-0,091	61	No

Table S2. Mitochondrial membrane proteins with determined structures. PDB+Chain = the PDB structure identifier and the specific chain of the protein; Encoded = the genome in which the gene is located (mitochondrial or nuclear); Hydro = hydrophobicity (free insertion energy (dG, kcal/mol)) of the selected transmembrane domain; Uniprot ID = the Uniprot identifier of the protein; TMD = length of the transmembrane domain; Tail = length of the selected transmembrane domain and the C-terminal tail; SRP Target = the prediction of whether the protein would be a target by SRP.

PDB+chain	**Encoded**	**Hydro**	**Uniprot ID**	**TMD**	**Tail**	**SRP Target**
1PP9:C	Mito	-2,009	P00157	23	349	Yes
1V55:A	Mito	-0,705	P00396	23	497	Yes
1V55:B	Mito	-2,178	P68530	23	199	Yes
1V55:C	Mito	-2,083	P00415	23	182	Yes
2XOK:K	Mito	-0,927	P61830	23	64	No
3CX5:C	Mito	-0,889	P00163	23	357	Yes
3H1J:C	Mito	-1,127	P18946	23	350	Yes
1V55:L	Nuclear	-0,901	P00430	23	28	No
1V55:J	Nuclear	-0,475	P07470	23	29	No
1V55:M	Nuclear	0,518	P10175	20	30	No
1YQ3:D	Nuclear	0,715	Q5ZIS0	20	33	No
3VR8:C	Nuclear	-1,497	P92506	22	36	No
1ZOY:D	Nuclear	0,109	A5GZW8	20	36	No
2FYU:K	Nuclear	0,286	P07552	23	36	No
1L0L:K	Nuclear	0,621	P07552	23	36	No
2YBB:k	Nuclear	1,785	P07552	23	36	No
3H1J:G	Nuclear	2,932	D0VX32	19	36	No
3H1J:D	Nuclear	-0,554	D0VX26	20	38	No
1L0L:D	Nuclear	-0,483	P00125	20	38	No
3VR8:D	Nuclear	-1,311	P92507	23	39	No
1V55:K	Nuclear	0,904	P13183	19	40	No
1L0L:G	Nuclear	1,043	P13271	23	40	No
3CX5:D	Nuclear	-0,012	P07143	20	41	No
3CX5:H	Nuclear	0,777	P08525	23	45	No
2LOS:A	Nuclear	0,997	Q9P0S9	20	50	No
3H1J:J	Nuclear	-0,553	N/A	23	51	No
3CX5:I	Nuclear	1,002	P22289	23	56	No
1V55:I	Nuclear	-0,384	P04038	23	58	No
1L0L:E	Nuclear	2,574	P13272	19	64	No
1V55:D	Nuclear	-1,332	P00423	21	69	No
2YBB:R	Nuclear	-0,142	P07471	23	69	No
1OKC:A	Nuclear	1,956	P02722	22	88	No
1ZOY:C	Nuclear	0,47	D0VWV4	21	97	No
1YQ3:C	Nuclear	0,127	D0VWW3	21	98	No
1YQ3:B	Nuclear	2,152	Q9YHT2	22	100	No
4C9G:A	Nuclear	0,249	P18239	23	101	No
3VR8:B	Nuclear	1,89	O44074	19	103	No
2YBB:8	Nuclear	1,968	Q56224	21	132	No

2YBB:6	Nuclear	2,225	Q56218	19	145	No
2YBB:3	Nuclear	2,971	Q56223	20	146	No
3H1J:E	Nuclear	1,989	Q5ZLR5	19	162	No
3CX5:E	Nuclear	2,172	P08067	19	162	No
1YQ3:A	Nuclear	2,147	Q9YHT1	22	213	No
1ZOY:A	Nuclear	2,598	Q0QF01	22	213	No
1L0L:B	Nuclear	2,829	P23004	19	240	No
2YBB:4	Nuclear	0,887	Q56220	22	279	No
2LCK:A	Nuclear	1,73	P70406	22	302	No
2YBB:1	Nuclear	1,934	Q56222	23	325	No
4AYX:A	Nuclear	-0,576	Q9NRK6	23	534	Yes
4MYC:A	Nuclear	-0,32	P40416	19	589	Yes
3VR8:A	Nuclear	2,371	Q33862	20	604	No

Table S3. Proteins localized to the mitochondrion with the aid of green fluorescent protein tags. Uniprot ID = the Uniprot identifier of the protein; Species = the species in which the protein was identified; Hydro = hydrophobicity (free insertion energy (dG, kcal/mol)) of the selected transmembrane domain; Tail = length of the selected transmembrane domain and the C-terminal tail; SRP Target = the prediction if the protein would be a targeted by SRP.

Uniprot ID	**Species**	**Hydro**	**Tail**	**SRP Target**
Q8WVC6	Human	-1,309	29	No
Q9Y5T4	Human	-1,25	114	No
Q96LL9	Human	-2,214	20	No
Q9NNZ3	Human	-1,048	85	No
Q9H4I9	Human	-0,129	42	No
Q9HD23	Human	-2,135	71	No
Q9NWS8	Human	-0,258	23	No
Q9H1U9	Human	-0,19	86	No
Q96IX5	Human	-0,21	28	No
Q86XE5	Human	0,989	101	No
Q9NVI7	Human	1,362	541	No
Q5T9A4	Human	1,414	401	No
Q8NE86	Human	0,131	117	No
Q8IYU8	Human	2,227	401	No
Q96EY8	Human	2,546	238	No
Q9H8M1	Human	2,797	206	No
Q96I36	Human	1,879	43	No
Q9P0S2	Human	0,789	91	No
Q5RI15	Human	2,552	84	No
Q86YH6	Human	2,802	293	No
Q96CU9	Human	1,287	421	No
Q14296	Human	2,59	396	No
B2RBJ8	Human	2,802	362	No
Q6P1Q0	Human	2,664	279	No
Q8IV42	Human	2,992	332	No
Q9H7H0	Human	2,519	448	No
H7BXW7	Human	0,562	59	No
Q96BW9	Human	2,304	21	No
Q567V2	Human	1,72	188	No
O75414	Human	2,885	120	No
Q86X67	Human	2,619	339	No
P49590	Human	2,64	203	No
Q7L3T8	Human	2,793	142	No
Q96ND0	Human	1,089	137	No
Q3KR05	Human	2,916	431	No
Q96CQ1	Human	1,902	305	No
Q9NX18	Human	2,48	154	No
P17152	Human	1,652	109	No
Q9BUB7	Human	2,109	259	No

Q96GJ1	Human	2,154	491	No
Q8WWC4	Human	1,178	252	No
Q9BQ52	Human	2,988	718	No
Q96LL9	Human	-2,214	20	No
A4D1T6	Human	-1,23	529	Yes
Q8IW45	Human	-0,619	342	Yes
Q96E52	Human	-1,343	331	Yes
Q9BVV7	Human	-2,004	140	Yes
P43617	Yeast	2,029	257	No
Q03829	Yeast	1,698	258	No
Q12251	Yeast	1,919	201	No
P40098	Yeast	0,856	217	No
Q06089	Yeast	1,529	155	No
P38172	Yeast	1,545	256	No
Q12082	Yeast	0,56	43	No
P36101	Yeast	0,252	441	No
Q3E7A6	Yeast	2,628	22	No
P40166	Yeast	2,67	99	No
Q12106	Yeast	2,913	250	No
P47131	Yeast	2,145	100	No
P32792	Yeast	0,312	309	No
P07806	Yeast	0,878	442	No
P10662	Yeast	2,404	184	No
P28817	Yeast	2,31	300	No
P39525	Yeast	2,806	441	No
P32378	Yeast	1,659	279	No
P36521	Yeast	2,853	54	No
P53163	Yeast	2,871	181	No
P36523	Yeast	2,472	64	No
Q12487	Yeast	2,152	137	No
P36516	Yeast	2,453	137	No
P09950	Yeast	2,631	185	No
Q06567	Yeast	1,823	538	No
P32874	Yeast	1,315	200	No
P04710	Yeast	1,236	192	No
P18238	Yeast	2,781	295	No
P07246	Yeast	2,933	204	No
P38884	Yeast	1,111	269	No
P40502	Yeast	2,249	81	No
Q03798	Yeast	0,074	194	No
P50945	Yeast	1,624	133	No
P38885	Yeast	1,783	266	No
P38341	Yeast	0,519	99	No
P25345	Yeast	2,604	350	No
Q01802	Yeast	1,322	328	No
P32580	Yeast	2,746	736	No

P05626	Yeast	2,425	182	No
P81449	Yeast	1,191	90	No
Q06405	Yeast	1,244	29	No
Q12233	Yeast	2,654	31	No
Q12349	Yeast	2,533	120	No
P81450	Yeast	1,745	50	No
P81451	Yeast	1,384	51	No
Q06485	Yeast	1,915	194	No
P32451	Yeast	2,59	78	No
P09440	Yeast	2,511	671	No
Q07560	Yeast	1,046	210	No
P43635	Yeast	2,934	439	No
P00890	Yeast	1,893	57	No
P07390	Yeast	0,796	415	No
P37299	Yeast	0,874	47	No
P08525	Yeast	0,777	45	No
P22289	Yeast	1,002	56	No
P08067	Yeast	2,172	162	No
P07252	Yeast	1,19	501	No
P14066	Yeast	2,997	45	No
P14905	Yeast	2,5	382	No
P38909	Yeast	0,369	336	No
Q3E7B2	Yeast	0,403	53	No
P39103	Yeast	0,573	54	No
P04039	Yeast	0,691	29	No
Q04935	Yeast	0,6	118	No
P32799	Yeast	0,666	79	No
P00431	Yeast	0,439	343	No
P54783	Yeast	1,609	357	No
P07275	Yeast	2,467	364	No
P09624	Yeast	1,665	294	No
P32891	Yeast	1,479	544	No
P46681	Yeast	2,388	274	No
P04819	Yeast	2,717	445	No
P25846	Yeast	1,435	180	No
P15801	Yeast	2,803	1071	No
P25039	Yeast	2,02	600	No
P39112	Yeast	1,779	231	No
P40215	Yeast	1,494	490	No
P36112	Yeast	1,089	505	No
P33893	Yeast	2,379	465	No
P32191	Yeast	1,286	644	No
P26364	Yeast	1,917	146	No
P0CS90	Yeast	2,15	241	No
P49367	Yeast	2,118	252	No
P28834	Yeast	2,55	111	No

P48526	Yeast	2,348	736	No
P38169	Yeast	0,571	455	No
P11325	Yeast	2,916	341	No
Q06510	Yeast	0,939	358	No
Q01926	Yeast	1,334	155	No
P53077	Yeast	0,351	36	No
P22438	Yeast	2,221	264	No
P32785	Yeast	2,376	47	No
Q03028	Yeast	1,964	299	No
Q99297	Yeast	2,753	295	No
Q12289	Yeast	2,332	225	No
P38702	Yeast	0,968	224	No
P53320	Yeast	2,4	351	No
P38127	Yeast	2,703	324	No
P32839	Yeast	1,013	405	No
P38323	Yeast	2,232	254	No
Q06143	Yeast	2,212	283	No
P18409	Yeast	2,436	260	No
Q92328	Yeast	1,882	111	No
P42834	Yeast	0,808	145	No
Q04013	Yeast	2,74	37	No
P40464	Yeast	1,108	238	No
P48237	Yeast	2,123	423	No
P38988	Yeast	2,951	236	No
P39515	Yeast	2,663	141	No
Q08749	Yeast	2,145	109	No
P32897	Yeast	1,393	125	No
Q01852	Yeast	2,809	411	No
Q02776	Yeast	0,398	364	No
P47045	Yeast	1,368	440	No
P33448	Yeast	0,547	29	No
P53507	Yeast	2,592	33	No
P07213	Yeast	1,026	608	No
P32795	Yeast	1,063	517	No
P46972	Yeast	1,058	168	No
P53239	Yeast	1,261	265	No
P35999	Yeast	2,48	252	No
P36046	Yeast	0,537	356	No
P46998	Yeast	0,355	146	No
P36163	Yeast	0,457	280	No
Q03218	Yeast	1,71	349	No
P53944	Yeast	2,358	196	No
P39953	Yeast	1,981	296	No
Q12375	Yeast	2,191	185	No
P38325	Yeast	2,261	63	No
P16547	Yeast	1,093	392	No

P32332	Yeast	1,795	302	No
P23641	Yeast	1,971	100	No
P40035	Yeast	2,841	283	No
P32898	Yeast	2,89	134	No
P53157	Yeast	1,93	77	No
P38857	Yeast	2,33	77	No
P53311	Yeast	1,882	94	No
P10566	Yeast	1,935	132	No
P23500	Yeast	2,185	132	No
P53257	Yeast	1,272	300	No
P10849	Yeast	2,732	119	No
P53070	Yeast	2,637	668	No
P38297	Yeast	2,725	163	No
Q02883	Yeast	0,084	416	No
P36013	Yeast	2,545	416	No
P36060	Yeast	0,007	290	No
Q06892	Yeast	2,266	249	No
Q12374	Yeast	0,825	517	No
P38969	Yeast	0,269	225	No
P32522	Yeast	2,898	838	No
P33751	Yeast	2,298	232	No
P08425	Yeast	2,279	138	No
P38812	Yeast	2,19	42	No
P46367	Yeast	2,551	324	No
P36147	Yeast	1,878	128	No
P53153	Yeast	0,111	262	No
Q12029	Yeast	2,734	228	No
Q12204	Yeast	2,78	194	No
P27697	Yeast	1,82	271	No
P40961	Yeast	2,372	277	No
P09368	Yeast	1,329	403	No
P32317	Yeast	2,854	294	No
Q01217	Yeast	2,166	747	No
P21560	Yeast	2,078	183	No
Q06011	Yeast	0,276	76	No
P38300	Yeast	2,321	204	No
P28737	Yeast	1,185	353	No
P38958	Yeast	0,001	93	No
P08468	Yeast	2,199	673	No
P10355	Yeast	2,194	209	No
P38796	Yeast	2,527	206	No
P23833	Yeast	0,479	224	No
P38825	Yeast	1,608	624	No
P38851	Yeast	2,514	726	No
P21592	Yeast	0,502	310	No
P40012	Yeast	1,547	528	No

Q06683	Yeast	2,703	684	No
P80235	Yeast	1,975	419	No
P38921	Yeast	2,212	283	No
Q06236	Yeast	0,723	107	No
P40530	Yeast	2,477	320	No
P53170	Yeast	2,948	314	No
P32473	Yeast	2,97	273	No
Q03713	Yeast	1,168	126	No
P53721	Yeast	1,434	204	No
P53259	Yeast	0,511	239	No
Q02773	Yeast	2,464	935	No
P25270	Yeast	2,547	155	No
Q07534	Yeast	2,282	293	No
P14693	Yeast	1,935	168	No
P50110	Yeast	1,935	312	No
P37298	Yeast	1,136	112	No
Q00711	Yeast	2,683	587	No
P21801	Yeast	2,152	93	No
P53312	Yeast	2,949	76	No
P00927	Yeast	1,707	346	No
P46943	Yeast	2,621	490	No
P38152	Yeast	2,997	193	No
P43122	Yeast	2,668	157	No
D6W196	Yeast	1,289	270	No
P04803	Yeast	2,033	252	No
P48527	Yeast	2,259	394	No
P53318	Yeast	2,677	452	No
O13525	Yeast	1,484	316	No
P53208	Yeast	1,252	221	No
P43557	Yeast	-0,86	23	No
P37267	Yeast	-0,181	138	Yes
P40452	Yeast	-0,467	127	Yes
P35180	Yeast	-0,493	176	Yes
P25573	Yeast	-1,086	360	Yes
Q04472	Yeast	-0,176	424	Yes
P08466	Yeast	-0,542	327	Yes
Q06417	Yeast	-0,249	321	Yes
Q02981	Yeast	-1,426	538	Yes
P33310	Yeast	-0,189	594	Yes
P33311	Yeast	-0,929	657	Yes
P19516	Yeast	-0,912	217	Yes
P40086	Yeast	-1,602	399	Yes
P40416	Yeast	-0,32	584	Yes
Q06493	Yeast	-0,338	319	Yes
Q06668	Yeast	-0,89	366	Yes
Q12171	Yeast	-1,185	492	Yes

Q08179	Yeast	-0,298	436	Yes
P32843	Yeast	-1,079	564	Yes
P53220	Yeast	-2,153	165	Yes
P39952	Yeast	-0,549	276	Yes
P39925	Yeast	-0,619	645	Yes
P40341	Yeast	-0,717	647	Yes
P21375	Yeast	-1,861	496	Yes
P39006	Yeast	-0,509	423	Yes
P50085	Yeast	-0,257	273	Yes
Q08023	Yeast	-0,669	502	Yes

Table S4. Chloroplast proteins in the photosynthetic apparatus of *Arabidopsis thaliana*. Uniprot ID = the Uniprot identifier of the protein; Encoded = the genome in which the gene is located (plastid or nuclear); Hydro = hydrophobicity (free insertion energy (dG, kcal/mol)) of the selected transmembrane domain; Tail = length of the selected transmembrane domain and the C-terminal tail; SRP Target = the prediction of whether the protein would be a target by SRP.

Uniprot ID	**Encoded**	**Hydro**	**Tail**	**SRP Target**
P56773	Plastid	-0.150	183	Yes
P56774	Plastid	-0.930	124	Yes
P56771	Plastid	-1.377	36	No
Q9ZR03	Nuclear	0.707	161	No
P61039	Plastid	-1.587	26	No
P56775	Plastid	-2.118	32	No
P83755	Plastid	-0.922	326	Yes
P56761	Plastid	-0.646	325	Yes
P56778	Plastid	-2.242	313	Yes
P56777	Plastid	-0.262	412	Yes
P56779	Plastid	-0.291	66	No
P62095	Plastid	1.902	30	No
P60129	Plastid	-3.299	22	No
P56781	Plastid	-2.472	30	No
P56782	Plastid	-0.261	25	No
P62109	Plastid	-3.023	29	No
P56780	Plastid	-0.565	36	No
P62100	Plastid	-3.234	34	No
Q9S841	Nuclear	2.221	267	No
Q42029	Nuclear	1.400	166	No
Q9XI73	Nuclear	0.914	147	No
P27202	Nuclear	-1.206	29	No
Q9XF91	Nuclear	-1.515	133	Yes
P61839	Plastid	-2.970	30	No
Q39194	Nuclear	-2.586	33	No
O49347	Nuclear	-0.363	34	No
P56790	Plastid	-1.425	61	No
Q9LR64	Nuclear	2.233	135	No
Q8W0Y8	Nuclear	N/A	0	No
P56766	Plastid	-2.200	594	Yes
P56767	Plastid	-1.801	599	Yes
P62090	Plastid	N/A	N/A	No
Q9SA56	Nuclear	N/A	N/A	No
Q9S714	Nuclear	2.957	144	No
Q9SHE8	Nuclear	-0.104	71	No
Q9S7N7	Nuclear	1.155	98	No
Q9SUI6	Nuclear	0.657	43	No
P56768	Plastid	0.340	30	No
P56769	Plastid	0.898	31	No

Q9SUI5	Nuclear	1.127	75	No
Q9SUI4	Nuclear	-0.803	85	No
P49107	Nuclear	1.497	118	No
Q949Q5	Nuclear	-1.677	30	No
P42699	Nuclear	1.259	125	No
O04090	Nuclear	N/A	N/A	No
Q8W493	Nuclear	2.063	132	No
Q93VA3	Nuclear	2.642	132	No
P19366	Plastid	N/A	N/A	No
P56757	Plastid	N/A	N/A	No
Q01909	Nuclear	N/A	N/A	No
Q9SSS9	Nuclear	2.904	49	No
P09468	Plastid	2.246	84	No
P56760	Plastid	-1.622	28	No
P56758	Plastid	-1.434	208	Yes
P93303	Plastid	-2.179	151	Yes
Q01667	Nuclear	0.591	108	No
P0CJ48	Nuclear	1.336	109	No
Q9SKT0	Nuclear	1.945	104	No
A0A023PKQ4	Plastid	-1.736	25	No
B6ZHE7	Nuclear	1.885	178	No

Table S5. SF-domains assigned to MitoCOGs. Proteins which were grouped as hypothetical proteins according to MitoCOG clustering are highlighted in bold

MitotCOG-ID	Protein length min	max	MitoCOG Description	SF Classification	SF Description
MitoCOG0001	166	805	NADH dehydrogenase subunit 2	f.33.1	Calcium ATPase, transmembrane domain M
MitoCOG0002	400	873	cytochrome c oxidase subunit 1	f.24.1	Cytochrome c oxidase subunit I-like
MitoCOG0002	400	873	cytochrome c oxidase subunit 1	b.6.1	Cupredoxins
MitoCOG0002	400	873	cytochrome c oxidase subunit 1	f.17.2	Cytochrome c oxidase subunit II-like, transmembrane region
MitoCOG0002	400	873	cytochrome c oxidase subunit 1	d.95.2	Homing endonucleases
MitoCOG0003	130	1296	cytochrome c oxidase subunit 2	b.6.1	Cupredoxins
MitoCOG0003	130	1296	cytochrome c oxidase subunit 2	f.17.2	Cytochrome c oxidase subunit II-like, transmembrane region
MitoCOG0003	130	1296	cytochrome c oxidase subunit 2	d.95.2	Homing endonucleases
MitoCOG0003	130	1296	cytochrome c oxidase subunit 2	h.1.5	Tropomyosin
MitoCOG0004	112	410	ATP synthase subunit 6	f.18.1	F1F0 ATP synthase subunit A
MitoCOG0005	212	441	cytochrome c oxidase subunit 3	f.25.1	Cytochrome c oxidase subunit III-like
MitoCOG0006	86	567	NADH dehydrogenase subunit 3	f.24.1	Cytochrome c oxidase subunit I-like
MitoCOG0008	134	770	NADH dehydrogenase subunit 4	f.4.3	Porins
MitoCOG0011	196	497	cytochrome b	f.21.1	Transmembrane di-heme cytochromes
MitoCOG0011	196	497	cytochrome b	f.32.1	a domain/subunit of cytochrome bc1 complex (Ubiquinol-cytochrome c reductase)
MitoCOG0013	138	1057	LAGLIDADG endonuclease	d.95.2	Homing endonucleases
MitoCOG0013	138	1057	LAGLIDADG endonuclease	f.24.1	Cytochrome c oxidase subunit I-like
MitoCOG0013	138	1057	LAGLIDADG endonuclease	f.21.1	Transmembrane di-heme cytochromes
MitoCOG0013	138	1057	LAGLIDADG endonuclease	e.8.1	DNA/RNA polymerases
MitoCOG0014	54	203	ATP synthase subunit 9	f.17.1	F1F0 ATP synthase subunit C
MitoCOG0015	979	1021	DNA mismatch repair protein MutS	a.113.1	DNA repair protein MutS, domain III
MitoCOG0015	979	1021	DNA mismatch repair protein MutS	c.37.1	P-loop containing nucleoside triphosphate hydrolases
MitoCOG0015	979	1021	DNA mismatch repair protein MutS	d.75.2	DNA repair protein MutS, domain I
MitoCOG0016	745	905	reverse transcriptase and maturase	e.8.1	DNA/RNA polymerases
MitoCOG0017	**501**	**677**	**hypothetical protein**	**d.95.2**	**Homing endonucleases**
MitoCOG0018	106	1042	DNA polymerase	e.8.1	DNA/RNA polymerases
MitoCOG0018	106	1042	DNA polymerase	c.55.3	Ribonuclease H-like
MitoCOG0019	100	1065	reverse transcriptase	e.8.1	DNA/RNA polymerases
MitoCOG0019	100	1065	reverse transcriptase	f.24.1	Cytochrome c oxidase subunit I-like
MitoCOG0019	100	1065	reverse transcriptase	f.21.1	Transmembrane di-heme cytochromes
MitoCOG0020	558	745	reverse transcriptase	e.8.1	DNA/RNA polymerases
MitoCOG0025	151	371	ribosomal protein S3	d.53.1	Ribosomal protein S3 C-terminal domain
MitoCOG0025	151	371	ribosomal protein S3	d.52.3	Prokaryotic type KH domain (KH-domain type II)
MitoCOG0026	64	130	ribosomal protein S19	d.28.1	Ribosomal protein S19
MitoCOG0027	133	521	ribosomal protein L2	b.40.4	Nucleic acid-binding proteins
MitoCOG0027	133	521	ribosomal protein L2	b.34.5	Translation proteins SH3-like domain
MitoCOG0029	154	182	NADH dehydrogenase subunit 10	e.19.1	HydA/Nqo6-like
MitoCOG0030	79	215	ribosomal protein S12	b.40.4	Nucleic acid-binding proteins
MitoCOG0031	155	443	NADH dehydrogenase subunit 7	e.8.1	HydB/Nqo4-like
MitoCOG0032	**175**	**179**	**hypothetical protein**	**d.141.1**	**Ribosomal protein L6**
MitoCOG0033	**234**	**330**	**hypothetical protein**	**d.53.1**	**Ribosomal protein S3 C-terminal domain**
MitoCOG0036	**273**	**314**	**hypothetical protein**	**a.75.1**	**Ribosomal protein S7**
MitoCOG0039	133	302	ribosomal protein L16	d.41.4	Ribosomal protein L16p/L10e
MitoCOG0043	177	287	NADH dehydrogenase subunit 9	d.307.1	Nqo5-like
MitoCOG0043	177	287	NADH dehydrogenase subunit 9	g.39.1	Glucocorticoid receptor-like (DNA-binding domain)
MitoCOG0044	**1260**	**1636**	**hypothetical protein**	**d.53.1**	**Ribosomal protein S3 C-terminal domain**
MitoCOG0048	100	178	ribosomal protein L14	b.39.1	Ribosomal protein L14
MitoCOG0051	95	282	succinate dehydrogenase subunit 3	f.21.2	Fumarate reductase respiratory complex transmembrane subunits
MitoCOG0052	235	270	succinate dehydrogenase subunit 2	a.1.2	alpha-helical ferredoxin
MitoCOG0052	235	270	succinate dehydrogenase subunit 2	d.15.4	2Fe-2S ferredoxin-like
MitoCOG0053	99	307	ribosomal protein L6	d.141.1	Ribosomal protein L6
MitoCOG0054	113	593	ribosomal protein S11	c.55.4	Translational machinery components
MitoCOG0055	87	119	ribosomal protein S14	g.39.1	Glucocorticoid receptor-like (DNA-binding domain)
MitoCOG0056	186	1775	ribosomal protein S3	d.53.1	Ribosomal protein S3 C-terminal domain
MitoCOG0056	186	1775	ribosomal protein S3	d.52.3	Prokaryotic type KH domain (KH-domain type II)
MitoCOG0058	119	919	LAGLIDADG endonuclease	d.95.2	Homing endonucleases
MitoCOG0058	119	919	LAGLIDADG endonuclease	d.285.1	DNA-binding domain of intron-encoded endonucleases
MitoCOG0058	119	919	LAGLIDADG endonuclease	f.21.1	Transmembrane di-heme cytochromes

MitoCOG0059	490	816	ATP synthase subunit 1	c.37.1	P-loop containing nucleoside triphosphate hydrolases
MitoCOG0059	490	816	ATP synthase subunit 1	a.69.1	C-terminal domain of alpha and beta subunits of F1 ATP synthase
MitoCOG0059	490	816	ATP synthase subunit 1	b.49.1	N-terminal domain of alpha and beta subunits of F1 ATP synthase
MitoCOG0060	140	339	ribosomal protein L11	d.47.1	Ribosomal L11/L12e N-terminal domain
MitoCOG0060	140	339	ribosomal protein L11	a.4.7	Ribosomal protein L11, C-terminal domain
MitoCOG0061	111	407	ribosomal protein S7	a.75.1	Ribosomal protein S7
MitoCOG0062	**389**	**843**	**hypothetical protein**	**d.52.3**	**Prokaryotic type KH domain (KH-domain type II)**
MitoCOG0063	122	159	ribosomal protein S8	d.140.1	Ribosomal protein S8
MitoCOG0064	102	276	ribosomal protein S13	a.156.1	S13-like H2TH domain
MitoCOG0066	200	749	NADH dehydrogenase subunit 11	d.58.1	4Fe-4S ferredoxins
MitoCOG0066	200	749	NADH dehydrogenase subunit 11	c.81.1	Formate dehydrogenase/DMSO reductase, domains 1-3
MitoCOG0066	200	749	NADH dehydrogenase subunit 11	d.15.4	2Fe-2S ferredoxin-like
MitoCOG0067	153	1641	ribosomal protein S4	d.66.1	Alpha-L RNA-binding motif
MitoCOG0068	121	773	LAGLIDADG endonuclease	d.95.2	Homing endonucleases
MitoCOG0069	159	170	LAGLIDADG endonuclease	d.95.2	Homing endonucleases
MitoCOG0070	74	118	ribosomal protein L20	a.144.2	Ribosomal protein L20
MitoCOG0071	145	163	NADH dehydrogenase subunit 8	d.58.1	4Fe-4S ferredoxins
MitoCOG0072	91	377	ribosomal protein S10	d.58.15	Ribosomal protein S10
MitoCOG0073	129	278	ribosomal protein L5	d.77.1	RL5-like
MitoCOG0074	105	473	ribosomal protein S1	b.40.4	Nucleic acid-binding proteins
MitoCOG0075	79	144	succinate dehydrogenase subunit 4	f.21.2	Fumarate reductase respiratory complex transmembrane subunits
MitoCOG0077	**139**	**143**	**hypothetical protein**	**d.58.62**	**Ribosomal protein L10-like**
MitoCOG0080	182	559	ribosomal protein S2	c.23.15	Ribosomal protein S2
MitoCOG0082	66	76	ribosomal protein L31	d.325.1	L28p-like
MitoCOG0088	401	949	maturase	e.8.1	DNA/RNA polymerases
MitoCOG0090	**199**	**315**	**hypothetical protein**	**f.17.1**	**F1F0 ATP synthase subunit C**
MitoCOG0090	**199**	**315**	**hypothetical protein**	**f.21.1**	**Transmembrane di-heme cytochromes**
MitoCOG0091	202	845	reverse transcriptase	e.8.1	DNA/RNA polymerases
MitoCOG0092	126	193	ribosomal protein L10	d.58.62	Ribosomal protein L10-like
MitoCOG0093	49	699	LAGLIDADG endonuclease	d.226.1	GIY-YIG endonuclease
MitoCOG0093	49	699	LAGLIDADG endonuclease	d.285.1	DNA-binding domain of intron-encoded endonucleases
MitoCOG0096	104	1204	RNA polymerase	e.8.1	DNA/RNA polymerases
MitoCOG0101	**111**	**490**	**hypothetical protein**	**e.8.1**	**DNA/RNA polymerases**
MitoCOG0116	173	473	LAGLIDADG endonuclease	d.226.1	GIY-YIG endonuclease
MitoCOG0117	206	763	LAGLIDADG endonuclease	d.95.2	Homing endonucleases
MitoCOG0117	206	763	LAGLIDADG endonuclease	f.24.1	Cytochrome c oxidase subunit I-like
MitoCOG0118	174	399	LAGLIDADG endonuclease	d.226.1	GIY-YIG endonuclease
MitoCOG0118	174	399	LAGLIDADG endonuclease	f.24.1	Cytochrome c oxidase subunit I-like
MitoCOG0119	222	267	RNA polymerase subunit alpha	d.181.1	Insert subdomain of RNA polymerase alpha subunit
MitoCOG0119	222	267	RNA polymerase subunit alpha	d.74.3	RBP11-like subunits of RNA polymerase
MitoCOG0120	1362	2640	RNA polymerase subunit beta	e.29.1	beta and beta-prime subunits of DNA dependent RNA-polymerase
MitoCOG0121	1316	2166	RNA polymerase subunit beta-Prime	e.29.1	beta and beta-prime subunits of DNA dependent RNA-polymerase
MitoCOG0122	518	654	RNA polymerase sigma factor	a.177.1	Sigma2 domain of RNA polymerase sigma factors
MitoCOG0122	518	654	RNA polymerase sigma factor	a.4.13	Sigma3 and sigma4 domains of RNA polymerase sigma factors
MitoCOG0124	70	70	ribosomal protein L35	d.301.1	L35p-like
MitoCOG0126	41	41	ribosomal protein L36	g.42.1	Ribosomal protein L36
MitoCOG0127	84	84	ribosomal protein S16	d.27.1	Ribosomal protein S16
MitoCOG0128	176	222	cytochrome oxidase assembly protein 11	b.146.1	Ctag/Cox11
MitoCOG0129	410	419	SecY-type transporter protein	f.41.1	Preprotein translocase SecY subunit
MitoCOG0130	394	400	elongation factor Tu	c.37.1	P-loop containing nucleoside triphosphate hydrolases
MitoCOG0130	394	400	elongation factor Tu	b.44.1	EF-Tu/eEF-1alpha/eIF2-gamma C-terminal domain
MitoCOG0130	394	400	elongation factor Tu	b.43.3	Translation proteins
MitoCOG0132	193	222	cytochrome c biogenesis protein ccmA	c.37.1	P-loop containing nucleoside triphosphate hydrolases
MitoCOG0133	275	288	ATP synthase subunit 3	c.49.2	ATP synthase (F1-ATPase), gamma subunit
MitoCOG0134	48	62	ribosomal protein L32	g.41.8	Zn-binding ribosomal proteins
MitoCOG0136	81	102	ribosomal protein L27	b.84.4	Ribosomal L27 protein-like
MitoCOG0137	86	118	ribosomal protein L19	b.34.5	Translation proteins SH3-like domain
MitoCOG0138	93	118	ribosomal protein L18	c.55.4	Translational machinery components
MitoCOG0139	218	226	ribosomal protein L1	e.24.1	Ribosomal protein L1

Table S6. Source species from which protein SF-domains were analysed in this study

	Genome	Genome Size (kb)	Number of proteins
A. Flux analysis in Eukaryotes and Bacteria			
1	*R. prowazekii*	-	843
2	*Ca. Pelagibacter*	-	1447
3	*C. Puniceispirillum*	-	2543
4	*D. vulgaris*	-	3091
5	*G. daltonii*	-	3798
6	*T. parva*	-	4079
7	*E. coli*	-	4228
8	*M. magneticum*	-	4561
9	*P. brasiliensis*	-	4750
10	*R. palustris*	-	4878
11	***S. pombe***	-	5035
12	*P. pastoris*	-	5040
13	*B. mallei*	-	5184
14	*A. tumefaciens*	-	5355
15	***P. falciparum***	-	5385
16	*Y. lipolytica*	-	6448
17	*U. maydis*	-	6522
18	***S. cerevisiae***	-	6692
19	*M. xanthus*	-	7316
20	*O. sp*	-	7773
21	*C. Solibacter*	-	7826
22	*T. gondii*	-	7993
23	*B. bovis*	-	8029
24	*L. major*	-	8408
25	*A. niger*	-	8592
26	*B. dendrobatidis*	-	8818
27	*M. brevicollis*	-	9196
28	*T. brucei*	-	9826
29	*C. vulgaris*	-	9994
30	*P. tricornutum*	-	10402
31	*A. anophagefferens*	-	11501
32	*T. adhaerens*	-	11520
33	*T. pseudonana*	-	11776
34	*P. pallidum.*	-	12356
35	*S. mansoni*	-	13191
36	*D. discoideum*	-	13263
37	*P. blakesleeanus*	-	14792
38	*A. castellan*	-	14944
39	*V. carteri*	-	15544
40	*N. gruberi*	-	15753
41	*I. scapularis*	-	20486
42	*L. bicolor*	-	20614
43	*B. malayi*	-	21252
44	***D. melanogaster***	-	23849
45	***T. thermophila***	-	24725
46	*A. mellifera*	-	27755
47	*S. purpuratus*	-	29129
48	*C. elegans*	-	31234
49	*S. moellendorffii*	-	34697
50	***A. thaliana***	-	35381
51	*P. patens*	-	35938
52	*P. tetraurelia*	-	39525
53	*B. floridae*	-	50817
54	*O. sativa*	-	67393
55	***H. sapiens***	-	97407
56	*M. pusilla*	-	112079
B. MitoMiner source organisms			
1	***B. taurus***		
2	***M. musculus***		
3	***R. norvegicus***		
4	***N. crassa***		
5	***T. thermophila***		
6	***G. lamlia***		

Note: In addition to list B, species in bold in list A were part of the MitoMiner reference set

Table S7. Description of OXPHOS genes, the complex they are found in and the SF-domain assigned to them as in fig. 4

Human Gene	**Complex**	**SF classification**	**SF desription**
Central			
NDUFS1;SDHB	C1, C2	d.15.4	2Fe-2S ferredoxin-like
NDUFA10;ATP5A1;ATP5	C1, C5	c.37.1	P-loop containing nucleoside triphosphate hydrolases
NDUFA9	C1	c.2.1	NAD(P)-binding Rossmann-fold domains
NDUFAB1	C1	a.28.1	ACP-like
NDUFS1;NDUFS8	C1	d.58.1	4Fe-4S ferredoxins
NDUFS1	C1	c.81.1	Formate dehydrogenase/DMSO reductase, domains 1-3
NDUFS2	C1	e.18.1	HydB/Nqo4-like
NDUFS3	C1	d.307.1	Nqo5-like
NDUFS7	C1	e.19.1	HydA/Nqo6-like
NDUFV1	C1	c.142.1	Nqo1 FMN-binding domain-like
NDUFV1	C1	d.15.13	Nqo1 middle domain-like
NDUFV1	C1	a.29.12	Nqo1C-terminal domain-like
SDHA	C2	c.3.1	FAD/NAD(P)-binding domain
SDHA	C2	a.7.3	Succinate dehydrogenase/fumarate reductase flavoprotein C-terminal domain
SDHA	C2	d.168.1	Succinate dehydrogenase/fumarate reductase flavoprotein, catalytic domain
SDHB	C2	a.1.2	alpha-helical ferredoxin
SDHC;SDHD	C2	f.21.2	Fumarate reductase respiratory complex transmembrane subunits
CYC1	C3	a.3.1	Cytochrome c
CYC1	C3	f.23.11	Cytochrome c1 subunit of cytochrome bc1 complex, transmembrane anchor
MT-CYB	C3	f.32.1	a domain/subunit of cytochrome bc1 complex
MT-CYB	C3	f.21.1	Transmembrane di-heme cytochromes
UQCRC1;UQCRC2	C3	d.185.1	LuxS/MPP-like metallohydrolase
UQCRFS1	C3	b.33.1	ISP domain
UQCRFS1	C3	f.23.12	ISP transmembrane anchor
UQCRFS1	C3	d.184.1	Non-globular alpha+beta subunits of globular proteins
MT-CO1	C4	f.24.1	Cytochrome c oxidase subunit I-like
MT-CO2	C4	b.6.1	Cupredoxins
MT-CO2	C4	f.17.2	Cytochrome c oxidase subunit II-like, transmembrane region
MT-CO3	C4	f.25.1	Cytochrome c oxidase subunit III-like
ATP5A1;ATP5B	C5	a.69.1	C-terminal domain of alpha and beta subunits of F1 ATP synthase
ATP5A1;ATP5B	C5	b.49.1	N-terminal domain of alpha and beta subunits of F1 ATP synthase
ATP5C1	C5	c.49.2	ATP synthase (F1-ATPase), gamma subunit
ATP5D	C5	a.2.10	Epsilon subunit of F1F0-ATP synthase C-terminal domain
ATP5D	C5	b.93.1	Epsilon subunit of F1F0-ATP synthase N-terminal domain
ATP5E	C5	a.137.8	Epsilon subunit of mitochondrial F1F0-ATP synthase
ATP5F1	C5	f.52.1	ATP synthase B chain-like
ATP5G1;ATP5G2;ATP5G	C5	f.17.1	F1F0 ATP synthase subunit C
ATP5O	C5	a.70.1	N-terminal domain of the delta subunit of the F1F0-ATP synthase
MT-ATP6	C5	f.18.1	F1F0 ATP synthase subunit A
Accessory			
NDUFA2;NDUFV2	C1	c.47.1	Thioredoxin-like
UQCRB	C3	f.27.1	14 kDa protein of cytochrome bc1 complex
UQCRH	C3	f.28.1	Non-heme 11 kDa protein of cytochrome bc1 complex
COX5A	C4	a.118.11	Cytochrome c oxidase subunit E
COX5B	C4	g.41.5	Rubredoxin-like
COX6B1;COX6B2	C4	a.51.1	Cytochrome c oxidase subunit h
ATP5H	C5	f.53.1	ATP synthase D chain-like
ATP5J	C5	f.45.1	Mitochondrial ATP synthase coupling factor 6
ATPIF1	C5	h.4.8	F1 ATPase inhibitor, IF1, C-terminal domain
Accessory STMD			
UQCR10	C3	f.23.14	Subunit X (non-heme 7 kDa protein) of cytochrome bc1 complex
UQCR11	C3	f.23.15	Subunit XI (6.4 kDa protein) of cytochrome bc1 complex
UQCRQ	C3	f.23.13	Ubiquinone-binding protein QP-C of cytochrome bc1 complex
COX4I1;COX4I2	C4	f.23.1	Mitochondrial cytochrome c oxidase subunit IV
COX6A1;COX6A2	C4	f.23.2	Mitochondrial cytochrome c oxidase subunit VIa
COX6C	C4	f.23.3	Mitochondrial cytochrome c oxidase subunit VIc
COX7A1;COX7A2	C4	f.23.4	Mitochondrial cytochrome c oxidase subunit VIIa
COX7B;COX7B2	C4	f.23.5	Mitochondrial cytochrome c oxidase subunit VIIb
COX7C	C4	f.23.6	Mitochondrial cytochrome c oxidase subunit VIIc (aka VIIIa)
COX8A;COX8C	C4	f.23.7	Mitochondrial cytochrome c oxidase subunit VIIIb (aka IX)
MT-ATP8	C5	a.118.3	Sec7 domain

Table S8. Description of mitochondria associated SF-domans that do not have any known bacterial homologs. The columns in order from left to right indicate the broad function al category of a SF-domain, the SCOP classification, SCOP description of SF, number of speices in which a protein with the corresponding SF-domain was experimentally identified and the corresponding species. The species names are described in the legend at the end of the table.

Function	**SF classification**	**SF description**	**No. Species**	**Species**
105 mitochondria associated SFs traced back to MCA that do not have any known bacterial homologs as in fig 5				
Information	d.58.12	eEF-1beta-like	9	BOVIN, DROME, NEUCR, ARATH, SCHPO, YEAST, RAT, HUMAN, MOUSE
Information	d.58.46	eEF1-gamma domain	9	BOVIN, DROME, NEUCR, ARATH, YEAST, RAT, HUMAN, PLAF7, MOUSE
Information	a.144.1	PABC (PABP) domain	9	BOVIN, HUMAN, YEAST, ARATH, SCHPO, RAT, NEUCR, PLAF7, MOUSE
Information	b.34.9	Tudor/PWWP/MBT	8	BOVIN, DROME, NEUCR, ARATH, SCHPO, RAT, HUMAN, MOUSE
Information	c.52.2	tRNA-intron endonuclease catalytic domain-like	7	BOVIN, HUMAN, YEAST, SCHPO, RAT, NEUCR, MOUSE
Information	a.11.2	Second domain of FERM	5	BOVIN, RAT, MOUSE, HUMAN, DROME
Information	a.160.1	PAP/OAS1 substrate-binding domain	5	BOVIN, RAT, MOUSE, HUMAN, DROME
Information	d.23.1	Tubby C-terminal domain-like	2	SCHPO, YEAST
Information	g.44.1	RING/U-box	1	SCHPO
Information	g.50.1	FYVE/PHD zinc finger	1	YEAST
Metabolism	f.23.12	ISP transmembrane anchor	10	BOVIN, DROME, NEUCR, ARATH, SCHPO, YEAST, RAT, HUMAN, PLAF7, MOUSE
Metabolism	b.69.6	Clathrin heavy-chain terminal domain	9	BOVIN, DROME, NEUCR, ARATH, SCHPO, YEAST, RAT, HUMAN, MOUSE
Metabolism	a.29.5	alpha-ketoacid dehydrogenase kinase, N-terminal domain	9	BOVIN, DROME, NEUCR, ARATH, SCHPO, YEAST, RAT, HUMAN, MOUSE
Metabolism	a.51.1	Cytochrome c oxidase subunit h	9	BOVIN, DROME, NEUCR, ARATH, SCHPO, YEAST, RAT, HUMAN, MOUSE
Metabolism	d.120.1	Cytochrome b5-like heme/steroid binding domain	9	BOVIN, DROME, NEUCR, ARATH, SCHPO, YEAST, RAT, HUMAN, MOUSE
Metabolism	f.27.1	14 kDa protein of cytochrome bc1 complex (Ubiquinol-cytochrome c reductase)	8	BOVIN, DROME, NEUCR, YEAST, SCHPO, RAT, HUMAN, MOUSE
Metabolism	a.24.15	FAD-dependent thiol oxidase	8	BOVIN, DROME, NEUCR, YEAST, SCHPO, RAT, HUMAN, MOUSE
Metabolism	c.131.1	Peptidyl-tRNA hydrolase II	8	BOVIN, DROME, YEAST, ARATH, SCHPO, RAT, HUMAN, MOUSE
Metabolism	a.227.1	ERO1-like	8	BOVIN, HUMAN, YEAST, ARATH, SCHPO, RAT, NEUCR, MOUSE
Metabolism	a.11.1	Acyl-CoA binding protein	7	BOVIN, DROME, YEAST, SCHPO, RAT, HUMAN, MOUSE
Metabolism	f.28.1	Non-heme 11 kDa protein of cytochrome bc1 complex (Ubiquinol-cytochrome c reductase)	6	BOVIN, YEAST, SCHPO, RAT, HUMAN, MOUSE
Metabolism	d.50.4	Peptidyl-tRNA hydrolase domain-like	6	BOVIN, DROME, YEAST, ARATH, HUMAN, MOUSE
Metabolism	b.74.1	Carbonic anhydrase	5	BOVIN, RAT, MOUSE, HUMAN, DROME
Metabolism	a.137.8	Epsilon subunit of mitochondrial F1F0-ATP synthase	4	SCHPO, MOUSE, HUMAN, YEAST
Metabolism	a.118.6	Protein prenylyltransferase	2	RAT, MOUSE
Regulation	a.235.1	ATP-dependent DNA ligase DNA-binding domain	10	BOVIN, DROME, NEUCR, ARATH, SCHPO, TETTS, YEAST, RAT, HUMAN, MOUSE
Regulation	b.55.1	PH domain-like	9	BOVIN, DROME, NEUCR, ARATH, SCHPO, YEAST, RAT, HUMAN, MOUSE
Regulation	a.59.1	PAH2 domain	8	BOVIN, DROME, YEAST, ARATH, SCHPO, RAT, HUMAN, MOUSE
Regulation	a.74.1	Cyclin-like	8	BOVIN, DROME, YEAST, ARATH, SCHPO, RAT, HUMAN, MOUSE
Regulation	g.37.1	C2H2 and C2HC zinc fingers	8	BOVIN, DROME, NEUCR, ARATH, YEAST, RAT, HUMAN, MOUSE
Regulation	b.7.1	C2 domain (Calcium/lipid-binding domain, CaLB)	8	BOVIN, HUMAN, YEAST, ARATH, SCHPO, RAT, NEUCR, MOUSE
Regulation	b.34.2	SH3-domain	8	BOVIN, DROME, NEUCR, YEAST, SCHPO, RAT, HUMAN, MOUSE
Regulation	g.83.1	Tim10-like	7	BOVIN, DROME, YEAST, SCHPO, RAT, HUMAN, MOUSE
Regulation	a.21.1	HMG-box	7	BOVIN, DROME, YEAST, SCHPO, RAT, HUMAN, MOUSE
Regulation	a.22.1	Histone-fold	6	BOVIN, ARATH, SCHPO, RAT, HUMAN, MOUSE
Regulation	a.140.2	SAP domain	5	BOVIN, MOUSE, HUMAN, DROME, YEAST
Regulation	a.69.2	Ypt/Rab-GAP domain of gyp1p	4	BOVIN, MOUSE, HUMAN, YEAST
Regulation	a.117.1	Ras GEF	4	MOUSE, HUMAN, DROME, YEAST
Regulation	d.332.1	RGC domain-like	3	RAT, MOUSE, HUMAN
Regulation	g.66.1	CCCH zinc finger	3	BOVIN, MOUSE, HUMAN
Regulation	a.66.1	Transducin (alpha subunit), insertion domain	3	RAT, MOUSE, HUMAN
Regulation	a.116.1	GTPase activation domain, GAP	3	RAT, MOUSE, HUMAN
Regulation	d.93.1	SH2 domain	3	BOVIN, MOUSE, HUMAN
Regulation	g.85.1	HIT/MYND zinc finger-like	2	TETTS, NEUCR
Regulation	b.34.14	PAZ domain	1	DROME

Regulation	a.217.1	Surp module (SWAP domain)	1	YEAST
Regulation	a.183.1	Nop domain	1	YEAST
Regulation	a.60.1	SAM/Pointed domain	1	YEAST
Regulation	a.91.1	Regulator of G-protein signaling, RGS	1	HUMAN
Regulation	a.141.1	Frizzled cysteine-rich domain	1	HUMAN
Regulation	d.198.2	Arp2/3 complex subunits	1	YEAST
Intra-cellularProcesses	f.42.1	Mitochondrial carrier	11	BOVIN, DROME, NEUCR, ARATH, SCHPO, TETTS, YEAST, RAT, HUMAN, PLAF7, MOUSE
Intra-cellularProcesses	d.25.1	Mitochondrial glycoprotein MAM33-like	9	BOVIN, DROME, NEUCR, ARATH, SCHPO, YEAST, RAT, HUMAN, MOUSE
Intra-cellularProcesses	b.104.1	P-domain of calnexin/calreticulin	9	BOVIN, DROME, NEUCR, ARATH, SCHPO, YEAST, RAT, HUMAN, MOUSE
Intra-cellularProcesses	b.2.7	Second domain of Mu2 adaptin subunit (ap50) of ap2 adaptor	9	BOVIN, GIAIC, DROME, NEUCR, ARATH, SCHPO, RAT, HUMAN, MOUSE
Intra-cellularProcesses	h.1.15	SNARE fusion complex	9	BOVIN, DROME, NEUCR, ARATH, SCHPO, YEAST, RAT, HUMAN, MOUSE
Intra-cellularProcesses	a.207.1	Formin homology 2 domain (FH2 domain)	8	BOVIN, DROME, NEUCR, ARATH, YEAST, RAT, HUMAN, MOUSE
Intra-cellularProcesses	b.64.1	Mannose 6-phosphate receptor domain	8	BOVIN, DROME, NEUCR, ARATH, SCHPO, RAT, HUMAN, MOUSE
Intra-cellularProcesses	g.12.1	LDL receptor-like module	8	BOVIN, DROME, NEUCR, ARATH, SCHPO, RAT, HUMAN, MOUSE
Intra-cellularProcesses	h.1.26	Myosin rod fragments	7	BOVIN, HUMAN, ARATH, SCHPO, RAT, NEUCR, MOUSE
Intra-cellularProcesses	a.64.1	Saposin	6	BOVIN, DROME, ARATH, RAT, HUMAN, MOUSE
Intra-cellularProcesses	a.47.2	t-snare proteins	6	BOVIN, ARATH, SCHPO, RAT, HUMAN, MOUSE
Intra-cellularProcesses	d.109.1	Actin depolymerizing proteins	5	BOVIN, RAT, ARATH, MOUSE, HUMAN
Intra-cellularProcesses	e.25.1	Sec1/munc18-like (SM) proteins	5	BOVIN, RAT, MOUSE, HUMAN, DROME
Intra-cellularProcesses	d.110.1	Profilin (actin-binding protein)	4	BOVIN, RAT, MOUSE, HUMAN
Intra-cellularProcesses	a.2.5	Prefoldin	4	BOVIN, RAT, MOUSE, HUMAN
Intra-cellularProcesses	e.1.1	Serpins	4	BOVIN, RAT, MOUSE, HUMAN
Intra-cellularProcesses	b.1.10	Clathrin adaptor appendage domain	4	BOVIN, RAT, MOUSE, HUMAN
Intra-cellularProcesses	b.60.1	Lipocalins	4	BOVIN, RAT, MOUSE, HUMAN
Intra-cellularProcesses	d.105.1	Subdomain of clathrin and coatomer appendage domain	4	BOVIN, RAT, MOUSE, HUMAN
Intra-cellularProcesses	c.8.4	PA domain	3	RAT, MOUSE, HUMAN
Intra-cellularProcesses	a.7.14	MIT domain	2	SCHPO, YEAST
Intra-cellularProcesses	b.69.5	RCC1/BLIP-II	1	YEAST
Intra-cellularProcesses	a.118.17	Cullin repeat-like	1	YEAST
Intra-cellularProcesses	e.40.1	Cullin homology domain	1	YEAST
Intra-cellularProcesses	a.14.1	VHP, Villin headpiece domain	1	ARATH
Intra-cellularProcesses	c.10.3	Outer arm dynein light chain 1	1	DROME
Intra-cellularProcesses	a.222.1	VPS9 domain	1	YEAST
Extra-cellularProcesses	c.10.1	RNI-like	8	BOVIN, HUMAN, YEAST, ARATH, SCHPO, RAT, NEUCR, MOUSE
Extra-cellularProcesses	a.7.1	Spectrin repeat	6	BOVIN, DROME, YEAST, RAT, HUMAN, MOUSE
Extra-cellularProcesses	g.3.11	EGF/Laminin	5	BOVIN, RAT, MOUSE, HUMAN, DROME
Extra-cellularProcesses	g.14.1	Kringle-like	2	MOUSE, HUMAN
Extra-cellularProcesses	a.118.24	Pseudo ankyrin repeat-like	1	NEUCR
General	d.15.1	Ubiquitin-like	10	BOVIN, DROME, NEUCR, ARATH, SCHPO, YEAST, RAT, HUMAN, PLAF7, MOUSE
General	d.110.4	SNARE-like	10	BOVIN, GIAIC, DROME, NEUCR, ARATH, SCHPO, YEAST, RAT, HUMAN, MOUSE
General	a.24.7	FKBP12-rapamycin-binding domain of FKBP-rapamycin-associated protein (FRAP)	9	BOVIN, GIAIC, DROME, YEAST, ARATH, SCHPO, RAT, HUMAN, MOUSE
General	d.31.1	Cdc48 domain 2-like	9	BOVIN, DROME, YEAST, ARATH, SCHPO, RAT, HUMAN, PLAF7, MOUSE
General	a.40.1	Calponin-homology domain, CH-domain	9	BOVIN, DROME, NEUCR, ARATH, SCHPO, YEAST, RAT, HUMAN, MOUSE
General	a.189.1	XPC-binding domain	8	BOVIN, HUMAN, YEAST, ARATH, SCHPO, RAT, NEUCR, MOUSE
General	a.118.7	14-3-3 protein	8	BOVIN, GIAIC, DROME, ARATH, SCHPO, RAT, HUMAN, MOUSE
General	g.41.2	Microbial and mitochondrial ADK, insert "zinc finger" domain	6	BOVIN, DROME, NEUCR, RAT, HUMAN, MOUSE
General	a.158.1	F-box domain	6	BOVIN, YEAST, ARATH, SCHPO, RAT, NEUCR
General	d.42.1	POZ domain	5	BOVIN, RAT, MOUSE, HUMAN, YEAST
General	c.10.2	L domain-like	4	BOVIN, RAT, MOUSE, HUMAN

General	a.238.1	BAR/IMD domain-like	4	SCHPO, NEUCR, HUMAN, YEAST
General	d.289.1	WWE domain	1	YEAST
Other	a.17.1	Cysteine alpha-hairpin motif	7	BOVIN, DROME, YEAST, SCHPO, RAT, HUMAN, MOUSE
Other	b.165.1	MOSC N-terminal domain-like	5	BOVIN, RAT, ARATH, MOUSE, HUMAN
Other	a.71.1	ERP29 C domain-like	5	BOVIN, RAT, MOUSE, HUMAN, DROME
Other	b.2.8	beta-sandwich domain of Sec23/24	5	BOVIN, RAT, MOUSE, NEUCR, HUMAN
Other	g.40.1	Retrovirus zinc finger-like domains	4	BOVIN, ARATH, MOUSE, DROME
Not Annotated	a.289.1		10	BOVIN, DROME, NEUCR, ARATH, SCHPO, YEAST, RAT, HUMAN, PLAF7, MOUSE
Not Annotated	d.355.1		9	BOVIN, DROME, YEAST, ARATH, SCHPO, RAT, HUMAN, PLAF7, MOUSE
Not Annotated	f.53.1		9	BOVIN, DROME, NEUCR, ARATH, SCHPO, YEAST, RAT, HUMAN, MOUSE
Not Annotated	d.81.4		1	ARATH
55 mitochondria associated SFs which are species specific and do not have any known bacterial homologs as in fig 5				
Information	d.75.1	tRNA-intron endonuclease N-terminal domain-like	1	SCHPO
Information	d.95.2	Homing endonucleases	1	YEAST
Metabolism	a.118.11	Cytochrome c oxidase subunit E	8	BOVIN, DROME, NEUCR, YEAST, SCHPO, RAT, HUMAN, MOUSE
Metabolism	f.23.1	Mitochondrial cytochrome c oxidase subunit IV	8	BOVIN, DROME, NEUCR, YEAST, SCHPO, RAT, HUMAN, MOUSE
Metabolism	b.34.8	Fumarylacetoacetate hydrolase, FAH, N-terminal domain	7	BOVIN, DROME, NEUCR, ARATH, RAT, HUMAN, MOUSE
Metabolism	f.23.13	Ubiquinone-binding protein QP-C of cytochrome bc1 complex (Ubiquinol-cytochrome c reductase)	7	BOVIN, DROME, YEAST, SCHPO, RAT, HUMAN, MOUSE
Metabolism	f.23.2	Mitochondrial cytochrome c oxidase subunit VIa	7	BOVIN, DROME, YEAST, SCHPO, RAT, HUMAN, MOUSE
Metabolism	f.23.14	Subunit X (non-heme 7 kDa protein) of cytochrome bc1 complex (Ubiquinol-cytochrome c reductase)	6	BOVIN, YEAST, SCHPO, RAT, HUMAN, MOUSE
Metabolism	f.23.6	Mitochondrial cytochrome c oxidase subunit VIIc (aka VIIIa)	6	BOVIN, YEAST, SCHPO, RAT, HUMAN, MOUSE
Metabolism	a.2.10	Epsilon subunit of F1F0-ATP synthase C-terminal domain	5	BOVIN, RAT, MOUSE, HUMAN, DROME
Metabolism	f.23.15	Subunit XI (6.4 kDa protein) of cytochrome bc1 complex (Ubiquinol-cytochrome c reductase)	5	BOVIN, RAT, MOUSE, HUMAN, YEAST
Metabolism	f.23.3	Mitochondrial cytochrome c oxidase subunit VIc	5	BOVIN, RAT, SCHPO, MOUSE, HUMAN
Metabolism	h.4.8	F1 ATPase inhibitor, IF1, C-terminal domain	5	BOVIN, RAT, MOUSE, HUMAN, YEAST
Metabolism	a.83.1	Guanido kinase N-terminal domain	4	BOVIN, RAT, MOUSE, HUMAN
Metabolism	d.184.1	Non-globular alpha+beta subunits of globular proteins	4	BOVIN, RAT, MOUSE, HUMAN
Metabolism	f.23.4	Mitochondrial cytochrome c oxidase subunit VIIa	4	BOVIN, RAT, MOUSE, HUMAN
Metabolism	f.23.5	Mitochondrial cytochrome c oxidase subunit VIIb	4	BOVIN, RAT, MOUSE, HUMAN
Metabolism	d.36.1	Chalcone isomerase	2	SCHPO, YEAST
Metabolism	a.191.1	Methenyltetrahydrofolate cyclohydrolase-like	1	DROME
Regulation	g.16.2	Plexin repeat	5	BOVIN, RAT, MOUSE, HUMAN, DROME
Regulation	a.137.3	Transducin (heterotrimeric G protein), gamma chain	4	BOVIN, RAT, MOUSE, HUMAN
Regulation	a.7.4	Smac/diablo	4	BOVIN, RAT, MOUSE, HUMAN
Regulation	d.170.1	SRCR-like	4	BOVIN, RAT, MOUSE, HUMAN
Regulation	d.285.1	DNA-binding domain of intron-encoded endonucleases	1	NEUCR
Intra-cellularProcesses	a.23.4	Mitochondrial import receptor subunit Tom20	8	BOVIN, DROME, NEUCR, YEAST, SCHPO, RAT, HUMAN, MOUSE
Intra-cellularProcesses	b.61.5	Dipeptidyl peptidase I (cathepsin C), exclusion domain	6	BOVIN, GIAIC, RAT, HUMAN, PLAF7, MOUSE
Intra-cellularProcesses	a.137.5	Moesin tail domain	5	BOVIN, RAT, MOUSE, HUMAN, DROME
Intra-cellularProcesses	a.65.1	Annexin	5	BOVIN, RAT, MOUSE, HUMAN, DROME
Intra-cellularProcesses	f.1.4	Bcl-2 inhibitors of programmed cell death	4	BOVIN, RAT, MOUSE, HUMAN
Intra-cellularProcesses	h.1.25	Troponin coil-coiled subunits	4	BOVIN, RAT, MOUSE, HUMAN
Intra-cellularProcesses	a.216.1	I/LWEQ domain	3	RAT, MOUSE, HUMAN
Intra-cellularProcesses	a.48.2	Transferrin receptor-like dimerisation domain	3	RAT, MOUSE, HUMAN
Intra-cellularProcesses	b.34.3	Myosin S1 fragment, N-terminal domain	3	SCHPO, HUMAN, YEAST
Intra-cellularProcesses	f.36.1	Neurotransmitter-gated ion-channel transmembrane pore	1	MOUSE

Extra-cellularProcesses	b.1.15	Integrin domains	5	BOVIN, RAT, MOUSE, HUMAN, DROME
Extra-cellularProcesses	b.3.5	Cna protein B-type domain	5	BOVIN, RAT, ARATH, MOUSE, HUMAN
Extra-cellularProcesses	d.200.1	Integrin beta tail domain	5	BOVIN, RAT, MOUSE, HUMAN, DROME
Extra-cellularProcesses	b.1.1	Immunoglobulin	4	BOVIN, RAT, MOUSE, HUMAN
Extra-cellularProcesses	b.1.16	Lamin A/C globular tail domain	4	BOVIN, RAT, MOUSE, HUMAN
Extra-cellularProcesses	d.171.1	Fibrinogen C-terminal domain-like	4	BOVIN, RAT, MOUSE, HUMAN
Extra-cellularProcesses	g.18.1	Complement control module/SCR domain	4	BOVIN, RAT, MOUSE, HUMAN
Extra-cellularProcesses	a.215.1	A middle domain of Talin 1	3	RAT, MOUSE, HUMAN
Extra-cellularProcesses	a.24.9	alpha-catenin/vinculin-like	3	RAT, MOUSE, HUMAN
Extra-cellularProcesses	g.27.1	FnI-like domain	2	MOUSE, HUMAN
Extra-cellularProcesses	b.30.6	V-region of surface antigen I/II (SA I/II, PAC)	1	YEAST
Extra-cellularProcesses	b.40.2	Bacterial enterotoxins	1	MOUSE
General	a.2.16	Calcyclin-binding protein-like	7	BOVIN, DROME, ARATH, RAT, HUMAN, PLAF7, MOUSE
General	a.126.1	Serum albumin-like	4	BOVIN, RAT, MOUSE, HUMAN
General	f.45.1	Mitochondrial ATP synthase coupling factor 6	4	BOVIN, RAT, MOUSE, HUMAN
Other	h.1.20	Intermediate filament protein, coiled coil region	4	BOVIN, RAT, MOUSE, HUMAN
Other	h.1.8	Fibrinogen coiled-coil and central regions	4	BOVIN, RAT, MOUSE, HUMAN
Other	e.34.1	NSP3 homodimer	1	SCHPO
Other	h.4.10	C-terminal domain of PLC-beta	1	HUMAN
Not Annotated	f.52.1		8	BOVIN, DROME, NEUCR, YEAST, SCHPO, RAT, HUMAN, MOUSE
Not Annotated	d.382.1		3	SCHPO, MOUSE, HUMAN

Species legend:
ARATH:*A. thaliana*; BOVIN:*B. taurus*; DROME:*D. melanogaste*r; GIAIC:*G. lamlia*; HUMAN:*H. sapien*s; MOUSE:*M. musculus*; NEUCR:*N. crassa*; PLAF7:*P. falciparum*; RAT:*R. norvegicus*; SCHPO:*S. pombe*; TETTS:*T. thermophila*; YEAST:S. cerevisiae

References

1. Gray KA, *et al.* (2013) Genenames.org: the HGNC resources in 2013. *Nucleic Acids Res.* 41, doi:10.1093/nar/gks1066.
2. Lomize MA, *et al.* (2006) OPM: Orientations of proteins in membranes database. *Bioinformatics* 22, 623–625.
3. Tanabe M, *et al*. (2012) Using the KEGG database resource. Curr. Protoc. Bioinforma., doi:10.1002/0471250953.bi0112s38.
4. The UniProt Consortium. (2015) UniProt: a hub for protein information. *Nucleic Acids Res*. 43, D204–212.
5. Race HL, Herrman RG, Martin W (1999) Why have organelles retained genomes?. *Trends Genetics* 15, 364-70.
6. Hessa T, *et al.* (2007) Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* 450, 1026–1030.
7. He D*, et al.* (2014) An Alternative Root for the Eukaryote Tree of Life. *Curr Biol* 24(4):465-470.
8. 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(1):237.
9. Smith AC, Blackshaw JA, & Robinson AJ (2012) MitoMiner: a data warehouse for mitochondrial proteomics data. *Nucleic Acids Res* 40(D1):D1160-D1167.
10. Gough J, Karplus K, Hughey R, & Chothia C (2001) Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *J Mol Biol* 313(4):903-919.
11. Eddy SR (2011) Accelerated profile HMM searches. *PLoS Comput Biol* 7(10).
12. Vogel C, Berzuini C, Bashton M, Gough J, & Teichmann SA (2004) Supra-domains: Evolutionary Units Larger than Single Protein Domains. *J Mol Biol* 336(3):809-823.
13. Swofford DL (2003) *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4* (Sinauer Associates, Sunderland, Massachusetts).
14. Chothia C, Gough J, Vogel C, & Teichmann SA (2003) Evolution of the protein repertoire. *Science* 300(5626):1701-1703.