Supporting Information - Text S1

Atkinson et al.

SI Methods

IF2 and IF3 Sequence retrieval and phylogenetic analysis

Sequences homolologous to IF2 were retrieved by BlastP searching the eukaryotic NCBI genome page with human IF2 as the query and an E value cut off of $1e^{-16}$ against eukaryotic taxa. The search was repeated on the prokaryotic genome page, against a select, but broad distribution of bacteria and archaea. As plants are not listed on the genome page, an additional BlastP search was carried out against the RefSeq database, with taxa limited to *Arabidopsis lyrata*, *Chlamydomonas reinhardtii*, *Oryza sativa* and *Vitis vinifera*. This formed the "full IF2" dataset, which also included sequences of eukaryotic cytoplasmic ortholog eIF5B and archaeal ortholog aIF5B. From this, a smaller dataset "cutdown IF2" just comprising representatives from just the IF2 part of the tree was constructed. For each dataset, sequences were aligned with MAFFT v6.234b with strategy L-INS-I (1), and sites for phylogenetic analysis were selected using GBlocks (at most 15 contiguous nonconserved positions allowed, minimum block length of 3 amino acids, and 50% gap characters accepted in a column) and checked by eye. The final IF2 data set dimensions were 216 characters from 274 taxa for "full IF2" and 405 characters from 92 taxa for "cutdown IF2."

IF3 homologs are too divergent to be retrieved by BlastP alone, therefore NCBI PSI-Blast was used to retrieve these sequences, starting with *S. pombe* mIF3 as the query. 10 iterations were carried out against all eukaryotes in the RefSeq database. In order to ensure all possible sequences were retrieved from fungi, the PSI-blast search was repeated against RefSeq, limited to fungi alone. After 6 iterations, no new sequences were found. Sequences were aligned with MAFFT v6.234b with strategy L-INS-I (1). Sequences that appeared unalignable or ambiguously aligned across their length were removed and the remaining sequences were realigned. This formed the dataset "full IF3", which was cut down to a smaller dataset "cutdown IF3" comprising all fungi, plus representatives from other eukaryotic groups (excluding long branched protist and plastid sequences) and bacteria. As this is a complicated alignment resulting from the presence of divergent sequences with many insertions and deletions, the alignment was inspected and sites were selected by eye for phylogenetic analysis. The final IF3 database dimensions were 139 characters from 169 taxa for "full IF3" and 156 characters from 77 taxa for "cutdown IF3".

Maximum likelihood (ML) phylogenetic analysis was carried out using RAxML v 7.0.4 (2) for both the "full IF2" and "full IF3" datasets. The program was run on the CIPRES portal v2.2 (http://www.phylo.org/sub_sections/portal/) with the PROTCATWAG model and 100 bootstrap replicates. The cutdown IF2 and cutdown IF3 datasets were subject to Bayesian inference (BI) phylogenetic analyses using MrBayes v3.1.2 also on the CIPRES portal v2.2 (3). MrBayes was run for 2 million generations, sampling every 100 generations, with 2 parallel runs of 8 chains each. A gamma distribution of amino acid substitution rates was used, and the substitution model set to "mixed", with the program converging on WAG with a posterior probability of 1.0 for IF3 and rtREV with a posterior probability of 1.0 for IF2. In

each case, a consensus tree was generated after discarding the first 200,000 generations from each run as a burn in. The standard deviation of split frequencies (SDSF) values are shown in Figure legends. RaxML trees with 100 bootstrap replicates were also generated for "cutdown IF3" and "cutdown IF2", using substitution models corresponding to those selected by MrBayes: PROTCATWAG for IF3 and PROTCATRTREV for IF2.

Sequence searching for yeast mitochondrial translational activator orthologs

PSI-Blast was used to search for orthologs of seventeen *S. cerevisiae* translational activators. In each case, the PSI-Blast iterations against the RefSeq database were run until no new sequences were found, or until >5000 sequences were found. Alignments were produced with MAFFT v6.234b (1), and trees generated with the fast and approximate ML program FastTree version 2.0.1 (4). Where hits formed a clade with the *S. cerevisiae* query and were apparent orthologs, these hits were entered into Table S2.

Aim23p in vivo complementation experiments

Plasmid pYCG_YJL131c carrying the *AIM23* ORF with 541bp upstream and 839 bp downstream genomic flanking regions in the background of the pRS416 vector (URA3 marker gene) strain was ordered from EUROSCARF (http://web.uni-frankfurt.de/fb15/mikro/euroscarf/). The *AIM23* ORF in pYCG_YJL131c was replaced with a XhoI restriction site using QuikChange site-directed mutagenesis kit (Stratagene) resulting in plasmid pYCG_YJL131c_ΔORF. The SPBC18E5.13 ORF encoding *S. pombe* mIF3 (*S.p.MIF3*) was synthesized *in vitro* and cloned into pJExpress414-T7-amp vector using DNA20 facility (https://www.dna20.com/). The *S.p.MIF3*/SPBC18E5.13 ORF was inserted into pYCG_YJL131c_ΔORF under the genomic flanks of *AIM23* via *in vivo* recombination in yeast (for detailed information see (5)). In short, the *S.p.MIF3*/SPBC18E5.13 ORF was PCR amplified using primers containing 30bp long 3'-proximal regions homologous to corresponding ends of pYCG_YJL131c_ΔORF linearized by XhoI. The wildtype yeast strain W303 was simultaneously transformed with these two DNA fragments using the high efficiency lithium acetate/PEG method (6) and yeast cells containing the product of homologous recombination were selected on selective medium lacking uracil. The resulting plasmid was isolated using QiaPrep plasmid purification kit (QIAgen).

To construct the haploid *AIM23* knock-out strain (*AIM23Δ*) we induced sporulation of heterozygous diploid strain Y21294 purchased from EUROSCARF [BY4743; Mat a/α; his3D1/his3D1; leu2D0/leu2D0; lys2D0/LYS2; MET15/met15D0; ura3D0/ura3D0; YJL131c::kanMX4/YJL131c] carrying one allele of *AIM23* disrupted with a geneticin (G418) resistance cassette, followed by dissection of resulting tetrads. Spores were selected by their ability to grow on YPD containing G418. In order to create haploid strains with an *AIM23* genomic deletion, carrying either *S. cerevisiae AIM23* (the 'wild-type' strain) or *S.p.MIF3* (the *AIM23Δ+S.p.MIF3* strain) on a plasmid, the Y21294 strain was transformed with the corresponding URA3-containing plasmids, sporulated and the resulting tetrads were dissected. Spores were selected by their ability to grow on minimal selective media without uracil and YPD containing G418. In order to ensure the haploid state of the resulting stains, the mating type of all putative haploids was determined by ability to

cross with either MATa or MATα reference strain during 5 hours in YPD media at 30°C (7).

Measurement of mitochondrial oxygen uptake

Mitochondrial respiration was assayed *in vivo* by direct measurement of oxygen uptake using a Clark-type oxygen electrode (Oxygraph, Hansatech Instruments, Norfolk, UK) at 30°C (8).

Measurement of mitochondrial membrane potential by fluorescence microscopy and flow cytometry

The yeast mitochondrial membrane potential was investigated *in vivo* by two methods: 1) Rhodamine 123 staining followed by fluorescence microscopy (9) and 2) DiOC6 staining followed by flow cytometry (10). Briefly, $10E^5$-$10E^6$ cells were treated with 50 mM Rhodamine-123 (Sigma) for 10 minutes at 30°C, followed by two washing steps with growth medium (CSM-URA or CSM, MP Biomedicals) without Rhodamine-123. The epifluorescence imaging was performed using Axiovert 200M microscope (Carl Zeiss) equipped with a 50 W mercury lamp and an excitation filter BP 450-490. Digital images were acquired with a Photometrics CoolSNAP HQ2 Video Camera (Roper Scientific) and using AxioVision4.6.3, Carl Zeiss software for digital image processing.

For flow cytometry, yeast cultures were grown using galactose as a carbon source until OD_{600} 0.6-0.7, 1 ml was spun down at 3000 g for 5 minutes, washed with filtered PBS and dissolved in 100 μl PBS. 1μl DiOC6 (Life Technologies) was added to a final concentration of 175 nM and cells were incubated in darkness at room temperature for 20-30 minutes, subjected to sonication for 10 seconds, diluted 40X in PBS and analyzed on a BD SLR II flow cytometer using LSR II Blue Laser Saphire (488 nm) for excitation and FITC-A (530 ± 15 nm) channel for detection. Data analysis was preformed in FlowJo software package (Tree Star, Inc).

Cloning, overexpression and purification of S. cerevisiae Aim23

The Aim23 ORF lacking the mitochondrial import signal predicted by MitoProt (II) (11) (amino acids 1-32) was amplified from *S. cerevisiae* genomic DNA and cloned into the pET28a 6His expression vector between NcoI and XhoI sites using forward primer 5' ACCATGGACAATGCATCATCTACCACAG 3' and reverse primer 5' GCTCGAGGCCCATTTCATTCATTTTTTTC 3', resulting in expression construct *AIM23Δ1-32* pET28a . The protein was expressed in *E. coli* BL21 (DE3) in LB media supplemented with kanamycin 25μg/μl at 37°C. At culture density OD_{600} 0.5 the temperature was decreased to 25°C, and after 15-20 min expression was induced by addition of 0.1 mM IPTG. After 3 hours cells were harvested by centrifugation (4500 RPM, 15 min at 4°C) and stored at -85°C.

For protein preparation, the frozen pellet was thawed on ice and resuspended in buffer A (25 mM Tris pH7.5, 300 mM NaCl, 20 mM imidazole, 5 mM βME, 0.2% Tween20, 10% glycerol) supplemented with 1 mM PMSF. Cells were disrupted at 4°C by six successive passes through EmulsiFlex-C3 microfluidizer. Lysate was cleared by two rounds of centrifugation at 14 000 RPM for 30 minutes at 4°C. Next, clarified lysate

was loaded on 1 ml HisTrapFF column (GE Healthcare), washed with 50 column volumes of 50 ml buffer B (A + 25 mM imidazole) and eluted with 30 ml gradient of imidazole concentration up to 500 mM. Fractions were analyzed on PAGE, pooled, diluted 3 times and loaded 1ml/min on 5 ml ion-exchange column SFF (GE Healthcare) pre-equilibrated in $A_{I\text{-}ex}$ buffer (20 mM Hepes pH 7.5, 0.2% Tween 20, 10% glycerol). The column was washed with buffer $B_{I\text{-}ex}$ ($A_{I\text{-}ex}$ + 100 mM NaCl) until the baseline and protein was eluted with NaCl gradient of buffers $B_{I\text{-}ex}$ and $C_{I\text{-}ex}$ ($A_{I\text{-}ex}$ + 1M NaCl). Peak fractions were analyzed on PAGE, pooled, concentrated using MWCO 30 000 Amicon Ultra (Millipore) concentrators. Protein preparations were stored in LLP buffer (10 mM Tris pH 7.5, 2 mM MgOAc, 60 mM KCl, 60 mM NH_4Cl, 6 mM βME) at -20°C.

Detection of Aim23p binding to the E. coli 30S ribosomal subunit

Complex formation was assayed in 100 μl of 1 x LLP buffer (10 mM Tris pH 7.5, 2 mM MgOAc, 60 mM KCl, 60 mM NH_4Cl, 6 mM βME) supplemented with 20% glycerol with final concentration of 1μM for 30S and 2.5 μM for Aim23p and IF3. Reaction mixtures were mixed on ice, preincubated at 37°C for 15 min and loaded on Illustra MicroSpin TM S-300 HR gel-filtration columns (GE Healthcare). The prior to use storage buffer was removed from gel filtration spin columns by centrifugation 1 min 700 g at +4°C, and after they were equilibrated with 1X LLP buffer supplemented with 1 mg/ml BSA. The columns were subjected to successive washes by 100 μl 1xLLP for 1 min at 700 g. 100 μl fractions were collected and analyzed by Western blotting using HisProbeTM-HRP reagent (Thermo Scientific) for 6-His tag detection. The protein fractions were transferred to Hybond-ECL nitrocellulose membrane (Amersham Biosciences) by vacuum filtration manifold system (Life Technologies). Thermo Scientific SuperSignal West Pico Chemiluminescent Substrate Working Solution was used to detect the signal using Agfa X-ray film.

E. coli 70S ribosome splitting assay

The ribosomal ribosomal subunit splitting assay was carried out according to (12) with minor modifications. 1 OD_{260} unit of *E. coli* 70S ribosomes and variable amounts of *E. coli* IF3 or yeast mitochondrial protein Aim23-6His were applied in molar excess ratios 1:10 and 1:20 were combined in 100 μL reaction volume in LLP buffer (10 mM Tris pH 7.5, 2 mM MgOAc, 60 mM KCl, 60 mM NH_4Cl, 6 mM βME). The reaction was mixed on ice and incubated at 37°C for 15 min. The Mg^{2+} concentration was adjusted to 7 mM and the sample was placed on ice before centrifugation analysis by 10-25% linear sucrose gradient in LLP buffer with 7 mM MgOAc. Untreated *E. coli* 70S were used as a control.

SI Figure Legends

Figure S1: Phylogeny of the "full IF2" dataset. The tree is generated from a RaxML maximum likelihood analysis of 216 alignment positions. Branches show maximum likelihood bootstrap support (MLBP). Branch support, GI numbers and substitutions per site are indicated as per Fig. 1. Eukaryotic IF2 sequences are mitochondrial unless otherwise labeled.

Figure S2: Phylogeny of the "full IF3" dataset. The tree is generated from a RaxML maximum likelihood analysis of 139 alignment positions. Branches show maximum likelihood bootstrap support (MLBP). Branch support, GI numbers and substitutions per site are indicated as per Fig. 1. Eukaryotic sequences are mitochondrial unless otherwise labeled.

Figure S3: ***Effects of S. cerevisiae Aim23 in heterologous E. coli translational system.*** (**A**) Detection of Aim23p binding to *E. coli* 30S ribosomal subunit. Aim23p and IF3 in 100 μl fractions from S-300 HR gel-filtration columns were detected by anti-6His Western dot-blot. Elution profiles of 30S (empty circles) and IF3 (solid circles) alone followed by UV absorbance are provided underneath the WB panels. *E. coli* 30S showed no detectible anti-6His signal (data not shown). (**B**) Effects of *E. coli* IF3 and *S. cerevisiae* Aim23p on the dissociation of *E. coli* 70S. *E. coli* IF3 and *S. cerevisiae* Aim23p were added in 10-fold or 20-fold excess over *E. coli* 70S ribosomes, and the mixture was analyzed by sucrose density gradient centrifugation. Untreated *E. coli* 70S in the absence of added protein served as a control.

Table S1: Strains and plasmids.
Names and descriptions of strains and plasmids used

Table S2: Distribution of mitochondrial initiation factors and activator proteins across a wide distribution of eukaryotes. For each factor and activator in each organism, the NCBI GI number of the sequence is shown. Following the activator columns, the remaining columns describe the taxonomy of the source organism. Text in rows corresponding to Saccharomycotina is colored in turquoise. The divergent Kinetoplastida mIF2-2s are not included. Where a gene was not identified in the exact taxon, but was found in another member of the same genus, the GI number is followed by the species name in parentheses.

Table S3: Amino acid content in initiation factors. Percentage composition for each amino acid were calculated with the ProtParam server (http://expasy.org/tools/protparam.html). Composition of the complete Uniprot database (as of 10 January 2010) was retrieved from http://expasy.org/sprot/relnotes/relstat.html.

SI References:

1. Katoh, K., Kuma, K., Toh, H. and Miyata, T. (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res*, **33**, 511-518.
2. Stamatakis, A. (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22**, 2688-2690.
3. Huelsenbeck, J.P. and Ronquist, F. (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, **17**, 754-755.
4. Price, M.N., Dehal, P.S. and Arkin, A.P. (2009) FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol*, **26**, 1641-1650.

5. Oldenburg, K.R., Vo, K.T., Michaelis, S. and Paddon, C. (1997) Recombination-mediated PCR-directed plasmid construction in vivo in yeast. *Nucleic Acids Res*, **25**, 451-452.
6. Gietz, R.D. and Woods, R.A. (2006) Yeast transformation by the LiAc/SS Carrier DNA/PEG method. *Methods Mol Biol*, **313**, 107-120.
7. Kaiser, C., Michaelis, S. and Mitchell, A. (1994) *Methods in Yeast Genetics: A Cold Spring Harbor Laboratory Course Manual*. CSHL Press.
8. Ocampo, A., Zambrano, A. and Barrientos, A. (2010) Suppression of polyglutamine-induced cytotoxicity in Saccharomyces cerevisiae by enhancement of mitochondrial biogenesis. *FASEB J*, **24**, 1431-1441.
9. Ludovico, P., Sansonetty, F. and Corte-Real, M. (2001) Assessment of mitochondrial membrane potential in yeast cell populations by flow cytometry. *Microbiology*, **147**, 3335-3343.
10. Shapiro, H.M. (2000) Membrane potential estimation by flow cytometry. *Methods*, **21**, 271-279.
11. Claros, M.G. and Vincens, P. (1996) Computational method to predict mitochondrially imported proteins and their targeting sequences. *Eur J Biochem*, **241**, 779-786.
12. Koc, E. and Spremulli, L. (2002) Identification of mammalian mitochondrial translational initiation factor 3 and examination of its role in initiation complex formation with natural mRNAs. *J Biol Chem*, **277**, 35541-35549.

Figure S1

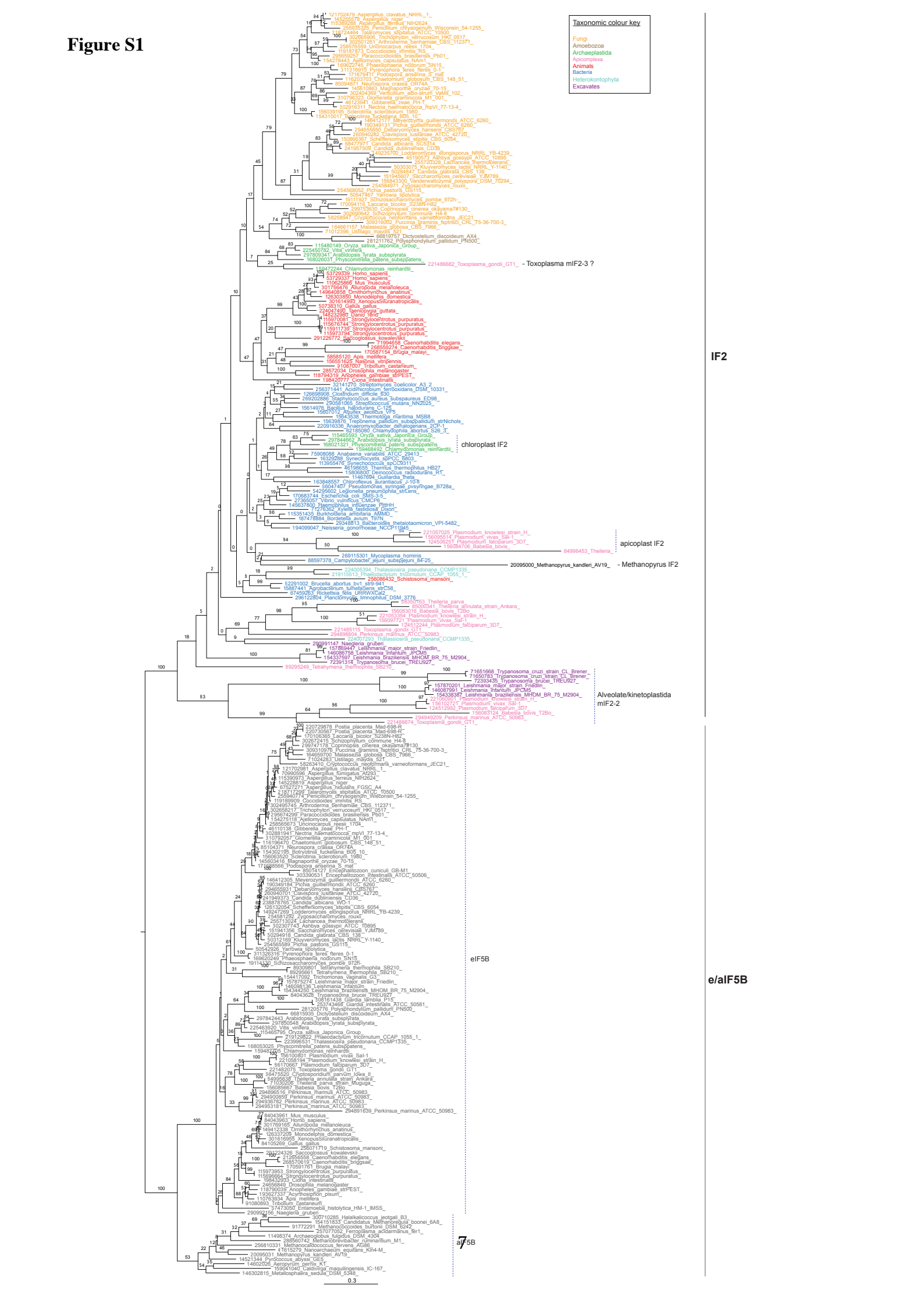
Taxonomic colour key
Fungi
Amoebozoa
Archaeplastida
Apicomplexa
Animals
Bacteria
Heterokontophyta
Excavates
- Toxoplasma mIF2-3 ?
IF2
chloroplast IF2
apicoplast IF2
- Methanopyrus IF2
Alveolate/kinetoplastida mIF2-2
eIF5B
e/aIF5B
aIF5B
0.3

Figure S2

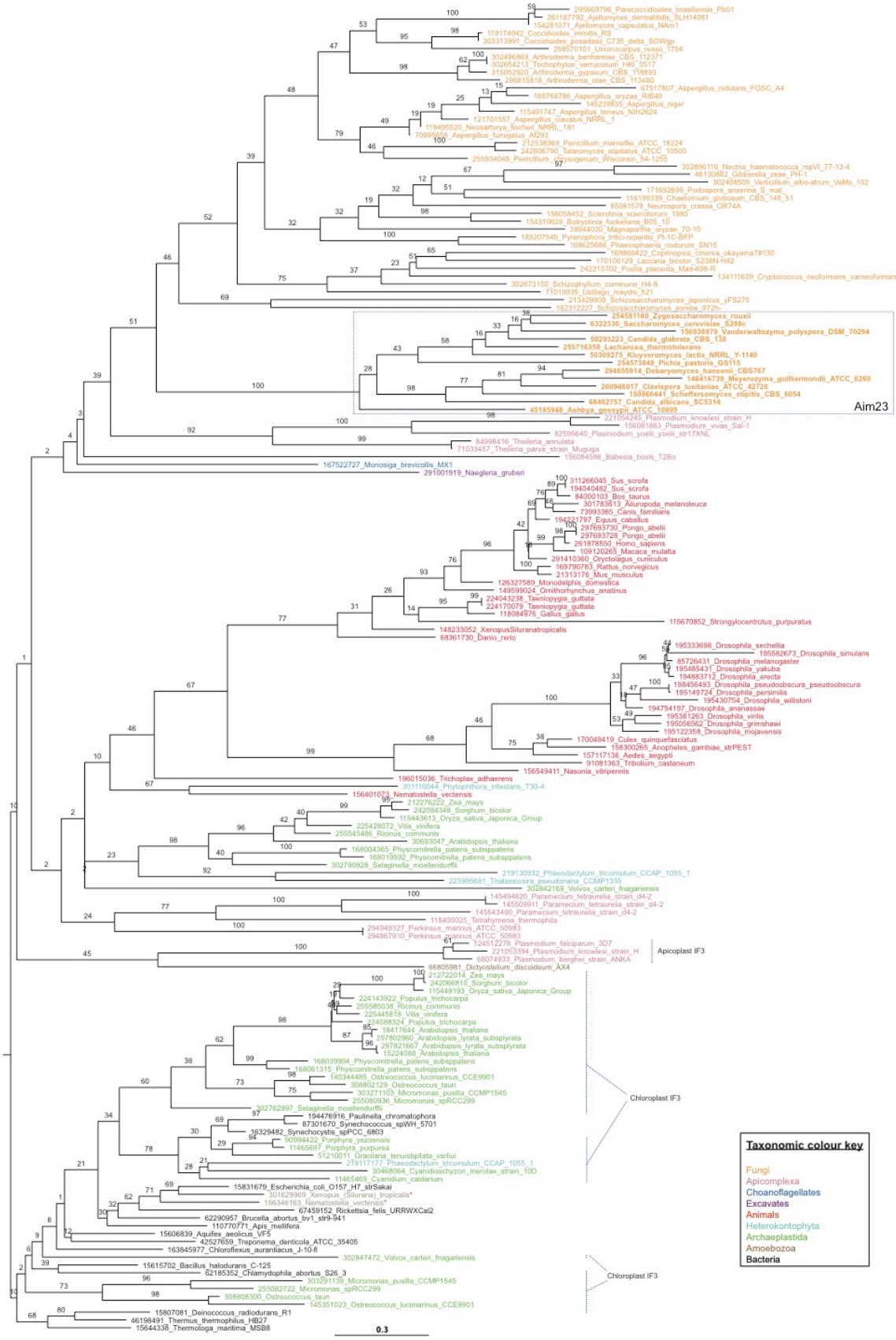
Aim23
Apicoplast IF3
Chloroplast IF3
Chloroplast IF3
Taxonomic colour key
Fungi
Apicomplexa
Choanoflagellates
Excavates
Animals
Heterokontophyta
Archaeplastida
Amoebozoa
Bacteria
0.3

Figure S3

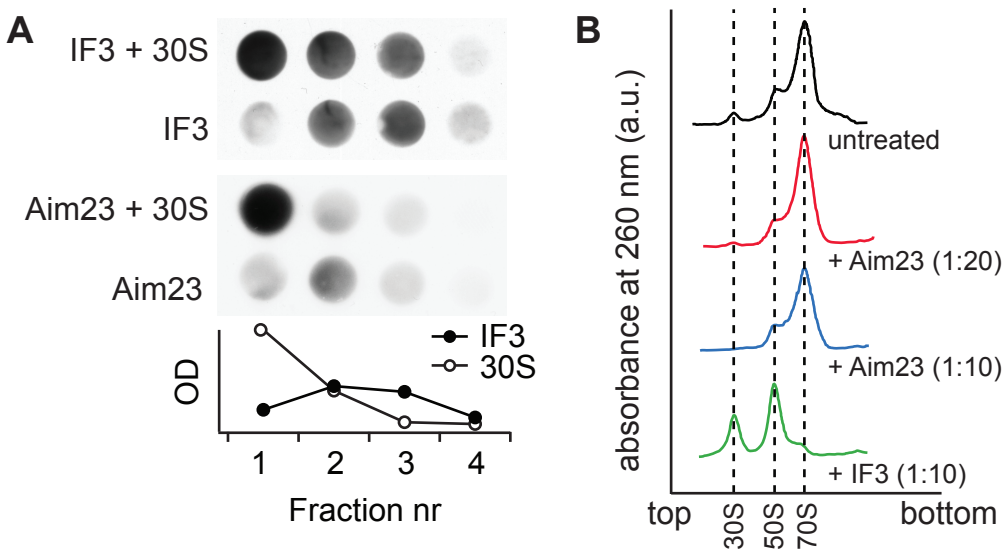

Table S1: Strains and plasmids.

Strain or plasmid	Comment
Y21294	Heterozygous diploid with one allele of *AIM23/YJL131c* replaced by KanMX4 cassette (BY4743; Mat a/α; his3D1/his3D1; leu2D0/leu2D0; lys2D0/LYS2; MET15/met15D0; ura3D0/ura3D0; YJL131c::kanMX4/YJL131c). (EUROSCARF).
HKA201	Haploid strain derivative of Y21294 with *AIM23/YJL131c* genomic deletion bearing *AIM23/YJL131c* on a plasmid (pYCG_YJL131c).
HKA202	Haploid strain derivative of Y21294 with *AIM23/YJL131c* genomic deletion (*AIM23Δ*) bearing *S.pMIF3*/SPBC18E5.13 on a plasmid (pSPBC18E5a).
HKA203	Haploid strain derivative of Y21294 with *AIM23/YJL131c* genomic deletion.
pYCG_YJL131c	Yeast centromeric plasmid encoding *AIM23/YJL131c* (URA3 marker). (EUROSCARF).
pSPBC18E5a	Yeast centromeric plasmid encoding *S.p.MIF3/SPBC18E5.13* (URA3 marker)
AIM23Δ1-32 pET28a	pET28a-based expression construct for 6His-tagged *S. cerevisiae* Aim23p lacking mitochondrial localization signal (amino acids 1-32)