Published in final edited form as:

Proteomics. 2009 January ; 9(2): 434–450. doi:10.1002/pmic.200800477.

A comprehensive analysis of *Trypanosoma brucei* mitochondrial proteome

Aswini K. Panigrahi, Yuko Ogata, Alena Zíková, Atashi Anupama, Rachel A. Dalley, Nathalie Acestor, Peter J. Myler, and Kenneth D. Stuart*

Seattle Biomedical Research Institute, 307 Westlake Ave N, Suite 500, Seattle, WA 98109, USA

Abstract

The composition of the large, single, mitochondrion of *T. brucei* was characterized by mass spectrometry (2D-LC-MS/MS and gel-LC-MS/MS) analyses. A total of 2,897 proteins representing a substantial proportion of procyclic form cellular proteome were identified, which confirmed the validity of the vast majority of gene predictions. The data also showed that the genes annotated as hypothetical (species specific) were over-predicted and that virtually all genes annotated as hypothetical, unlikely are not expressed. By comparing the mass spectrometry data with genome sequence, 40 genes were identified that were not previously predicted. The data are placed in a publicly available web-based database (www.TrypsProteome.org). The total mitochondrial proteome is estimated at 1,008 proteins, with 401, 196, and 283 assigned to the mitochondrion with high, moderate, and lower confidence, respectively. The remaining mitochondrial proteins were estimated by statistical methods although individual assignments could not be made. The identified proteins have predicted roles in macromolecular, metabolic, energy generating, and transport processes providing a comprehensive profile of the protein content and function of the *T. brucei* mitochondrion.

Keywords

Database; Mass spectrometry; Mitochondrion; Organelle fractionation; Proteomics

1 INTRODUCTION

Trypanosomes are protozoan parasites that cause enormous disease burden, African trypanosomiasis caused by *Trypanosoma brucei* contributes to 1.5 million DALYs (Disability Adjusted Life Years), Chagas disease caused by *T. cruzi* contributes to 667,000 DALYs, and Leishmaniasis contribute to 2 million DALYs (World Health Report, 2004). Sequencing of the *T. brucei*, *T. cruzi* and *L. major* (the TriTryps) genomes is essentially complete [1–3] and accumulation of extensive genome sequence information poses new challenges as well as opportunities for post-genomic research. The TriTryp genomes have substantial sequence, gene content, and gene order conservation [4], and most of the basic cellular processes are shared among these trypanosomatids. Bioinformatics and comparative genomics play powerful roles in identifying putative genes, and defining the potential functions and relationships of many genes in the repertoire. However about 2/3rd of the predicted genes in these organisms have no known function and are currently annotated as encoding hypothetical proteins (www.genedb.org). *T. brucei* genome is predicted to encode 9,211 proteins, of which only 35.7% have been assigned functional roles based on experimental data (5.1%) or sequence

*Corresponding Author: ken.stuart@sbri.org, Ph: 206-256-7316, Fax: 206-256-7229.

similarities to proteins of known function in other organisms (30.6%) (www.genedb.org/genedb/tryp/index.jsp). The gene predictions in the TriTryps have not been systematically tested to determine if the predicted protein is present, let alone whether the predicted functions are accurate. The Trypanosome genomes have an unusual organization. They consist of clusters of numerous genes on the same DNA strand (directional clusters), each of these appear to be transcribed from single promoter-like elements, and RNA abundance is primarily regulated by transcript processing and turnover [4,5]. Most biological processes are controlled at the protein rather than RNA level and this may be especially true in *T. brucei* where regulation of transcription is rare and regulation of translation has been demonstrated [6,7]. Thus experimental evidence for gene expression at the protein level is important in defining the potential function of Trypanosomatid genomes. Progress in the development of mass spectrometric proteomics technologies enables proteins to be analyzed in a high throughput, automated manner. Fortunately, essentially all trypanosomatid genes lack introns, which simplifies gene identification and aids proteomic characterization. Such an approach can identify the molecular components of organelles, sub-cellular structures, and biological macromolecular complexes, as well as determining levels of protein expression between two different cell states, and various post-translational modifications that control regulatory pathways. Thus while the availability of the TriTryp genome sequences has accelerated research progress in many laboratories, only limited information has been generated at the proteome level for these organisms [8–14].

In this study we used a shotgun proteomics approach to identify proteins present in the mitochondrion of *T. brucei* procyclic form (PF) cells. The resultant profile was compared to the genome database and used to assess the validity of gene annotation. The results substantially and efficiently advance the annotation of trypanosomatid genomes. The proteomic data also enabled us to identify a set of new genes in *T. brucei*. Identified proteins were assigned to mitochondrion (mt) by criteria including enrichment in the organelle fraction, demonstrated or putative role in relevant biological processes, and association with known mitochondrial complexes, especially for those with unknown functions. We also identified a large set of proteins with unknown function that are likely associated with multi-protein mt complexes. We have created a web-based database “www.TrypsProteome.org” for dissemination of the proteomic data from these analyses.

2 MATERIALS AND METHODS

2.1 Cell growth, isolation of mitochondrial vesicles and lysis

Trypanosoma brucei procyclic form (PF) cells IsTaR 1.7a were grown at 27°C in SDM-79 media containing hemin (7.5 mg/ml) and 10% FBS. The cells were harvested at mid-log phase of growth by centrifugation at $6,000 \times g$ for 10 min at 4°C. The mitochondrial vesicles were isolated from PF cells by hypotonic lysis followed by Percoll gradient floatation as described [15]. Briefly, $\sim 2 \times 10^{10}$ PF cells were harvested at mid-log phase of growth and washed with 30 ml of SBG buffer (20 mM phosphate buffer, pH 7.9, 150 mM NaCl, 6 mM glucose). The cells were resuspended in 20 ml of DTE buffer (1 mM Tris-HCl, pH 8.0, 1 mM EDTA), disrupted by 5 strokes in Dounce homogenizer and immediately sucrose was added to a final concentration of 0.25 M (3.34 ml of 60% sucrose solution). After mixing the lysate was centrifuged at $15,000 \times g$ for 10 min at 4°C. The organelle enriched pellet was resuspended in 3.9 ml of STM buffer (20 mM Tris-HCl pH 8.0, 250 mM sucrose, 2 mM $MgCl_2$) and treated with DNase (9 µg/ml final concentration). The sample was incubated in ice for 60 min following which equal volume of STE buffer (20 mM Tris-HCl pH 8.0, 250 mM sucrose, 2 mM EDTA) was added, mixed, and centrifuge as above. The pellet was resuspended in 4 ml of 70% Percoll using a small Dounce homogenizer with tight fitting pestle B for 5 strokes, layered at bottom of a 32 ml of 20–35% linear Percoll gradient and centrifuged at $103,900 \times g$ for 60 min at 4°C.

C. The mitochondria enriched fraction that appears in the density range of 1.052 to 1.069 g/ml was collected using a syringe and 18-gauge needle and washed 4 times with STE buffer and the mitochondrial vesicles were pelleted by centrifugation at $32,530 \times g$ for 15 min.

The PF cells were washed with 1X PBS and lysed with 1% Triton X-100 with bi-directional mixing for 15 min at 4 °C. The lysed samples were separated to soluble supernatant and insoluble pellet fractions by centrifugation at $17,500 \times g$ for 30 min at 4 °C. The pellet was washed thrice with 1X PBS, 1% Triton X-100 solution, and the cleared supernatant and pellet fractions were collected and analyzed by mass spectrometry (see 2.4). Similarly mitochondrial vesicles were lysed with 1% Triton X-100 and separated to cleared supernatant and pellet fractions as above.

2.4 Mass spectrometry

The proteins in detergent soluble fractions were digested with sequencing grade modified trypsin (Promega) and the resulting peptides were analyzed by two-dimensional liquid chromatography tandem mass spectrometry (2D-LC-MS/MS). In first dimension the peptides were fractionated by off-line strong cation exchange (SCX) chromatography, multiple fractions were collected and in second dimension the peptides were further fractionated by on-line reverse phase (RP) chromatography. Briefly, 200 µg of proteins from detergent soluble fractions of PF cells and PF mitochondria were precipitated separately with 6 volumes of Acetone. The precipitates were dissolved in 8M Urea, 1 mM DTT and incubated at 50 °C for 1 h. After 4 fold dilution with 50 mM ammonium bicarbonate the proteins were digested with 2 µg trypsin O/N. The peptide samples were diluted 1:8 with 5% acetonitrile in 0.4% acetic acid buffer and loaded onto a 10 cm long x 2.1 mm ID polysulfoethyl column (PolyLC Inc) at a flow rate of 200 µl/min. The unbound peptides were washed away with 5% acetonitrile in 0.4% acetic acid at 200 µl/min flow rate for 10–20 min until A_{280} of the flow through reached the base line. The peptides were eluted with a 20 min linear gradient of 0–200 mM of ammonium acetate in 5% acetonitrile and 0.4% acetic acid, followed by a 5 min linear gradient of 200–500 mM of ammonium acetate in 5% acetonitrile and 0.4% acetic acid at 200 µl/min flow rate. Fractions of 200 µl were collected and dried in Speed Vac. The peptides in each fraction were dissolved in 10 µl of 5% acetonitrile, 0.4% acetic acid buffer and loaded onto a 10 cm long x 75 µm ID C18 capillary column at a flow rate of 200 nl/min. The peptide elution from C18 column was achieved by 5 min isocratic flow of 5% acetonitrile and 0.4% acetic acid followed by a 45 min linear gradient of 5–40% acetonitrile in 0.4% acetic acid, and a 5 min linear gradient of 40–80% acetonitrile in 0.4% acetic acid. The eluted peptides were analyzed on-line by electrospray ionization tandem mass spectrometry using a LTQ mass spectrometer (Thermo Electron) that was tuned for optimal performance at 2.2 kV spray voltage and 200°C capillary temperature using MRFA ion 524.3 at monthly interval. Xcalibur 1.4 SR1 version software was used to collect mass spectrometry data and the mass range for the MS survey scan was m/z 400–1400. Each MS scan was followed by 5 MS/MS scans and the data was collected using a dynamic exclusion method where a specific ion was sequenced twice at a maximum and is excluded from the list for 45 seconds.

The proteins in insoluble fractions were dissolved in 1X SDS-PAGE buffer, separated on 10% SDS-PAGE gels and stained with SYPRO Ruby stain (Invitrogen). Each gel lane was divided into 12 approximately equivalent pieces, the proteins were digested in-gel with trypsin O/N, and the resulting peptides were extracted (with 50% acetonitrile, 5% formic acid) and dried in Speed Vac [15]. The peptides were fractionated by C18 RP chromatography and analyzed online by mass spectrometry as above.

2.5 MS/MS Data analysis

The mass spectrometry data was analyzed against *T. brucei* sequence databases using TurboSEQUEST program in BioworksBrowser 3.1 software package (Thermo Electron) in a multi-processor cluster platform. The peak lists were generated using the Sequest module of Bioworks 3.1, cluster version SR1 using the default parameters (MW range: 400–3500, precursor mass tolerance: 1.4, group scan: 25 and minimum ion count: 15). The MS/MS data was compared with v4.0 predicted protein sequence database [1] (www.genedb.org). The database contained 9,211 *T. brucei* nuclear encoded protein sequences of which 612 are annotated as hypothetical unlikely plus 18 mitochondrial encoded protein sequences (we also included mouse immunoglobulin heavy and light chains, bovine serum albumin and human keratin sequences in the database). Parallel data analysis was carried out with a polypeptide database that contained all polypeptides of ≥ 50 amino acids (STOP codon to STOP codon) from six-frame translated *T. brucei* genome sequence (total 271,892 entries) (ftp://ftp.sanger.ac.uk/pub/databases/T.brucei_sequences/T.brucei_genome_v4/). No enzyme was specified during the SEQUEST search, peptide mass tolerance was set at 1.4 and fragment ion-tolerance at 0.0 (per the default parameters recommended by the manufacturer for good quality data). No fixed modification was set for any of the amino acids but differential modification for 'M' was set at 15.994. The output from SEQUEST search was filtered and compiled using PeptideProphet and ProteinProphet programs [16,17] using a local semiautomated platform built upon Trans-Proteomic Pipeline (TPP) (<http://tools.proteomecenter.org/software.php>).

The dataset presented here include only the doubly tryptic peptides that have minimum peptide identification probability of 0.9 and have a minimum SEQUEST X-correlation value of 1.5 for +1 ions, 1.8 for +2 ions, and 2.5 for +3 ions. We excluded any peptide containing more than one missed trypsin cleavage site in the sequence and that containing cysteine amino acid since alkylation step was not carried out during sample preparation. Proteins containing these peptides and with minimum identification probability of 0.9 were considered positive.

2.6 Sequence Analysis

In selected cases the homology searches with TriTryp databases was carried out using OmniBLAST to identify any homologous or related proteins/genes. The probable functions of the proteins were assigned based on GeneDB annotation and for proteins with unknown function possible motifs and domains were searched in the PROSITE, InterPro and CDD databases.

2.7 Proteomics Database

The data from different experiments were stored and accessed via the Proteomics module of the SBEAMS (Systems Biology Experiment Analysis Management System) database (<http://www.sbeams.org/Proteomics/>) built on MS SQL Server 2000. The results from SEQUEST, PeptideProphet and ProteinProphet analyses were imported into the database using built-in Perl scripts. The results from single or sets of experiments were filtered with specific parameters (as above and in Results section) and compiled using SQL queries, and the output data saved in new Tables. A web based database (www.TrypsProteome.org) was developed using Microsoft .net framework 1.1. It is using Web Service to pull out data from compiled Tables and locally stored *T. brucei* GeneDB and Gene Ontology (GO) databases in .xml format.

3 RESULTS

3.1 Proteome coverage

We used a combination of cellular fractionation (non-ionic detergent soluble and insoluble) and sub-cellular fractionation (enrichment of mitochondrial vesicles) followed by protein fractionation (1D-gel) and peptide fractionation (SCX/RP chromatography) techniques for enhancing the coverage of peptides in mass spectrometry analyses. The peptides from Triton X-100 soluble supernatant fractions of whole cell and isolated mitochondria were fractionated by two-dimensional liquid chromatography and analyzed by tandem mass spectrometry (2D-LC-MS/MS) (see Supplementary Figure 1 for representative results). Reciprocally the proteins in insoluble pellet fractions were fractionated based on size in 1D SDS-PAGE gel, peptides were generated from multiple fractions and analyzed by RP-LC-MS/MS. The mass spectrometry data was analyzed and compiled as described in Methods section, and uploaded to SBEAMS database. In whole cell detergent soluble supernatant 1,689 proteins were identified by 2D-LC-MS/MS analysis, and 810 proteins were identified in pellet fraction by 1D-gel-LC-MS/MS analysis. There were 477 proteins in common between these two datasets, thus 2,022 proteins were identified in whole cell sample by MS/MS analysis. Similar analyses identified 1,548 proteins in mitochondrial enriched fraction of which 673 proteins were also identified in whole cell analysis (Figure 1). Thus 875 additional proteins were identified by analysis of mitochondrial enriched fraction compared to the whole cell fraction, and in total 2,897 proteins were identified in these analyses of which 1,333 have been assigned to known or putative function(s) (Supplementary Table 1). These results represent a substantial proportion of the *T. brucei* PF cellular proteome at mid-log phase of growth, and it also showed in a complex proteome like *T. brucei* analyses of sub-cellular/organelle fractions is required for maximal proteome coverage. The compiled results from multiple mass spectrometry runs and from different sample preparation methods enhanced the protein coverage [18]. It resulted in detection of 65–75% of the same proteins between runs depending on sample complexity and second runs yielding 13–32% increase in proteins identified in highly complex samples, as seen by others [19].

In these analyses 12,131 unique peptides were identified (additionally 727 of these peptides were also identified in modified form), and 916 proteins were matched to these by a single peptide hit and 1,981 proteins by two or more peptide hits. The entire latter group of proteins had a very high protein identification probability (≥ 0.99 for 1,972 proteins and 0.98 for the other 9 proteins). Of the proteins identified with one peptide match 664 (72%) had protein identification probability of ≥ 0.99 , 111 (12%) had 0.98 and the rest between 0.9 and 0.97. The identified peptide sequences and associated probability values are presented in Supplementary Table 1.

3.2 Assessing validity of gene prediction

The *T. brucei* genome, excluding the pseudogenes, has had 9,211 protein coding sequences predicted (v4.0 database), 36% of which are annotated as encoding proteins with known or putative function, 51% as hypothetical conserved, 6% as hypothetical and 7% as hypothetical unlikely. However, of the 2,897 proteins identified in our mass spectrometry analysis representing the partial proteome 46% have assigned functions, 53% annotated as hypothetical conserved, ~1% as hypothetical and we did not identify any protein from hypothetical unlikely group (Figure 2). We observed a similar proportion in analyses of the BF cellular proteome (Results not shown). These results indicate that only a small proportion of the genes annotated as hypothetical and virtually none of the genes annotated as hypothetical unlikely are actually expressed in *T. brucei* cells. It also shows that a majority (but probably not all) of the genes annotated as hypothetical conserved are expressed in the cell. If we assume that almost all of the proteins annotated with assigned functions are expressed in cell during some stage then

based on the observed ratio to proteins of unknown function it would extrapolate to ~7,210 proteins being expressed in *T. brucei* cells.

The acquired MS/MS data was also compared with predicted polypeptide sequences from nucleic acid (NA) database. While there was a very good concordance compared to the results obtained from v4.0 protein database some discrepancies were also apparent, especially in probability values of the identified peptides, and using the cut-off described above it missed some of the peptides identified by comparison to protein database (Results not shown). We did not identify 243 of the peptides that were identified in comparison to v4.0 protein database upon comparison of the MS/MS data to NA database (based on the best hit criteria, line 1 of .out file). It resulted in non-detection of 23 proteins, all originally identified with only one peptide hit against protein database. This predicts an error rate of 2% at peptide assignment level and 0.8% at protein assignment level.

We identified 146 unique peptides that correspond to predicted polypeptide sequences that are either not annotated as predicted genes (n=53) or to annotated genes (n=22). Four of the latter 22 genes more recently have been annotated as predicted genes; and one or more peptides in the other 18 annotated genes matched to predicted amino acid sequences upstream of currently annotated AUG start codon (Supplementary Figure 2). In 14 of these genes a start codon could be predicted upstream of the identified peptide(s). However, the other 4 lack a start codon upstream of the identified peptide sequence indicating possible sequencing error or an alternative start codon in these proteins. It is also possible that the strain used for proteomic studies may have slight differences in genomic sequences that could be reflected in different start codon upstream of the identified peptide sequences. Sequencing error appears to be most likely, especially since the homologs of 3 of these proteins (Tb927.3.2740, Tb927.3.4920 and Tb10.70.3350) are larger in both *L. major* and *T. cruzi* and span the polypeptide region identified upstream of the annotated start codon. Thus, this proteomic study identifies the start codon for a group of genes.

Of the 53 polypeptides identified in this analysis that do not map to currently annotated genes, 13 had no predicted start codon upstream of the matched peptide sequences and all were identified by single peptide hit, indicating that they may be false hits. As above these results are well within the estimated error range. Thus at an increased confidence level, 19 new ORFs were identified by two or more peptide matches and are likely bona-fide ORFs that were missed in the GeneDB annotation (Supplementary Figure 3A). Five of these proteins belong to the retrotransposon hot spot (RHS) protein group and 11 others have varying degrees of homology to predicted *T. cruzi* and/or *L. major* proteins, and the 3 other have homology to polypeptides predicted from *T. cruzi* (and *L. major* in 2 of the cases) genome sequences (Supplementary Table 2A). Eleven other polypeptides (including 3 belonging to RHS protein group), were each identified by only one peptide match (Supplementary Figure 3B) but have similarities to predicted *T. cruzi* and *L. major* proteins (Supplementary Table 2B) and thus are also likely bona-fide ORFs. The other ten polypeptides were identified by single peptide hits (Supplementary Figure 3C) but have no significant homology to annotated *T. cruzi* or *L. major* proteins. However, eight of those have some similarities to predicted polypeptides from *T. cruzi* and/or *L. major* contig sequences (Supplementary Table 2c). Thus this study identified 40 additional ORFs (30 with high confidence and 10 possible) that were missed in the GeneDB annotation.

3.3 Preliminary assignment of proteins to mitochondrion

We assessed the ability of available software [Mitoprot (<http://ihg.gsf.de/ihg/mitoprot.html>), SignalP (<http://www.cbs.dtu.dk/services/SignalP/>), Predotar (<http://urgi.versailles.inra.fr/predotar/predotar.html>), TargetP (<http://www.cbs.dtu.dk/services/TargetP/>) and PSORT (<http://psort.nibb.ac.jp/form2.html>)] in

predicting the localization of *T. brucei* proteins to the mitochondrion using a set of known mitochondrial and non-mitochondrial proteins. The results showed different programs have different level of sensitivity vs. specificity and the correlations between different programs were poor. Overall Mitoprot and SignalP performed better than the other programs (Results not shown). Representative results obtained from Mitoprot and SignalP programs are shown in Figure 3, where the relative scores obtained for each proteins are plotted from a set of known mitochondrial proteins (ten proteins each from editosome [20] and MRB complex 1 [21] in panel A and non-mitochondrial (ten proteins each from glycosome and cytoplasm) in panel B. The results showed that while the Mitoprot program was able to identify most of the known mitochondrial proteins, it also had the highest false positive rate by predicting known non-mitochondrial proteins as being mitochondrial. Even combining both the programs failed to identify approximately 25% of the mitochondrial proteins (see Supplementary Table 4 for relative scores of proteins assigned to mitochondria in this study). Thus the available programs have limited use in predicting localization of proteins to mitochondrion in *T. brucei* and additional qualifying criteria are required for sub-cellular assignment of proteins in this organism.

In this study we identified 1,548 proteins in mitochondria enriched fraction of which 607 (39%) have been assigned with a function. A randomized statistical approach was used to assess the coverage of mitochondrial proteome achieved in this analysis. We selected the proteins having 'mitochondrion/mitochondrial' 'text' in GO cellular component assignment, and calculated the proportion of those proteins that were identified in our shotgun proteomic analysis. The results indicate that we have identified ~86% of mitochondrial proteome (results not shown). This is supported by our observation that we only missed detecting one of the twenty annotated editosome proteins which are in low abundance in the mitochondrion.

We anticipate that not all of the proteins identified in this fraction will be mitochondrial. This is due in part to the fact that the single structurally complex mitochondrion is disrupted and reseals as vesicles during isolation, and also due to sample cross contamination. Based on available GO annotation, key-word search in protein description and literature references [11,13] 139 of 607 proteins that are assigned with function(s) (23%) appear to be non-mitochondrial. These proteins are assigned to other compartments of cell such as cytoplasm (8.7%), glycosome (6.6%), cytoskeleton and flagellum (4.4%), nucleus (2%) and others (Supplementary Table 3C). Additionally 52 other proteins (8.6%) may localize to membrane (Supplementary Table 3A), and we anticipate a large proportion of those would be associated with mitochondrial membrane. While 194 (32%) proteins could be assigned to mitochondrion (Table 1), the other 222 proteins (36.6%) have not been assigned to any cellular compartment (Supplementary Table 3B). It is likely that the majority of the 941 proteins that were identified in mitochondrial enriched fraction that have no known function will also be mitochondrial and the above ratio will be reflected in this group. Overall of the 385 proteins that are assignable (demonstrated or putative) to a specific sub-cellular compartment ~44% appear to be non-mitochondrial. We anticipated part of the proteins that are assigned to membrane will be non-mitochondrial. Thus, the other 56% are likely mitochondrial proteins and hence by extrapolation from our current coverage the *T. brucei* mitochondrial proteome is predicted to consist of ~1,008 proteins.

In general the mitochondrial and glycosomal proteins were identified with higher peptide coverage (peptide count) in organelle enriched fraction compared to the whole cell fraction (Table 1, Supplementary Table 3C). Reciprocally cytoplasmic and nuclear proteins were identified with higher peptide coverage in the whole cell fraction (Supplementary Table 3C). Similarly in recent proteomic studies from *T. brucei* large sets of mitochondrial proteins were identified in glycosome enriched fraction in addition to proteins from other cellular compartments [11,13]. Thus solely based on the identification of proteins in a specific organelle

enriched fraction, they may not be assignable to the specific sub-cellular compartment and additional studies are required to determine their true localization. However, qualitatively proteins identified only in mitochondrial enriched fraction or at significantly higher peptide coverage compared to the whole cell fraction can be assigned to mitochondrion with higher confidence, especially for those lacking a predicted glycosomal localization signal.

3.4 Proteins likely associated with mitochondrial complexes

To increase the confidence of protein assignments to the mitochondrion we carried out glycerol gradient experiment in which the Triton X-100 soluble fraction of enriched mitochondria was fractionated, and fractions from high 'S' value region (~20S, ~40S and ~80S) were analyzed by mass spectrometry. We anticipate that a majority of the proteins identified in these fractions (see Supplementary figure 4 for SDS-PAGE protein profile of the fractions) are likely to be associated with multi-protein complexes which are probably mitochondrial. We identified 633 proteins by analysis of these 3 different fractions that were also identified in mitochondrial enriched fraction. In this group 336 (53%) proteins are currently annotated as hypothetical (no known function) (Table 2 and Supplementary Table 5), and 297 (47%) have been assigned with function(s). In the later group 134 proteins (45%) are assignable to the mitochondrion (Table 1), 31 (10.4%) to the glycosome, 16 (5.4%) to membrane, and 101 proteins (34%) have not been assigned to any cellular compartment (Supplementary Table 3). Only 5% of the proteins are assignable to another cellular compartment such as the cytoplasm or nucleus compared to 16.3% in the mitochondrial enriched fraction. Indeed some proteins that are assigned to cytoplasm, such as dihydrolipoamide dehydrogenase proteins (Tb927.3.4390 and Tb927.8.7380) and heat shock protein HslVU (Tb927.5.1520 and Tb11.01.4050), were identified only in mitochondrial fractions or with significantly higher peptide coverage in this fraction compared to the whole cell fraction. The data indicate that these proteins are mitochondrial although they had been assigned to the cytoplasm (Supplementary Table 3). Similarly, FBPase fructose-1,6-bisphosphate which is annotated as cytosolic (Tb09.211.0540) may localize to the glycosome. TOP2 DNA topoisomerase II (Tb09.160.4090) which localizes to the mitochondrion [22] and was only identified in the mitochondrial enriched fraction is currently assigned to the nucleus in the GO database. In addition, preliminary results from our lab indicate that heat shock proteins HslVU (Tb927.5.1520 and Tb11.01.4050) are mitochondrial (Acestor N, unpublished results). Thus, results from this study are substantially refining the sub-cellular assignment of a large set of proteins. We estimate most (>95%) of the 633 proteins identified in the glycerol gradient fractions localize to the mitochondrion or glycosome.

We assigned 194 of the proteins identified by mass spectrometry analyses to the mitochondrion based on known/putative function, keyword search, GO annotation and publications [11,13, 21,23] and grouped those in Table 1 by association with various biological processes. The results showed a large sub-set of the proteins that we identified in glycerol gradient samples are associated with mitochondrial multi-protein complexes, e.g. we identified protein components of respiratory complexes I-V, RNA editing complex, and ribosome etc. In recent studies from our laboratory we have determined the composition of several mitochondrial complexes using affinity tag, monoclonal antibody affinity purification and mass spectrometry analyses [20,21,23]. Of the 336 proteins identified in glycerol gradient sample that have no known function 20 proteins are associated with respiratory complex I and the MRB complex 1 [21], 72 proteins with mt ribosomes [23], and 40 others are associated with other mitochondrial complexes (Alena Zikova, manuscripts in preparation). In total we have identified 207 proteins with unknown function that were in the mitochondrial enriched fraction and are associated with mitochondrial complexes (Table 2). These proteins are currently annotated as hypothetical but conserved motifs/domains that are indicative of possible function (s) were identified in 59 of these proteins (Table 2).

We anticipate that the majority of the 204 proteins with unknown function that were identified in glycerol gradient fraction but have not been assigned to any complexes based on our current knowledge are potential mitochondrial proteins (Supplementary Table 5). Six of these proteins appear to be glycosomal based on motifs and targeting signal and 2 other are non-mitochondrial based on peptide coverage compared to whole cell fraction. Furthermore, more than 500 proteins with unknown function were identified in the mitochondrial enriched fraction. These have not been included in the high confidence list, but a substantial proportion of them are likely to be mitochondrial. As indicated in section 3.3, peptide count information would be applicable for preliminary assignment to a sub-set of those proteins, however further qualifying criteria are required to assign them to mitochondria. In Supplementary Table 6, we provide a list of 283 proteins that are likely mitochondrial. These include proteins identified with at least 2 peptides in mitochondrial enriched fraction and detected only or with higher peptide count compared to whole cell fraction. We also excluded any proteins that have putative a glycosomal targeting signal.

Overall, in this study 401 proteins were assigned to *T. brucei* mitochondrion (Table 1 and 2) based on their assigned function, GO annotation and/or specific association to mitochondrial complexes. It also provides a list of 196 high confidence candidate proteins (Supplementary Table 5) majority of which are expected to be associated with mt complexes, and 283 likely mitochondrial proteins (Supplementary Table 6) that need further follow up for definitive assignment.

3.5 TrypsProteome database

The dataset presented here and that generated in our *T. brucei* mitochondrial proteome project is being available via a website (www.TrypsProteome.org). The database is searchable by several fields and is also linked to GeneDB database. Currently the information on protein identification, the identified peptides and their respective probability values are available on the website. The mass spectrometry identification of peptides and assignment to specific proteins were carried out by well established and widely accepted criteria, however, it is important to note that they are based on statistical confidence levels and possible error rate as explained above should be taken into consideration when qualifying the data. The database is planned to include detailed information on complex composition, sedimentation and immunolocalization as generated.

4 DISCUSSION

A combination of sub-cellular, protein, and peptide fractionation was used along with high throughput tandem mass spectrometry for a comprehensive characterization of the mitochondrial proteome of PF *T. brucei*. Analysis of the acquired data confirmed the validity of the very large majority of predicted genes in *T. brucei* and identified several genes that were previously not predicted. This analysis also showed that the set of genes annotated as hypothetical (species specific) is over-predicted and that virtually all genes annotated as hypothetical, unlikely are not expressed. Overall, the proteomic analysis extrapolated to a total of 1,008 mitochondrial proteins. Of these, specific assignments were made with progressively diminishing stringent criteria for 401, 196, and 283 proteins. The balance was estimated by statistical methods but mitochondrial assignments could not be made to individual proteins. The data have been placed in a publicly available web-based database that we constructed. Analyses of the data reveal a complex and divergent gene expression system, divergent energy production machinery, and a less divergent metabolic system.

The study generated high quality data on a substantial proportion (more than third) of the cellular proteome and was useful in assessing the initial genome annotation. Further studies are required for complete coverage of *T. brucei* cellular proteome that would analyze other

sub-cellular fractions such as cytoplasm and nucleus etc., and also BF stage of the parasite. The use of stringent cut-off criteria and statistical approaches for peptide and protein assignments [16,17] and identifying the large proportion proteins with two or more peptides allowed protein assignment at high confidence. The error rate was <1% and may have missed some proteins for which mass spectrometry data was obtained. For example, two proteins with similarities to VSGs, Tb11.24.0011 and Tb11.43.0001, were identified, each with a single different peptide match, and thus either these proteins are expressed in PF, which is not expected for true VSGs, or these assignments are incorrect. Protein assignments were based on assessment of the gene prediction models by comparison to the predicted protein database and six-frame STOP-to-STOP translated genome sequence database of ≥ 50 amino acids length. It resulted in identification of protein coding ORFs that were not originally predicted which is similar to the experience with *Plasmodium falciparum* [24]. The occurrences of homologs of the newly identified ORFs in *L. major* and/or *T. cruzi* genomes enhance the confidence in gene identification. While a large proportion of the newly identified genes were in the typical size range some were small, for example three encode 49 (Tb11.11283), 55 (Tb02.7762) and 57 (Tb11.29019) amino acids (Supplementary Figure 3). This predicts that the whole genome encodes additional smaller ORFs that have not yet been annotated. The proteome data also allowed reassignment of start codons for some genes.

No proteins that are annotated as hypothetical, unlikely were detected suggesting that these are not functional genes or minimally not expressed in the stages examined. Less than 1% of the proteins identified were predicted from genes that were annotated as hypothetical (brucei specific) group while protein coverage for genes annotated as hypothetical, conserved was very high. Thus most genes annotated as hypothetical, conserved are expressed in PF *T. brucei* while only a small fraction of genes annotated as species specific are expressed. None of the 18 kinetoplast encoded proteins were identified, possibly due to the characteristics of these proteins which include high hydrophobicity, membrane association, non migration into SDS-PAGE gels and few potential trypsin cleavage sites making their peptide identification by routine mass spectrometry analysis difficult.

The large complex tubular mitochondrion ruptures and reseals upon cell breakage and thus isolation by the established protocol that was used [15,25] may result in mitochondrial vesicle preparations that contain soluble abundant non-mitochondrial proteins, and organelles which have similar physical characteristics to the mitochondrial vesicles. Indeed, glycosome preparations have been reported to contain a substantial proportion of mitochondrial proteins [11,13] and in this study glycosomes were commonly present in purified mitochondrial preparations. Proteins from other compartments of cell such as cytoplasm and nucleus which were identified via their predicted functions were diminished in mitochondria compared to whole cells as determined by peptide frequency (Supplementary Table 3C). The enrichment of mitochondrial and glycosomal proteins is more evident in the glycerol gradient fractions compared to the total mitochondrial lysates. In general, peptide coverage of mitochondrial proteins was greater in mitochondrial fraction and lower for cytoplasmic and nuclear proteins when compared to whole cell fraction analyzed under similar conditions.

The shotgun proteomic experiments described here do not provide direct experimental evidence for localization of proteins to specific organelle; however, it provided candidate proteins for downstream analyses. The results from glycerol gradient fraction analysis provided a large set of candidate proteins for further characterization such as identification of associating complexes by TAP-tag analyses that is being carried out in our laboratory in a high throughput manner. In recent studies we have isolated and determined the composition of several multi-protein complexes from mitochondria [20,21,23]. We considered any proteins specifically associated with a mitochondrial complex or mitochondrial process would be mitochondrial, e.g. core sub-units of respiratory complex I and its associated protein components would be

mitochondrial. Using this criterion in our analyses we extended the assignment of proteins to mitochondrion for a large set of proteins with known function beyond currently available GO annotation (Table 1) and also to a large set of proteins with unknown function (Table 2).

The mitochondrial proteome of *T. brucei* is estimated to have just over one thousand proteins, and 880 proteins were assigned to the mitochondrion at varying levels of confidence (Table 1 and 2; and Supplementary Table 5 and 6). This coverage correlates well with the proteomic analyses of yeast mitochondria in which 851 proteins were identified in the purified organelles and detected 84% of the confirmed mitochondrial proteins [18,26]. Thus the Trypanosomes mitochondrial proteome is larger than that of yeast. This might reflect additional biological processes such as RNA editing and metabolic complexities associated with the life cycles. The protein import process may also be more complex as seen by our analysis where mitochondrial signal prediction tools designed for yeast and human have limited capabilities in *T. brucei* (Figure 3). It is possible multiple mechanisms exist for import of proteins to Trypanosome mitochondria. The high confidence protein dataset provided in this study (Tables 1 and 2) can be used as a reference for creating new or modifying existing tools for prediction of mitochondrial import in this group of parasites.

Trypanosome mitochondrion contains complex machinery for energy metabolism through oxidative phosphorylation, iron-sulfur clusters and network of metabolic processes. The proteomic analyses identified many of the core components of respiratory complexes I-V. The presence of functional complex I in *T. brucei* is controversial; however, recently we reported the ability to purify a complex by mAb affinity and TAP-tag purification that contains proteins with homology to those in complex I including some of the core sub-units [21]. This study in addition identified other potential core components of complex I including NDUFA5 (Tb10.70.3150), NDUFA9 (Tb09.244.2620 and Tb10.05.0070), NDUFA13 (Tb11.01.0640), and NDUFS7 (Tb11.47.0017) in mitochondrial enriched fraction. Further NDUFB9 (Tb11.01.7460) have been identified by mass spectrometry in a mitochondrial membrane enriched fraction (Nathalie Acestor, manuscript in preparation). Thus we were able to identify peptides corresponding to all of the core sub-units of complex I that could be identified by homology search against human/bovine protein database [21]. Interestingly, we identified another proteins Tb927.4.1130 in this study which has domain related to the LYR family complex I protein indicating that it may be part of this complex. Complex I also contains numerous proteins of unknown function [21], illustrating its divergence in Trypanosomes. We have not detected peptides from the core components of complex I in the BF cell proteome (unpublished results), suggesting that either they are absent in BF or present at much lower level compared to PF cells. Alternatively, the TAO had higher peptide coverage in BF cells compared to PF cells (unpublished results) as did the Tb10.6k15.0550 protein which is currently annotated as hypothetical but has an alternative oxidase (AOX) domain (termed as TbAOX2 in [27]), suggesting a role in energy metabolism. Numerous enzymes associated with various metabolic processes as well as components associated with amino acid metabolism, protein biosynthesis and turnover, and lipid metabolism were identified in trypanosome mitochondria (Table 1). The functional assignment for most of these proteins has been possible based on sequence conservation with other organisms (www.genedb.org).

The mitochondrial DNA network termed kDNA is composed of few dozen maxicircles and few thousand minicircles. The protein components associated with kDNA replication such as polymerases and topoisomerase II were identified, as is RNA polymerase associated with transcription [28,29]. The mt RNAs are post-transcriptionally edited by RNA editing, a process that has been well characterized [30]. Three different endonucleases have been identified that are required for editing process [31]. However, how the poly-cistronic transcripts are processed prior to editing is not known, and it is anticipated at least one endonuclease will be required for cleavage of such RNAs. In this study numerous proteins of unknown function that have

motifs indicating association with nucleic acids and candidate nucleases were identified, some of which may play role in above processes.

In summary, 1,548 proteins were identified in *T. brucei* mitochondrial enriched fraction, of which 1,008 are estimated to be mitochondrial with 880 specific proteins assigned to the mitochondrion at varying level of confidence. Of the 880 proteins 401 were assigned to the mitochondrion at high confidence, and additional studies are required for definitive sub-cellular assignment for the remaining proteins. Importantly, the results showed that a large proportion of proteins assigned to the mitochondrion have no known function. This is consistent with our recent data on the compositions of multi-protein mitochondrial complexes which show a large fraction of the components in mitochondrial complexes are unique to trypanosomes [21,23]. This reflects the complexity, and structural and functional divergence of multi-protein complexes and biological processes in trypanosome mitochondria. A web-based database www.TrypsProteome.org has been created for dissemination of proteomic data from this and ongoing proteomic studies in our laboratory.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Eric Deutsch for help with SBEAMS Database and Christiane Hertz-Fowler for providing six-frame translated genome database. *T. brucei* protein sequence was downloaded from GeneDB (www.genedb.org) site. Research conducted using equipment made possible by Economic Development Administration - US Department of Commerce and the M.J. Murdock Charitable Trust. This work was supported by NIH grant AI065935.

Abbreviations

PF	procyclic form
BF	blood stream form
mt	mitochondrion
2D-LC-MS/MS	two dimensional liquid chromatography tandem mass spectrometry
GO	Gene Ontology

References

1. Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, et al. The genome of the African trypanosome, *Trypanosoma brucei*. *Science* 2005;309:416–422. [PubMed: 16020726]
2. El-Sayed NMA, Myler PJ, Bartholomeu D, Nilsson D, et al. The genome sequence of *Trypanosoma cruzi*, etiological agent of Chagas' disease. *Science* 2005;309(5733):409–415. [PubMed: 16020725]
3. Ivens AC, Peacock CS, Worthey EA, Murphy L, et al. The genome of the kinetoplastid parasite, *Leishmania major*. *Science* 2005;309(5733):436–442. [PubMed: 16020728]
4. El-Sayed NMA, Myler PJ, Blandin G, Berriman M, et al. Comparative genomics of trypanosomatid parasitic protozoa. *Science* 2005;309(5733):404–409. [PubMed: 16020724]
5. Martinez-Calvillo S, Yan S, Nguyen D, Fox M, et al. Transcription of *Leishmania major* Friedlin chromosome 1 initiates in both directions within a single region. *Mol Cell* 2003;11(5):1291–1299. [PubMed: 12769852]
6. Parsons M, Hill T. Elevated phosphoglycerate kinase mRNA but not protein in monomorphic *Trypanosoma brucei*: Implications for stage- regulation and post-transcriptional control. *Mol Biochem Parasitol* 1989;33:215–228. [PubMed: 2704387]

7. Gale M Jr, Carter V, Parsons M. Translational control mediates the developmental regulation of the *Trypanosoma brucei* Nrk protein kinase. *J Biol Chem* 1994;269 (50):31659–31665. [PubMed: 7989338]
8. Atwood JA III, Weatherly DB, Minning TA, Bundy B, et al. The *Trypanosoma cruzi* proteome. *Science* 2005;309(5733):473–476. [PubMed: 16020736]
9. Jones A, Faldas A, Foucher A, Hunt E, et al. Visualisation and analysis of proteomic data from the procyclic form of *Trypanosoma brucei*. *Proteomics* 2006;6(1):259–267. [PubMed: 16302277]
10. Broadhead R, Dawe HR, Farr H, Griffiths S, et al. Flagellar motility is required for the viability of the bloodstream trypanosome. *Nature* 2006;440(7081):224–227. [PubMed: 16525475]
11. Colasante C, Ellis M, Ruppert T, Voncken F. Comparative proteomics of glycosomes from bloodstream form and procyclic culture form *Trypanosoma brucei brucei*. *Proteomics* 2006;6 (11): 3275–3293. [PubMed: 16622829]
12. Rosenzweig D, Smith D, Opperdoes F, Stern S, et al. Retooling Leishmania metabolism: from sand fly gut to human macrophage. *FASEB J* 2008;22(2):590–602. [PubMed: 17884972]
13. Vertommen D, Van RJ, Szikora JP, Rider MH, et al. Differential expression of glycosomal and mitochondrial proteins in the two major life-cycle stages of *Trypanosoma brucei*. *Mol Biochem Parasitol* 2008;158(2):189–201. [PubMed: 18242729]
14. Bridges DJ, Pitt AR, Hanrahan O, Brennan K, et al. Characterisation of the plasma membrane subproteome of bloodstream form *Trypanosoma brucei*. *Proteomics* 2008;8(1):83–99. [PubMed: 18095354]
15. Panigrahi, AK.; Schnauffer, A.; Stuart, KD. *Methods in Enzymology*. Gott, JM., editor. Elsevier Inc; 2007. p. 3-24.
16. Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* 2002;74 (20):5383–5392. [PubMed: 12403597]
17. Nesvizhskii AI, Keller A, Kolker E, Aebersold R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* 2003;75 (17):4646–4658. [PubMed: 14632076]
18. Reinders J, Zahedi RP, Pfanner N, Meisinger C, et al. Toward the complete yeast mitochondrial proteome: multidimensional separation techniques for mitochondrial proteomics. *J Proteome Res* 2006;5(7):1543–1554. [PubMed: 16823961]
19. Elias JE, Haas W, Faherty BK, Gygi SP. Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nat Methods* 2005;2 (9):667–675. [PubMed: 16118637]
20. Panigrahi AK, Ernst NL, Domingo GJ, Fleck M, et al. Compositionally and Functionally Distinct Editosomes in *Trypanosoma brucei*. *RNA* 2006;12(6):1038–1049. [PubMed: 16611942]
21. Panigrahi AK, Zikova A, Dalley RA, Acestor N, et al. Mitochondrial complexes in *Trypanosoma brucei*: a novel complex and a unique oxidoreductase complex. *Mol Cell Proteomics* 2008;7(3):534–545. [PubMed: 18073385]
22. Wang Z, Englund PT. RNA interference of a trypanosome topoisomerase II causes progressive loss of mitochondrial DNA. *EMBO J* 2001;20 (17):4674–4683. [PubMed: 11532932]
23. Zikova A, Panigrahi AK, Dalley RA, Acestor N, et al. *Trypanosoma brucei* mitochondrial ribosomes: affinity purification and component identification by mass spectrometry. *Mol Cell Proteomics* 2008;7.7:1286–1296. [PubMed: 18364347]
24. Lasonder E, Ishihama Y, Andersen J, Vermunt A, et al. Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry. *Nature* 2002;419:537–542. [PubMed: 12368870]
25. Harris ME, Moore DR, Hajduk SL. Addition of uridines to edited RNAs in trypanosome mitochondria occurs independently of transcription. *J Biol Chem* 1990;265:11368–11376. [PubMed: 1694178]
26. Andreoli C, Prokisch H, Hortnagel K, Mueller JC, et al. MitoP2, an integrated database on mitochondrial proteins in yeast and man. *Nucleic Acids Res* 2004;32(Database issue):D459–D462. [PubMed: 14681457]
27. Chaudhuri M, Ott RD, Hill GC. Trypanosome alternative oxidase: from molecule to function. *Trends Parasitol* 2006;22 (10):484–491. [PubMed: 16920028]
28. Liu B, Liu Y, Motyka SA, Agbo EC, et al. Fellowship of the rings: the replication of kinetoplast DNA. *Trends Parasitol* 2005;21(8):363–369. [PubMed: 15967722]

29. Lukes J, Hashimi H, Zikova A. Unexplained complexity of the mitochondrial genome and transcriptome in kinetoplastid flagellates. *Curr Genet* 2005;48 (5):277–299. [PubMed: 16215758]
30. Stuart KD, Schnauffer A, Ernst NL, Panigrahi AK. Complex management: RNA editing in trypanosomes. *Trends Biochem Sci* 2005;30 (2):97–105. [PubMed: 15691655]
31. Carnes J, Trotter JR, Peltan A, Fleck M, et al. RNA editing in *Trypanosoma brucei* requires three different editosomes. *Mol Cell Biol* 2008;28(1):122–130. [PubMed: 17954557]

Figure 1. Summary of protein identification from cellular and sub-cellular fractions. Number of proteins identified in detergent soluble (supernatant) and insoluble fractions (pellet) from whole cell and mitochondrial enriched fraction and comparison between them are shown.

Figure 2.

Comparison of proportions of protein groups annotated in GeneDB database to that identified by proteomics. Percent of proteins with known or assigned function are denoted by filled line, hypothetical conserved by gray, hypothetical (species specific) by black, and hypothetical, unlikely by white areas in the pie-chart.

Figure 3.

Assessing the ability of available software for prediction of mitochondrial proteins in *T. brucei*. Schematics showing representative scores obtained for presence of mitochondrial import signal in a group of (A) known mitochondrial proteins and (B) non-mitochondrial proteins using Mitoprot (open bar) and SignalP (gray bar) programs.

Table 1

Proteins with known or putative function that were identified in mitochondrial enriched fraction and are assigned to mitochondrion. They are grouped based on their association with demonstrated/predicted biological processes. The proteins were identified by indicated number of peptides (number of unique peptides in parenthesis) in whole cell (PF) or mitochondrial (Mt) enriched fraction. The proteins that were also identified in the analyzed glycerol gradient fractions (GG) are indicated (Y). Blank columns indicate the proteins were not identified by mass spectrometry in the analyzed samples.

Protein_ID	GeneDB_Annotation	PF	Mt	GG
Mitochondrial Respiration				
Complex I				
Tb10.70.5510	adrenodoxin reductase, putative		8(8)	
Tb10.389.1140	electron transfer protein, NDUFS1	3(3)	5(5)	Y
Tb927.3.860	acyl carrier protein, NDUFAB1	3(2)	4(3)	Y
Tb927.5.3350	iron superoxide dismutase, putative	3(3)	7(7)	Y
Tb09.160.5260	oxidoreductase, putative	2(2)	11(9)	Y
Tb927.7.7410	oxidoreductase, putative	1(1)	14(13)	Y
Tb11.01.7480	superoxide dismutase, putative		3(3)	Y
Tb11.47.0017	NADH-ubiquinone oxidoreductase 20 kDa subunit, NDUFS7		1(1)	Y
Tb11.01.0640	NADH-ubiquinone oxidoreductase complex I subunit, NDUFA13		1(1)	Y
Tb11.01.8630	NADH-ubiquinone oxidoreductase complex I subunit, NDUFA2		1(1)	Y
Tb10.70.3150	NADH-ubiquinone oxidoreductase complex I subunit, NDUFA5		5(5)	Y
Tb10.61.1790	NADH-ubiquinone oxidoreductase complex I subunit, NDUFA6	2(2)	19(19)	Y
Tb10.05.0070	NADH-ubiquinone oxidoreductase complex I subunit, NDUFA9		9(9)	Y
Tb09.244.2620	NADH-ubiquinone oxidoreductase complex I subunit, NDUFA9		5(5)	
Tb927.5.450	NADH-ubiquinone oxidoreductase, NDUFV1	1(1)	9(8)	Y
Tb927.7.6350	NADH-ubiquinone oxidoreductase, NDUFV2		4(4)	
Tb09.211.0330	chaperone protein DNAJ, putative	4(4)	5(5)	Y
Complex II				
Tb927.8.6580	succinate dehydrogenase flavoprotein, putative	6(6)	22(20)	Y
Tb09.160.4380	succinate dehydrogenase, putative	2(2)	6(6)	Y
Complex III				
Tb927.8.1890	cytochrome c1, heme protein, mitochondrial precursor	5(4)	9(8)	Y
Tb09.211.4700	RISP reiske iron-sulfur protein, mitochondrial precursor	4(4)	13(13)	Y
Complex IV				
Tb10.6k15.2180	COX IX cytochrome c oxidase subunit IX		2(2)	
Tb09.160.1820	COX V cytochrome c oxidase subunit V, putative	2(2)	8(7)	Y
Tb927.3.1410	COX VII cytochrome c oxidase VII, putative	1(1)	7(7)	Y
Tb927.4.4620	COX VIII cytochrome c oxidase VIII, putative	1(1)	3(3)	
Tb10.100.0160	COXVI cytochrome C oxidase subunit VI, putative	1(1)	6(6)	Y
Tb11.01.4702	cytochrome c oxidase subunit 10, putative	2(2)	6(5)	Y
Tb927.1.4100	cytochrome C oxidase subunit IV	6(6)	13(11)	Y

Protein_ID	GeneDB_Annotation	PF	Mt	GG
Complex V				
Tb927.7.7420	ATP synthase alpha chain, mitochondrial precursor	19(15)	29(23)	Y
Tb927.3.1380	ATP synthase beta chain, mitochondrial precursor	17(13)	33(21)	Y
Tb10.100.0070	ATP synthase F1 subunit gamma protein, putative	9(9)	18(16)	Y
Tb927.6.4990	ATP synthase, epsilon chain, putative	4(4)	8(8)	Y
Tb927.5.1710	ribonucleoprotein p18, mitochondrial precursor, putative	3(2)	6(5)	Y
Others in aerobic respiration				
Tb927.8.3580	ATP-dependent chaperone, putative		5(5)	Y
Tb10.70.4380	cytochrome C oxidase assembly protein, putative		4(3)	Y
Electron transport				
Tb10.6k15.0960	54 NDH2 NADH dehydrogenase	5(5)	13(12)	Y
Tb10.6k15.3640	AOX TAO alternative oxidase	3(3)	12(12)	Y
Tb10.70.7840	choline dehydrogenase, putative	2(2)	10(10)	Y
Tb927.8.5120	cytochrome c	2(2)	2(2)	Y
Tb927.1.1580	cytochrome c oxidase assembly factor, putative		11(11)	Y
Tb11.03.0950	electron transfer flavoprotein, putative	3(3)	8(6)	Y
Tb927.8.1240	electron transfer flavoprotein-ubiquinone oxidoreductase, putative		7(7)	
Tb11.02.5530	electron-transfer-flavoprotein, alpha polypeptide, putative	2(2)	6(5)	
Tb11.47.0012	glutaredoxin, putative		1(1)	
Tb11.55.0026	isovaleryl-coA dehydrogenase, putative	1(1)	8(7)	Y
Tb10.406.0490	monothiol glutaredoxin, putative	1(1)	2(2)	Y
Tb11.02.4800	NADH-cytochrome b5 reductase-like protein		3(3)	
Tb927.8.1990	TRYP2 tryparedoxin peroxidase	6(6)	13(12)	Y
mRNA Editing				
Tb09.160.2970	KREL1 mitochondrial RNA editing ligase 1		8(8)	Y
Tb927.1.3030	KREL2, REL2 RNA editing ligase 2	1(1)	9(9)	Y
Tb927.1.1690	KREN1, MP90 RNA-editing complex protein		3(3)	Y
Tb10.70.1580	KREN2, MP61 RNA-editing complex protein		2(2)	
Tb10.70.1750	KREN3, MP67 RNA-editing complex protein		2(2)	
Tb927.2.2470	KREPA1 MP81, RNA-editing complex protein		11(11)	Y
Tb10.6k15.2310	KREPA2, MP63 RNA-editing complex protein	1(1)	9(8)	Y
Tb927.8.620	KREPA3, RNA-editing complex protein MP42		3(2)	Y
Tb927.8.680	KREPA5, MP19 RNA-editing complex protein		1(1)	
Tb10.70.2090	KREPA6, MP18 RNA-editing complex protein		2(2)	Y
Tb11.02.0490	KREPB4, MP46 RNA-editing complex protein		6(6)	Y
Tb11.03.0180	KREPB5, MP44 RNA-editing complex protein		5(4)	Y
Tb09.160.4130	KREPB7, RNA-editing complex protein		5(4)	
Tb927.8.5690	KREPB8, MP41 RNA editing complex protein		1(1)	
Tb927.7.1550	KRET2, TbMP57 RET2 RNA-editing complex protein		5(5)	Y
Tb927.7.1070	KREX1, KREPC1 RNA-editing complex protein		7(7)	Y

Protein_ID	GeneDB_Annotation	PF	Mt	GG
Tb10.70.3850	KREX2, MP99 RNA-editing complex protein	4(4)	18(18)	Y
Tb11.01.0610	KREH1, MHEL61 DEAD box protein	6(6)	18(16)	Y
Nucleic acid metabolic process				
Tb927.7.3950	KRET1, TUTase RNA-editing complex protein		7(7)	Y
Tb11.02.5770	RBP16 mitochondrial RNA binding protein	3(3)	6(5)	Y
Tb11.55.0009	GBP21 mitochondrial RNA binding protein 1	1(1)	11(10)	Y
Tb11.01.4860	guide RNA binding protein	6(5)	9(8)	Y
Tb927.8.2740	TbRBP38 mitochondrial RNA binding protein	4(4)	19(15)	Y
Tb10.406.0050	RNA-binding protein, putative		3(3)	Y
Tb927.6.1680	zinc finger protein, putative		7(7)	
Tb927.8.3830	ATP-dependent DEAD/H RNA helicase, putative		4(4)	
Tb927.8.4040	endonuclease G, putative		1(1)	
Tb11.01.7110	methionyl-trna formyltransferase		1(1)	
Tb09.160.5280	mitochondrial exoribonuclease DSS-1		1(1)	
Tb10.100.0220	mitochondrial structure specific endonuclease I (SSE-1), putative		1(1)	
Tb11.02.3530	MTRNAP mitochondrial DNA-directed RNA polymerase		10(10)	
Tb11.01.7300	poly(A) polymerase, putative		3(3)	
Tb10.70.3680	rRNA dimethyltransferase, putative		3(3)	
Tb11.01.8090	kinetoplast DNA-associated protein, putative		2(2)	
Tb09.160.4090	TOP2 DNA topoisomerase II		13(13)	
DNA replication/packaging				
Tb927.5.2780	mitochondrial DNA polymerase beta	3(3)	16(16)	Y
Tb11.02.2300	POLIB mitochondrial DNA polymerase I protein B, putative		5(5)	
Tb11.02.0770	POLID mitochondrial DNA polymerase I protein D, putative	1(1)	11(11)	Y
Tb927.7.3990	TbPOL1C mitochondrial DNA polymerase I protein C		5(5)	Y
Tb10.6k15.1460	kinetoplast DNA-associated protein, putative		2(2)	
Tb10.6k15.1600	kinetoplast DNA-associated protein, putative		1(1)	
Tb927.8.7260	kinetoplast-associated protein, putative		3(3)	Y
Amino acid metabolic process				
Tb927.8.6060	2-amino-3-ketobutyrate coenzyme A ligase, putative	15(13)	23(18)	Y
Tb11.01.1440	aminomethyltransferase, putative	1(1)	9(9)	Y
Tb11.02.2740	aspartate aminotransferase, mitochondrial	4(3)	9(8)	Y
Tb927.2.4590	branched-chain amino acid aminotransferase, putative	1(1)	5(3)	Y
Tb927.2.4610	branched-chain amino acid aminotransferase, putative		9(8)	
Tb09.160.4310	GDH glutamate dehydrogenase	17(16)	43(40)	Y
Tb09.211.1380	glycine cleavage system H protein, putative		4(4)	Y
Tb927.6.2790	L-threonine 3-dehydrogenase, putative	16(13)	31(19)	Y
Protein translation				
Tb927.4.1070	50S ribosomal protein L13, putative		2(2)	
Tb927.7.3960	50S ribosomal protein L16, putative	1(1)	2(2)	Y

Protein_ID	GeneDB_Annotation	PF	Mt	GG
Tb927.8.5860	50S ribosomal protein L17, putative		4(4)	
Tb927.5.3360	50S ribosomal protein L2, putative		4(4)	
Tb927.2.4890	ribosomal protein L11, putative		5(5)	
Tb927.3.5610	ribosomal protein L3 mitochondrial, putative		7(6)	Y
Tb09.211.2580	30S Ribosomal protein S17, putative		5(5)	Y
Tb10.389.0130	30S ribosomal protein S8, putative		3(3)	
Tb10.70.2190	Mitochondrial elongation factor G, putative	6(6)	14(14)	Y
Protein folding/modification/transport				
Tb927.6.3740	heat shock 70 kDa protein, mitochondrial precursor, putative	22(19)	48(43)	Y
Tb11.02.0250	heat shock protein, mitochondrial precursor, putative	16(14)	42(35)	Y
Tb927.7.1340	HSP10 10 kDa heat shock protein, putative	4(4)	6(6)	Y
Tb10.70.0430	HSP60 chaperonin Hsp60, mitochondrial precursor	25(20)	63(46)	Y
Tb11.01.6680	HSP60 chaperonin Hsp60, mitochondrial precursor, putative	2(2)	22(19)	Y
Tb11.01.1160	lipoate-protein ligase, putative		1(1)	
Tb927.8.630	lipoate-protein ligase, putative		3(3)	
Tb11.01.4870	inner membrane preprotein translocase Tim17, putative	2(1)	1(1)	Y
Protein catabolic process				
Tb927.2.3030	ATP-dependent Clp protease subunit, heat shock protein 78 (HSP78)	12(10)	30(28)	Y
Tb11.01.2000	hslVU complex proteolytic subunit, putative	3(3)	10(10)	Y
Tb10.6k15.3030	mitochondrial ATP-dependent zinc metallopeptidase, putative		8(8)	Y
Tb927.4.3300	mitochondrial ATP-dependent zinc metallopeptidase, putative	2(2)	16(15)	Y
Tb10.70.2630	mitochondrial inner membrane signal peptidase, putative		1(1)	
Tb10.6k15.0490	mitochondrial intermediate peptidase, putative	5(5)	9(9)	Y
Tb11.02.1480	mitochondrial processing peptidase alpha subunit, putative	10(10)	21(20)	Y
Tb927.2.4110	mitochondrial processing peptidase alpha subunit, putative	2(2)	10(9)	Y
Tb927.5.1060	mitochondrial processing peptidase, beta subunit, putative	13(11)	22(19)	Y
Tb09.160.3110	mitochondrial processing peptide beta subunit, putative	5(5)	15(15)	Y
Tb927.8.1860	pitrilysin-like metalloprotease	6(6)	30(28)	Y
Lipid metabolic process				
Tb927.8.1060	malonyl-CoA decarboxylase, mitochondrial precursor, putative		3(3)	
Tb927.6.4540	3-hydroxy-3-methylglutaryl-CoA reductase, putative	5(5)	13(11)	Y
Tb10.61.1530	lipoic acid synthetase, mitochondrial precursor, putative		5(5)	
Regulation				
Tb927.8.4810	prohibitin	9(8)	15(13)	Y
Tb09.211.4511	kinetoplastid membrane protein KMP-11	3(3)	2(2)	Y
Tb927.8.7170	inositol polyphosphate 1-phosphatase, putative		10(9)	Y
Various Metabolic processes				
Tb10.70.7550	2-oxoisovalerate dehydrogenase alpha subunit, putative	2(2)	16(16)	Y
Tb927.3.5630	3,2-trans-enoyl-CoA isomerase, mitochondrial precursor, putative		2(2)	Y

Protein_ID	GeneDB_Annotation	PF	Mt	GG
Tb927.4.4910	3,2-trans-enoyl-CoA isomerase, mitochondrial precursor, putative	3(2)	5(4)	Y
Tb927.8.7530	3,2-trans-enoyl-CoA isomerase, mitochondrial precursor, putative	4(4)	10(9)	Y
Tb927.8.2540	3-ketoacyl-CoA thiolase, putative	5(5)	14(11)	Y
Tb927.8.1420	acyl-CoA dehydrogenase, mitochondrial precursor, putative	1(1)	4(4)	Y
Tb10.70.4280	delta-1-pyrroline-5-carboxylate dehydrogenase, putative	9(9)	27(21)	Y
Tb11.01.8200	enoyl-CoA hydratase/isomerase family protein, putative	3(3)	16(13)	Y
Tb09.160.4300	farnesyl pyrophosphate synthetase, putative	2(2)	10(9)	Y
Tb11.02.2700	fumarate hydratase class I, putative	5(5)	26(23)	Y
Tb11.02.5280	glycerol-3-phosphate dehydrogenase, putative	13(11)	29(23)	Y
Tb11.02.0440	glycerolphosphate mutase, putative		3(3)	
Tb927.8.3690	IDH isocitrate dehydrogenase [NADP], mitochondrial precursor, putative	10(9)	23(20)	Y
Tb11.02.2070	long-chain-fatty acid-CoA ligase protein, putative		6(6)	
Tb11.02.3130	malic enzyme, putative	8(8)	22(18)	Y
Tb10.70.3280	methylglutaconyl-CoA hydratase, mitochondrial precursor, putative	1(1)	7(6)	Y
Tb10.70.5110	mMDH mitochondrial malate dehydrogenase	10(9)	14(14)	Y
Tb11.01.7290	NUDIX hydrolase, conserved		9(8)	Y
Tb10.70.1410	PP2C protein phosphatase 2C, putative		5(5)	Y
Tb11.01.1130	prenyltransferase, putative		7(7)	
Tb11.02.4890	pterin-4-alpha-carbinolamine dehydratase, putative		2(2)	
Tb11.02.2390	pyruvate dehydrogenase (lipoamide) kinase, putative	2(2)	7(7)	Y
Tb10.6k15.3250	succinyl-CoA ligase [GDP-forming] beta-chain, putative	7(7)	25(20)	Y
Tb927.3.2230	succinyl-CoA synthetase alpha subunit, putative	5(5)	15(12)	Y
Tb11.02.0290	succinyl-coA:3-ketoacid-coenzyme A transferase, putative	6(5)	20(16)	Y
Tb11.01.5180	ubiquinone biosynthesis protein-like protein		8(7)	
Tb11.02.4480	3-methylcrotonoyl-CoA carboxylase beta subunit, putative	4(4)	15(12)	Y
Tb927.8.6970	3-methylcrotonyl-CoA carboxylase, putative		10(9)	Y
Tb11.01.1740	2-oxoglutarate dehydrogenase E1 component, putative	24(24)	56(49)	Y
Tb11.47.0004	2-oxoglutarate dehydrogenase subunit, putative	21(21)	45(39)	Y
Tb11.01.3550	2-oxoglutarate dehydrogenase, E2 component	6(6)	20(16)	Y
Tb11.01.8470	dihydrolipoyl dehydrogenase	12(11)	27(24)	Y
Tb10.6k15.3080	dihydrolipoamide acetyltransferase precursor, putative	8(7)	17(15)	
Tb10.70.5380	dihydrolipoamide acetyltransferase, putative	3(3)	11(9)	Y
Tb927.3.1790	pyruvate dehydrogenase E1 beta subunit, putative	11(10)	19(14)	Y
Tb10.389.0890	pyruvate dehydrogenase E1 component alpha subunit, putative	7(6)	30(20)	Y
Transport				
Tb09.211.1750	mitochondrial carrier protein, putative	10(9)	13(12)	Y
Tb09.211.2370	mitochondrial carrier protein, putative		1(1)	
Tb09.211.3200	mitochondrial carrier protein, putative		3(3)	Y
Tb10.389.0340	mitochondrial carrier protein, putative		3(3)	

Protein_ID	GeneDB_Annotation	PF	Mt	GG
Tb10.389.0690	mitochondrial carrier protein, putative	4(4)	10(9)	Y
Tb10.61.0610	mitochondrial carrier protein, putative	3(3)	7(7)	Y
Tb10.61.1820	mitochondrial carrier protein, putative	11(8)	18(11)	Y
Tb10.61.2510	mitochondrial carrier protein, putative		2(2)	
Tb10.70.2290	mitochondrial carrier protein, putative		2(2)	
Tb11.01.5040	mitochondrial carrier protein, putative		2(1)	
Tb11.01.5960	mitochondrial carrier protein, putative		7(6)	
Tb11.02.2960	mitochondrial carrier protein, putative	3(2)	10(8)	Y
Tb11.03.0870	mitochondrial carrier protein, putative		5(5)	
Tb927.2.2970	mitochondrial carrier protein, putative	5(5)	7(7)	Y
Tb927.3.2980	mitochondrial carrier protein, putative		2(2)	
Tb927.4.1660	mitochondrial carrier protein, putative		6(6)	
Tb927.5.1550	mitochondrial carrier protein, putative		2(2)	
Tb927.7.3940	mitochondrial carrier protein, putative	1(1)	4(4)	
Tb927.8.3330	mitochondrial carrier protein, putative		2(2)	
Tb927.8.4440	mitochondrial carrier protein, putative		1(1)	
Tb10.61.0090	potassium voltage-gated channel, putative		6(6)	Y
Miscellaneous				
Tb927.3.1000	frataxin-like, mitochondrial precursor, putative		2(2)	
Tb927.7.7080	mitochondrial glycoprotein-like protein	1(1)	5(5)	
Tb927.6.2420	p22 protein precursor	4(4)	12(10)	Y
Tb927.7.3470	p22 protein precursor, putative		5(5)	Y
Tb10.26.0860	HIRA-interacting protein 5, putative		2(2)	Y

Table 2

Proteins currently annotated as hypothetical, conserved that were assigned to mitochondrion. These proteins were identified in mitochondrial enriched fraction (Mt) by indicated number of peptides (number of unique peptides in parenthesis) and some of them as indicated by number of peptides were identified in whole cell sample (PF). Proteins identified in glycerol gradient sample (GG) are indicated (Y). These proteins are assigned to mitochondrion based on their specific association with mitochondrial complexes. Predictable motifs and domains in these proteins are also shown.

Protein_ID	Domains/motifs	PF	Mt	GG	Protein_ID	Domains/motifs	PF	Mt	GG
Respiratory Complex I									
Tb09.160.0760			2(2)		Tb927.2.4380			3(3)	Y
Tb09.244.2670			5(5)		Tb927.3.3660			3(3)	Y
Tb10.6k15.3040	Vacuolar H ⁺ -ATPase V1 sector, subunit C	1(1)	17(16)	Y	Tb927.6.1410			6(6)	
Tb1.01.5480			2(2)		Tb927.6.2010	Acyl-CoA synthase	2(2)	20(18)	Y
Tb1.02.4810		5(5)	21(19)	Y	Tb927.7.7330		1(1)	5(5)	Y
Tb927.1.730	NB-ARC domain		1(1)	Y	Tb927.8.4250	Tetratricopeptide-like helical		8(7)	
MRB Complex 1									
Tb10.6k15.0150			9(9)	Y	Tb927.4.1500	ATP-dependent RNA helicase	1(1)	23(23)	Y
Tb1.01.8620		1(1)	6(6)	Y	Tb927.4.4160		1(1)	11(11)	Y
Tb1.02.5390			9(9)	Y	Tb927.5.3010	PRK09169, hypothetical protein	3(3)	14(13)	Y
Tb927.2.1860			12(12)	Y	Tb927.7.2570		4(4)	16(16)	Y
Tb927.2.3800		2(2)	12(12)	Y	Tb927.7.800			7(7)	Y
Tb927.2.6070	RBZ Zinc finger, RanBP2-type	1(1)	1(1)	Y	Tb927.8.8170		1(1)	12(11)	Y
Tb927.3.1590	Predicted ATPase of the ABC class	7(7)	4(4)		Tb927.8.8180		2(2)	13(12)	Y
Ribosome large subunit									
Tb09.160.2250		1(1)	1(1)	Y	Tb927.3.820			1(1)	Y
Tb09.160.5240	MRP-L47, Mitochondrial 39-S ribosomal protein L47		3(3)	Y	Tb927.4.1810	Ribosomal protein L33		2(2)	
Tb09.211.0230			2(2)		Tb927.4.2720	SrmB, Superfamily II DNA and RNA helicases		5(5)	
Tb09.211.3800	SpoU rRNA Methylase family		3(3)		Tb927.4.4600	L51_S25_C1-B8, mt ribosomal protein L51/S25/C1-B8 domain	1(1)	7(4)	Y
Tb09.211.4580			2(2)	Y	Tb927.4.4610			6(6)	
Tb10.26.0600			4(4)		Tb927.5.2070			3(3)	Y
Tb10.26.1000			2(2)	Y	Tb927.5.3110			5(4)	Y
Tb10.389.1710		1(1)	6(6)	Y	Tb927.5.3410			5(5)	Y

Protein_ID	Domains/motifs	PF	Mt	GG	Protein_ID	Domains/motifs	PF	Mt	GG
Tb10.6k15.3290		1(1)	4(4)	Y	Tb927.5.3870			1(1)	Y
Tb10.70.5970			5(5)		Tb927.5.3980	COG0200, RplO, Ribosomal protein L15	1(1)	11(10)	Y
Tb10.70.7650		1(1)	12(12)	Y	Tb927.5.4120			6(6)	
Tb10.70.7960	Pentatricopeptide repeat	1(1)	5(5)	Y	Tb927.6.1440			3(3)	Y
Tb11.01.1600			4(3)	Y	Tb927.6.3600	Predicted SAM-dependent methyltransferases		3(3)	
Tb11.01.1810			1(1)		Tb927.6.3930	Tetrapetide repeat domain	2(2)	10(10)	Y
Tb11.01.1840		4(3)	6(6)	Y	Tb927.6.4040			1(1)	
Tb11.01.1930		1(1)	3(3)	Y	Tb927.6.4080			4(3)	Y
Tb11.01.2340			12(9)		Tb927.6.4200			9(9)	
Tb11.01.3500			3(3)	Y	Tb927.7.2760	Ribosomal protein L22/L17e		5(3)	
Tb11.01.6620			9(9)		Tb927.7.2990			6(6)	Y
Tb11.01.7140			2(2)		Tb927.7.3460	Tetrapetide-like helical		3(3)	Y
Tb11.02.1110		2(2)	4(4)		Tb927.7.3510			3(3)	Y
Tb11.02.2250			4(4)		Tb927.7.4140	Ribosomal prokaryotic L21 protein		2(2)	
Tb11.02.3670			6(5)		Tb927.7.4710	MRP_L46, Mitochondrial ribosomal protein L46		3(3)	
Tb11.02.3800	PseudoU_synth_ScPUS7, Pseudouridine synthase	1(1)	9(9)	Y	Tb927.7.6800			8(7)	Y
Tb11.02.3810	rplD, 50S ribosomal protein L4		4(4)		Tb927.7.7010		1(1)	2(2)	Y
Tb11.03.0260			4(4)	Y	Tb927.8.1880			4(4)	Y
Tb11.55.0016			9(9)	Y	Tb927.8.2760	YihA_EngB subfamily, GTPase of unknown function		1(1)	
Tb927.1.1160	Pentatricopeptide repeat		2(2)	Y	Tb927.8.3170	Pentatricopeptide repeat		5(5)	
Tb927.3.1710	rplX, 50S ribosomal protein L24		4(4)	Y	Tb927.8.3300		2(2)	12(11)	
Ribosome small subunit									
Tb09.160.3800			4(4)	Y	Tb927.1.2990	Pentatricopeptide repeat	2(2)	12(12)	Y
Tb09.160.4750		2(2)	25(21)	Y	Tb927.2.4400		3(3)	26(26)	Y
Tb09.211.2430		1(1)	14(13)	Y	Tb927.3.2260	UBQ, Ubiquitin homologues		6(5)	Y
Tb09.211.3000			11(10)	Y	Tb927.3.5240			12(11)	Y
Tb09.211.4510			4(4)	Y	Tb927.3.770		1(1)	4(4)	Y
Tb10.26.0050	PPR repeat		3(3)		Tb927.3.970	NAD_binding_2, NAD binding domain of 6-phosphogluconate dehydrogenase	1(1)	8(8)	Y

Protein_ID	Domains/motifs	PF	Mt	GG	Protein_ID	Domains/motifs	PF	Mt	GG
Tb10.26.0750	IGR protein motif, unknown function		2(2)	Y	Tb927.4.3690	Tcp10_C, T-complex protein 10 C-terminus, Sod_Fe_C, Iron/manganese superoxide dismutases		1(1)	Y
Tb10.406.0510	30S ribosomal protein S11		7(7)	Y	Tb927.5.1510		1(1)	8(8)	Y
Tb10.61.0190			3(3)		Tb927.5.1790			7(6)	Y
Tb10.61.3110			2(2)	Y	Tb927.5.3640		2(2)	4(4)	Y
Tb10.6k15.3900		1(1)	12(10)	Y	Tb927.5.4040	Peptidoglycan-binding LysM		21(20)	Y
Tb10.70.0530	RpsE, Ribosomal protein S5		13(12)	Y	Tb927.6.1250			4(4)	
Tb10.70.3840		1(1)	2(1)	Y	Tb927.6.2080			5(5)	Y
Tb10.70.4220			5(4)	Y	Tb927.6.2180			6(5)	
Tb10.70.4850		1(1)	4(4)	Y	Tb927.6.4560			2(2)	
Tb11.01.1910	Enoyl-CoA hydratase/isomerase family		20(19)	Y	Tb927.6.4580			1(1)	
Tb11.01.2170			4(4)		Tb927.6.4930	RHOD_HSP67B2, Member of the Rhodanese Homology Domain superfamily		7(4)	
Tb11.01.3300		1(1)	6(6)	Y	Tb927.7.2620			4(4)	
Tb11.01.3710			2(2)		Tb927.7.3050		1(1)	15(13)	Y
Tb11.01.7510			6(5)	Y	Tb927.7.3240			3(2)	
Tb11.02.0130		1(1)	18(16)	Y	Tb927.8.1430			3(3)	Y
Tb11.02.2710	Mitochondrial small ribosomal subunit Ram22	1(1)	16(16)	Y	Tb927.8.3110	rps9, ribosomal protein S9		9(9)	
Tb11.02.3180			19(18)	Y	Tb927.8.4550			1(1)	
Tb11.02.5670			1(1)	Y	Tb927.8.4860	PPR repeat		7(7)	Y
Tb11.47.0024		4(4)	22(20)	Y	Tb927.8.5200		1(1)	28(27)	Y
Tb927.1.1200	S15/NS1, RNA-binding		10(10)	Y	Tb927.8.5280			5(5)	Y
Other mt complexes									
Tb09.211.1600			1(1)		Tb927.1.4230			3(3)	
Tb09.211.1900			7(7)	Y	Tb927.2.2940			2(2)	
Tb09.211.4400		1(1)	11(11)	Y	Tb927.2.3610		3(3)	3(3)	Y
Tb09.211.4740	ubiE, ubiquinone/menaquinone biosynthesis methyltransferase	2(2)	11(10)	Y	Tb927.2.5930			5(5)	
Tb09.244.2840			3(3)		Tb927.3.1690		3(3)	9(7)	Y
Tb09.v1.0420		1(1)	5(4)	Y	Tb927.3.1890			7(6)	
Tb10.05.0050			3(3)	Y	Tb927.3.1940	DUF477 (domain of unknown function)	2(2)	9(9)	Y

Protein_ID	Domains/motifs	PF	Mt	GG	Protein_ID	Domains/motifs	PF	Mt	GG
Tb10.389.1785		1(1)	5(4)	Y	Tb927.3.2180		1(1)	2(2)	
Tb10.389.1880	TRX_family, Thioredoxin fold	1(1)	2(2)		Tb927.3.2880		2(2)	5(3)	Y
Tb10.6k15.1300			2(2)		Tb927.3.750	SET (Su(var)3-9, Enhancer-of-zeste, Trithorax) domain; Putative methyl transferase	2(2)	10(10)	Y
Tb10.6k15.2510	ATP synthase delta (OSCP) subunit	10(9)	17(17)	Y	Tb927.4.3450		1(1)	3(3)	
Tb10.70.0625		1(1)	3(3)	Y	Tb927.4.4300			7(7)	
Tb10.70.1520			2(2)		Tb927.4.440			1(1)	
Tb10.70.1640		2(2)	1(1)		Tb927.4.720		1(1)	1(1)	
Tb10.70.1890		1(1)	8(8)	Y	Tb927.5.2580			5(5)	Y
Tb10.70.2155			1(1)		Tb927.5.2930		10(9)	25(24)	Y
Tb10.70.2320		3(2)	4(3)	Y	Tb927.5.3040			2(2)	
Tb10.70.4500		1(1)	5(5)	Y	Tb927.5.3090		1(1)	2(2)	Y
Tb10.70.6930			2(2)		Tb927.5.3300			1(1)	
Tb10.70.7760	LETM1, inner mitochondrial membrane protein	1(1)	14(12)	Y	Tb927.6.4130		1(1)	5(4)	Y
Tb11.01.1690			6(6)		Tb927.6.590		1(1)	1(1)	
Tb11.01.1900	Macro domain, Appr-1-p processing	4(4)	8(8)	Y	Tb927.7.3590		1(1)	2(2)	Y
Tb11.01.3860		3(1)	7(5)	Y	Tb927.7.3910			3(3)	
Tb11.01.5390	DnaJ domain		3(3)	Y	Tb927.7.5260	CHCH domain		3(3)	
Tb11.01.7090	Complex 1 protein (LYR family)		3(3)		Tb927.7.6990			6(6)	Y
Tb11.02.1440			7(7)	Y	Tb927.7.7440		1(1)	4(4)	Y
Tb11.02.2280	Surfeit locus 1, COG3346		8(7)		Tb927.7.840		1(1)	3(3)	
Tb11.02.2460	DnaJ domain		5(5)	Y	Tb927.8.1490	DUF1674 (domain of unknown function)		3(3)	Y
Tb11.02.2480			3(3)		Tb927.8.2530	COX17, Cytochrome C oxidase copper chaperone		1(1)	
Tb11.02.4120		4(4)	7(7)	Y	Tb927.8.3320	IQ calmodulin-binding motif		5(5)	Y
Tb11.02.5160		1(1)	7(6)	Y	Tb927.8.5560	SpeE, Spermidine synthase [Amino acid transport and metabolism]		12(10)	
Tb11.02.5660		6(6)	8(8)	Y	Tb927.8.5640		2(2)	5(5)	Y
Tb11.03.0475		1(1)	5(5)	Y	Tb927.8.580			3(3)	
Tb11.0400 ^{HP}		3(3)	9(9)	Y	Tb927.8.6080	Glycerophosphoryl diester phosphodiesterase family	4(4)	8(8)	Y
Tb11.46.0006			5(5)	Y	Tb927.8.6960			2(2)	

Protein_ID	Domains/motifs	PF	Mt	GG	Protein_ID	Domains/motifs	PF	Mt	GG
Tb11.47.0022		3(3)	9(8)	Y					