

Generation of a Mitochondrial Protein Compendium in *Dictyostelium Discoideum*

Anna V Freitas

National Institutes of Health

Jake T Herb

National Institutes of Health

Miao Pan

National Institutes of Health

Yong Cheng

National Institutes of Health

Marjan Gucek

National Institutes of Health

Tian Jin

National Institutes of Health

Hong Xu (✉ hong.xu@nih.gov)

National Institutes of Health

Research Article

Keywords: Dictyostelium discoideum, mitochondria, metabolism, MitoCarta, proteomic

Posted Date: November 17th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1054199/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

The social amoeba *Dictyostelium discoideum* is a well-established model to study numerous cellular processes including cell motility, chemotaxis, and differentiation. As energy metabolism is involved in these processes, mitochondrial genetics and bioenergetics are of interest, though many features of *Dictyostelium* mitochondria differ from metazoans. A comprehensive inventory of mitochondrial proteins is critical to understanding mitochondrial processes and their involvement in various cellular pathways. Here, we utilized high-throughput multiplexed protein quantitation and homology analyses to generate a high-confidence mitochondrial protein compendium. Our proteomic approach, which utilizes quantitative mass spectrometry in combination with mathematical modeling, was validated through mitochondrial targeting sequence prediction and live-cell imaging. Our final compendium consists of 1082 proteins. Within our *D. discoideum* mitochondrial proteome, we identify many proteins that are not present in humans, yeasts, or the ancestral alpha-proteobacteria, which can serve as a foundation for future investigations into the unique mitochondria of *Dictyostelium*. Additionally, we leverage our compendium to highlight the complexity of metabolic reprogramming during starvation-induced development. Our compendium lays a foundation to investigate mitochondrial processes that are unique in protists, as well as for future studies to understand the functions of conserved mitochondrial proteins in health and diseases using *D. discoideum* as the model.

Introduction

Dictyostelium discoideum, a social amoeba, is a well-established model organism to study eukaryotic cellular processes such as cell motility, chemotaxis, and differentiation (Bozzaro, 2013). Under normal nutrient conditions, *D. discoideum* grows axenically through binary fission (Kessin, 2001). However, upon starvation, amoebae secrete cAMP, which attracts neighboring cells to aggregate together and form a multicellular mound. Cells in a mound move collectively as a slug toward light, heat, or humidity to find a suitable environment. The slug eventually matures into a fruiting body consisting of two major types of differentiated cells, spore cells that will start a new life cycle and stalk cells that form a stalk to hold the spore aloft (Kay, 1982). As many of the aforementioned biological processes are intertwined with cellular energetics, investigation of mitochondrial biogenesis and functions is an emerging area in *D. discoideum* research (Francione *et al.*, 2011; Pearce *et al.*, 2019).

The *D. discoideum* mitochondrial genome is ~56 kb, circular, double-stranded DNA that encodes two ribosomal RNAs, 18 transfer RNAs (tRNAs), five open reading frames without annotated function, and 38 proteins including 18 subunits of the electron transport chain complexes and 15 ribosomal proteins (Ogawa *et al.*, 2000). Phylogenetic studies reveal that Amoebozoa diverged before Opisthokonta, but after the divergence of Plantae (Baldauf and Doolittle, 1997), and are more closely related to animals than plants. Notably, the *Dictyostelium* mitochondrial genetic system possesses a few differences from metazoans (Pearce *et al.*, 2019). *D. discoideum* mitochondrial DNA (mtDNA) has four introns in *cox1/2* genes and utilizes universal codons (Angata *et al.*, 1995; Ogawa *et al.*, 2000), a common feature of most plants' mitochondria (Jukes and Osawa, 1990; Cho *et al.*, 1998). The universal genetic code and the lack

of a full set of tRNA genes on the *Dictyostelium* mitochondrial genome indicate that some nuclear-encoded tRNAs are likely imported into mitochondria to support the organellar translation. Additionally, the electron transport chain in *Dictyostelium* contain an additional component compared to its metazoan counterparts: an alternative oxidase (AOX) (Pearce *et al.*, 2019), which is found across eukaryotic clades besides animals (McDonald *et al.*, 2008). AOX is highly expressed during vegetative growth, but its expression level is markedly reduced upon starvation, suggesting a potential metabolic reprogramming occurs during starvation-induced development (Jarmuszkiwicz *et al.*, 2002). Interestingly, either reduction of mtDNA content or disruption of the *rps4* locus (encoding mt-ribosomal protein S4) on mtDNA impairs aggregation and slug phototaxis but has no impact on vegetative growth (Chida, 2004; Chida *et al.*, 2008), suggesting that mtDNA, and most likely an intact oxidative phosphorylation system is essential to initiate the development program. On the other hand, pharmacological inhibitions of either Complex I or Complex V can induce aggregation, even though mitochondrial respiration appears to increase at the beginning of starvation (Kelly *et al.*, 2021). Therefore, the interplay between mitochondrial function and *Dictyostelium* development remains to be explored.

Despite the growing interest in using *D. discoideum* as a model organism to study many conserved mitochondrial processes and some unique biology, a comprehensive list of the mitochondrial proteins has yet to be established. A recent proteomic study detected 294 proteins in *D. discoideum* mitochondria (Mazur *et al.*, 2021), which we believe is far from complete. Nuclear-encoded mitochondrial proteins, which constitute over 90% of the total mitochondrial proteome, are synthesized in the cytoplasm and then imported to mitochondria. It is estimated that the import of ~60% of these proteins relies on a positively charged, N-terminal mitochondrial targeting sequence (MTS) (Vögtle *et al.*, 2009). Computational approaches that integrate machine learning and known biological data are frequently used to predict mitochondrial targeting based on the presence of an MTS (Almagro Armenteros *et al.*, 2019). However, this method is insufficient to capture all mitochondrial proteins, as most proteins on the outer membrane and in the inner membrane space, and some inner membrane proteins rely on alternate translocation mechanisms. An alternative computational approach leverages sequence homology to known mitochondrial protein compendiums that were generated using mass spectrometry (MS)-based proteomic discovery (Pagliarini *et al.*, 2008; Morgenstern *et al.*, 2017). However, to compensate for the rapid evolution of the mitochondrial genome, nuclear-encoded mitochondrial proteins evolve faster than other nuclear-encoded proteins (Cole *et al.*, 1995; Sloan *et al.*, 2014; Havird *et al.*, 2015; Li *et al.*, 2017; Yan *et al.*, 2019). Thus, some *Dictyostelium* mitochondrial proteins may escape the homology search, and protist-specific mitochondrial proteins will certainly be missed.

In this study, we combined quantitative proteomics and mathematical modeling to identify over 900 high-confidence mitochondrial proteins, which were validated through both *in silico* and fluorescent microscopy analyses. We further complemented the proteomics-based mitochondrial protein discovery with bioinformatic approaches to create a compendium of 1082 *D. discoideum* mitochondrial proteins. We also discuss conserved *D. discoideum* mitochondrial proteins that may be used as the basis of validating mitochondrial proteins in other organisms, as well as unique features of the mitochondrial proteome in *D. discoideum*.

Results And Discussion

Mitochondrial protein discovery using quantitative proteomics

To identify putative *D. discoideum* mitochondrial proteins, we searched for *Dictyostelium* homologs of 1136 human mitochondrial proteins listed in the Human MitoCarta 3.0 (Morgenstern *et al.*, 2017; Rath *et al.*, 2021), and retrieved 616 proteins (Figure 1A, Table S1). This number is much less than known mitochondrial proteins in humans (1136) and baker's yeast (901) (Pagliarini *et al.*, 2008; Rath *et al.*, 2021). We posited that mitochondrial proteome might be highly divergent between *D. discoideum* and humans, and many *Dictyostelium* mitochondrial proteins might be missed from this bioinformatic curation. We, therefore, took a proteomic approach to directly identify mitochondrial proteins in *D. discoideum* (Figure 1A). From AX2 axenic cultures, we prepared mitochondria isolates—both crude and highly purified—through Percoll gradient ultracentrifugation. We performed tandem mass tag (TMT) liquid chromatography-mass spectrometry (LC-MS) on both mitochondria isolates and included AX2 whole-cell lysate as the control. A total of 6,892 proteins were captured in all samples (Table S2).

A limitation of identifying organellar proteins from their subcellular fractions alone is that high-abundance contaminants are often co-purified and result in false-positive hits. To address this issue, we assessed the probability of a protein localizing to mitochondria by comparing its relative enrichment in mitochondrial preparations to a list of 47 authentic mitochondrial proteins that includes components of electron transport chain complexes and conserved enzymes in citrate cycles (Table S3). We first calculated the ratio of a protein's abundance in the mitochondria isolates, both crude and highly purified, versus its abundance in the whole-cell lysate. The resulting value, indicating its enrichment in mitochondrial preparations, was further normalized to the average enrichment ratio of the 47 reference mitochondrial proteins, to compute the relative enrichment ratio (RER). Overall, a protein's RER in crude mitochondria isolate is in accordance with that in purified mitochondria (Figure 2A). However, the distribution of RERs appears continuous in crude mitochondria (Figure 1B), but clusters into two distinct populations in purified mitochondria (Figure 1B), which allowed us to determine a proper threshold of RER for mitochondrial proteins using mathematical modeling. Thus, we proceeded to analyze the RER for purified mitochondria only.

In principle, a true mitochondrial protein would be co-purified with the reference mitochondrial proteins in pure mitochondrial isolates, and its RER should be 1.0. However, the RER distribution of these 47 reference proteins (Figure 2B) appears as a normal curve centered around 1.0, suggesting that many mitochondrial proteins may have an RER below 1.0. Among all proteins profiled using TMT-based LC-MS, only 259 have an RER higher than 1.0 in purified mitochondria (Figure 2C). We posit that different mitochondria proteins might be degraded to different extents, based on their intrinsic stability, during the procedure of mitochondrial purification, which involves overnight ultracentrifugation. Therefore, it is necessary to determine a proper RER value to differentiate mitochondrial proteins from non-mitochondrial proteins. We applied the expectation-maximization (EM) approach to a gaussian mixture model (GMM) to

bin all proteins into two clusters: non-mitochondrial and mitochondrial proteins based on their RER values (Figure 2D). We chose an RER cutoff of 0.343 (Figure 2D) and assigned a total of 908 proteins having an RER higher than 0.343 as putative mitochondrial proteins (Table S4). GMM predicts that less than 17% proteins in the mitochondrial cluster, and only 0.1% proteins in the non-mitochondrial cluster would spill over to the other group (Figure 2D), which corresponds to an 83% recovery rate and 7% false discovery rate, respectively.

Validation of mitochondrial protein discovery based on quantitative proteomics

To validate the accuracy of RER-based mitochondrial protein discovery, we first assessed the recovery rate of putative mitochondrial proteins *in silico*. Many mitochondrial proteins possess an N-terminus mitochondrial targeting sequence (MTS) that directs the import of nuclear-encoded mitochondrial proteins into the mitochondrial matrix (Backers, 2017). Overall, 24% of all proteins retrieved in the proteomics discovery experiment contain an MTS (Table S4). Importantly, 94% of these MTS-bearing proteins had an RER greater than 0.343 (Figure 3A). On the contrary, 96% of proteins that were destined to other organelles such as the ER, Golgi, lysosomes, vacuoles, or secretory pathway had an RER less than 0.343. These analyses demonstrate that a cutoff value of an RER at 0.343 effectively separates mitochondrial proteins from non-mitochondrial proteins.

We also surveyed the localization of 81 proteins recovered in LC-MS (Table S5), using fluorescent microscopy. These proteins were selected on the basis that their subcellular localization has not been annotated previously as mitochondrial, and their RERs are randomly distributed from 0.1 to 1.5. Each protein was tagged with GFP at its C-terminus and co-expressed with an MTS-mCherry fusion protein, which marks mitochondria in *D. discoideum* AX2 cells. Among the 81 proteins, 90% of proteins with an RER higher than 0.343 showed complete or partial mitochondrial localization (Figure 3B, 3D), whereas only 5% of proteins with an RER less than 0.343 showed mitochondrial localization (Figure 3B, 3D), demonstrating a strong positive correlation between RER value and probability of mitochondrial localization (Figure 3B). Moreover, logistic regression analysis on the localization pattern of these 81 proteins predicts that a protein has more than a 78% probability of localizing to the mitochondria if its RER is higher than 0.343 (Figure 3C).

A comprehensive mitochondrial protein compendium in *D. discoideum*.

To further improve the coverage and accuracy of the mitochondrial protein discovery, we revised the list based on the *in vivo* microscopy validation by removing four non-mitochondrial localizing proteins and adding two mitochondrial localizing proteins. We also integrated three sets of mitochondrial protein discovery: the aforementioned list of mitochondrial proteins identified from quantitative proteomics analyses, those retrieved from homology detection, and those retrieved during a gene ontology search for mitochondrial genes. Among the 616 *D. discoideum* homologs of human mitochondrial proteins (Table S1), 352 proteins have an RER higher than 0.343 and hence were already included in the list, 223 proteins have an RER lower than 0.343, and 41 proteins were not captured in LC-MS. Among the 264 proteins that

were not included in the list, 113 proteins do not have a predicted MTS (Table S1), whereas their human homologs have MTSs, suggesting these proteins might localize to other cellular compartments in *D. discoideum*. An exception is ribosomal protein S14 (O21035), which is encoded in the nuclear genome in humans but is encoded in the mitochondrial genome in *D. discoideum*, and thus contains an MTS in human cells but lacks one in *D. discoideum*. We added those remaining 152 proteins to the list, as well as 32 proteins with mitochondrial gene ontologies that had not emerged during the proteomic or homology analysis. The final compendium consists of 1082 high-confidence mitochondrial proteins in *D. discoideum* (Table S6).

Characterization of the *D. discoideum* mitochondrial proteome

Out of the 1082 *D. discoideum* mitochondrial proteins, there are 627 and 458 proteins that have homologs in the mitochondrial proteome of human and *Saccharomyces cerevisiae*, respectively (Figure 4A), indicating that *D. discoideum* mitochondria are more closely related to mitochondria in metazoans than fungi. Only 324 *D. discoideum* mitochondrial proteins have homologs in *Rickettsia prowazekii* (Figure 4A), an α -proteobacteria that is closely related to the mitochondrial ancestor. Overall, a total of 313 proteins, representing 28.9% of the *D. discoideum* mitochondrial proteome, have no homologs in the whole proteome of humans, *S. cerevisiae* or *R. prowazekii* (Figure 4A, Table S6), indicating that a large fraction of *D. discoideum* mitochondrial proteins was evolved *de novo* after the divergence of Amoebozoa. Moreover, 75 *D. discoideum* mitochondrial proteins (6.9%) have no homologs in *D. purpureum* (Figure 4A), a closely related species of social amoeba, further substantiating the fast-evolving nature of the amoeba mitochondrial proteome.

There are 89 *D. discoideum* mitochondrial proteins (8.2%) with human homologs that had not been annotated as mitochondrial proteins (Figure 4B, Table S6). Among these 89 proteins, 74 were also not annotated as mitochondrial proteins in yeast, including 32 that had homologs in *S. cerevisiae*. Given the estimated false-discovery rate of our compendium, the localization of these proteins needs to be experimentally accessed. Nonetheless, there are a few examples, such as the RNB domain-containing protein (DDB_G0288469) and tRNA-binding domain-containing protein (DDB_G0349377), both of which have predicted MTSs and are likely targeted to the mitochondrial matrix. The human homologs of DDB_G0288469, DIS3-like exonuclease 2, and DDB_G0349377, rhomboid-related protein 4, were not included in the human compendium (Rath *et al.*, 2021), despite evidence that the yeast homolog of DIS3-like exonuclease 2 localizes to the mitochondria (Pagliarini *et al.*, 2008), and that rhomboid-related protein 4 has been partially shown to localize to the mitochondria. The mitochondrial localization of their *D. discoideum* homologs substantiates these two proteins might indeed localize to the mitochondria and indicates that our compendium can complement previous studies toward a more comprehensive discovery of mitochondrial proteins in other organisms.

Additionally, we categorized the *D. discoideum* mitochondrial proteome using PANTHER biological function or protein family classifications (Figure 4B, Table S6). Proteins involved in mitochondrial gene expression and metabolism comprise the largest fractions of all mitochondrial proteins, over 20% for

each category. Other proteins are involved in mitochondrial protein homeostasis, the electron transport chain, redox signaling and metabolism, and regulation of mitochondrial morphology and dynamics. A large fraction of *D. discoideum* mitochondrial proteins, approximately 15%, have no classified functions based on PANTHER analyses (Figure 4B).

***D. discoideum* -specific mitochondrial proteins**

Proteins involved in gene expression consisted of a large fraction of *D. discoideum* specific mitochondrial proteome (Figure 4B), reflecting that the *D. discoideum* mitochondrial genome is more complex than human mtDNA. On the contrary, few metabolism proteins emerged in the list (Figure 4B), suggesting that metabolic processes are highly conserved between *D. discoideum* and metazoans. Here, we expand upon a few of the unique features of the *D. discoideum* mitochondrial protein compendium.

Mosaic nature of mitochondrial ribosomes

Mitochondrial ribosomes (mitoribosomes), ribosomal assembly factors, and other proteins involved in translation represented 8.9% and 9.3% of the overall and unique mitochondrial protein compendium, respectively. While mitoribosomes are thought to be evolved from bacterial ribosomes, these two differ greatly with regards to their structure, function, as well as their composition of proteins and RNAs. We identified 51 proteins that are predicted to be mitoribosomal proteins, including 13 proteins that did not share significant homology with any *H. sapiens*, *S. cerevisiae*, or *R. prowazekii* proteins (Table S7). Interestingly, *D. discoideum* mitoribosomal proteins belong to families across several taxonomic groups: 35 proteins belong to mammalian mitoribosomal protein families (28s and 39s), 2 belong to eukaryotic cytosolic ribosomal protein families (60s), 9 belong to yeast mitoribosomal protein families (37s and 54s), 2 belong to chloroplast or bacterial ribosomal protein families (30s and 50s), 1 is from archaea, and 2 are universally conserved among prokaryotes and eukaryotes. It has previously been shown that cytosolic ribosomes tether to the mitochondrial outer membrane. Hence, the recovery of the 60S ribosomal protein L22 could be the result of the association of cytoplasmic ribosome with the mitochondrial outer membrane rather than its localization in the matrix (Gold *et al.*, 2017). Nonetheless, the presence of proteins representing multiple mitoribosome lineages suggests that there may be *D. discoideum* or protist-specific mechanisms to process mitochondrial transcripts and to regulate mitochondrial translation. Further validation of these findings is necessary as the composition and structure of the *D. discoideum* mitoribosome have yet to be resolved.

Mitochondrial DNA and RNA processing factors

Among the list of unique proteins are 24 candidate mtDNA and mtRNA processing factors including five endonucleases and a pentatricopeptide repeat (PPR)-containing protein A (PtcA). Bioinformatic analysis suggests that PtcA belongs to the mitochondrial group I intron splicing family. PPR proteins, defined by tandem PPR domains, are implicated in several different mitochondrial gene expression processes including translation initiation, and ribosomal stabilization (Manna, 2015). The number of PPR proteins that are encoded in an organism varies greatly: terrestrial plants, such as *Arabidopsis thaliana*, have

upwards of 450 PPR proteins, while humans have 7 (Lurin *et al.*, 2004; Lightowlers and Chrzanowska-Lightowlers, 2013). *D. discoideum* has 12 PPR-domain containing peptides, including PtcA, reflecting a greater complexity of *D. discoideum*'s mitochondrial genome compared to that of metazoans (Manna *et al.*, 2013).

Divergent evolution path of lipid biosynthesis

The mevalonate pathway, which produces five-carbon blocks for the synthesis of diverse biomolecules such as cholesterol and coenzyme Q10, is an essential and highly conserved process in eukaryotes, archaea, and some bacteria. In animals and fungi, the mevalonate pathway takes place in ER, and 3-hydroxy-3-methylglutaryl (HMG)-coenzyme A (CoA) reductase (HMGR), a key enzyme in this pathway that converts HMG-CoA to mevalonate, localizes in the ER and peroxisomes (Chin *et al.*, 1984; Keller *et al.*, 1986; Burg and Espenshade, 2011). HMGR2, one of two HMG reductases in *D. discoideum*, is recovered in our compendium and contains a predicted MTS, suggesting that it likely localizes to the mitochondrial matrix. Additionally, HGSA, one of the two HMG-CoA synthases, also emerged as a mitochondrial protein. Our mitochondrial protein discovery suggests that mevalonate metabolism may take place in mitochondria in *D. discoideum*, highlighting the evolutionary divergence of some metabolic pathways that originated from the common mitochondrial ancestor.

Implication of mitochondrial function in multicellular development

D. discoideum with reduced mtDNA or a disruption of the gene encoding mt-ribosomal protein S4 display no defect in vegetative growth but have impaired starvation-induced development, suggesting that mitochondrial respiration is necessary for multicellularity (Chida, 2004; Chida *et al.*, 2008). However, contrasting evidence has demonstrated a significant decrease in mitochondrial respiration after respiration, and accordingly, a decreasing expression of many respiration complexes (Kelly *et al.*, 2021). To understand potential regulations of mitochondrial function in multicellular development, we retrieved RNA sequencing data using the *Dictyostelium* gene expression database, dictyExpress (Parikh *et al.*, 2010; Stajdohar *et al.*, 2017).

Overall, there was a decrease in the expression of mitochondrial genes within our compendium over the 24-hr development time course (Figure 5A). A similar pattern is observed in proteins that are involved in mitochondrial DNA maintenance and gene expression. Interestingly, despite the decrease in gene expression machinery (Figure 5B), over half of the mitochondria-encoded genes in the dataset (19 of 35) were upregulated ($\log_2FC \geq 1$) after starvation induction (Figure 5C). Further, in examining all respiratory chain complexes, 12 nuclear-encoded ETC subunits had a higher expression level ($\log_2FC \geq 1$) at or after 12 hours of starvation (Figure 5D), besides the 10 nuclear or mitochondrial-encoded subunits that show a burst of expression in the first 4 hours after the starvation (Figure 5D). The complex pattern of mitochondrial gene expression, particularly the upregulation of electron transport chain complex subunits during the development suggests potential roles of mitochondrial respiration in *Dictyostelium*

development, and that both nuclear and mitochondrial-encoded proteins are likely implicated in these processes.

Conclusion

Here, we generated the most comprehensive list of mitochondrial proteins in *D. discoideum* to date. Our compendium lays the foundation for future studies to understand the functions of conserved mitochondrial proteins in health and diseases using *D. discoideum* as the model. It also provides an entry to study many fascinating mitochondrial processes that are unique in protists. Additionally, thorough comparative genomics, our compendium will complement mitochondrial protein discovery in other organisms and may shed light on the evolution of mitochondrial proteome and processes.

Materials And Methods

Cell culture and transformation

Dictyostelium discoideum AX2 cultures were maintained in HL5 medium at 22°C (Fey *et al.*, 2007). Transformants were generated via electroporation as previously described (Gaudet *et al.*, 2007), with modifications. After electroporation (BioRad Genepulser), cells were incubated on ice for 10 minutes. Subsequently, cells were transferred from the cuvette with 2 mL of HL5 and plated onto 12-well tissue culture plates. After 24 hours, transformants were selected with Genectin and/or Blasticidin S (Thermo Fisher, 10 µg/mL each) in HL5.

Protein mass spectrometry

Mitochondrial isolation

Cells were harvested at a concentration of $1-3 \times 10^6$ cells/mL and resuspended at 2×10^7 cells/mL in 800 µL of Reagent A from the Mitochondrial Isolation Kit for Cultured Cells (Thermo Fisher #89874) on ice. Cell lysis and crude mitochondrial preparation were performed as previously described with modifications (Graham, 1999; Glancy and Balaban, 2011). Cells were lysed with 35 strokes of a Dounce homogenizer followed by the addition of an equal volume of Reagent C. Whole-cell lysate samples were stored at -80°C or were centrifuged three times (700 x g, 10 minutes, 4°C) to purify the mitochondria. For each centrifugation step, the supernatant was transferred to a fresh 1.5 mL tube. The crude mitochondrial lysate was used immediately for purification or was stored at -80°C.

To generate purified mitochondrial isolates, Percoll gradient centrifugation was performed as follows. Lysis suspension (1-2 mL) was added to the top of a Percoll (Cytiva) and Development Buffer (DB) (5 mM Na₂HPO₄, 5 mM KH₂PO₄, 1 mM CaCl₂, 2 mM MgCl₂, pH = 6.5) solution (8 mL, 30% Percoll) in a 10 mL ultracentrifuge tube. Ultracentrifugation (68,000 x g, 40 minutes) yielded three distinct layers. The top layer, containing contaminants, was discarded. The middle layer, containing mitochondria, was transferred into a fresh 2 mL tube. Aliquots of the mitochondrial suspension were topped off with 500 µL

of DB, then centrifuged (13,000 x g, 10 minutes, 4°C). Following centrifugation, the supernatant was aspirated, and the mitochondria-containing pellet was maintained on ice. To lyse the mitochondria, the pellet was washed with 2 mL of DB, centrifuged (10,000 x g, 10 minutes, 4°C), and resuspended in 2 mL of DB with urea (8M). Protein yield was quantified via Bradford Assay (BioRad) according to the manufacturer's protocol.

Relative protein quantification

Resuspended cell pellets were lysed via pulsed sonication, then sequentially reduced, alkylated, and digested overnight with trypsin. The protein digests were labeled with 10-plex Tandem Mass Tag (TMT) reagents (Thermo Fisher Scientific) (Dayon *et al.*, 2008), then were pooled and desalted. To separate the peptide mixtures into 24 fractions, high pH reversed-phase liquid chromatography was performed (Yang *et al.*, 2012). Each fraction was analyzed on an Orbitrap Lumos (Thermo Fisher Scientific) nanoLCMS system.

Peptide and proteins were identified as described in He *et al.*, (2020). In brief, the resulting LCMS raw data were searched against a database downloaded from dictybase.org using the Sequest HT algorithm on the Proteome Discoverer 2.4 platform (Thermo Fisher Scientific). Three groups of samples were normalized to 47 reference mitochondrial proteins.

The relative enrichment (RE) was defined as the ratio of a protein's enrichment from a crude or purified mitochondria sample over its enrichment from a whole-cell lysate sample. To calculate the relative enrichment ratio (RER), REs were normalized such that the average RER of 47 known mitochondrial (TCA cycle, ETC, or OXPHOS) proteins is 1 (Table S3). The RER presented is a median value of three biological replicates.

Mathematical modeling

To classify proteins as mitochondrial or non-mitochondrial based on their RER, the RER distribution of isolated proteins was fit to a Gaussian mixture model (GMM) using the Expectation-Maximization (EM) algorithm in R (Benaglia *et al.*, 2009). The cutoff value (RER = 0.343) was four times the standard deviation plus the mean of curve 1 (representing non-mitochondrial proteins), such that 99.9% of the proteins below the cutoff were contained within curve 1.

Bioinformatic analyses

Homology analyses

Protein sequence homology was established by BlastP expect < 0.001 and bit-score > 40 (Pearson, 2013), or by HMMER sequence e-value < 0.01. Subcellular localization was predicted using TargetP-2.0 (Almagro Armenteros *et al.*, 2019). Biological functions for all proteins in the *D. discoideum* and human mitochondrial proteome were manually categorized from biological function or protein family classifications provided from PANTHER (The Gene Ontology Consortium *et al.*, 2021).

In silico dataset correction

Two strategies were implemented to supplement the mitochondrial protein discovery. The top *D. discoideum* homolog of human mitochondrial proteins (Table S1, Rath *et al.*, 2021) were curated. Additionally, *D. discoideum* proteins annotated with the gene ontology term “mitochondr*” on AmiGO were selected (Carbon *et al.*, 2009). Proteins within these lists were integrated into the final mitochondrial compendium so long as they had a predicted mitochondrial targeting sequence if their human homolog also had a predicted mitochondrial targeting sequence, except in cases where the *D. discoideum* protein was mitochondrial-encoded.

RNA sequencing data visualization

Normalized RNA-seq data from Parikh *et al.* (2010) was retrieved using dictyExpress (Stajdohar *et al.*, 2017). For the 1082 proteins in the mitochondrial compendium, only 1075 corresponding genes were present in the dataset. To compare the gene profiles, data were scaled to a mean of 0 and standard deviation of 1 using the scale function in R. For the overall mitochondrial expression profile, genes and timepoints were ordered using hierarchical clustering (heatmaps.2). For the profiles of individual biological processes, scaled data (Table S8) were imported into the matrix visualization software Morpheus (<https://software.broadinstitute.org/morpheus>) and ordered via hierarchical clustering with one minus Pearson’s correlation as the distance metric and average as the linkage method. Gene upregulation was determined by a log₂ fold-change ≥ 1 .

Library generation for imaging verification

To evaluate the efficacy of our mitochondrial protein identification, 98 proteins were selected to be GFP tagged so that their localization could be assessed via fluorescence microscopy (Table S5). None of the proteins selected had a gene ontology annotation that indicated mitochondrial localization. Four proteins selected for verification had homologs listed in the human mitochondrial proteome.

All genes were synthesized by Gene Universal. Of the 98 genes submitted for synthesis, 85 were generated as inserts in pDM323, a *D. discoideum* extrachromosomal expression vector with G418 resistance and a C-terminal GFP tag (Veltman *et al.*, 2009); 6 genes were generated as inserts in the shuttle vector puC57 and were subsequently cloned into pDM323 between BglIII and SpeI sites using the In-Fusion® HD Cloning Kit (Takara Bio USA) and confirmed by sequencing. The other 7 genes were unable to be synthesized, such that only 92 proteins were screened. Of these 92 genes, only 81 were successfully expressed in *D. discoideum*. The RER of the proteins that were verified were as follows: 14 proteins with a RER > .75, 11 proteins with a RER = 0.75-0.5, 24 proteins with a RER = .5-0.25, and 32 proteins with RER < 0.25. To observe mitochondrial localization, the mitochondrial targeting sequence of respiratory cytochrome oxidase c subunit IV fused with mCherry (CoxIV-mCherry) was cloned into pDM326, a *D. discoideum* extrachromosomal expression vector with Blasticidin S resistance (Veltman *et al.*, 2009). All primers for cloning are listed in Table S9.

Live-cell imaging

To image cells in the axenic phase, cells (200 μ l) were transferred to 8-well glass chambers 7 to 10 days post-transformation. Cells were allowed to adhere to the bottom of the chamber for 30 minutes before the media was aspirated. The media was replaced with 1x PBS after three washes (200 μ l for all). Confocal images were collected on a PerkinElmer Ultraview system (Zeiss Plan-apochromat 63x/1.4 oil lens, Volocity acquisition software, Hamamatsu Digital Camera C10600 ORCA-R2, Immersol immersion oil). Images (0.5 μ m z-step) were analyzed with ImageJ (National Institutes of Health) and formatted in Adobe Photoshop.

Code and data availability

Proteomics data are deposited at ProteomXchange (PXD029101). R code is available from Anna Freitas's GitHub repo <https://github.com/freitasav/DD-mitoproteome>.

Quantification and Statistical Analyses

All data were presented as the mean \pm SD unless otherwise indicated. P values were calculated in R using a one-way analysis of variance (ANOVA) followed by Tukey's post-hoc test to test for the effect of RER on mitochondrial localization. Statistical significance of difference was considered when $p < 0.05$.

To predict the probability of localization based on RER, outliers were identified and removed from the microscopy validation dataset based on the interquartile method (median + 1.5 SD). Data were analyzed using a binomial logistic regression (glm function in R) with *excluded from the mitochondria* as the reference level, and *partial mitochondrial localization*, *mitochondrial localization*, or *combined* (in which the partial and mitochondrial outcomes are collapsed) as the outcome levels. Predicted probabilities and 95% confidence intervals were calculated (predict function in R) to compare outcomes.

Declarations

ACKNOWLEDGEMENTS

We thank Dr. Edward Korn for his advice and reagents on *D. discoideum* culturing; Dr. Raúl Covian Garcia for his advice on mitochondria purification; and dictyBase for various plasmids. This work was supported by the Intramural Research Program of National Heart, Lung, and Blood Institute.

DECLARATION OF INTERESTS

The authors declare no competing interests.

References

1. Almagro Armenteros, JJ, Salvatore, M, Emanuelsson, O, Winther, O, von Heijne, G, Elofsson, A, and Nielsen, H (2019). Detecting sequence signals in targeting peptides using deep learning. *Life Sci Alliance* 2, e201900429.

2. Angata, K, Kuroe, K, Yanagisawa, K, and Tanaka, Y (1995). Codon usage, genetic code and phylogeny of *Dictyostelium discoideum* mitochondrial DNA as deduced from a 7.3-kb region. *Curr Genet* 27, 249–256.
3. Benaglia, T, Chauveau, D, Hunter, DR, and Young, D (2009). mixtools: An R Package for Analyzing Finite Mixture Models. *J Stat Soft* 32.
4. Bozzaro, S (2013). The Model Organism *Dictyostelium discoideum*. In: *Dictyostelium Discoideum Protocols*, ed. L Eichinger, and F Rivero, Totowa, NJ: Humana Press, 17–37.
5. Burg, JS, and Espenshade, PJ (2011). Regulation of HMG-CoA reductase in mammals and yeast. *Progress in Lipid Research* 50, 403–410.
6. Carbon, S, Ireland, A, Mungall, CJ, Shu, S, Marshall, B, Lewis, S, the AmiGO Hub, and the Web Presence Working Group (2009). AmiGO: online access to ontology and annotation data. *Bioinformatics* 25, 288–289.
7. Chida, J (2004). The necessity of mitochondrial genome DNA for normal development of *Dictyostelium* cells. *Journal of Cell Science* 117, 3141–3152.
8. Chida, J, Amagai, A, Tanaka, M, and Maeda, Y (2008). Establishment of a new method for precisely determining the functions of individual mitochondrial genes, using *Dictyostelium* cells. *BMC Genet* 9, 25.
9. Chin, DJ, Gil, G, Russell, DW, Liscum, L, Luskey, KL, Basu, SK, Okayama, H, Berg, P, Goldstein, JL, and Brown, MS (1984). Nucleotide sequence of 3-hydroxy-3-methyl-glutaryl coenzyme A reductase, a glycoprotein of endoplasmic reticulum. *Nature* 308, 613–617.
10. Cho, Y, Qiu, Y-L, Kuhlman, P, and Palmer, JD (1998). Explosive invasion of plant mitochondria by a group I intron. *Proceedings of the National Academy of Sciences* 95, 14244–14249.
11. Cole, RA, Slade, MB, and Williams, KL (1995). *Dictyostelium discoideum* mitochondrial DNA encodes a NADH:Ubiquinone oxidoreductase subunit which is nuclear encoded in other eukaryotes. *J Mol Evol* 40, 616–621.
12. Dayon, L, Hainard, A, Licker, V, Turck, N, Kuhn, K, Hochstrasser, DF, Burkhard, PR, and Sanchez, J-C (2008). Relative Quantification of Proteins in Human Cerebrospinal Fluids by MS/MS Using 6-Plex Isobaric Tags. *Anal Chem* 80, 2921–2931.
13. Fey, P, Kowal, AS, Gaudet, P, Pilcher, KE, and Chisholm, RL (2007). Protocols for growth and development of *Dictyostelium discoideum*. *Nat Protoc* 2, 1307–1316.
14. Francione, LM, Annesley, SJ, Carilla-Latorre, S, Escalante, R, and Fisher, PR (2011). The *Dictyostelium* model for mitochondrial disease. *Seminars in Cell & Developmental Biology* 22, 120–130.
15. Gaudet, P, Pilcher, KE, Fey, P, and Chisholm, RL (2007). Transformation of *Dictyostelium discoideum* with plasmid DNA. *Nat Protoc* 2, 1317–1324.
16. Glancy, B, and Balaban, RS (2011). Protein composition and function of red and white skeletal muscle mitochondria. *300*, 11.

17. Gold, VA, Chroscicki, P, Bragoszewski, P, and Chacinska, A (2017). Visualization of cytosolic ribosomes on the surface of mitochondria by electron cryo-tomography. *EMBO Rep* 18, 1786–1800.
18. Graham, JM (1999). Purification of a Crude Mitochondrial Fraction by Density-Gradient Centrifugation. *Current Protocols in Cell Biology* 4.
19. Havird, JC, Whitehill, NS, Snow, CD, and Sloan, DB (2015). Conservative and compensatory evolution in oxidative phosphorylation complexes of angiosperms with highly divergent rates of mitochondrial genome evolution: MITONUCLEAR COEVOLUTION IN *SILENE* OXPHOS COMPLEXES. *Evolution* 69, 3069–3081.
20. He, Y, Chen, Y, Morris, DL, Lee, D-Y, and Tjandra, N (2020). Bax expression is optimal at low oxygen tension and constant agitation. *Protein Expression and Purification* 165, 105501.
21. Jarmuszkiewicz, W, Behrendt, M, Navet, R, and Sluse, FE (2002). Uncoupling protein and alternative oxidase of *Dictyostelium discoideum*: occurrence, properties and protein expression during vegetative life and starvation-induced early development. *FEBS Letters* 532, 459–464.
22. Jukes, TH, and Osawa, S (1990). The genetic code in mitochondria and chloroplasts. *Experientia* 46, 1117–1126.
23. Kay, RR (1982). cAMP and spore differentiation in *Dictyostelium discoideum*. *Proceedings of the National Academy of Sciences* 79, 3228–3231.
24. Keller, GA, Pazirandeh, M, and Krisans, S (1986). 3-Hydroxy-3-methylglutaryl coenzyme A reductase localization in rat liver peroxisomes and microsomes of control and cholestyramine-treated animals: quantitative biochemical and immunoelectron microscopical analyses. *Journal of Cell Biology* 103, 875–886.
25. Kelly, B et al. (2021). Sulfur sequestration promotes multicellularity during nutrient limitation. *Nature* 591, 471–476.
26. Kessin, RH (2001). *Dictyostelium: Evolution, Cell Biology, and the Development of Multicellularity*, Cambridge: Cambridge University Press.
27. Li, Y, Zhang, R, Liu, S, Donath, A, Peters, RS, Ware, J, Misof, B, Niehuis, O, Pfrender, ME, and Zhou, X (2017). The molecular evolutionary dynamics of oxidative phosphorylation (OXPHOS) genes in Hymenoptera. *BMC Evol Biol* 17, 269.
28. Lightowers, RN, and Chrzanowska-Lightowers, ZM (2013). Human pentatricopeptide proteins: Only a few and what do they do? *RNA Biology* 10, 1433–1438.
29. Lurin, C et al. (2004). Genome-Wide Analysis of Arabidopsis Pentatricopeptide Repeat Proteins Reveals Their Essential Role in Organelle Biogenesis[W]. *The Plant Cell* 16, 2089–2103.
30. Manna, S (2015). An overview of pentatricopeptide repeat proteins and their applications. *Biochimie* 113, 93–99.
31. Manna, S, Brewster, J, and Barth, C (2013). Identification of Pentatricopeptide Repeat Proteins in the Model Organism *Dictyostelium discoideum*. *International Journal of Genomics* 2013, 1–8.

32. Mazur, M, Wojciechowska, D, Sitkiewicz, E, Malinowska, A, Świdarska, B, Kmita, H, and Wojtkowska, M (2021). Mitochondrial Processes during Early Development of *Dictyostelium discoideum*: From Bioenergetic to Proteomic Studies. *Genes* 12, 638.
33. Morgenstern, M et al. (2017). Definition of a High-Confidence Mitochondrial Proteome at Quantitative Scale. *Cell Reports* 19, 2836–2852.
34. Ogawa, S et al. (2000). The mitochondrial DNA of *Dictyostelium discoideum*: complete sequence, gene content and genome organization. *Mol Gen Genet* 263, 514–519.
35. Pagliarini, DJ et al. (2008). A Mitochondrial Protein Compendium Elucidates Complex I Disease Biology. *Cell* 134, 112–123.
36. Parikh, A et al. (2010). Conserved developmental transcriptomes in evolutionarily divergent species. *Genome Biol* 11, R35.
37. Pearce, XG, Annesley, SJ, and Fisher, PR (2019). The *Dictyostelium* model for mitochondrial biology and disease. *Int J Dev Biol* 63, 497–508.
38. Pearson, WR (2013). An Introduction to Sequence Similarity (“Homology”) Searching. *Current Protocols in Bioinformatics* 42.
39. Rath, S et al. (2021). MitoCarta3.0: an updated mitochondrial proteome now with sub-organelle localization and pathway annotations. *Nucleic Acids Research* 49, D1541–D1547.
40. Sloan, DB, Triant, DA, Wu, M, and Taylor, DR (2014). Cytonuclear Interactions and Relaxed Selection Accelerate Sequence Evolution in Organelle Ribosomes. *Molecular Biology and Evolution* 31, 673–682.
41. Stajdohar, M, Rosengarten, RD, Kokosar, J, Jeran, L, Blenkus, D, Shaulsky, G, and Zupan, B (2017). dictyExpress: a web-based platform for sequence data management and analytics in *Dictyostelium* and beyond. *BMC Bioinformatics* 18, 291.
42. The Gene Ontology Consortium et al. (2021). The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Research* 49, D325–D334.
43. Veltman, DM, Akar, G, Bosgraaf, L, and Van Haastert, PJM (2009). A new set of small, extrachromosomal expression vectors for *Dictyostelium discoideum*. *Plasmid* 61, 110–118.
44. Vögtle, F-N et al. (2009). Global Analysis of the Mitochondrial N-Proteome Identifies a Processing Peptidase Critical for Protein Stability. *Cell* 139, 428–439.
45. Yan, Z, Ye, G, and Werren, JH (2019). Evolutionary Rate Correlation between Mitochondrial-Encoded and Mitochondria-Associated Nuclear-Encoded Proteins in Insects. *Molecular Biology and Evolution* 36, 1022–1036.
46. Yang, F, Shen, Y, Camp, DG, and Smith, RD (2012). High-pH reversed-phase chromatography with fraction concatenation for 2D proteomic analysis. *Expert Review of Proteomics* 9, 129–134.

Figures

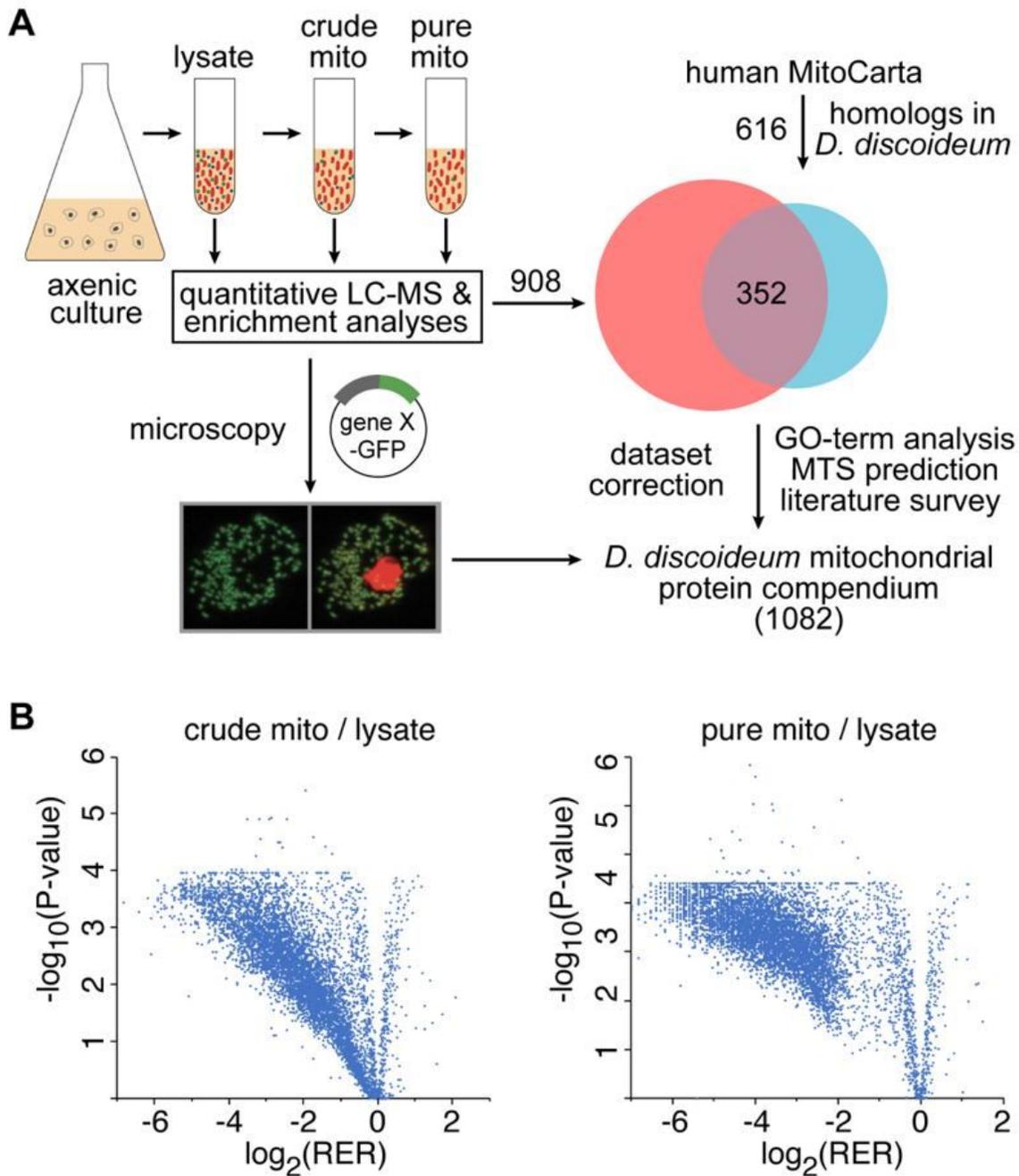


Figure 1

Curation of a comprehensive mitochondrial proteome in *Dictyostelium discoideum*. (A) A homology search against human mitochondrial protein sequences yielded a list of 616 putative *D. discoideum* mitochondrial proteins. To identify additional mitochondrial proteins, we performed quantitative MS for proteins identified in the whole-cell lysate, as well as in crude and purified mitochondrial samples. Microscopy was used to validate MS and enrichment analyses. After correcting the proteomic dataset based on microscopy results, we incorporated the homology and proteomic analyses to yield a comprehensive mitochondrial protein compendium. (B) Volcano plots displaying the $-\log_{10}(\text{p-value})$

versus log₂ (relative enrichment ratios) of proteins in crude mitochondrial or pure mitochondrial versus whole-cell lysate samples. Data are presented as means (n = 3).

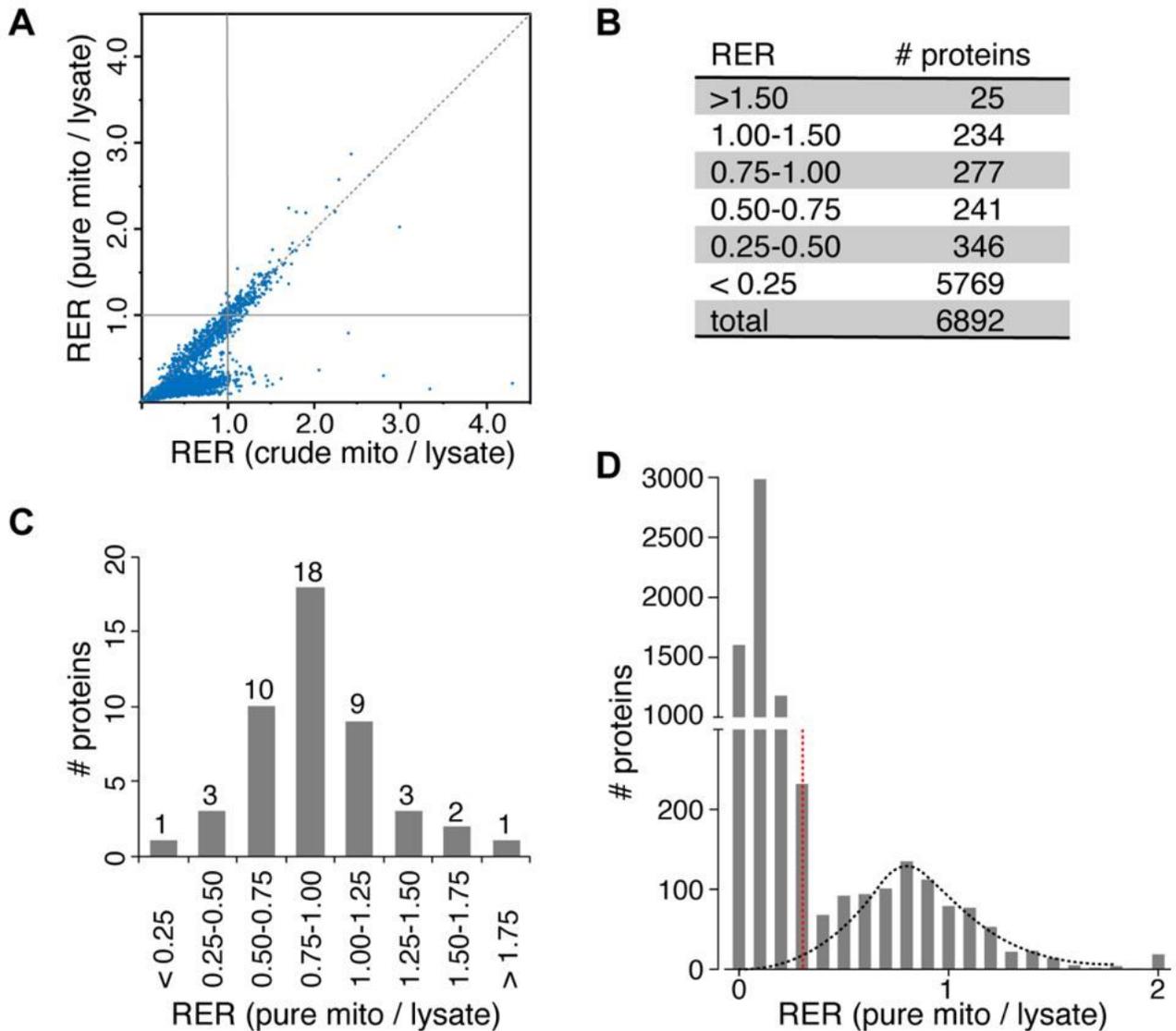


Figure 2

Prediction of mitochondrial localization based on relative enrichment analysis. (A) Crude mitochondrial RER versus pure mitochondrial RER for all samples quantified in the proteomic discovery experiment. The dashed line represents equal crude and pure RERs. Data are presented as means (n = 3). (B) Summary table of pure mitochondrial relative enrichment ratios for all proteins in the protein discovery experiment. (C) Distribution of core *D. discoideum* mitochondrial ETC, OXPHOS, and TCA cycle proteins (n = 47) based on RER. (D) The distribution of mitochondrial (n = 908) and non-mitochondrial proteins (n = 5984) based on RER follow a Gaussian mixture model, in which a cutoff value (red, RER = 0.343) separates two normal curves.

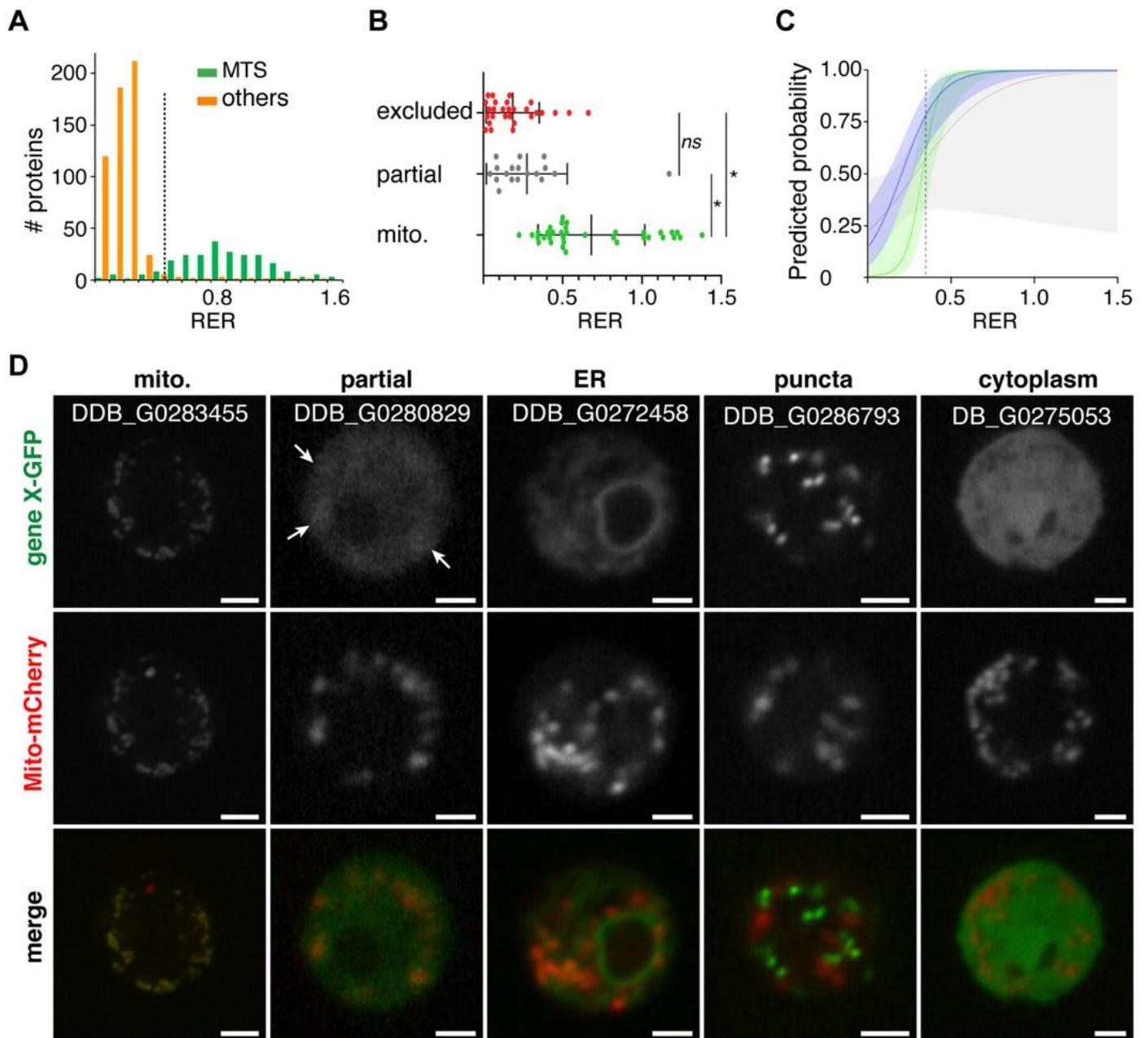


Figure 3

Validation of proteomic discovery and enrichment analyses. (A) Distribution of all proteins quantified with a predicted mitochondrial targeting sequence (MTS) or signal peptide sequence. The dashed line represents the RER cutoff for predicting mitochondrial localization (0.343). (B) Dot plot with interquartile ranges of proteins that had mitochondrial localization (green), partial mitochondrial localization (gray), or were excluded from the mitochondria (red). *, $p \leq 5 \times 10^{-5}$. (C) Predicted probabilities and confidence intervals as analyzed by logistic regression for mitochondrial (green), partial mitochondrial (gray), and either mitochondrial or partial (blue) localization. The dashed line represents the RER cutoff for predicting mitochondrial localization (0.343). (D) Representative live-cell confocal images of axenic stage D.

discoideum expressing GFP-tagged genes of interest leveled with cytochrome oxidase c subunit IV tagged with mCherry (Mito-mCherry). Arrowheads denote areas of partial localization.

A

	<i>D. discoideum</i> (1082)
<i>H. sapiens</i> (1136)	627
<i>S. cerevisiae</i> (901)	458
<i>R. prowazekii</i> (832)	324
<i>D. purpureum</i> (?)	1007

B

Biological Processes	overall 1082	unique 367	diff. localized 89
protein import & biogenesis	26	8	2
protein maturation & folding	26	13	0
protein degradation	30	1	8
DNA related processes	46	16	7
mRNA related processes	11	8	1
tRNA related processes	52	7	2
ribosome & translation	96	34	3
others of gene expression	32	15	1
respiratory chain assembly	42	11	0
respiratory chain subunits	62	14	2
energy metabolism	49	6	2
lipid metabolism	61	5	6
nucleotides metabolism	13	0	1
amino acids metabolism	45	2	2
coenzyme/cofactor biosynthesis	17	1	1
Fe/S proteins biosynthesis	15	0	0
carriers & channels	48	5	1
regulation & signaling	42	6	13
morphology & dynamics	35	3	5
homeostasis & stress response	40	6	2
developmental processes	24	5	2
others	105	65	13
unclassified	165	136	15

Figure 4

Categorization of the *D. discoideum* mitochondrial protein compendium. (A) *D. discoideum* mitochondrial proteins with homologs in *H. sapiens* or *S. cerevisiae* mitochondrial proteomes, or the complete proteomes of the alphaproteobacteria *R. prowazekii* or *D. purpureum*. (B) The functional categorization of

all *D. discoideum* mitochondrial proteins (overall, n = 1082), those that lack human homologs (unique, n = 367), and those that have human homologs, but homologs were not annotated as mitochondrial proteins (diff. localized, n = 89).

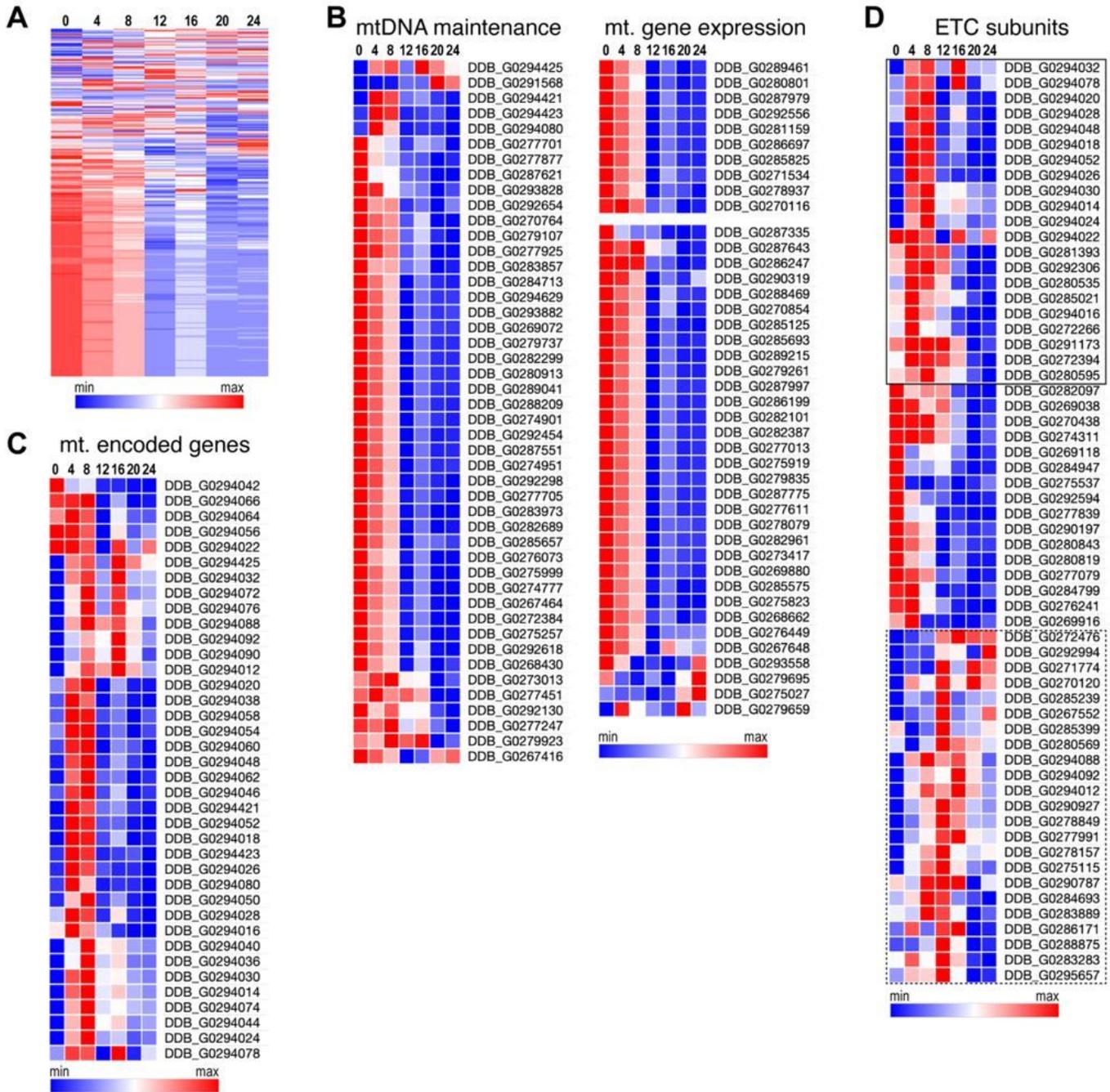


Figure 5

Expression profile of mitochondrial genes during *D. discoideum* development. Heatmaps showing normalized RNA-seq analysis values of (A) all mitochondrial proteins within the compendium and (B-D) specific mitochondrial proteins during the 24-hour starvation-induced development cycle. Each column of heat maps represents the time in hours after developmental induction. Rows are clustered by similarity in

gene expression profiles. Boxes outline ETC subunits showing increased expression 4 hours after starvation (solid) and 12 hours after starvation (dashed).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTables.xlsx](#)