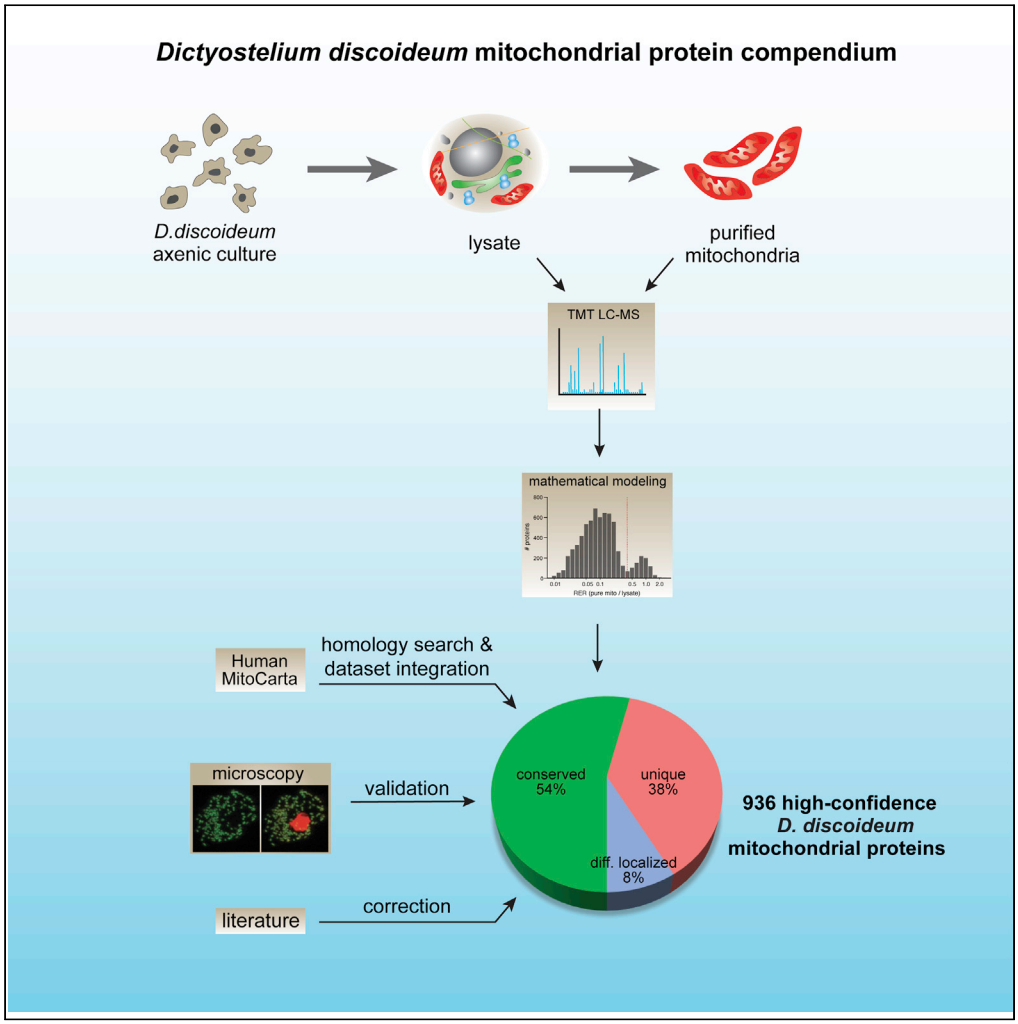
iScience

CellPress
OPEN ACCESS

Article

Generation of a mitochondrial protein compendium in *Dictyostelium discoideum*



Anna V. Freitas, Jake T. Herb, Miao Pan, Yong Chen, Marjan Gucek, Tian Jin, Hong Xu

hong.xu@nih.gov

Highlights

Identification of 936 high-confidence mitochondrial proteins in *D. discoideum*

Half of mitochondrial proteins are conserved between *Dictyostelium discoideum* and human

Many unique mitochondrial proteins are involved in mtDNA gene expression

Mitochondrial proteins are dynamically expressed during development

Freitas et al., iScience 25, 104332
May 20, 2022
https://doi.org/10.1016/j.isci.2022.104332

iScience

CellPress
OPEN ACCESS

Article

Generation of a mitochondrial protein compendium in *Dictyostelium discoideum*

Anna V. Freitas,[1,3] Jake T. Herb,[1,3] Miao Pan,[2] Yong Chen,[1] Marjan Gucek,[1] Tian Jin,[2] and Hong Xu[1,4,*]

SUMMARY

The social ameba *Dictyostelium discoideum* has emerged as a powerful model to study mitochondrial genetics and bioenergetics. However, a comprehensive inventory of mitochondrial proteins that is critical to understanding mitochondrial processes has yet to be curated. Here, we utilized high-throughput multiplexed protein quantitation and homology analyses to generate a high-confidence mitochondrial protein compendium consisting of 936 proteins. Our proteomic approach, which utilizes mass spectrometry in combination with mathematical modeling, was validated through mitochondrial targeting sequence prediction and live-cell imaging. Our final compendium consists of 936 proteins. Nearly, a third of *D. discoideum* mitochondrial proteins do not have homologs in humans, budding yeasts, or an ancestral alphaproteobacteria. 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 ameba and to understand the functions of conserved mitochondrial proteins in *D. discoideum*.

INTRODUCTION

Dictyostelium discoideum, a social ameba, 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, amebae 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 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* mtDNA has four introns in the *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* contains an additional component compared to its mammalian counterparts: an alternative oxidase (AOX) (Pearce et al., 2019), which is found across eukaryotic clades except vertebrates (McDonald et al., 2009). AOX is highly expressed during vegetative growth, but its expression level is markedly reduced upon starvation, suggesting that metabolic reprogramming potentially occurs during starvation-induced development (Jarmuszkiewicz 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

[1]National Heart, Lung and Blood Institute, National Institute of Health, Bethesda, MD 20892, USA

[2]National Institute of Allergy and Infectious Diseases, National Institute of Health, Rockville, MD 20852, USA

[3]These authors contributed equally

[4]Lead contact

*Correspondence: hong.xu@nih.gov

https://doi.org/10.1016/j.isci.2022.104332

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 (Bolender et al., 2008). 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 Amoebozoa-specific mitochondrial proteins will certainly be missed.

In this study, we combined quantitative proteomics and mathematical modeling to identify over 850 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 936 *D. discoideum* mitochondrial proteins. We also discuss conserved *D. discoideum* mitochondrial proteins that may be used as the basis for 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 1,136 human mitochondrial proteins listed in the Human MitoCarta 3.0 (Rath et al., 2021), and retrieved 616 proteins (Figure 1A, Table S1). This number is much less than known mitochondrial proteins in humans (1,136) and baker's yeast (901) (Morgenstern et al., 2017; Rath et al., 2021). We posited that the mitochondrial proteomes 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 include components of electron transport chain complexes and enzymes in the citric acid cycle (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, reflecting their different enrichments in purified mitochondria (Figure 1B), which allowed us to determine a proper RER threshold for mitochondrial proteins using mathematical modeling. Thus, we proceeded to only analyze the purified mitochondria RER.

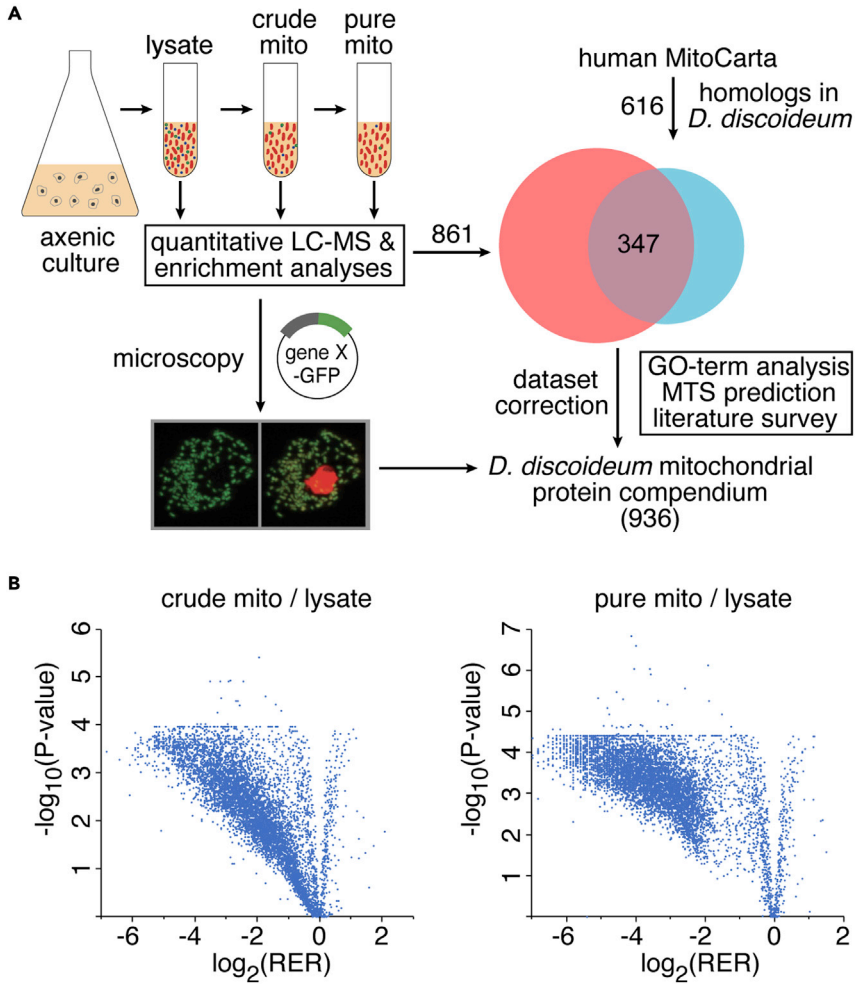


Figure 1. Curation of a comprehensive mitochondrial proteome in *Dictyostelium discoideum*

(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 −log10 (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).

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 mitochondrial 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 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 0.392 (Figure 2D, Table S4), such that all proteins with an RER greater than or equal to this value had a 100% probability of belonging to the mitochondrial cluster. Based on this analysis, a total of 861 proteins were assigned as putative mitochondrial proteins (Table S5).

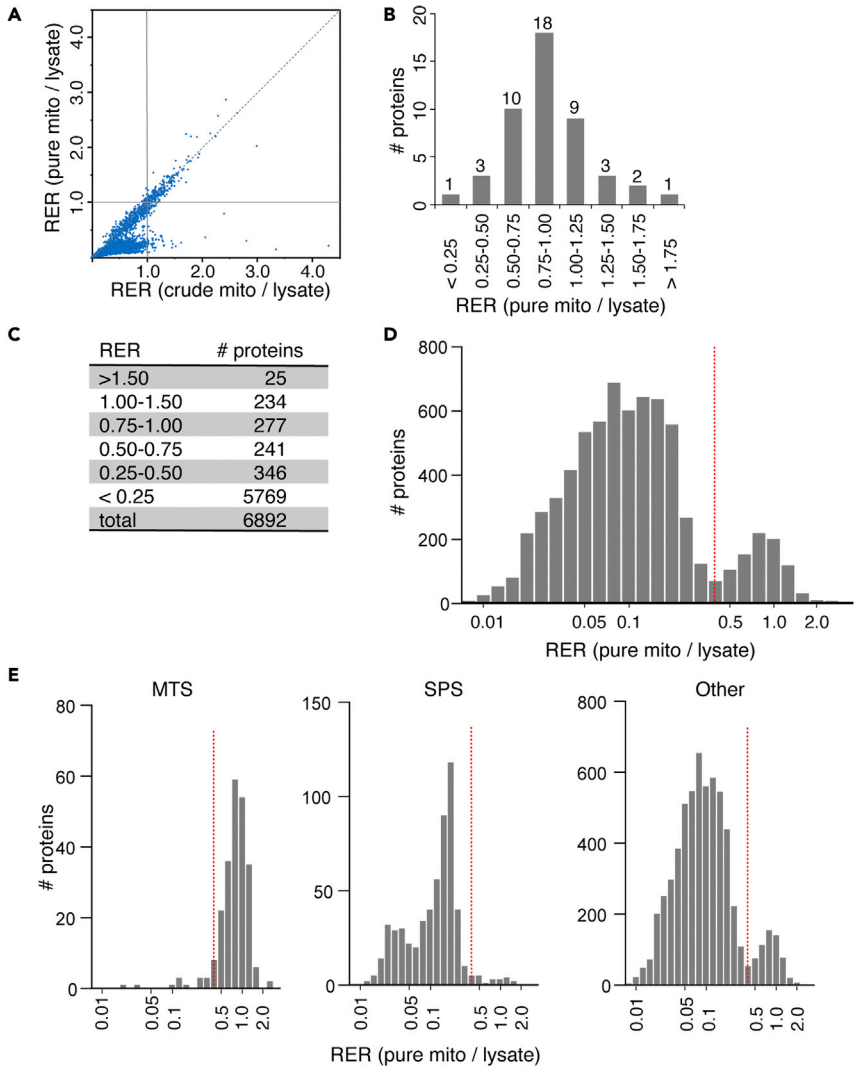


Figure 2. Prediction of mitochondrial localization based on relative enrichment analysis

(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) Distribution of core *D. discoideum* mitochondrial ETC, OXPHOS, and TCA cycle proteins (n = 47) based on RER.

(C) Summary table of pure mitochondrial relative enrichment ratios for all proteins in the protein discovery experiment.

(D) The distribution of mitochondrial (n = 861) and non-mitochondrial proteins (n = 6031) based on RER follows a Gaussian mixture model, in which a cutoff value (red, RER = 0.392) separates two normal curves.

(E) An RER cutoff of 0.392 (red) effectively separates proteins based on the distribution of proteins with a predicted mitochondrial targeting sequence (MTS, n = 235) or, as a negative control, a signal peptide sequence (SPS, n = 565). The clustering also fits the distribution of proteins with no sequence prediction (Other, n = 6092).

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 (Backes and Herrmann, 2017). Overall, 3.4% of all proteins recovered in the proteomics experiment contain an MTS (Table S5). Importantly, we recovered 94% of these MTS-bearing proteins using the RER cutoff of 0.392 (Figure 2E, MTS). As a negative control, 2.2% of proteins that were destined to other organelles such as the ER, Golgi, lysosomes, vacuoles, or secretory pathway (SP) had an RER greater than or equal to 0.392 (Figure 2E, SPS). Furthermore, all remaining proteins with no

sequence prediction also fit into these clusters (Figure 2E, other). These analyses demonstrate that an RER cutoff value of 0.392 effectively separates mitochondrial proteins from non-mitochondrial proteins.

We also surveyed the localization of 81 proteins recovered in LC-MS (Table S6), 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, 91% of proteins with an RER higher than 0.392 showed complete or partial mitochondrial localization (Figures 3A and 3C), demonstrating a strong positive correlation between RER value and probability of mitochondrial localization (Figure 3A). Moreover, logistic regression analysis on the localization pattern of these 81 proteins predicts that an individual protein has more than a 79% probability of localizing to the mitochondria if its RER is higher than 0.392 (Figure 3B). Although the microscopy analysis demonstrated that few non-mitochondrial proteins were recovered in our compendium, 12% of proteins with an RER less than 0.392 showed mitochondrial localization (Figure 3A).

A comprehensive mitochondrial protein compendium in *D. discoideum*

To further improve the coverage and accuracy of our mitochondrial protein discovery, we first revised the proteomics list based on the *in vivo* microscopy validation. We removed three non-mitochondrial localizing proteins and added five mitochondrial localizing proteins.

We also examined the proteins with a 95% or greater predicted probability of localizing to the mitochondria based on the GMM (RER Table S4). For these 66 proteins, we retrieved five that had a predicted MTS and/or were homologs of human mitochondrial proteins (Tables S1 and S5). Additionally, there were 41 *D. discoideum* homologs of human mitochondrial proteins that were not captured in LC-MS. For these 46 proteins, we removed any that demonstrated more sequence homology to a non-mitochondrial human protein than a protein in the Human MitoCarta3.0. We posit that these *D. discoideum* proteins may share a conserved domain, and therefore emerged in the homology analyses, but do not share the same function. We also removed any proteins that were non-mitochondrial based on previous evidence. We added the remaining 21 proteins to the list, as well as 52 proteins with mitochondrial gene ontologies that had not emerged during the proteomic or homology analysis. The final compendium consists of 936 high-confidence mitochondrial proteins in *D. discoideum* (Table S7).

Characterization of the *D. discoideum* mitochondrial proteome

Out of the 936 *D. discoideum* mitochondrial proteins, 504 and 404 proteins have homologs in the mitochondrial proteome of humans and *Saccharomyces cerevisiae*, respectively (Table 1). Only 286 *D. discoideum* mitochondrial proteins have homologs in *Rickettsia prowazekii* (Table 1), an α-proteobacterium that is closely related to the mitochondrial ancestor. Overall, a total of 300 proteins, representing 32.1% of the *D. discoideum* mitochondrial proteome, have no homologs in the whole proteome of humans, *S. cerevisiae*, or *R. prowazekii* (Table S7), indicating that a large fraction of *D. discoideum* mitochondrial proteins were evolved *de novo* after the divergence of Amoebozoa. Moreover, 74 *D. discoideum* mitochondrial proteins (7.9%) have no homologs in *Dictyostelium purpureum* (Table 1), a closely related species of social ameba, further substantiating the fast-evolving nature of the ameba mitochondrial proteome.

There are also 77 *D. discoideum* mitochondrial proteins (8.2%) with human homologs that had not been annotated as mitochondrial proteins (Table 2, Table S7). Among these 77 proteins, 62 were also not annotated as mitochondrial proteins in budding yeast, including 27 that had homologs in *S. cerevisiae*. Given the estimated false-discovery rate of our compendium, the localization of these proteins needs to be experimentally assessed. 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 a predicted MTS 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 two localizes to the mitochondria (Morgenstern et al., 2017), 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.

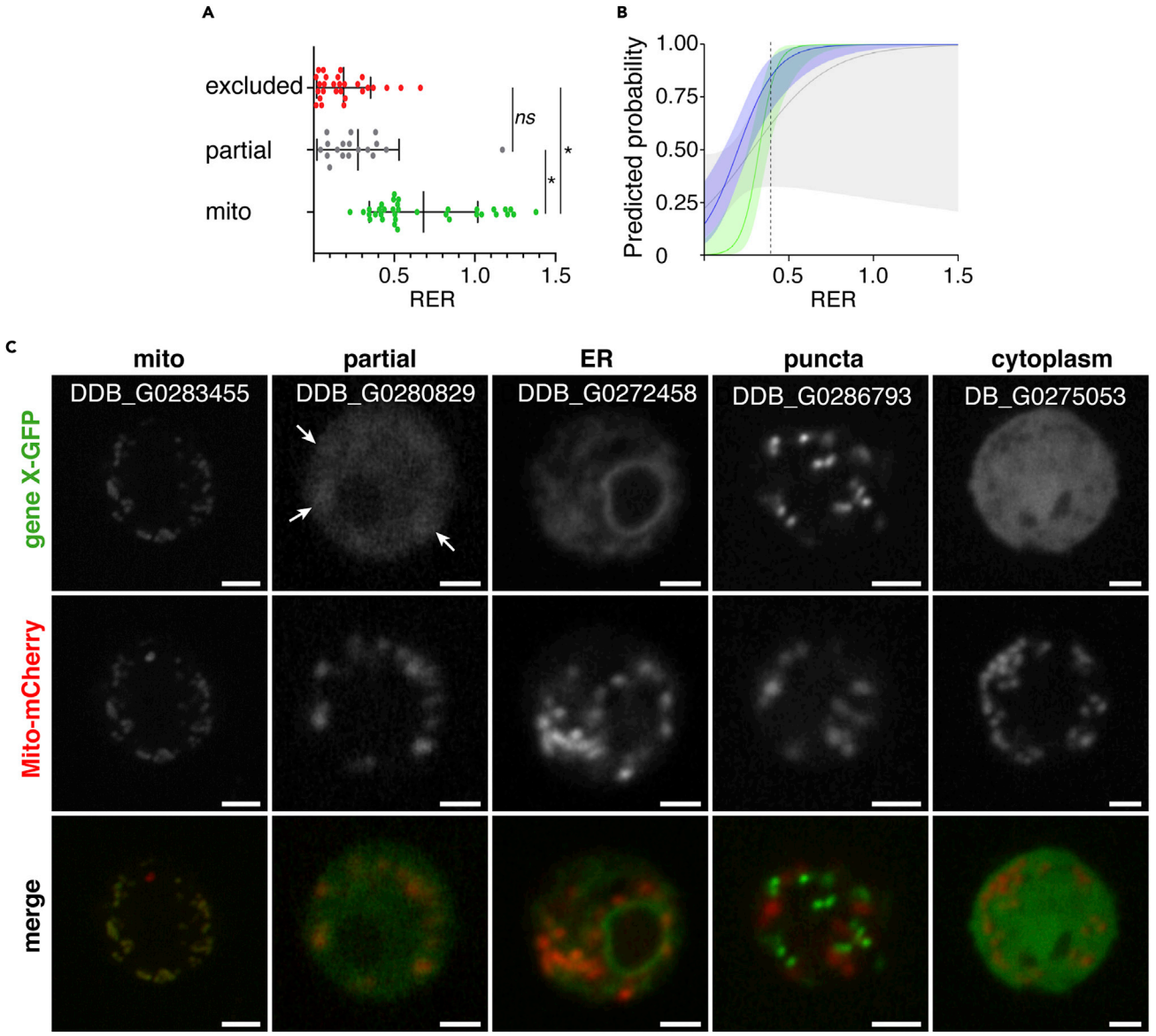


Figure 3. Validation of proteomic discovery and enrichment analyses

(A) Dot plot with interquartile ranges of proteins that had mitochondrial localization (green, n = 35), partial mitochondrial localization (gray, n = 18), or were excluded from the mitochondria (red, n = 28). *, $p \leq 5 \times 10^{-5}$.

(B) Predicted probabilities and confidence intervals as analyzed by logistic regression for mitochondrial (green), partial mitochondrial (gray), and either complete or partial mitochondrial (blue) localization. The dashed line represents the RER cutoff for predicting mitochondrial localization (0.392).

(C) Representative live-cell confocal images of vegetative stage *D. discoideum* expressing GFP-tagged genes of interest leveled with cytochrome oxidase c subunit IV tagged with mCherry (Mito-mCherry). Scale bars: 5.00 μm. Arrowheads denote areas of partial localization.

Additionally, we categorized the *D. discoideum* mitochondrial proteome using PANTHER biological function or protein family classifications (Table 2, Table S7). Proteins involved in mitochondrial gene expression and metabolism comprise the largest fractions of all mitochondrial proteins, over 22% 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 16%, have no classified functions based on PANTHER analyses (Table 2).

Table 1. Comparison of mitochondrial proteome in *D. discoideum* to other organisms

	D. discoideum (936)
H. sapiens (1,136)	504
S. cerevisiae (901)	404
R. prowazekii (832)	286
D. purpureum (?)	862

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*. "?" reflects that the number mitochondrial proteins in *D. purpureum* is currently unknown.

D. discoideum-specific mitochondrial proteins

Different from human mtDNA, *D. discoideum* mitochondrial genome contains several introns and has numerous overlapping genes, which require complex post-transcriptional processing. Consistent with this notion, proteins involved in gene expression consisted of a large fraction of *D. discoideum*-specific mitochondrial proteome (Table 2). On the contrary, few metabolism proteins emerged in the list (Table 2), 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 9.7% and 9.6% of the overall and unique mitochondrial protein compendium, respectively. While mitoribosomes are thought to be evolved from bacterial ribosomes, these two differ greatly regarding their structure, function, as well as their composition of proteins and RNAs. We identified 50 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 S8). Interestingly, *D. discoideum* mitoribosomal proteins belong to families across several taxonomic groups: 35 proteins belong to mammalian mitoribosomal protein families (28 and 39S), one belongs to eukaryotic cytosolic ribosomal protein families (60S), nine belong to yeast mitoribosomal protein families (37 and 54S), two belong to chloroplast or bacterial ribosomal protein families (30 and 50S), one is from archaea, and two are universally conserved among prokaryotes and eukaryotes. In *D. discoideum*, it has previously been shown that cytosolic ribosomes tether to the mitochondrial outer membrane for cotranslational import (Ahmed et al., 2006). Hence, the recovery of the 60S ribosomal protein L6 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 Amoebozoa-specific mechanisms to process mitochondrial transcripts and 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 49 candidate mtDNA and mtRNA processing factors including six endonucleases and a pentatricopeptide repeat (PPR)-containing protein A (PtcA, DDB_G0293882). 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).

Table 2. The functional categorization of all *D. discoideum* mitochondrial proteins

	overall	unique	diff. localized
protein import & biogenesis	25	8	2
protein maturation & folding	24	13	0
protein degradation	26	1	6
DNA-related processes	39	16	8
mRNA-related processes	10	8	1
tRNA-related processes	44	6	2
ribosome & translation	91	34	2
others of gene expression	29	14	1
respiratory chain assembly	41	11	0
respiratory chain subunits	62	14	2
energy metabolism	44	6	2
lipid metabolism	45	5	5
nucleotide metabolism	5	0	0
amino acid metabolism	37	2	2
coenzyme/cofactor biosynthesis	17	1	1
Fe/S proteins biosynthesis	15	0	0
carriers & channels	44	5	0
regulation & signaling	26	6	10
morphology & dynamics	25	3	6
homeostasis & stress response	29	6	1
developmental processes	14	5	2
Others	95	63	11
Unclassified	150	128	13
Total	936	355	77

The number of proteins in each biological process for all mitochondrial proteins (overall, n = 936), those that lack human homologs (unique, n = 355), and those that have human homologs, but homologs were not annotated as mitochondrial proteins (diff. localized, n = 77) are indicated.

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. Likewise, HGSA, one of the two HMG-CoA synthases, also emerged as a mitochondrial protein. Additionally, the isozymes of HMGR2 and HGSA—HMGR1 and HGSB, respectively—were not recovered in our compendium and have predicted signal peptide sequences. Taken together, this suggests that mevalonate metabolism may redundantly take place in the mitochondria in addition to in the ER 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 displays no defect in vegetative growth but has 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 starvation, 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 (Table S9; 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-h developmental time course (Figure 4A). A similar pattern is observed in proteins that are involved in mitochondrial DNA maintenance and gene expression including many ribosomal proteins. Interestingly, despite the decrease in gene expression machinery (Figures 4B and 4C), 80% of the mitochondria-encoded

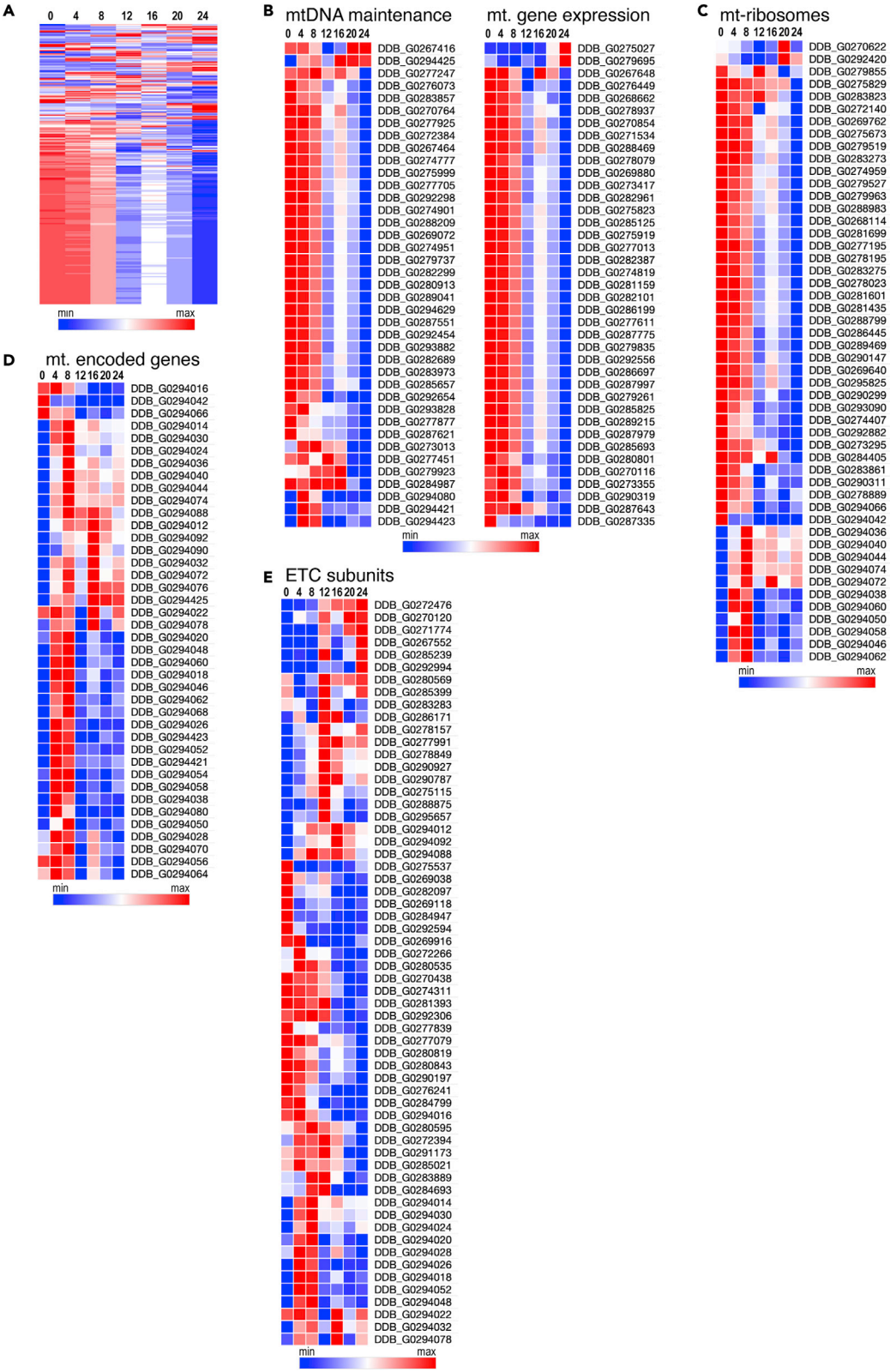

A
0 4 8 12 16 20 24
min
max
B
mtDNA maintenance
0 4 8 12 16 20 24
DDB_G0267416
DDB_G0294425
DDB_G0277247
DDB_G0276073
DDB_G0283857
DDB_G0270764
DDB_G0277925
DDB_G0272384
DDB_G0267464
DDB_G0274777
DDB_G0275999
DDB_G0277705
DDB_G0292298
DDB_G0274901
DDB_G0288209
DDB_G0269072
DDB_G0274951
DDB_G0279737
DDB_G0282299
DDB_G0280913
DDB_G0289041
DDB_G0294629
DDB_G0287551
DDB_G0292454
DDB_G0293882
DDB_G0282689
DDB_G0283973
DDB_G0285657
DDB_G0292654
DDB_G0293828
DDB_G0277877
DDB_G0287621
DDB_G0273013
DDB_G0277451
DDB_G0279923
DDB_G0284987
DDB_G0294080
DDB_G0294421
DDB_G0294423
min
max
mt. gene expression
0 4 8 12 16 20 24
DDB_G0275027
DDB_G0279695
DDB_G0267648
DDB_G0276449
DDB_G0268662
DDB_G0278937
DDB_G0270854
DDB_G0271534
DDB_G0288469
DDB_G0278079
DDB_G0269880
DDB_G0273417
DDB_G0282961
DDB_G0275823
DDB_G0285125
DDB_G0275919
DDB_G0277013
DDB_G0282387
DDB_G0274819
DDB_G0281159
DDB_G0282101
DDB_G0286199
DDB_G0277611
DDB_G0287775
DDB_G0279835
DDB_G0292556
DDB_G0286697
DDB_G0287997
DDB_G0279261
DDB_G0285825
DDB_G0289215
DDB_G0287979
DDB_G0285693
DDB_G0280801
DDB_G0270116
DDB_G0273355
DDB_G0290319
DDB_G0287643
DDB_G0287335
C
mt-ribosomes
0 4 8 12 16 20 24
DDB_G0270622
DDB_G0292420
DDB_G0279855
DDB_G0275829
DDB_G0283823
DDB_G0272140
DDB_G0269762
DDB_G0275673
DDB_G0279519
DDB_G0283273
DDB_G0274959
DDB_G0279527
DDB_G0279963
DDB_G0288983
DDB_G0268114
DDB_G0281699
DDB_G0277195
DDB_G0278195
DDB_G0283275
DDB_G0278023
DDB_G0281601
DDB_G0281435
DDB_G0288799
DDB_G0286445
DDB_G0289469
DDB_G0290147
DDB_G0269640
DDB_G0295825
DDB_G0290299
DDB_G0293090
DDB_G0274407
DDB_G0292882
DDB_G0273295
DDB_G0284405
DDB_G0283861
DDB_G0290311
DDB_G0278889
DDB_G0294066
DDB_G0294042
DDB_G0294036
DDB_G0294040
DDB_G0294044
DDB_G0294074
DDB_G0294072
DDB_G0294038
DDB_G0294060
DDB_G0294050
DDB_G0294058
DDB_G0294046
DDB_G0294062
min
max
D
mt. encoded genes
0 4 8 12 16 20 24
DDB_G0294016
DDB_G0294042
DDB_G0294066
DDB_G0294014
DDB_G0294030
DDB_G0294024
DDB_G0294036
DDB_G0294040
DDB_G0294044
DDB_G0294074
DDB_G0294088
DDB_G0294012
DDB_G0294092
DDB_G0294090
DDB_G0294032
DDB_G0294072
DDB_G0294076
DDB_G0294425
DDB_G0294022
DDB_G0294078
DDB_G0294020
DDB_G0294048
DDB_G0294060
DDB_G0294018
DDB_G0294046
DDB_G0294062
DDB_G0294068
DDB_G0294026
DDB_G0294423
DDB_G0294052
DDB_G0294421
DDB_G0294054
DDB_G0294058
DDB_G0294038
DDB_G0294080
DDB_G0294050
DDB_G0294028
DDB_G0294070
DDB_G0294056
DDB_G0294064
min
max
E
ETC subunits
0 4 8 12 16 20 24
DDB_G0272476
DDB_G0270120
DDB_G0271774
DDB_G0267552
DDB_G0285239
DDB_G0292994
DDB_G0280569
DDB_G0285399
DDB_G0283283
DDB_G0286171
DDB_G0278157
DDB_G0277991
DDB_G0278849
DDB_G0290927
DDB_G0290787
DDB_G0275115
DDB_G0288875
DDB_G0295657
DDB_G0294012
DDB_G0294092
DDB_G0294088
DDB_G0275537
DDB_G0269038
DDB_G0282097
DDB_G0269118
DDB_G0284947
DDB_G0292594
DDB_G0269916
DDB_G0272266
DDB_G0280535
DDB_G0270438
DDB_G0274311
DDB_G0281393
DDB_G0292306
DDB_G0277839
DDB_G0277079
DDB_G0280819
DDB_G0280843
DDB_G0290197
DDB_G0276241
DDB_G0284799
DDB_G0294016
DDB_G0280595
DDB_G0272394
DDB_G0291173
DDB_G0285021
DDB_G0283889
DDB_G0284693
DDB_G0294014
DDB_G0294030
DDB_G0294024
DDB_G0294020
DDB_G0294028
DDB_G0294026
DDB_G0294018
DDB_G0294052
DDB_G0294048
DDB_G0294022
DDB_G0294032
DDB_G0294078
min
max

Figure 4. Expression profile of mitochondrial genes during *D. discoideum* development

Heatmaps showing normalized RNA-seq analysis (*Z* score) values of (A) all mitochondrial proteins within the compendium and (B–D) specific mitochondrial proteins during the 24-h 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.

genes in the dataset (32 of 40) were upregulated (log2FC $\geq$ 1) in the first 4 h after starvation induction (Figure 4D). Furthermore, in examining all respiratory chain complexes, seven nuclear-encoded ETC subunits had a higher expression level (log2FC $\geq$ 1) at or after 8 h of starvation, besides the 11 nuclear-encoded subunits that show a burst of expression in the first 4 h after the starvation (Figure 4E, Table S9). 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. Future characterization of the mitochondrial proteome at different developmental stages would be valuable to complement these findings, as changes in a gene's expression do not always correspond with its protein level or activity (Kotsifas et al., 2002).

In conclusion, 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 ameba. Additionally, through comparative genomics, our compendium will complement mitochondrial protein discovery in other organisms and may shed light on the evolution of the mitochondrial proteome and processes.

Limitations of the study

To reduce the false discovery of non-mitochondrial proteins, we set a stringent RER cutoff at 0.392 based on the mathematical modeling. Our microscopy-based validation demonstrates that several proteins with RERs less than 0.392 actually localize to mitochondria (Figure 3A), while a few others with RER above this threshold do not (Figure 3A). Although the *in silico* dataset correction improved the coverage of our compendium, some bona fide mitochondrial proteins without homologs in the Human MitoCarta 3.0 that have RER values below the cutoff would be missed in the compendium. In addition, our proteomic discovery method does not effectively discern mitochondria-associated proteins from bona fide mitochondrial proteins that localize to the matrix, intermembrane space, or are inserted in the inner or outer membrane. Future updates to our compendium should examine the submitochondrial localization of the inventoried proteins. Lastly, our mitochondria protein discovery is limited to the vegetative stage of *D. discoideum*. Future characterization of the mitochondrial proteome in other developmental stages would lead to an improved understanding of mitochondrial regulation of *D. discoideum* development and differentiation.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Material availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - *D. discoideum* transformation
  - Protein mass spectrometry
  - Mathematical modeling
  - Bioinformatic analyses
  - Library generation for imaging verification
  - Live-cell imaging
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2022.104332.

ACKNOWLEDGMENTS

We thank E. D. Korn and S. Shu for reagents and advice on *D. discoideum* culture; R. S. Balaban, R. C. Garcia, and L. Edwards for advice on mitochondrial isolation. Funding: This work is supported by NHLBI Intramural Research Program.

AUTHOR CONTRIBUTIONS

Conceptualization, H.X.; Methodology, A.V.F., J.T.H., and H.X.; Software, A.V.F.; Validation, A.V.F.; Formal Analysis, A.V.F.; Investigation, J.T.H., M.P., and Y.C.; Resources, M.G., T.J., and H.X.; Data Curation, A.V.F. and Y.C.; Writing – Original Draft, A.V.F. and H.X.; Writing – Review & Editing, A.V.F. and H.X.; Visualization, A.V.F. and H.X.; Supervision, M.G., T.J., and H.X.; Project Administration, M.G., T.J., and H.X.; Funding Acquisition, H.X.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 10, 2022
Revised: March 29, 2022
Accepted: April 26, 2022
Published: May 20, 2022

REFERENCES

Ahmed, A.U., Beech, P.L., Lay, S.T., Gilson, P.R., and Fisher, P.R. (2006). Import-associated translational inhibition: Novel *in vivo* evidence for cotranslational protein import into *Dictyostelium discoideum* mitochondria. Eukaryot. Cell *5*, 1314–1327. https://doi.org/10.1128/ec.00386-05.

Almagro Armenteros, J.J., 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. https://doi.org/10.26508/lsa.201900429.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic Local Alignment Search Tool. J. Mol. Biol. *8*, 403–410.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402. https://doi.org/10.1093/nar/25.17.3389.

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. https://doi.org/10.1007/bf00326157.

Backes, S., and Herrmann, J.M. (2017). Protein Translocation into the Inter- membrane Space and Matrix of Mitochondria: Mechanisms and Driving Forces. Front. Mol. Biosci. *4*, 83.

Baldauf, S.L., and Doolittle, W.F. (1997). Origin and evolution of the slime molds (Mycetozoa). Proceedings of the National Academy of Sciences *94*, 12007–12012.

Bolender, N., Sickmann, A., Wagner, R., Meisinger, C., and Pfanner, N. (2008). Multiple pathways for sorting mitochondrial precursor proteins. EMBO Rep. *9*, 42–49. https://doi.org/10.1038/sj.embor.7401126.

Bozzaro, S. (2013). The model organism Dictyostelium discoideum. In Dictyostelium Discoideum Protocols, L. Eichinger, F. Rivero, and N.J. Totowa, eds. (Humana Press), pp. 17–37.

Burg, J.S., and Espenshade, P.J. (2011). Regulation of HMG-CoA reductase in mammals and yeast. Prog. Lipid Res. *50*, 403–410. https://doi.org/10.1016/j.plipres.2011.07.002.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinformatics *10*, 421. https://doi.org/10.1186/1471-2105-10-421.

Carbon, S., Ireland, A., Mungall, C.J., Shu, S., Marshall, B., and Lewis, S.; The AmiGO hub; The web presence working group (2009). AmiGO: online access to ontology and annotation data. Bioinformatics *25*, 288–289. https://doi.org/10.1093/bioinformatics/btn615.

Chida, J., Yamaguchi, H., Amagai, A., and Maeda, Y. (2004). The necessity of mitochondrial genome DNA for normal development of Dictyostelium cells. J. Cell Sci. *117*, 3141–3152. https://doi.org/10.1242/jcs.01140.

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. https://doi.org/10.1186/1471-2156-9-25.

Chin, D.J., Gil, G., Russell, D.W., Liscum, L., Luskey, K.L., Basu, S.K., Okayama, H., Berg, P., Goldstein, J.L., and Brown, M.S. (1984). Nucleotide sequence of 3-hydroxy-3-methyl-glutaryl coenzyme A reductase, a glycoprotein of endoplasmic reticulum. Nature *308*, 613–617. https://doi.org/10.1038/308613a0.

Cho, Y., Qiu, Y.-L., Kuhlman, P., and Palmer, J.D. (1998). Explosive invasion of plant mitochondria by a group I intron. Proc. Natl. Acad. Sci. *95*, 14244–14249. https://doi.org/10.1073/pnas.95.24.14244.

Cole, R.A., Slade, M.B., and Williams, K.L. (1995). Dictyostelium discoideum mitochondrial DNA encodes a NADH:Ubiquinone oxidoreductase subunit which is nuclear encoded in other eukaryotes. J. Mol. Evol. *40*, 616–621. https://doi.org/10.1007/bf00160509.

Dayon, L., Hainard, A., Licker, V., Turck, N., Kuhn, K., Hochstrasser, D.F., Burkhard, P.R., 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. https://doi.org/10.1021/ac702422x.

Fey, P., Kowal, A.S., Gaudet, P., Pilcher, K.E., and Chisholm, R.L. (2007). Protocols for growth and development of Dictyostelium discoideum. Nat. Protoc. *2*, 1307–1316. https://doi.org/10.1038/nprot.2007.178.

Francione, L.M., Annesley, S.J., Carilla-Latorre, S., Escalante, R., and Fisher, P.R. (2011). The Dictyostelium model for mitochondrial disease. Semin. Cell Develop. Biol. *22*, 120–130. https://doi.org/10.1016/j.semcdb.2010.11.004.

Gaudet, P., Pilcher, K.E., Fey, P., and Chisholm, R.L. (2007). Transformation of Dictyostelium discoideum with plasmid DNA. Nat. Protoc. *2*, 1317–1324. https://doi.org/10.1038/nprot.2007.179.

Glancy, B., and Balaban, R.S. (2011). Protein composition and function of red and white skeletal muscle mitochondria. Am. J. Physiology-Cell Physiol. *300*, C1280–C1290. https://doi.org/10.1152/ajpcell.00496.2010.

Gold, V.A., 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. https://doi.org/10.15252/embr.201744261.

Graham, J.M. (1999). Purification of a crude mitochondrial fraction by density-gradient centrifugation. Curr. Protoc. Cell Biol. *4*, 3–4. https://doi.org/10.1002/0471143030.cb0304s04.

Havird, J.C., Whitehill, N.S., Snow, C.D., and Sloan, D.B. (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. https://doi.org/10.1111/evo.12808.

He, Y., Chen, Y., Morris, D.L., Lee, D.-Y., and Tjandra, N. (2020). Bax expression is optimal at low oxygen tension and constant agitation. Protein Expression and Purification *165*, 105501.

Jarmuszkiewicz, W., Behrendt, M., Navet, R., and Sluse, F.E. (2002). Uncoupling protein and alternative oxidase of *Dictyostelium discoideum*: occurrence, properties and protein expression during vegetative life and starvation-induced early development. FEBS Lett. *532*, 459–464. https://doi.org/10.1016/s0014-5793(02)03734-1.

Jukes, T.H., and Osawa, S. (1990). The genetic code in mitochondria and chloroplasts. Experientia *46*, 1117–1126. https://doi.org/10.1007/bf01936921.

Kay, R.R. (1982). cAMP and spore differentiation in Dictyostelium discoideum. Proc. Natl. Acad. Sci. *79*, 3228–3231. https://doi.org/10.1073/pnas.79.10.3228.

Keller, G.A., 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. J. Cell Biol. *103*, 875–886. https://doi.org/10.1083/jcb.103.3.875.

Kelly, B., Carrizo, G.E., Edwards-Hicks, J., Sanin, D.E., Stanczak, M.A., Priesnitz, C., Flachsmann, L.J., Curtis, J.D., Mittler, G., Musa, Y., et al. (2021). Sulfur sequestration promotes multicellularity during nutrient limitation. Nature *591*, 471–476. https://doi.org/10.1038/s41586-021-03270-3.

Kessin, R.H. (2001). Dictyostelium: Evolution, Cell Biology, and the Development of Multicellularity (Cambridge University Press).

Kotsifas, M., Barth, C., De Lozanne, A., Lay, S.T., and Fisher, P.R. (2002). Chaperonin 60 and mitochondrial disease in Dictyostelium. J. Muscle Res. Cell Motil. *23*, 839–852. https://doi.org/10.1023/a:1024444215766.

Li, Y., Zhang, R., Liu, S., Donath, A., Peters, R.S., Ware, J., Misof, B., Niehuis, O., Pfrender, M.E., and Zhou, X. (2017). The molecular evolutionary dynamics of oxidative phosphorylation (OXPHOS) genes in Hymenoptera. BMC Evol. Biol. *17*, 269. https://doi.org/10.1186/s12862-017-1111-z.

Lightowlers, R.N., and Chrzanowska-Lightowlers, Z.M. (2013). Human pentatricopeptide proteins: only a few and what do they do? RNA Biol. *10*, 1433–1438. https://doi.org/10.4161/rna.24770.

Lurin, C., Andrees, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyère, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B., 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. https://doi.org/10.1105/tpc.104.022236.

Manna, S. (2015). An overview of pentatricopeptide repeat proteins and their applications. Biochimie *113*, 93–99. https://doi.org/10.1016/j.biochi.2015.04.004.

Manna, S., Brewster, J., and Barth, C. (2013). Identification of pentatricopeptide repeat proteins in the model organism *Dictyostelium discoideum*. Int. J. Genomics *2013*, 1–8. https://doi.org/10.1155/2013/586498.

Mazur, M., Wojciechowska, D., Sitkiewicz, E., Malinowska, A., Świderska, B., Kmita, H., and Wojtkowska, M. (2021). Mitochondrial processes during early development of Dictyostelium discoideum: from bioenergetic to proteomic studies. Genes *12*, 638. https://doi.org/10.3390/genes12050638.

Morgenstern, M., Stiller, S.B., Lubbert, P., Peikert, C.D., Dannenmaier, S., Drepper, F., Weill, U., HoB, P., Feuerstein, R., Gebert, M., et al. (2017). Definition of a high-confidence mitochondrial proteome at quantitative scale. Cell Rep. *19*, 2836–2852. https://doi.org/10.1016/j.celrep.2017.06.014.

Ogawa, S., Yoshino, R., Angata, K., Iwamoto, M., Pi, M., Kuroe, K., Matsuo, K., Morio, T., Urushihara, H., Yanagisawa, K., and Tanaka, Y. (2000). The mitochondrial DNA of Dictyostelium discoideum: complete sequence, gene content and genome organization. Mol. Gen. Genet. *263*, 514–519. https://doi.org/10.1007/pl00008685.

Pagliarini, D.J., Calvo, S.E., Chang, B., Sheth, S.A., Vafai, S.B., Ong, S.E., Walford, G.A., Sugiana, C., Boneh, A., Chen, W.K., et al. (2008). A mitochondrial protein compendium elucidates complex I disease biology. Cell *134*, 112–123. https://doi.org/10.1016/j.cell.2008.06.016.

Parikh, A., Miranda, E.R., Katoh-Kurasawa, M., Fuller, D., Rot, G., Zagar, L., Curk, T., Sucgang, R., Chen, R., Zupan, B., et al. (2010). Conserved developmental transcriptomes in evolutionarily divergent species. Genome Biol. *11*, R35. https://doi.org/10.1186/gb-2010-11-3-r35.

Pearce, X.G., Annesley, S.J., and Fisher, P.R. (2019). The Dictyostelium model for mitochondrial biology and disease. Int. J. Dev. Biol. *63*, 497–508. https://doi.org/10.1387/ijdb.190233pf.

Pearson, W.R. (2013). An Introduction to Sequence Similarity ("Homology") Searching. Current Protocols in Bioinformatics, 42.

Rath, S., Sharma, R., Gupta, R., Ast, T., Chan, C., Durham, T.J., Goodman, R.P., Grabarek, Z., Haas, M.E., Hung, W.H.W., et al. (2021). MitoCarta3.0: an updated mitochondrial proteome now with sub-organelle localization and pathway annotations. Nucleic Acids Res. *49*, D1541–D1547. https://doi.org/10.1093/nar/gkaa1011.

Sloan, D.B., Triant, D.A., Wu, M., and Taylor, D.R. (2014). Cytonuclear interactions and relaxed selection accelerate sequence evolution in organelle ribosomes. Mol. Biol. Evol. *31*, 673–682. https://doi.org/10.1093/molbev/mst259.

Stajdohar, M., Rosengarten, R.D., 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. https://doi.org/10.1186/s12859-017-1706-9.

The Gene Ontology Consortium, Carbon, S., Douglass, E., Good, B.M., Unni, D.R., Harris, N.L., Mungall, C.J., Basu, S., Chisholm, R.L., Dodson, R.J., et al. (2021). The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res. *49*, D325–D334. https://doi.org/10.1093/nar/gkaa1113.

Veltman, D.M., Akar, G., Bosgraaf, L., and Van Haastert, P.J.M. (2009). A new set of small, extrachromosomal expression vectors for Dictyostelium discoideum. Plasmid *61*, 110–118. https://doi.org/10.1016/j.plasmid.2008.11.003.

Vögtle, F.-N., Wortelkamp, S., Zahedi, R.P., Becker, D., Leidhold, C., Gevaert, K., Kellermann, J., Voos, W., Sickmann, A., Pfanner, N., and Meisinger, C. (2009). Global analysis of the mitochondrial N-proteome identifies a processing peptidase critical for protein stability. Cell *139*, 428–439. https://doi.org/10.1016/j.cell.2009.07.045.

Yan, Z., Ye, G., and Werren, J.H. (2019). Evolutionary rate correlation between mitochondrial-encoded and mitochondria-associated nuclear-encoded proteins in insects. Mol. Biol. Evol. *36*, 1022–1036. https://doi.org/10.1093/molbev/msz036.

Yang, F., Shen, Y., Camp, D.G., and Smith, R.D. (2012). High-pH reversed-phase chromatography with fraction concatenation for 2D proteomic analysis. Expert Rev. Proteomics *9*, 129–134. https://doi.org/10.1586/epr.12.15.

Yu, Y. (2022). mixR: an R package for finite mixture modeling for both raw and binned data. J. Open Source Softw. *7*, 4031. https://doi.org/10.21105/joss.04031.

McDonald, A.E., Vanlerberghe, G.C., and Staples, J.F. (2009). Alternative oxidase in animals: unique characteristics and taxonomic distribution. Journal of Experimental Biology *212*, 2627–2634.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
5α Competent *E. coli* (High Efficiency)	New England Biolabs	Cat#C2987U
Chemicals, peptides, and recombinant proteins		
HL5 medium including glucose	Formedium	Cat#HLG0102
Genectin Selective Antibiotic	Thermo Fisher Scientific	Cat#10131035
Blasticidin S HCl	Thermo Fisher Scientific	Cat#A1113902
Percoll density gradient media	Cytiva	Cat#17089101
Sodium Phosphate Dibasic	Sigma-Aldrich	Cat#S0876
Potassium Phosphate Monobasic	Sigma-Aldrich	Cat#P5655
Calcium Chloride 2M	Quality Biologicals	Cat#351-130-721
Magnesium Chloride 1M	Quality Biologicals	Cat#351-033-721
Urea	Sigma-Aldrich	Cat#U5128
Critical commercial assays		
Mitochondrial Isolation Kit for Cultured Cells	Thermo Fisher Scientific	Cat#89874
Protein Assay Kit	Bio-Rad	Cat#5000001
TMT10plex Isobaric Label Reagent Set	Thermo Fisher Scientific	Cat#90110
In-Fusion HD Cloning Kit	Takara Bio USA	Cat#638920
QIAprep Spin Miniprep Kit	Qiagen	Cat#27106X4
Zymoclean Gel DNA Recovery Kit	Zymo Research	Cat#D4002
Deposited data		
Raw and analyzed proteomic data	This paper	PRIDE ProteomeXchange: PXD029101
Code for statistics and mathematical modeling	This paper	Zenodo: http://doi.org/10.5281/zenodo.6471311
Original images and supplementary tables	This paper	Mendeley Data: https://doi.org/10.17632/yhjnbnsgz3.1
Experimental models: Organisms/strains		
D. discoideum: AX2 Parent strain	Jin lab, National Institute of Allergy and Infectious Diseases, NIH	N/A
Oligonucleotides		
Primers for cloning	This paper, Integrated DNA Technology	Table S10
Primers for sequencing	This paper, Integrated DNA Technology	Table S10
Recombinant DNA		
GFP plasmid	dictyBase	pDM323
mCherry plasmid	dictyBase	pDM326
Gene libraries for verification	This paper, Gene Universal	Accession numbers in Table S6
Software and algorithms		
Adobe Photoshop and Illustrator	Adobe	http://www.adobe.com/creativecloud.html
AmiGO	Carbon et al. (2009)	http://amigo.geneontology.org
BLAST+	Camacho et al., 2009; Altschul et al., 1990, 1997	http://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
dictyExpress (RNA sequencing data)	Stajdohar et al., 2017; Parikh et al., 2010	http://dictyexpress.research.bcm.edu/landing/
GraphPad Prism 9	GraphPad	http://graphpad.com
HMMER v3.3.2	Howard Hughes Medical Institute	http://http://hmmer.org
ImageJ	National Institutes of Health	http://imagej.nih.gov/ij/
mixR	Yu (2022)	http://github.com/GaryBAYLOR/mixR.git
Morpheus	The Broad Institute	http://software.broadinstitute.org/morpheus
PANTHER	The Gene Ontology Consortium	http://pantherdb.org
Proteome Discoverer v2.4	Thermo Fisher Scientific	http://www.thermofisher.com/us/en/home/industrial/mass-spectrometry/liquid-chromatography-mass-spectrometry-lc-ms/lc-ms-software/multi-omics-data-analysis/proteome-discoverer-software.html
RStudio v1.4.1717	RStudio	http://www.rstudio.com
TargetP-2.0	Almagro Armenteros et al., 2019	http://services.healthtech.dtu.dk/service.php?TargetP-2.0
Volocity Acquisition	Quorum Technologies Inc.	http://www.volocity4d.com/volocity-acquisition

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Hong Xu (hong.xu@nih.gov).

Material availability

All plasmids generated in this study are available upon request.

Data and code availability

- Proteomics data have been deposited at ProteomeXchange (PXD029101). Original microscopy images and supplementary tables have been deposited at Mendeley (10.17632/yhjnbnsgz3.1). All data is publicly available as of the date of publication.
- All code and data used in the analysis are available on GitHub (10.5281/zenodo.6471311).
- This paper also analyzes existing, publicly available RNA sequencing data retrieved from DictyExpress (see key resources table: Software and algorithms).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

D. discoideum AX2 cells were obtained from the Jin lab and originally derived from the Carole Parent lab, National Cancer Institute, NIH. Cultures were maintained in HL5 medium (Formedium) at 22°C (Fey et al., 2007).

METHOD DETAILS

D. discoideum transformation

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 min. Subsequently, cells were transferred from the cuvette with 2 mL of HL5 and plated onto 12-well tissue culture plates. After 24 h, transformants were selected with Genectin and/or Blasticidin S (Thermo Fisher Scientific, 10 μg/mL each) in HL5.

Protein mass spectrometry

Mitochondrial isolation

Cells were harvested at a concentration of 1-3 x 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 Scientific) on ice. Cell lysis and crude mitochondrial preparation were performed as previously described with modifications (Glancy and Balaban, 2011; Graham, 1999). 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 min, 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 Na_2HPO_4, 5 mM KH_2PO_4, 1 mM $CaCl_2$, 2 mM $MgCl_2$, pH = 6.5) solution (8 mL, 30% Percoll) in a 10 mL ultracentrifuge tube. Ultracentrifugation (68,000 x g, 40 min) 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 min, 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 min, 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.

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 two-component Gaussian mixture model (GMM) in R (Yu, 2022). All proteins with an RER greater or equal to the cutoff value (RER = 0.392) have a 100% posterior probability of belonging to the mitochondrial cluster (Table S4).

Bioinformatic analyses

Homology analyses

Protein sequence homology was established by BLAST+ (Altschul et al., 1990, 1997; Camacho et al., 2009) 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 by 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. Proteins within this list were integrated into the final mitochondrial compendium if they also had a 95% or greater predicted probability of localizing to the mitochondria based on the mixture modeling (Table S4), unless one of the following criteria were met. First, the protein had a human homolog that was non-mitochondrial with a higher degree of sequence identity, indicating that the *D. discoideum* protein likely had a similar domain as the human MitoCarta protein but had a different function. Second, the literature indicated that the protein was nonmitochondrial or the homology evidence was a poor indicator of a similar function. Additionally, *D. discoideum* proteins annotated with the gene ontology term "mitochond*" on AmiGO (Carbon et al., 2009) or dictyBase were selected.

RNA sequencing data normalization and visualization

Normalized RNA-seq data from Parikh et al. (2010) was retrieved using dictyExpress (Stajdohar et al., 2017). For the 936 proteins in the mitochondrial compendium, only 932 corresponding genes were present in the dataset. To compare the gene profiles, data were normalized (z-score) 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 S9) were imported into the matrix visualization software Morpheus (http://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 $\geq$ 1 of non-normalized values.

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 S6). 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 cloned into pDM323 between BglII and SpeI sites using the In-Fusion HD Cloning Kit (Takara Bio USA) and confirmed by sequencing. Fragment or plasmid purification was performed using the Zymoclean Gel DNA Recovery Kit (Zymo Research) or High Efficiency 5α Competent *E. coli* (New England Biolabs) in combination with the QIAprep Spin Miniprep Kit (Qiagen).

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 (Mito-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 S10.

Live-cell imaging

To image cells in the axenic phase, cells (200 μL) 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 min before the media was aspirated. The media was replaced with 1x PBS after three washes (200 μL 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 μm z-step) were analyzed with ImageJ (National Institutes of Health) and formatted in Adobe Photoshop and Adobe Illustrator.

QUANTIFICATION AND STATISTICAL ANALYSIS

All data were presented as the mean ± 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. The fit of each binomial logistic regression was compared via McFadden's pseudo-R^2 (nagelkerke function in R) to determine which predicted probability to report. Predicted probabilities and 95% confidence intervals were calculated (predict function in R) to compare outcomes and were reported in the text for the model with the highest pseudo-R^2.