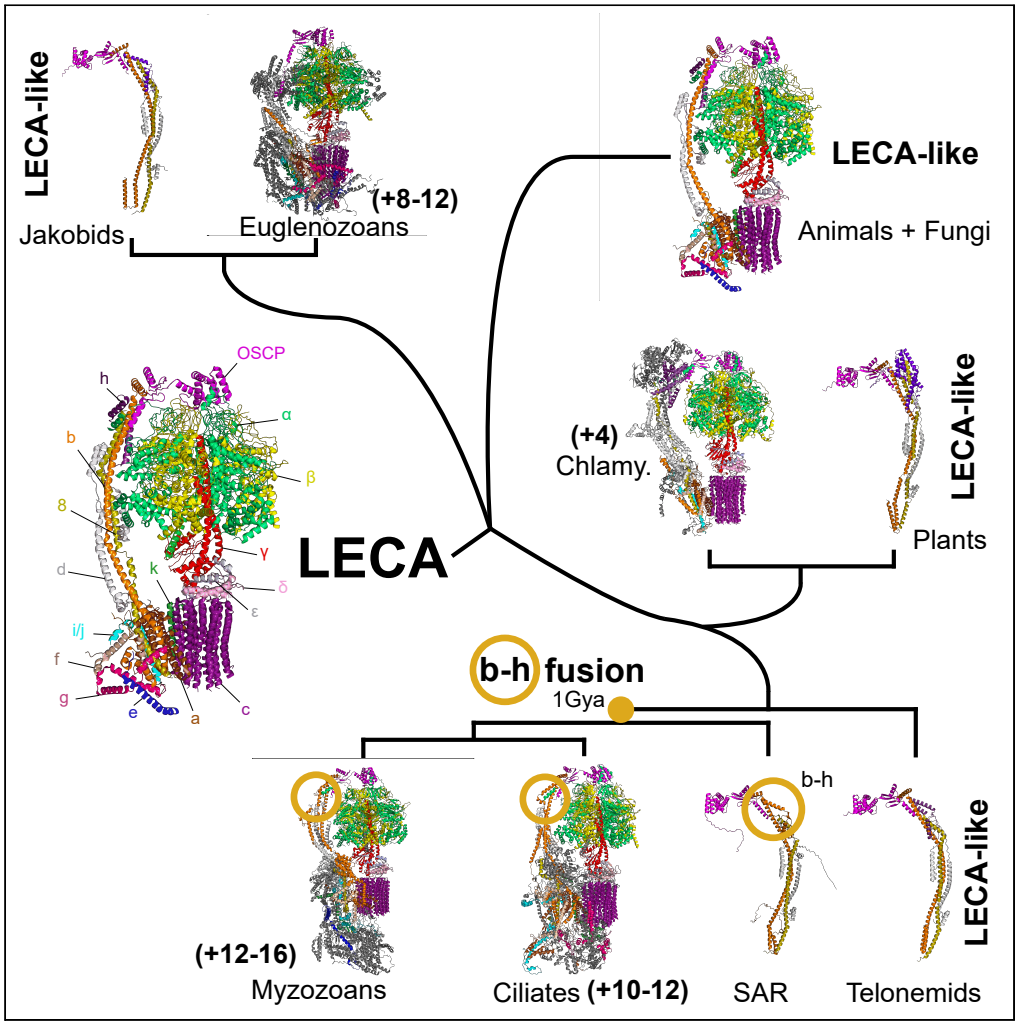
iScience

CellPress
OPEN ACCESS

Article

The persistent homology of mitochondrial ATP synthases



Savar D. Sinha, Jeremy G. Wideman

jeremy.wideman@asu.edu

Highlights

The 17 ATP synthase subunits of animals and fungi are ancestral to eukaryotes

Accreted ATP synthase subunits are usually additions not replacements

ATP synthase subunits b and h fused >1 billion years ago in the SAR lineage

Sinha & Wideman, iScience 26, 106700
May 19, 2023
https://doi.org/10.1016/j.isci.2023.106700

iScience

CellPress
OPEN ACCESS

Article

The persistent homology of mitochondrial ATP synthases

Savar D. Sinha[1] and Jeremy G. Wideman[1,2,*]

SUMMARY

Relatively little is known about ATP synthase structure in protists, and the investigated ones exhibit divergent structures distinct from yeast or animals. To clarify the subunit composition of ATP synthases across all eukaryotic lineages, we used homology detection techniques and molecular modeling tools to identify an ancestral set of 17 ATP synthase subunits. Most eukaryotes possess an ATP synthase comparable to those of animals and fungi, while some have undergone drastic divergence (e.g., ciliates, myzozoans, euglenozoans). Additionally, a ~1 billion-year-old gene fusion between ATP synthase stator subunits was identified as a synapomorphy of the SAR (Stramenopila, Alveolata, Rhizaria) supergroup (stramenopile, alveolate, rhizaria). Our comparative approach highlights the persistence of ancestral subunits even amidst major structural changes. We conclude by urging that more ATP synthase structures (e.g., from jakobids, heteroloboseans, stramenopiles, rhizarians) are needed to provide a complete picture of the evolution of the structural diversity of this ancient and essential complex.

INTRODUCTION

Mitochondria arose from an alphaproteobacterial endosymbiont over 1.5 billion years ago.[1–3] Despite its prokaryotic origins, only ~10-20% of the mitochondrial proteome is directly descended from homologs in alphaproteobacteria.[4] All other mitochondrial proteins were either acquired through horizontal gene transfer (HGT) from other bacteria or originated *de novo* in the eukaryotic nuclear genome during eukaryogenesis.[5] The most famous and perhaps the most important function of mitochondria is their involvement in the formation of a proton gradient via Complexes I-IV of the electron transport chain (ETC) and the ultimate production of ATP via ATP synthase.[6] The eukaryotic versions of the ETC and ATP synthase retain many of the basic features of their prokaryotic counterparts; however, eukaryotes have accreted numerous supernumerary protein subunits.[7,8] It is currently unclear which of these additional components are ancestral to eukaryotes and which are lineage specific.

All studied mitochondrial ATP synthases form dimers, whereas their counterparts in bacteria and chloroplasts do not.[7] It has been proposed that the eukaryotic supernumerary ATP synthase subunits evolved to facilitate their dimerization as well as the formation and maintenance of cristae,[9–13] which are present in every aerobic eukaryote.[14] In fact, variation in the angle between ATP synthase monomers across eukaryotes[14] as well as dimer row formation along crista ridges[9,15–17] correlate well with variations in cristae morphology and ultrastructure, providing further support linking molecular structure to cell biological variation. In 2019, Kühlbrandt categorized mitochondrial ATP synthases into four different types,[7] which correlate with different cristae ultrastructures.[12,15–20] We will reference the types here but urge the field to move away from typological thinking and toward an evolutionary framework. We will instead refer to Type I-like ATP synthases of animals and fungi and plants as the canonical ATP synthase structure. The Type II, both Type III (two distinct structures from related, but distinct lineages), and Type IV ATP synthases, will be referred to collectively as divergent ATP synthases, but individually by the lineages to which they belong (chlorophyceans, ciliates, myzozoans, and euglenozoans, respectively) (Table 1).

Recent investigations of animal and fungal ATP synthases demonstrated that the same set of 17 subunits compose the dimeric complexes in both lineages[27,28] (Figure 1). These ATP synthases comprise two sectors which can be separated biochemically. The soluble F_1 sector is composed of five subunits (α, β, γ, δ, and ε) and is the site of ATP synthesis from ADP and P_i via the rotation of the γ-subunit. The membrane-bound F_o sector comprises the peripheral stalk and proton half-channels. The a-subunit facilitates the transfer of protons down

[1]Center for Mechanisms of Evolution, Biodesign Institute, School of Life Sciences, Arizona State University, Tempe, AZ 85281, USA

[2]Lead contact

*Correspondence: jeremy.wideman@asu.edu

https://doi.org/10.1016/j.isci.2023.106700

Table 1. Nomenclatural cheat sheet: Alternative names for classical ATP synthase subunits in other eukaryotes with solved structures

Canonical Type I Animal/Fungal	Chlorophycean Type II *Polytomella*	Ciliate Type IIIc *Tetrahymena*	Myzozoan Type IIIm *Toxoplasma*	Euglenozoan Type IV *Trypanosoma*
α	α	α	α	α
β	β	β	β	β
γ	γ	γ	γ	γ
δ	δ	δ	δ	δ
ε	ε	ε	ε	ε
OSCP	OSCP	OSCP	OSCP	OSCP
a	a	ymf66	a	a
b	ASA6	b-h	b-h	b
c	c	c	c	c
d	ASA1	d	d	ATBTB2/d
e	–	–	ATPTG15	ATBTB8/ATPEG2
f	ASA9	f	f	ATPTB5/f
g	g[a]	ATPTT6	–	ATPTB9/ATPEG1
h	ASA4	(b-h)	(b-h)	–
i/j	ASA8	i/j	i/j	ATPTB10/i/j
k	k[b]	k	k	ATPTB7/k
8	ASA5	ymf56	8	ATPTB13/8

Due to the significant divergence in sequence similarity across various lineages, several subunits initially labeled as novel subunits in earlier studies[19,20,21,22,23,24] were later found to be ancestral subunits in future analyses[15,16,18,25] including this investigation. In the euglenozoan column, the name before the vertical line represents the kinetoplastid nomenclature while the name after displays the euglenid names for the subunits.

[a]g-Subunit in chlorophyceans is only conserved in select lineages.

[b]The Cryo-EM structure of *Polytomella parva* does not contain a k-subunit[26]; however, a k-like subunit was identified in certain chlorophyceans.

their concentration gradient via the rotation of the c-ring, an oligomer of 10 c-subunits in most mitochondria except animals, which have an 8-member c-ring.[29,30] The remaining subunits (b, d, e, f, g, h, i/j, k, 8, and OSCP) compose the peripheral stalk, which links the F_o to the F_1 sector. In animal and fungal structures, subunits e, g, and k increase dimer stability by forming a hydrophobic complex on both sides of the dimer that induces membrane curvature.[7,31–33] Many of these subunits are present in divergent ATP synthases; however, the ancestry of several subunits in these complexes remains unresolved. Furthermore, due to the short length (<100 aa) of many of these protein subunits, an extensive comparative genomics survey including representatives from across eukaryotic diversity has not been undertaken.

Here, we use diverse bioinformatic tools to identify ATP synthase subunits in a broad diversity of 219 eukaryotic predicted proteomes. We identify several divergent proteins in solved chlorophycean, ciliate, myzozoan, and euglenozoan ATP synthases as divergent homologues of animal and fungal subunits and conclude that the ATP synthase present in the Last Eukaryotic Common Ancestor (LECA) contained the 17 canonical ATP synthase subunits. Furthermore, we show that most ATP synthase subunits are conserved across most eukaryotic lineages and suggest that novel subunits tend to be accreted and do not usually replace ancestral subunits. Finally, we identified a ~1 billion-year-old gene fusion between subunits b and h of the peripheral stalk in the SAR (Stramenopila, Alveolata, Rhizaria) supergroup. Collectively, these data provide a broad picture of the evolution of eukaryotic ATP synthases.

RESULTS

The ancestral eukaryotic ATP synthase contained all 17 subunits shared by animal and fungal complexes

To determine the ancestral subunit complement of eukaryotic ATP synthase, we first collected representative protein sequences from eukaryotic species with solved or previously characterized structures (Data S1).

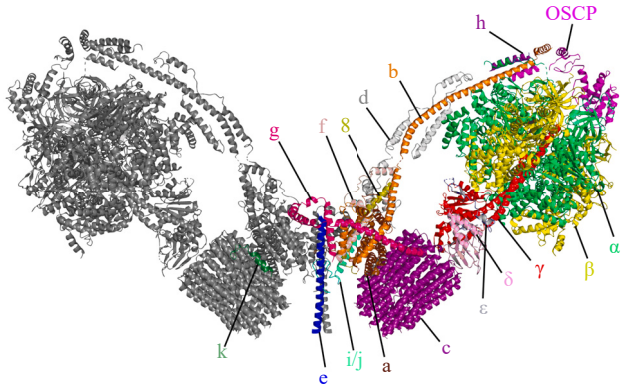


Figure 1. Structure of *Saccharomyces cerevisiae* F_1F_o ATP synthase dimer

The Cryo-EM Type I structural model from yeast could exemplify a typical eukaryotic ATP synthase (PDB: 6B8H).[28] The F_1 sector comprises the catalytic head of the complex including the αβ hexamer and the central rotor γ + δ, and ε. The F_o sector comprises the peripheral stalk (8, b, d, f, h, i/j, OSCP), H^+ channel (a and c), and intermembrane dimerization subunits (e, g, k).

We used these sequences as BLAST[34,35] queries into an expanded version of The Comparative Set (TCS), which is a subset of EukProt v3[36] predicted proteomes selected based on their phylogenetic importance and BUSCO completion (See STAR Methods for more details and Data S1 for a complete list of proteomes used in this study). We added a few extra species not included in the TCS for a total of 219 predicted proteomes. Top hits with E-values below a threshold of 0.01 were used as reciprocal BLAST queries[34,35] into predicted proteomes of species with solved ATP synthase structures. Orthologues were considered validated if *bona fide* ATP synthase subunits were retrieved as top hits of reciprocal searches. Orthologue sets of each ATP synthase subunit were then aligned to build Hidden Markov Models (HMMs) to use in HMMER[37,38] searches to identify more divergent orthologues. In several extreme cases, divergent orthologues could not be validated using reciprocal searching, and structural comparisons using Phyre2,[39] SWISS-MODEL[40] and AlphaFold2[41] were used to infer orthology (See Figures S1-S7). Using a mixture of these approaches, we identified all 17 canonical animal/fungal ATP synthase subunits across diverse lineages spanning the entire tree of eukaryotes (Figure 2). From these data, we conclude that, under nearly any eukaryotic rooting hypothesis,[42–44] the ancestral ATP synthase contained all 17 animal/fungal subunits. Similarly, a complex complement of ATP synthase assembly factors is also retained in most eukaryotic lineages (Figure S8). Certain factors such as subunit s and Atp25 were confirmed to be animal- or fungal-specific, respectively. All other assembly factors are retained in most lineages with notable absences of Atp10 in Chlorophyceans; Atp10, Nca2, TMEM70, and TMEM242 in select alveolate lineages; and TMEM70 in most discicristates. Similar to animals and fungi, lineage-specific assembly factors may have evolved in certain groups such as kinetoplastids,[21] but further experimental investigation in diverse lineages will be required to uncover other lineage-specific factors.

Divergent ATP synthases retain most ancestral protein subunits

While all investigated eukaryotic ATP synthases exhibit dimeric structures,[47] one of the most conspicuous structural differences between lineages is their dimer angles.[9,12,15–18,20] These dimer angle variations influence mitochondrial cristae shape as they correlate with differences in mitochondrial ultrastructure.[9] Alongside divergences in cristae morphology and dimer angles, ancestral ATP synthase subunits have diverged so greatly in certain lineages that several subunits were originally mis-identified as novel lineage-specific accretions or replacements (Table 1, Data S2 and[15,16,18]). Here, we were able to bioinformatically identify several putatively lineage-specific proteins in divergent ATP synthases as orthologues of canonical subunits from animals and fungi (Figures 2 and S9).

Unlike other green algae, chlorophycean ATP synthases like that of *Polytomella parva* have divergent structures.[26] They possess an enlarged peripheral stalk with a smaller dimer angle of 56° (compared to 86° in yeast), resulting in club-shaped cristae[9] and include several F_o subunits widely conserved across Chlorophyceans (ASA1-ASA10), originally thought to be lineage-specific.[26,48] Here, we identified most as orthologues of ancestral subunits. We corroborated previous studies which suggested ASA4 as homologous to subunit h.[25] Through HHpred searches and structural comparisons of helix location in assembled complexes (Figure 3),

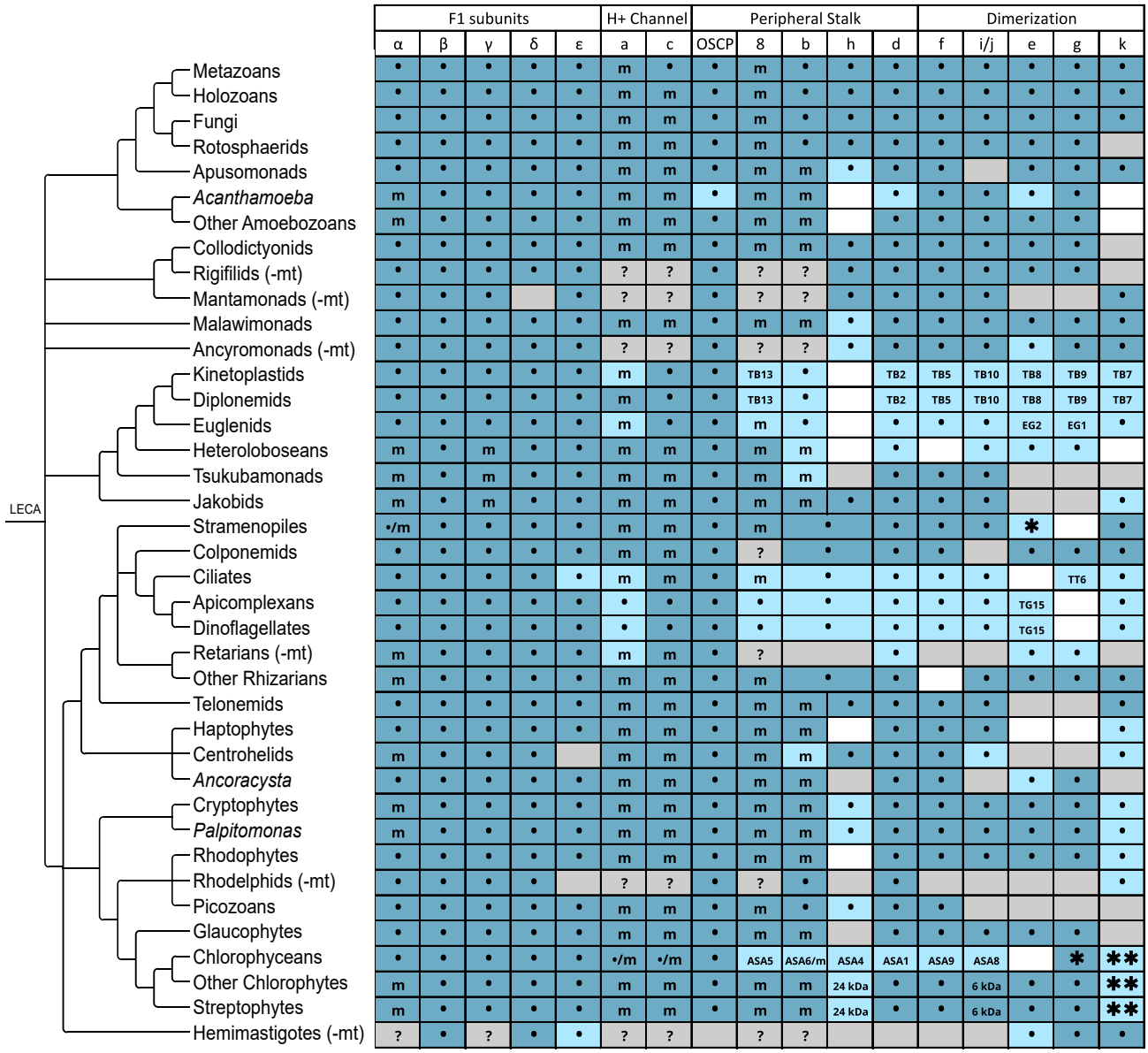


Figure 2. Most eukaryotic ATP synthases retain the core set of 17 ancestral subunits

ATP synthase orthologues were identified in 219 EukProt3 proteomes from major eukaryotic taxa. Subunits were collected via BLAST and HMMER. Dark blue indicates verification via reciprocal BLAST or phmmer[34,35,38] and light blue via HHpred or structural modeling.[39–41,45] White indicates that the subunit is either lost or diverged beyond recognition. Gray stipulates the lack of genomic data. Subunits misidentified as novel are indicated by their species-specific names (e.g. TB5). ·/^{m}Subunits are nuclear/mitochondrially encoded respectively. *Presence is restricted to certain lineages. **Subunit may no longer associate with ATP synthase. $^{(-mt)}$Taxa lack published mitochondrial genomes. $^{?}$Presence in mitochondrial genome is unknown. Retarian mitochondrial genomes are from Macher et al. 2023.[46]

we identified ASA1 as subunit d (Figure S3), ASA5 as subunit 8 (Figure S4), ASA6 as subunit b (Figure S2), ASA8 as subunit i/j (Figure S6), and ASA9 as subunit f (Figure S5). We note that ASA10 matches the location but runs in the opposite direction of an N-terminal helix lost from chlorophycean subunit a, suggesting a partial replacement (Figures 3 and S10). HMMER also identified divergent subunit k orthologues in green plants and algae that appear to be absent from the solved *P. parva* structure (Figure 2).

It should be noted, however, that the composition of chlorophycean ATP synthases is not uniform. In *Tetradesmus obliquus*, we identified an ortholog of subunit g, which appears to have been lost in related

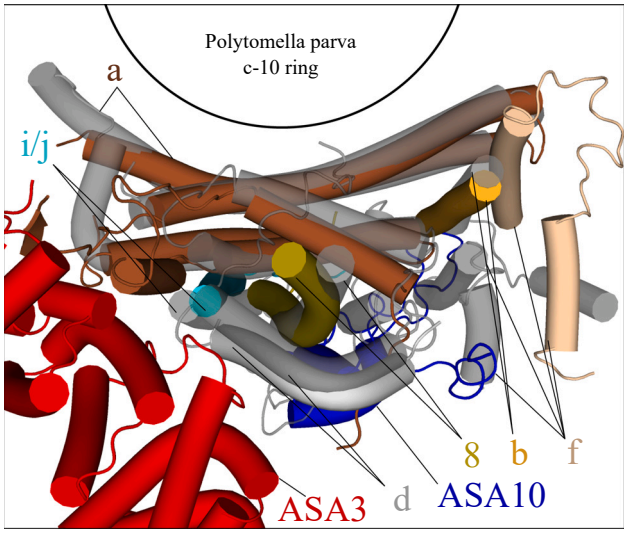


Figure 3. Superpositioning of *Polytomella parva* and yeast ATP synthase F_o components confirms orthology relationships between subunits

Yeast helices are shown in gray while *Polytomella* helices are shown in color. By aligning the yeast (PDB: 6B8H)[28] and *Polytomella* (PDB: 6RD4) through PyMOL2[49–51] obtained from PDB[52] we found that helices from *Polytomella* subunits ASA1, ASA5, ASA6, ASA8, and ASA9 aligned well with yeast subunits d, 8, b, i/j, and f respectively—perhaps with the exception of ASA9, which is tilted away from f due to extra b subunit helices (not shown). While the structure of the ASA10 aligns closely with the H1 helix of bovine subunit a, the helix runs in the opposite direction and no sequence similarity could be detected between the two sections.

algae. Furthermore, while all ATP synthase subunits are nuclear encoded in chlamydomonad algae such as *P. parva* and *Chlamydomonas reinhardtii*, our study revealed the presence of subunits a, c, and ASA6 in the mitochondrial genome of *T. obliquus*, supporting our identification of ASA6 as the orthologue of subunit b (see above), which is mitochondrially encoded across many early diverging lineages of eukaryotes. Since no protein comparable to ASA5/subunit 8 was identified in the nuclear genome of *T. obliquus*, it is possible that subunit 8 is also mitochondrially encoded, though may have diverged beyond recognition. These observations support the gradual evolution of chlorophycean ATP synthases from a more typical canonical ancestor.

Alveolate ATP synthases within the major lineages of myzozoans (which includes apicomplexans, such as *Toxoplasma,* and dinoflagellates), and ciliates (including *Tetrahymena*) are among the most divergent structures investigated. Though ciliates and myzozoans are sister lineages, their ATP synthases are diverged even from one another. Although both alveolate structures are diverged from one another, ATP synthase monomers are both positioned approximately parallel to each other (~0° dimer angles), producing vastly different cristae structures from both one another and other ATP synthases. The tetramers of ciliate ATP synthases produce tubular cristae, while the hexamers of myzozoan ATP synthases form bulbous cristae.[15,16] Both ciliate and myzozoan ATP synthases present a wide range of conserved ancestral subunits; however, none are shared between the two lineages. Ciliate complexes contain 12 lineage-specific subunits universally conserved across ciliates, whereas apicomplexan complexes contain 16 unrelated lineage-specific subunits conserved in all apicomplexans, 12 of which are also conserved in dinoflagellates (Data S1).[15,16,22,23] Through HHpred sequence comparison with alignments of subunits e and g obtained from the 219 EukProt3 genomes and transcriptomes, we identified *Toxoplasma* ATPTG15 as a putative orthologue of subunit e and *Tetrahymena* ATPTT6 as a putative subunit g sequence (Figure 2); however, we were unable to identify a candidate subunit g in apicomplexans or a subunit e in ciliates (see Data S1). The widespread conservation of ATP synthases among ciliates and myzozoans indicates that these structural changes occurred early in the evolution of these lineages and have remained relatively stable for over 500 million years.[53]

Extensive research on *Trypanosoma brucei* and *Euglena gracilis* has revealed the incredibly divergent structure of euglenozoan ATP synthases. Like other divergent complexes, the euglenid and kinetoplastid structures possess smaller dimer angles of 45° and 60° respectively, with novel subunits accumulated

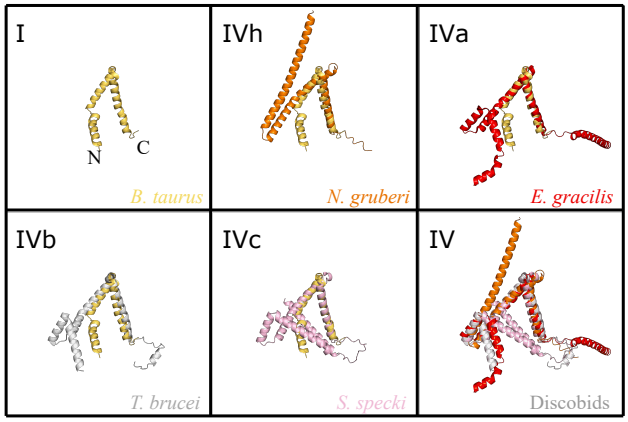


Figure 4. Heterolobosean subunit g confirms the orthology of euglenozoan subunits

[I]Structure of *Bos taurus* (bovine) subunit g (PDB: 6ZPO).[27] [IVh]*Naegleria gruberi* subunit g (plDDT: 82.2). [IVa]*Euglena gracilis* subunit g (PDB: 6TDU).[20] [IVb]*Trypanosoma brucei* subunit g (PDB: 8AP6).[18] [IVc]*Sulcionema specki* subunit g (plDDT: 76.6). [IV]Euglenozoan g-subunits overlayed. Sequences for *Naegleria gruberi* and *Sulcionema specki* were obtained via HMMER and structurally modeled using AlphaFold2.[38,41] Structures were aligned through PyMOL2.[49–51] Bovine subunit is in yellow. N-terminus and C-terminus of bovine structure notated by N and C, respectively.

around the periphery of both the F_1 and F_o sectors.[18–20,54] All three aerobic euglenozoan lineages (kinetoplastids, diplonemids, and euglenids) share 8 lineage-specific ATP synthase subunits, with an additional four present only in euglenids and one specific to kinetoplastids and diplonemids (Data S1). During article preparation, a subset of our bioinformatic predictions was validated by structural and biochemical analyses of the *T. brucei* ATP synthase.[18,19] Using HMMsearch, HHpred, and structural validation, we were able verify the putative identities of ancestral subunits originally classified as novel, including ATPTB2 as subunit d (Figure S3), ATPTB5 as subunit f (Figure S5), ATPTB7 as subunit k (Figure S7), ATPTB8/ATPEG2 as subunit e, ATPTB9/ATPEG1 as subunit g (Figure 4), ATPTB10 as subunit i/j (Figure S6), and ATPTB13 as subunit 8 (Figure S4).[18,19,55] However, even though we were able to identify many subunits, kinetoplastid subunit b could not be found bioinformatically. Two recent reportsidentified two unrelated proteins as candidates for kinetoplastid subunit b.[18,21] Our homology searches revealed that both candidate subunit b proteins are universally conserved in sequenced kinetoplastids and diplonemids; however, they bear little resemblance to each other, their euglenid counterpart, or any other subunit b orthologue (Figure 5). The putative b subunit identified by Dewar et al. (2022) was shown not to associate with the Cryo-EM structure,[18] thus, it was concluded that this subunit serves as an assembly factor given its pivotal role in the assembly of the complex.[21] The protein identified in the cryo-EM structure by Gahura et al. (2022) appears to share homology to the first helix of the yeast subunit b[18] and can be considered the *bona fide* kinetoplastid subunit b. Finally, no orthologue of subunit h was identified in either the *Trypanosoma*[18] or *Euglena*[20] structures, suggesting either rapid divergence or total replacement by subunits ATPTB3 and ATPTB4 which occupy the relative location of subunit h.

Other discobids related to euglenozoans lack the shared characteristics of euglenozoan ATP synthases. For example, jakobids, which contain the most gene-rich and bacteria-like mitochondrial genome[8] in addition to tubulo-vesicular cristae,[14] appear to have more ancestral-like ATP synthase structures and virtually contain a canonical ATP synthase lacking detectable e and g subunits (Figure 2). Whereas tsukubamonads and heteroloboseans, which branch nearest euglenozoans, may possess ATP synthases of intermediate divergence. While subunits a, c, d, i/j, and 8 are readily identifiable in tsukubomonads and heteroloboseans, other components including subunits h and k could not be identified and subunit f could only be found in *Tsukubamonas globosa*. Initially, b subunits could not be identified in either Tsukubamonas nor heteroloboseans, but structural modeling of mitochondria-encoded proteins revealed a subunit b in each lineage (Figure 5). Moreover, a euglenozoan-like g-subunit orthologue which structurally resembles ATPEG1 according to the SWISS-MODEL, was identified in heteroloboseans (Figure 4), and could provide clues to the formation of the discoid cristae shapes conserved in both heteroloboseans and euglenozoans.[19,20,56] Taken together, these data suggest that a gradual divergence of ATP synthase may have occurred in the discicristate lineage. What explains this divergence remains mysterious, but both adaptive

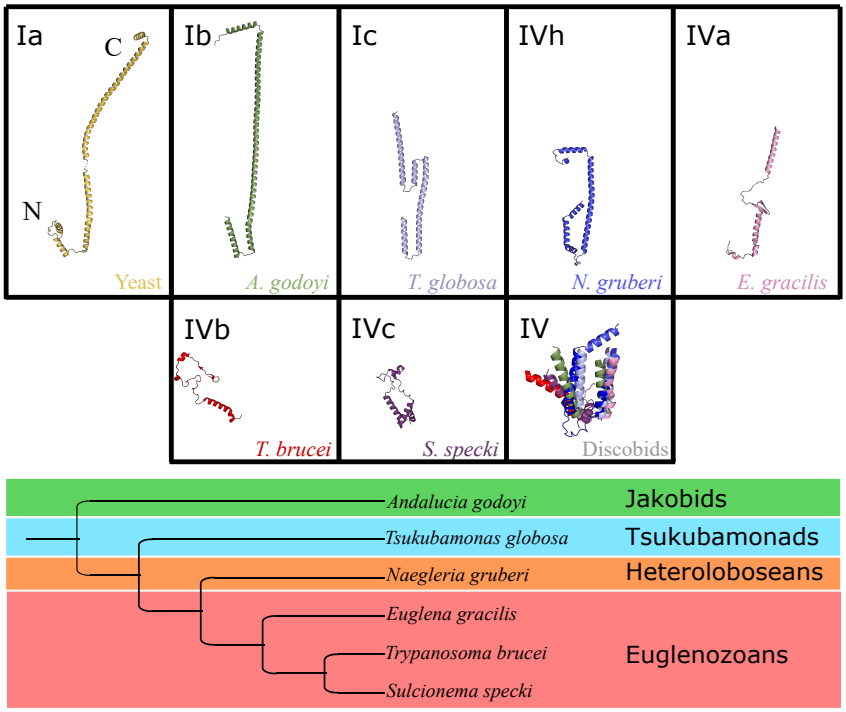


Figure 5. Truncation of subunit b in the transition from canonical to divergent ATP synthase structures in the lineage leading to Euglenozoans

[Ia]*Saccharomyces cerevisiae* (yeast) subunit b (PDB: 6B8H).[28] [Ib]*Andalucia godoyi* subunit b (plDDT: 86.6). [Ic]*Tsukubamonas globosa* subunit b (plDDT: 67.6). [IVh]*Naegleria gruberi* subunit b (plDDT: 68.0). [IVa]*Euglena gracilis* subunit b PDB: 6TDU).[20] [IVb]*Trypanosoma brucei* subunit b (PDB: 8AP6).[18] [IVc]*Sulcionema specki* subunit b (plDDT: 65.6). [IV]Alignment of discobid subunit b structure demonstrates the conservation of N-terminus helices. Tree shown below depicts the evolution of discobids.[36] Sequences for *Andalucia godoyi*, *Tsukubamonas globosa*, *Naegleria gruberi* and *Sulcionema specki* were obtained via HMMER and structurally modeled using AlphaFold2.[38,41] Structures were aligned through PyMOL2.[49–51] N-terminus and C-terminus of yeast structure notated by N and C, respectively.

and constructive neutral evolutionary scenarios could be at play.[57] More intermediate lineages that branch at the base of euglenozoans (e.g., EU17/18[58]) are required to better understand the evolution of these ATP synthases.

SAR lineage synapomorphy revealed in billion-year-old gene fusion between subunits b and h

The SAR supergroup links stramenopiles, alveolates, and rhizarians together into a single clade identified only by multigene phylogenetics.[59] Despite this phylogenetic support, to our knowledge, no definitive organismal or molecular traits that characterize the SAR supergroup have been identified. Here, we present the first synapomorphy of the SAR supergroup—a gene fusion of the ATP synthase b and h subunits (Figure 6).

The ancestral b subunit was encoded in the mitochondrial genome of LECA and is still encoded in the mitochondrial genome of many extant eukaryotes, including the sister group to SAR, the telonemids.[63] All SAR representatives encode subunit b in their nuclear genome (Figure 2). This means that ~1 billion years ago, prior to the diversification of SAR, subunit b was likely transferred from the mitochondrial to the nuclear genome. In our investigations of SAR b subunits, we noticed consistent C-terminal extensions. HHpred searches[45] and structural modeling[41] revealed the C-terminal portion of SAR b subunits to share homology with h-subunits, which otherwise cannot be identified in any SAR species. For example, HHpred searches with *Aureococcus anophagefferens* (stramenopile), *Colponemidia sp. Colp-15* (alveolate), and *Plasmodiophora brassicae* (rhizarian) subunit b targeted against the Pfam database all hit the domain for subunit h—ATP-synt_F6—with low p values of 0.00067, 0.00012, and 0.0021, respectively. Furthermore, hmmsearches using all the h-subunits collected in the study targeted against all stramenopiles, alveolates, and rhizarians

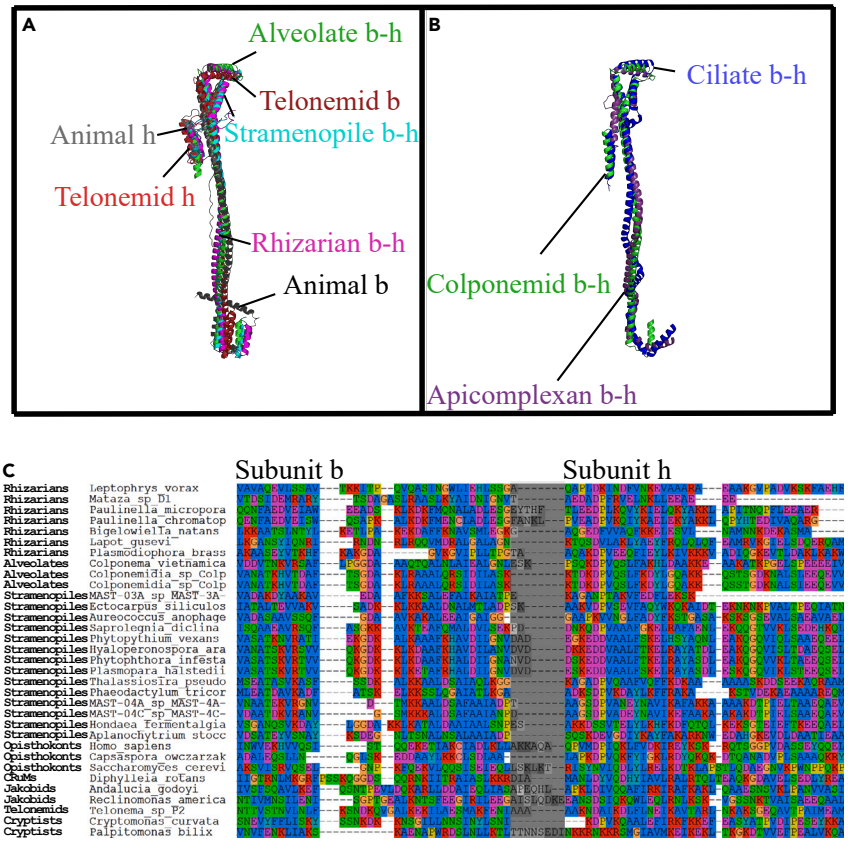


Figure 6. Billion-year-old subunit b-h fusion is a synapomorphy of the eukaryotic clade comprising stramenopiles, alveolates, and rhizarians (SAR)

(A) Structural alignment of predicted stramenopile, alveolate, and rhizarian b-h fusions with b and h subunits from typical structures. Since some SAR ATP synthases are extremely divergent (e.g., ciliates and myzozoans), we chose minimally divergent representatives from each SAR clade for structural modeling using AlphaFold2 and subsequent alignment via PyMOL2.[41,49–51] Animal: *Bos taurus* (gray) obtained from (PDB: 6ZPO).[27,52] All other sequences were predicted from AlphaFold2. Stramenopile: *Ectocarpus siliculosus* (blue) (plDDT: 73.6). Alveolate: *Colponema vietnamica* (green) (plDDT: 73.4). Rhizarian: *Lapot gusevi* (pink) (plDDT: 71.5). Telonemid: *Telonema sp. P-2* (red) (plDDT: 87.6 and 72.9 for subunits b and h respectively).

(B) Alveolate ATP synthases retain highly divergent b-h fusion proteins. *C. vietnamica* b-h subunit was structurally aligned to known structures of b subunits from *Tetrahymena thermophila* (PDB: 6YNY) and *Toxoplasma gondii* (PDB: 6TMK).[15,16,52]

(C) SAR b-h aligned with concatenated b and h subunits from diverse eukaryotes outside the SAR supergroup. N-terminal subunit h mitochondrial targeting sequences were identified using TargetP 2.0[60] and removed. These trimmed h subunits were appended to b subunits from the same species. These artificially fused b and h subunit sequences were then aligned to b-h subunits from stramenopiles, alveolates, and rhizarians using MUSCLE[61] and visualized using Seaview Version 5.[62] The approximate b-h boundary is shown by the gray box.

in the EukProt3 TCS revealed various SAR subunit b C-termini as hits, including those from *Phytophthora infestans* (stramenopile), *Colponema vietnamica* (alveolate), and *P. brassicae* (rhizarian) with e-values of 1.5e-08, 6.9e-09, and 7.2e-05, respectively. Thus, we suggest that upon its transfer to the nucleus, the subunit b gene fused with the subunit h gene, producing the b-h subunit of stramenopiles, alveolates, and rhizarians. These data help explain the h-like structure observed at the C-terminus of *Toxoplasma gondii* subunit b.[16]

Given the high conservation of the b-h subunit, questions remain about the evolutionary benefits of this fusion. Previous studies have suggested that subunit h may be involved the stability of the F_1 sector along with the OSCP subunit and the C-terminus of the b-subunit.[64] We were unable to identify subunit h in several lineages. While hmmsearches corroborated orthology between ASA4 and subunit h in green algae,

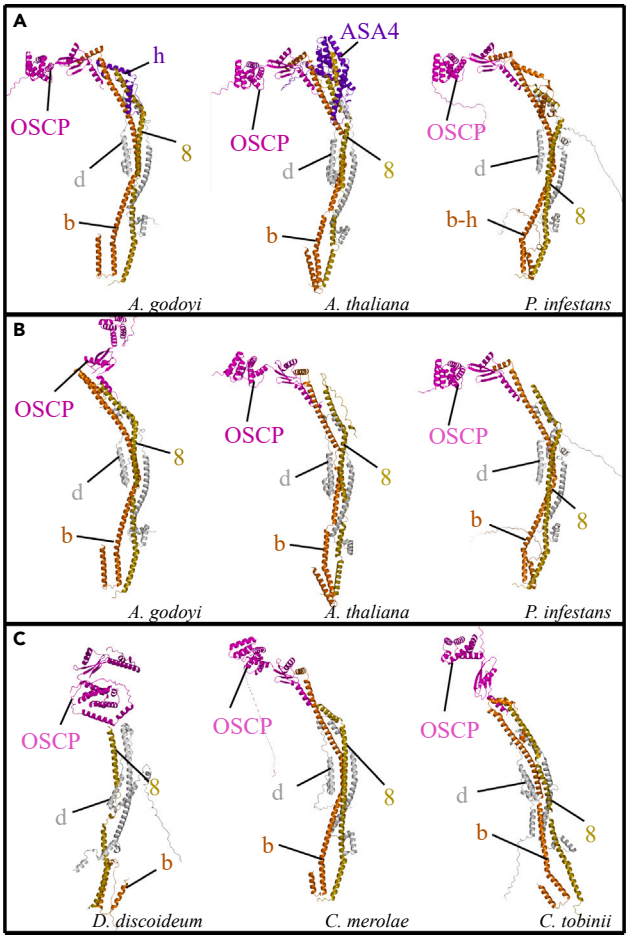


Figure 7. Multi-subunit modeling of ATP synthase peripheral stalk structures provides insight into subunit h function. Peripheral stalk structures were modeled from diverse eukaryotes

(A) *Andalucia godoyi* (plDDT: 72.8), *Arabidopsis thaliana* (plDDT: 76.0), and *Phytophthora infestans* (plDDT: 72.0) subunits interact to form a classical peripheral stalk architecture.

(B) Peripheral stalks from *Andalucia godoyi* (plDDT: 66.5), *Arabidopsis thaliana* (plDDT: 68.7), and *Phytophthora infestans* (plDDT: 68.1) have difficulty folding correctly without subunit h.

(C). Predicted structures from *Dictyostelium discoideum* (plDDT: 51.5), *Cyanidioschyzon merolae* (plDDT: 61.5), and *Chrysochromulina tobinii* (plDDT: 67.8) fold incorrectly, similar to structures lacking subunit h. Sequences from each organism were obtained from BLAST and HMMER searches into the EukProt3 and NCBI databases to obtain nuclear and mitochondrial encoded proteins, respectively.[34–36,38,66] Peripheral stalk structures were assembled from our findings via the "Colab-Fold: AlphaFold2 using MMseqs2" notebook.[41,65] Structures with canonical b, d, h, and 8 subunits folded into structures resembling animal or fungal ATP synthase (e.g., *Andalucia godoyi*, *Arabidopsis thaliana*, and *Phytophthora infestans*). Conversely, any structure lacking the h subunit could not be modeled into classical peripheral stalk structures (e.g., *Dictyostelium discoideum*, *Cyanidioschyzon merolae*, *Chrysochromulina tobinii*).

no subunit h orthologues could be identified in red algae or glaucophytes. Similarly, while all other F_o subunits were identified in the Cryo-EM structures of *T. brucei* and *E. gracilis*, subunit-h could not be identified. We modeled the b, d, h, 8, and OSCP subunits of several eukaryotes. Interestingly, the peripheral stalk proteins from *Andalucia godoyi*, *Arabidopsis thaliana*, or *Phytophthora infestans* were predicted to form realistic structures using Colab-Fold[65] (Figure 7A). However, some structural aberrations in OSCP folding could be seen when the h subunit was omitted from the predictions (Figure 7B). Similarly, in lineages where the h subunit could not be found, the peripheral stalk did not form realistic structures when modeled with Colab-Fold (Figure 7C). Given these structural predictions, subunit h appears to play an important structural role in canonical ATP synthases. Thus, it is possible that taxa in which we were unable to find an orthologue of subunit h may possess extremely divergent h-subunits or novel accretions that perform a similar function.

Absence of evidence is not evidence of absence: Small subunits e, f, g, h, and k are likely present but not detected in most predicted proteomes

The recent euglenozoan structural data and ambiguous orthology of subunits e and g in alveolates indicate the requirement of structural data to identify core ATP synthase subunits in some lineages. While our investigation has uncovered the broad conservation of all ATP synthase subunits in several major eukaryotic lineages, even the most sensitive homology detection tools cannot unambiguously identify orthologues of small divergent subunits (e.g., subunits e, f, g, h, k). Furthermore, since the default parameters of many gene prediction programs eliminate proteins shorter than 100 aa, short ATP synthase subunits are often left unpredicted. Thus, many of these smaller subunits could only be detected in the transcriptomes or raw reads in the SRA database (See Data S2 cells in orange). Since HMM searches could not be conducted on these databases due to computational limitations, some more divergent small subunits may remain unidentified.

In animals and fungi, both subunits e and g are necessary for dimer stability.[67] However, we were unable to identify both subunits in numerous lineages. In stramenopiles and haptophytes, we identified putative e subunits among a select few taxa but could not identify dimerization subunits in most species in these lineages. Furthermore, no probable subunit g was found in myzozoans and no subunit e was identified in ciliates. Neither subunit e nor g could be identified in jakobids, and in tsukubamonads, we were unable to uncover any orthologues of subunit e (Figure 2). Thus, only structural investigations will uncover the dimerization mechanisms of these lineages.

Subunit f is a larger subunit present among almost all taxa investigated, with notable absences in heterolobosean and rhizarians. Given the presence of divergent subunit f sequences in euglenozoans, it is likely that subunit f is similarly divergent in heterolobosean. The absence of subunit f in all rhizarians may also indicate minor alterations to the general structure of rhizarian ATP synthase (Figure 2).

DISCUSSION

While the amino acid composition of ATP synthase subunits varies across eukaryotes, the conservation of core fungal and animal subunits implies that the ancestral structure of the LECA ATP synthase resembled the canonical complex present in animals and fungi. This would appear true for nearly any proposed rooting hypothesis.[36,68–74] The ancestral-like qualities of *A. godoyi* ATP synthase and the apparent lack of dimerization subunits e and g in jakobids (Figure 2) is consistent with a Jakobid root,[68] implying a subsequent addition of these subunits after the divergence of jakobids from all other eukaryotes. However, very few jakobids have associated sequence data and many more jakobids must be sequenced to strengthen this suggestion. The core subunit composition of ATP synthase is largely conserved in divergent ATP synthases as we and others identified subunits originally thought to be novel and lineage specific as divergent orthologues of ancestral subunits (Figure 2, Datas S1, S2, S3, and S4, and[15,16,18–20,25,75]). This indicates that the accretion of novel subunits is more prevalent than the replacement of ancestral subunits. In fact, very few replacements were detected, with the only notable instances including subunits e and g, which are completely replaced in some chlorophycean structures and subunit h, which is possibly replaced in euglenozoan structures. A wide range of taxa that lack identifiable ancestral subunits are prime targets for future structural studies. These include amoebozoans (lack subunits h and k), jakobids/tsukubamonads (lack subunits e and g), heterolobosean (lack subunits f, h, i/j, and k), stramenopiles (lack subunits e and g), retarians (lack subunits b-h, f, 8, i/j, and k), other rhizarians (lack subunit f), haptophytes (lack subunits e, g, and h), and rhodophytes (lack subunit h).

The widespread conservation of subunits implies that virtually all uniquely mitochondrial features of ATP synthase (e.g., dimerization) evolved prior to the divergence of eukaryotes. Previous work showed that the Mitochondrial Contact Site and Cristae Organizing System (MICOS) originated from homologs in bacteria,[76] and that mitochondrial cristae are likely homologous to alphaprotebacterial intracytoplasmic membranes (ICMs).[77] While cristae formation requires ATP synthase dimerization,[10,78] it is unclear how ATP synthase might impact alphaproteobacterial ICM formation.[79] Closer investigation of ATP synthases in diverse alphaproteobacteria will shed light on the early evolution of mitochondrial cristae.

Limits of the study

As more complete genomes and transcriptomes become available for important eukaryotic lineages, a better picture of ATP synthase evolution will become clear. In addition, as more structures are solved,

more ancestral subunits lost in animals and fungi could be discovered. Since we are only working with known subunits from solved structures, we cannot predict novel subunits in uninvestigated lineages. Similarly, as mentioned above, the absence of evidence is not evidence of absence. There are limits to homology detection and some subunits may be present but not detected in some lineages. Finally, our work was done during the infancy of structural modeling and as more structures and better modeling software becomes available, better structural predictions can be made.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Database selection and homology searching
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.106700.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation, DBI-2119963, BII: Mechanisms of Cellular Evolution.

AUTHOR CONTRIBUTIONS

Savar D. Sinha: Data curation, visualization, investigation, and writing. Jeremy G. Wideman: Conceptualization, supervision, and writing.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 30, 2022
Revised: January 24, 2023
Accepted: April 14, 2023
Published: April 19, 2023

REFERENCES

1. Fan, L., Wu, D., Goremykin, V., Xiao, J., Xu, Y., Garg, S., Zhang, C., Martin, W.F., and Zhu, R. (2020). Phylogenetic analyses with systematic taxon sampling show that mitochondria branch within Alphaproteobacteria. Nat. Ecol. Evol. *4*, 1213–1219. https://doi.org/10.1038/s41559-020-1239-x.

2. Muñoz-Gómez, S.A., Hess, S., Burger, G., Lang, B.F., Susko, E., Slamovits, C.H., and Roger, A.J. (2019). An updated phylogeny of the Alphaproteobacteria reveals that the parasitic Rickettsiales and Holosporales have independent origins. Elife *8*, e42535. https://doi.org/10.7554/eLife.42535.

3. Muñoz-Gómez, S.A., Susko, E., Williamson, K., Eme, L., Slamovits, C.H., Moreira, D., López-García, P., and Roger, A.J. (2022). Site-and-branch-heterogeneous analyses of an expanded dataset favour mitochondria as sister to known Alphaproteobacteria. Nat. Ecol. Evol. *6*, 253–262. https://doi.org/10.1038/s41559-021-01638-2.

4. Gabaldón, T., and Huynen, M.A. (2007). From endosymbiont to host-controlled organelle: the hijacking of mitochondrial protein synthesis and metabolism. PLoS Comput. Biol. *3*, e219. https://doi.org/10.1371/journal.pcbi.0030219.

5. Ku, C., Nelson-Sathi, S., Roettger, M., Garg, S., Hazkani-Covo, E., and Martin, W.F. (2015). Endosymbiotic gene transfer from prokaryotic pangenomes: inherited chimerism in eukaryotes. Proc. Natl. Acad. Sci. USA *112*, 10139–10146. https://doi.org/10.1073/pnas.1421385112.

6. Walker, J.E. (2018). Chapter 13 Structure, Mechanism and Regulation of ATP Synthases (The Royal Society of Chemistry), pp. 338–373. https://doi.org/10.1039/9781788010405-00338.

7. Kühlbrandt, W. (2019). Structure and mechanisms of F-type ATP synthases. Annu. Rev. Biochem. *88*, 515–549. https://doi.org/10.1146/annurev-biochem-013118-110903.

8. Gray, M.W., Burger, G., Derelle, R., Klimeš, V., Leger, M.M., Sarrasin, M., Vlček, Č., Roger, A.J., Eliáš, M., and Lang, B.F. (2020). The draft nuclear genome sequence and predicted mitochondrial proteome of Andalucia godoyi, a protist with the most gene-rich and bacteria-like mitochondrial genome. BMC

Biol. *18*, 22. https://doi.org/10.1186/s12915-020-0741-6.

9. Blum, T.B., Hahn, A., Meier, T., Davies, K.M., and Kühlbrandt, W. (2019). Dimers of mitochondrial ATP synthase induce membrane curvature and self-assemble into rows. Proc. Natl. Acad. Sci. USA *116*, 4250–4255. https://doi.org/10.1073/pnas.1816556116.

10. Paumard, P., Vaillier, J., Coulary, B., Schaeffer, J., Soubannier, V., Mueller, D.M., Brèthes, D., di Rago, J.-P., and Velours, J. (2002). The ATP synthase is involved in generating mitochondrial cristae morphology. EMBO J. *21*, 221–230. https://doi.org/10.1093/emboj/21.3.221.

11. Arnold, I., Pfeiffer, K., Neupert, W., Stuart, R.A., and Schägger, H. (1998). Yeast mitochondrial F1F0-ATP synthase exists as a dimer: identification of three dimer-specific subunits. EMBO J. *17*, 7170–7178. https://doi.org/10.1093/emboj/17.24.7170.

12. Davies, K.M., Strauss, M., Daum, B., Kief, J.H., Osiewacz, H.D., Rycovska, A., Zickermann, V., and Kühlbrandt, W. (2011). Macromolecular organization of ATP synthase and complex I in whole mitochondria. Proc. Natl. Acad. Sci. USA *108*, 14121–14126. https://doi.org/10.1073/pnas.1103621108.

13. Strauss, M., Hofhaus, G., Schröder, R.R., and Kühlbrandt, W. (2008). Dimer ribbons of ATP synthase shape the inner mitochondrial membrane. EMBO J. *27*, 1154–1160. https://doi.org/10.1038/emboj.2008.35.

14. Pánek, T., Eliáš, M., Vancová, M., Lukeš, J., and Hashimi, H. (2020). Returning to the fold for lessons in mitochondrial crista diversity and evolution. Curr. Biol. *30*, R575–R588. https://doi.org/10.1016/j.cub.2020.02.053.

15. Flygaard, R.K., Mühleip, A., Tobiasson, V., and Amunts, A. (2020). Type III ATP synthase is a symmetry-deviated dimer that induces membrane curvature through tetramerization. Nat. Commun. *11*, 5342. https://doi.org/10.1038/s41467-020-18993-6.

16. Mühleip, A., Kock Flygaard, R., Ovciarikova, J., Lacombe, A., Fernandes, P., Sheiner, L., and Amunts, A. (2021). ATP synthase hexamer assemblies shape cristae of Toxoplasma mitochondria. Nat. Commun. *12*, 120. https://doi.org/10.1038/s41467-020-20381-z.

17. Davies, K.M., Anselmi, C., Wittig, I., Faraldo-Gómez, J.D., and Kühlbrandt, W. (2012). Structure of the yeast F1Fo-ATP synthase dimer and its role in shaping the mitochondrial cristae. Proc. Natl. Acad. Sci. USA *109*, 13602–13607. https://doi.org/10.1073/pnas.1204593109.

18. Gahura, O., Mühleip, A., Hierro-Yap, C., Panicucci, B., Jain, M., Hollaus, D., Slapničková, M., Zíková, A., and Amunts, A. (2022). An ancestral interaction module promotes oligomerization in divergent mitochondrial ATP synthases. Nat. Commun. *13*, 5989. https://doi.org/10.1038/s41467-022-33588-z.

19. Gahura, O., Hierro-Yap, C., and Zíková, A. (2021). Redesigned and reversed: architectural and functional oddities of the trypanosomal ATP synthase. Parasitology *148*, 1151–1160. https://doi.org/10.1017/S0031182021000202.

20. Mühleip, A., McComas, S.E., and Amunts, A. (2019). Structure of a mitochondrial ATP synthase with bound native cardiolipin. Elife *8*, e51179. https://doi.org/10.7554/eLife.51179.

21. Dewar, C.E., Oeljeklaus, S., Wenger, C., Warscheid, B., and Schneider, A. (2022). Characterization of a highly diverged mitochondrial ATP synthase Fo subunit in Trypanosoma brucei. J. Biol. Chem. *298*, 101829. https://doi.org/10.1016/j.jbc.2022.101829.

22. Balabaskaran Nina, P., Dudkina, N. v, Kane, L.A., van Eyk, J.E., Boekema, E.J., Mather, M.W., and Vaidya, A.B. (2010). Highly divergent mitochondrial ATP synthase complexes in Tetrahymena thermophila. PLoS Biol. *8*, e1000418. https://doi.org/10.1371/journal.pbio.1000418.

23. Salunke, R., Mourier, T., Banerjee, M., Pain, A., and Shanmugam, D. (2018). Highly diverged novel subunit composition of apicomplexan F-type ATP synthase identified from Toxoplasma gondii. PLoS Biol. *16*, e2006128. https://doi.org/10.1371/journal.pbio.2006128.

24. Mühleip, A.W., Dewar, C.E., Schnaufer, A., Kühlbrandt, W., and Davies, K.M. (2017). In situ structure of trypanosomal ATP synthase dimer reveals a unique arrangement of catalytic subunits. Proc. Natl. Acad. Sci. USA *114*, 992–997. https://doi.org/10.1073/pnas.1612386114.

25. Röhricht, H., Schwartzmann, J., and Meyer, E.H. (2021). Complexome profiling reveals novel insights into the composition and assembly of the mitochondrial ATP synthase of Arabidopsis thaliana. Biochim. Biophys. Acta. Bioenerg. *1862*, 148425. https://doi.org/10.1016/j.bbabio.2021.148425.

26. Murphy, B.J., Klusch, N., Langer, J., Mills, D.J., Yildiz, Ö., and Kühlbrandt, W. (2019). Rotary substates of mitochondrial ATP synthase reveal the basis of flexible F1-Fo coupling. Science *364*, eaaw9128. https://doi.org/10.1126/science.aaw9128.

27. Spikes, T.E., Montgomery, M.G., and Walker, J.E. (2020). Structure of the dimeric ATP synthase from bovine mitochondria. Proc. Natl. Acad. Sci. USA *117*, 23519–23526. https://doi.org/10.1073/pnas.2013998117.

28. Guo, H., Bueler, S.A., and Rubinstein, J.L. (2017). Atomic model for the dimeric FO region of mitochondrial ATP synthase. Science *358*, 936–940. https://doi.org/10.1126/science.aao4815.

29. Song, J., Pfanner, N., and Becker, T. (2018). Assembling the mitochondrial ATP synthase. Proc. Natl. Acad. Sci. USA *115*, 2850–2852.

30. Watt, I.N., Montgomery, M.G., Runswick, M.J., Leslie, A.G.W., and Walker, J.E. (2010). Bioenergetic cost of making an adenosine triphosphate molecule in animal mitochondria. Proc. Natl. Acad. Sci. USA *107*, 16823–16827. https://doi.org/10.1073/pnas.1011099107.

31. Baker, L.A., Watt, I.N., Runswick, M.J., Walker, J.E., and Rubinstein, J.L. (2012). Arrangement of subunits in intact mammalian mitochondrial ATP synthase determined by cryo-EM. Proc. Natl. Acad. Sci. USA *109*, 11675–11680. https://doi.org/10.1073/pnas.1204935109.

32. Jiko, C., Davies, K.M., Shinzawa-Itoh, K., Tani, K., Maeda, S., Mills, D.J., Tsukihara, T., Fujiyoshi, Y., Kühlbrandt, W., and Gerle, C. (2015). Bovine F1Fo ATP synthase monomers bend the lipid bilayer in 2D membrane crystals. Elife *4*, e06119. https://doi.org/10.7554/eLife.06119.

33. Hahn, A., Parey, K., Bublitz, M., Mills, D.J., Zickermann, V., Vonck, J., Kühlbrandt, W., and Meier, T. (2016). Structure of a complete ATP synthase dimer reveals the molecular basis of inner mitochondrial membrane morphology. Mol. Cell *63*, 445–456. https://doi.org/10.1016/j.molcel.2016.05.037.

34. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

35. Madden, T. (2003). The BLAST Sequence Analysis Tool (National Center for Biotechnology Information (US)).

36. Richter, D.J., Berney, C., Strassert, J.F.H., Poh, Y.-P., Herman, E.K., Muñoz-Gómez, S.A., Wideman, J.G., Burki, F., and de Vargas, C. (2022). EukProt: a database of genome-scale predicted proteins across the diversity of eukaryotes. Peer Community J. *2*, e56. https://doi.org/10.24072/pcjournal.173.

37. Finn, R.D., Clements, J., and Eddy, S.R. (2011). HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. *39*, W29–W37. https://doi.org/10.1093/nar/gkr367.

38. Eddy, S.R. (1998). Profile hidden Markov models. Bioinformatics *14*, 755–763. https://doi.org/10.1093/bioinformatics/14.9.755.

39. Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modelling, prediction and analysis. Nat. Protoc. *10*, 845–858. https://doi.org/10.1038/nprot.2015.053.

40. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., et al. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. *46*, W296–W303. https://doi.org/10.1093/nar/gky427.

41. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold.

Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

42. Rochette, N.C., Brochier-Armanet, C., and Gouy, M. (2014). Phylogenomic test of the hypotheses for the evolutionary origin of eukaryotes. Mol. Biol. Evol. *31*, 832–845. https://doi.org/10.1093/molbev/mst272.

43. Cerón-Romero, M.A., Fonseca, M.M., de Oliveira Martins, L., Posada, D., and Katz, L.A. (2021). Phylogenomic analyses of 2,786 genes in 158 lineages support a root of the eukaryotic tree of life between opisthokonts (animals, fungi and their microbial relatives) and all other lineages. Preprint at bioRxiv. https://doi.org/10.1101/2021.02.26.433005.

44. He, D., Fiz-Palacios, O., Fu, C.-J., Fehling, J., Tsai, C.-C., and Baldauf, S.L. (2014). An alternative root for the eukaryote tree of life. Curr. Biol. *24*, 465–470. https://doi.org/10.1016/j.cub.2014.01.036.

45. Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F., Söding, J., Lupas, A.N., and Alva, V. (2018). A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. J. Mol. Biol. *430*, 2237–2243. https://doi.org/10.1016/j.jmb.2017.12.007.

46. Macher, J.-N., Coots, N.L., Poh, Y.-P., Girard, E.B., Langerak, A., Muñoz-Gómez, S.A., Sinha, S.D., Jirsová, D., Vos, R., Wissels, R., et al. (2023). Single-cell genomics reveals the divergent mitochondrial genomes of retaria (foraminifera and radiolaria). mBio *0*, e0030223. https://doi.org/10.1128/mbio.00302-23.

47. Dudkina, N. v, Heinemeyer, J., Keegstra, W., Boekema, E.J., and Braun, H.-P. (2005). Structure of dimeric ATP synthase from mitochondria: an angular association of monomers induces the strong curvature of the inner membrane. FEBS Lett. *579*, 5769–5772. https://doi.org/10.1016/j.febslet.2005.09.065.

48. Vázquez-Acevedo, M., Cardol, P., Cano-Estrada, A., Lapaille, M., Remacle, C., and González-Halphen, D. (2006). The mitochondrial ATP synthase of chlorophycean algae contains eight subunits of unknown origin involved in the formation of an atypical stator-stalk and in the dimerization of the complex. J. Bioenerg. Biomembr. *38*, 271–282. https://doi.org/10.1007/s10863-006-9046-x.

49. Schrödinger, L.L.C. (2015). The PyMOL molecular graphics System. Version∼1.8.

50. Schrödinger, L.L.C. (2015). The JyMOL molecular graphics development component. Version∼1.8.

51. Schrödinger, L.L.C. (2015). The AxPyMOL molecular graphics plugin for microsoft PowerPoint. Version∼1.8.

52. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The protein data bank. Nucleic Acids Res. *28*, 235–242. https://doi.org/10.1093/nar/28.1.235.

53. Dorrell, R.G., Butterfield, E.R., Nisbet, R.E.R., and Howe, C.J. (2013). Evolution: unveiling early alveolates. Curr. Biol. *23*, R1093–R1096. https://doi.org/10.1016/j.cub.2013.10.055.

54. Montgomery, M.G., Gahura, O., Leslie, A.G.W., Zíková, A., and Walker, J.E. (2018). ATP synthase from Trypanosoma brucei has an elaborated canonical F1-domain and conventional catalytic sites. Proc. Natl. Acad. Sci. USA *115*, 2102–2107. https://doi.org/10.1073/pnas.1720940115.

55. Šubrtová, K., Panicucci, B., and Zíková, A. (2015). ATPaseTb2, a unique membrane-bound FoF1-ATPase component, is essential in bloodstream and dyskinetoplastic trypanosomes. PLoS Pathog. *11*, e1004660. https://doi.org/10.1371/journal.ppat.1004660.

56. Dyková, I., Kyselová, I., Pecková, H., Oborník, M., and Lukeš, J. (2001). Identity of Naegleria strains isolated from organs of freshwater fishes. Dis. Aquat. Org. *46*, 115–121.

57. Muñoz-Gómez, S.A., Bilolikar, G., Wideman, J.G., and Geiler-Samerotte, K. (2021). Constructive neutral evolution 20 Years later. J. Mol. Evol. *89*, 172–182. https://doi.org/10.1007/s00239-021-09996-y.

58. Záhonová, K., Lax, G., Sinha, S.D., Leonard, G., Richards, T.A., Lukeš, J., and Wideman, J.G. (2021). Single-cell genomics unveils a canonical origin of the diverse mitochondrial genomes of euglenozoans. BMC Biol. *19*, 103. https://doi.org/10.1186/s12915-021-01035-y.

59. Burki, F., Shalchian-Tabrizi, K., Minge, M., Skjæveland, Å., Nikolaev, S.I., Jakobsen, K.S., and Pawlowski, J. (2007). Phylogenomics reshuffles the eukaryotic supergroups. PLoS One *2*, e790. https://doi.org/10.1371/journal.pone.0000790.

60. 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.

61. Edgar, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinf. *5*, 113. https://doi.org/10.1186/1471-2105-5-113.

62. Gouy, M., Tannier, E., Comte, N., and Parsons, D.P. (2021). Seaview version 5: a multiplatform software for multiple sequence alignment, molecular phylogenetic analyses, and tree reconciliation. In Multiple Sequence Alignment: Methods and Protocols, K. Katoh, ed. (Springer US), pp. 241–260. https://doi.org/10.1007/978-1-0716-1036-7_15.

63. Strassert, J.F.H., Jamy, M., Mylnikov, A.P., Tikhonenkov, D. v, and Burki, F. (2019). New phylogenomic analysis of the enigmatic phylum telonemia further resolves the eukaryote tree of life. Mol. Biol. Evol. *36*, 757–765. https://doi.org/10.1093/molbev/msz012.

64. Lee, H.G. (2015). Mammalian Atp Synthase: Novel Insights Iinto the Roles of its Supernumerary Subunits. Doctoral dissertation (The Johns Hopkins University).

65. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. Nat. Methods *19*, 679–682. https://doi.org/10.1038/s41592-022-01488-1.

66. Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., and Madden, T.L. (2008). NCBI BLAST: a better web interface. Nucleic Acids Res. *36*, W5–W9. https://doi.org/10.1093/nar/gkn201.

67. He, J., Ford, H.C., Carroll, J., Douglas, C., Gonzales, E., Ding, S., Fearnley, I.M., and Walker, J.E. (2018). Assembly of the membrane domain of ATP synthase in human mitochondria. Proc. Natl. Acad. Sci. USA *115*, 2988–2993.

68. Kannan, S., Rogozin, I.B., and Koonin, E.V. (2014). MitoCOGs: clusters of orthologous genes from mitochondria and implications for the evolution of eukaryotes. BMC Evol. Biol. *14*, 237. https://doi.org/10.1186/s12862-014-0237-5.

69. He, D., Fiz-Palacios, O., Fu, C.-J., Fehling, J., Tsai, C.-C., and Baldauf, S.L. (2014). An alternative root for the eukaryote tree of life. Curr. Biol. *24*, 465–470. https://doi.org/10.1016/j.cub.2014.01.036.

70. Tria, F.D.K., Brueckner, J., Skejo, J., Xavier, J.C., Kapust, N., Knopp, M., Wimmer, J.L.E., Nagies, F.S.P., Zimorski, V., Gould, S.B., et al. (2021). Gene duplications trace mitochondria to the onset of eukaryote complexity. Genome Biol. Evol. *13*, evab055. https://doi.org/10.1093/gbe/evab055.

71. Roger, A.J., and Simpson, A.G.B. (2009). Evolution: revisiting the root of the eukaryote tree. Curr. Biol. *19*, R165–R167. https://doi.org/10.1016/j.cub.2008.12.032.

72. Strassert, J.F.H., Irisarri, I., Williams, T.A., and Burki, F. (2021). A molecular timescale for eukaryote evolution with implications for the origin of red algal-derived plastids. Nat. Commun. *12*, 1879. https://doi.org/10.1038/s41467-021-22044-z.

73. Rogozin, I.B., Basu, M.K., Csürös, M., and Koonin, E.V. (2009). Analysis of rare genomic changes does not support the unikont–bikont phylogeny and suggests cyanobacterial symbiosis as the point of primary radiation of eukaryotes. Genome Biol. Evol. *1*, 99–113. https://doi.org/10.1093/gbe/evp011.

74. Katz, L.A., Grant, J.R., Parfrey, L.W., and Burleigh, J.G. (2012). Turning the crown upside down: gene tree parsimony roots the eukaryotic tree of life. Syst. Biol. *61*, 653–660. https://doi.org/10.1093/sysbio/sys026.

75. Rudy, C.L., Ondřej, G., Brian, P., Alena, Z., and Hassan, H. (2021). Mitochondrial contact site and cristae organization System and F1FO-ATP synthase crosstalk is a fundamental property of mitochondrial cristae. mSphere *6*, e003277-21. https://doi.org/10.1128/mSphere.00327-21.

76. Roger, A.J., Muñoz-Gómez, S.A., and Kamikawa, R. (2017). The origin and diversification of mitochondria. Curr. Biol. *27*, R1177–R1192. https://doi.org/10.1016/j.cub.2017.09.015.

77. Muñoz-Gómez, S.A., Wideman, J.G., Roger, A.J., and Slamovits, C.H. (2017). The origin of mitochondrial cristae from alphaproteobacteria. Mol. Biol. Evol. *34*, 943–956. https://doi.org/10.1093/molbev/msw298.

78. Arselin, G., Vaillier, J., Salin, B., Schaeffer, J., Giraud, M.-F., Dautant, A., Brèthes, D., and Velours, J. (2004). The Modulation in Subunits e and g Amounts of Yeast ATP Synthase Modifies Mitochondrial Cristae Morphology. J. Biol. Chem. *279*, 40392–40399. https://doi.org/10.1074/jbc.M404316200.

79. Anand, R., Reichert, A.S., and Kondadi, A.K. (2021). Emerging roles of the MICOS complex in cristae dynamics and biogenesis. Biology *10*, 600. https://doi.org/10.3390/biology10070600.

80. Menardo, F., Loiseau, C., Brites, D., Coscolla, M., Gygli, S.M., Rutaihwa, L.K., Trauner, A., Beisel, C., Borrell, S., and Gagneux, S. (2018). Treemmer: a tool to reduce large phylogenetic datasets with minimal loss of diversity. BMC Bioinf. *19*, 164. https://doi.org/10.1186/s12859-018-2164-8.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Data S1	This study.	N/A
Data S2	This study.	N/A
Data S3	This study.	(https://doi.org/10.6084/m9.figshare.21936459)
Data S4	This study.	(https://doi.org/10.6084/m9.figshare.21056146)
Software and algorithms		
BLAST	Altschul et al.; Madden T.; Johnson et al.[34,35,66]	https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs&DOC_TYPE=Download
MUSCLE	Edgar R.C.[61]	https://drive5.com/muscle/downloads_v3.htm
hmmsearch v3.3.2	Finn et al. and Eddy S.R.[37,38]	http://hmmer.org/
HHpred v3.3.0	Zimmermann et al.[45]	https://toolkit.tuebingen.mpg.de/tools/hhpred
Phyre2	Kelley et al.[39]	http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index
SWISS-MODEL	Waterhouse et al.[40]	https://swissmodel.expasy.org/
AlphaFold2	Jumper et al.[41]	https://github.com/deepmind/alphafold
ColabFold	Mirdita et al.[65]	https://github.com/sokrypton/ColabFold
PyMOL v2.5	Schrödinger LLC[49–51]	https://pymol.org/2/
Other		
EukProt3 v3	Richter et al.[36]	https://doi.org/10.6084/m9.figshare.12417881.v3
EukProt3 TCS BLAST server	Richter et al.[36]	https://evocellbio.com/SAGdb/TCS/

RESOURCE AVAILABILITY

Lead contact

Further information and requests should be directed to and will be fulfilled by the lead contact, Jeremy Wideman (jeremy.wideman@asu.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- All data are available as supplementary files or has been deposited at FigShare and is publicly available as of the date of publication. DOIs are listed in the key resources table.
- No original code was generated in this study.
- Any additional information required to reanalyse the data reported in this work is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

No experimental models or subjects were used.

METHOD DETAILS

Database selection and homology searching

Sequences from organisms with well-annotated genomes or solved structures of ATP synthases (e.g., *Saccharomyces cerevisiae*, *Homo sapiens*, *Arabidopsis thaliana*, *Acanthamoeba castellani*, *Polytomella*

parva, *Tetrahymena thermophila*, *Toxoplasma gondii*, *Trypanosoma brucei*, and *Euglena gracilis*) were retrieved from the NCBI sequence database to be used in initial queries against a database of 219 EukProt3[36] proteomes selected based on their taxonomic breadth and estimated genome completion.[80] BLAST[34,35,66] (v2.12.0, word size 3) queries were conducted into 219 selected proteomes. Top BLAST hits with e values below a threshold of 0.01 were then used as reciprocal BLAST queries against the proteomes with experimentally validated orthologues. A protein sequence was considered a validated orthologue if the experimentally validated ATP synthase subunits were returned as top hits in the reciprocal BLAST searches. To detect divergent homologues, validated orthologues were aligned using MUSCLE[61] to make HMMs to search all predicted proteomes lacking validated orthologues using hmmsearch v3.3.2.[37,38] Top hits from HMM searches were validated via reciprocal BLAST, phmmer, and HHpred v3.3.0[45] searches into predicted proteomes with known ATP synthase orthologues. In some cases, putative orthologues could not be validated and an in silico structural approach was taken. Homology modelling tools like Phyre2[39] and SWISS-MODEL[40] were initially used to identify proteins that have predicted structural similarities to ATP synthase components of *S. cerevisiae* and *H. sapiens*. Finally, AlphaFold2[41] was used to predict the structures of divergent ATP synthase subunits whose orthology could not be validated by other means. These predicted structures were aligned with solved structures PyMOL v2.5,[49–51] to identify similarities in protein structure. All ATP synthase subunit EukProt accessions and sequences are in Datas S1, S2, and S3, respectively.

QUANTIFICATION AND STATISTICAL ANALYSIS

No quantification or statistical analysis was performed.