Medical Research Archives

ESMED
European Society of Medicine

OPEN ACCESS

Published: January 31, 2023

Citation: Wong T, Chan C, et al., 2023. An Ectosymbiosis-Based Mechanism of Eukaryogenesis, Medical Research Archives, [online] 11(1). https://doi.org/10.18103/mra.v11i1.3365

DOI
https://doi.org/10.18103/mra.v11i1.3365

ISSN: 2375-1924

RESEARCH ARTICLE

An Ectosymbiosis-Based Mechanism of Eukaryogenesis

Tze-Fei Wong*, Chung-Kwon Chan, Xi Long, and Hong Xue

Division of Life Science, Hong Kong University of Science and Technology, Hong Kong, China.

* Correspondence: **bcjtw@ust.hk**

ABSTRACT

The mechanisms proposed for eukaryogenesis are divisible into mitochondria-early and mitochondria-late ones, where the mitochondriate-eukaryotes were evolutionary precursors or products of the amitochondriate-eukaryotes respectively. Analysis of prokaryote-to-eukaryote gene transfers in eukaryogenesis showed two tranches of high-intensity transfers from prokaryotes to eukaryotes mediated by the endosymbioses that gave rise to mitochondria and chloroplasts, and hundreds of medium-intensity transfers which included the transfer of hydrogenase and pyruvate: ferredoxin oxidoreductase genes from the *Thermoanaerobacter-Hungateiclostridium-Sporanaerobacter* group of bacteria to the amitochondriate eukaryotes. Since 94.5% of these medium-intensity transfers generated more than 100 inter-proteome similarity hits between each donor-recipient pair, they were not readily explicable by horizontal gene transfers or endosymbioses, pointing instead to the participation of a huge number of ectosymbiotic transfers. The euryarchaeon *Aciduliprofundum boonei* and the gammaproteobacterium *Escherichia coli* were among the foremost contributors of archaeal and bacterial genes to the eukaryotic DNA-apparati respectively, and the ratios of the genes in different eukaryotes indicated that *Microsporidia* have retained more of the genomic imprint of *Aciduliprofundum* than all other eukaryotes. These findings supported an ectosymbiosis-based mechanism of eukaryogenesis with *Aciduliprofundum* as the Archaeal Parent of Eukarya, and *Microsporidia* as the eukaryotes phylogenetically closest to the Last Eukaryotic Common Ancestor.

Keywords: *Aciduliprofundum*, archaeal parent, ectosymbiosis, endosymbiosis, eukaryogenesis, valyl-tRNA synthetase

INTRODUCTION

For over a century, endosymbiosis has been regarded as a core participant in eukaryogenesis[1]. The reason is effectiveness: unlike horizontal gene transfer (HGT) which can bring a small number of exogenous prokaryotic genes into the eukaryotic lineages per transfer, or gradual autogenous development of a postulated phagotrophic Archezoan similar to *Mastigamoeba*[2], a single endosymbiotic event can introduce a novel organelle into the host cell. There are more than twenty different endosymbiotic models, including the usage of an Archaeal Parent as host[3,4]; a hydrogen-dependent archaeon as host to a hydrogen-producing bacterial symbiont[5]; a progressive integration of a methanogenic archaeon and a delta-proteobacterium[6]; or a chimeric fusion of bacterium and archaeon with a single nucleus and a single kinetosome[7].

Different lines of evidence have favored an Archaeal Parent for the Eukarya domain: *Aciduliprofundum boonei* (or Abo; see three-letter abbreviations in Table 1) furnished a well-endowed candidate Archaeal Parent with top inter-proteome similarity bitscores among archaea toward *Giardia* and *Trichomonas*[8]; the Asgard and TACK archaeons provided essential genes for eukaryotic signature proteins[9] to the eukaryotes[10,11]; a large excess of archaea-derived over bacteria-derived ribosomal proteins was found in *Giardia*, *Trichomonas*, yeast, and humans[8]; and archaeal genes were more important than bacterial genes for the eukaryotes[12]. However, the question of whether the mitochondriate to eukaryotes (MTEs) emerged early prior to the amitochondriate eukaryotes (AMIs), or late following the AMIs has to be resolved. Although studies on SSU rRNA, elongation factor EF-1alpha and other proteins favored older ages of AMIs relative to MTEs[13-16], the discoveries of mitochondrial genes in the AMIs have gained momentum for the *degeneration theory* that AMIs arose from the degeneration of MTEs, thereby favoring older ages of MTEs in comparison with AMIs[17-23]. Recently, compromises between these views have also been introduced to the effect that, if observations were not completely compatible with a Last Eukaryotic Common Ancestor (LECA) with upfront mitochondria, the formation of some Pre-endosymbiont[24] or First Eukaryotic Common Ancestor (FECA)[10] might render possible a less abrupt initiation of eukaryogenesis.

Notably, even upfront mitochondria may not meet more than a limited fraction of the exogenous prokaryotic protein-coding genes required by the developing eukaryote lineages as suggested by the large varieties of such genes in the MTEs[25,26] or AMIs[8], for most mitochondrial and chloroplast DNA sections observed in the cell nucleus were gene fragments often less than 150 bp, and their transfer to the nucleus could be a complex process involving RNA intermediates or the acquisition of targeting signals[27]. Since exogenous prokaryotic protein-coding genes could be important to the development of LECA, the present study was directed to an examination of their biological sources and avenues of entry into the eukaryotes, in order to determine how the extremely narrow scope of major gene-donor endosymbionts consisting of only proteobacteria and cyanobacteria might be overcome by the evolving eukaryotes.

Table 1. Species names and their three-letter abbreviations. See Supplementary Table 1 for descriptions of species.

ABBR.	SPECIES NAME	ABBR.	SPECIES NAME
ARCHAEA		Bpr	*Bathycoccus prasinos*
Abo	*Aciduliprofundum boonei*	Bs1	*Blastocystis sp. subtype 1*
Acf	*Aciduliprofundum sp. MAR08-339*	Bs4	*Blastocystis sp. subtype 4*
Afu	*Archaeoglobus fulgidus*	Cel	*Caenorhabditis elegans*
Aia	*Acidilobus sp. 7A*	Cme	*Cyanidioschyzon merolae*
Alt	*C.Altiarchaeales archaeon*	Cne	*Cryptococcus neoformans*
Ape	*Aeropyrum pernix*	Cpa	*Cryptosporidium parvum*
Bat	*C.Bathyarchaeota archaeon*	Ddi	*Dictyostelium discoideum*
Csu	*C.Caldiarchaeum subterraneum*	Dme	*Drosophila melanogaster*

Csy	*Cenarchaeum symbiosum*
Fac	*Ferroplasma acidiphilum*
Ffo	*Fervidicoccus fontis*
Hal	*Halobacterium salinarum*
Hei	*C.Heimdallarchaeota archaeon*
Hgi	*Haloferax gibbonsii*
Hla	*Halobiforma lacisalsi*
Kcr	*C.Korarchaeum cryptofilum*
Lok	*Lokiarchaeum sp. GC14_75*
Mac	*Methanosarcina acetivorans*
Mar	*C. Marsarchaeota G2 archaeon*
Mbo	*Methanoregula boonei*
Mco	*Methanocella conradii*
Mes	*C.Methanosuratus sp.*
Min	*C.Methanomassiliicoccus intestinalis*
Mja	*Methanocaldococcus jannaschii*
Mka	*Methanopyrus kandleri*
Mnt	*Methanonatronarchaeum thermophilum*
Mph	*Methanophagales archaeon*
Mte	*C.Methanoplasma termitum*
Nca	*C.Nitrosocaldus cavascurensis*
Nga	*C.Nitrososphaera gargensis*
Odi	*C.Odinarchaeota archaeon*
Pae	*Pyrobaculum aerophilum*
Pfu	*Pyrococcus furiosus*
Psy	*Prometheoarchaeum syntrophicum*
Sso	*Saccharolobus solfataricus*
Tac	*Thermoplasma acidophilum*
Tho	*C.Thorarchaeota archaeon*
Tvo	*Thermoplasma volcanium*
Woa	*C.Woesearchaeota archaeon*
EUKARYA	
Aal	*Anncaliia algerae*
Acc	*Acanthamoeba castellanii*
Ago	*Andalucia godoyi*
Asu	*Acytostelium subglobosum*
Bbo	*Babesia bovis*
Bde	*Batrachochytrium dendrobatidis*
Bho	*Blastocystis hominis*
BACTERIA	
Aae	*Aquifex aeolicus*
Aba	*Acinetobacter baumannii*
Atu	*Agrobacterium tumefaciens*
Azo	*Azospirillum sp. M2T2B2*
Bap	*Buchnera aphidicola*
Bja	*Bradyrhizobium japonicum*
Blo	*Bifidobacterium longum*
Bsu	*Bacillus subtilis*
Cex	*Caldisericum exile*
Cje	*Campylobacter jejuni*
Cpo	*Cloacibacillus porcorum*
Dpu	*Dictyostelium purpureum*
Dre	*Danio rerio*
Eae	*Edhazardia aedis*
Ebi	*Enterocytozoon bieneusi*
Ecu	*Encephalitozoon cuniculi*
Ein	*Entamoeba invadens*
Enh	*Entamoeba histolytica*
Esi	*Ectocarpus siliculosus*
Gin	*Giardia intestinalis*
Gla	*Giardia lamblia*
Gth	*Guillardia theta*
Hsa	*Homo sapiens*
Imu	*Ichthyophthirius multifiliis*
Lbi	*Laccaria bicolor*
Mbr	*Monosiga brevicollis*
Mci	*Mucor circinelloides*
Mon	*Monocercomonoides sp. PA203*
Mpa	*Marchantia paleacea*
Nbo	*Nosema bombycis*
Ngr	*Naegleria gruberi*
Oco	*Ordospora colligate*
Per	*Perkinsela sp. CCAP 1560/4*
Pfa	*Plasmodium falciparum*
Pte	*Paramecium tetraurelia*
Ram	*Reclinomonas americana*
Sap	*Saprolegnia parasitica*
Sce	*Saccharomyces cerevisiae*
Slo	*Spraguea lophii*
Spo	*Schizosaccharomyces pombe*
Spu	*Spizellomyces punctatus*
Sra	*Strongyloides ratti*
Ssa	*Spironucleus salmonicida*
Tbr	*Trypanosoma brucei*
Tgo	*Toxoplasma gondii*
Tps	*Thalassiosira pseudonana*
Trh	*Trachipleistophora hominis*
Trv	*Trichomonas vaginalis*
Tth	*Tetrahymena thermophila*
Ttr	*Thecamonas trahens*
Vcu	*Vavraia culicis*
Kol	*Kosmotoga olearia*
Mau	*Mahella australiensis*
Mca	*Macrococcus caseolyticus*
Mtu	*Mycobacterium tuberculosis*
Pde	*Paracoccus denitrificans*
Pel	*Pelobacter sp. SFB93*
Pmo	*Petrotoga mobilis*
Rpr	*Rickettsia prowazekii*
Rru	*Rhodospirillum rubrum*
Rso	*Ralstonia solanacearum*
Spn	*Streptococcus pneumoniae*

Cur	*Curvibacter sp.*
Cvo	*Chelativorans sp. BNC1*
Dra	*Deinococcus radiodurans*
Dth	*Dictyoglomus thermophilum*
Eco	*Escherichia coli*
Hth	*Hungateiclostridium thermocellum*
Kae	*Klebsiella aerogenes*
Ssp	*Sporanaerobacter sp. NJN-17*
Syn	*Synechocystis sp. PCC 6803*
Tht	*Thermobaculum terrenum*
Tis	*Tistrella mobilis*
Tma	*Thermotoga maritima*
Tte	*Thermoanaerobacter tengcongensis*
Xca	*Xanthomonas campestris*

Analysis of the genes transferred from prokaryotes to eukaryotes yielded a heatmap that displayed not only the large endosymbiotic transfers associated with the formation of mitochondria and chloroplast but also a host of medium-intensity transfers from a wide spectrum of prokaryotic species that were in all likelihood mediated by ectosymbiosis. These ectosymbiotically transferred genes vastly enriched the variety of proteins in the eukaryote proteomes, and furnished an answer to the long-standing question regarding the origins of the hydrogenase and pyruvate:ferredoxin oxidoreductase (PFO) genes in the amitochondriate eukaryotes by tracing them to the *Thermoanaerobacter-Hungateiclostridium-Sporanaerobacter* group of bacteria. They also enabled multiple genera of prokaryotes to contribute genes to the DNA apparati of different eukaryotes; and these contributions indicated a particularly strong presence of *Aciduliprofundum* genes in the DNA apparati of *Microsporidia* and other AMIs, which supported an Archaeal Parent role for *Aboliprofundum*, the proximity of *Microsporidia* to the Last Eukaryotic Common Ancestor (LECA), and an ectosymbiosis-based mechanism of eukaryogenesis.

METHODS

Source of sequences

VARS sequences from different species were retrieved from NCBI GenBank release (ftp://ftp.ncbi.nlm.nih.gov/genomes/)[28]. Mitochondrial DNA-encoded protein sequences were retrieved from the RefSeq mitochondrial reference genomes in the NCBI Protein Database (https://www.ncbi.nlm.nih.gov/protein)[29].

Estimation of inter-proteome or inter-protein similarity bitscores

Inter-proteome and inter-protein similarity bitscores were performed as described[8]. The proteomes of various species were employed to construct a local BLAST database using makeblastdb[30], and query proteomes or proteins were searched against the local database using BLASTP with a BLOSUM62 matrix and thresholds set to e-value $<1x10^{-5}$, >25% percent identity and >50% query coverage. Only when the query and subject sequences that were the best match of each other, viz. when query sequence *n* from species 1 exhibited the highest bitscore toward subject sequence *m* among all proteins of species 2 and *vice versa*, were the data included in the estimation of similarity.

RESULTS

Heatmap of prokaryote-to-eukaryote gene transfers

To survey the gene transfers from prokaryotes to eukaryotes in the course of eukaryogenesis, a heatmap of the similarity bitscores between different prokaryote (on the x-axis) and eukaryote (on the y-axis) protein-coding genes was constructed. It showed numerous plausible transfers of genes from prokaryotes to eukaryotes at various x-y junctions. The large transfers from proteobacteria to the MTEs, and from Syn (representing cyanobacteria) to the algae Cme, Esi, Gth, Tps, and Bpr would correspond to the genes originating from the mitochondria- and chloroplast-yielding bacterial endosymbionts respectively. The transfer of Syn genes mainly to the algae with limited spillover to other eukaryotes attested to the high specificity of the similarity bitscores in identifying cognate pairs of gene donors and recipients. There were a wide variety of medium-intensity transfers from archaea and bacteria into both the AMI and MTE eukaryotes (Figure 1). These medium-intensity transfers were unlikely to be the outcome of HGTs, because 94.5% of them consisted of more than one hundred similarity hits each (Supplementary Table S2), whereas HGTs usually bring about transfers of single or a small number of genes. Instead, that they consisted largely of

ectosymbiotic gene transfers was suggested by the parallel usage of endosymbiosis in legume-*Rhizobium* interactions, and ectosymbiosis as in the binding of *Nostoc* to specialized leaf cavities on *Anthoceros* host in nitrogen fixation[31]; the critical role of ectosymbiosis in determining benthic biodiversity in the Arctic deep sea[32]; protist-spirochete interactions in the termite gut[33]; the exchanges of genes between DPANN and *Thermoplasma*[34]; and the repeatedly evolved host-specific ectosymbiosis between amphipods and sulfur-oxidizing bacteria in a cave ecosystem[35]. As well, the results in the heatmap were in accord with the findings of a wide variety of archaeal and bacterial genes in the genome of yeast[25] and Gla and Trv[8], as well as the apparent lack of any Pre-endosymbiont, FECA, or endosymbionts aside from the proteobacteria and cyanobacteria that led to the formation of mitochondria and chloroplast respectively[26]. Since the total number of prokaryotes that could contribute genes to the eukaryotes through ectosymbiosis would exceed by far the species included in the heatmap, they ensured an ample supply of exogenous prokaryotic genes to meet the logistic demand of eukaryotic development on an unprecedented scale.

Notably, the display by some archaea of an *accelerated gene adoption* (AGA) phenotype that enriched their genomes with more bacteria-derived genes than other archaeons indicated that highly AGA-active archaeons could recruit exogenous genes efficiently through non-HGT or hyper-HGT mechanisms, except for genes belonging to a subset of bacteria exemplified by Rpr, Bap, Cje, and Blo which resisted recruitment via AGA[8]. Surprisingly, on the heatmap in Figure 1, few Rpr, Bap, Cje and Blo genes were transferred to the AMI or MTE archaeons, suggesting that the Archaeal Parent excelled, like Abo or Tvo, in AGA-activity, making it a superb recruiter of ectosymbiosis-transferred genes in the course of eukaryogenesis.

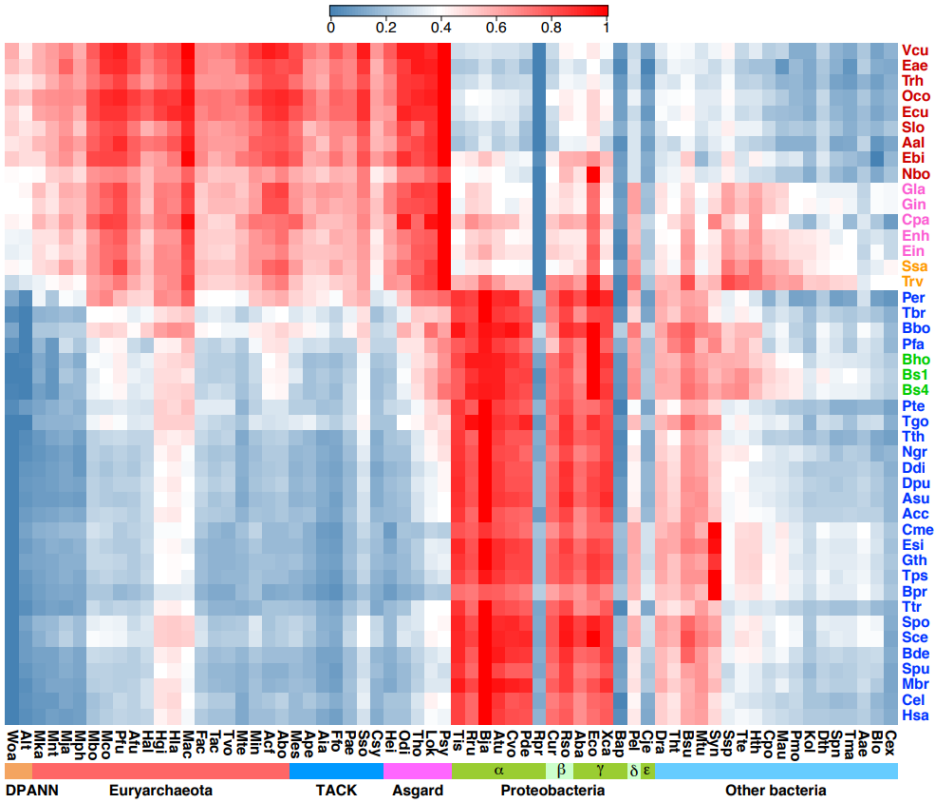

Figure 1. Heatmap of inter-proteome similarity bitscores between eukaryotes and prokaryotes. For each eukaryotic proteome, its similar bitscore toward a prokaryotic proteome is represented by the square at the intersection between a eukaryotic row and a prokaryotic column, and scaled linearly from 0 to 1.0 for each row according to the thermal scale, with zero marking the minimum and 1.0 the maximum bitscores. On the y-axis, AMIs with mitosomes are placed at the top followed by AMIs with hydrogenosomes and MTEs. Different MTEs are ordered on the y-axis broadly by

biological groups and according to the Abo/Eco ratios for the different groups (Figure 7). The aggregate similarity bitscores for individual prokaryote-eukaryote pairs are shown in Supplementary Table 2.

Identification of Archaeal Parent

Since it was probable that the earliest eukaryotes inherited some elements of the information system of the Archael Parent, the eukaryotic DNA-replication apparatus would be an attractive site to look for its footprint. Accordingly, the similarity bitscores exhibited by a range of prokaryotes toward the DNA-apparatus genes, viz. Cluster of Ortholog Groups[36] for replication, recombination and repair (COG-RRR), and for replication and repair (COG-L), of a variety of eukaryotes were estimated (Figure 2). The results demonstrated that:

a) A sizable number of archaeons and bacteria exhibited top similarity bitscores toward constituents of the DNA apparatus in different eukaryotes, in keeping with the indication by the heatmap that both archaea and bacteria donated protein-coding genes to the eukaryotes through ectosymbiosis.

b) Abo exhibited the largest number of top bitscores along with Eco toward the Microsporidia Eae and Vcu, the *Giardias* Gla and Gin, the *Trichomonas* Trv, and a number of top bitscores toward *Fungi* and C. *elegans* but less so toward the algae.

c) *B. subtilis* exhibited prominent bitscores toward the algae.

d) The Asgard and TACK archaeons exhibited relatively few top bitscores, except for *Methanosuratus* (Mes) which shared the largest number of top bitscores with Abo toward Eae. Interestingly, while Abo was co-prominent with Mes in top bitscores toward microsporidian Eae, it was co-prominent with *Pyrococcus* (Pfu) in top bitscores toward *Giardia*, suggesting that the phylogenetic branching of AMIs between the Microsporidia and Excavata groups was accompanied by divergent contents of prokaryotic proteins.

These results were indicative of the foremost contributions made by Abo to the DNA-apparatus proteins of eukaryotes, especially the AMIs, in support of Abo as the leading candidate Archaeal Parent. In contrast, because Eco and Bsu genes were continually recruited into both AMIs and MTEs in the course of eukaryogenesis, their genes were more eminent than Abo genes among the non-fungal and non-animal MTEs.

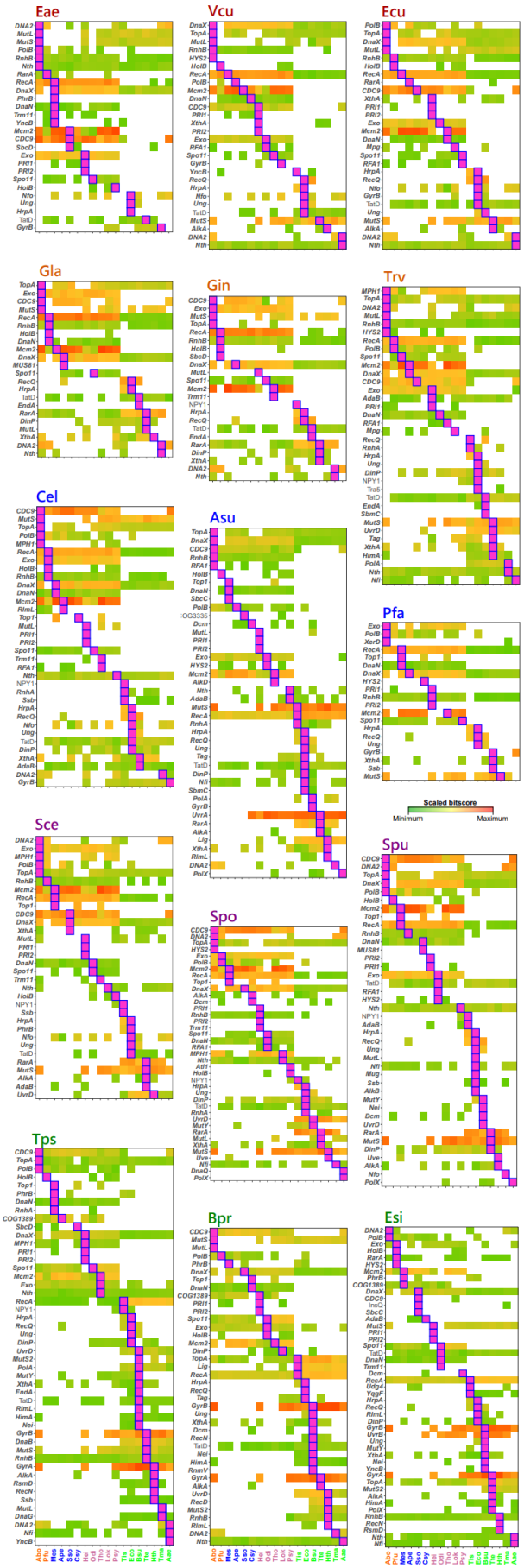

Figure 2. Similarity bitscores of COG groups in the eukaryotic DNA apparatus. Eukaryote panels are divided into Microsporidia (labeled red), Excavata (orange), mixed MTEs (blue), Fungi (purple) and Algae (green). Similarity bitscores displayed by different prokaryotes (x-axis) toward various COG groups (y-axis) are color-coded according to the thermal scale. Within each row, a purple rectangle inside a blue box marks the top bitscore of the row.

Mitochondrial DNA-encoded and mitochondria-like organelle proteins

While there was convincing evidence on the alphaproteobacterial origin of mitochondria[37], the large influxes of multiple proteobacterial proteins into the MTEs in the heatmap were consistent with chimerism among mitochondrial proteins[38-40]. Numerous species of mtDNA-encoded proteins showed top or high similarity bitscores toward the alphaproteobacterium *Tistrella* (Tis) (Figure 3 top panel), in agreement with the proximity between Tis DNA and mtDNA on the phylogenetic tree for alphaproteobacteria[41]. However, proteins from other alphaproteobacteria such as *Orientia*, *Pelagibacter* and *Paracaedibacter* also displayed top bitscores toward some species of mitochondria. For human mtDNA-encoded proteins HM1-11, Tis likewise shared top bitscores with *Rickettsia* and *Ehrlichia* (Figure 3 lower panel). Since *Rickettsia* and *Ehrlichia* are infectious agents of humans, their acquisitions of top bitscores for HM3 and HM2 respectively might have occurred during the infection process. In any case, the entry of the genomic sequences from multiple alphaproteobacteria into the same species of mitochondria was reminiscent of the competition between distinct strains of *Nostoc* ectosymbionts for binding to the leaf cavities of their *Anthoceros* host in nitrogen fixation[31], suggesting that different alphaproteobacterial DNAs could gain access to the mtDNA through ectosymbiosis, where they would be inserted possibly via non-homologous recombination at double-stranded breaks in the alphaproteobacterial DNA already positioned as the resident mtDNA, as in the insertion of mtDNA fragments into nuclear DNA[27].

The origins of the hydrogenase and PFO genes required for the activities of the mitosomes and hydrogenosomes of AMIs remained to be determined. Hydrogenosome resembled *Clostridium* metabolically[42]; and the PFOs from *Trichomonas*, *Giardia*, *Spironucleus*, and *Entamoeba* shared a single bacterial origin, even though the data were insufficient to identify that origin[43]. Since the heatmap showed medium-intensity ectosymbiotic gene transfers from the Clostridial/Firmicute species Tte, Hth and Ssp to *Microsporidia*, Gla, Gin, Enh, Ein, Ssa, Trv as well as MTEs like Bbo and Bho, the hydrogenase and PFO gene sequences from Tte were employed as specific probes for these genes, and the positive responses from AMIs and MTEs with mitosomes or hydrogenosomes (Figure 4, lines 9-10) suggest that their hydrogenase and PFO genes originated from bacteria related to the Tte-Hth-Ssp group; the AMI species with positive responses to the dehydrogenase or PFO probes were devoid of the subunits of the electron-transport proteins succinate dehydrogenase and fumarate reductase. In contrast, a variety of MTEs responded positively to the probes from Pde for these enzymes (lines 11-12) in line with their possession of an electron-transport chain[44].

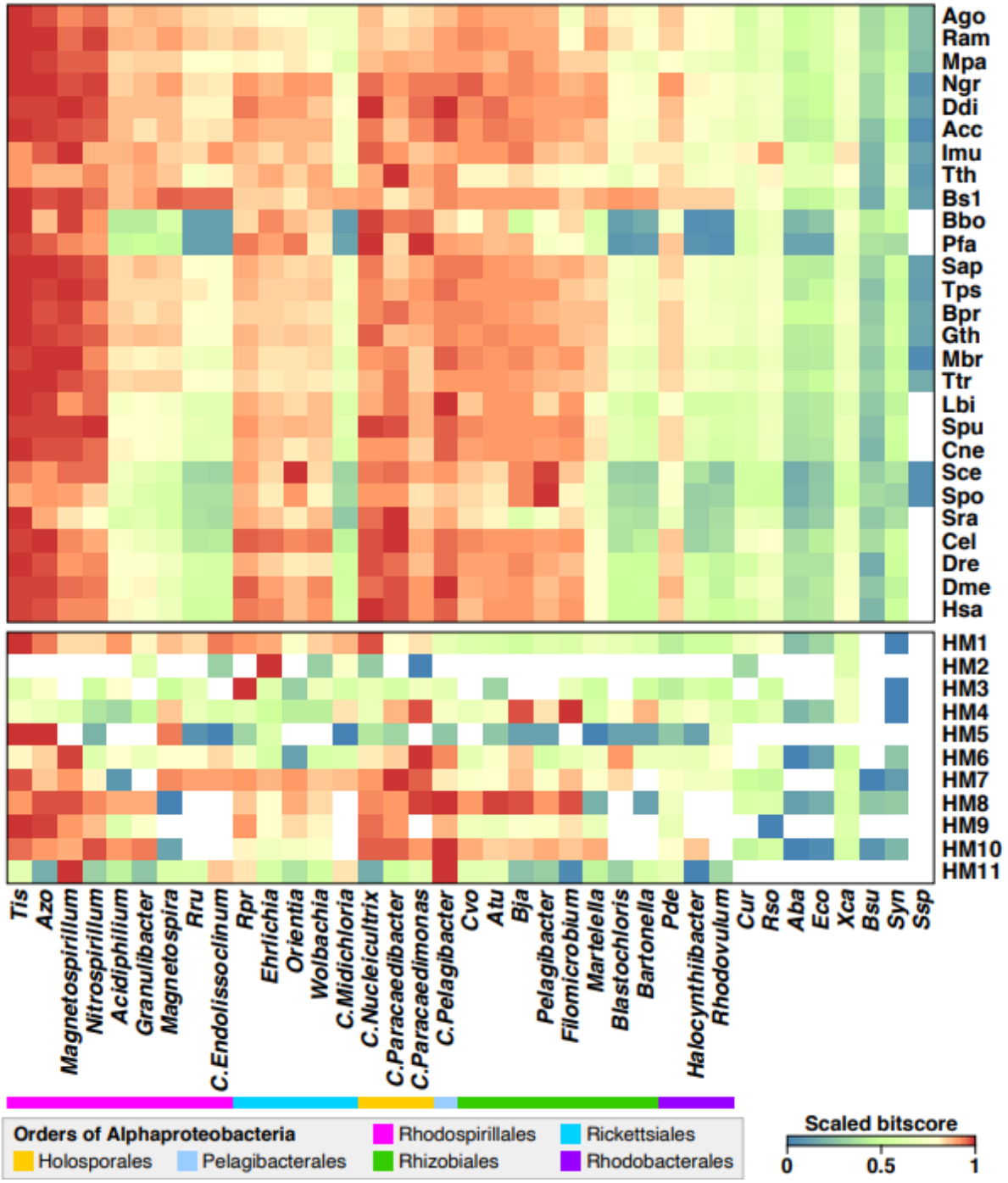

Figure 3. Similarity bitscores of mitochondrial constituents. (Upper panel) Bitscores between mtDNA-encoded proteins of different eukaryotes and various bacterial proteomes. (Lower panel) Bitscores between human mtDNA-encoded proteins (HM1-11, viz. NADH dehydrogenase subunits 1-4, 4L and 5, cytochrome b, cytochrome oxidase subunits I-III, and ATP synthase F0 subunit-6 respectively) and various bacterial proteomes.

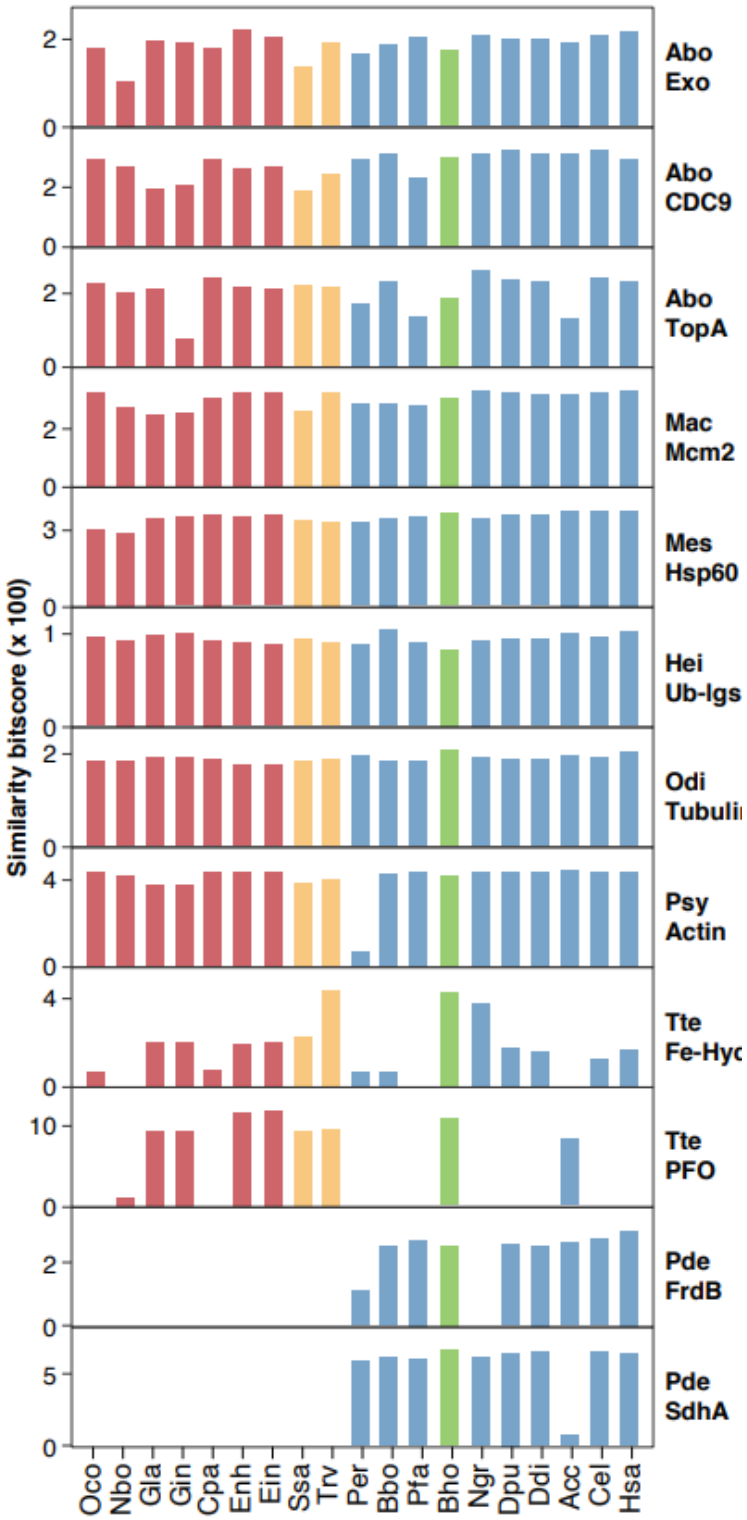

Figure 4. Similarity bitscores between prokaryotic probes and different eukaryotes. The bitscores pertained to twelve different probes (Exo, 5'-3' exonuclease; CDC9, ATP-dependent DNA ligase; TopA, DNA topoisomerase IA; Mcm2, DNA replicative helicase MCM subunit 2; Ub-lgs, ubiquitin-protein ligase; Fe-Hyd, iron only hydrogenase; FrdB Fe-S protein subunit, and SdhA flavoprotein subunit, of succinate dehydrogenase/fumarate reductase). Lines 1-8 show fairly uniform responses from eukaryotes to some non-electron transport protein probes, while lines 9-12 show the varied responses from distinct groups of eukaryotes to electron-transport protein probes.

Degeneration theory

Diplomonads, parabasalids and *Microsporidia*, all AMIs devoid of mitochondria, were postulated to be primitive eukaryotes that evolved prior to the advent of alphaproteobacteria-derived mitochondria based on the sequences of SSU rRNAs and protein markers such as elongation factor EF-1alpha[13-16]. However, this postulate was beset by possible long branch artefacts of rapidly evolving SSU rRNA, and an insert in EF-1alpha sequences, thereby favoring the degeneration theory that the AMIs were formed from degenerating *Fungi*[45,46]. For instance, the discovery of mitochondrial-like chaperonin 60 (cpn 60) genes in Trv, Gla and *Microsporidia* was regarded as evidence for the derivation of the cpn60 genes of AMIs from mitochondria[47]. However, the AMI and MTE cpn 60 genes were separated into distinct divisions on the cpn 60 phylogenetic tree without clear indication of a mitochondrial origin of the cpn 60 gene in AMIs. The application of the degeneration theory to cpn 60 was accordingly burdened with unresolved directional ambiguity, for genes could migrate not only from mitochondria to nucleus but also from nucleus to mitochondria[48].

In another application of the degeneration theory, the VARs sequences of Trv, Gla, a variety of MTEs and the bacteria Eco and *H. influenzae* were found to display a 37-aa insert just downstream from the KMSK motif, which was explained in terms of the derivation of Trv and Gla from degenerating MTEs[49]. However, the absence of this insert from all the archaea analyzed ruled against its bequeathal to the eukaryotes by an Archaeal Parent (Supplementary Figure 1); and the presence of an IKDENG insert in the VARS from Eco but not *Xanthomonas* (Xca) (Figure 5) pointed to an Xca-like origin of eukaryotic VARS. This was supported by the conservation of amino acid residues (marked by red asterisks) between Xca and eukaryotic VARS in the vicinity of the KMSK motif. When a maximum parsimony phylogenetic tree rooted by Xca was built for the eukaryotic VARS sequences, it allocated the microsporidian VARS sequences to a distinct division on the tree, and Trv VARS to a particularly low-branching position near the root (Figure 6). These tree features were consistent with the ectosymbiotic transfer of the 37-aa insert from Xca to Trv, which in turn passed it to the microsporidian and non-microsporidian divisions of the tree. Within the non-microsporidian division, the two *Giardia* species were lower branching than the MTEs that included the *Fungi*. Therefore the application of the degeneration theory to VARS led to the erroneous conclusion that the VARS of AMIs arose from degenerating *Fungi*.

```
BACTERIA
Xca LVTGFDIIFFWVARMIMATDSF------TGQVPFRDVYITGLIRDAQGQKMSKSKGNVLDPLDIIDG
Eco MVSGFDIIFFWIARMIMMTMHFIKDENGKPQVPFHTVYMTGLIRDDEGQKMSKSKGNVIDPLDMVDG
MICROSPORIDIAN
Vcu LETGSDILFFWVARMVMLSVEL------TGKVPFNTILLHGIVRDANGRKMSKSLGNVIDPLYVIEG
Eae LETGYDILFFWVARMTMLSIEL------TQTIPFKKILLHGIIRDAHGKKMSKSLGNVIDPLFIING
Trh LETGSDILFFWVARMVMLSLEL------TGKVPFNTILLHGIVRDANGRKMSKSLGNVIDPLYVIEG
Oco LETGSDILFFWVARMAMLGIEL------TRKIPFDQVLLHGIVRDAHGRKMSKSLGNVIDPIFVIDG
Ecu LETGSDILFFWVARMVMLGLEL------TGKVPFSQVLLHGIVRDAHGRKMSKSLGNVIDPIFVIDG
Slo LETGSDILFFWVARMVMMSLEL------TGKIPFKQILLHGIVRDTHGRKMSKSLGNVIDPIFVIEG
Aal LETGSDILFFWVARMIFMNVLL------TNTIPFTSVLLHGIIRDSMGRKMSKSLGNVIDPLFIING
Ebi LETGSDILFFWVARMVMLSYEL------IGTKPFNTILLHGIVRDAHGKKMSKSLGNVIDPIFVIDG
Nbo LETGSDILFFWVARMVMMSLEL------TGSIPFKQVLLHGIVRDAHGRKMSKSLGNVIDPLFVIDG
OTHER EUKARYOTES
Gla LETGNDIIFFWVARMVMLSLEL------YNVLPFKEVYLHALVRDAHGAKMSKSKGNVVDPIDVIKG
Gin LETGNDIIFFWVARMVMLSLEL------YNVLPFREVYLHALVRDAHGAKMSKSKGNVVDPIDVIKG
Enh LETGNDIIFFWVARMVMMSLEL------MDCLPFKEVLFHSIVRDAQGRKMSKSLGNIIDPIDVIEG
Ein LETGHDIIFFWVARMVMMSLEL------TDKIPFKQVLFHSIIRDAQGRKMSKSLGNVIDPIDVIEG
Trv METGSDILVFWVARMVMLCTHLA----PGGELPFKHIDLHQMVRDKSGRKMSKSLGNVLDPADLIQG
Bbo LETGNDIIFFWVARMVMMSLHL------VGKLPFNTVYLHPLVRDARGEKMSKSKGNVLDPLEVIEG
Tgo LETGHDILFFWVARMVMMSLQL------TDKLPFDTVFLHAMVRDAHGQKMSKSKGNVIDPLEVISG
Acc LETGHDILFFWVARMVMMGLNL------TGQLPFSQVLLHAMVRDAHGRKMSKSLGNVVDPIDVTEG
Sce LETGWDILFFWVTRMILLGLKL------TGSVPFKEVFCHSLVRDAQGRKMSKSLGNVIDPLDVITG
Bde LETGWDILFFWVARMVMMSLKF------NGVVPFKQVFCHAMIRDAHGRKMSKSLGNVIDPIDVIEG
Spu LETGRDILFFWVARMVMLGQKL------TGQIPFKQVFCHAMVRDAHGRKMSKSLGNVIDPMDVING
Cel LETGHDILFFWVARMVFMAQEL------TGKLPFKEILLHAMIRDAHGRKMSKSLGNVIDPLDVIRG
Hsa LETGHDILFFWVARMVMLGLKL------TGRLPFREVYLHAIVRDAHGRKMSKSLGNVIDPLDVIYG
          * **  **  **            **          **  * ***** **  **     *
```

Figure 5. Alignment of VARS sequences bearing a 37-aa insert downstream from the KMSK motif. Amino acid residues that were totally conserved between all the aligned bacterial, microsporidial, and other eukaryotic VARS sequences are marked by red asterisks below the Hsa sequence. The amino acids in the aligned sequences shown are numbered according to the VARS sequence of Xca (see Supplementary Figure 1), starting from residue 699 and ending with residue 765. The KMSK motif is located at residues 748 – 751.

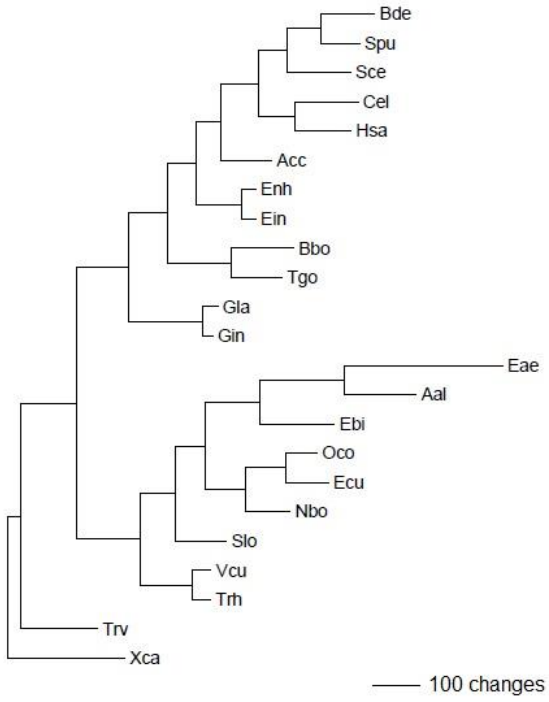

Figure 6. Phylogenetic tree of eukaryotic VARS sequences. The tree was rooted by the VARS of Xanthomonas campestris.

Ectosymbiosis-based mechanism of eukaryogenesis

The heatmap in Figure 1 showed widespread influx of prokaryotic genes into the eukaryotes, which would include the genes bequeathed to the eukaryotes by Abo as Archaeal Parent, and genes derived from archaea and bacteria through ecto- and endo-symbioses. Among the ectosymbiosis-derived genes, the Eco genes were outstanding for their continual adoption by both AMIs and MTEs, as indicated by the prominence of Eco genes in the DNA apparati of a variety of eukaryotes (Figure 2). Accordingly, the ratio between the Abo-derived and Eco-derived genes in a eukaryote genome would provide a measure of how far the eukaryote had evolved from Abo based on the relative abundance of inherited and ectosymbiotically-transferred genes. As shown in Figure 7, *Microsporidia* displayed the highest Abo/Eco ratios and therefore the closest phylogenetic relationship with Abo among the eukaryotes; the exceptionally low ratio displayed by Nbo among the *Microsporidia* stemmed from a particularly evident Eco-to-Nbo gene transfer on the heatmap. They exceeded the Gla and Gin ratios which in turn exceeded the Ssa and Trv ratios. The MTE ratios were all smaller than the Trv ratio, pointing to the remote relationship between Abo and the MTEs. These findings suggest that the decreasing Abo/Eco ratio could be employed to mark the order of emergence of different categories of eukaryotes, as outlined in the multistage eukaryogenesis mechanism in Figure 8.

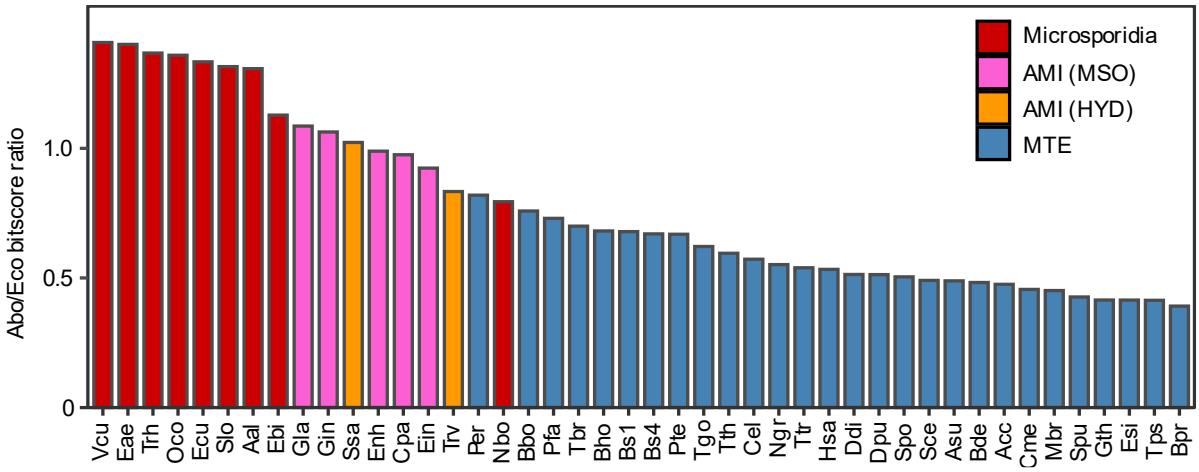

Figure 7. Ratios of eukaryotic similarity bitscores toward Abo relative to Eco. The eukaryotes (x-axis) are arranged in descending order of their Abo/Eco ratios. Columns of color-coded eukaryotes include *Microsporidia*; AMIs bearing mitosome (MSO); AMIs bearing hydrogenosome (HYD); and MTEs. See Supplementary Table 3 for the numerical ratios.

In this mechanism, Abo launched eukaryogenesis at Stage 0. At Stages 1 and 2, genes were recruited from archaeal and bacterial ectosymbionts, the ether-lipid membranes of Abo were gradually replaced by ester lipids, and Abo underwent maturation as Archaeal Parent under the stimulus of mainly archaeal ectosymbionts. At Stage 2, the continual influx of exogenous genes brought about the formation of cell nucleus, marking the birth of LECA. Stages 3 and 4 witnessed the accelerated uptake and adoption of bacterial genes giving rise to the mitosome and hydrogenosome, and continual expansion of host cell volume as in the case of *Thaumarchaeota* to giant cell size through association with a gammaproteobacterial ectosymbiont[50]. The expanded cell volume enabled the capture and accomodation of proteobacterial ectosymbionts through processes such as phagocytosis[51] or enwrapment with cytoplasmic membranes[52], turning them into endosymbionts and subsequently mitochondria at Stages 5 and 6, and finally mitochondria with different lineages of alphaproteobacteria mtDNA at Stage 7. Because the formation of cell nucleus preceded the capture

of endosymbionts, this mechanism represented an ectosymbiosis-based mechanism of eukaryogenesis.

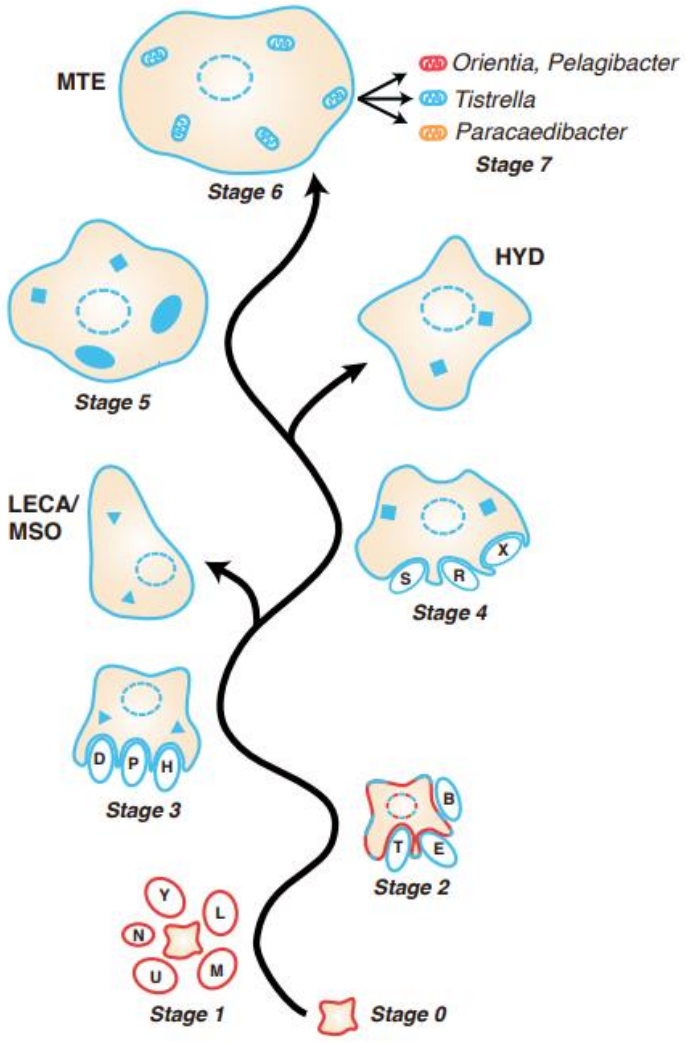

Figure 8. Ectosymbiotic mechanism of eukaryogenesis. Stages 0-7 are developmental stages of the Archaeal Parent lineage. Archaeal membranes are colored red, and bacterial membranes colored blue. Ellipses enclosed by dashed line represent cell nucleus; solid triangles mitosomes, solid squares hydrogenosomes, solid elipses endosymbionts, and elipses with cristae mitochondria. The letters Y, L, M, U, N, T, E, B, D, P, H, S, R and X stand for *Prometheoarchaeum*, *Lokiarchaeum*, *Methanosuratus*, *Pyrococcus*, DPANN, *Thermoanaerobacter*, *E. coli*, *B. subtilis*, *Paracoccus*, *Pelobacter*, *Hungateiclostridium*, *Tistrella*, *Ralstonia* and *Xanthomonas* respectively, representing just a small fraction of the archaea and bacteria taking part in eukaryogenesis as ectosymbiotic gene donors to the eukaryotes.

DISCUSSION

Eukaryotes are known to contain a series of ESPs that are fundamental to all eukaryotes, and the Asgard archaeons were major sources of the ESPs in Gla, amounting to 39% of total ESPs with Tho, Odi, and Lok as the foremost contributors[8], even though the Asgards might not have bridged the gap between prokaryotes and eukaryotes[53]. However, because there was no coalescence of the diverse Asgard ESP contributions through a single Asgard donor, Abo became the Archaeal Parent based on several criteria. First, the similarity bitscores displayed by Abo toward *Giardia* and *Trichomonas* exceeded those displayed by the Euryarchaeota archaeons *Ferroplasma*, *Halobiforma*, *Methanosarcina*, *Methanocella*,

Methanonatronarchaeum, *Methanoplasma*, and *Thermoplasma*, the TACK archaeons *Bathyarchaeota*, *Cenarchaeum*, *Korarchaeum*, *Marsarchaeota*, *Nitrososphaera*, and *Saccharolobus*, and the Asgard archaeons, *Heimdallarchaeota*, *Lokiarchaeum*, *Odinarchaeota*, *Prometheoarchaeum* and *Thorarchaeota* despite the small genome size (1.4 Mb) of Abo[8]. Secondly, most proteins involved in DNA precursor synthesis or DNA replication exhibited puzzling phylogenetic patterns that might be repeated 'inventions'[54]. The results in Figure 2 suggest that at least some of these repeated 'inventions' were changes in the prokaryotic source of the proteins, as illustrated by the prominence of *B. subtilis* proteins in the algae but not in the AMIs. Yet in the midst of these puzzling patterns, the Abo proteins were dominant in the DNA apparati of *Microsporidia, Giardia* and *Trichomonas,* with an unsurpassed number of top similar bitscores in each of the Eae, Vcu, Gla, Gin, and Trv DNA-apparati in accord with unusual proximity between Abo and the AMIs.

Thirdly, Abo was 83% similar in 16S rRNA to its cloest relative *Thermoplasma volcanium* (Tvo)[55], and both of these archaeons were devoid of a firm cell wall, for which reason *Thermoplasma* was the first proposed candidate Archaeal Parent for eukaryotes[3]. Both of them were also outstanding in the strength of their AGA phenotype, which would expedite the capture of bacterial genes[8]. However, Abo was a far more active gene contributor to the AMIs than Tvo in accord with the Archaeal Parenthood of Abo (Figure 1). Fourthly, the usefulness of the Abo/Eco ratio as a measure of the evolutionary distance traversed by any eukaryote from the Archaeal Parent also substantiated the status of Abo as the starting point of eukaryogenesis.

The highest Abo/Eco ratios exhibited by *Microsporidia* among the eukaryotes constituted key evidence for a *Microsporidia*-proximal LECA. In addition, within the non-*Microsporidian* division of the VARS tree, the *Giardia* species Gla and Gin, bearing smaller Abo/Eco ratios than *Microsporidia,* were lower branching than the MTEs in keeping with a core *Microsporidia-Giardia-Fungi*-Animalia evolutionary sequence that was also supported by the presence of four or more top bitscores for Abo-derived proteins in the *Microsporidia, Giardia, Trichomonas, Fungi* and C. *elegans*. Such a sequence indicates that there was probably no major parallel eukaryogenic lineage besides the Abo-initiated lineages. Notably, two independent multi-protein based fungal phylogenies also showed that *Microsporidia* were primitive *Fungi* rather than the end products from the degeneration of other *Fungi*[56.57].

Between different *Microsporidia*, there were such divergence, genome reduction and high evolutionary rates that the utility of their sequence-based phylogenies has been questioned[58]. However, *Microsporidia* developed hexokinases bearing secretion signal sequences, and an elaborate extrusion device consisting of a coiled polar tube with an anchoring disc for their spores; and there are 150 genera and more than 1,200 species of *Microsporidia* capable of infecting virtually all animal phyla accompanied by rapid proliferation within the varied hosts[59]. They can even direct host biology to the formation of cyst-like xenoma that provided high concentrations of energy and nutrients to support massive growth of the microsporidian parasites[60]. Such vitality suggests that the divergence and rapid evolution of *Microsporidia* could be due to their adaptations to widely different hosts more than intrinsic genomic instability. This possibility was supported strongly by conserved aspects of microsporidian molecular biology, in the relative uniformity of their elevated Abo/Eco ratios, the thoroughly conserved amino acid residues of their VARS in the vicinity of the KMSK motif together with other eukaryotes including humans as marked by red asterisks in the sequence alignment (Figure 5), and the coherent clustering of all the *Microsporidia* species within a separate division on the VARS tree (Figure 6).

In the ectosymbiosis-based mechanism (Figure 8), the proteome of Abo at Stage 0 contained 1,500 proteins, while a mitosome-containing eukaryote at Stage 3 exemplified by *Giardia* contained 5,000 proteins. This rate of proteome expansion would likely exceed the capability of HGT, but not ectosymbiosis especially when Abo the Archaeal Parent was endowed with top-ranked

AGA activity. Thus the evolutionary events mediated by ectosymbiosis in the present study could include the recruitment of distinct groups of archaeal and bacterial genes into the DNA apparati of different kinds of eukaryotes (Figure 2); the appearance of multiple species of alphaproteobacterial gene segments within the same mtDNAs (Figure 3); influx of the genes for hydrogenase and PFO into the AMIs (Figure 4); and the entry of a VARS with a 37-aa insert from Xca into the eukaryotes (Figure 5). These diverse ectosymbioses were in accord with the historical utilization of distinct groups of bacterial genes by different higher taxa of archaea[61], suggesting that ectosymbiosis might have long been employed by archaea to enhance their biodiversity with bacterial genes.

In this mechanism, Stages 1-7 of eukaryogenesis followed the curve of Abo/Eco decreases, with *Microsporidia* emerging first, then the other AMIs and finally the MTEs, which was opposite to the supposition made by the degeneration theory that the MTEs emerged first on the basis of two unjustified assumptions. First, the detection of a given protein in both the AMIs and the MTEs was interpreted by the theory as evidence for the evolutionary derivation of the AMI from the MTE. Such an interpretation would be valid if protein-coding genes were unidirectionally transferable from mitochondria to nucleus but not *vice versa*. However, there were not only mitochondria-to-nucleus transfers, but also nucleus-to-mitochondria transfers among living organisms. For instance, the present-day *Arabidopsis* mitochondrial tRNA repertoire includes 12 'native' tRNAs, and six plastid-derived tRNAs (now mitochondria-encoded)[48]. Secondly, the theory assumes that the hydrogenosomes in AMIs evolved from the mitochondria of the MTE precursors of the AMIs[21], which was inconsistent with the non-utilization of hydrogenases by present-day mitochondria, suggesting that the AMI hydrogenosomes must have acquired the hydrogenase genes from some other genome[37]. This expectation was fulfilled in the present study by the finding that the AMIs obtained their hydrogenase genes from bacteria related to *Thermoanaerobacter tengcongensis* via ectosymbiosis (Figure 4 line 9).

CONCLUSION

The three-domain structure of life proposed by Woese et al[62] represents the core of biology. To understand the functional significance of the separation of the three domains, elucidation of their origins becomes essential. In this regard, the identification by Xue et al[63] of a LUCA proximal to *Methanopyrus kandleri* (Mka), a resident of deep-sea hydrothermal vents based on analysis of alloacceptors tRNAs, has been confirmed by the top VARS-IARS bitscore of Mka among 5,000 species of organisms[8]; the hydrothermal vent-like habitat of LUCA[64]; the oldest age of the *Methanopyrus* lineage among archaea dating back to 2.8 Gya[65]; and the invention of the wobble rules of translation by Mka in using uniformly the GNN and UNN anticodon duo to decode the four codons in all family tetracodon boxes, and employment of tRNAs(Ser) from adjacent sequence space to read the non-contiguous UCN and AGY codons of Ser[66]. These findings verified the prediction that such vents represented the birthplace of life on Earth[67,68], which has turned the search for possible hydrothermal vents on Mars and other planets into a priority of exobiological exploration, and focused the search for an origin of eukaryotes within the Archaea.

Recently, evidence based on the use of dark fermentation by both *Thermococci* and primitive *Clostridia* for the production of hydrogen, and the sister-clade relationship between them in the minor-Themococcal division of the VARS tree has led to the proposal by Wong et al of a *Thermococci*-to-*Clostridia* evolutionary pathway for the emergence of Bacteria from Archaea[69]. As suggested by Nierhaus[70], this emergence was propelled by bacterial innovations such as the use of ester-lipids instead of ether-lipids, initiation of translation by N-formylMet, and most importantly the acquisition of elongation factor 4 (EF4/LepA). EF4/LepA, highly conserved in bacteria but absent from archaea, catalyzed ribosomal back-translocation and remobilization of stalled ribosomes, thereby enhancing protein synthesis five-fold under conditions of high intracellular magnesium ion or low temperature[71]. Therefore it would increase the survival rates of bacterial derivatives of archaea compared to the archaea themselves in crossing the

ice-cold seawater surrounding their original deep sea hydrothermal-vent habitats to spread to diverse niches in the mesothermal zone.

While the origin of Archaea shaped an interface between the biological and inorganic worlds, and the origin of Bacteria from Archaea consisted of a vertical transmission of genes accompanied by extensive mutations, genome reductions and possible influx of exogenous genes, present evidence suggests that Eukarya depended on a completely different kind of origin that deployed a continual stream of ectosymbioses and a small number of endosymbioses. On account of the small size of the Abo genome, MTEs with upfront mitochondria generated by engulfment of proteobacterial endosymbionts were extremely difficult to achieve owing to spatial constraint. As a result, the development of Abo as Archaeal Parent had to rely for a prolonged period on prokaryotic genes recruited through ectosymbiosis. This enforced reliance presented a remarkable opportunity for the eukaryotes. Instead of extracting the genes from a handful of endosymbionts and becoming just highly capable archaeons, Abo and its offspring were free to pick and choose any number or variety of ectosymbiotically transmitted genes for adoption from a vast pool of prokaryotic genes equal to the accumulated innovations from eons of archaeal and bacterial evolution. Consequently, they kept enriching themselves with all varieties of genes until they were large enough in cell size to capture endosymbionts like alphaproteobacteria and cyanobacteria, and transform them into power-house organelles to drive their unending expansion. Without the immense advantages of ectosymbiosis, one can only ponder how many more billions of years it would take the Eukarya to attain such accomplishments as language and use of tools.

Data availability

All data generated or analyzed during this study are included in this article and its supplementary information files. The following are available online. Supplementary Table 1: Descriptions of species analyzed. Supplementary Table 2: Inter-proteome similarity bitscores between eukaryotes and prokaryotes. Supplementary Table 3: Ratios between *Aciduliprofundum*- and *Escherichia coil*-derived protein-coding genes in different eukaryotes. Supplementary Figure 1: Organisms with or without a 37-aa insert in valyl-tRNA synthetase downstream of the KMSK motif.

Acknowledgments

This research was funded by grant number 5112-703-0110D-42002 from the University Grants Council of Hong Kong to Applied Genomics Center of Hong Kong University of Science and Technology.

Author contributions

Conceptualization, T.F.W. and H.X.; data analysis, C.K.C. and X.L.; writing, T.F.W., C.K.C., X.L. and H.X.. All authors have read and agreed to the published version of the manuscript.

Conflicts of interest

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; or in the writing of the manuscript.

References

1. Martin WF, Garg S, Zimorski V. Endosymbiotic theories for eukaryote origin. Phil Trans Roy Soc London B Biol Sci. 2015; 370: 20140330. doi: 10.1098/rstb.2014.0330.

2. Cavalier-Smith T. The origin of eukaryotic and archaebacterial cells. Ann N Y Acad Sci. 1987;503:17-54. doi:10.1111/j.1749-6632.1987.tb40596.x.

3. Searcy DG, Stein DB, Searcy KB. A mycoplasma-like archaebacterium possibly related to the nucleus and cytoplasms of eukaryotic cells. Ann N Y Acad Sci 1981;361:312-324. doi:10.1111/j.1749-6632.1981.tb46527.x.

4. Cox CJ, Foster PG, Hirt RP, Harris SR, Embley TM. The archaebacterial origin of eukaryotes. Proc Natl Acad Sci U S A 2008;105(51):20356-20361. doi:10.1073/pnas.0810647105.

5. Martin W, Müller M. The hydrogen hypothesis for the first eukaryote. Nature 1998;392(6671):37-41. doi:10.1038/32096.

6. Moreira D, Lopez-Garcia P. Symbiosis between methanogenic archaea and delta-proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. *J Mol Evol* 1998;47(5):517-530. doi:10.1007/pl00006408

7. Margulis L, Dolan MF, Guerrero R. The chimeric eukaryote: origin of the nucleus from the karyomastigont in amitochondriate protists. Proc Natl Acad Sci U S A 2000;97(13):6954-6959. doi:10.1073/pnas.97.13.6954.

8. Long X, Xue H, Wong JTF. Descent of bacteria and eukarya from an archaeal root of life. Evol Bioinform Online 2020;16:1176934320908267. Published 2020 Jun 23. doi:10.1177/1176934320908267.

9. Hartman H, Fedorov A. The origin of the eukaryotic cell: a genomic investigation. Proc Natl Acad Sci U S A 2002;99(3):1420-1425. doi:10.1073/pnas.032658599.

10. Eme L, Spang A, Lombard J, Stairs CW, Ettema TJG. Archaea and the origin of eukaryotes. Nat Rev Microbiol 2018;16(2):120. doi:10.1038/nrmicro.2017.154.

11. Raymann K, Brochier-Armanet C, Gribaldo S. The two-domain tree of life is linked to a new root for the Archaea. Proc Natl Acad Sci U S A 2015;112(21):6670-6675. doi:10.1073/pnas.1420858112.

12. Cotton JA, McInerney JO. Eukaryotic genes of archaebacterial origin are more important than the more numerous eubacterial genes, irrespective of function. Proc Natl Acad Sci U S A 2010;107(40):17252-17255. doi:10.1073/pnas.1000265107.

13. Vossbrinck CR, Maddox JV, Friedman S, Debrunner-Vossbrinck BA, Woese CR. Ribosomal RNA sequence suggests microsporidia are extremely ancient eukaryotes. Nature 1987;326(6111):411-414. doi:10.1038/326411a0.

14. Sogin ML, Silberman JD. Evolution of the protists and protistan parasites from the perspective of molecular systematics. Int J Parasitol 1998;28(1):11-20. doi:10.1016/s0020-7519(97)00181-1.

15. Kamaishi T, Hashimoto T, Nakamura Y, et al. Protein phylogeny of translation elongation factor EF-1 alpha suggests microsporidians are extremely ancient eukaryotes. J Mol Evol 1996;42(2):257-263. doi:10.1007/BF02198852.

16. Pittis AA, Gabaldón T. Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. Nature

2016;531(7592):101-104. doi:10.1038/nature16941.

17. Horner DS, Hirt RP, Kilvington S, Lloyd D, Embley TM. Molecular data suggest an early acquisition of the mitochondrion endosymbiont. Proc Biol Sci 1996;263(1373):1053-1059. doi:10.1098/rspb.1996.0155.

18. Roger AJ. Reconstructing early events in eukaryotic evolution. Am Nat 1999;154(S4):S146-S163. doi:10.1086/303290.

19. Van de Peer Y, Ben Ali A, Meyer A. Microsporidia: accumulating molecular evidence that a group of amitochondriate and suspectedly primitive eukaryotes are just curious fungi. Gene 2000;246(1-2):1-8. doi:10.1016/s0378-1119(00)00063-9.

20. Arisue N, Sánchez LB, Weiss LM, Müller M, Hashimoto T. Mitochondrial-type hsp70 genes of the amitochondriate protists Giardia intestinalis, Entamoeba histolytica and two microsporidians. Parasitol Int 2002;51(1):9-16. doi:10.1016/s1383-5769(01)00093-9.

21. de Graaf RM, Duarte I, van Alen TA, et al. The hydrogenosomes of Psalteriomonas lanterna. *BMC Evol Biol*. 2009;9:287. Published 2009 Dec 9. doi:10.1186/1471-2148-9-287

22. Degli Esposti M. Late mitochondrial acquisition, Really? Genome Biol Evol 2016;8(6):2031-2035. doi:10.1093/gbe/evw130.

23. Martin WF, Roettger M, Ku C, Garg SG, Nelson-Sathi S, Landan G. Late mitochondrial origin is an artifact. Genome Biol Evol 2017;9(2):373-379. doi:10.1093/gbe/evx027.

24. Gray MW. The pre-endosymbiont hypothesis: a new perspective on the origin and evolution of mitochondria. Cold Spring Harb Perspect Biol 2014;6(3):a016097. Published 2014 Mar 1. doi:10.1101/cshperspect.a016097.

25. Esser C, Ahmadinejad N, Wiegand C, Rotte C, Sebastiani F, Gelius-Dietrich G, Henze K, Kretschmann E, Richly E, Leister D, et al. A genome phylogeny for mitochondria among alpha-proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. Mol Biol Evol 2004; 21: 1643-1660. doi: 10.1093/molbev/msh160.

26. Rochette NC, Brochier-Armanet C, Gouy M. Phylogenomic test of the hypotheses for the evolutionary origin of eukaryotes. Mol Biol Evol 2014; 31: 832-845. doi: 10.1093/molbev/mst27.

27. Blanchard JL, Lynch M. Organellar genes: why do they end up in the nucleus?. Trends Genet 2000;16(7):315-320. doi:10.1016/s0168-9525(00)02053-9.

28. Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, W. Sayers E, et al. GenBank. Nucleic Acids Res. 2016; 44 D67-72. doi: 10.1093/nar/gkv127.

29. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucl Acid Res. 2016; 44: D733-745. doi: 10.1093/nar/gkv1189.

30. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinform. 2009; 10: 421. doi: 10.1186/1471-2105-10-421.

31. Enderlin CS, Meeks JC. Pure culture and reconstitution of the Anthoceros-Nostoc symbiotic association. Planta. 1983; 158: 157-165. Doi: 10.1007/BF00397709.

32. Hétérier V, David B, De Ridder C, Rigaud T. Ectosymbiosis is a critical factor in the local benthic biodiversity of the Antarctic deep sea. Mar Ecol Prog Ser. Published online 2008. doi:10.3354/meps07487.

33. Noda S, Ohkuma M, Yamada A, Hongoh Y, Kudo T. Phylogenetic position and in situ identification of ectosymbiotic spirochetes on protists in the termite gut. Appl Environ Microbiol. 2003;69(1):625-633. doi:10.1128/AEM.69.1.625-633.2003.

34. Golyshina OV, Toshchakov SV, Makarova KS, et al. 'ARMAN' archaea depend on association with euryarchaeal host in culture and in situ. Nat Commun. 2017;8(1):60. Published 2017 Jul 5. doi:10.1038/s41467-017-00104-7.

35. Bauermeister J, Ramette A, Dattagupta S. Repeatedly evolved host-specific ectosymbioses between sulfur-oxidizing bacteria and amphipods living in a cave ecosystem. PLoS One. 2012;7(11):e50254. doi:10.1371/journal.pone.0050254.

36. Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 2000;28(1):33-36. doi:10.1093/nar/28.1.33.

37. Andersson SG, Kurland CG. Origins of mitochondria and hydrogenosomes. Curr Opin Microbiol. 1999;2(5):535-541. doi:10.1016/s1369-5274(99)00013-2.

38. Abhishek A, Bavishi A, Bavishi A, Choudhary M. Bacterial genome chimaerism and the origin of mitochondria. Can J Microbiol. 2011;57(1):49-61. doi:10.1139/w10-099.

39. Wu B, Buljic A, Hao W. Extensive horizontal transfer and homologous recombination generate highly chimeric mitochondrial genomes in Yeast. Mol Biol Evol. 2015;32(10):2559-2570. doi:10.1093/molbev/msv127.

40. Ku C, Nelson-Sathi S, Roettger M, Garg S, Hazkani-Covo E, Martin WF. Endosymbiotic gene transfer from prokaryotic pangenomes: Inherited chimerism in eukaryotes. Proc Natl Acad Sci U S A. 2015;112(33):10139-10146. doi:10.1073/pnas.1421385112.

41. Martijn J, Vosseberg J, Guy L, Offre P, Ettema TJG. Deep mitochondrial origin outside the sampled alphaproteobacteria. Nature. 2018;557(7703):101-105. doi:10.1038/s41586-018-0059-5.

42. Whatley JM, John P, Whatley FR. From extracellular to intracellular: the establishment of mitochondria and chloroplasts. Proc R Soc Lond B Biol Sci. 1979;204(1155):165-187. doi:10.1098/rspb.1979.0020.

43. Horner DS, Hirt RP, Embley TM. A single eubacterial origin of eukaryotic pyruvate: ferredoxin oxidoreductase genes: implications for the evolution of anaerobic eukaryotes. Mol Biol Evol. 1999;16(9):1280-1291. doi:10.1093/oxfordjournals.molbev.a026218 .

44. Tielens AG, Rotte C, van Hellemond JJ, Martin W. Mitochondria as we don't know them. Trends Biochem Sci. 2002;27(11):564-572. doi:10.1016/s0968-0004(02)02193-x.

45. Hirt RP, Logsdon JM Jr, Healy B, Dorey MW, Doolittle WF, Embley TM. Microsporidia are related to Fungi: evidence from the largest subunit of RNA polymerase II and other proteins. Proc Natl Acad Sci U S A. 1999;96(2):580-585. doi:10.1073/pnas.96.2.580.

46. Keeling PJ. Congruent evidence from α-tubulin and β-tubulin gene phylogenies for a zygomycete origin of Microsporidia. Fungal Genet Biol. 2003: 298-309. doi:10.1016/S1087-1845(02)00537-6.

47. Roger AJ, Svärd SG, Tovar J, Clark CG, Smith MW, et al. A mitochondrial-like chaperonin 60 gene in Giardia lamblia: Evidence that diplomonads once harbored an endosymbiont related to the progenitor of mitochondria. Proc Nat Acad Sci USA 1998; 95: 229-234. doi:10.1073/pnas.95.1.229.

48. Marienfeld J, Unseld M, Brennicke A. The mitochondrial genome of Arabidopsis is

composed of both native and immigrant information. Trends Plant Sci. 1999;4(12):495-502. doi:10.1016/s1360-1385(99)01502-2.

49. Hashimoto T, Sánchez LB, Shirakura T, Müller M, Hasegawa M. Secondary absence of mitochondria in Giardia lamblia and Trichomonas vaginalis revealed by valyl-tRNA synthetase phylogeny. Proc Natl Acad Sci U S A. 1998;95(12):6860-6865. doi:10.1073/pnas.95.12.6860.

50. Muller F, Brissac T, Le Bris N, Felbeck H, Gros O. First description of giant Archaea (Thaumarchaeota) associated with putative bacterial ectosymbionts in a sulfidic marine habitat. Environ Microbiol. 2010;12(8):2371-2383. doi:10.1111/j.1462-2920.2010.02309.x.

51. Yutin N, Wolf MY, Wolf YI, Koonin EV. The origins of phagocytosis and eukaryogenesis. Biol Direct. 2009;4:9. Published 2009 Feb 26. doi:10.1186/1745-6150-4-9.

52. Baum DA, Baum B. An inside-out origin for the eukaryotic cell. BMC Biol. 2014;12:76. Published 2014 Oct 28. doi:10.1186/s12915-014-0076-2.

53. Da Cunha V, Gaia M, Gadelle D, Nasir A, Forterre P. Lokiarchaea are close relatives of Euryarchaeota, not bridging the gap between prokaryotes and eukaryotes. PLoS Genet. 2017;13(6):e1006810. Published 2017 Jun 12. doi:10.1371/journal.pgen.1006810.

54. Forterre P. The origin of DNA genomes and DNA replication proteins. *Curr Opin Microbiol*. 2002;5(5):525-532. doi:10.1016/s1369-5274(02)00360-0

55. Reysenbach AL, Liu Y, Banta AB, et al. A ubiquitous thermoacidophilic archaeon from deep-sea hydrothermal vents. Nature. 2006;442(7101):444-447. doi:10.1038/nature04921.

56. Capella-Gutiérrez S, Marcet-Houben M, Gabaldón T. Phylogenomics supports microsporidia as the earliest diverging clade of sequenced fungi. BMC Biol. 2012;10:47. Published 2012 May 31. doi:10.1186/1741-7007-10-47.

57. James TY, Kauff F, Schoch CL, et al. Reconstructing the early evolution of Fungi using a six-gene phylogeny. Nature. 2006;443(7113):818-822. doi:10.1038/nature05110.

58. Corradi N, Keeling PJ. Microsporidia: a journey through radical taxonomical revisions. Fungal Biol Rev. Published online 2009. doi:10.1016/j.fbr.2009.05.001.

59. Cuomo CA, Desjardins CA, Bakowski MA, et al. Microsporidian genome analysis reveals evolutionary strategies for obligate intracellular growth. Genome Res. 2012;22(12):2478-2488. doi:10.1101/gr.142802.112.

60. Li T, Fang Z, He Q, Yu B, Zhou Z. Characterizing the xenoma of Varimorpha necatrix provides insights into the most efficient mode of Microsporidian proliferation. Front Cell Infect Microbiology 2021. Doi: 10.3389/fcimb.2021. 699239.

61. Nelson-Sathi S, Sousa FL, Roettger M, et al. Origins of major archaeal clades correspond to gene acquisitions from bacteria. *Nature*. 2015;517(7532):77-80. doi:10.1038/nature13805.

62. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A. 1990;87(12):4576-4579. doi:10.1073/pnas.87.12.4576.

63. Xue H, Tong KL, Marck C, Grosjean H, Wong JTF. Transfer RNA paralogs: evidence for genetic code-amino acid biosynthesis coevolution and an archaeal root of life. Gene.

2003;310:59-66. doi:10.1016/s0378-1119(03)00552-3.

64. Weiss MC, Preiner M, Xavier JC, Zimorski V, Martin WF. The last universal common ancestor between ancient Earth chemistry and the onset of genetics. PLoS Genet. Published online 2018. doi:10.1371/journal.pgen.1007518.

65. Blank CE. Low rates of lateral gene transfer among metabolic genes define the evolving biogeochemical niches of archaea through deep time. Archaea 2012; 2012: 23. Doi: 10.1155/2012/843539.

66. Wong JTF, Ng SK, Mat WK, Hu T, Xue H. Coevolution theory of the genetic code at age forty: pathway to translation and synthetic life. Life. 2016;6(1):12. Published 2016 Mar 16. doi:10.3390/life6010012.

67. Corliss JB, Dymond J, Gordon LI, et al. Submarine thermal springs on the Galápagos Rift. Science. Published online 1979. doi:10.1126/science.203.4385.1073

68. Baross JA, Hoffman SE. Submarine hydrothermal vents and associated gradient environments as sites for the origin and evolution of life. Orig Life Evol Biosph. Published online 1985. doi:10.1007/BF01808177

69. Wong TF, Chan CK, Xue H. Thermococcus-to-Clostridia pathway for the evolution of the Bacteria domain. bioRxiv. Preprint posted online January 11, 2023. https://doi.org/10.21203/rs.3.rs-2461311/v1

70. Nierhaus KH. Cited in Wong JTF, Root of Life. In Prebiotic Evolution and Astrobiology; Wong, JTF., Lazcano, A., eds.; Landes Bioscience: Austin, TX, USA, 2009; pp. 120–144. doi:10.1201/9781498713986.

71. Pech M, Karim Z, Yamamoto H, Kitakawa M, Qin Y, Nierhaus KH. Elongation factor 4 (EF4/LepA) accelerates protein synthesis at increased Mg2+ concentrations. *Proc Natl Acad Sci U S A*. 2011;108(8):3199-3203. doi:10.1073/pnas.1012994108

SUPPLEMENTARY MATERIAL

Supplementary Tables 1, 2, and 3 are available here

Supplementary Figure 1. Species or organisms with or without a 37-aa insert in Valyl-tRNA synthetase downstream of the KMSK motif. All sequences are numbered according to the Xanthomonas campestris sequence which begins with residue 741 and ends with residue 816 in the alignment below. Notably, the algae Cme and Bpr among the eukaryote were devoid of the insert.

```
 ARCHAEA
Abo ILAPDGRPMHTSWGNV--VDPLEIID----------------------------------------EYGADALRFFAA
Afu VFGEDGRKMSKSLGNV--IVPEEVVE----------------------------------------KYGVDALRQWAA
Hei VVDSKGEKLSKSKGTD--VQPEKMIE----------------------------------------KYGGDAVRFYGA
Mac VLGPDGHKMSKSLGNV--ISPEEVTT----------------------------------------QYSADAFRQWGA
Mbo VLGEDGFKMSKSRGNV--IVPEDLVG----------------------------------------RYGADALRQWAA
Mco GLDPHGKAMHKSKGNI--VEPLPIVD----------------------------------------KYSADALRWWAA
Mja VFGEDGHKMSKSRGNV--VEPDEIIA----------------------------------------KYGADALRLWAS
Odi VLDEHGRAMHKSLGNI--VWVEPLLK----------------------------------------KYGADALRLFGC
Pfu VAGPDGRKMSKSYGNV--VSPEEVIP----------------------------------------KYGADALRLWTA
Psy IRDAKGQKISKSMENIEDYDPLKIIE----------------------------------------NVGADSLRYALI
Sso VLGPDGTRMSKSKGNV--VSPLDRVN----------------------------------------DFGADAIRMALL
Tho VVDENGETMSKSKGNS--PPPMPFVE----------------------------------------KYGADAMRMFGI
Tvo VFDMYGEKMSKSKGNI--VDIYAITD----------------------------------------KYGADALRFWAS
 BACTERIA
Aba VRDGEGQKMSKSKGNV--LDPLDLIDGIDLESLVAKRTTGLMNPKDAAKIEKSTRKEFPEGINAYGTDAVRFTFC
Atu VRDKNGQKMSKSKGNV--IDPLELID----------------------------------------EYGADALRFTLA
Bja VRDEKGAKMSKSKGNV--IDPLNLID----------------------------------------EYGADALRFTLA
Cje VKDEQGRKMSKSLGNV--IDPNESIK----------------------------------------EYSADILRFTLA
Mtu IRDESGRKMSKSKGNV--IDPLDWVE----------------------------------------MFGADALRFTLA
Pde VRDEKGAKMSKSKGNV--IDPLTLID----------------------------------------EYGADALRFTLT
Pel VRDASGQKMSKSKGNV--IDPLTVID----------------------------------------EYGTDAFRFTLA
Rru VRDEKGQKMSKSKGNV--IDPLDMTD----------------------------------------QYGTDALRFTLI
Rso VRDSEGKKMSKSEGNT--LDPVDLIDGIALEPLLVKRTTGLRRPKDAPNVEKRTRKEFPDGIPAFGADALRFTFA
Ssp VRDSQGRKMSKSLGNG--IDPLDVID----------------------------------------KYGADALRFTLV
Syn VRDENGKKMSKSANNG--IDPLLLIN----------------------------------------KYGTDALRYTLI
Tis VRDEKGQKMSKSKGNV--IDPIDLID----------------------------------------KYGADAVRFTLL
Tte VRDALGRKMSKSLGNG--IDPLEVIE----------------------------------------KYGADTLRFTLV
Eco IRDDEGQKMSKSKGNV--IDPLDMVDGISLPELLEKRTGNMMQPQLADKIRKRTEKQFPNGIEPHGTDALRFTLA
Xca IRDAQGQKMSKSKGNV--LDPLDIIDGISIEDLVAKRTSGLMQPRMAEKIEKATRKEFPDGIIAHGADALRFTIA
 MICROSPORIDIAN
Vcu VRDANGRKMSKSLGNVIDPLYVIEGIELDELAKSVT----STNLEPREVKTALEGQKKDFPMGIPRCGSDALRFT
Eae IRDAHGKKMSKSLGNVIDPLFIINGIKLSEMNNILKESHNNGYISNQELLRALDSQKKDFPRGVANCGADALRFA
Trh VRDANGRKMSKSLGNVIDPLYVIEGVQLDELAKSIT----ATNLDPKEIKAALEGQRKDFPMGIPRCGSDALRFT
Oco VRDAHGRKMSKSLGNVIDPIFVIDGCSLNELIATMK----SGNLDEKEVKVAEAALRKDFPNGIPRCGADALRFT
Ecu VRDAHGRKMSKSLGNVIDPIFVIDGCSLEKLISTMR----SGNLDEREVKRAEAVLRQDFPNGISRCGADALRFA
Slo VRDTHGRKMSKSLGNVIDPIFVIEGISLKGLNESI-----MTNLDKDEIKKAIEGQKKEYPNGIPQCGADALRFA
Aal IRDSMGRKMSKSLGNVIDPLFIINGCELKELNDSI-----SSTLSKKERDISLTYQKKTFPNGIKKCGADALRFC
```

```
Ebi VRDAHGKKMSKSLGNVIDPIFVIDGASQEELISKI-----SINVSNEEKKRAIASIKLDYPNGIPKCGADALRFA
Nbo VRDAHGRKMSKSLGNVIDPLFVIDGSSLENLIEVMK----SGNLALSEIKLAEKNLRKDFASGIAKCGADALRFT
 OTHER EUKARYOTES
Gla VRDAHGAKMSKSKGNVVDPIDVIKGITLQEMGDKVR----ATNLPPKEIERALELQSKDFPIGIPECGTDALRFA
Gin VRDAHGAKMSKSKGNVVDPIDVIKGITLQEMGDKVR----ATNLPPKEIERALELQSKDFPIGIPECGTDALRFA
Cpa VRDSQGRKMSKSLGNVIDPIEIIEGISFDDLNKKLD----QGNLPLQEIKKSKENNLKDFPDGIPECGADALRIG
Enh VRDAQGRKMSKSLGNIIDPIDVIEGISLKGLNDKLY----TYNLPEKECVIAAEGQKKDFPNGIEECGTDAMRFA
Ein IRDAQGRKMSKSLGNVIDPIDVIEGISLEGLNEKLK----IYNLSEKEIAIATKGQQMNFPHGIEECGTDAMRFA
Ssa VRDSKGEKMSKSKGNVIDPLDCIFGISLKDLHARLR----EGNLSENEIKLAEKLQKAEFPAGIAQCGTDALRMA
Trv VRDAQGRKMSKSLGNVIDPRHVINGIELEDLVAEIE----NSTFDDKEKKIAIDGRKADFPNGIPQCGTDAMRLA
Tbr VRDKNGEKMSKSKGNVIDPLFIISGVSLEALHDTVR----SGNLDEKEVSRALKLQRETFPNGIPECGSDALRFG
Bbo VRDARGEKMSKSKGNVLDPLEVIEGATLDSLIDKIN----NSSLPQGEIKKAIVLQKQQFPQGIPACGTDALRLG
Pfa IRDSRGEKMSKSKGNVVDPLDIIDGISLNKLHEKLY----EGNLPEKEIKRAIELQKKEFPKGIPECGTDALRFG
Bho IRDKYGRKMSKSLGNVIDPLEIINGCDLESMLEKIR----HGNLDPAEVERASQGKRQDFPEGIPMCGTDALRFG
Pte IRDKDGKKMSKSLGNVIDPLEIIDGTSLENLKSKIY----EGNLSKDEVERAIKQKEEEFPNGIPECGGDALRFG
Tgo VRDAHGQKMSKSKGNVIDPLEVISGISLQDLQAKLH----KGNLPEKEIKRAEEVLKKEFPKGIEACGCDALRLG
Tth IRDSQGEKMSKSKGNVIDPLEIIDGCNLQTLIQKIQ----EGNLDKKEMNRAVQLKSKEYPEGFPECGSDALRYG
Ngr VRDKQGRKMSKSLGNVIDPIDMIKGTTFEDLKRGIE---KNTNITKQEMKKALQGVQQEFPNGIPQCGTDALRFT
Ddi VRDSHGRKMSKSLGNVIDPNDVIKGISLDELIAKLY----EGNLDSKEIEKATSGVKSDFPTGIAECGTDAMRFA
Dpu IRDSQGRKMSKSLGNVIDPLNVINGITLKELKDNVL----SSNLTDKEKSIATKGLDKEFPQGIPQCGTDSLRLS
Asu VRDSHGRKMSKSLGNVIDPLEVISGITLDQLVENLK----KGNLDPVELDRATLGLKQDYPEGITECGTDAMRFA
Acc VRDAHGRKMSKSLGNVVDPIDVTEGIRLTDMHQKLR----EGNLEASEVEKAIKGQQKDFPNGISECGTDAMRFA
Cme VRDANGRKMSKTLGNVTDPLEVI----------------------------------------SKYGTDALRFT
Esi VRDKFGRKMSKTLGNVIDPLEVIYGCDLDTLHKKLE----VGNLPAKEMQKAKEGQKMDFPKGIPECGADALRFG
Tps VRDKDGRKMSKSLGNVIDPLEVINGCTLETLLEKLE----GGNLPPKEVARAKKDQEADFPEGIPECGSDALRFG
Bpr VRDEQGRKMSKSLGNVVDPLGVI----------------------------------------GDVGCDALRFT
Ttr VRDKSGRKMSKSLGNVLDPADLIQGASLDKLLAALE----GGNLPAGERARASSDLKAAFPDGFPAFGTDALRYA
Spo VRDSEGRKMSKSLGNVIDPMDIINGVTLENMKKALL----EGNLPISEVHKSSKQMEKAFPNGIPAQGIDIFRYG
Sce VRDAQGRKMSKSLGNVIDPLDVITGIKLDDLHAKLL----QGNLDPREVEKAKIGQKESYPNGIPQCGTDAMRFA
Bde IRDAHGRKMSKSLGNVIDPIDVIEGVTLQLLQERLE----KGNLDPRELVRARDGQKKDFPNGIPECGTDALRFG
Spu VRDAHGRKMSKSLGNVIDPMDVINGISLPLLHKRLE----EGNLDAREIKKAQEGQKRDFPNGIPQCGTDAMRFA
Mbr IRDKEGRKMSKSLGNVVDPMDVRNGITLEDLHERLL----EGNLDPAELERAKEGQKRQFPDGIKECGVDALRFA
Cel IRDAHGRKMSKSLGNVIDPLDVIRGISLNDLQAQLL----GGNLDEKEIAVAKEGQARDYPDGIPECGVDALRFA
Has VRDAHGRKMSKSLGNVIDPLDVIYGISLQGLHNQLL----NSNLDPSEVEKAKEGQKADFPAGIPECGTDALRFG
```