Bacterial genome chimaerism and the origin of mitochondria

Ankur Abhishek, Anish Bavishi, Ashay Bavishi, and Madhusudan Choudhary

Abstract: Many studies have sought to determine the origin and evolution of mitochondria. Although the *Alphaproteobacteria* are thought to be the closest relatives of the mitochondrial progenitor, there is dispute as to what its particular sister group is. Some have argued that mitochondria originated from ancestors of the order *Rickettsiales*, or more specifically of the *Rickettsiaceae* family, while others believe that ancestors of the family *Rhodospirillaceae* are also equally likely the progenitors. To resolve some of these disputes, sequence similarity searches and phylogenetic analyses were performed against mitochondria-related proteins in *Saccharomyces cerevisiae*. The 86 common matches of 5 *Alphaproteobacteria* (*Rickettsia prowazekii*, *Rhodospirillum rubrum*, *Rhodopseudomonas palustris*, *Rhodobacter sphaeroides*, and *Ochrobactrum anthropi*) to yeast mitochondrial proteins were distributed fairly evenly among the 5 species when sorted by highest identity or score. Moreover, exploratory phylogenetic analyses revealed that among these common matches, 44.19% (38) had branched most closely with *O. anthropi*, while only 34.88% (30) corresponded with *Rickettsia prowazekii*. More detailed phylogenetic analyses with additional *Alphaproteobacteria* and including genes from the mitochondria of *Reclinomonas americana* found matches of mitochondrial genes to those of members of the *Rickettsiaceae*, *Anaplasmataceae*, and *Rhodospirillaceae* families. The results support the idea that notable bacterial genome chimaerism has occurred en route to the formation of mitochondria.

Key words: mitochondria origin, endosymbiotic theory, *Alphaproteobacteria*.

Résumé : Plusieurs études ont tenté de déterminer l'origine et l'évolution de la mitochondrie. Même si l'on croit que les *Alphaproteobacteria* puissent être les parents les plus proches du progéniteur mitochondrial, l'identité de son groupe sœur est contestée. Certains ont soutenu que la mitochondrie provenait d'ancêtres de l'ordre de *Rickettsiales* ou plus spécifiquement de la famille des *Rickettsiaceae,* alors que d'autres croient que des ancêtres de la famille des *Rhodospirillaceae* constituent des progéniteurs tout aussi plausibles. Afin de résoudre certains de ces conflits, des recherches de similarité de séquences et des analyses phylogéniques ont été réalisées en prenant comme référence des protéines mitochondriales apparentées chez *Saccharomyces cerevisiae*. Les 86 correspondances communes de 5 *Alphaproteobacteria* (*Rickettsia prowazekii*, *Rhodospirillium rubrum*, *Rhodopseudomonas palustris*, *Rhodobacter sphaeroides* et *Ochrobactrum anthropi*) avec les protéines mitochondriales de levure étaient distribuées de façon assez uniforme parmi les 5 espèces lorsqu'elles étaient triées selon l'identité la plus élevée ou le score. De plus, des analyses phylogénique exploratoires ont révélé que parmi ces correspondances communes, 44,19 % (38) se greffaient davantage à *O. anthropi* alors que 34,88 % seulement correspondaient à *Rickettsia prowazekii*. Des analyses phylogéniques plus détaillées d'*Alphaproteobacteria* supplémentaires en incluant aussi des gènes mitochondriaux de *Reclinomonas americana* ont révélé des correspondances de gènes mitochondriaux avec des membres des familles des *Rickettsiaceae*, *Anaplasmataceae* et *Rhodospirillaceae.* Ces résultats appuient l'idée qu'un chimérisme important du génome bactérien est survenu au cours de la formation de la mitochondrie.

Mots-clés : origine de la mitochondrie, théorie endosymbiotique, *Alphaproteobacteria*.

[Traduit par la Rédaction]

Introduction

Ancestors of the members of the current-day *Alphaproteobacteria* subgroup are thought to be the progenitors of mitochondria (Gray et al. 1999; Lang et al. 1999*b*; Embley and Martin 2006), and species that belong to this group inhabit a large variety of environments, from coastal regions to deep-sea sediments (Giovannoni et al. 2005). Members of this subgroup are pathogenic or free-living and have diverse metabolic capabilities such as the use of a variety of organic compounds and the production of secondary metabolites (Kersters et al. 2006) under aerobic, anoxygenic photosynthetic, or semiaerobic growth conditions. Other members of this subgroup are anaerobic heterotrophs or possess the abil-

Received 5 July 2010. Revision received 2 November 2010. Accepted 10 November 2010. Published on the NRC Research Press Web site at cjm.nrc.ca on 7 January 2011.

A. Abhishek, A. Bavishi, A. Bavishi, and M. Choudhary.[1] Department of Biological Sciences, Sam Houston State University, Huntsville, TX 77431, USA.

[1]Corresponding author (e-mail: mchoudhary@shsu.edu).

ity to grow under such conditions (Imhoff and Trüper 1992). As such, these factors are supportive of the origin of mitochondria from *Alphaproteobacteria* precursors.

However, the determination of the organism most related to eukaryotic mitochondria and the placement of the mitochondrial tree branch are contested (Wu et al. 2004). Some argue that organisms in the order *Rickettsiales* are the closest relatives to mitochondria (Fitzpatrick et al. 2006). Other studies have more specifically submitted that the *Rickettsiaceae* family (Emelyanov 2001) or the *Pelagibacter*, *Rickettsiaceae*, and *Anaplasmataceae* families together (Williams et al. 2007) constitute the sister group to mitochondria. Consequently, *Rickettsia prowazekii*, an organism in the *Rickettsiaceae* family, is thought to be very closely related to mitochondria for several reasons, including the finding that several gene clusters in the mitochondrial genome are reminiscent of those found in *Rickettsia prowazekii*, including similar sets of proteins involved in ATP production and transport functions (Andersson et al. 1998). However, the relationship of mitochondria to organisms in the order *Rickettsiales* has been challenged based on phylogenomics, which demonstrates a close relationship of mitochondria to *Rhodospirillum rubrum* (Esser et al. 2004). More recently, another study compared *Chlamydomonas* proteins to a total of 354 sequenced genomes, including 286 eubacteria, 24 archaebacteria, and 44 eukaryotes, and found that members of the *Rhizobiales* and *Rhodobacterales* have many proteins that link them closely to mitochondria (Atteia et al. 2009).

Over time, mitochondria have undergone 3 important forms of evolution: the loss of genes from the mitochondrial genome, the transfer of genes from the mitochondrial to the nuclear genome, and the specialization of nuclear genes to serve mitochondria-related functions (Karlberg et al. 2000; Andersson et al. 2003). Now, most mitochondria contain up to 20 protein-coding genes with exceptions, like the mitochondrion of *Reclinomonas americana*, which possesses 67 protein-coding genes (Lang et al. 1997; Andersson et al. 2003). Other organisms do not possess mitochondria, but rather evolutionary homologs such as hydrogenosomes, which oxidize pyruvate and make ATP by substrate-level phosphorylation, and mitosomes, which do not seem to play any direct roles in ATP synthesis (Embley and Martin 2006). Many studies seeking to find the closest relative to the progenitor of mitochondria have only looked at the genes internal to the present-day, largely reduced mitochondria and correspondingly neglected the large portion of genes present in the nuclear genomes of eukaryotes. Since many genes have been deleted from the transitional mitochondrial genome and eukaryotic nuclear genomes contain large proportions of genes encoding for mitochondria-related functions, these analyses may be somewhat incomplete or limited in truly ascertaining mitochondrial origin. In this study, protein homology analysis was conducted on 5 species from the *Alphaproteobacteria* against 773 proteins of *Saccharomyces cerevisiae* that perform mitochondrial functions to obtain a clearer picture of the origin of mitochondria. The common matches of these 5 *Proteobacteria* species exhibited about the same level of gene homologies with yeast mitochondrial genes and represented similar functional categories. However, exploratory phylogenetic analyses demonstrated significant gene matches to both *Rickettsia prowazekii* and *Ochrobactrum anthropi*. Subsequent phylogenetic analyses among a larger group of *Alphaproteobacteria* found matches of mitochondrial genes to those of members of the *Rickettsiaceae*, *Anaplasmataceae*, and *Rhodospirillaceae* families. These results reveal that mitochondria may have originated before the separation of these bacterial lineages and that many of its original genes could have been lost or muddled by gene deletion and horizontal gene transfer to and from distantly related microorganisms. Thus, genome chimaerism, which has been ongoing among prokaryotes, should be taken into account in discussing the origin of mitochondria. The impact of the current results on theories of mitochondrial origin and potential modifications to such theories are discussed.

Materials and methods

Sequence similarity search

Sequence similarity searches are one method by which to discover the phylogenetic connections among organisms as well as to estimate the time elapsed from the origin of a particular biological event. A total of 773 mitochondria-related proteins (both internal to mitochondria and nuclear genes) of *Saccharomyces cerevisiae* were batch downloaded from http://www.yeastgenome.org. Including nuclear-coded mitochondrial proteins in analyses of the origin of mitochondria is necessary, especially as most mitochondrial functions are now coded for by nuclear genes, presumably through the transfer or slippage of genes from mitochondria to the nucleus. Also, *Saccharomyces cerevisiae* has been used before in mitochondria origin studies (Esser et al. 2004). Preliminary search data revealed similarity between yeast mitochondrial proteins and 5 *Alphaproteobacteria* species (*Rickettsia prowazekii*, *Rhodospirillum rubrum*, *Rhodopseudomonas palustris*, *Rhodobacter sphaeroides*, and *O. anthropi*). The selection of these species was not completely arbitrary but rather calculated, as these species belong to orders to which mitochondria have been proposed to have diverged from. As such, the mitochondrial proteins were blasted against the proteins of these 5 species using BLASTP (Altschul et al. 1997), which is conservative in its sequence searches. The results were filtered by setting a cutoff bit-score value of 100, which corresponds approximately to an E-value cutoff of 10^{-22}. In the cases that more than one match was found to a mitochondrial gene for a given bacterium, only the highest match was retained. Each homologous bacterial protein was then matched to its clusters of orthologous groups (COG) functional group to examine the roles of these genes in relation to their mitochondrial counterparts.

Phylogenetic analyses

Given the ancient origin of mitochondria and significant protein divergence over such a large time span, protein similarity analysis alone is most probably insufficient to ascertain the closest relatives to mitochondria, although it can be extremely informative. Two sets of phylogenetic analyses were performed. The first set of analyses was exploratory. For this set, the common matches from the BLASTP results for the 5 bacterial species (*Rickettsia prowazekii*, *Rhodospirillum rubrum*, *Rhodopseudomonas palustris*, *Rhodobacter*

Table 1. Organism characteristics and similarity to yeast mitochondrial proteins.

Lifestyle	Chromosome or plasmid	Genome size (nt)	No. of genes	No. of hits (score ≥100)	% Genome-encoding mitochondrial proteins	G+C content (%)
***Rickettsia prowazekii* Madrid E.**						
Free-living and pathogenic	CI	1 111 523	886	109	12.30	29.00
***Rhodospirillum rubrum* ATCC 11170**						
Free-living	CI	4 352 825	3870	136	3.45	65.40
	Plasmid (unnamed)	53 732	50	0	3.45	59.80
***Rhodobacter sphaeroides* 2.4.1**						
Free-living	CI	3 188 609	3092	132	3.30	69.00
	CII	943 016	866	10	3.30	69.00
	Plasmid A	114 045	90	0	3.30	69.20
	Plasmid B	114 178	102	2	3.30	70.10
	Plasmid C	105 284	90	0	3.30	63.80
	Plasmid D	100 828	98	0	3.30	63.80
	Plasmid E	37 100	32	0	3.30	66.80
***Rhodopseudomonas palustris* CGA009**						
Free-living	CI	5 459 213	4891	148	3.02	65.00
	Plasmid pRPA	8 427	7	0	3.02	60.40
***Ochrobactrum anthropi* ATCC 49188**						
Free-living and pathogenic	CI	2 888 297	2803	123	2.96	56.10
	CII	1 895 911	1725	16	2.96	56.20
	Plasmid pOANT01	170 351	180	4	2.96	56.20
	Plasmid pOANT02	101 491	96	0	2.96	58.50
	Plasmid pOANT03	93 589	91	3	2.96	54.30
	Plasmid pOANT04	57 138	37	0	2.96	55.30

sphaeroides, and *O. anthropi*) were then subjected to further phylogenetic analysis. Phylogenetic analysis of the individual genes is more robust and sensitive to developmental factors. Although constructing trees for concatenated mitochondrial proteins can be useful in ascertaining mitochondrial origin, the extremely ancient and dynamic nature of mitochondrial evolution suggests that such an approach can mask evolutionary features and variability in individual proteins. Since only the common genes among all 5 *Alphaproteobacteria* were selected, these chosen yeast mitochondrial proteins must be homologous to their ancient eubacterial counterparts and were most probably present in the originator bacteria of mitochondria. For the second set of phylogenetic analyses, a BLASTP search was run against the mitochondrion of *Reclinomonas americana* and the genomes of 49 additional *Alphaproteobacteria* using the common yeast genes that matched those of the original 5 selected *Alphaproteobacteria*. The *Reclinomonas americana* mitochondrion contains 67 protein-coding genes, 30 structural RNAs, and a large sequence of mtDNA (Lang et al. 1997; Plasterer et al. 2001). A list of all the *Alphaproteobacteria* used for this set can be found in Supplementary Table S1[2]. The BLASTP results were filtered using a bit-score value of 100, and the remaining genes were subjected to phylogenetic examination. Geneious version 4.8 was used to create alignments and construct phylogenetic trees (Drummond et al. 2010). Protein sequences were aligned using MUSCLE (Edgar 2004), an algorithm noted for both its speed and precision. PhyML (Guindon and Gascuel 2003) along with the WAG model (Whelan and Goldman 2001) was used to construct unrooted maximum likelihood trees. Bootstrap values were computed for all trees using 100 replications.

Results and discussion

Characteristics of the 5 bacteria are shown in Table 1. The total number of yeast mitochondrial protein matches with a score ≥100 for each of the 5 bacteria, also shown in Table 1, reflects that the genome of *Rickettsia prowazekii* has the lowest number of protein matches (109), representing ~12% of its genome. The other 4 species (*Rhodospirillum rubrum*, *Rhodobacter sphaeroides*, *Rhodopseudomonas palustris*, and *O. anthropi*) showed a significantly higher number of protein matches (136–148) to the yeast mitochondrial homologs, although these numbers reflect only ~3% of their respective genomes, as their genomes are approximately 5–6 times larger than the genome of *Rickettsia prowazekii*. The strength of homologies among 86 total common protein matches, shared by all 5 species with the yeast mitochondrial proteins as shown in Table 2, reflect the diverse distribution of the common gene matches. In essence, *Rhodospirillum rubrum* and *Rhodopseudomonas palustris* contained the largest amount of genes with the highest percent identity and (or) score values, as reflected

[2] Supplementary data for this article are available on the Journal Web site (http://cjm.nrc.ca).

Table 2. Common *Alphaproteobacteria* gene matches to *Saccharomyces cerevisiae* mitochondrial genes based on percent identity (I) and score (S).

		Rickettsia prowazekii		*Rhodobacter sphaeroides*		*Rhodospirillum rubrum*		*Rhodopseudomonas palustris*		*Ochrobactrum anthropi*	
S. cerevisiae ID*	Gene†	% I‡	S§	% I‡	S§	% I‡	S§	% I‡	S§	% I‡	S§
		37.75	153	31.87	119	35.34	150	34.77	129	33.33	147
Q0105	COB	55.74	402	47.77	345	54.62	388	53.12	389	50.94	347
Q0110	BI2	53.44	161	55.12	160	48.34	164	52.11	167	51.47	159
Q0115	BI3	59.87	209	54.60	193	56.36	209	56.10	206	51.53	189
Q0120	BI4	59.23	297	50.21	257	54.31	274	52.87	283	50.82	249
YBL099W	ATP1	60.95	620	68.88	668	70.98	670	68.79	675	70.12	688
YIL125W	KGD1	37.32	650	42.46	800	44.35	811	44.79	834	44.46	828
YJR121W	ATP2	74.07	687	75.85	705	76.10	711	76.86	711	75.59	707
YOR065W	CYT1	44.13	210	37.78	132	42.23	181	42.38	188	45.00	175
YOR142W	LSC1	51.34	259	52.86	259	53.72	236	54.21	261	51.33	243
YBR263W	SHM1	41.10	327	44.71	345	46.78	352	43.54	343	47.61	359
YEL030W	ECM10	61.13	694	60.23	691	60.53	715	62.46	736	59.93	721
YER087W		39.80	166	28.10	134	35.96	146	32.33	150	34.05	157
YIL094C	LYS12	32.43	153	36.44	195	34.36	142	36.29	205	33.24	162
YJL200C	ACO2	26.23	174	26.45	159	27.51	128	27.34	118	26.54	169
YJR045C	SSC1	63.91	724	60.92	712	64.06	743	64.74	766	63.46	744
YLR304C	ACO1	29.53	213	27.62	176	29.68	208	26.85	182	28.38	187
YLR369W	SSQ1	50.00	587	48.98	570	50.32	602	51.28	611	50.24	594
YBR084W	MIS1	42.86	216	41.46	428	44.83	239	44.41	230	44.98	238
YDR226W	ADK1	34.13	132	46.19	184	43.66	175	47.59	168	41.90	156
YER170W	ADK2	32.84	119	40.10	135	34.36	110	37.56	113	32.66	105
YGR244C	LSC2	44.53	311	46.46	348	47.98	339	44.81	323	46.33	336
YPL262W	FUM1	61.09	575	63.79	592	57.39	513	61.64	560	63.79	600
YDR232W	HEM1	44.84	358	49.06	395	50.59	419	47.63	400	46.74	374
YOR274W	MOD5	32.99	141	25.26	102	28.96	113	30.23	128	28.05	114
YBR221C	PDB1	59.82	399	60.42	417	63.44	425	59.33	389	58.28	393
YDR036C	EHD3	28.30	127	28.25	134	33.90	152	32.95	155	30.36	159
YDR148C	KGD2	56.09	278	43.32	326	59.48	288	59.21	295	60.09	301
YMR207C	HFA1	28.15	230	32.03	225	34.22	238	27.18	235	32.34	222
YBL080C	PET112	35.10	274	31.98	206	32.16	219	31.94	230	31.87	235
YDR268W	MSW1	38.26	229	38.53	227	40.00	232	37.88	214	37.09	227
YDR341C		25.83	122	25.55	135	25.39	111	25.17	120	24.96	113
YEL050C	RML2	43.88	204	42.05	202	43.19	203	42.56	204	43.88	212
YFL022C	FRS2	31.42	132	31.09	127	29.85	118	32.47	133	27.92	117
YGL143C	MRF1	42.97	239	43.43	246	42.00	241	41.38	242	43.04	232
YGR094W	VAS1	27.53	261	43.07	379	39.26	634	37.07	587	38.36	592
YGR220C	MRPL9	44.08	184	45.97	187	45.5	187	45.97	184	46.19	184
YHL004W	MRP4	35.19	142	33.91	134	31.58	130	32.43	122	31.84	132
YHR011W	DIA4	31.50	182	27.46	175	32.52	186	33.53	182	32.87	180
YLR382C	NAM2	37.69	541	37.61	535	39.79	581	39.07	545	39.14	562
YMR293C		35.10	243	37.29	254	37.35	249	36.11	239	35.19	247
YNR036C		50.00	110	54.55	134	56.91	137	61.11	134	56.91	133
YOL033W	MSE1	33.88	244	37.07	214	33.94	239	30.67	233	30.82	236
YPL040C	ISM1	29.69	239	31.18	476	30.78	445	32.22	469	33.56	447
YPL097W	MSY1	35.38	206	37.61	217	31.48	206	30.52	214	36.22	239
YPL104W	MSD1	30.38	230	29.97	224	30.49	232	27.87	207	28.31	219
YDL104C	QRI7	28.30	127	30.85	121	31.48	120	29.23	120	36.53	166
YLR259C	HSP60	56.46	557	54.94	554	56.65	565	56.65	578	56.65	554
YOR232W	MGE1	38.36	104	35.40	102	65.91	474	39.49	122	44.44	134
YPL132W	COX11	43.35	157	39.88	126	30.49	232	38.95	138	40.64	154
YHR120W	MSH1	31.41	352	30.45	343	31.22	356	31.01	365	31.33	386
YER061C	CEM1	38.20	273	39.41	294	40.59	309	40.63	286	47.30	238
YBL022C	PIM1	37.22	422	38.36	478	39.24	479	41.12	483	42.02	338

Table 2 (*concluded*).

S. cerevisiae ID*	Gene†	*Rickettsia prowazekii*		*Rhodobacter sphaeroides*		*Rhodospirillum rubrum*		*Rhodopseudomonas palustris*		*Ochrobactrum anthropi*	
		% I‡	S§	% I‡	S§	% I‡	S§	% I‡	S§	% I‡	S§
YBR024W	SCO2	41.18	151	37.93	116	41.61	138	40.29	117	36.08	115
YBR037C	SCO1	42.25	130	38.51	114	40.71	126	40.71	111	33.94	104
YBR227C	MCX1	37.72	232	38.83	232	37.8	231	37.36	236	37.17	231
YCL017C	NFS1	59.70	478	35.16	203	41.97	264	36.08	224	36.68	182
YDL033C	SLM3	32.81	159	30.33	142	31.33	154	29.49	150	32.51	167
YDR194C	MSS116	30.75	163	29.84	164	31.59	172	29.34	163	31.07	158
YDR258C	HSP78	53.48	759	55.02	757	56.07	789	57.27	808	56.33	790
YEL024W	RIP1	57.25	172	48.34	145	52.47	192	52.11	158	45.61	163
YEL052W	AFG1	28.76	120	34.59	179	32.16	194	32.67	188	31.06	192
YER017C	AFG3	49.16	406	45.70	386	48.83	403	47.37	407	49.26	416
YGL064C	MRH4	24.66	100	24.71	104	25.53	108	24.59	101	26.57	115
YGL236C	MTO1	45.81	503	47.58	496	44.88	471	47.36	502	46.28	503
YGR028W	MSP1	39.13	128	39.67	139	38.17	135	35.74	155	38.43	135
YHL014C	YLF2	37.89	229	33.16	203	38.42	234	36.41	226	36.68	226
YHL035C	VMR1	27.06	156	25.90	157	35.43	156	26.86	162	27.20	156
YHR024C	MAS2	25.42	128	26.34	131	27.27	148	28.70	160	27.54	147
YHR106W	TRR2	45.74	277	46.84	274	44.38	260	47.63	281	43.96	263
YHR168W	MTG2	42.60	129	44.00	129	38.14	125	43.79	127	41.07	128
YJL102W	MEF2	32.65	371	31.17	323	33.12	375	34.19	362	31.35	343
YLR069C	MEF1	42.90	573	41.25	525	41.19	540	41.65	529	43.80	564
YLR163C	MAS1	32.67	194	36.66	239	35.17	241	33.64	240	33.57	224
YLR188W	MDL1	30.06	218	34.94	237	33.33	194	36.40	274	36.01	242
YLR289W	GUF1	44.48	506	45.06	510	44.88	520	46.78	506	44.87	512
YML110C	COQ5	43.80	206	48.06	227	43.56	208	46.12	213	48.83	228
YMR023C	MSS1	35.35	258	31.54	220	35.10	249	32.66	239	33.00	251
YMR089C	YTA12	48.83	436	47.77	422	42.70	427	46.80	444	50.22	458
YMR301C	ATM1	44.69	506	45.97	541	43.62	509	44.81	508	41.86	492
YOL023W	IFM1	38.98	382	37.38	376	38.40	387	36.71	382	35.94	384
YOR187W	TUF1	65.91	471	64.99	467	65.91	474	67.34	479	66.33	477
YPL091W	GLR1	28.64	151	36.64	274	34.91	263	36.07	273	37.83	273
YPL270W	MDL2	31.04	222	33.55	239	30.65	215	34.83	300	36.27	281
YPR024W	YME1	46.98	447	47.87	454	47.55	477	47.65	457	49.80	462
YKL194C	MST1	35.21	262	33.90	269	36.43	280	35.26	270	32.91	262

**Saccharomyces cerevisiae* mitochondria-related gene ID.

†Gene name for the given *S. cerevisiae*. A blank means that no gene name was available or given.

‡Identity value for the highest organism gene match to the corresponding *S. cerevisiae* gene.

§Score value for the highest organism gene match to the corresponding *S. cerevisiae* gene.

in Fig. 1. The distribution of the COG functions for protein matches with a bit-score ≥100, as shown in Fig. 2, further suggests that the most abundant group of proteins are shared among all 5 species. The COGs represent proteins involved in energy production, translation, post-translation, and amino acid metabolism, which all are theoretically necessary for the establishment of mitochondria after the occurrence of an earlier symbiotic association event. A χ^2 test on a contingency table detailing the COG distributions among the organisms was insignificant (χ^2 = 19.76, $p > 0.05$), demonstrating that the distributions of the COGs do not differ markedly between the organisms. Therefore, one species' functional distributions were not necessarily more favorable or specialized than that of another for permitting the eventual origin of mitochondria. Although it must be noted that sequence similarity analysis does not provide the ability to make phylogenetic inferences, it does provide a foundation from which further analyses can be performed.

A sample phylogenetic tree for the first set of phylogenetic analyses is depicted in Fig. 3. The analyses, detailed in Table 3, revealed a concentration of gene matches with 2 organisms. More specifically, of the 86 genes, 44.19% (38) had branched most closely with *O. anthropi*, while only 34.88% (30) corresponded with *Rickettsia prowazekii*, as shown in Fig. 4. *Rhodobacter sphaeroides* and *Rhodospirillum rubrum* had 4 (4.65%) and 1 (1.16%) highest gene matches, respectively. The remaining genes had different combinations of bacteria matching as sister groups, with 8 (9.30%) matching on a *Rhodobacter sphaeroides – Rhodopseudomonas palustris – Rhodospirillum rubrum* cluster, 2 (2.32%) matching on a *Rhodobacter sphaeroides – Rhodopseudomonas palustris – O. anthropi* cluster, and 1 (1.16%)

Fig. 1. Distribution of the 86 common gene matches based on highest identity and highest score. The figure reveals that there is no concentration of gene matches in any of the organisms based on the BLAST results. I, identity; S, score.

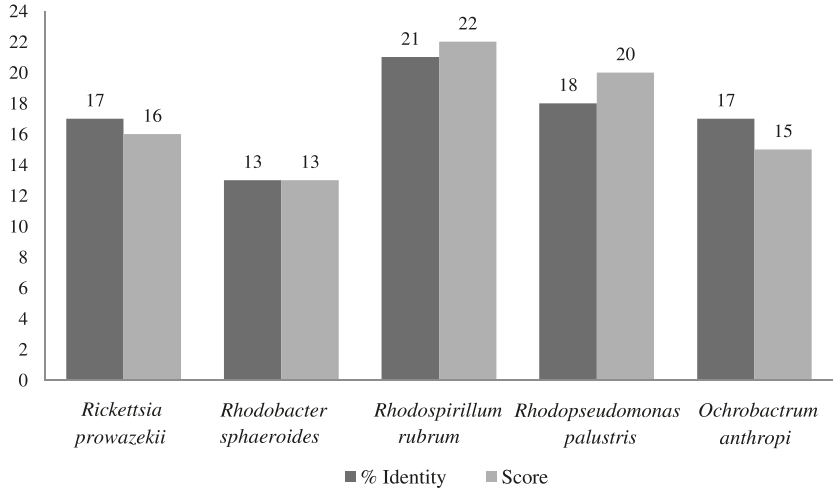


Fig. 2. Mitochondrial homologs distributed by cluster of orthologous groups (COG) for 5 *Alphaproteobacteria*. The letters on the *x*-axis represent the abbreviations for specific COG functions: translation, ribosome structure, and biogenesis (J); post-translational modification, protein turnover, and chaperone (O); cell division and chromosome partitioning (D); energy production and conservation (C); amino acid transport and metabolism (E); nucleotide transport and metabolism (F); coenzyme metabolism (H); lipid metabolism (I); inorganic transport and metabolism (P); secondary metabolite synthesis (Q); carbohydrate transport and metabolism (G); transcription (K); DNA replication, recombination, and repair (L); cell envelope biogenesis, outer membrane (M); function unknown (S); defense mechanisms (V); general function predicted only (R).

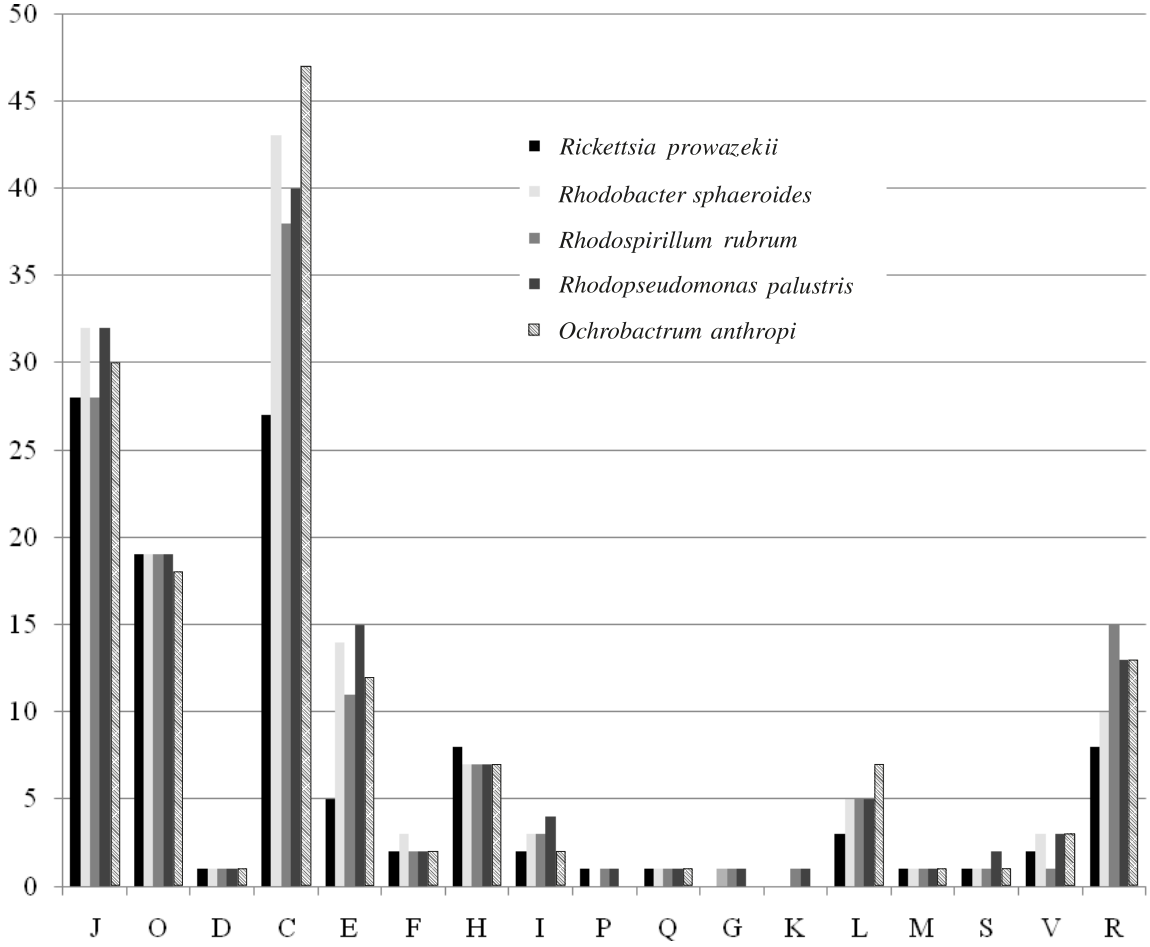


Fig. 3. A sample phylogenetic tree for the *Saccharomyces cerevisiae* gene denoted Q0085. The highest matching genes to Q0085 from the organisms *Rickettsia prowazekii*, *Rhodobacter sphaeroides*, *Rhodospirillum rubrum*, *Rhodopseudomonas palustris*, and *Ochrobactrum anthropi* were aligned with the corresponding yeast gene and a maximum likelihood tree was constructed. The numbers on the branches represent the bootstrap support for those nodes. The scale bar provides a measure of the substitutions per site for each branch.

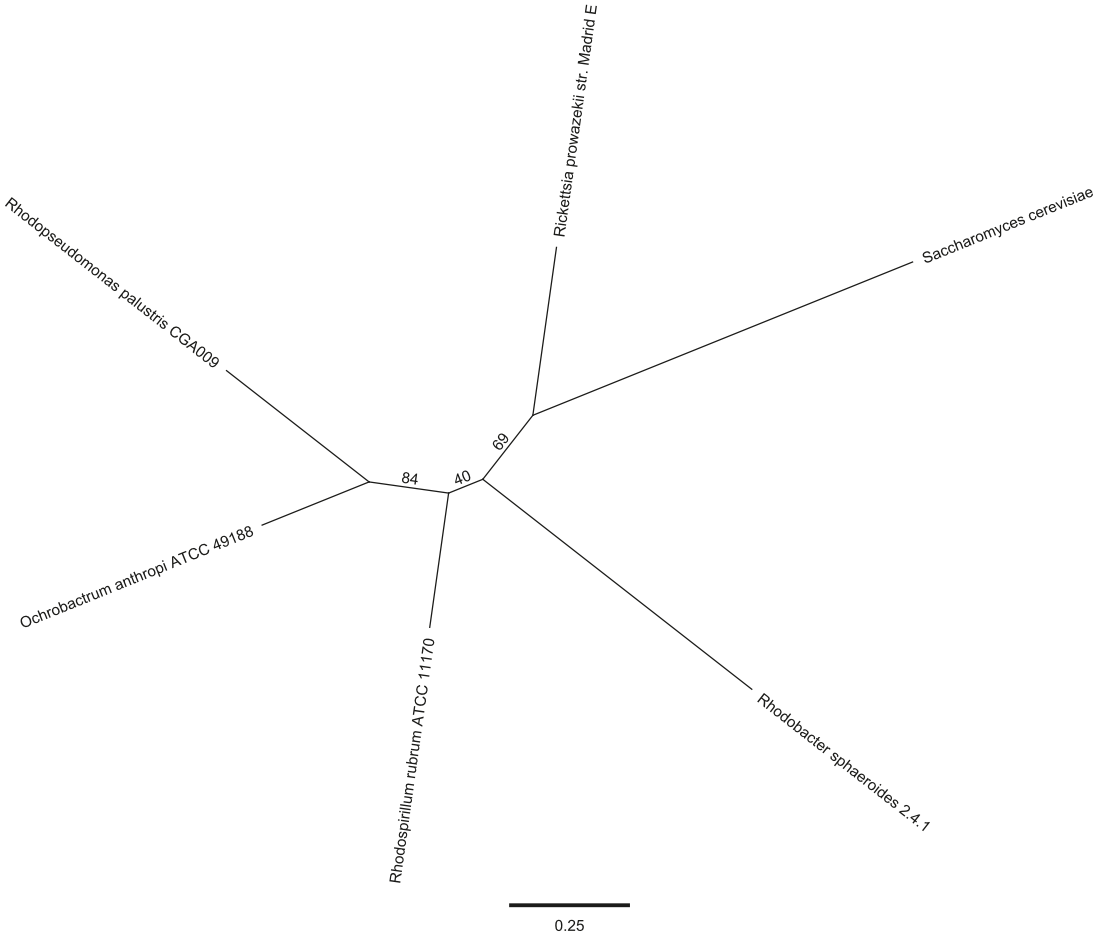
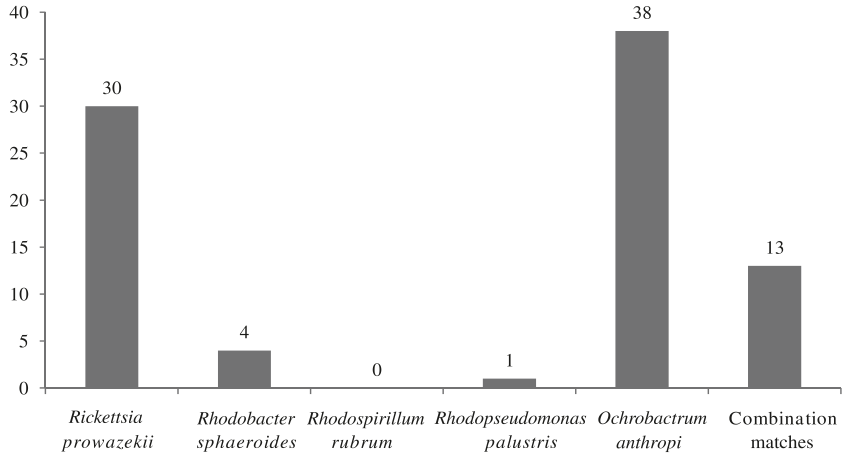


Fig. 4. Distribution of the 86 common genes based on matching sister group. The majority of genes matched to the organisms *Ochrobactrum anthropi* and *Rickettsia prowazekii*.



each matching to a *O. anthropi* – *Rickettsia prowazekii* – *Rhodospirillum rubrum* cluster, *Rickettsia prowazekii* – *O. anthropi* cluster, and *Rhodobacter sphaeroides* – *Rickettsia prowazekii* – *Rhodospirillum rubrum* cluster. Of the 38 *O. anthropi* matches, 19 (50%) had bootstrap values ≥90, while of the 30 *Rickettsia prowazekii* matches, 12 (40%) had bootstrap values ≥90. Of the 18 remaining matches, only 5 (27.78%) had similarly high bootstrap values. The

Table 3. Mitochondrial gene sister groups.

Saccharomyces cerevisiae ID	G+C content (%)*	Sister group†	Bootstrap value‡
Q0085	23.7	*Rickettsia prowazekii*	69
Q0105	18.9	*Rhodobacter sphaeroides – Rhodopseudomonas palustris – O. anthropi*	74
Q0110	19.4	*Rhodobacter sphaeroides*	55
Q0115	17.6	*Rickettsia prowazekii*	49
Q0120	18.6	*Rickettsia prowazekii*	83
YBL022C	40.7	*O. anthropi*	88
YBL080C	41.1	*O. anthropi*	96
YBL099W	43.7	*Rhodobacter sphaeroides – Rhodopseudomonas palustris – O. anthropi*	69
YBR024W	40.0	*O. anthropi*	45
YBR037C	44.9	*Rickettsia prowazekii*	24
YBR084W	41.6	*Rhodobacter sphaeroides*	97
YBR221C	43.1	*Rhodobacter sphaeroides – Rhodopseudomonas palustris – Rhodospirillum rubrum*	63
YBR227C	37.8	*O. anthropi*	97
YBR263W	44.7	*O. anthropi*	89
YCL017C	43.2	*Rickettsia prowazekii*	100
YDL033C	45.6	*O. anthropi*	95
YDL104C	39.1	*O. anthropi*	86
YDR036C	39.4	*Rhodobacter sphaeroides – Rhodopseudomonas palustris – Rhodospirillum rubrum*	100
YDR148C	43.0	*O. anthropi*	98
YDR194C	35.6	*O. anthropi*	90
YDR226W	43.8	*O. anthropi*	87
YDR232W	47.0	*Rickettsia prowazekii*	38
YDR258C	38.8	*O. anthropi*	88
YDR268W	39.2	*O. anthropi*	55
YDR341C	38.7	*O. anthropi*	88
YEL024W	42.1	*O. anthropi*	51
YEL030W	41.0	*O. anthropi*	88
YEL050C	43.0	*O. anthropi*	93
YEL052W	41.7	*Rhodobacter sphaeroides – Rhodopseudomonas palustris – Rhodospirillum rubrum*	81
YER017C	43.1	*O. anthropi*	88
YER061C	41.8	*Rickettsia prowazekii*	85
YER087W	38.5	*Rickettsia prowazekii*	100
YER170W	41.9	*Rhodobacter sphaeroides – Rhodopseudomonas palustris – Rhodospirillum rubrum*	54
YFL022C	43.8	*Rickettsia prowazekii*	93
YGL064C	39.4	*O. anthropi*	90
YGL143C	41.5	*Rhodobacter sphaeroides*	44
YGL236C	47.0	*Rickettsia prowazekii*	87
YGR028W	37.9	*R. palustris*	88
YGR094W	39.6	*Rickettsia prowazekii*	100
YGR220C	39.6	*Rickettsia prowazekii*	88
YGR244C	42.1	*O. anthropi*	62
YHL004W	43.1	*Rickettsia prowazekii*	100
YHL014C	37.7	*Rickettsia prowazekii*	98
YHL035C	35.8	*O. anthropi*	93
YHR011W	39.7	*Rickettsia prowazekii*	91
YHR024C	42.7	*Rickettsia prowazekii*	33
YHR106W	43.2	*O. anthropi*	75
YHR120W	37.0	*O. anthropi*	63
YHR168W	42.7	*O. anthropi*	100
YIL094C	42.2	*O. anthropi – Rickettsia prowazekii – Rhodospirillum rubrum*	98

Table 3 (*concluded*).

Saccharomyces cerevisiae ID	G+C content (%)*	Sister group†	Bootstrap value‡
YIL125W	40.6	*Rhodobacter sphaeroides – Rhodopseudomonas palustris – Rhodospirillum rubrum*	35
YJL102W	38.2	*Rickettsia prowazekii*	78
YJL200C	40.0	*O. anthropi*	100
YJR045C	43.4	*O. anthropi*	97
YJR121W	43.0	*Rhodobacter sphaeroides*	46
YKL194C	39.0	*Rhodobacter sphaeroides – Rickettsia prowazekii – Rhodospirillum rubrum*	100
YLR069C	40.2	*Rickettsia prowazekii*	66
YLR163C	39.6	*Rhodobacter sphaeroides – Rhodopseudomonas palustris – Rhodospirillum rubrum*	85
YLR188W	41.4	*Rickettsia prowazekii – O. anthropi*	74
YLR259C	43.2	*Rickettsia prowazekii*	86
YLR289W	38.6	*O. anthropi*	100
YLR304C	41.4	*O. anthropi*	100
YLR369W	37.9	*O. anthropi*	100
YLR382C	39.8	*Rickettsia prowazekii*	98
YML110C	40.6	*Rickettsia prowazekii*	53
YMR023C	38.0	*Rickettsia prowazekii*	87
YMR089C	40.0	*O. anthropi*	98
YMR207C	38.6	*O. anthropi*	82
YMR293C	38.3	*Rickettsia prowazekii*	99
YMR301C	41.3	*O. anthropi*	80
YNR036C	49.4	*Rickettsia prowazekii*	83
YOL023W	36.9	*Rickettsia prowazekii*	100
YOL033W	38.7	*Rickettsia prowazekii*	100
YOR065W	45.5	*O. anthropi*	48
YOR142W	41.8	*O. anthropi*	96
YOR187W	42.0	*O. anthropi*	95
YOR232W	38.1	*Rickettsia prowazekii*	63
YOR274W	39.1	*O. anthropi*	68
YPL040C	38.8	*Rhodobacter sphaeroides – Rhodopseudomonas palustris – Rhodospirillum rubrum*	98
YPL091W	40.2	*Rhodobacter sphaeroides – Rhodopseudomonas palustris – Rhodospirillum rubrum*	42
YPL097W	38.9	*O. anthropi*	100
YPL104W	38.9	*O. anthropi*	97
YPL132W	39.3	*Rickettsia prowazekii*	92
YPL262W	43.8	*Rickettsia prowazekii*	63
YPL270W	40.0	*O. anthropi*	49
YPR024W	42.9	*Rickettsia prowazekii*	68

*G+C content of the *S. cerevisiae* gene.

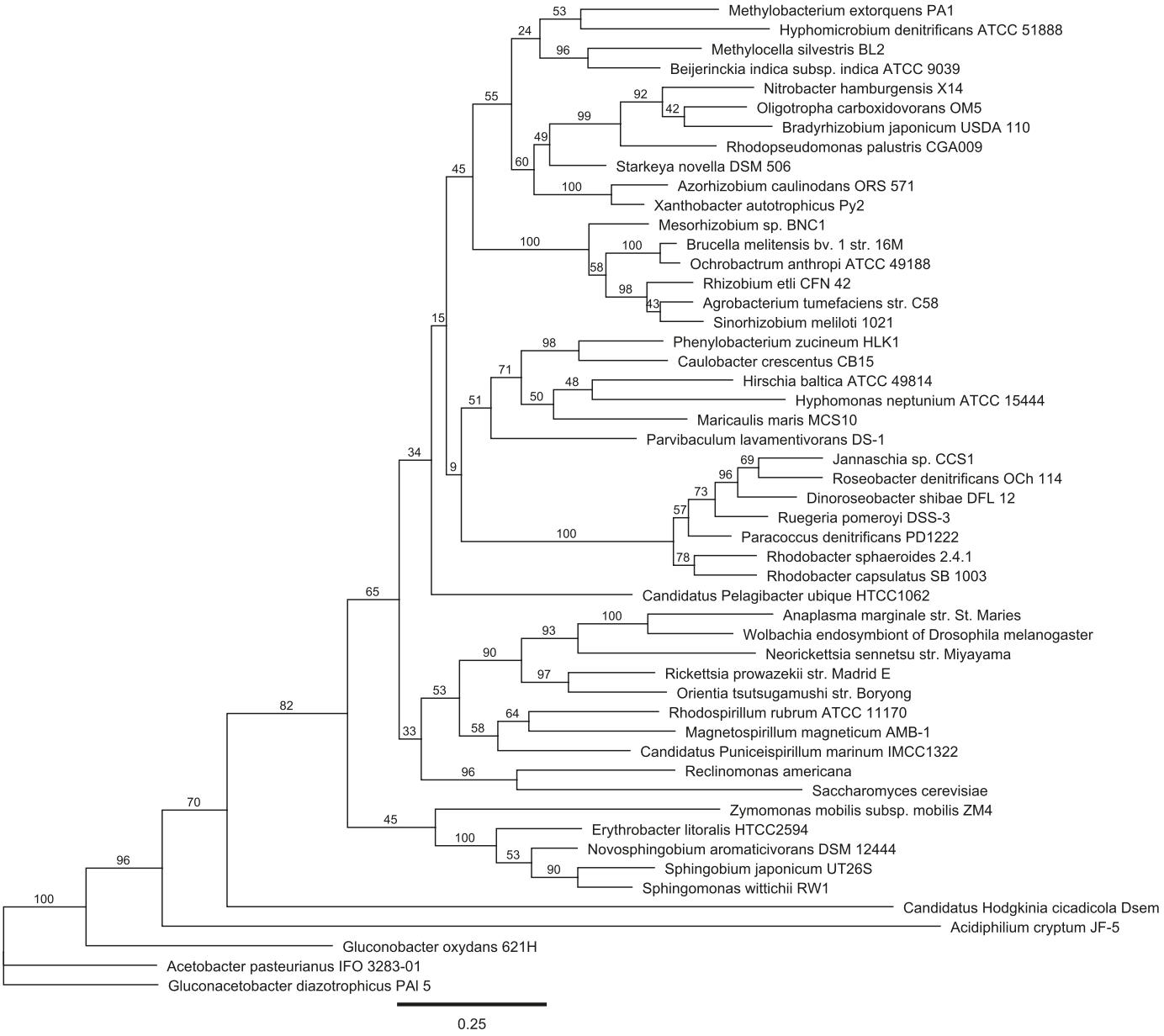
†Branched sister group that matches highest to the given *S. cerevisiae* gene.

‡Bootstrap support value for the sister group branch.

mean bootstrap value for the *O. anthropi* matches (84.60) was higher than that for the *Rickettsia prowazekii* matches (79.13) and the remaining matches (72.39).

Sample phylogenetic trees for the second set of phylogenetic analyses, which included more than 40 more alphaproteobacterial species, are depicted in Figs. 5 and 6. The BLASTP searches of the common yeast genes to the *Reclinomonas americana* mitochondrion yielded 11 significant (bitscore >100) matches. Since these 11 matches represent proteins that have been conserved for mitochondrial functions in both organisms, trees were constructed for these matches with additional *Alphaproteobacteria*. These trees were optimized for topology, length, and rate. Three of the 11 trees were unresolvable, as mitochondrial genes of the yeast and *Reclinomonas americana* branched to the entire *Alphaproteobacteria* as a sister group or the 2 mitochondrial genes branched to separate locations on a tree. Of the 8 remaining trees, 3 showed branching of mitochondrial genes to the members of the *Rickettsiales* order, from the families *Rickettsiaceae* and *Anaplasmataceae*, while 3 other trees showed branching of mitochondrial genes to a combination of not only members of the *Rickettsiaceae* and *Anaplasmataceae* families, but also members of the *Rhodospirillaceae* family. Of the remaining 2 trees, 1 showed branching of mitochondrial genes to bacteria

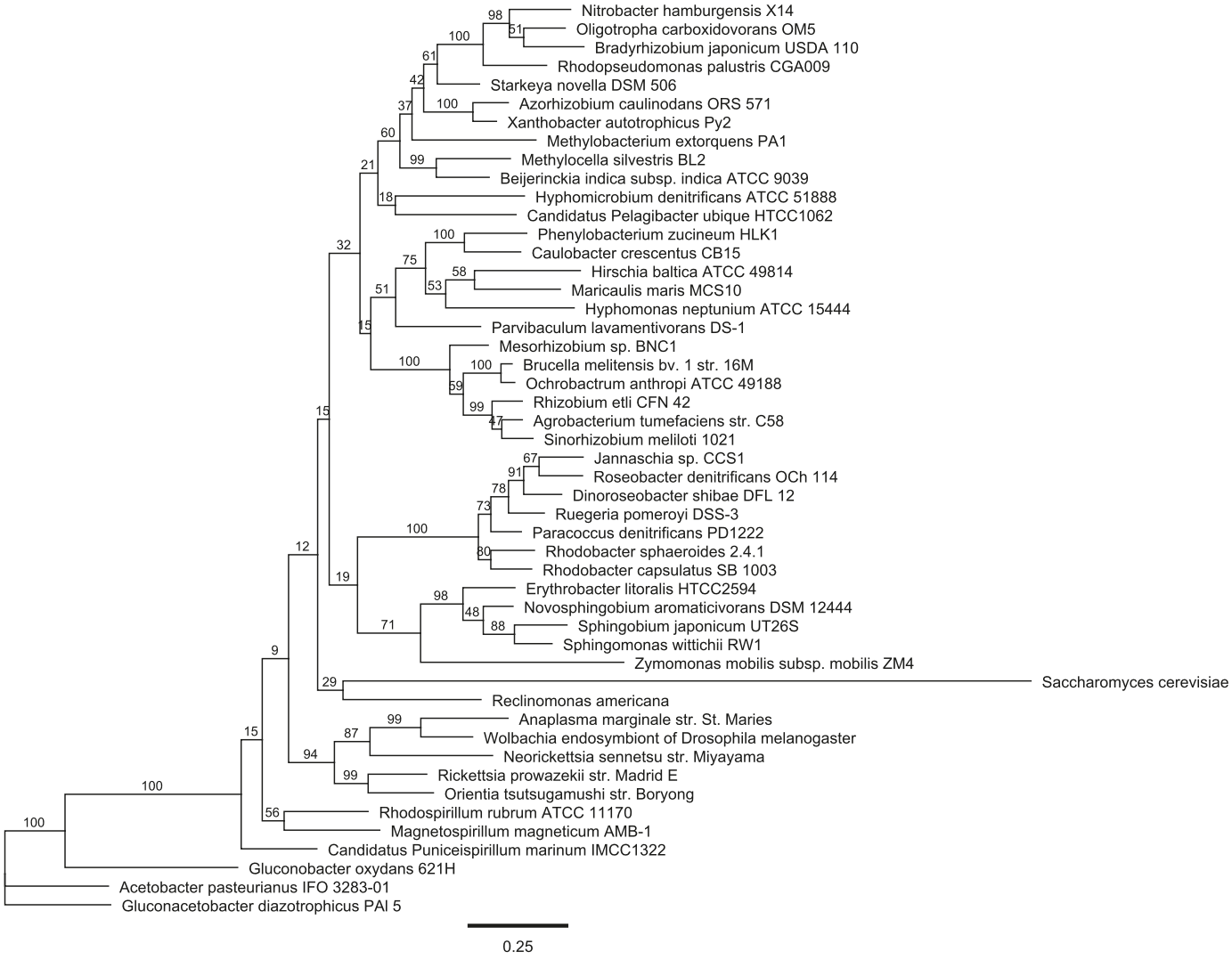
Fig. 5. The highest matching genes to Q0105 from several *Alphaproteobacteria* were aligned with the corresponding yeast gene and a matching mitochondrial gene from *Reclinomonas americana*, and a maximum likelihood tree was constructed. This tree shows the highest match of the mitochondrial genes to members of the *Rickettsiales* and *Rhodospirillales* orders. The tree was rooted to provide a better visual representation of the groupings. The numbers on the branches represent the bootstrap support for those nodes. The scale bar provides a measure of the substitutions per site for each branch.



outside of the *Rickettsiaceae*, *Anaplasmataceae*, and *Rhodospirillaceae* families, while one showed branching to bacteria outside the *Rickettsiaceae* and *Anaplasmataceae* families.

It has been estimated that *Proteobacteria* first diverged into its own clade close to 2.8 billion years ago, while the origin of the *Alphaproteobacteria* group is close to 2 billion years ago (Battistuzzi et al. 2004). Moreover, the beginning of a symbiotic relationship between bacteria and their proto-eukaryotic host cells has been estimated to be in the vicinity of 2 billion years ago (Feng et al. 1997), while the minimum age of the eukaryotic group is thought to be at least 1.45 billion years (Javaux et al. 2001). Thus, the split of *Alphaproteobacteria* from *Gammaproteobacteria* and the start of a symbiotic relationship between bacteria and eukaryotes seem to be relatively close occurrences. However, this ~2 billion year mark dating the origin of mitochondria has been challenged based on recent observations (Embley and Martin 2006). Although current-day *Rickettsia prowazekii* and *O. anthropi* are animal pathogens, the pathogenic nature of either of these 2 species could have possibly evolved much later through horizontal gene transfer of pathogenic genes. Both also differ significantly in their G+C contents, which are 29% and 56%, respectively. The other 3 species are free-living and contain significantly higher G+C contents, ~65%. Additionally, mitochondria in different species exhibit varying sizes and G+C contents. *Reclinomonas*

Fig. 6. The highest matching genes to Q0110 from several *Alphaproteobacteria* were aligned with the corresponding yeast gene and a matching mitochondrial gene from *Reclinomonas americana*, and a maximum likelihood tree was constructed. This tree shows the highest match of the mitochondrial genes to members outside the *Rickettsiales* and *Rhodospirillales* orders. The tree was rooted to provide a better visual representation of the groupings. The numbers on the branches represent the bootstrap support for those nodes. The scale bar provides a measure of the substitutions per site for each branch.



americana (69 304 nt) and *Drosophila melanogaster* (19 517 nt) both have low G+C contents of 26% and 17%, respectively, while *Arabidopsis thaliana* (366 324 nt) and *Beta vulgaris* subsp. *vulgaris* (368 801 nt) have higher G+C contents of 44% and 43%, respectively. The variability in these mitochondrial G+C contents could suggest that different selection mechanisms were operating in their incorporation or that mitochondria have originated from multiple symbiotic events with bacteria with varied G+C genome composition. Since the G+C contents of the nuclear genome and mitochondrial genome are different within each of these organisms, differences in G+C contents among different mitochondrial genomes could possibly support the latter hypothesis. However, this line of reasoning is circumstantial at best, as distantly related organisms also contain similar G+C content to that of eukaryotic mitochondria and G+C content comparison is a very simplistic method of comparing genomes, especially since selective constraints, lifestyle pressures, and environmental conditions could have determined the outcome of G+C composition in many organisms.

There is significant evidence for the transfer of genes from the bacterial genome to the nuclear genome of protoeukaryotic cells (Gray et al. 1999). These genes were then later selected for and remained necessary for mitochondrial functions. Some organisms lack mitochondria, but equivalent mitochondrial functions are coded for by their nuclear genomes (Chose et al. 2003*a*, 2003*b*). Other organisms possess mitochondria-derived organelles, such as mitosomes and hydrogenosomes, that, in the vast majority of cases, do not possess coding genomes (Embley and Martin 2006). Also, in *Cryptosporidium parvum*, there is a compartment that resembles a mitochondrion, but this compartment lacks a genome, suggesting that the organelle underwent both an extreme reduction of its mitochondrial genome and a transfer of remaining genes into the nucleus (Henriquez et al. 2005). Although the slippage of bacterial DNA into the nucleus allowed for

direct and spontaneous gene (DNA) transfer, these gene functions were sorted out through early symbiotic processes and the development of mitochondria. Recently, it has also been shown that there is direct DNA transfer to the nucleus from organelles (Timmis et al. 2004; Hazkani-Covo et al. 2010).

Even though many factors make *Rickettsia prowazekii* a probable ancestral connection to mitochondria, this does not rule out a significant influence of the ancestors of *O. anthropi* and other bacteria on the ultimate origin and development of mitochondria. This is especially true given that a greater number of the common genes among the yeast and the 5 bacteria phylogenetically matched to those of *O. anthropi*, a member of the *Rhizobiales* order, compared with those of *Rickettsia prowazekii*, a member of the *Rickettsiales* order. Moreover, of the 8 resolvable large phylogenetic trees, only 3 showed distinct branching of *Reclinomonas americana* and *S. cerevisiae* to members solely of the order *Rickettsiales*. As such, the results could suggest that mitochondria may have originated either from many independent species-specific associations with precursor eukaryotic cells or from a single species that existed before the divergence of many *Alphaproteobacteria* orders. Also, it must be noted that the first set of phylogenetic analyses contained only 5 bacteria, a number that some would consider to be a small sample. Although taxon sampling is thought to be necessary to reduce the phylogenetic error (Pollock et al. 2002; Zwickl and Hillis 2002) that may be present with small sample sizes, there is some evidence to the contrary (Rosenberg and Kumar 2001, 2003). The data sets for the tree constructed in this study also used sizeable protein, not nucleotide, sequences, thereby providing a decent amount of information that could be used for the tree construction.

Although horizontal gene transfer can muddle phylogenies and tree topologies (Bergthorsson et al. 2003), possibly skewing the interpretations that have been put forth concerning the origin of mitochondria, the reverse concept must always be considered, namely that the potent force of horizontal gene transfer in the evolution of genomes also cannot be underestimated. It is quite possible that specialized metabolic processes, such as aerobic respiration and photosynthesis, did not simply originate from a single massive gene transfer from one organism to other but rather by many smaller gene transfers among multiple organisms (Lang et al. 1999*a*). Such an understanding presents 2 scenarios. On one hand, before the mitochondrial progenitor invaded a protoeukaryotic organism, many bacterial species might have gone through multiple gene transfers to create a cluster of genes selected for mitochondrial or mitochondria-related functions in a single bacterial species. Afterwards, this bacterium that contained genes from multiple bacteria formed a symbiotic relationship with a precursor eukaryotic cell, eventually leading to the formation of mitochondria. The alternative scenario is that after a symbiotic relationship between a bacterium and a protoeukaryotic cell was initiated, many bacterial gene transfers to the precursor eukaryotic cell further strengthened mitochondrial functions and processes. However, it may be that a single ancestor of these *Alphaproteobacteria* may have possessed a varied compilation of genes, whether by horizontal gene transfer or some other means, which led to mitochondria, but these genes do not belong to the known set of modern *Alphaproteobacteria* genes, making any sort of analyses in the search for the mitochondrial progenitor much more complex (Esser et al. 2007).

In conclusion, although the origin of mitochondria seems more possible from the earlier symbiotic event of an ancestor of an *Alphaproteobacteria* species and a protoeukaryotic cell, more data are needed to validate that this was the sole event. Otherwise, it seems that a great deal of horizontal gene transfer would have had to occur, with a eukaryotic cell possessing an engulfed bacterium, to facilitate such diverse matches and eventual formation of mitochondria. On the other hand, attempting to identify such gene transfer concretely would be difficult if not impossible given the length of time elapsed since the symbiotic events that led to mitochondria. Regardless, the results clearly demonstrate the significant role of other bacteria outside the order *Rickettsiales* in the formation of mitochondria. As such, our results support either understanding, as in light of the data, a level of bacterial genome chimaerism has occurred en route to the formation of mitochondria.

Acknowledgements

We thank the Office of Research and Special Programs at Sam Houston State University for funding this work through the award of an Enhancement Grant for Research (EGR) to Madhusudan Choudhary. We also thank the 3 anonymous reviewers for their helpful comments and suggestions.

References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**(17): 3389–3402. doi:10.1093/nar/25.17.3389. PMID:9254694.

Andersson, S.G., Zomorodipour, A., Andersson, J.O., Sicheritz-Pontén, T., Alsmark, U.C., Podowski, R.M., et al. 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature (Lond.), **396**(6707): 133–140. doi:10.1038/24094. PMID:9823893.

Andersson, S.G., Karlberg, O., Canbäck, B., and Kurland, C.G. 2003. On the origin of mitochondria: a genomics perspective. Philos. Trans. R. Soc. Lond. B Biol. Sci. **358**(1429): 165–179. doi:10.1098/rstb.2002.1193. PMID:12594925.

Atteia, A., Adrait, A., Brugière, S., Tardif, M., van Lis, R., Deusch, O., et al. 2009. A proteomic survey of *Chlamydomonas reinhardtii* mitochondria sheds new light on the metabolic plasticity of the organelle and on the nature of the α-proteobacterial mitochondrial ancestor. Mol. Biol. Evol. **26**(7): 1533–1548. doi:10.1093/molbev/msp068. PMID:19349646.

Battistuzzi, F.U., Feijao, A., and Hedges, S.B. 2004. A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. BMC Evol. Biol. **4**(1): 44. doi:10.1186/1471-2148-4-44. PMID:15535883.

Bergthorsson, U., Adams, K.L., Thomason, B., and Palmer, J.D. 2003. Widespread horizontal transfer of mitochondrial genes in flowering plants. Nature (Lond.), **424**(6945): 197–201. doi:10.1038/nature01743. PMID:12853958.

Chose, O., Sarde, C.O., Gerbod, D., Viscogliosi, E., and Roseto, A. 2003*a*. Programmed cell death in parasitic protozoans that lack mitochondria. Trends Parasitol. **19**(12): 559–564. doi:10.1016/j.pt.2003.09.016. PMID:14642765.

Chose, O., Sarde, C.O., Noël, C., Gerbod, D., Jimenez, J.C., Brenner, C., et al. 2003*b*. Cell death in protists without mitochondria. Ann. N. Y. Acad. Sci. **1010**(1): 121–125. doi:10.1196/annals.1299.021. PMID:15033707.

Drummond, A., Ashton, B., Buxton, S., Cheung, M., Heled, J., Kearse, M., et al. 2010. Geneious, version 4.8. Available from http://www.geneious.com [accessed 17 June 2009].

Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32**(5): 1792–1797. doi:10.1093/nar/gkh340. PMID:15034147.

Embley, T.M., and Martin, W. 2006. Eukaryotic evolution, changes and challenges. Nature (Lond.), **440**(7084): 623–630. doi:10.1038/nature04546. PMID:16572163.

Emelyanov, V.V. 2001. Rickettsiaceae, *Rickettsia*-like endosymbionts, and the origin of mitochondria. Biosci. Rep. **21**(1): 1–17. doi:10.1023/A:1010409415723. PMID:11508688.

Esser, C., Ahmadinejad, N., Wiegand, C., Rotte, C., Sebastiani, F., Gelius-Dietrich, G., et al. 2004. A genome phylogeny for mitochondria among α-proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. Mol. Biol. Evol. **21**(9): 1643–1660. doi:10.1093/molbev/msh160. PMID:15155797.

Esser, C., Martin, W., and Dagan, T. 2007. The origin of mitochondria in light of a fluid prokaryotic chromosome model. Biol. Lett. **3**(2): 180–184. doi:10.1098/rsbl.2006.0582. PMID:17251118.

Feng, D.F., Cho, G., and Doolittle, R.F. 1997. Determining divergence times with a protein clock: update and reevaluation. Proc. Natl. Acad. Sci. U.S.A. **94**(24): 13028–13033. doi:10.1073/pnas.94.24.13028. PMID:9371794.

Fitzpatrick, D.A., Creevey, C.J., and McInerney, J.O. 2006. Genome phylogenies indicate a meaningful α-proteobacterial phylogeny and support a grouping of the mitochondria with the Rickettsiales. Mol. Biol. Evol. **23**(1): 74–85. doi:10.1093/molbev/msj009. PMID:16151187.

Giovannoni, S.J., Tripp, H.J., Givan, S., Podar, M., Vergin, K.L., Baptista, D., et al. 2005. Genome streamlining in a cosmopolitan oceanic bacterium. Science, **309**(5738): 1242–1245. doi:10.1126/science.1114057. PMID:16109880.

Gray, M.W., Burger, G., and Lang, B.F. 1999. Mitochondrial evolution. Science, **283**(5407): 1476–1481. doi:10.1126/science.283.5407.1476. PMID:10066161.

Guindon, S., and Gascuel, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. **52**(5): 696–704. doi:10.1080/10635150390235520. PMID:14530136.

Hazkani-Covo, E., Zeller, R.M., Martin, W., and Malik, H.S. 2010. Molecular poltergeists: mitochondrial DNA copies (*numts*) in sequenced nuclear genomes. PLoS Genet. **6**(2): e1000834. doi:10.1371/journal.pgen.1000834. PMID:20168995.

Henriquez, F.L., Richards, T.A., Roberts, F., McLeod, R., and Roberts, C.W. 2005. The unusual mitochondrial compartment of *Cryptosporidium parvum*. Trends Parasitol. **21**(2): 68–74. doi:10.1016/j.pt.2004.11.010. PMID:15664529.

Imhoff, J.F., and Trüper, H.G. 1992. The genus *Rhodospirillum* and related genera. *In* The prokaryotes. *Edited by* A. Balows, H.G. Trüper, M. Dworkin, W. Harder, and K.-H. Schleifer. Springer-Verlag, New York. pp. 2141–2159.

Javaux, E.J., Knoll, A.H., and Walter, M.R. 2001. Morphological and ecological complexity in early eukaryotic ecosystems. Nature (Lond.), **412**(6842): 66–69. doi:10.1038/35083562. PMID:11452306.

Karlberg, O., Canbäck, B., Kurland, C.G., and Andersson, S.G. 2000. The dual origin of the yeast mitochondrial proteome. Yeast, **17**(3): 170–187. doi:10.1002/1097-0061(20000930)17:3<170::AID-YEA25>3.0.CO;2-V. PMID:11025528.

Kersters, K., Devos, P., Gillis, M., Swings, J., Vandamme, P., and Stackebrandt, E. 2006. Introduction to the proteobacteria. *In* The prokaryotes: a handbook on the biology of bacteria. *Edited by* M. Dworkin, S. Falkow, E. Rosenberg, K.H. Schleifer, and E. Stackebrandt. Springer-Verlag, New York. pp. 3–37.

Lang, B.F., Burger, G., O'Kelly, C.J., Cedergren, R., Golding, G.B., Lemieux, C., et al. 1997. An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. Nature (Lond.), **387**(6632): 493–497. doi:10.1038/387493a0. PMID:9168110.

Lang, B.F., Gray, M.W., and Burger, G. 1999*a*. Mitochondrial genome evolution and the origin of eukaryotes. Annu. Rev. Genet. **33**(1): 351–397. doi:10.1146/annurev.genet.33.1.351. PMID:10690412.

Lang, B.F., Seif, E., Gray, M.W., O'Kelly, C.J., and Burger, G. 1999*b*. A comparative genomics approach to the evolution of eukaryotes and their mitochondria. J. Eukaryot. Microbiol. **46**(4): 320–326. doi:10.1111/j.1550-7408.1999.tb04611.x. PMID:10461380.

Plasterer, T.N., Smith, T.F., and Mohr, S.C. 2001. Survey of human mitochondrial diseases using new genomic/proteomic tools. Genome Biol. **2**(6): research0021.1–research0021.16. doi:10.1186/gb-2001-2-6-research0021. PMID:11423010.

Pollock, D.D., Zwickl, D.J., McGuire, J.A., and Hillis, D.M. 2002. Increased taxon sampling is advantageous for phylogenetic inference. Syst. Biol. **51**(4): 664–671. doi:10.1080/10635150290102357. PMID:12228008.

Rosenberg, M.S., and Kumar, S. 2001. Incomplete taxon sampling is not a problem for phylogenetic inference. Proc. Natl. Acad. Sci. U.S.A. **98**(19): 10751–10756. doi:10.1073/pnas.191248498. PMID:11526218.

Rosenberg, M.S., and Kumar, S. 2003. Taxon sampling, bioinformatics, and phylogenomics. Syst. Biol. **52**(1): 119–124. doi:10.1080/10635150390132894. PMID:12554445.

Timmis, J.N., Ayliffe, M.A., Huang, C.Y., and Martin, W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat. Rev. Genet. **5**(2): 123–135. doi:10.1038/nrg1271. PMID:14735123.

Whelan, S., and Goldman, N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol. Biol. Evol. **18**(5): 691–699. PMID:11319253.

Williams, K.P., Sobral, B.W., and Dickerman, A.W. 2007. A robust species tree for the *Alphaproteobacteria*. J. Bacteriol. **189**(13): 4578–4586. doi:10.1128/JB.00269-07. PMID:17483224.

Wu, M., Sun, L.V., Vamathevan, J., Riegler, M., Deboy, R., Brownlie, J.C., et al. 2004. Phylogenomics of the reproductive parasite *Wolbachia pipientis w*Mel: a streamlined genome overrun by mobile genetic elements. PLoS Biol. **2**(3): e69. doi:10.1371/journal.pbio.0020069. PMID:15024419.

Zwickl, D.J., and Hillis, D.M. 2002. Increased taxon sampling greatly reduces phylogenetic error. Syst. Biol. **51**(4): 588–598. doi:10.1080/10635150290102339. PMID:12228001.