

Pitfalls of jackhmmer and psiblast

- Danger: the inclusion of one error means your iterations spin out of control.
- most common: coiled coil, or low complexity
- Jackhmmer seems more prone to this than psi-blast

A very specific pitfall of jackhmmer

- Many insignificant domain hits make a significant protein hit; if your query has a vague and insignificant hit to a protein with almost identical repeats

	(Yeast) (<i>Monilia parapsilosis</i>)		
Membrane-anchored cell surface protein	<i>Burkholderia pseudomallei</i> (strain 1106a)		2.6e-07
Membrane-anchored cell surface protein	<i>Burkholderia pseudomallei</i> MSH-R346		4.1e-07
Putative membrane-anchored cell surface protein	<i>Burkholderia pseudomallei</i> (strain K96243)		2.2e-06
Membrane-anchored cell surface protein	<i>Burkholderia pseudomallei</i> 1106b		8.3e-06

contents

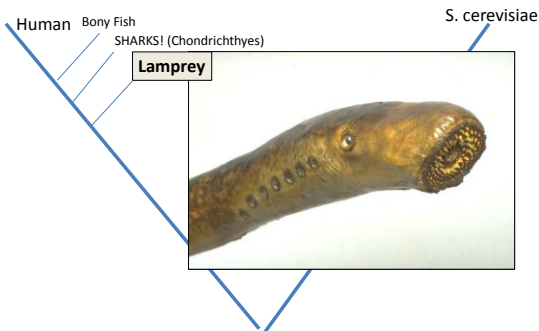
- Counting back from human (and *S. cerevisiae*) “crucial” / “early branching” genomes
- Eukaryotic supergroups (quick)
- Three kingdoms or two?, nature of the first eukaryote
- From FECA to LECA: duplications
- From FECA to LECA: endosymbiosis

(eukaryotic) Tree of Life, eukaryogenesis, LECA

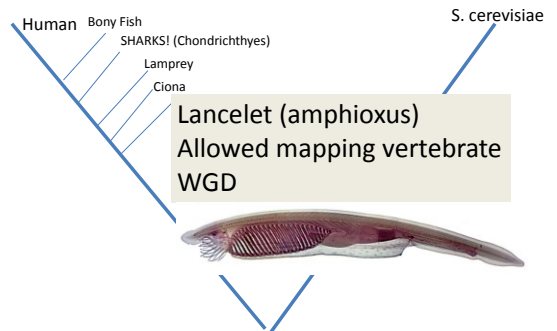
3/6/2014

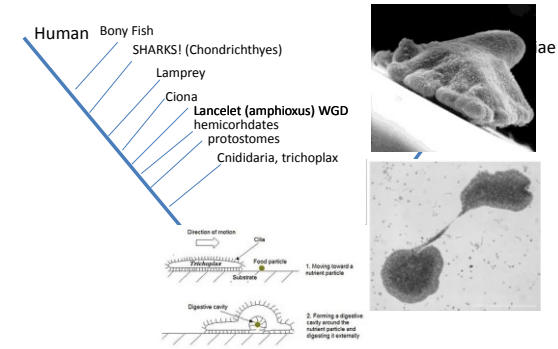
3

Crucial genomes fill gaps

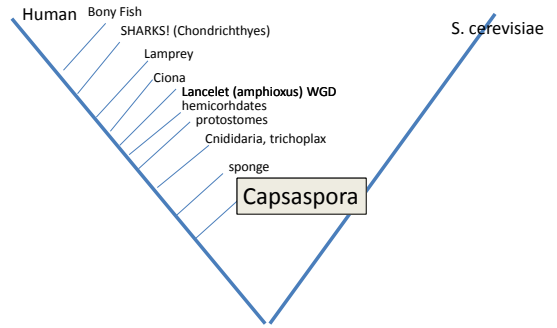


Crucial genomes fill gaps





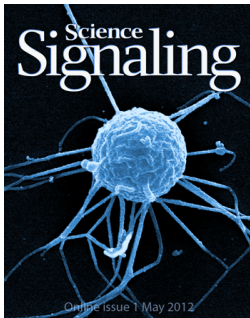
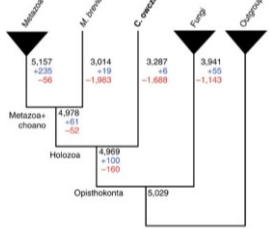
<http://www.trichoplax.com/mediafiles/115.jpg>
http://genome.jgi-psf.org/Triaad1/Trichoplax2FM_Saigster.jpg
http://upload.wikimedia.org/wikipedia/en/thumb/1/17/Exodigestion_in_Trichoplax_adhaerens.jpg/380px-Exodigestion_in_Trichoplax_adhaerens.jpg



ARTICLE
 Received 18 Jan 2012; Accepted 18 Jul 2012; Published 10 Aug 2012

The *Capsaspora* genome reveals a complex unicellular prehistory of animals

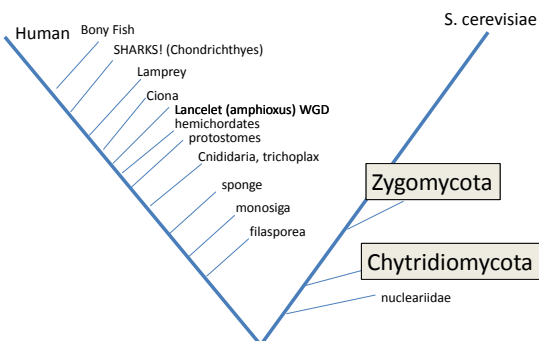
Frederick Kappe¹, Stefan Cheloufi², Anne de Heidegger³, Armin Sauer-Ruedel⁴, Matthias W. Bräse⁵, Eric Klomparens⁶, Martin Cui⁷, Pierre Krenkel⁸, Mehdi Vaziri-Nia⁹, Nora Sautter-Pohl¹⁰, Gökçe Tarcanli¹¹, Roman Denzlik¹², Gerard Manning¹³, E. Franz Lang¹⁴, Carsten Rüssel¹⁵, Brian J. Haas¹⁶, Andrew J. Roger¹⁷, Chad Nusbaum¹⁸ & Peter Raas-Johnsen¹⁹



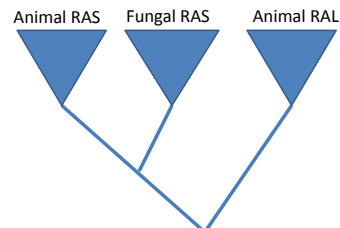
OPISTHOKONTA

	OTHER EUKARYOTES	AMOEBOZOA	FUNGI	<i>Capsaspora</i> <i>opercularis</i>	Morphogon <i>brevecolis</i>	Amphimedon <i>quercusiformis</i>	Trichoplax <i>adhaerens</i>	Neurospora <i>vegetalis</i>	Bla. ATGPA
p53 (PF00870)			1	2	1	2	3	0-12	
Runt (PF00853)			2	0	1	1	1	1-4	
RHD (PF00554)			1	0	1	1	1	1-2	

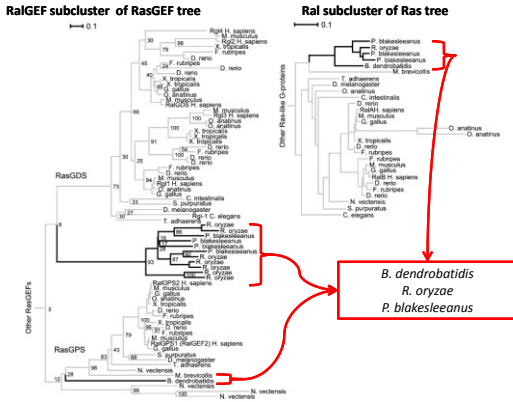
“acquisition of metazoan-specific developmental systems and the co-option of pre-existing genes drove the transition from unicellular protists to metazoans.” e.g. NFKappaB



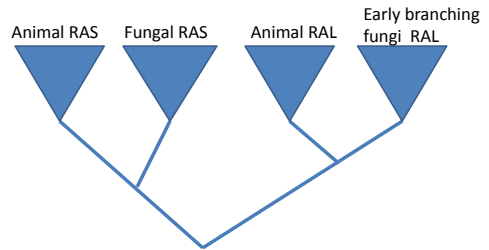
RAL evolution?



Animal invention and wrong tree (“consensus” in the RAS field)
 OR old duplication and loss



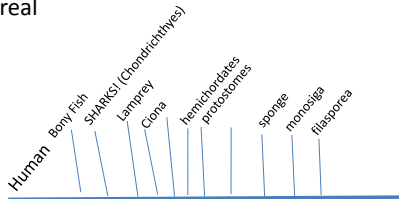
RAL evolution?



Old duplication and loss. No more OR.

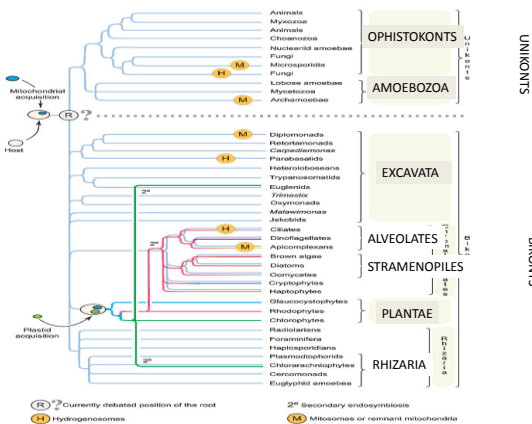
Is the asymmetry (comb) real?

- Part is perspective (protostomes!)
- Part is sampling
- Part is real



many genomes, many more underway

- Diversity at many levels
- Allow / needed for different questions
- Reveals more old diversity re: duplicates or OGs
- Fun biology (not directly applicable but helps to remember the names and relationships of the weird beasts) (a good taxonomy button like in jackhammer also helps)



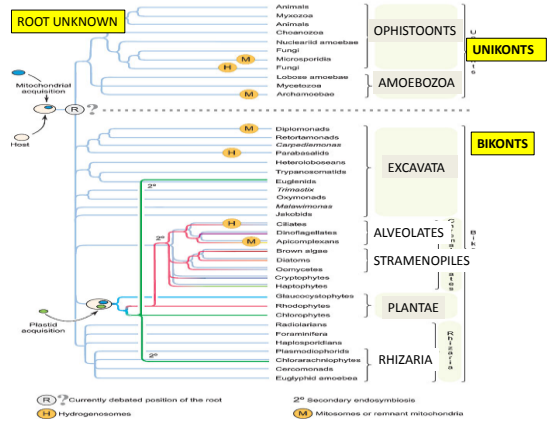
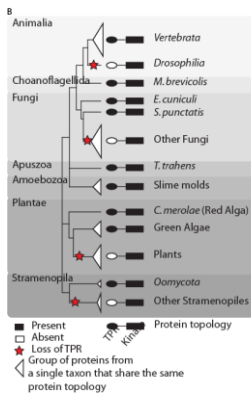
~5 Supergroups

- Current sampling hugely biased >> 500 opisthokonts, 1 rhizaria, 5(?) excavates
- Phylogenetic/ cellular/ protein diversity staggering as compared to e.g. human-fruitfly
- Especially relevant for “evolutionary cell biology”
- Mini project: one of each (super)group, fungi, animals, plantae, alveolates, amoebzoa, stramenopiles

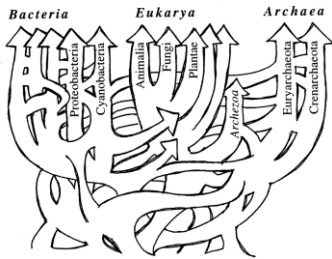
MPS1 parallel loss of TPR domain

Early branching / key genomes in supergroups gives beautiful stories

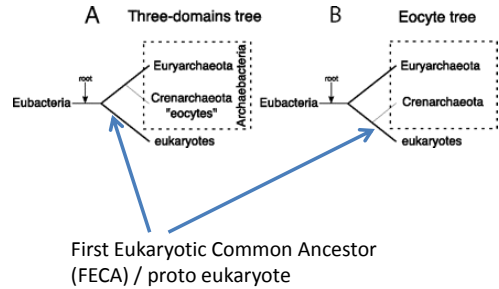
Tromer / kops in press



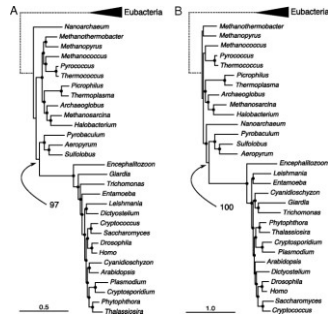
ToL & 3 kingdoms



Eocyte hypothesis

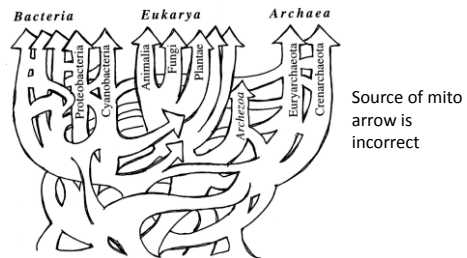


Latest trees suggest eocyte

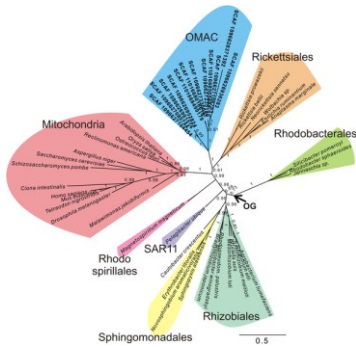


Proc Natl Acad Sci U S A. 2008 Dec 23;105(51):20156-61. Epub 2008 Dec 10. The archaeobacterial origin of eukaryotes. Cox CJ, Foster PG, Hirt BP, Harris SB, Embley TM.

Timing of alpha-proteobacterial endosymbiosis relative to bacterial radiations?

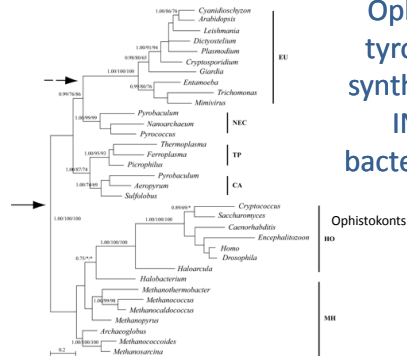


Timing of alpha-proteobacterial endosymbiosis relative to bacterial radiations?



First came: Emergence of proteobacteria, emergence of alpha proteobacteria within the proteobacteria and many speciations within the alpha-proteobacteria

I think this fits better with eocyte than with 3 kingdoms



Ophisthokont tyrosyl-tRNA synthetase falls INSIDE a bacterial genus!

Mol Biol Evol. 2005 Nov;22(11):2142-6.
The presence of a haloarchaeal type tyrosyl-tRNA synthetase marks the ophisthokonts as monophyletic.
[PubMed | Xu Y Gokarten JP](http://pubmed.ncbi.nlm.nih.gov/16444441/)

Late origin of eukaryotes?

- Proto eukaryote not ancient as bacterial and archael radiations, endosymbiosis was after many bacterial radiations, leca was thus also after these radiations, supported by ophisthokonts (supposedly "short" after LECA) being within archael genus
- Proto-eukaryote still seems to have had a long way to go to a eukaryote; "feca-2-leca"
- Nature of proto-eukaryote?

Novel archaea has operon with UBQ system

- [Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group](#), Nunoura T, Takaki Y, Kakuta J, Nishi S, Sugahara J, Kazama H, Chee GJ, Hattori M, Kanai A, Atomi H, Takai K, Takami H. Nucleic Acids Res. 2011 Apr;39(8):3204-23

The gene cluster of the Ub-like protein modifier system in *C. subterraneum*, eukaryotic "type" ubiquitin



Nunoura T et al. Nucl. Acids Res. 2011;39:3204-3223

© The Author(s) 2010. Published by Oxford University Press.

Nucleic Acids Research

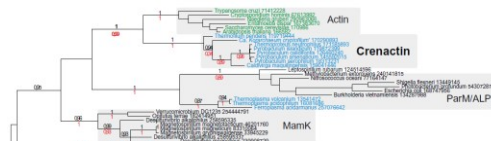
Molecular Microbiology (2011) 80(4), 1052–1061

doi:10.1111/j.1365-2958.2011.07635.x
First published online 6 April 2011

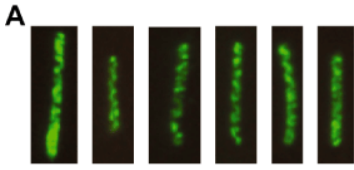
An actin-based cytoskeleton in archaea

Thijs J. G. Ettema,^{1*} Ann-Christin Lindås¹ and Rolf Bernander¹
¹Department of Molecular Evolution, Evolutionary Biology Center, Uppsala University, Norbyvägen 18C, SE-752 36, Uppsala, Sweden.

ments twisted around one another to form a right-handed double helix, which constitute well-characterized central components of cytoskeleton-dependent processes (Pollard and Cooper, 2009; van den Erii et al., 2001). As a result of the involvement in pivotal processes, the primary structure of actin has been extremely well conserved



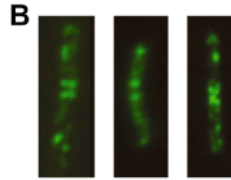
1:1 Orthologous to eukaryotic actin with limited phylogenetic distribution in archaea



In eukaryotic and bacterial cells, spatial organization is dependent upon cytoskeletal filaments

Actin is a main eukaryotic cytoskeletal element, cell shape determination, mechanical force generation and cytokinesis

Archaeal cytoskeleton of crenactin which forms helical structures within *Pyrobaculum calidifontis* cells, as shown by *in situ* immunostaining



Examples of subpopulation of cells displaying centrally located band-like structures.

Cytokinesis?

Eukaryotic features in archaea are present in subclade of archaea where also now the ToL places the eukaryotes

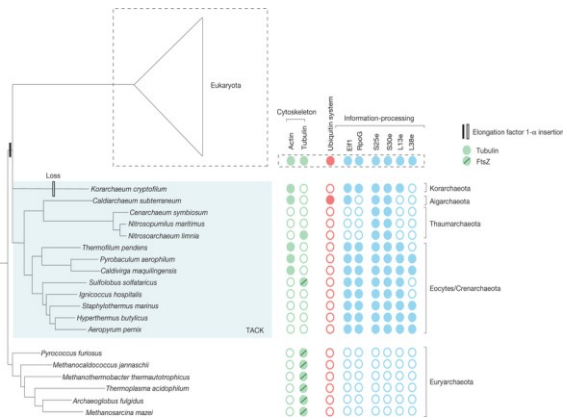
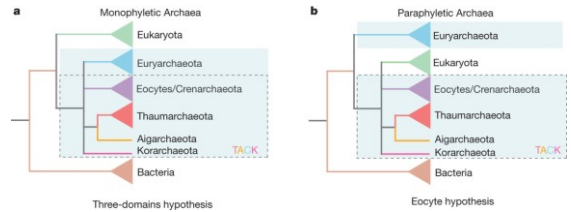
Proto-eukaryote is getting more complex as more archaeal diversity is sequenced and bioinformatically and biochemically characterized

REVIEW

doi:10.1038/nature12779

An archaeal origin of eukaryotes supports only two primary domains of life

Tom A. Williams¹, Peter G. Foster², Cymon J. Cox³ & T. Martin Embley¹



Eukaryogenesis: FECA to LECA

- Endosymbiosis
- Duplication

Similarity in membrane "topology" between a mitochondria and a eukaryotic cell that eats a bacterium: the double membrane topology



"Theory of endosymbiosis"

<http://en.wikipedia.org/wiki/Phagocytosis>

DNA

Mitochondria have their own chromosome ... but this chromosome is circular and not enveloped in a "nucleus"



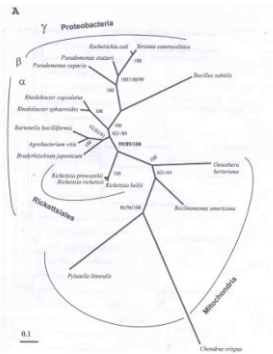
prokaryotes	eukaryotes
Circular chromosomes, no organelles	Linear chromosomes organelles

"Theory of endosymbiosis"

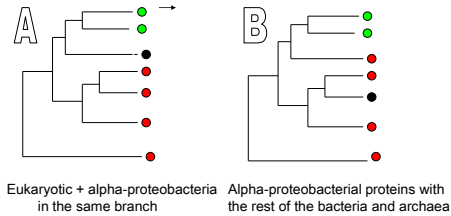
<http://home.nc.rr.com/ambient/site/mtdna.htm>

Phylogenetic trees

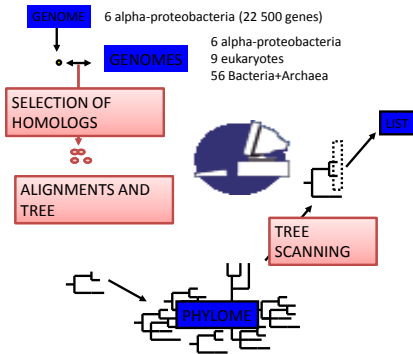
- Mitochondrial chromosome, genes, rRNA
- Similarity according to an established model of sequence change. Determine how organisms / genes are related: tree
- Tree: eukaryotic mitochondria cluster within bacteria, within alpha -proteobacteria, next to rickettsia, obligate intracellular parasites of eukaryotic cells



Identifying eukaryotic proteins with an alpha-proteobacterial origin based on their phylogeny



Detecting eukaryotic genes of alpha-proteobacterial ancestry



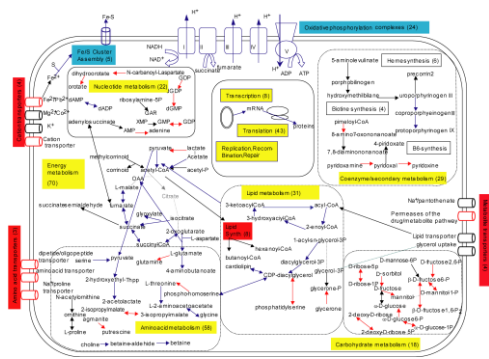
Benchmarking?

Species	Proteins*	NJ-Set ^b	ML-Set ^b
<i>Reclinomonas americana</i> (mitochondrion)	67	71.6	62.70
<i>Deinococcus radiodurans</i>	3,084	1.1	0.03
<i>Agrobacterium tumefaciens</i> (Cereon)	5,392	13.03	7.59
<i>Agrobacterium tumefaciens</i> (Washington)	5,298	13.11	11.76
<i>Bradyrhizobium japonicum</i>	8,257	11.18	8.25
<i>Brucella melitensis</i>	3,186	16.1	11.08
<i>Brucella suis</i>	3,247	15.86	9.67
<i>Caulobacter crescentus</i>	3,718	13.23	8.85
<i>Magnetococcus magnetotacticum</i>	4,280	11.36	8.74
<i>Rhizobium loti</i>	7,259	13.08	8.94
<i>Rhizobium meliloti</i>	6,149	13.67	9.17
<i>Rickettsia conorii</i>	1,374	20.3	16.59
<i>Rickettsia prowazekii</i>	834	25.06	19.78
Total selected ^c		1,026	842

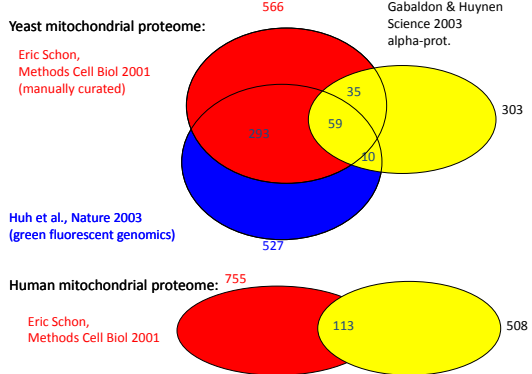
1. "a control"
2. ML works!

*Number of protein-coding genes per genome.
^bPercentage of selected proteins in each genome by each approach.
^cTotal number of selected OGs.
 An analysis that was based on six alpha-proteobacterial genomes [9] detected 630 proteins of alpha-proteobacterial origin that were in eukaryotes, and retrieved 49% of the *R. americana* mitochondrial genes and 1.3% of *D. radiodurans*. Increasing the number of analyzed genomes has thus substantially increased the number of proteins and the "completeness" of the proteome, while reducing the number of potential false positives. doi:10.1371/journal.pcbi.0030219.t001

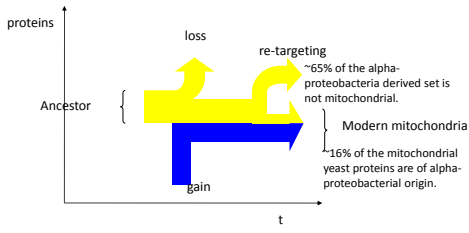
Reconstruction of the Proto-mitochondrial Cell



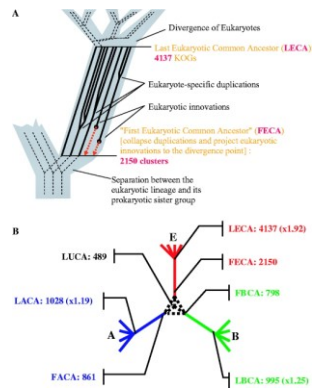
The majority of the proto-mitochondrial proteome is not mitochondrial (anymore)



From endosymbiont to organelle, not only loss and gain of proteins but also "re-targeting":



Gabalton and Huynen, Science 2004



FECA/eocyte to LECA

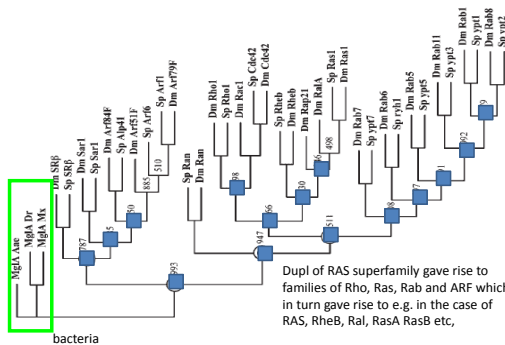
Duplications
 Inventions

Nucleic Acids Res. 2005 Aug; 33(31):41-46:36-38
 Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell.
 Makarova KS, Wolf YI, Mekhietov SA, Mirkin BG, Koonin EV

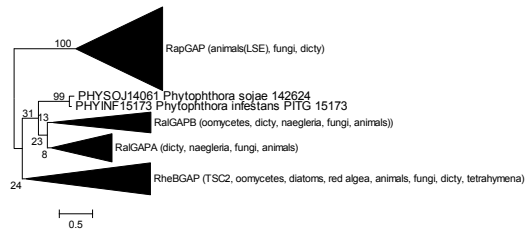
duplications: e.g. small GTPases

Small GTPases and the evolution of the eukaryotic cell

Gaspár Jékely^{1,2}



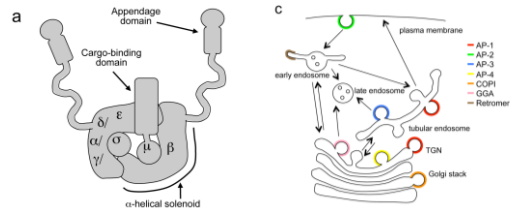
Not just the gtpases, also their activating proteins: Rap/Ral/RheB GAP tree: events from before the LECA



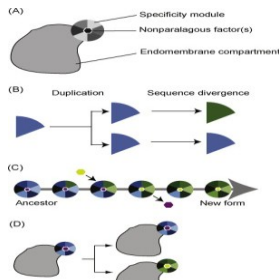
- Adaptor protein (AP) complexes sort cargo into vesicles for transport from one membrane compartment of the cell to another
- [Evolution of specificity in the eukaryotic endomembrane system](#). Dacks JB, Peden AA, Field MC. *Int J Biochem Cell Biol*. 2009 Feb;41(2):330-40.

Adaptor proteins arose via feca-2-leca duplications

Complex	Large chain 1	Large chain 2	Medium chain	Small chain
AP-1	γ	β 1	μ 1	σ 1
AP-2	α	β 2	μ 2	σ 2
AP-3	δ	β 3	μ 3	σ 3
AP-4	ϵ	β 4	μ 4	σ 4
F-COP	γ -COP	β -COP	μ -COP	σ -COP



Neo or sub functionalization ... for membrane identity?



Parallels discussion for protein Complexes, e.g. zipper model

Fig. 3 Organelle evolution driven by gene duplication of the identity-encoding machinery. (A) An initial endomembrane compartment is shown, with an as-yet undifferentiated set of identity-encoding machinery shown. The segmented circle indicate...

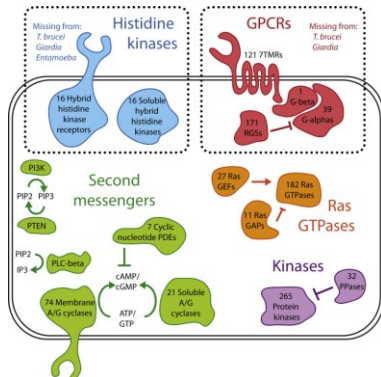
LECA “genome of eden”, “The Incredible Expanding Ancestor of Eukaryotes” (Eugene V. Koonin)

~4000 genes

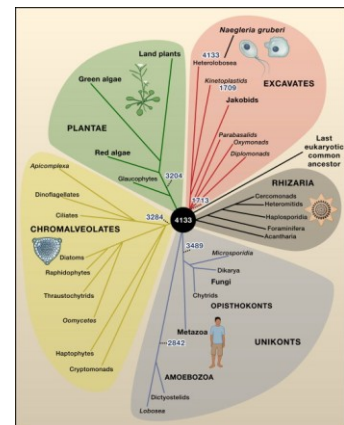
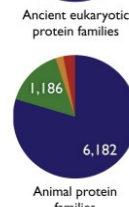
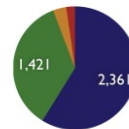
[The genome of Naegleria gruberi illuminates early eukaryotic versatility.](#)

Fritz-Laylin LK, Prochnik SE, Ginger ML, Dacks JB, Carpenter ML, Field MC, Kuo A, Paredez A, Chapman J, Pham J, Shu S, Neupane R, Cipriano M, Mancuso J, Tu H, Salamov A, Lindquist E, Shapiro H, Lucas S, Grigoriev IV, Cande WZ, Fulton C, Rokhsar DS, Dawson SC. *Cell*. 2010 Mar 5;140(5):631-42.

Signalling complexity



A

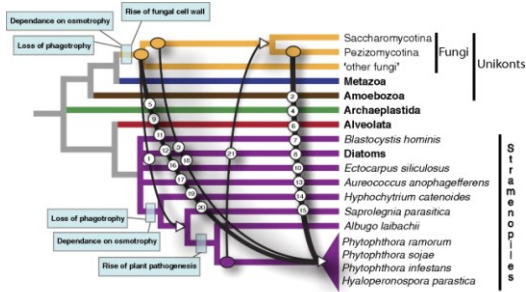
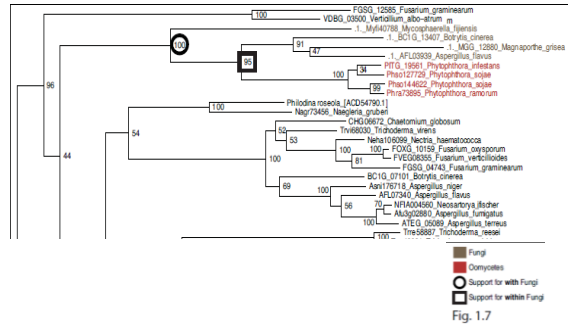


Euk ToL / Orthology complications

- HGT between eukaryotes
- Parallel HGT from bacteria
- Serial / secondary endosymbiosis
- (tertiary endosymbiosis)

HGT between eukaryotes

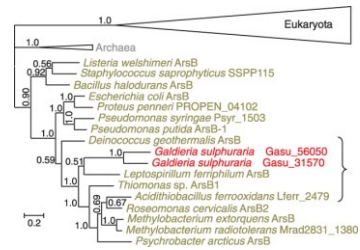
Proc Natl Acad Sci U S A. 2011 Sep 13;108(37):15258-63. Horizontal gene transfer facilitated the evolution of plant parasitic mechanisms in the oomycetes. Richards TA, Soanes DM, Jones MD, Vasieva O, Leonard G, Paszkiewicz K, Foster PG, Hall N, Talbot NJ.



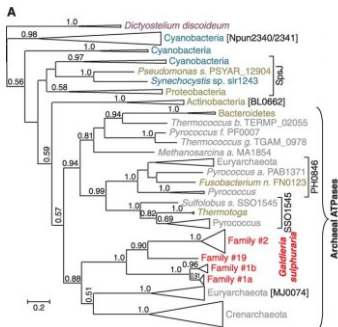
HGT from bacteria



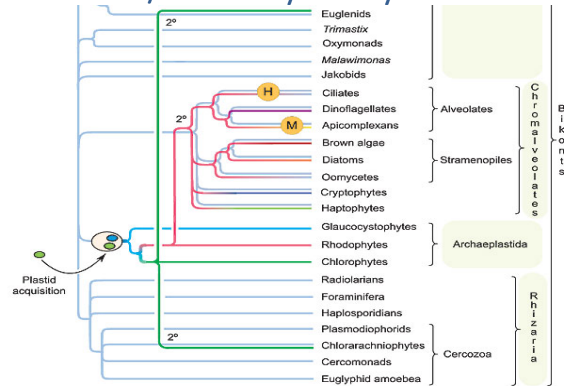
Gene Transfer from Bacteria and Archaea Facilitated Evolution of an Extremophilic Eukaryote
 Gerald Schönknecht et al.
 Science 339, 1207 (2013)
 DOI: 10.1126/science.1231707

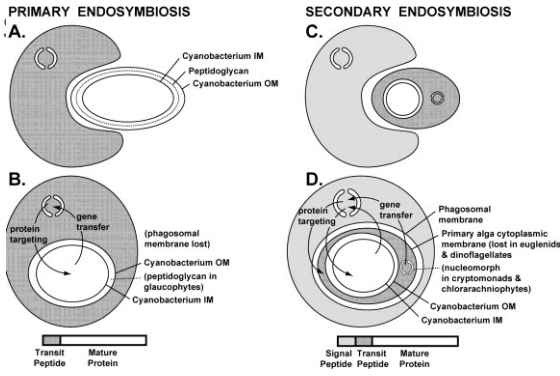


Parallel HGT from bacteria



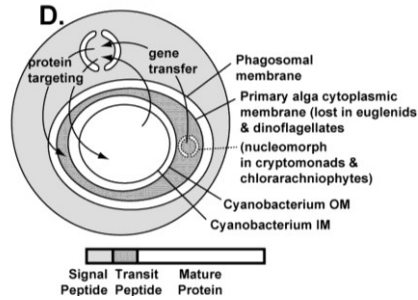
Serial / secondary endosymbiosis





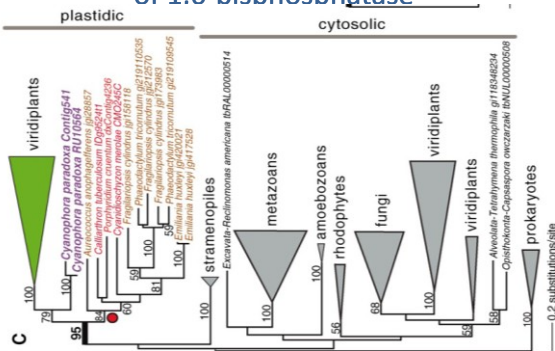
<http://www.amjbot.org/content/91/10/1481/F2.large.jpg>

Serial / secondary endosymbiosis (EGT, gene transfer, protein re-targeting)



<http://www.amjbot.org/content/91/10/1481/F2.large.jpg>

Serial / secondary endosymbiosis e.g. Tree of 1.6-bisphosphatase



A good KOG database would???

- How should it relate to COG's;
 - to endosymbiosis origin vs archaeal origin
 - Fusions and duplications (Big bang) during feca to leca
- (How) should it deal with serial endosymbiosis
- (How) should it deal with HGT between euks
- (How) should it deal with parallel HGT from bacs to euks