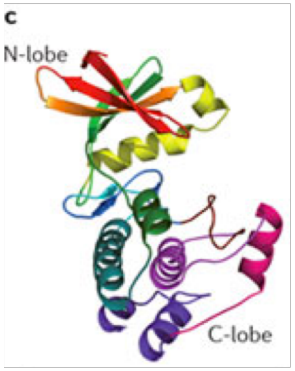
# Key Genomes, (eukaryotic) Tree of Life, implications for expectation of patterns that can appear in gene phylogenies (and OGs)



## contents

- Key Genomes: Counting back from human (and *S. cerevisiae*) “crucial” / “early branching” genomes
- Eukaryotic supergroups & Root
- After LECA: gene loss, secondary endosymbiosis

## What is the evolutionary history of this protein? What happened in its evolution? Which other organisms have **“it”**? And when did it arise in evolution?




In order to answer these question: do sensitive homology searches, making and interpreting trees. But also: the right genomes & what can you expect.

?

## How we revealed the trend of a complex ancestor and independent loss

A combination of:

- **New genomes at crucial positions**
- Improved sensitivity of sequence similarity searches (and homologs that are orthologs)
- Studying gene families with a lot of pre-LECA duplications

# Improved sensitivity of sequence similarity searches. Profile-based searches reveal ancient origins of CKK

**The CKK Domain (DUF1781) Binds Microtubules and Defines the CAMSAP/ *ssp4* Family of Animal Proteins**

*Anthony J. Baines,*† Paola A. Bignone,*[1] Mikayala D.A. King,* Alison M. Maggs,‡ Pauline M. Bennett,‡ Jennifer C. Pinder,‡[2] and Gareth W. Phillips*‡[3]*

*Department of Biosciences, University of Kent, Canterbury, Kent, United Kingdom; †Centre for Biomedical Informatics, University of Kent, Canterbury, Kent, United Kingdom; and ‡Randall Division of Cell and Molecular Biophysics, King's College London, New Hunt's House, London, United Kingdom

We describe a structural domain common to proteins related to human calmodulin-regulated spectrin-associated protein1 (CAMSAP1). Analysis of the sequence of CAMSAP1 identified a domain near the C-terminus common to CAMSAP1 and two other mammalian proteins KIAA1078 and KIAA1543, which we term a CKK domain. This domain was also present in invertebrate CAMSAP1 homologues and was found in all available eumetazoan genomes (including cnidaria), but not in the placozoan *Trichoplax adherens*, nor in any nonmetazoan organism. Analysis of codon alignments by the sitewise likelihood ratio method gave evidence for strong purifying selection on all codons of mammalian CKK domains, potentially indicating conserved function. Interestingly, the *Drosophila* homologue of the CAMSAP family is encoded by the *ssp4* gene, which is required for normal formation of mitotic spindles. To investigate function of the CKK domain, human CAMSAP1-enhanced green fluorescent protein (EGFP) and fragments including the CKK domain were ex-

We conclude that the CKK domain binds microtubules and represents a domain that evolved with the metazoa.

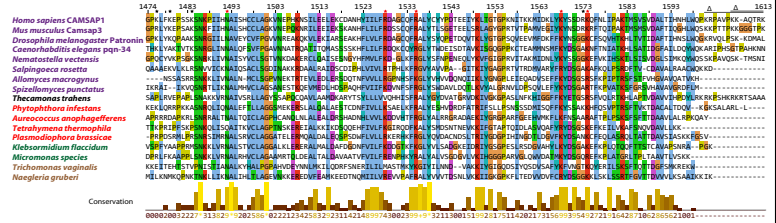
*Mol. Biol. Evol.* 26(9):2005–2014. 2009
doi:10.1093/molbev/msp115
Advance Access publication June 9, 2009

## Yet …

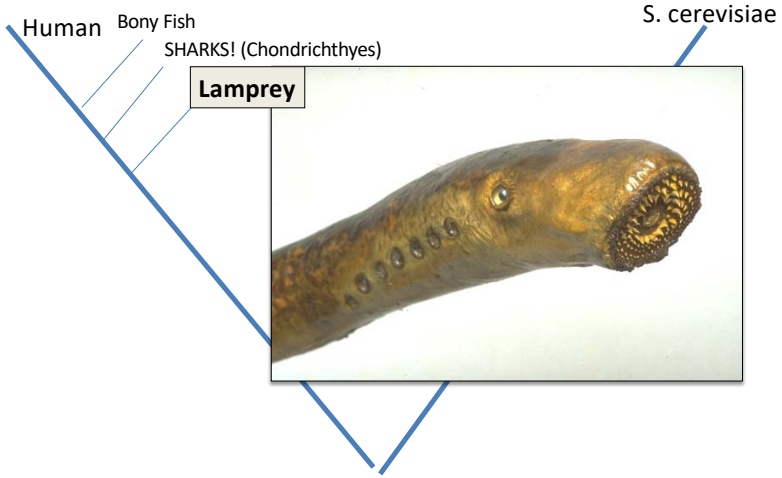
ARTICLES
nature structural & molecular biology

A structural model for microtubule minus-end recognition and protection by CAMSAP proteins

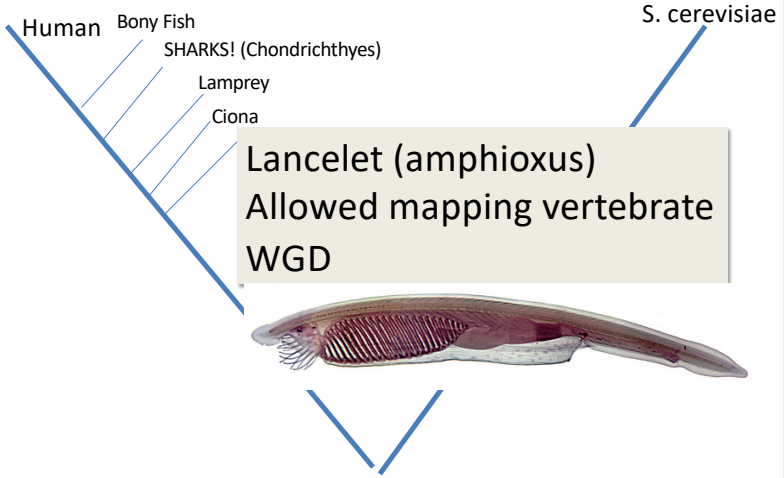
Joseph Atherton[1,11], Kai Jiang[2,11], Marcel M Stangier[3], Yanzhang Luo[4], Shasha Hua[2], Klaartje Houben[4], Jolien J E van Hooff[5–7], Agnel-Praveen Joseph[1], Guido Scarabelli[8], Barry J Grant[8], Anthony J Roberts[1], Maya Topf[1], Michel O Steinmetz[3,10], Marc Baldus[4] & Carolyn A Moores[1] & Anna Akhmanova[2]

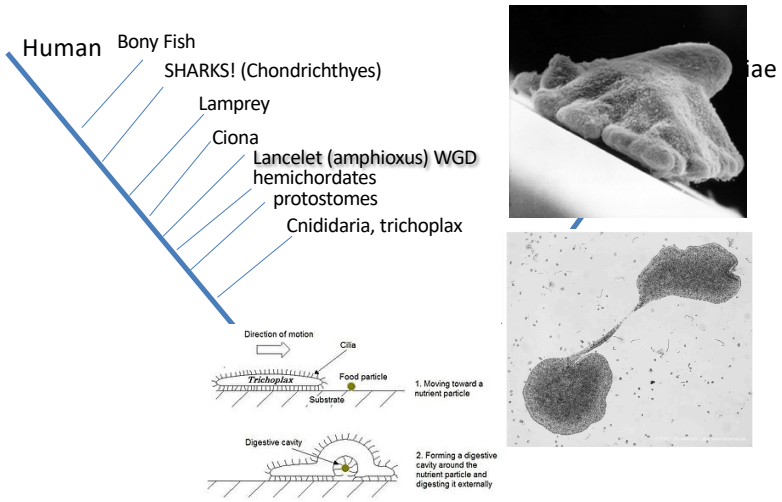



## Crucial genomes fill gaps

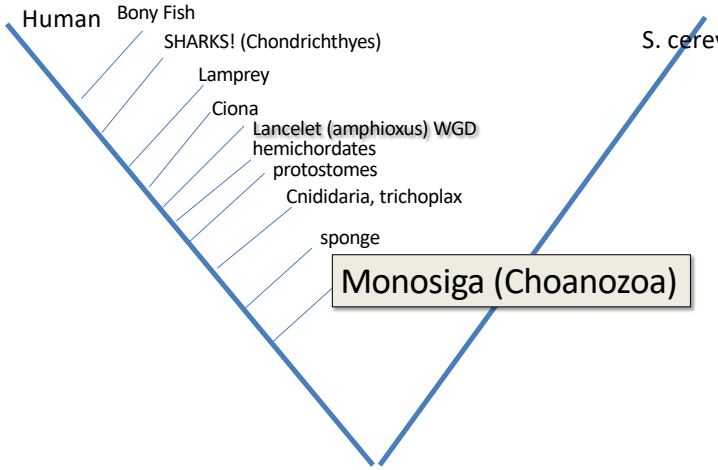



## Crucial genomes fill gaps



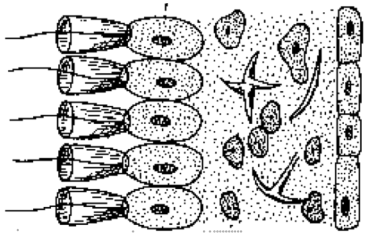



http://www.trichoplax.com/mediafiles/l15.jpg
http://genome.jgi-psf.org/Triad1/TrichoplaxSEM_Sagasser.jpg
http://upload.wikimedia.org/wikipedia/en/thumb/f/f7/Exodigestion_in_Trichoplax_adhaerens.jpg/380px-Exodigestion_in_Trichoplax_adhaerens.j




# Monosiga brevicollis, choanoflagelates single celled / colonial protists with a collar and flagellum to filter feed

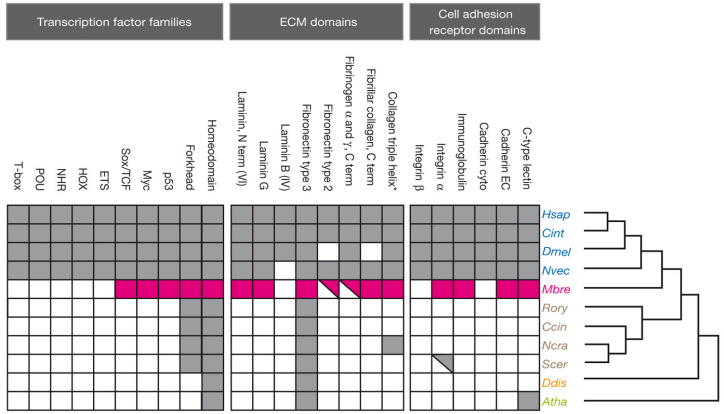
Monosiga



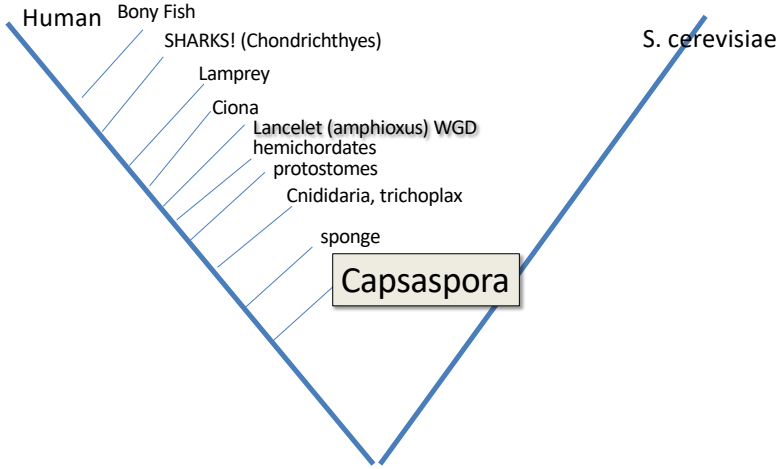
Portion of the body wall of a sponge

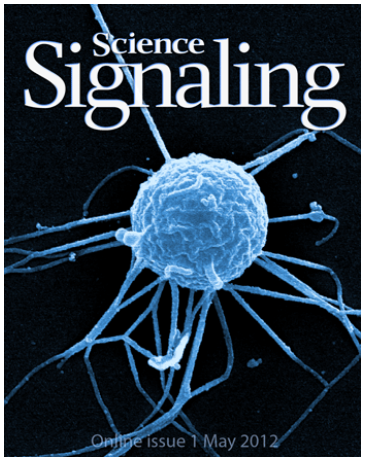
http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/C/choanoflagellates.gif

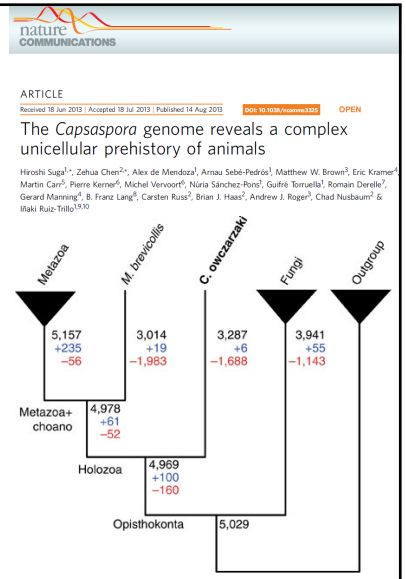
# Monosiga brevicolis

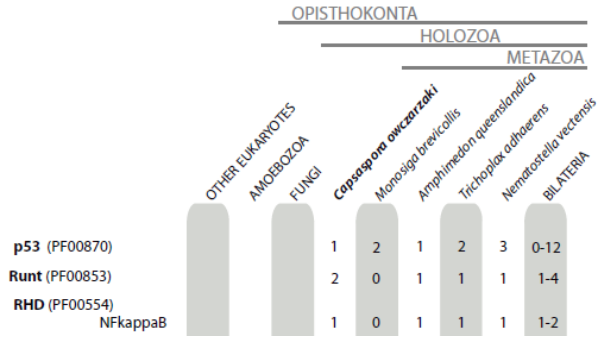



The genome of the choanoflagellate **Monosiga brevicollis** and the origin of metazoans.
**King** N, Westbrook MJ, Young SL, Kuo A, Abedin M, Chapman J, Fairclough S, Hellsten U, Isogai Y, Letunic I, Marr M, Pincus D, Putnam N, Rokas A, Wright KJ, Zuzow R, Dirks W, Good M, Goodstein D, Lemons D, Li W, Lyons JB, Morris A, Nichols S, Richter DJ, Salamov A, Sequencing JG, Bork P, Lim WA, Manning G, Miller WT, McGinnis W, Shapiro H, Tjian R, Grigoriev IV, Rokhsar D.
Nature. 2008 Feb 14;451(7180):783-8

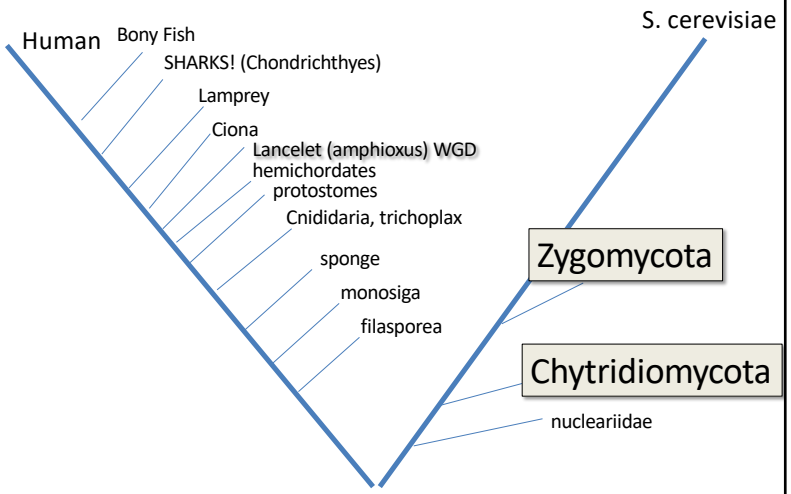


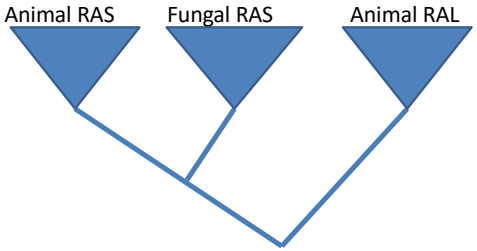



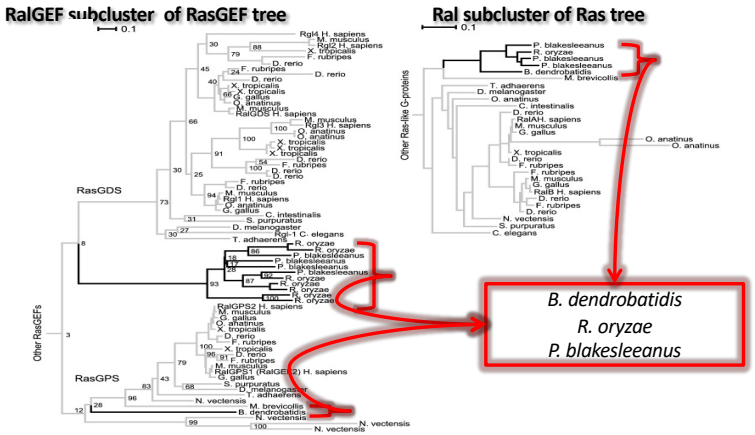

"acquisition of metazoan-specific developmental systems **and the co-option of pre-existing genes** drove the transition from unicellular protists to metazoans." e.g. NFkappaB



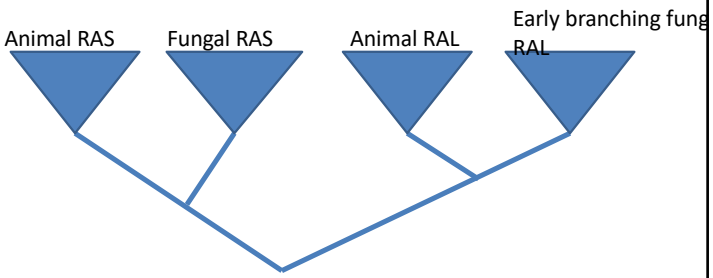
## However more data (better taxon sampling) >> tree reconciliation: the case of RAL evolution?




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



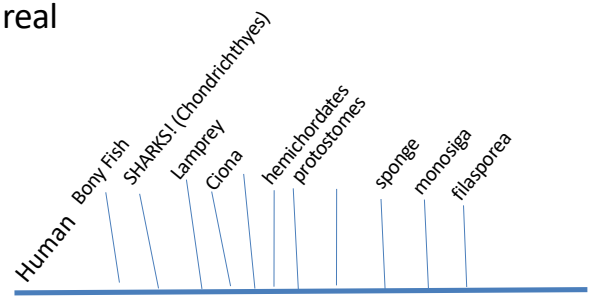

## However more data (better taxon sampling) >> tree reconciliation: the case of RAL evolution?



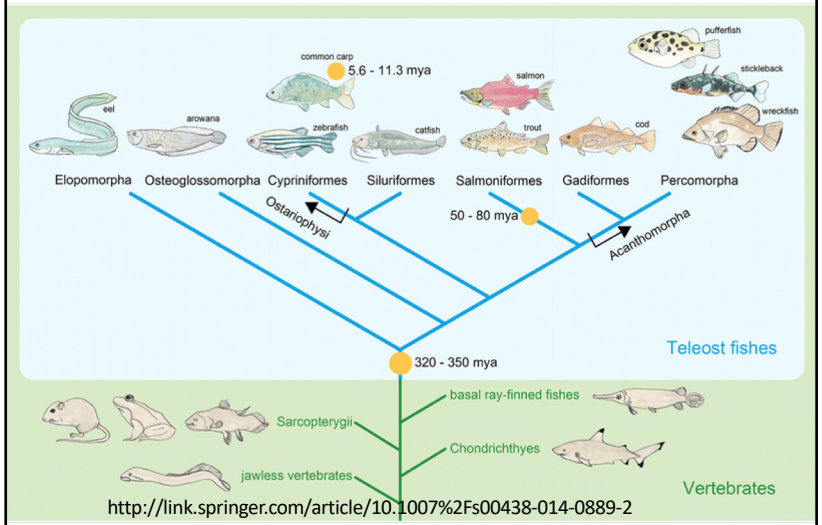

Old duplication **and** loss. The hypothesis of animal specific duplication and accelerated evolution & wrong gene tree can be rejected

## Is the asymmetry (comb) real?

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



# Matter of perspective or not?




http://link.springer.com/article/10.1007%2Fs00438-014-0889-2

# many genomes, many more underway

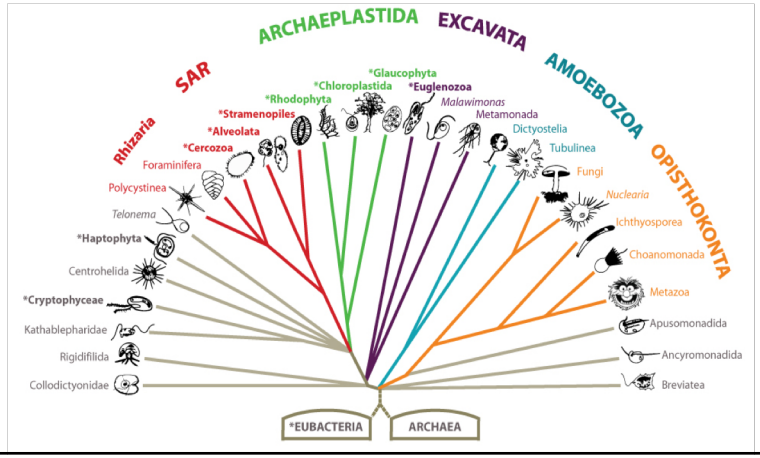
- Asgard archaea, tens of new bacterial phyla
- Diversity at many levels
- Allow / needed for different questions
- Reveals more old diversity re: duplicates or OGs

- Fun biology (not directly applicable but helps to remember the names and relationships of the weird beasties) (a good taxonomy button like in jackhmmer also helps)

# contents

- Key Genomes: Counting back from human (and *S. cerevisiae*) “crucial” / “early branching” genomes
- Eukaryotic supergroups & Root
- After LECA: gene loss, secondary endosymbiosis

# Outline of eukaryotic tree of life



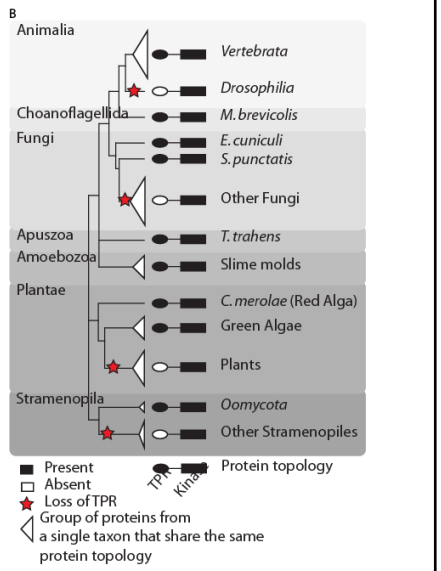
# ~5 Supergroups

- Current sampling hugely biased >> 1000 opistokonts, 2 rhizaria, 10(?) 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, amoebozoa, stramenopiles

# MPS1 parallel loss of TPR domain

Early branching / key genomes **in** supergroups gives beautiful stories

Tromer / kops in press




Current Biology 24, 465–470, February 17, 2014 ©2014 Elsevier Ltd All rights reserved http://dx.doi.org/10.1016/j.cub.2014.01.036

Report

## An Alternative Root for the Eukaryote Tree of Life

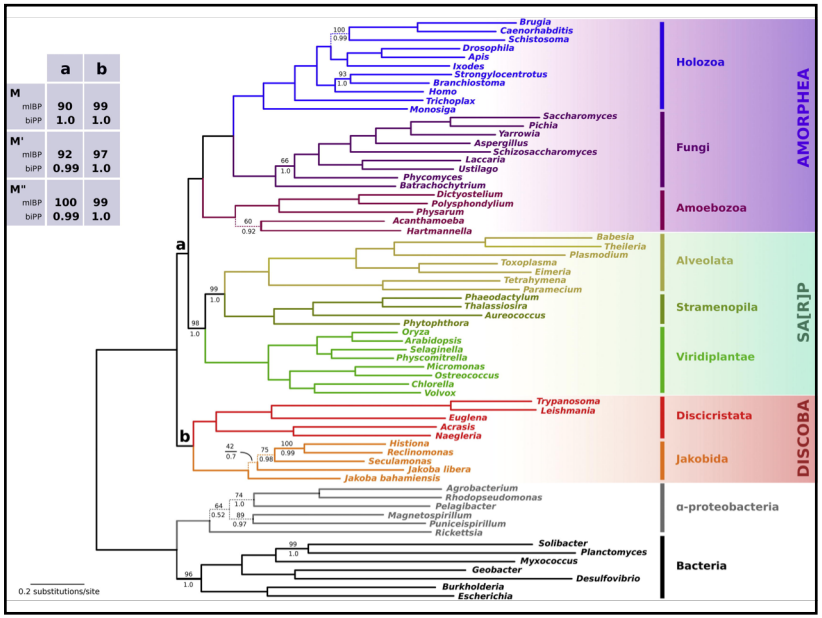
Ding He,[1,2] Omar Fiz-Palacios,[1,2] Cheng-Jie Fu,[1] Johanna Fehling,[1,3] Chun-Chieh Tsai,[1] and Sandra L. Baldauf[1,*]
[1]Program in Systematic Biology, Department of Organismal Biology, Uppsala University, Norbyvägen 18D, 75236 Uppsala, Sweden

Summary

The root of the eukaryote tree of life defines some of the most fundamental relationships among species. It is also critical for defining the last eukaryote common ancestor (LECA), the shared heritage of all extant species. The unikont-bikont root has been the reigning paradigm for eukaryotes for more than 10 years [1] but is becoming increasingly controversial [2–4]. We developed a carefully vetted data set, consisting of 37 nuclear-encoded proteins of close bacterial ancestry (euBacs) and their closest bacterial relatives, augmented by deep sequencing of the *Acrasis kona* (Heterolobosea, Discoba) transcriptome. Phylogenetic analysis of these data produces a highly robust, fully resolved global phylogeny of eukaryotes. The tree sorts all examined eukaryotes

Two parallel protocols employing a combination of homologous clustering and phylogenetic screening were used to identify proteins suitable for deep eukaryote phylogeny (see Supplemental Experimental Procedures and Figure S1 available online). Screening identified genes that appear to be (1) of bacterial origin, (2) present in the last eukaryote common ancestor (LECA) (universal or nearly universal among eukaryotes), and (3) with strong phylogenetic signal (out-paralog free and consistent with well-supported eukaryote phylogeny [5]). Of the 281 universal euBac proteins identified, most failed the latter criteria, primarily due to early gene duplication and lineage-specific losses.

Thirty-seven euBacs survived all screening protocols, 33 of which are known or predicted to function in the mitochondrion (Table S1). To increase sampling for Excavata, we sequenced the *Acrasis kona* (Heterolobosea, Discoba) transcriptome, yielding a full set of 37 euBac proteins. Outgroup taxa included the closest bacterial relatives of the 37 euBacs (Table S2). All euBacs in the final data set reproduce eight or more of the ten major eukaryote groups represented, and 36 euBacs show >60% maximum-likelihood bootstrap (mlBP) support for six or more of these major groups (Table S4).



PNAS PLUS

# Bacterial proteins pinpoint a single eukaryotic root

Romain Derelle[a,b,1], Guifré Torruella[c], Vladimír Klimeš[d], Henner Brinkmann[e], Eunsoo Kim[f], Čestmír Vlček[g], B. Franz Lang[h], and Marek Eliáš[d]

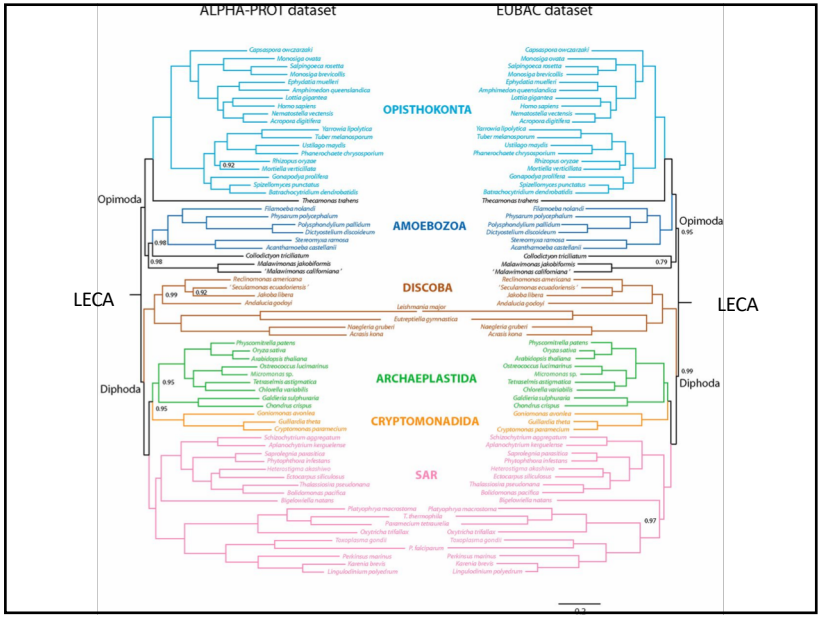
[a]Centre for Genomic Regulation, 08003 Barcelona, Spain; [b]Universitat Pompeu Fabra, 08003 Barcelona, Spain; [c]Institut de Biologia Evolutiva, Consejo Superior de Investigaciones Científicas–Universitat Pompeu Fabra, 08003 Barcelona, Spain; [d]Faculty of Science, Department of Biology and Ecology, University of Ostrava, 710 00 Ostrava, Czech Republic; [e]Leibniz-Institut DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, D-38124 Braunschweig, Germany; [f]Sackler Institute for Comparative Genomics and Division of Invertebrate Zoology, American Museum of Natural History, New York, NY 10024; [g]Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, 142 20 Prague 4, Czech Republic; and [h]Robert Cedergren Centre for Bioinformatics and Genomics, Département de Biochimie, Université de Montréal, Montreal, QC, Canada H3T 1J4

Edited by Thomas Martin Embley, University of Newcastle upon Tyne, Newcastle upon Tyne, United Kingdom, and accepted by the Editorial Board January 13, 2015 (received for review October 28, 2014)

**The large phylogenetic distance separating eukaryotic genes and their archaeal orthologs has prevented identification of the position of the eukaryotic root in phylogenomic studies. Recently, an innovative approach has been proposed to circumvent this issue: the use as phylogenetic markers of proteins that have been transferred from bacterial donor sources to eukaryotes, after their emergence from Archaea. Using this approach, two recent independent studies have built phylogenomic datasets based on bacterial sequences, leading to different predictions of the eukaryotic root. Taking advantage of additional genome sequences from the jakobid *Andalucia godoyi* and the two known malawimonad species (*Malawimonas jakobiformis* and *Malawimonas californiana*), we reanalyzed these two phylogenomic datasets. We show that both datasets pinpoint the same phylogenetic position of the eukaryotic root that is between "Unikonta" and "Bikonta," with malawimonad and collodic-**

constantly find fast evolving eukaryotes at the base of all other eukaryotes (9–12).

In the absence of a close outgroup, rare cytological and genomic changes specific to some eukaryotic lineages have also been considered for rooting of the eukaryotic tree. In this context, the leading hypothesis used to be the Unikonta–Bikonta dichotomy, in which unikonts and bikonts are ancestrally characterized by (arguably) either a single or two flagella, respectively. This subdivision seemed to be supported by the distribution of certain gene fusions (13), and a specific myosin paralog (14), but both characters later proved to have a more complex evolutionary history (2). Furthermore, the idea of the "uniflagellate" ancestry for unikonts became untenable (2). For this reason, the concept of Unikonta has been recently superseded by proposing a "megagroup" Amorphea, which embraces unikonts as well as




# New genomes = new phylogeney

LETTER

https://doi.org/10.1038/s41586-018-0708-8

## Hemimastigophora is a novel supra-kingdom-level lineage of eukaryotes

Gordon Lax[1,4], Yana Eglit[1,4], Laura Eme[2,3,4], Erin M. Bertrand[1], Andrew J. Roger[2] & Alastair G. B. Simpson[1*]

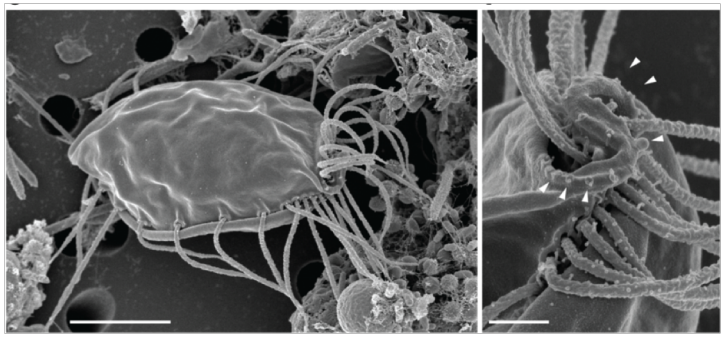
**Almost all eukaryote life forms have now been placed within one of five to eight supra-kingdom-level groups using molecular phylogenetics[1–4]. The 'phylum' Hemimastigophora is probably the most distinctive morphologically defined lineage that still awaits such a phylogenetic assignment. First observed in the nineteenth century, hemimastigotes are free-living predatory protists with two rows of flagella and a unique cell architecture[5–7]; to our knowledge, no molecular sequence data or cultures are currently available for this group. Here we report phylogenomic analyses based on high-coverage, cultivation-independent transcriptomics that place Hemimastigophora outside of all established eukaryote supergroups. They instead comprise an independent supra-kingdom-level lineage that most likely forms a sister clade to the 'Diaphoretickes' half of eukaryote diversity (that is, the 'stramenopiles, alveolates and Rhizaria' supergroup (Sar), Archaeplastida and Cryptista, as well as other major groups). The**

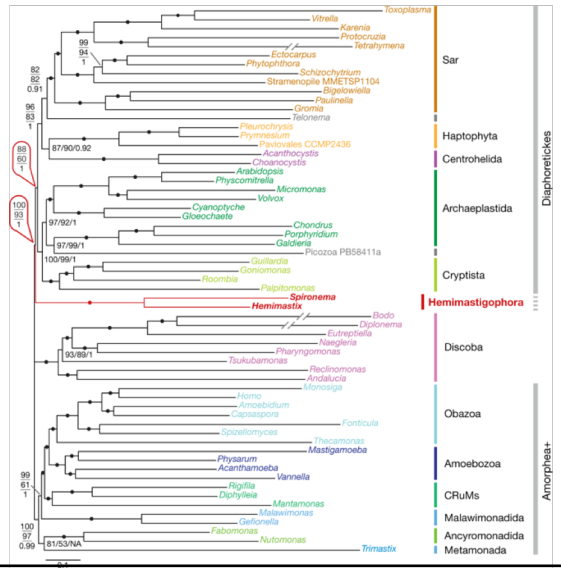
**Gene sequence.** The partial small subunit ribosomal RNA (SSU rRNA) gene sequence of strain BW2H has been deposited in GenBank, accession code MF682191.

**Comments.** Cells are larger and have several more flagella than *Hemimastix amphikineta*, the only previously described species (14-μm by 7-μm cell body, 12 flagella per row[6]).

Cells of *H. kukwesjijk* are oval in profile with a blunt anterior projection (the capitulum) and two rows of flagella along their whole length (Fig. 1b, Extended Data Fig. 1). In cultivation as strain BW2H, live cells were 16.5–20.5-μm long by 7–12.5-μm wide (18.3 ± 1 μm × 9.9 ± 1.2 μm; $n = 61$), with a sub-central, rounded nucleus and posterior contractile vacuole (Fig. 1c). Each row of 17–19 flagella (mean 18.4; $n = 25$) lay in a channel between the two thick thecal plates. The anteriormost 9 or 10 flagella were closely spaced, and the rest emerged from separate notches in the underlying plate (Fig. 1b, e). The capitulum was bordered by the overlapping anterior

# hemimastigophora






# Events after LECA not loss/duplication

- HGT between eukaryotes
- Parallel HGT from bacteria
- Serial / secondary endosymbiosis
- (tertriary 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.

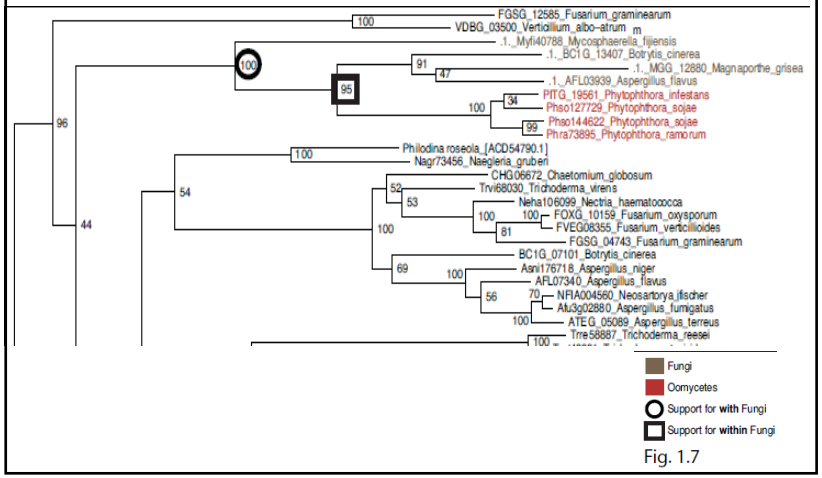


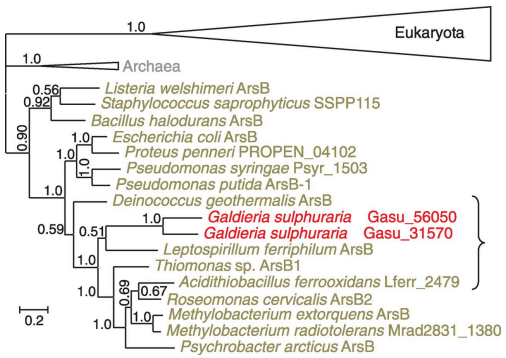

Fig. 1.7

# HGT from bacteria

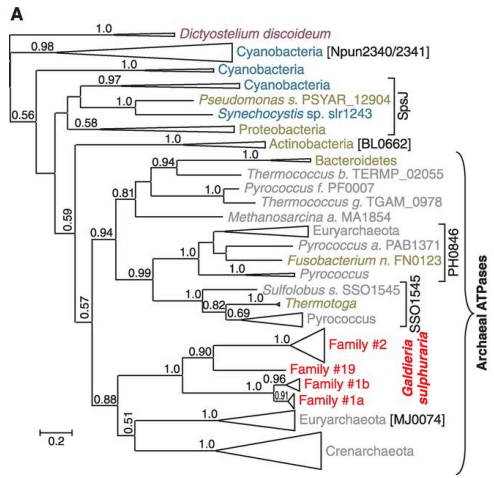


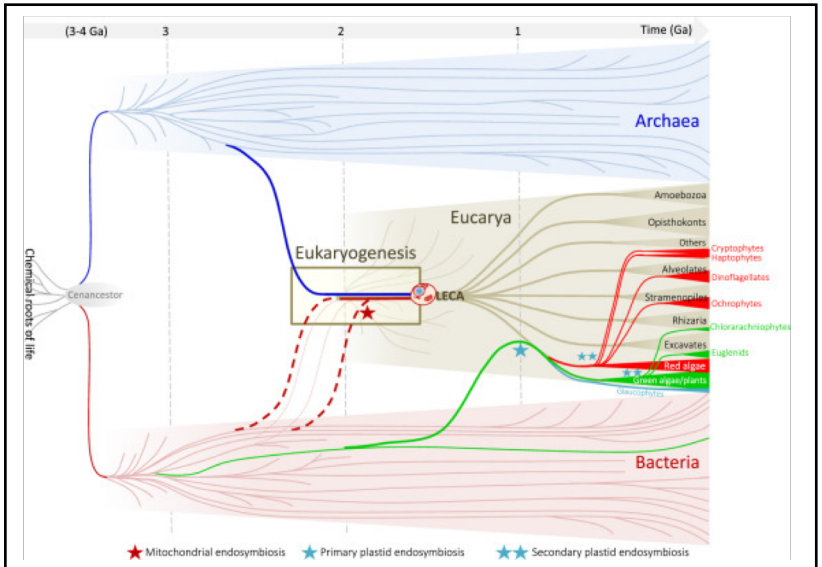
Science
AAAS

**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

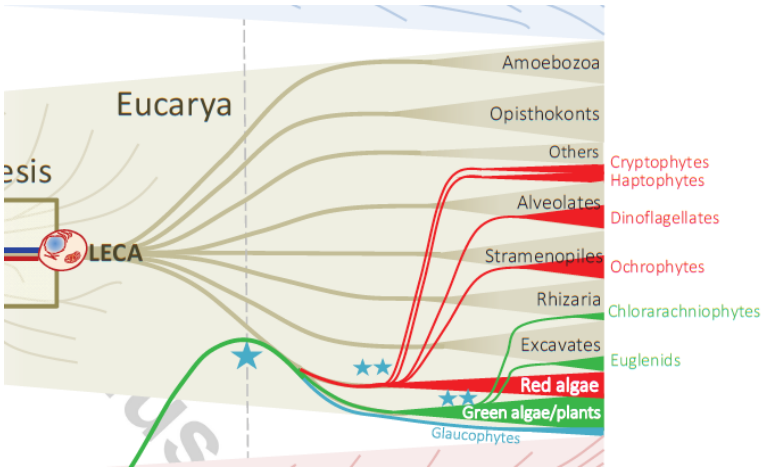


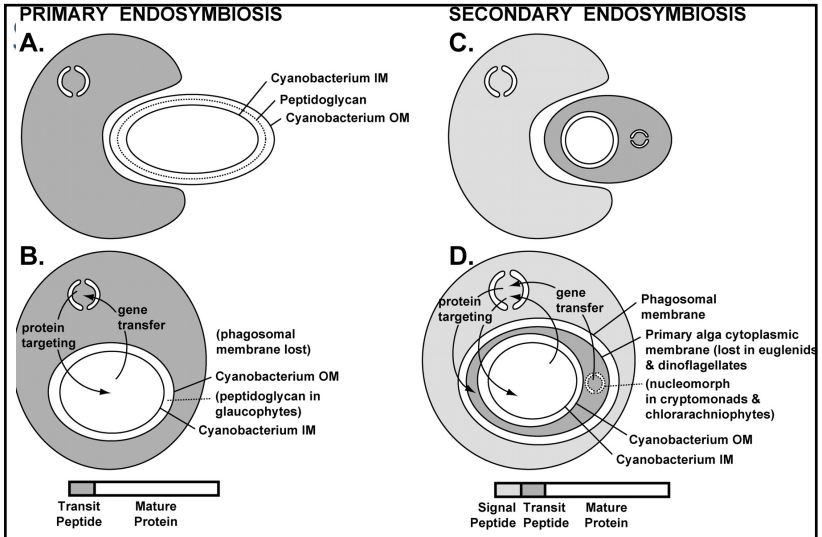
# Parallel HGT from bacteria



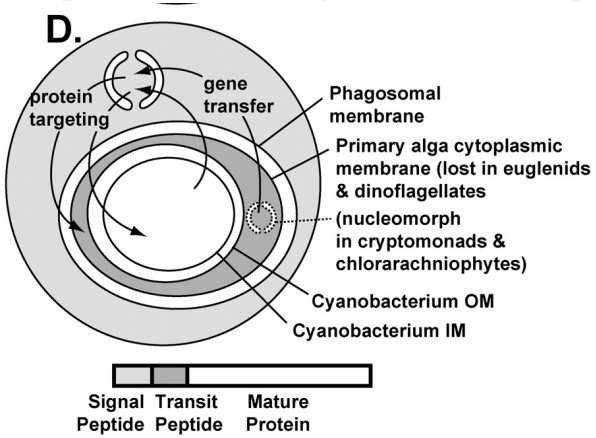

http://www.sciencedirect.com/science/article/pii/S0022519317300954



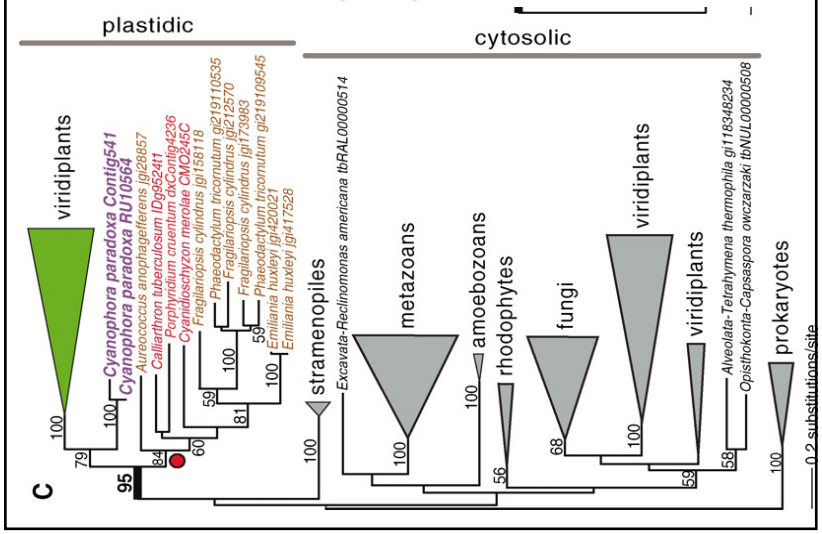

http://www.amjbot.org/content/91/10/1481/F2.large.jpg

# 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




# Secondary endosymbiosis explains position of diatoms in this tree