Origin and evolution of eukaryotic cellular processes

How new genomic data and new bioinformatic analyses have changed our view on the evolution of cellular complexity

Prof. dr. Berend Snel

Theoretical Biology and Bioinformatics

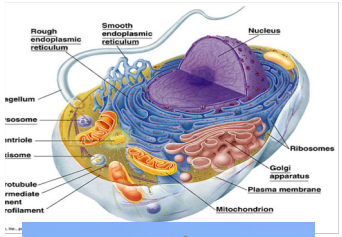
Utrecht University

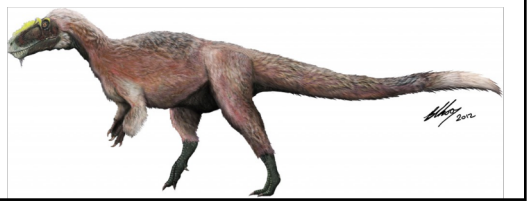
Utrecht Bioinformatics Center

Universiteit Utrecht

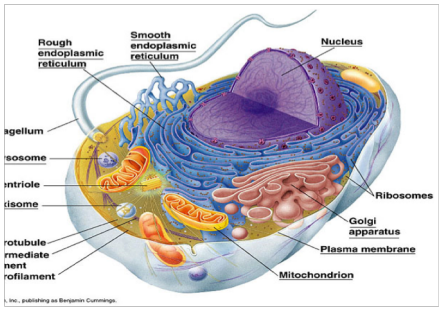
- Introduction: a complex ancestor and independent loss
- Developments that lead to this view:
  - new genomes
  - sensitive homology searches
  - analysis of large gene families
- A null model of genome evolution
  - innovations

Eukaryotes: most of the visible biosphere

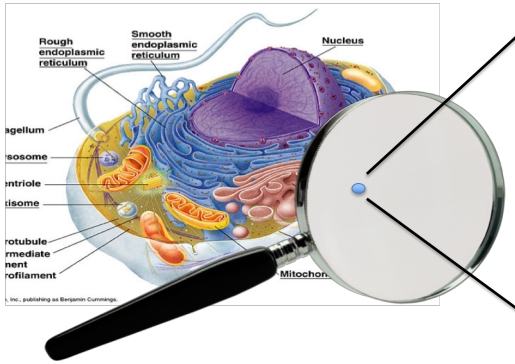


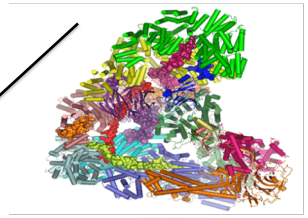


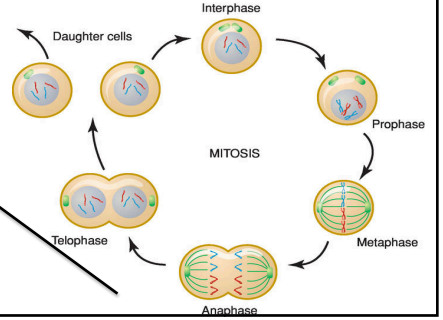
Eukaryotic cell



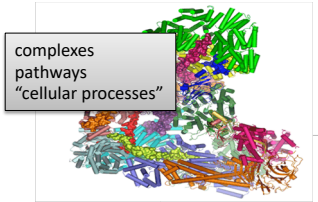
Protein complexes
Cellular processes



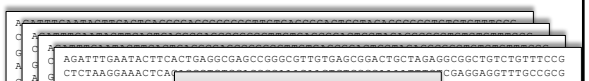




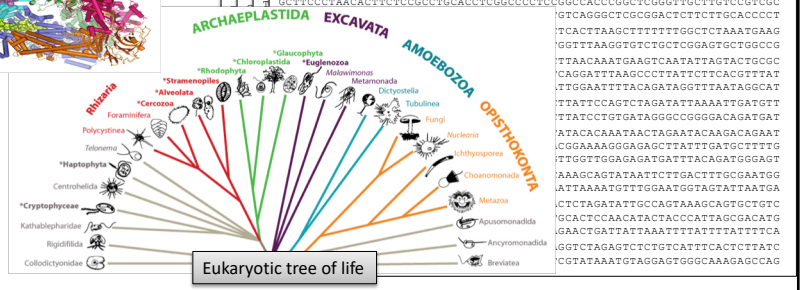
How do we study evolution of cellular processes & complexes



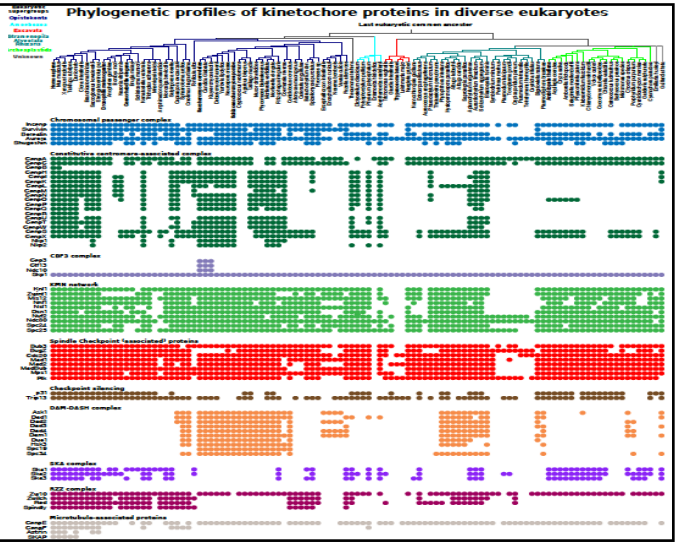
complexes
pathways
"cellular processes"

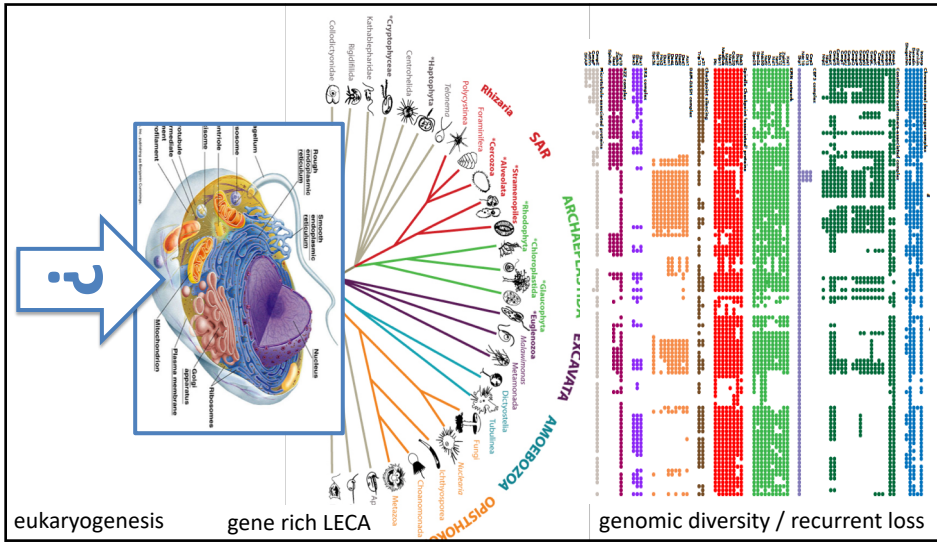


genomes (predicted proteomes)

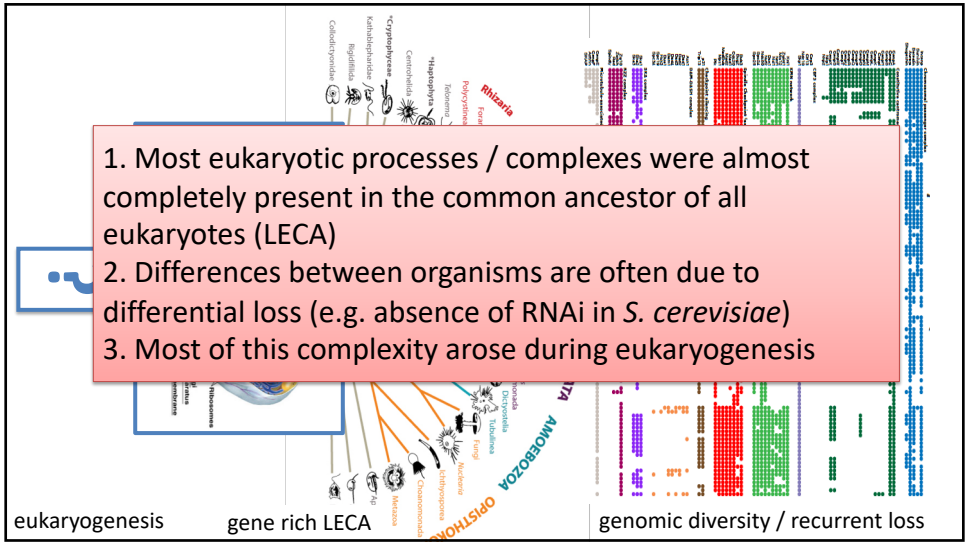


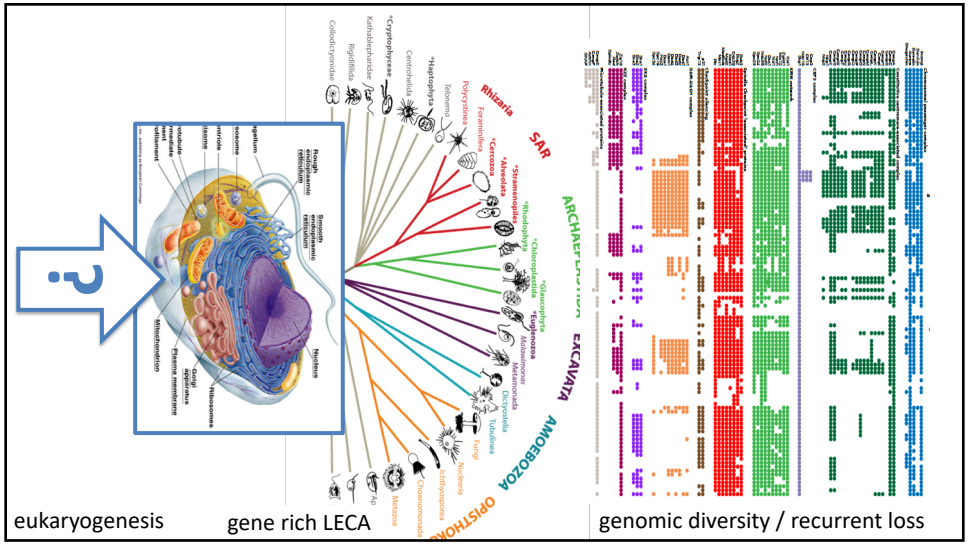
Hunt for orthologs to create phylogenetic profiles to infer their evolution

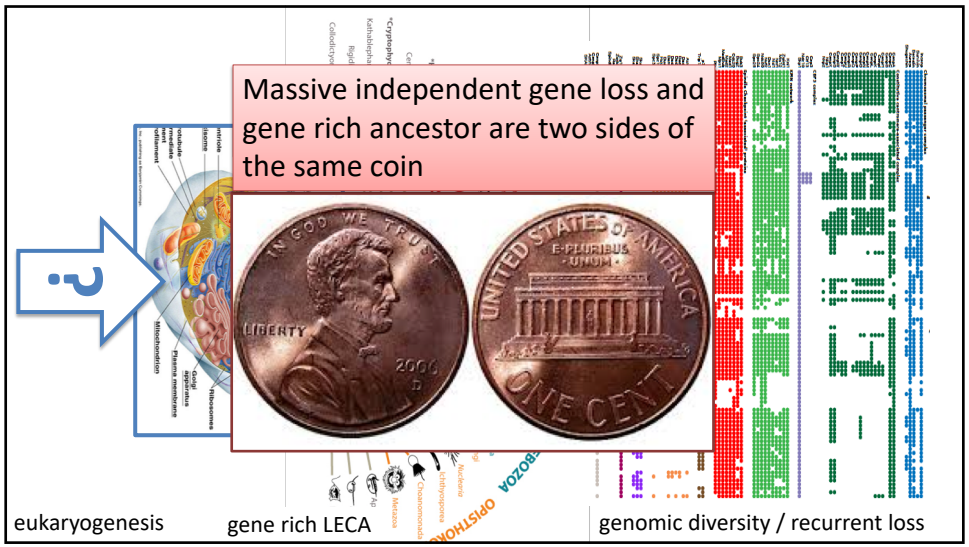




eukaryogenesis | gene rich LECA | genomic diversity / recurrent loss

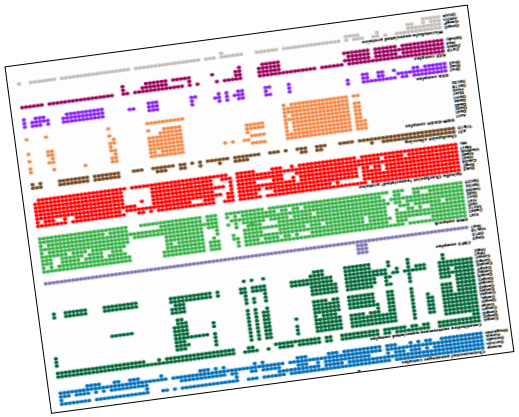






- Introduction: a complex ancestor and independent loss
- Developments that lead to this view:
  - new genomes
  - sensitive homology searches
  - analysis of large gene families
- Eukaryogenesis:
  - Endosymbiosis
  - Gene duplications
  - Gene inventions
  - Ancestral genes: new genomes
- A null model of genome evolution
  - innovations

How did we reveal the trend of a complex ancestor and independent loss?

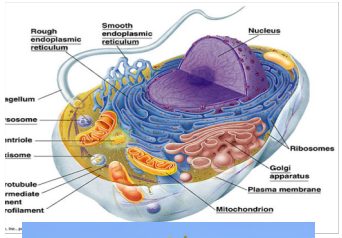


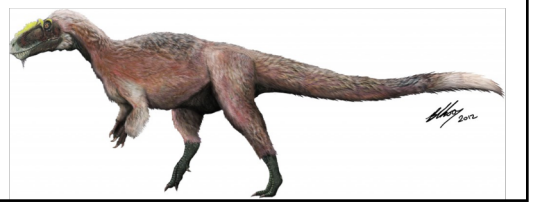
How we revealed the trend of a complex ancestor and independent loss.

A combination of:

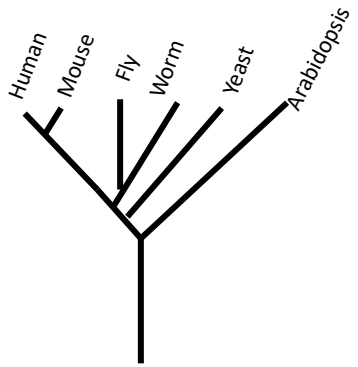
- New genomes at crucial positions
- Improved sensitivity of sequence similarity searches (and homologs that are orthologs)
- Studying gene families with a lot of pre-LECA duplications

Eukaryotes: most of the visible biosphere

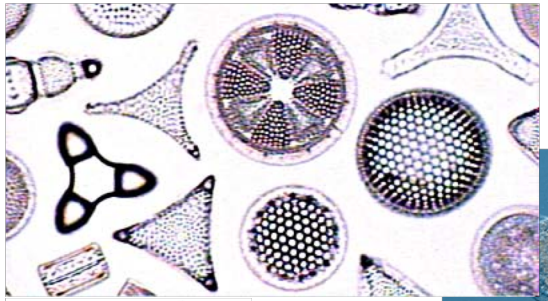


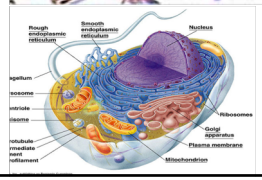


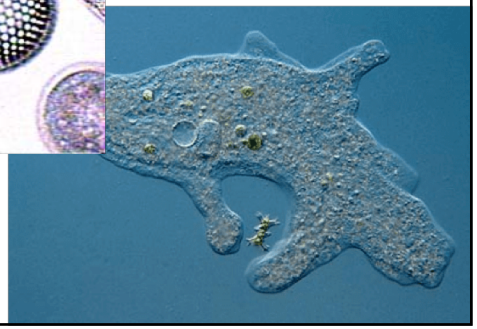
Many ideas / analyses originally based on human, mouse, fly, worm, yeast, Arabidopsis ….



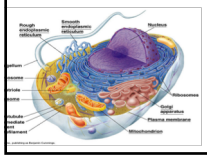
Eukaryotes: most eukaryotes are not visible



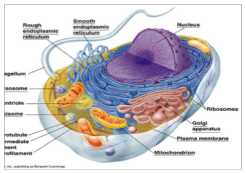


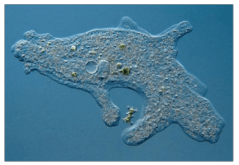


Eukaryotes: quite some visible eukaryotes are not fungi, animals or plants

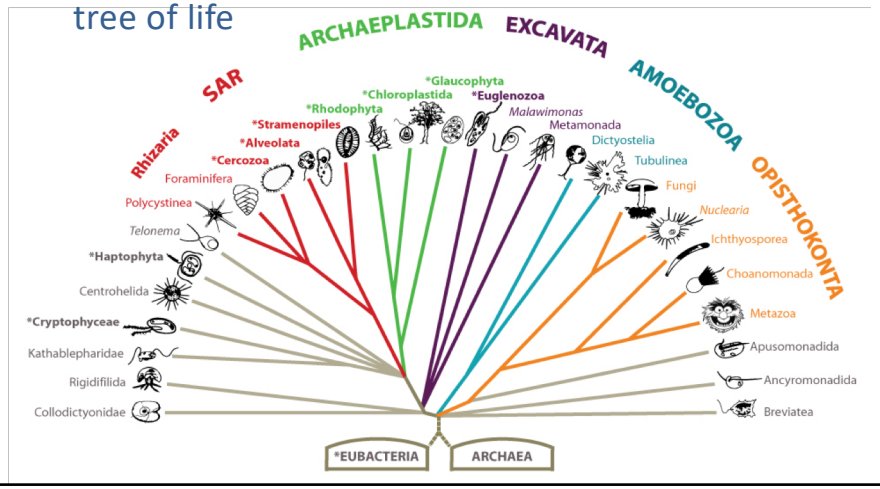


Much more eukaryotic evolutionary and cellular diversity than you might think …

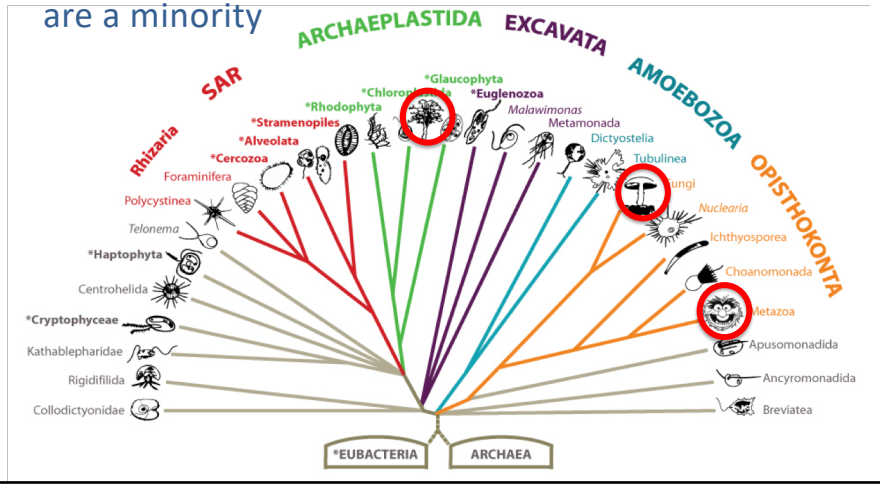


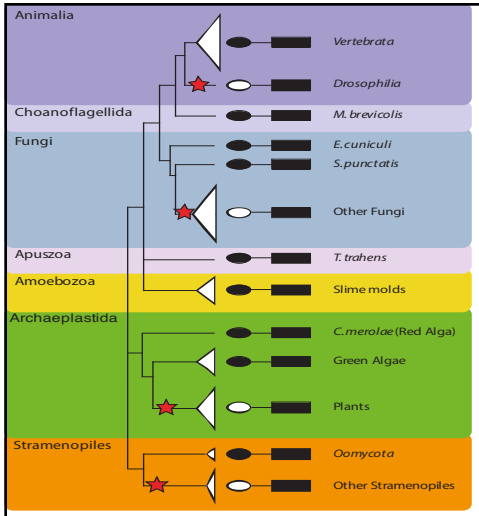


Outline of eukaryotic tree of life



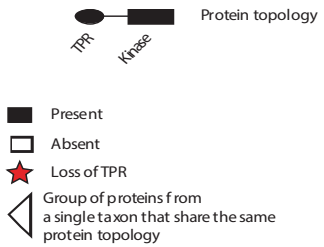
Plants, animal, fungi are a minority





http://jcb.rupress.org/content/201/2/217.long

Recurrent loss of TPR repeat in evolution in MPS1 revealed by additional eukaryotic diversity



Cell

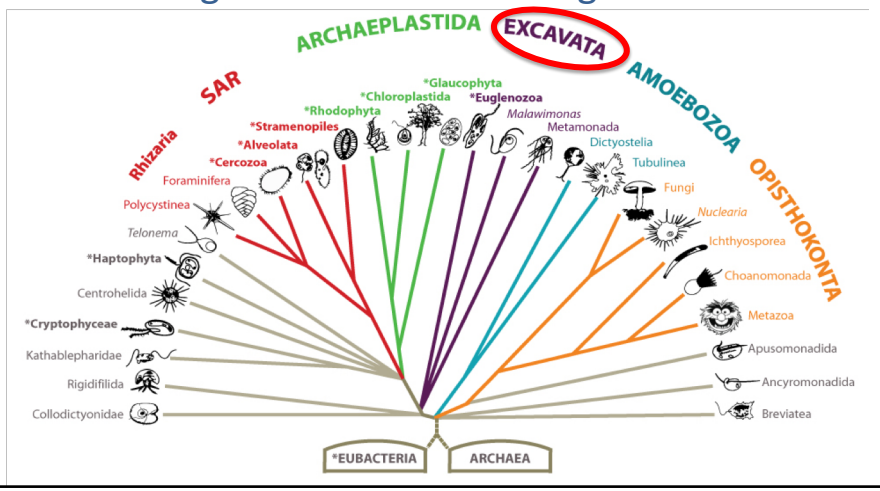
The Genome of *Naegleria gruberi* Illuminates Early Eukaryotic Versatility

Lillian K. Fritz-Laylin,[1,10] Simon E. Prochnik,[3,10] Michael L. Ginger,[4] Joel B. Dacks,[5,6] Meredith L. Carpenter,[1] Mark C. Field,[6] Alan Kuo,[3] Alex Paredez,[1] Jarrod Chapman,[3] Jonathan Pham,[7] Shengqiang Shu,[3] Rochak Neupane,[2] Michael Cipriano,[7] Joel Mancuso,[8] Hank Tu,[3,11] Asaf Salamov,[3] Erika Lindquist,[3] Harris Shapiro,[3] Susan Lucas,[3] Igor V. Grigoriev,[3] W. Zacheus Cande,[1] Chandler Fulton,[9] Daniel S. Rokhsar,[1,3,*] and Scott C. Dawson[7,*]

[1]Department of Molecular and Cell Biology
[2]Center for Integrative Genomics, 545 Life Sciences Addition
University of California, Berkeley, Berkeley, CA 94720, USA
[3]U.S. Department of Energy, Joint Genome Institute, Walnut Creek, CA 94598, USA
[4]School of Health and Medicine, Division of Biomedical and Life Sciences, Lancaster University, Lancaster LA1 4YQ, UK
[5]Department of Cell Biology, University of Alberta Edmonton, Alberta, Canada
[6]The Molteno Building, Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QT, UK
[7]Department of Microbiology, University of California, Davis, CA 95616, USA
[8]Gatan Inc., 5794 W. Las Positas Boulevard, Pleasanton, CA 94588, USA
[9]Department of Biology, Brandeis University, Waltham, MA 02454-9110, USA
[10]These authors contributed equally to this work
[11]Present address: Life Technologies, 850 Lincoln Center Drive, Foster City, CA 94404, USA
*Correspondence: dsrokhsar@gmail.com (D.S.R.), scdawson@ucdavis.edu (S.C.D.)

DOI 10.1016/j.cell.2010.01.032

First genome of a free living excavate



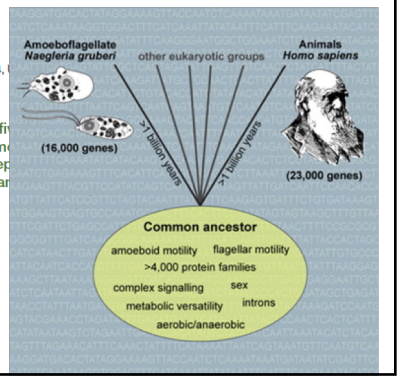
Cell

Leading Edge
Previews

The Incredible Expanding Ancestor of Eukaryotes

Eugene V. Koonin[1,*]
[1]National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD 20894,
*Correspondence: koonin@ncbi.nlm.nih.gov
DOI 10.1016/j.cell.2010.02.022

Comparing the genome sequences of free-living organisms in the fi
enables predictions to be made about the genome of the last comm
The genome sequence of the amoeboflagellate *Naegleria gruberi* rep
(2010) reveals the surprising complexity of this unicellular organism ar
common eukaryotic ancestor.

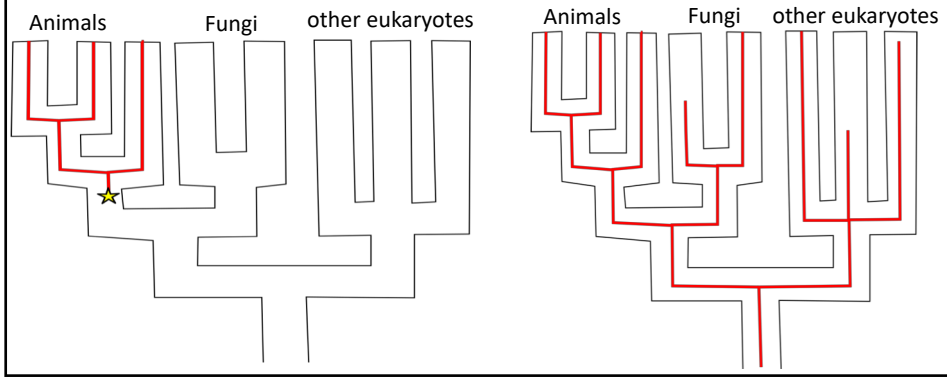


How we revealed the trend of a complex ancestor and independent loss

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 reveals lineage specific genes to be ancient with orthologs across eukaryotes



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

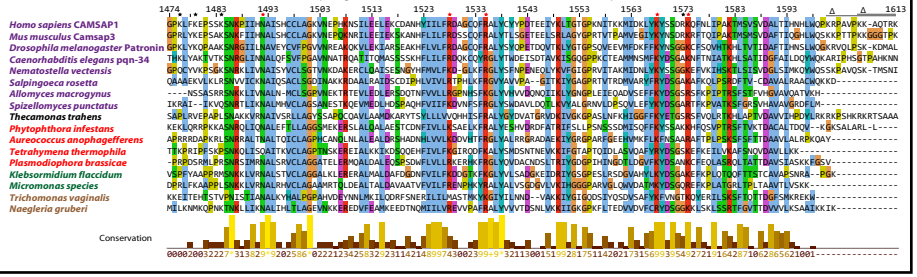
Yet …

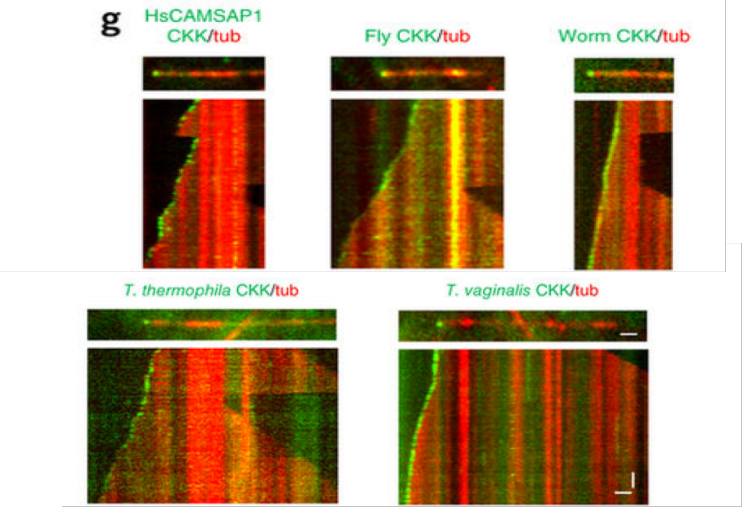
ARTICLES

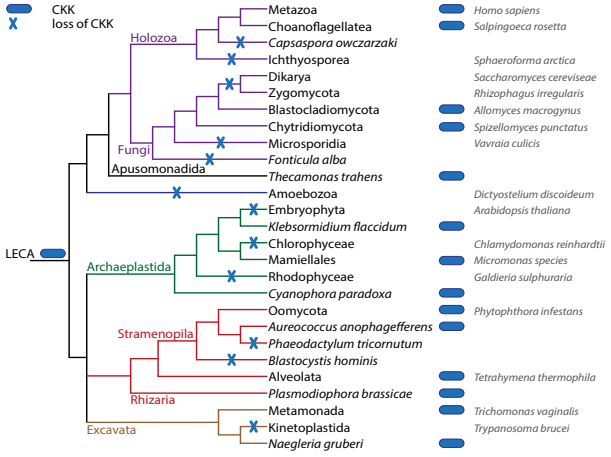
nature structural & molecular biology

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]







Improved sensitivity of sequence similarity searches. Homology modeling & crystal structure reveal ancient origins of DIX-domain

LETTERS
nature plants

A SOSEKI-based coordinate system interprets global polarity cues in *Arabidopsis*

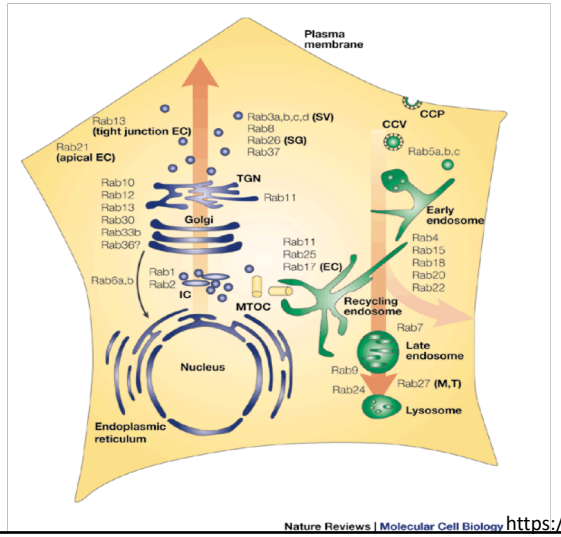
Saiko Yoshida, Alja van der Schuren, Maritza van Dop, Luc van Galen, Shunsuke Saiga, Milad Adibi, Barbara Möller, Colette A. ten Hove, Peter Marhavý, Richard Smith, Jiri Friml and Dolf Weijers

"We are about to submit the next paper in which we show the crystal structure and demonstrate that SOSEKI works in the same way as Dishevelled in animals." Dolf Weijers

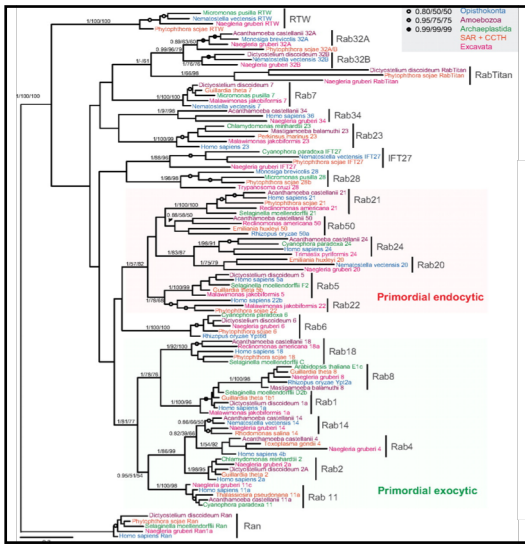
How we revealed the trend of a complex ancestor and independent loss

A combination of:

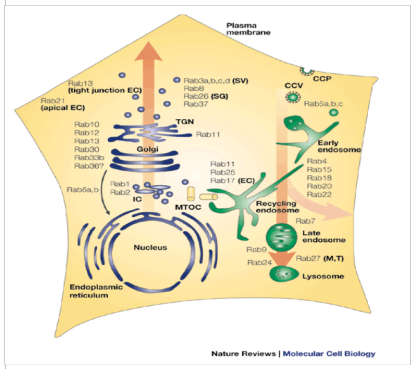
- New genomes at crucial positions
- Improved sensitivity of sequence similarity searches (and homologs that are orthologs)
- Studying gene families with a lot of pre-LECA duplications



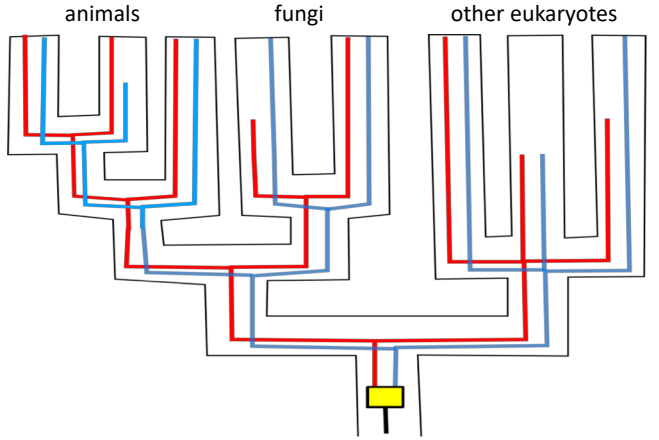
https://www.nature.com/articles/35052055/



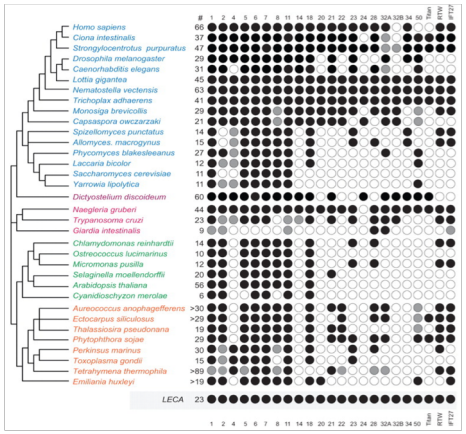
Endomembrane system /Rab



Elias *et al.* http://jcs.biologists.org/content/125/10/2500.long

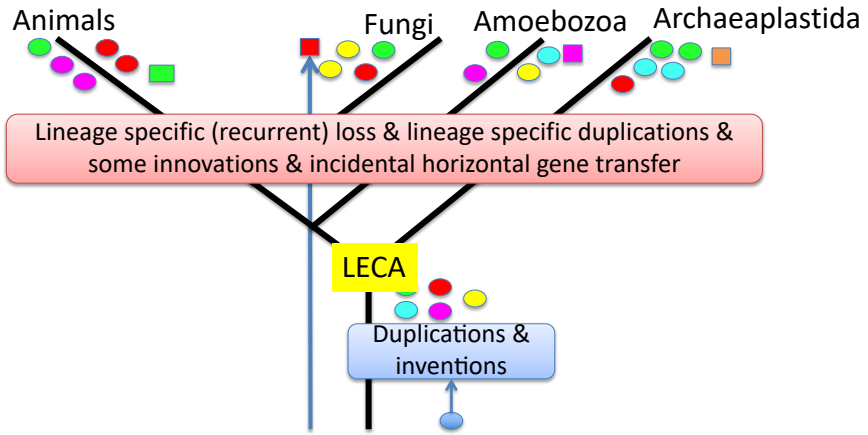


These 23 LECA Rab's were differentially retained in present-day eukaryotes -> massive loss



- Introduction: a complex ancestor and independent loss
- Developments that lead to this view:
  - new genomes
  - sensitive homology searches
  - analysis of large gene families
- A null model of genome evolution
  - innovations

What should you expect when studying a specific protein or a set of related genomes? "null model of gene family evolution"



REVIEWS

Evolution by gene loss

Ricard Albalat and Cristian Cañestro

Abstract | The recent increase in genomic data is revealing an unexpected perspective of gene loss as a pervasive source of genetic variation that can cause adaptive phenotypic diversity. This novel perspective of gene loss is raising new fundamental questions. How relevant has gene loss been in the divergence of phyla? How do genes change from being essential to dispensable and finally to being lost? Is gene loss mostly neutral, or can it be an effective way of adaptation? These questions are addressed, and insights are discussed from genomic studies of gene loss in populations and their relevance in evolutionary biology and biomedicine.

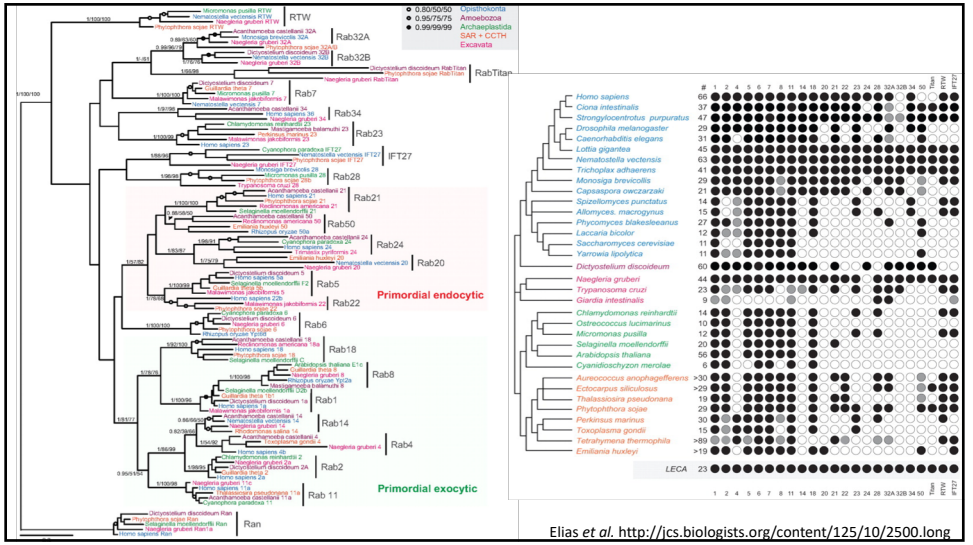
Pseudogenization
An evolutionary phenomenon whereby a gene loses its function, accumulates mutations and becomes a pseudogene.

Eumetazoan
Clade that classically includes all animals (metazoan) except sponges and Placozoa, although recent analyses of

Loss is nothing else but change, and change is Nature's delight — Marcus Aurelius, AD 121–180

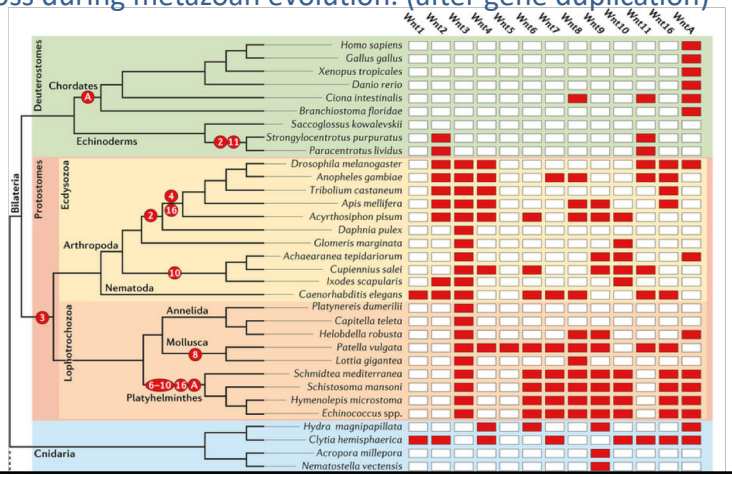
Great attention has in the past been paid to the mechanisms of evolution by gene duplication (that is, neofunctionalization and subfunctionalization)[1,2]. By contrast, gene loss has often been associated with the loss of redundant gene duplicates without apparent functional consequences, and therefore this process has mostly been neglected as an evolutionary force. However,

Here, we address some of the fundamental questions in evolutionary biology that have emerged from this novel perspective of evolution by gene loss. Examples from all life kingdoms are covered, from bacteria to fungi and from plants to animals, including key examples of gene loss in humans. We review how gene loss has affected the evolution of different phyla and address key questions, including how genes can become dispensable, how many of our current genes are actually dispensable, how patterns of gene loss are biased, and whether the effects of

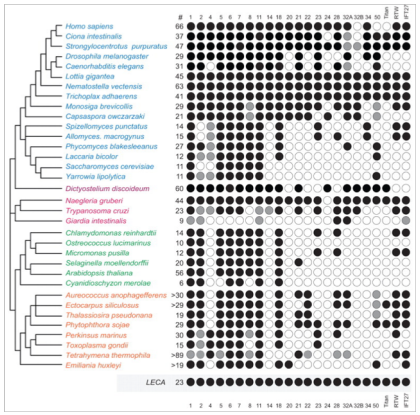


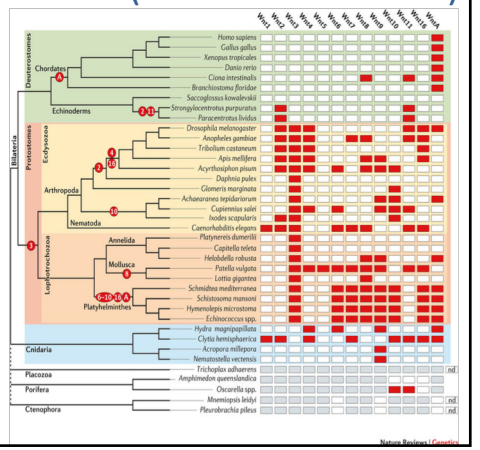
Elias *et al.* http://jcs.biologists.org/content/125/10/2500.long

wingless (Wnt) : a paradigmatic example of the pervasiveness of gene loss during metazoan evolution. (after gene duplication)



Duplication/Loss at different levels. e.g. origin of eukaryotes, origin of animals WNT (“*scale invariant*”?)





Prospects & Overviews

Problems & Paradigms

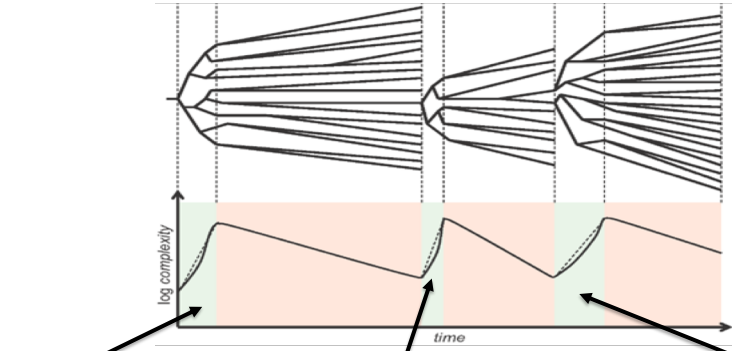
Genome reduction as the dominant mode of evolution

Yuri I. Wolf and Eugene V. Koonin*

A common belief is that evolution generally proceeds towards greater complexity at both the organismal and the genomic level, numerous examples of reductive evolution of parasites and symbionts notwithstanding. However, recent evolutionary reconstructions challenge this notion. Two notable examples are the reconstruction of the complex archaeal ancestor and the intron-rich ancestor of eukaryotes. In both cases, evolution in most

Introduction: Complexity can either increase or decrease during the evolution of various life forms

The textbook depiction of the evolution of life on earth is that of an ascent toward a steadily increasing organismal complexity: from primitive protocells to prokaryotic cells



- Introduction: a complex ancestor and independent loss
- Developments that lead to this view:
  - new genomes
  - sensitive homology searches
  - analysis of large gene families
- A null model of genome evolution
  - innovations

	Evolutionary history of a protein: how to & results from	Insights into genome and network evolution	Methods for the large scale study of genome and network evolution
Literature discussion	✓	✓	✓
Lectures	✓	✓	✓
COO, pen&paper	✓	X	X

Location		Date	Time	Lectures	Activities
KRUYT-O0123	Mon	11-3-2019	9:00-10:00	Introduction 1.	
	Mon	11-3-2019	10:00-15:00	Phylogenenies & Gene Phylogenenies: the basics*	Pen & paper and computer exercises*
KRUYT-O0123	Tue	12-3-2019	9:00-13:30	Introduction 2. Gene phylogenies. Large scale orthology.	Pen & paper and computer exercises
KRUYT-O0123	Wed	13-3-2019	9:00-15:00	Homology. Domains. Sequence evolution.	Journal club, Pen & paper and computer exercises
KRUYT-O0123	Thu	14-3-2019	9:00-15:00	Eukaryotic Tree of Life. Eukaryogenesis.	Paper exercise, Pen & paper and computer exercises
KRUYT-O0123	Fri	15-3-2019	9:00-15:00	Whole Genome Duplications	Journal club, Pen & paper and computer exercises
KRUYT-O0123	Mon	18-3-2019	9:00-15:00	Conservation of function of orthologs. Phylogenetic profiles. Complex/pathway membership evolution. Evolution of function after duplication.	Journal club, Pen & paper and computer exercises
KRUYT-O0123	Tue	19-3-2019	9:00-14:30	Exceptions to phylogenetic profiles similarity and implications for function evolution. Evolution of regulation. Summary.	Journal club, Pen & paper and computer exercises
	Wed	20-3-2019		self study	
	Thu	21-3-2019		self study	
RUPPERT-C	Fri	22-3-2019	13:30-16:30	Exam#	

* This is my part of Introduction to Bioinformatics. So it is not compulsory if you already completed this master course.

The exam is open book & first time this course will use a written exam of instead of a mini project

Unscheduled hours are expected to be used for reading papers and self study