Genome **Biology**

# Evolution of the protein domain repertoire of eukaryotes reveals strong functional patterns

Christian M Zmasek[*], Adam Godzik

## Background

Eukaryotes exhibit massive variation in their morphological complexity, ranging from protists to basal animals, such as *Trichoplax adhaerens* with only four different cell types, to mammals with ~ 210 different cell types [1]. Yet, the number of protein coding genes in eukaryotic genomes remains remarkably constant. Here, we attempt to shed some light on this paradox by analyzing eukaryotic genome evolution with a protein function centric view.
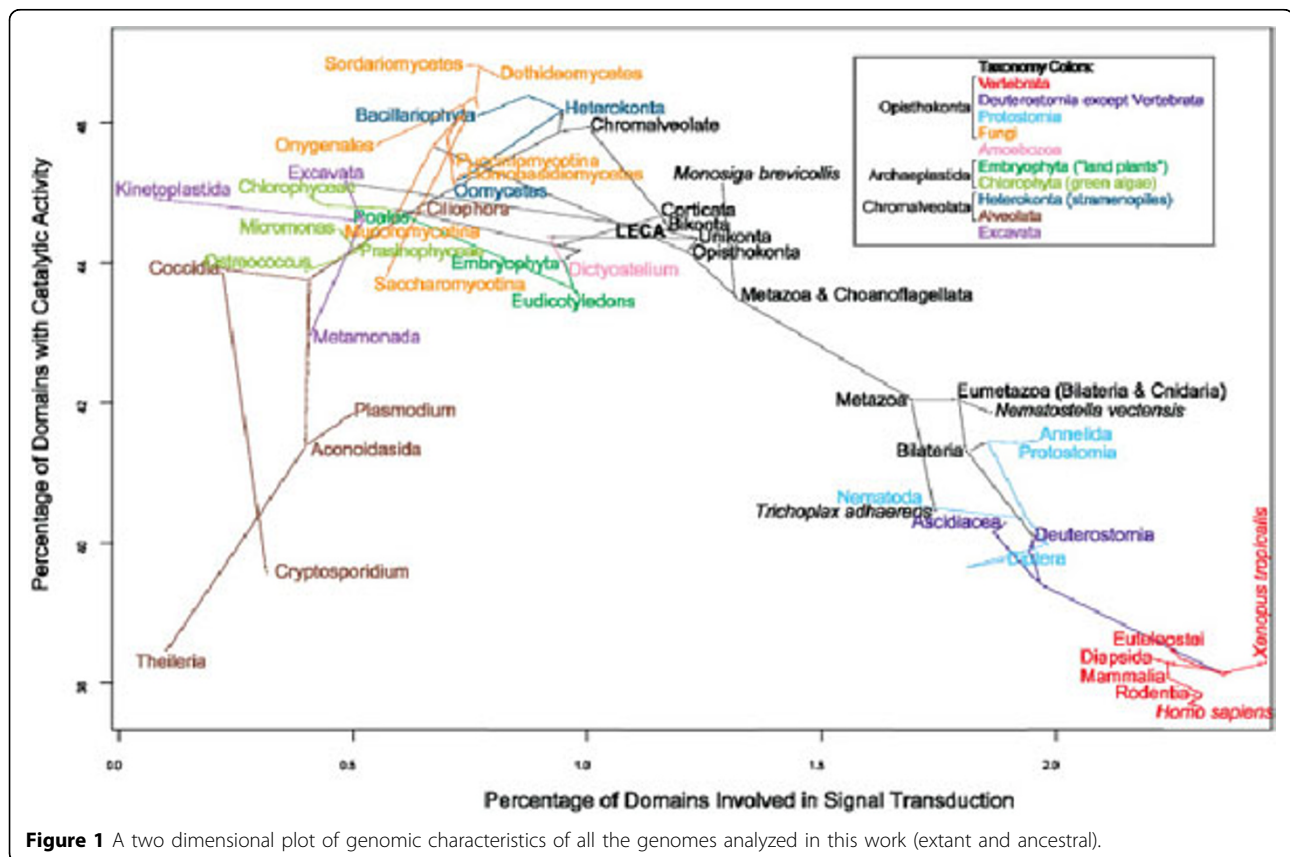


**Figure 1** A two dimensional plot of genomic characteristics of all the genomes analyzed in this work (extant and ancestral).

Bioinformatics and Systems Biology, Sanford-Burnham Medical Research Institute, La Jolla, CA 92037, USA
Full list of author information is available at the end of the article

## Materials and methods

Complete sets of predicted proteins for 114 eukaryotic genomes were analyzed with HMMER3 for Pfam domains. Ancestral domain sets were reconstructed using Dollo parsimony. Over- represented gene ontology (GO) terms were determined with Ontologizer 2.0.

## Results

*The evolution of most eukaryotes is dominated by domain losses.* By comparing ancestral and extant domainomes (the complete set of domains of a genome), we show that the evolution of eukaryotes is dominated by domain losses and not by domain gains. Two exceptions to this trend are the rise of multicellular animals and the origin of vertebrates. The last eukaryotic common ancestor (LECA) appears to have had an already large repertoire of ~ 4400 distinct domains. *During animal evolution, domains with regulatory functions are gained and domains with metabolic functions are lost.* By mapping Pfam domains to GO terms and applying enrichment analysis we are able to make inferences about the functional compositions of reconstructed domainomes and compare functional profiles of domain losses and gains, showing that functional profiles of these two differ drastically: animal evolution is associated with an gain of domains involved in various aspects of regulation, apoptosis, and immune system functions and loss of domains involved in amino acid biosynthesis and carbohydrate metabolism. *Classifying eukaryotes by the functional profiles of their genomes reproduces the tree of life.* By clustering according to their functional profiles of both extant and inferred ancestral genomes we are able to create a classification of eukaryotes which in large parts corresponds to the eukaryotic tree of life (Figure 1). *Gut microbes complement human reduced metabolic capacity.* Intriguingly, a 'meta-organism' composed of human protein domains and domains found in two common gut commensals, *Bacteroides thetaiotaomicron* and *Eubacterium rectale,* very closely resembles the LECA in its profile of metabolic domains. This potentially means that symbiotic microbes enabled animals (mammals, in particular) to lose large parts of their domains with metabolic functions.

### Reference
1. Carroll S: Chance and necessity: the evolution of morphological complexity and diversity. *Nature* 2001, **409**:1102-1109.