Gene family evolution of the chromalveolates based on complete genome sequences

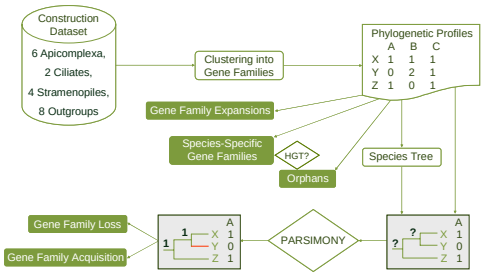
Cindy Martens[1,2], Klaas Vandepoele[1,2], and Yves Van de Peer[1,2]

[1] Department of Plant Systems Biology, VIB, Technologiepark 927, B-9052 Ghent, Belgium.
[2] Bioinformatics and Evolutionary Genomics, Department of Molecular Genetics, Technologiepark 927, Ghent University, B-9052 Ghent, Belgium.
E-mail: Cindy.Martens@psb.ugent.be

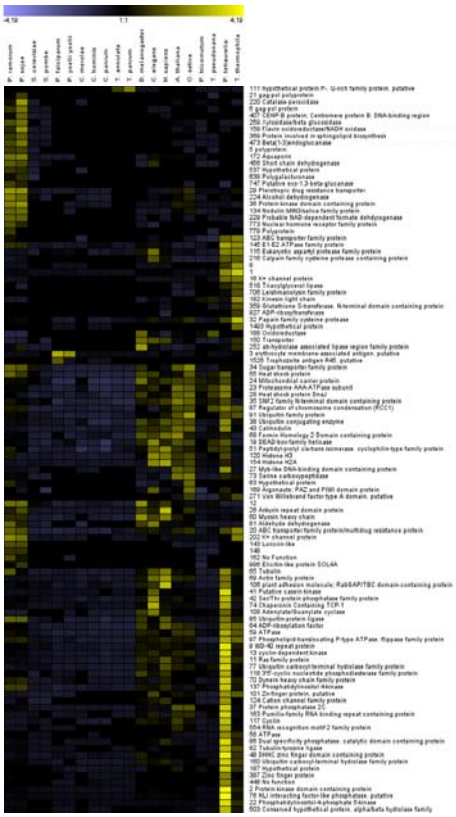
Introduction

The chromalveolate supergroup represents a large fraction of known eukaryotic diversity, ranging from tiny obligate intracellular parasites to free-living algae. Many of these protist species are of great medical (e.g. *Plasmodium* spp., causative agent of malaria), veterinary and agricultural importance because of their pathogenicity to man, cattle and crops. Others are of ecological interest such as the diatoms, which are responsible for approximately 40% of the marine primary production. With the genome sequence data of 12 chromalveolate genomes (6 within the Apicomplexa lineage, 2 ciliates and 4 within the Chromista), we have investigated the evolution of gene families by means of phylogenetic profiles. The aims of our study were threefold: (i) to study gene loss along the different chromalveolate lineages, (ii) to study gene gain and gene family expansions, and (iii) to investigate whether the previously mentioned evolutionary events can be correlated with the differences in lifestyle between the chromalveolate species.

Strategy and Objectives

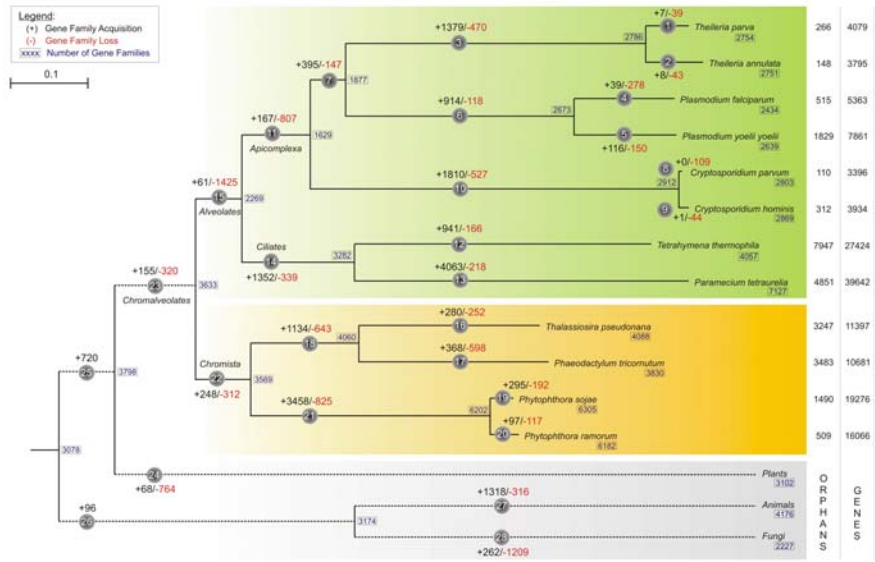


Gene Family Expansions



Species/Lineage-specific gene family expansions. The five percent most variable phylogenetic profiles, based on the standard deviations, were extracted. The matrix of these profiles was transformed into a matrix of z-scores to center and normalize the data. Subsequently, these profiles were hierarchically clustered. Yellow bars indicate expansions.

Gene Family Loss and Acquisition



A parsimonious scenario of gene loss and gene gain in chromalveolate evolution. At every time point (grey circles) in the chromalveolate tree (distance method, poisson correction, NJ, all nodes 100% BS-supported), gene loss (red) and gene acquisition (black) was counted based on the Dollo parsimony principle. The deduced ancestral gene sets are shown in boxes. Next to every species, the number of orphans and number of predicted genes is indicated.

Functional trends (Gene Ontology)

(The time points or TP are indicated in grey circles on the figure above)

GO-overrepresentation Gene Family Loss

TP	GO-label	GO-description	Q-value	Enrichment	# Gene Families
10	GO:0006099	tricarboxylic acid cycle	1,640E-06	9,5	12
10	GO:0008652	amino acid biosynthesis	2,384E-02	2,5	11
11	GO:0042384	cilium biogenesis	4,779E-04	15,5	6
11	GO:0008652	amino acid biosynthesis	9,501E-04	3,1	19
15	GO:0008652	amino acid biosynthesis	3,470E-17	4,9	58
18	GO:0042384	cilium biogenesis	1,826E-03	20,1	6
21	GO:0008652	amino acid biosynthesis	6,886E-03	2,6	17

GO-overrepresentation Gene Family Acquisition

TP	GO-label	GO-description	Q-value	Enrichment	# Gene Families
4	GO:0044406	adhesion to host	4,823E-02	52,3	1
6	GO:0009405	pathogenesis	3,420E-10	7,1	23
6	GO:0006508	proteolysis	0,0390997	1,7	32
7	GO:0009405	pathogenesis	1,86E-06	7,6	14
11	GO:0009405	pathogenesis	0,0009476	8,6	8

Conclusion

In this study, we were able to link the loss, acquisition and expansion of particular gene families in different chromalveolate species/lineages with their current lifestyles using a phylogenomics approach.