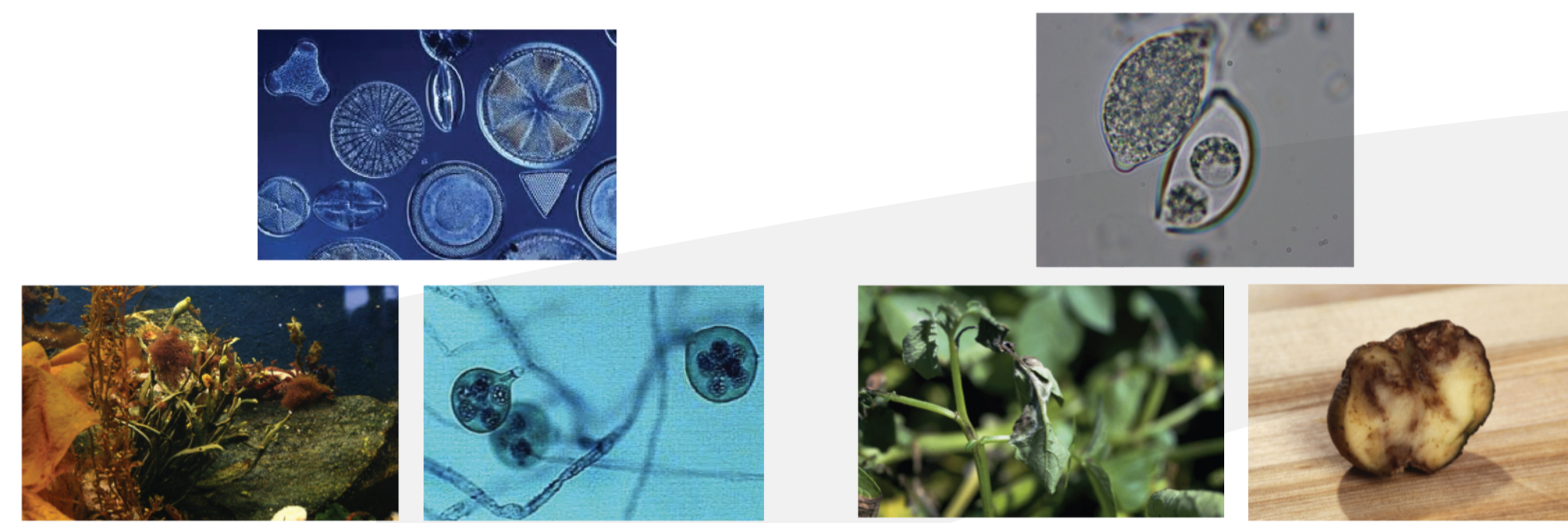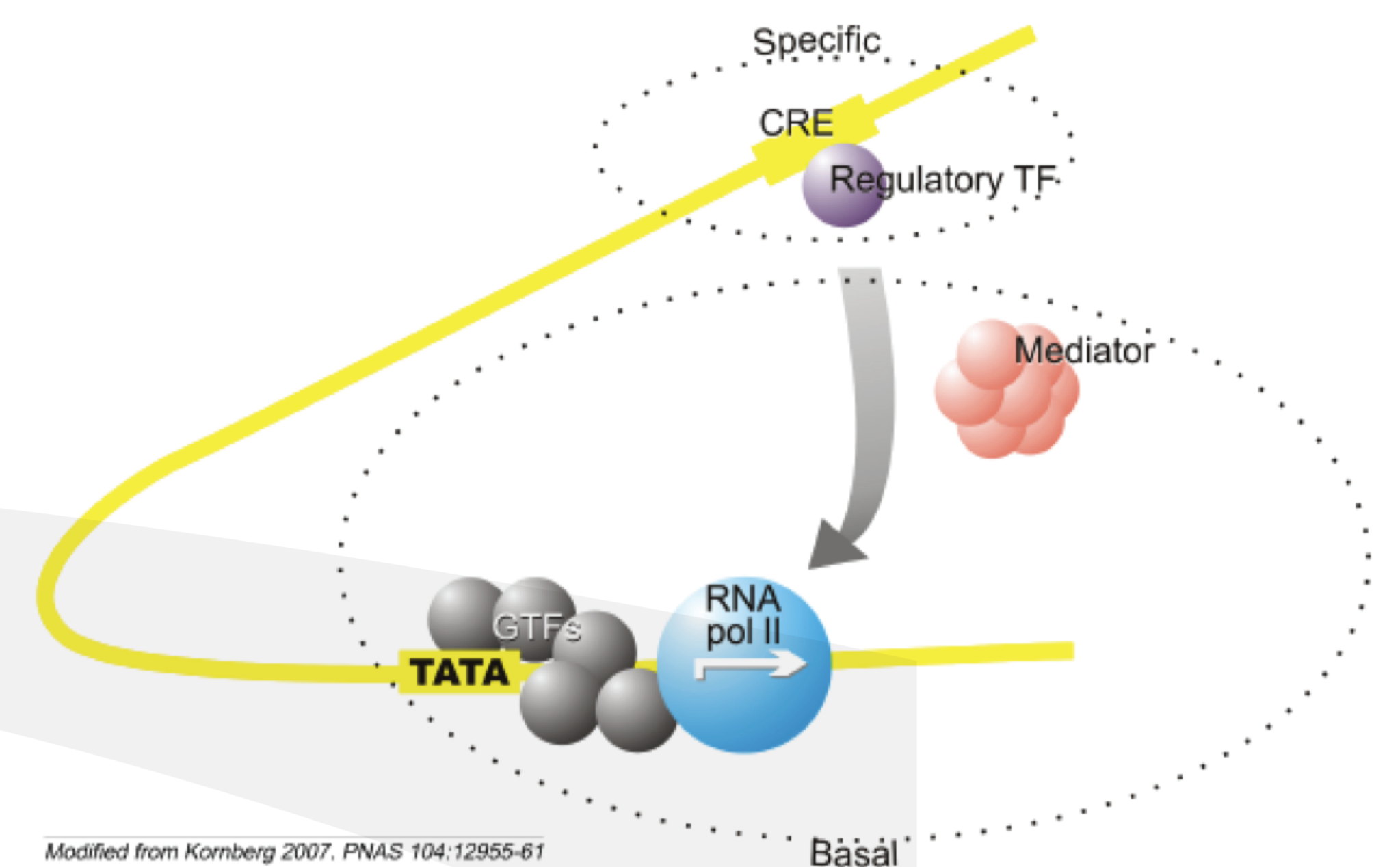# Identification of Transcription Factors and their Correlation with the High diversity in Stramenopiles

Francisco Buitrago-Florez, Silvia Restrepo, Diego M. Riaño-Pachón
E-Mail: fra-buit@uniandes.edu.co
URL: http://bce.uniandes.edu.co

High diversity among Stramenopiles

Stramenopiles is a Protist group that range from large multicellular seaweeds to tiny unicellular species; their members are present in freswater, marine, and terrestial habitats, and embrace many ecologycally important organisms (e.g, diatoms, brown algae) and many species of economically importance such as *Phytophthora infestans*, an important pathogen of potato, or *Saprolegnia parasitica*, a major fish pathogen.

Mechanisms of diversity?

Modified from Kornberg 2007. PNAS 104:12955-61

Transcription regulation model A) General transcription factors (GTFs) recruit RNA polymerase II in order to start the transcription of the gene, this type of transcription factors are comon among the genes transcribed by RNA-Pol II. B) Regulatory transcription factos (TFs) interact with the gene promoter to regulate gene transcription, this transcription factors are usually specific for each gene and have been found to be linked to the generation of diversity in organisms (Komberg, 2007). C) Other transcription regulators (TRs) as the mediator complex, participate in the process of transcription, but might not interact with DNA

A rules system for classification of TFs and TRs

■ Transcription factor family	● PFAM domain
▨ Other transcriptional regulator	◑ In house domain
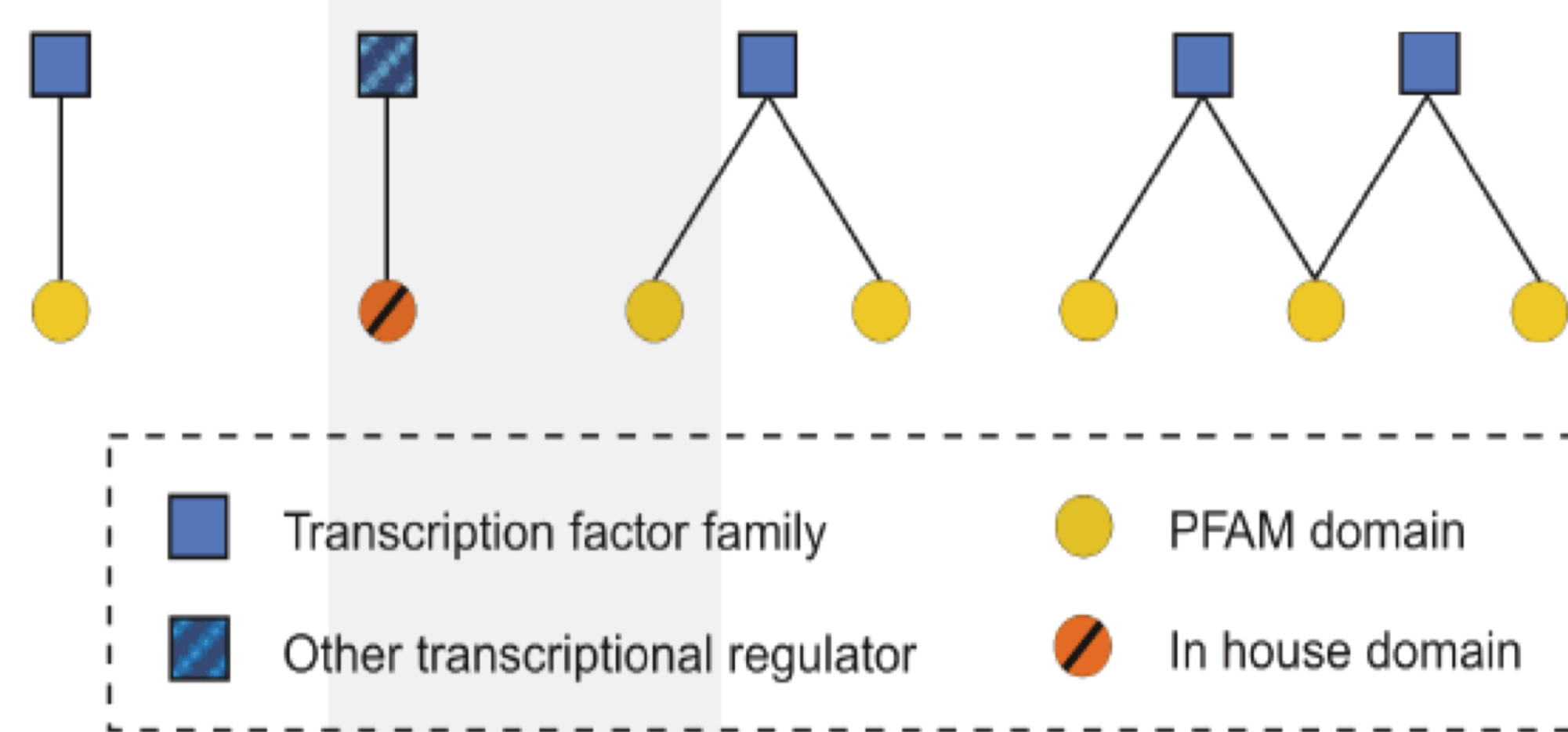
Modified from Riaño-Pachón et al. 2007. BMC Bioinformatics 8:42
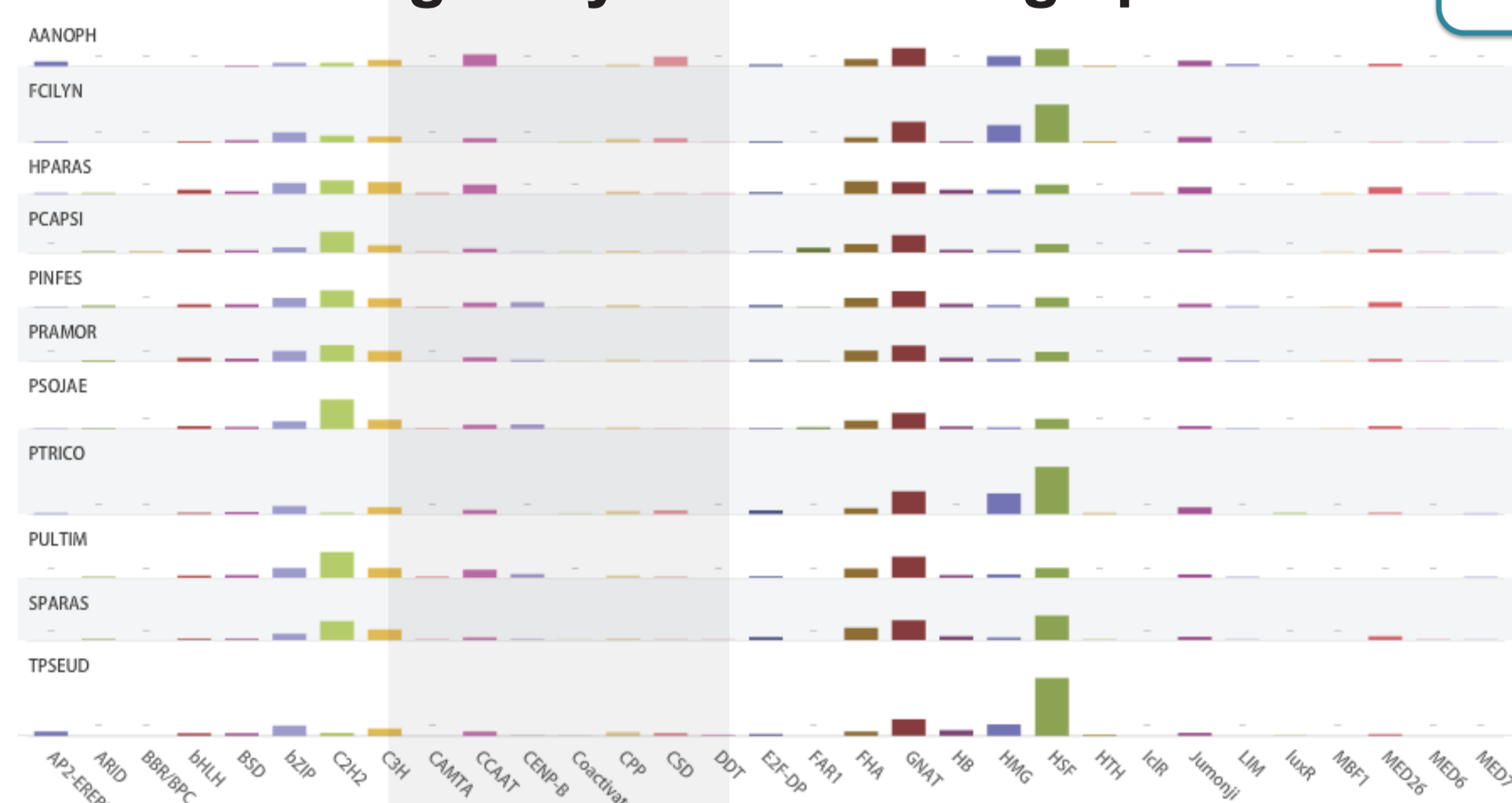
Rules for the classification of TFs and TRs depicted as a bipartite graph. Blue squares represent protein families, TFs are indicated in solid color and TRs are indicated by shaded squares.Yellow circles represent protein domains from the PFAM database (http://pfam. sanger.ac.uk/)orange circles represent domains generated in-house. In house models were generated based on seed alignments and profile HMMs were built using HMMER 3.0 (http://hmmer.janelia.org/)

Species included in this study

| Organism	Size	Genes	Proteins	Source
P. infestans	237 Mb	18179	18140	BROAD
P. ramorum	65 Mb	15743	16066	DOE-JGI
P. capsici	64 Mb	19805	15919	DOE-JGI
P. sojae	86 Mb	19027	19276	DOE-JGI
S. parasitica	53.09 Mb	20113	20088	BROAD
H. parasitica	82.05 Mb	14567	14565	BROAD
P. ultimum	44.91 Mb	15291	12614	BROAD
T. pseudonana	32 Mb	11397	11318	DOE-JGI
A. anophagefferens	56 Mb	11502	11501	DOE-JGI
F. cylindrus	80,5 Mb	22501	22501	DOE-JGI
P. Tricornutum	28 Mb	10489	10389	DOE-JGI

We use all the available genome sequences of Stramenopiles, there are large differences on genome size, number of genes and number of predicted proteins in this group.

Pfam V. 25 + In House Models → Proteomes → Search Hmmscan → Rules System → Assign Family → 38 TF and 25 TR Families

The Relative importance of TFs and TRs greatly varies among species

Columns represent selected families of TFs and TRs, rows represent the species included in this study and bars represent the relative importance: the proportion of a family over the total number of regulators per species. This analysis will highlight the importance of different TFs and TRs families. We can see that some families have greater representativity in some organisms. in some species however, there is homogeneity among families.

## Families are correlated with specific traits in Stramenopiles
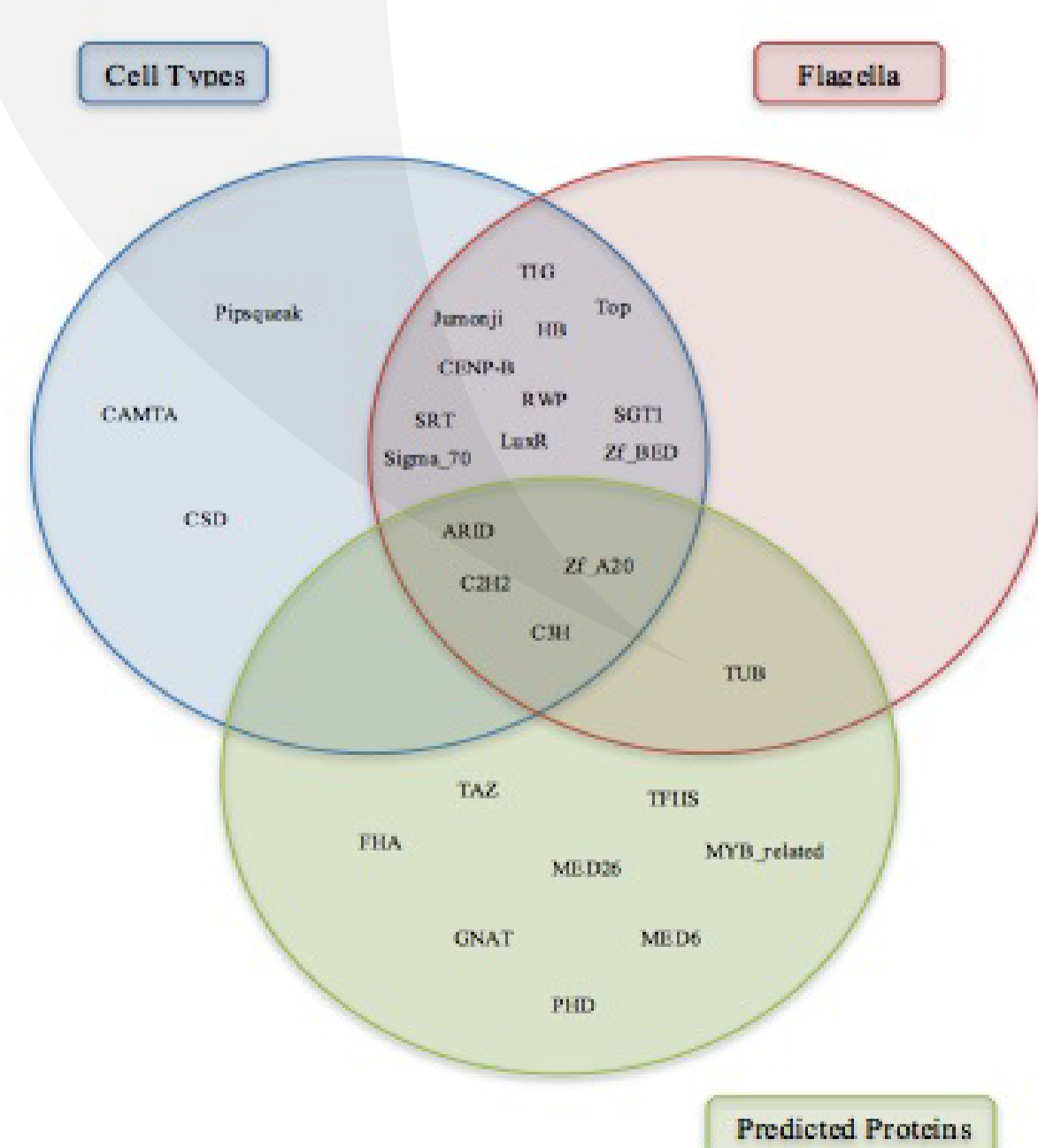
We performed a Phylogenetic Independent Contrast analysis between the size of the families and specific traits of Stramenopiles, this analisys shows a common set of families correlated with all traits, making them a candidate set responsible for the diversity of this group of organisms. A different subset of families is correlated with formation of flagella and cell types together. Moreover, there are several families specifically correlated with the number of predicted proteins and three families are specifically correlated with the number of cell types in the life cycle.
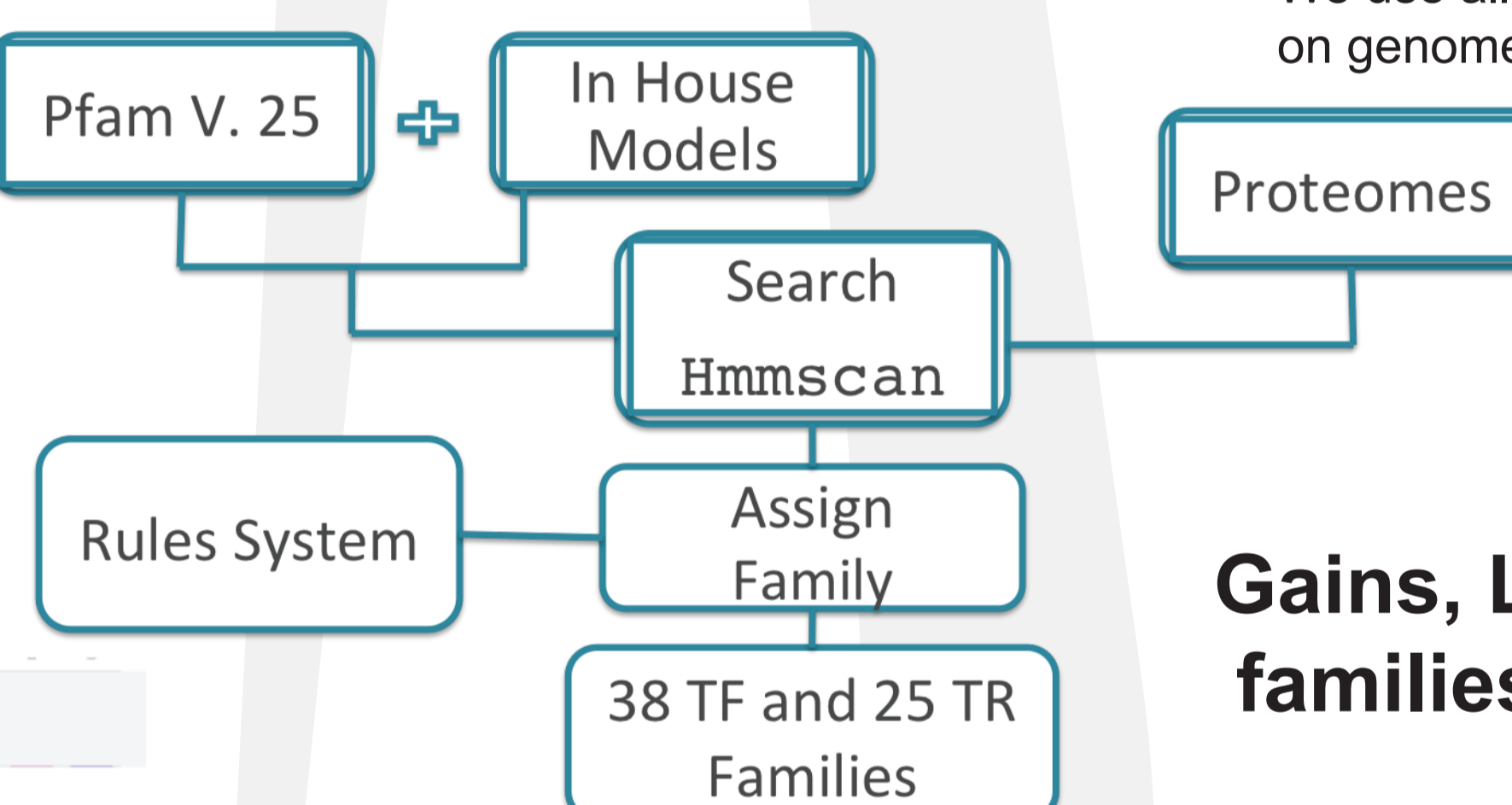
Gains, Losses and High abundance of TF and TR families traced in a phylogenetic reconstruction
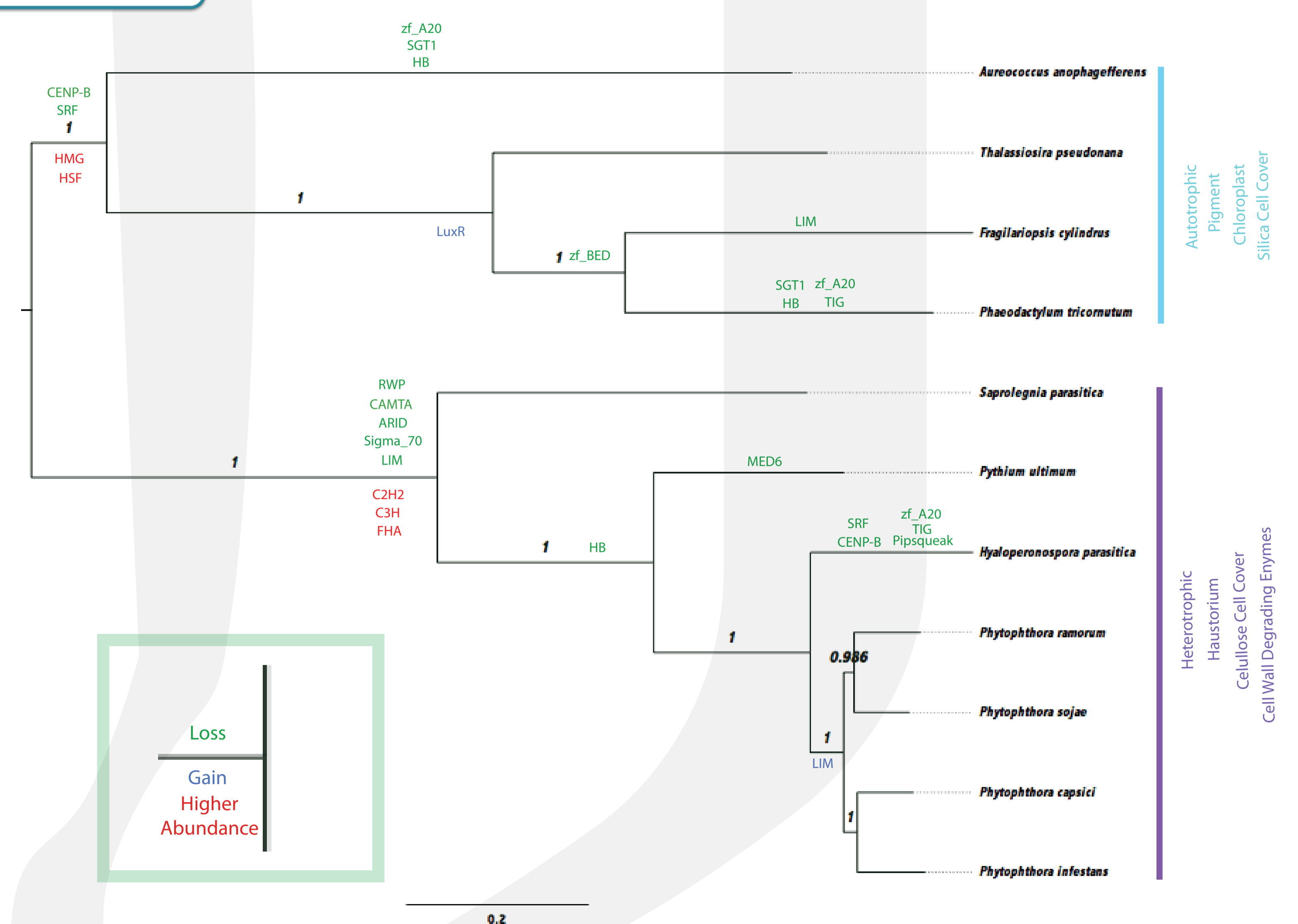
First, we recovered orthologous groups using OrthoMCL (http://www.orthomcl.org) finding168 gene clusters. The phylogenetic reconstruction was carried out with Maximum Likelihood analysis using RaxML and FastTree, both with 1000 bootstrap replicates. We found several families lost in specific species like the heterotrophic organism *Hyaloperonospora parasitica*, and the autotrophic organism *Phaeodactylum tricornutum*. These families need to be investigated in depth, in order to understand whether these established losses and gains have and impact in the diversity of specific groups. Moreover, we could suggest a gain of the LuxR family for some autotrophic species, most likely resulting from horizontal gene transfer between bacteria and Stramenopiles. Furthermore, we established high abundance of families taking into account the comparison between the two largest groups (autotrophic and heterotrophic organisms). We found high abundance of HMG and HSF families in the autotrophic organisms and C2H2, C3H and FHA families in the heterotrophic organisms, that could be implicated in the regulation of specific features of this group, making them of great interest for further investigations.

Bibliography

Komberg, R. 2007. The molecular basis of eukaryotic transcription. PNAS. 104:12955-12961.
Lang, D. et al., 2010. Genome-Wide phylogenetic comparative analysis of plant transcriptional regulation: A timeline of loss, gain, expansion, and correlation with complexity. Genome Biology and Evolution. 2:488-503.
Pérez-Rodríguez, P. et al., 2010. PlnTFDB: updated content and new features of the plant transcription factor database. Nucleic Acids Research. 38: Database issue doi:10.1093/nar/gkp805.
Riaño-Pachón, D. et al., 2007. PlnTFDB: an integrative plant transcription factor database. BCM Bioinformatics. 8:42.
Stamatakis, A. et al., 2005. RAxML-OMP: An efficient program for phylogenetic inference on SMPs. In proceedings of 8th international conference on parallel computing technologies, 38:288-302.