



# PlnTFDB Plant Transcription Factor Database Update

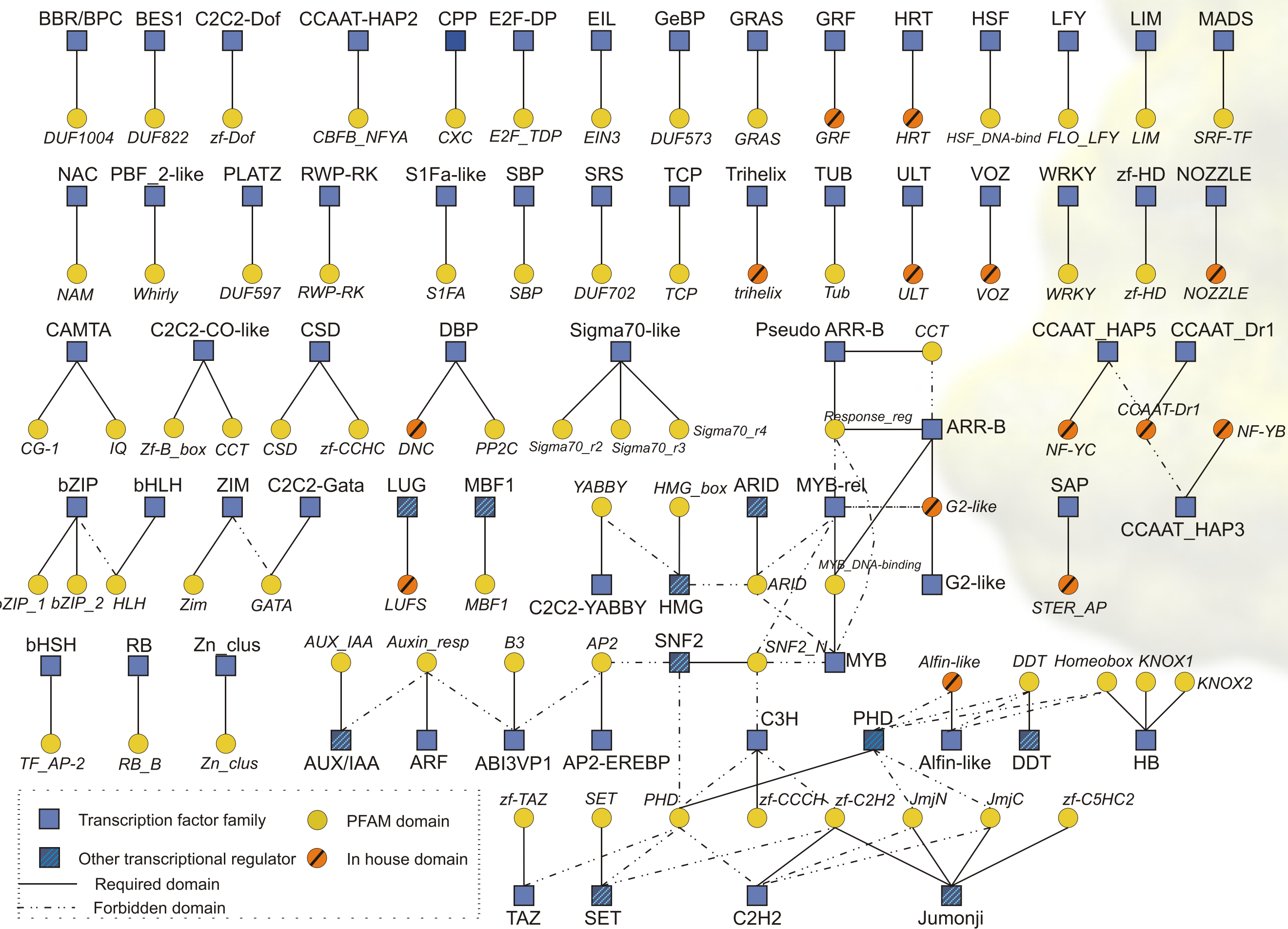


Riaño-Pachón DM<sup>1,2</sup>; Correa LGG<sup>1,2</sup>; Mueller-Roeber B<sup>1,2</sup>

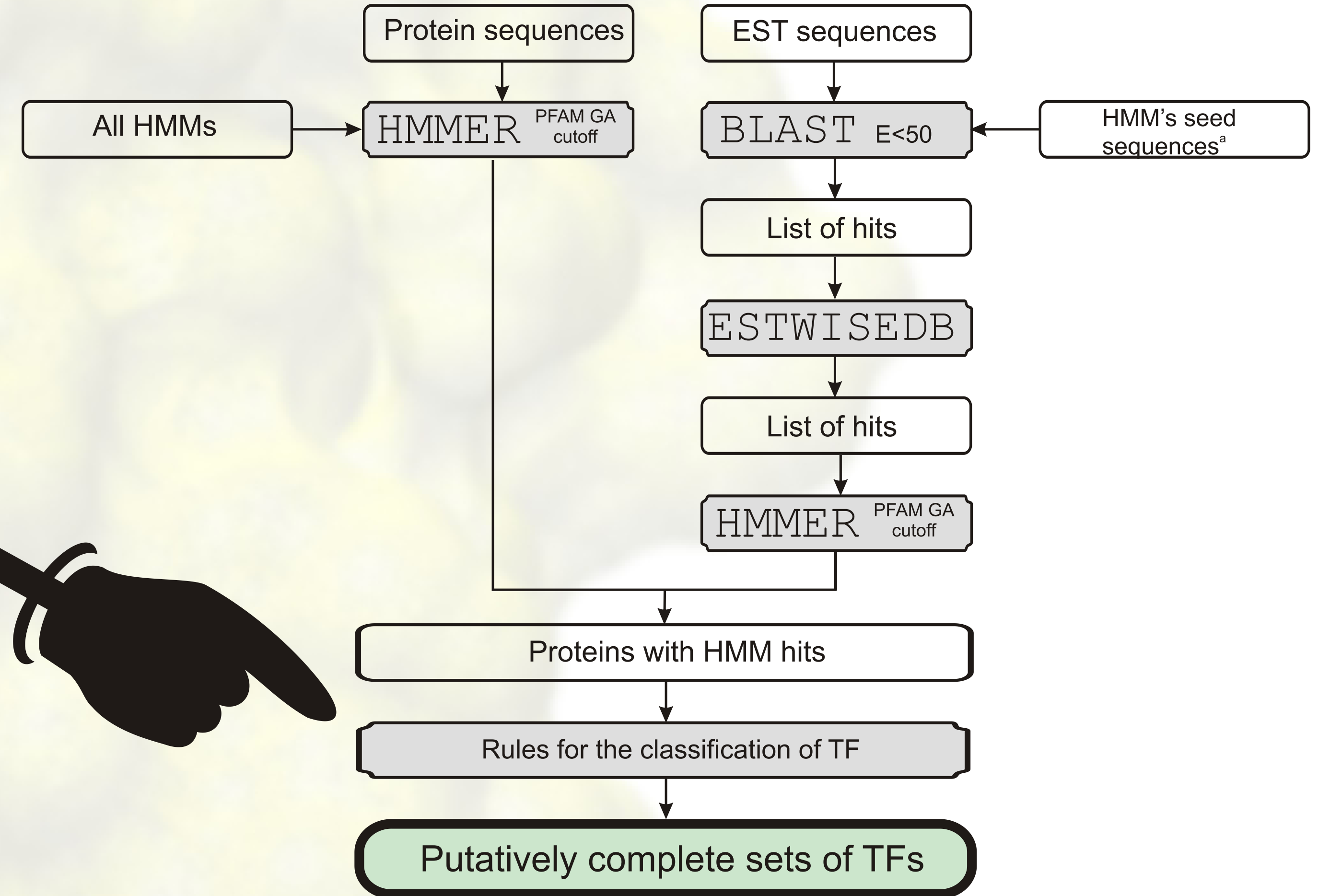
e-mail: {diriano, correa, bmr}@uni-potsdam.de

## 1 Rules for the Classification of Transcription Factors

## 2 Pipeline for the identification of Transcription Factors

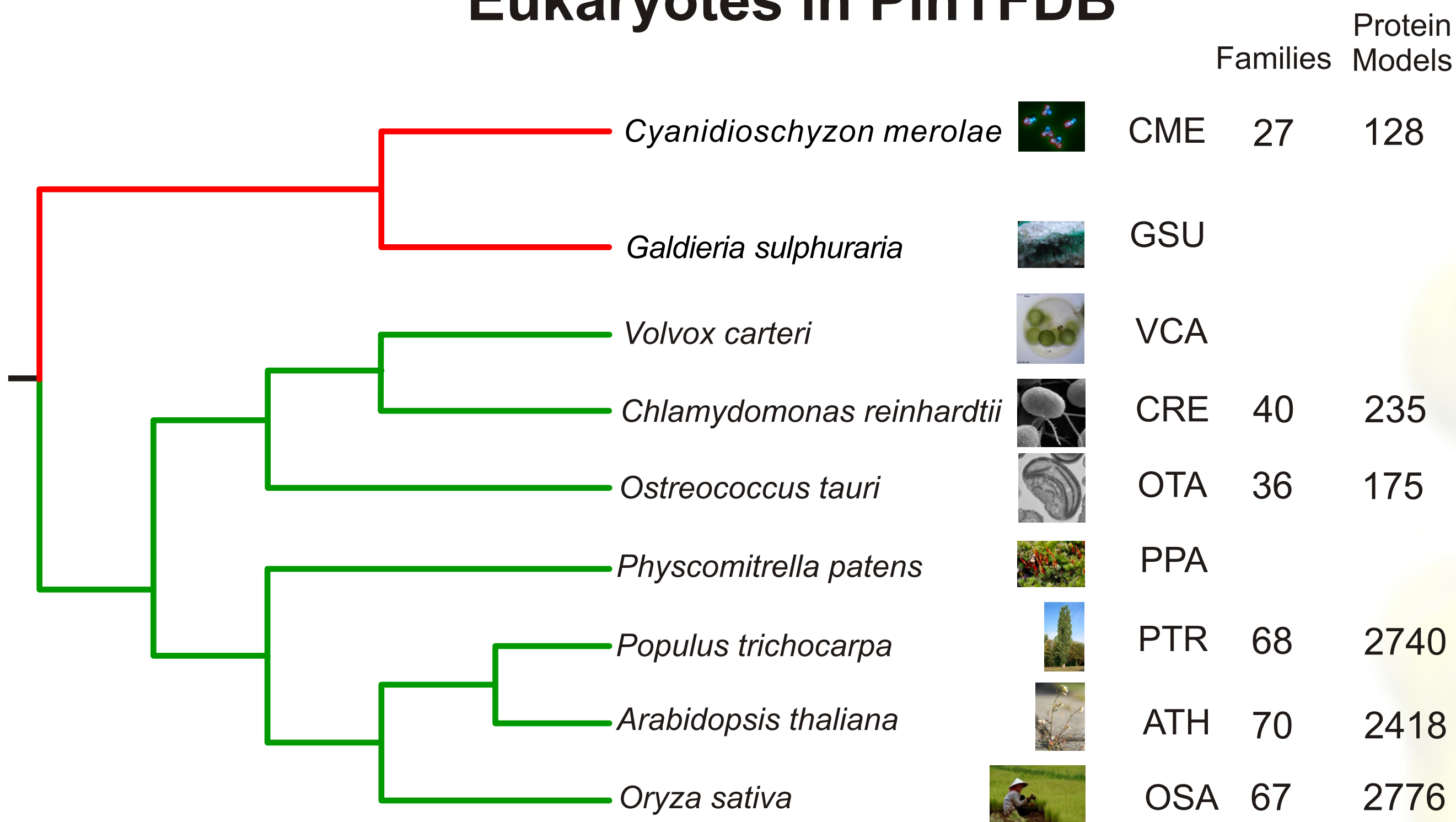


Blue squares represent families, TFs are indicated in solid color, other transcription regulators are indicated by shaded squares. Yellow circles represent protein domains from the PFAM database, orange circles represent domains generated in-house. Continuous edges appear when a domain must be present in members of the family. Discontinuous edges indicate that the domain must not appear in members of the family.



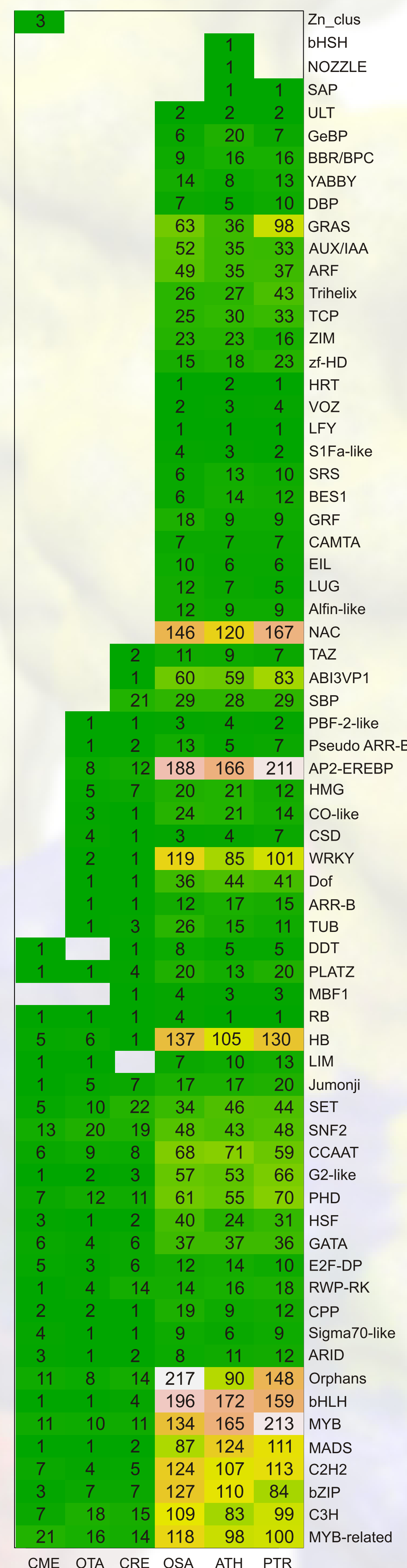
The pipeline for the identification of TFs was developed for completely sequenced genomes. We extended our approach in order to include those organisms for which only EST data is available. \*Only the seed sequences, from HMMs related or found in TFs were used.

## 3 Cladogram of Sequenced Photosynthetic Eukaryotes in PlnTFDB



The phylogenetic tree shows the evolutionary relationships among the photosynthetic eukaryotes currently available in PlnTFDB. Number of families and protein models are shown. In addition to the shown species, PlnTFDB includes *Hordeum vulgare* (barley) with TFs predicted out of ESTs. Soon further organisms will be added. There is a clear to trend to accumulate and diversify the set of TFs as new habitats are colonized and new functions emerge.

## 5 Number of TFs per Family per Species



## 6 Web Site



Through the web site we have developed, users have access to all protein, transcript and locus sequences for each of the predicted transcription factors. In addition, the domain structure of every sequence is available, together with putative orthologs found by INPARANOID in all species included in PlnTFDB.

## 4 Pairwise Number of Clusters of Orthologs

2nd Species	1st Species					
	CME	CRE	OTA	PTR	ATH	OSA
CME		45:47	43:47	77:64	76:63	99:64
CRE	53		78:82	152:99	156:99	186:103
OTA	48	89		151:107	181:111	178:110
PTR	70	123	134		1202:1011	1041:833
ATH	71	123	139	993		994:928
OSA	72	139	130	783	762	

Orthologs were determined in a pairwise manner using INPARANOID, which is an implementation of the best bidirectional hit approach using BLAST. The number of clusters of orthologs is shown below the diagonal. Above the diagonal the number of orthologous protein models in each species appear, first for the species in the column and then for the species in the row.

## 7 Coming soon

The following species will soon be added to PlnTFDB

- the diatom *Thalassiosira pseudonana*
- the fern *Selaginella moellendorffii*
- maize (*Zea mays*)
- grapevine, cv Pinot noir (*Vitis vinifera*)

The table in the left shows the phylogenetic profile of TF families in photosynthetic eukaryotes, showing also the number of protein models found of each family for each species.

<http://plntfdb.bio.uni-potsdam.de/v2.0/>