SUPPLEMENTARY MATERIALS

Sorting TFs into familes in GrassTFDB

The PlnTFDB has 68 families and AtTFDB has 50, there are TF families in PlnTFDB that contain protein domains that are sometimes present in transcription factors, but are associated for other types of transcriptional regulators. CCAAT family members in PlnTFDB were distributed among four CCAAT subfamilies as described in PlantTFDB and AtTFDB. The families that are present in PlnTFDB and PlantTFDB but excluded from GrassTFDB are MBF1, FHA, AUX/IAA, HMG, LIM, LUG, PLATZ, S1Fa-like, SET, SRS, TAZ, ULT, NOZZLE, bHSH, CSD, DBP, DDT, Pseudo ARR-B, RB, SAP, Sigma70-like, SNF2 and ZIM. The rules for family classification were extended for proteins that posses more than one domain, as identified by HMMPFAM, which characterize different families, as proteins with Homeobox and bZIP domains that was classified as Homeobox, the criteria do define de family was using the current literature and the highest e-value in HMMPFAM. We developed and applied a PERL script to set the established rules for the identification and classification of all TFs for the GRASSTFDB (Summarized in Supplementary Table I).

SUPPLEMENTAL FIGURE LEGENDS

Supplementary Figure 1. Interactive visualization of grass TF phylogenetic trees by ATV. The sample tree corresponds to SBP family members in GRASSIUS, the color coding indicates species origin of the TFs. ATV provides various options to investigate information about a node. The branches can be collapsed, re-rooted, viewed as subtrees and linked to the corresponding pages in GRASSIUS.

Supplementary Figure 2. Full-length transcription factor expression plasmid (TFome) collection. Data on each TFome clone consists of details of the corresponding PCR reaction used to generate the clone, its nucleotide sequence and corresponding translation into amino acids, the name of the matching TF in GrassTFDB and contact information for requesting clones or additional information.

Supplementary Figure 3. Detailed information of gene upstream regions. GrassPROMD utilizes BioPerl to construct information tracks for an upstream region of a gene, focusing on experimentally verified *cis*-regulatory elements (CRE). For a single CRE, the following information is currently visually available: The location of the CRE and other features relative to the transcription start site, the upstream region sequence, the name of the TF(s) targeting particular CREs, when known. The box representing the CRE provides a hyperlink to a page containing additional details and the TF name provides a link to the respective TF in GrassTFDB.

Supplementary Figure 4. GRASSIUS is open to community contribution. Researchers are welcome to contribute to GRASSIUS content by filling out the TF submission form. Another form is available under the "Community Contribution" page to submit promoter data as well.

Supplementary Figure 5. GrassPROMDB database structure. The MySQL tables contain data for both the promoter and the gene driven by that promoter. Location and

literature data for individual CRE is linked to promoter data.   Currently under development are features that permit to identify promoters with respect to the biological pathways in which a particular gene is involved. In addition, information on gene expression data will become soon available.